

**LIPIDOMIC ANALYSIS OF EXTRACELLULAR VESICLES
AND ITS POTENTIAL FOR CANCER DIAGNOSIS**

Erika Dorado Montezuma

Imperial College London

Department of Metabolism, Digestion and Reproduction

Thesis submitted for the degree of Doctor of Philosophy (PhD)

February 2021

Declaration of Originality

I certify that this thesis and the research reported herein was the product of my own work. Information derived from published research and any contribution from the work of other people has been fully acknowledged in the text and the references are given in accordance with standard referencing practices of this academic discipline.

Erika Dorado (2021)

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial-No Derivatives 4.0 International Licence (CC BY-NC-ND). Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work. When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

To my family, mentors, supervisors and friends

Acknowledgments

I would like to thank the STRATiGRAD PhD programme for this PhD opportunity, funding and trainings. Especially, Merck KGaA, Darmstadt, Germany, for its contribution to this programme.

I would like to thank my main supervisor Prof. Zoltan Takats and co-supervisors Prof. Molly Stevens and Prof. Jeremy Nicholson for their valuable feedback and guidance during my PhD.

I would like to thank all patients and volunteers that participated in this PhD thesis and kindly supported cancer research. I would like to thank Prof. Charles Coombes and his team for their collaboration in the sample collection from cancer patients and healthy volunteers.

I would like to thank Dr. Luisa Doria for her guidance with respect to lipidomics analysis and for her friendship. I would like also to thank Dr. Anika Nagalkerke for the valuable discussions about extracellular vesicles analysis.

I would like to thank my colleagues at Imperial for their exciting scientific discussions but most importantly for their friendship during my PhD journey.

I would like to dedicate and warmly thank my family for their support through my scientific career. Especially, I would like to thank my husband for his support and listening to my endless conversations about extracellular vesicles and cancer.

Abstract

Breast and ovarian cancer are among the deadliest types of cancer in women. The development of diagnostic methods for early diagnosis of cancer can improve patient survival rates. Cancer cells produce extracellular vesicles (EVs), including exosomes and microvesicles, which can be found in cancer patients' body fluids. Cancer-derived EVs play an important role in cancer development, progression and metastasis. In addition, they have a great potential as a source of biomarkers for the development of body fluid-based methods for cancer diagnosis. The aim of this PhD thesis was to study the lipid composition of cancer-derived EVs and its potential for the identification of lipid biomarkers for breast and ovarian cancer diagnosis.

The lipid composition of EVs secreted by cancerous and non-cancerous mammary and ovarian epithelial cells was studied by untargeted lipidomics. Cancer-derived EVs were found to be enriched in sphingolipids and glycerophospholipids compared to their parental cells. It was found that EVs produced by cancerous cells can be distinguished from those produced by non-cancerous cells based on their lipid content. Furthermore, EVs' lipid composition differs between breast cancer subtypes, as well as between ovarian cancer subtypes. EVs found in blood plasma from breast cancer patients and healthy individuals were also analysed by untargeted lipidomics and a combination of five specific lipid species was found to have a diagnostic value for breast cancer diagnosis. Importantly, the same combination of lipids was also able to differentiate between EVs found in blood plasma from primary breast cancer patients and EVs found in blood plasma from healthy individuals.

In conclusion, this study demonstrates the potential of the EVs' lipid content for the identification of lipid biomarkers for the early detection of breast and ovarian cancer and their use for the development of diagnostic methods, based on blood plasma analysis of cancer patients.

Contents

List of figures.....	11
List of tables.....	13
List of abbreviations.....	14
1. Background.....	19
1.1.Cancer.....	20
1.1.1. Breast cancer.....	23
1.1.2. Ovarian cancer.....	27
1.2.Lipids.....	29
1.2.1. Lipids metabolism and cancer.....	34
1.3.Extracellular vesicles.....	35
1.3.1. EVs' isolation methods.....	37
1.3.2. EVs' lipid composition.....	40
1.3.3. EVs and cancer.....	41
1.4.Lipidomics and Mass spectrometry.....	42
1.4.1. Mass spectrometer.....	43
1.4.2. Liquid chromatography mass spectrometry and lipidomics.....	47
1.4.3. Lipidomics and cancer.....	49
1.5.Bioinformatics and statistical analysis	51
2. Hypotheses and aims.....	54
3. Lipidomic analysis of mammary epithelial cells and their secreted EVs....	58
3.1.Background.....	59
3.2.Methods.....	60
3.2.1. Mammary cells evaluated.....	60
3.2.2. EVs' isolation.....	60
3.2.3. EVs' characterisation.....	61
3.2.4. Lipid extraction.....	61

3.2.5. LC-MS analysis.....	62
3.2.6. Data analysis.....	63
3.3.Results and Discussion.....	64
4. Lipidomic analysis of ovarian epithelial cells and their secreted EVs.....	83
4.1.Background.....	84
4.2.Methods.....	85
4.2.1. Cell culture of ovarian epithelial cells.....	85
4.2.2. EVs' isolation and characterisation.....	85
4.2.3. TEM.....	85
4.2.4. Lipid extraction.....	86
4.2.5. LC-MS analysis.....	86
4.2.6. Data analysis.....	87
4.3.Results and Discussion.....	87
5. Isolation of EVs from human blood plasma.....	106
5.1.Background.....	107
5.2.Methods.....	108
5.2.1. Sample collection.....	108
5.2.2. Sample preparation.....	108
5.2.3. Density gradient ultracentrifugation.....	109
5.2.4. Density gradient measurement.....	109
5.2.5. Size exclusion and bind-elute chromatography.....	109
5.2.6. EVs' characterisation.....	110
5.3.Results and Discussion.....	110
6. Lipidomic analysis of EVs found in blood plasma from breast cancer patients and healthy individuals.....	117
6.1.Background.....	118
6.2.Methods.....	119
6.2.1. Sample collection.....	119
6.2.2. Isolation of EVs from blood plasma.....	119

6.2.3. EVs' characterisation.....	120
6.2.4. Lipid extraction.....	120
6.2.5. LC-MS and data analysis.....	120
6.3. Results and Discussion.....	120
7. General discussion and conclusions.....	132
8. Bibliography.....	138
Appendix I.....	158
Appendix II.....	162

List of Figures

Figure 1.1.1. Main cancer characteristics.....	21
Figure 1.1.2. Cancer invasion and metastasis multistep process.....	22
Figure 1.1.3. Histological grade of breast cancer.....	25
Figure 1.1.4. Ovarian epithelial carcinomas subtypes and genetic aberrations associated to them.....	28
Figure 1.2.1. Major glycerophospholipids.....	31
Figure 1.2.2. Sphingosine and major sphingolipids.....	31
Figure 1.2.3. Lipid synthesis and composition of cell membranes.....	33
Figure 1.3.1. Extracellular vesicles biogenesis.....	36
Figure 1.4.1. Mass spectrometer main components.....	43
Figure 1.4.2. Ion formation in ESI.....	46
Figure 1.4.3. Lipids separation by reverse phase LC-MS.....	48
Figure 3.3.1. Workflow for the lipidomic analysis of mammary cells and their secreted EVs.....	65
Figure 3.3.2. Lipids enriched in EVs when compared to their parental cells.....	66
Figure 3.3.3. Lipid species highly correlated between EVs and cells.....	74
Figure 3.3.4. Statistical analysis of cancerous and non-cancerous EVs and cells...	78
Figure 3.3.5. Statistical analysis of cells and EVs obtained from different breast cancer subtypes.....	81
Figure 4.3.1. Workflow for the lipidomic analysis of the ovarian epithelial cancerous and non- cancerous epithelial cells and their secreted EVs.....	88
Figure 4.3.2. Ovarian cancer-derived EVs characterisation.....	89
Figure 4.3.3. Lipids enriched in EVs when compared to their parental cells.....	92
Figure 4.3.4. Lipid species highly correlated between EVs and cells.....	98
Figure 4.3.5. Statistical analysis of cancerous and non-cancerous EVs and cells..	102
Figure 4.3.6. Statistical analysis of cells and EVs obtained from endometrioid and high grade serous ovarian cancer subtypes	104
Figure 5.3.1. Workflow of the methodology used for isolation of EVs from human blood plasma.	111

Figure 5.3.2. Detection of EV protein markers in the fractions obtained by density gradient ultracentrifugation of human blood plasma.....	112
Figure 5.3.3. Characterisation of EVs isolated from human blood plasma by combination of density gradient ultracentrifugation and size exclusion/bind-elute chromatography.....	114
Figure 6.3.1. Workflow of the lipidomic analysis of EVs found in breast cancer patients and healthy volunteers.....	121
Figure 6.3.2. Protein characterisation and number of EVs isolated from blood plasma from breast cancer patients and healthy volunteers.....	123
Figure 6.3.3. Statistical analysis of the LC-MS data for EVs found in blood plasma from breast cancer patients and healthy volunteers.....	126
Figure 6.3.4. Statistical analysis of the LC-MS data for EVs found in blood plasma from primary breast cancer patients and healthy volunteers.....	131
Figure 7.1. Illustration of the lipids identified in EVs.....	134
Supplementary Figure 1. Characterisation of EVs isolated from mammary epithelial cells.....	159
Supplementary Figure 2. PCA of the LC-MS data acquired for the human mammary epithelial cells and their EVs.....	159
Supplementary Figure 3. PCA of the LC-MS data acquired for the human ovarian epithelial cells and their EVs	160

List of tables

Table 3.3.1. List of lipids species enriched in EVs when compared to their parental cells.....	67
Table 3.3.2. Lipids enriched in cells when compared to their secreted EVs.....	69
Table 3.3.3. Lipids with very strong correlation between EVs and their parental cells.....	74
Table 4.3.1. List of lipids species enriched in EVs when compared to their parental cells.....	93
Table 4.3.2. Lipids enriched in cells when compared to their secreted EVs.....	94
Table 4.3.3. Lipids with a very strong correlation between EVs and their parental cells.....	98
Table 6.3.1. Pathological characteristics of breast cancer samples evaluated.....	122
Supplementary Table 1. List of lipid standards analysed to test the quality of the LC-MS data and mass spectrometers.....	160

List of abbreviations

AA	Arachidonic Acid
APOA1	Apolipoprotein A1
ACN	Acetonitrile
AKT	Protein Kinase B
ANOVA	Analysis Of Variance
ARID1A	AT-Rich Interactive Domain-Containing Protein 1
AUROC	Area Under The Receiver Operating Characteristic
BMP	Bis(Monoacylglycero)Phosphate
BRAF	B-Raf
BRCA1	Breast Cancer 1
BRCA2	Breast Cancer 2
CA125	Cancer Antigen 125
CASP8	Caspase 8
CE	Cholesterol Esters
CER	Ceramide
CI	Chemical Ionisation
CL	Cardiolipins
C1P	Ceramide-1-Phosphate
DDA	Data Dependent Acquisition
DESI	Desorption Electrospray Ionisation
DCIS	Ductal Carcinoma In Situ
DG	Diacylglycerol
FA	Fatty Acid
EDTA	Ethylenediaminetetraacetic Acid
EI	Electron Ionisation
ER	Oestrogen Receptor
ER+	ER-Positive
ESI	Electrospray Ionisation
ESI-	Negative Electrospray Ionisation
ESI+	Positive Electrospray Ionisation

EVS	Extracellular Vesicles
FGFR2	Fibroblast Growth Factor Receptor 2
FT-ICR	Fourier Transform-Ion Cyclotron Resonance
HDL	High-Density Lipoproteins
HER2	Human Epidermal Growth Factor Receptor 2
HER2+	Human Epidermal Growth Factor Receptor 2 Positive
HEXCER	Hexosylceramides
HPLC	High-Performance Liquid Chromatography
IDC	Invasive Ductal Carcinoma
IDL	Intermediate-Density Lipoproteins
ILC	Invasive Lobular Carcinoma
IPA	Isopropanol
IT	Ion Traps
KRAS	K-Ras
LACCER	Lactosylceramide
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass Spectrometry
LCIS	Lobular Carcinoma In Situ
LDL	Low-Density Lipoproteins
LPA	Lysophosphatidic Acid
LPC	Lysophosphatidylcholine
LPE	Lysophosphatidylethanolamine
LPG	Lysophosphatidylglycerol
LPI	Lysophosphatidylinositol
LPS	Lysophosphatidylserine
LSP1	Lymphocyte-Specific Prote
LR	Logistic Regression
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MAP3K1	Mitogen-Activated Kinase Kinase Kinase 1
MTOR	Mammalian Target Of Rapamycin
MS	Mass Spectrometry

MSC	Mesenchymal Stem Cells
MSI	Mass Spectrometry Imaging
MS/MS	Tandem Mass Spectrometry
NTA	Nanoparticle Tracking Anal
OPLS-DA	Orthogonal Partial Least-Squares Discriminant Analysis
PA	Phosphatidic Acid
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PEG	Polyethylene Glycol
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PI3K	Phosphatidylinositol 3-Kinase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase, Catalytic Subunit Alpha
PIP3	Phosphatidylinositol-3,4,5-Trisphosphate
PLS	Partial Least-Squares
PR	Progesterone Receptor
PR+	Progesterone Receptor Positive
PS	Phosphatidylserine
PCA	Principal Component Analysis
PLS-DA	Partial Least-Squares Discriminant Analysis
PQN	Probabilistic Quotient Normalisation
Q	Quadrupole
QC	Quality Control
RAB-5B	Ras-Related Protein Rab-5B
REIMS	Rapid Evaporative Ionization Mass Spectrometry
RFE	Recursive Feature Elimination
S1P	Sphingosine-1-Phosphate
SEC	Size Exclusion Chromatography
SIMS	Secondary Ion Mass Spectrometry
SPC	Sphingosylphosphorylcholine

TEM	Transmission Electron Microscopy
TG	Triacylglycerol
TN	Triple Negative
TNRC9	Thymocyte Selection–Associated High Mobility Group Box 9
TOF	Time-Of-Flight
TP53	Tumour Protein 53
VLDL	Very-Low-Density Lipoproteins

Chapter 1

Background

1. Background

1.1. Cancer

More than two hundred different diseases are identified as cancer. This is mainly due to the fact that cells in any part of the body have the potential to become cancerous cells. Cancer is one of the deadliest diseases in the world after cardiovascular diseases. In 2017, the odds of people developing cancer were 1 in 3 for men and 1 in 4 for women. Moreover, cancer was responsible for 9.6 million deaths worldwide and 24.5 million new cases were reported. It is estimated that the number of cases will increase in the coming decades [1].

Most cancers are classified into three main groups, depending on the type of cell from which the disease arises; carcinomas, sarcomas, and lymphomas or leukemias. Carcinomas are malignancies of epithelial cells, and they are the most common human cancer. Sarcomas are cancers of the mesenchymal tissue; this includes bone, cartilage, muscles and other connective tissues [2]. Although they are rare in humans, sarcomas can occur practically everywhere in the body [2]. Lymphomas or leukemias are blood-cell cancers. Lymphomas arise from cells of the immune system affecting the lymphatic system, while leukemias arise from blood-forming cells, affecting the blood and bone marrow [3, 4].

Cancer can result from the chronic division and growth of cells as a consequence of the accumulation of mutations in the genes which control cell proliferation and maintain genome stability, including DNA repair genes. Cancer cells evolve strategies to coordinate their signalling network circuitry to successfully regulate cell proliferation while evading essential cell processes such as apoptosis. Moreover, they can reprogramme their cellular metabolism and evade the immune system. Genome instability is one of the factors leading to the development of cancer [5], but epigenetic aberrations and environmental factors can also contribute to cancer, or even an interaction of all these different factors [6] (Figure 1.1.1).

Cancer is characterised by: 1. Deregulation of signalling networks that promote cell growth; 2. Evasion of growth suppressors; 3. Avoidance of cell death mechanisms, 4. Unlimited replicative potential; 5. Induction of angiogenesis; 6. Tissue invasion and metastasis capabilities; 7. Resisting immune destruction and; 8. Deregulation of energy metabolism [5] (Figure 1.1.1).

Cancer cells adjust their energy metabolism to respond to the demand of the energy production required for the continuing cell growth and division. Energy metabolism in cancer cells is switched to glycolysis regardless of oxygen availability, process termed aerobic glycolysis or the Warburg effect [5, 7]. The increased rate of glucose consumption has already been documented in many human cancers [8]. This glycolytic switch is associated with mutations in oncogenes and tumour suppressors as well as up-regulation of the phosphatidylinositol 3-kinase (PI3K) signalling pathway [9].

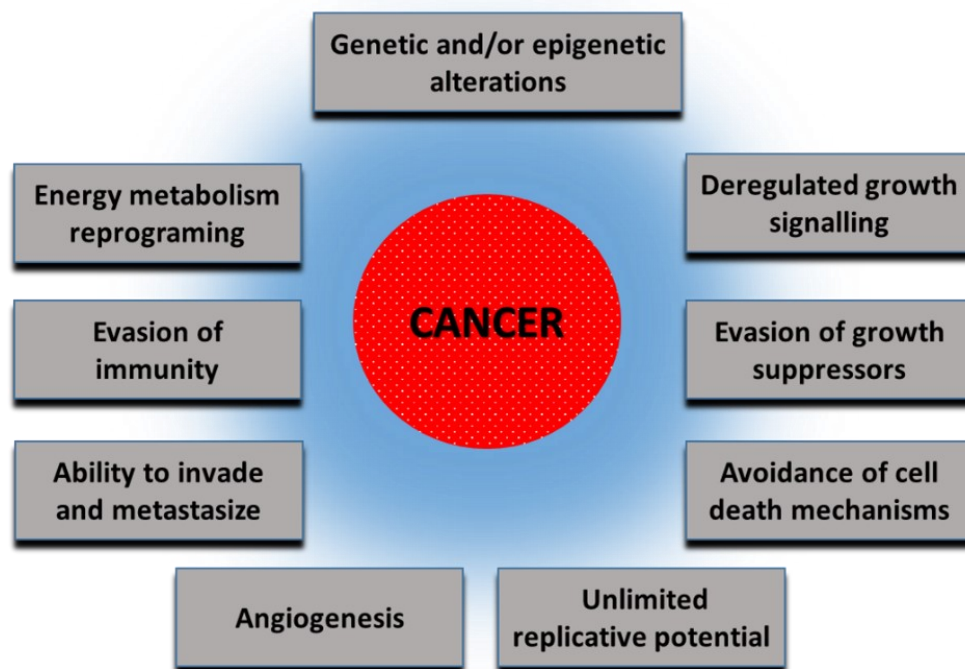


Figure 1.1.1. Main cancer characteristics.

Cancer cells have the ability to invade neighbouring tissues as well as spreading to distant tissues and other organs in the body (metastasis). This multistep process begins with the invasion of local cells, intravasation and the transportation of cancer cells through the lymphatic or vascular circulation [5, 10]. This is followed by extravasation into the parenchyma of tissues in a new location, where the cancer cells attach and develop a secondary tumour. Here cancer cells can become dormant [5, 10] Figure 1.1.2. Downregulation or inactivation of the adhesion protein E-cadherin in carcinomas has been associated with the enhancement of invasion and metastasis [5, 10]. Similarly, tumour associated macrophages can facilitate tumour invasion and metastasis as well as immunosuppression and tumour angiogenesis [11].

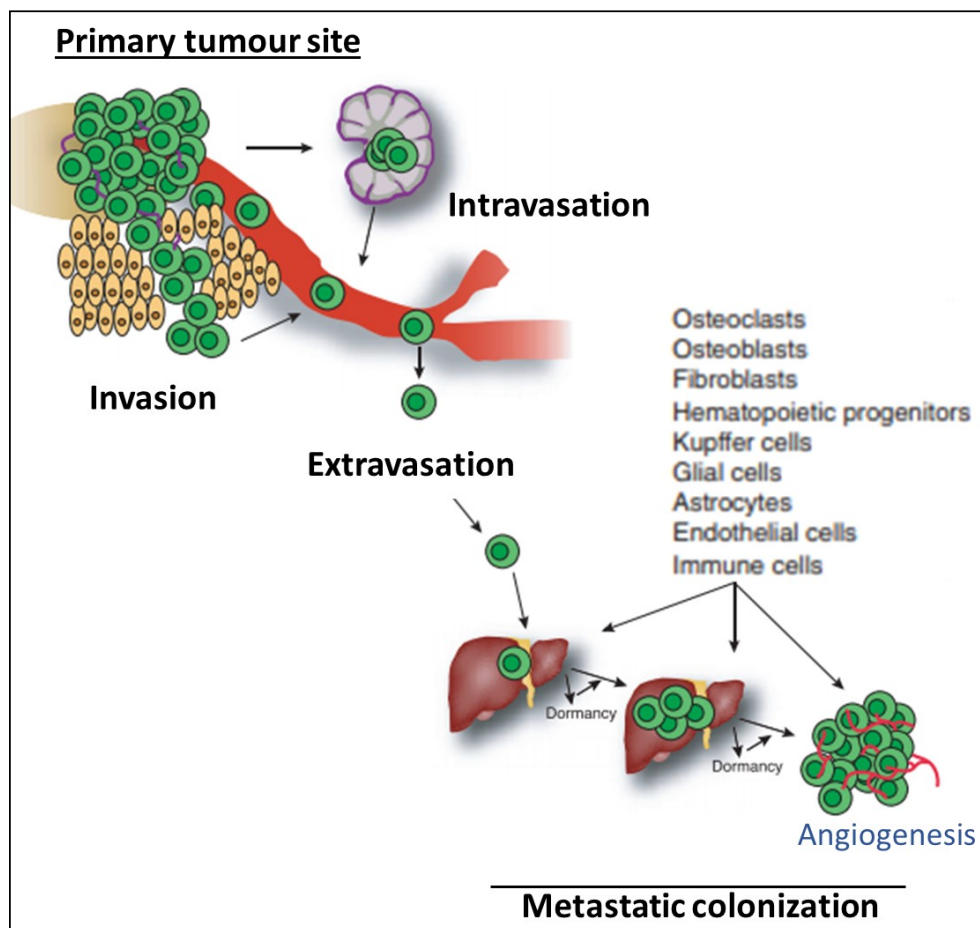


Figure 1.1.2. Cancer invasion and metastasis multistep process (Adapted from Steeg 2006 [10]).

Tumour microenvironment involves reciprocal interactions between cancer cells and cells recruited in the tumour site, and plays an essential role in cancer metastasis. Tumour cells interact with cells in distant organs where the metastatic colonisation takes place, including amongst others endothelial cells, immune cells, osteoblasts, astrocytes and hematopoietic progenitors [10] (Figure 1.1.2). The anti-tumour functions of immune cells are downregulated in the tumour site and they contribute to promote tumour growth. This includes cells that mediate adaptive immunity (mainly dendritic cells and T lymphocytes, B cells are not frequent) and innate immunity (mainly macrophages and polymorphonuclear leukocytes, natural killers are unusual) [12]. Understanding not only the role of the cells recruited by the tumour but also the molecular interactions in the tumour microenvironment could open a new path to the development of new anti-cancer therapies.

1.1.1. Breast cancer

Breast cancer is the leading cause of cancer deaths among women, and the second most common cancer among women, after non-melanoma skin cancer [1]. Worldwide, in 2017, breast cancer was the fifth leading cause of cancer deaths for both men and women [1]. The chances of women developing cancer is higher in high socio-demographic index countries when compared to low socio-demographic index countries [1].

Breast cancer patients are mainly symptomatic, a breast lump being the most common symptom. In addition, patients can experience no-lump symptoms such as nipple and breast skin abnormalities, breast pain and breast swelling, amongst others [13]. Initial breast cancer detection methods include breast self-examinations and clinical breast exams, followed by imaging and tissue biopsy. Imaging diagnostic tools include mammography, magnetic resonance imaging, and ultrasound. Mammography is the most common breast imaging tool for women aged 40 years or over, but the sensitivity of this tool can be affected by breast density [14, 15]. One of the biggest concerns related to mammography (especially

as a screening tool) is the potential health risks associated to radiation exposure. However, the benefits for women over 40 years old could outweigh the risk [16]. As an alternative to mammography, ultrasound is mainly recommended to women under 40 years old, pregnant or breastfeeding women, as well as for those with dense breast tissue [17]. Magnetic resonance imaging has demonstrated to be a highly sensitive tool for breast cancer detection and is not affected by breast density, but its specificity is variable and lower than a mammography [18, 19]. Although these imaging techniques are useful and widely applied as screening methods, they are not used as a stand-alone diagnosis method but a combination of these tools can increase their sensitivity to detected breast cancer [20]. In addition, after clinical and imaging analyses, in most cases needle biopsy is used to establish a diagnosis when malignancy is suspected [21].

Most breast cancers are carcinomas, which can be classified into different groups according to the breast tumour histopathology. Ductal tumours develop in breast ducts and they account for most of the breast cancer cases, after lobular tumours (Figure 1.1.3A). Carcinomas can be invasive or *in situ*. Invasive carcinoma is the most common type, and this includes invasive ductal carcinoma (IDC) and invasive lobular carcinoma (ILC). Meanwhile *in situ* carcinomas, include ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS). IDC refers to the abnormal proliferation of neoplastic cells spreading from the duct wall into the surrounding breast tissue. DCIS refers to those epithelial cells that remain confined to the epithelium by an intact basement membrane [22]. ILC is the second most common histologic subtype of breast cancer after IDC, but it is also more difficult to be diagnosed by mammography and clinical breast examination. However, a patient's prognosis is considered to be better for ILC patients than IDC patients [23]. ILC is morphologically characterised by small non-cohesive cells lacking E-cadherin protein expression [24], which are more likely hormone receptor-positive [23]. LCIS is considered a risk factor of breast carcinoma but not necessarily a precursor of the disease [25]. Breast carcinomas can also spread throughout the body and are called metastatic carcinomas. Although only a minority of women are initially diagnosed with a late stage (or metastasis), almost 30% of early-stage

breast cancer diagnosis will develop metastasis [26]. Metastatic breast cancer is responsible for most breast cancer patient deaths.

Breast cancers can be classified into three grades, according to the Nottingham Grading System, based on the degree of differentiation of the tumour tissue [27] (Figure 1.1.3). Grade 1 is a well-differentiated tumour with tubule formation ($>75\%$), mild degree of nuclear pleomorphism, and low mitotic count. Grade 2 is a moderately differentiated tumour. Grade 3 is poorly differentiated with no tubule formation ($<10\%$), a marked degree of cellular pleomorphism and frequent mitoses.

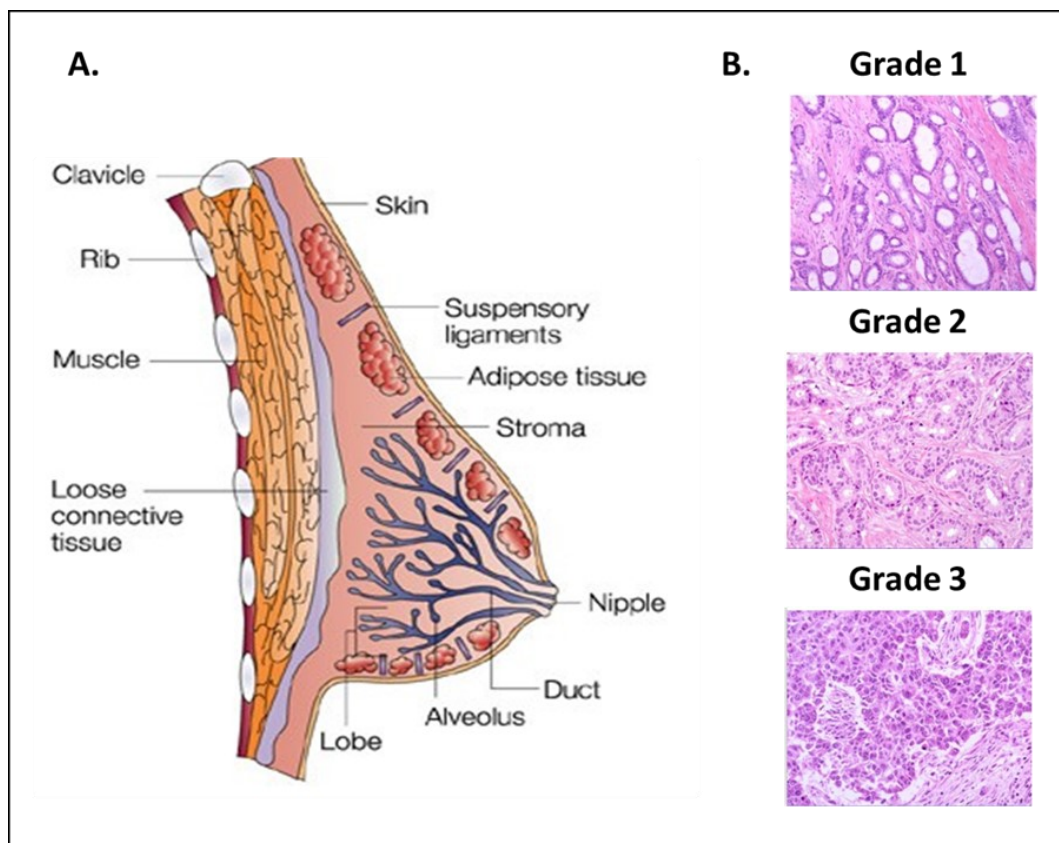


Figure 1.1.3. Histological grade of breast cancer. **A.** Anatomy of the human mammary gland (Figure taken from Ali and Coombes 2002 [28]). **B.** Histological grade of breast cancer as assessed by the Nottingham Grading System (Figure taken from Rakha *et al.* 2010 [27]).

Breast cancer refers to a group of diseases which can be classified into different molecular subtypes based on the expression of the oestrogen receptor (ER), progesterone receptor (PR) and the human epidermal growth factor receptor 2 (HER2), also known as HER2/neu or ErbB2 [29]. Breast cancer subtype luminal (luminal A or B) refers to breast cancers which are positive for ER (ER+) and/or PR (PR+). Subtype HER2 includes breast cancers overexpressing HER2 [29]. The basal-like subtype includes breast cancers that do not express any of ER, PR and HER2. This type of cancer is called triple negative (TN) breast cancer [29]. Inhibition of oestrogen signalling has been the main form of hormonal therapy for patients with ER+ breast cancer. Patients with hormone receptor-positive breast cancers (ER+/PR+) can be treated with hormonal therapy such as tamoxifen that blocks the effect of oestrogen on tissue receptors. Those patients generally have a better prognosis compared to those who are breast cancer positive for HER2 amplification [30, 31]. However, there are patients who develop resistance to endocrine therapies [28]. Breast cancer HER2-positive (HER2+) patients are treated with agents that block HER2, such as trastuzumab [31]. TN breast cancer have a poor prognosis and patients are treated mainly with chemotherapy because there are no approved targeted therapies available [30].

Breast cancer risk has been associated with prolonged exposure to oestrogens (early menarche, late first full-term pregnancy, late menopause, oral contraceptives and oestrogen-replacement therapy) [28]. Hormonal factors but also genetic factors have been associated with breast cancer risk. This includes mutations of the genes Breast cancer 1 (*BRCA1*), Breast cancer 2 (*BRCA2*), Tumour protein 53 (*TP53*), Fibroblast growth factor receptor 2 (*FGFR2*), Mitogen-activated kinase kinase kinase 1 (*MAP3K1*), Lymphocyte-specific protein (*LSP1*) and Caspase 8 (*CASP8*) [32].

1.1.2. Ovarian cancer

Ovarian cancer is one of the top five most common types of cancer in women worldwide and one of the leading causes of gynaecologic cancer-related deaths after cervical cancer [1]. Overall, ovarian cancer is the most lethal gynaecologic cancer as currently there is not an adequate screening scheme for its diagnosis and treatment. Ovarian cancer is diagnosed at an advanced stage because the early stage symptoms are similar to common gastrointestinal, genitourinary and gynaecological related illnesses [33]. Approximately 90% of women with ovarian cancer survive for five years after being diagnosed at an early stage, while less than 30% of women diagnosed at an advanced stage survive for that period of time [33].

Ovarian cancers (more than 90%) have predominantly an epithelial histology and can be classified into different subtypes which are considered different diseases responding differently to chemotherapy [34]. Ovarian epithelial cancer subtypes were initially classified into four groups based on their histopathology: clear cell, endometrioid, mucinous and serous. Later, five subtypes were proposed dividing the serous subtype into high grade serous and low-grade serous [35]. High grade serous ovarian cancer is the most commonly diagnosed subtype of ovarian cancer (70%), followed by endometrioid (10%) and clear cell (10%), while mucinous (3%) and low-grade serous carcinomas (<5%) count for less than 10 % of the cases together [34].

Several genetic mutations have been associated with these specific ovarian cancer subtypes [35, 36]. High grade serous ovarian carcinomas are aggressive cancers that are mainly detected in the advanced stage. They exhibit mutations in the TP53 gene and inactivation of the BRCA genes by promoter hypermethylation or mutations. Clear cell, mucinous, endometrioid and low grade serous carcinomas rarely exhibit TP53 mutations and BRCA inactivation is unknown. Clear cell and endometrioid ovarian cancers are associated with mutations in Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha (*PIK3CA*) and AT-rich interactive domain-containing protein 1 (*ARID1A*). Similarly, *PIK3CA* mutations have also

been found in low-grade serous cancer as well as mutations in the B-Raf (BRAF) and K-Ras (KRAS) proto-oncogenes. Genetics aberrations in KRAS also occur in most mucinous tumours (Figure 1.1.4).

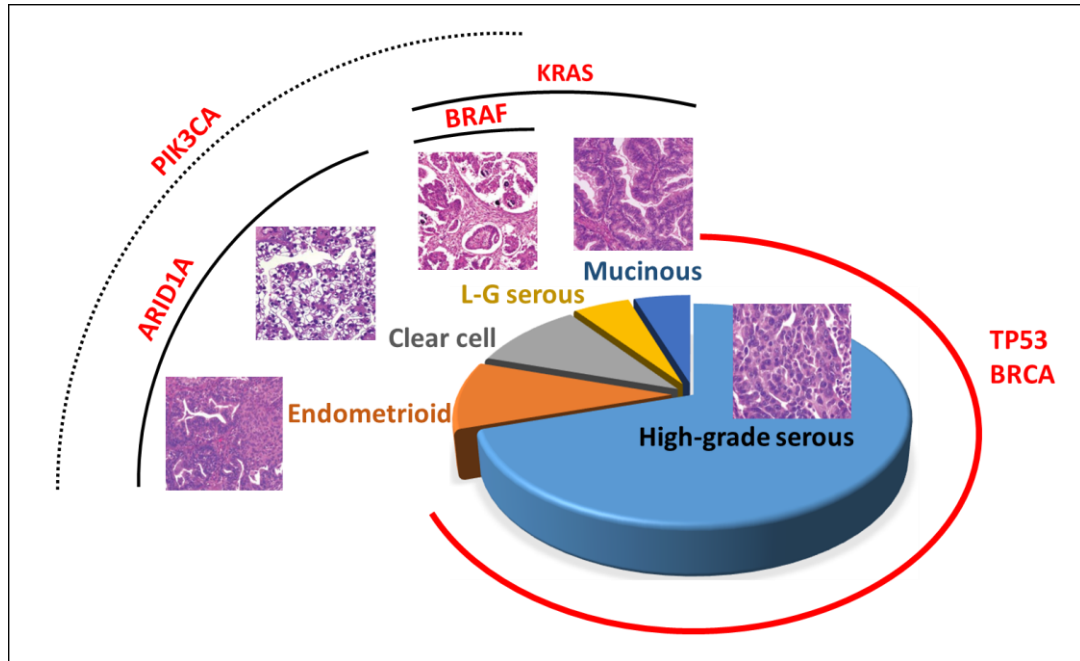


Figure 1.1.4. Ovarian epithelial carcinomas subtypes and genetic aberrations associated to them. L-G serous = low-grade serous ovarian cancer.

The origin of ovarian epithelial cancers is still controversial and it was initially attributed mainly to the ovarian surface epithelium [34]. However, recently, a non-ovarian origin has also been suggested for the distinct ovarian carcinoma subtypes, a Müllerian type-tissue origin [36]. Serous ovarian cancer could have an origin in the fallopian tube epithelium that embeds on the ovary, while endometrioid and clear cell ovarian cancers could have an origin in the endometrium and it could be due to endometriosis [36].

Currently, the main treatment for ovarian cancer is surgery, followed by chemotherapy, if diagnosed at an early stage. A combination of transvaginal

sonography and the cancer antigen 125 (CA125) serum marker have potential to detect ovarian cancer, and are currently the main tools used to detect this type of cancer [33]. However, transvaginal sonography is rarely used when symptoms are not evident, and most ovarian cancer patients are asymptomatic at an early-stage or experience nonspecific symptoms which are related to other pathologies. In addition, not all ovarian cancers express CA125 and high levels of this antigen are frequently found in women with other gynaecologic pathologies including endometriosis and fibrosis as well as during pregnancy and ovulation [37]. Biomarker discovery for the early detection of ovarian cancer is a critical factor to improve patient survival and treatment. Platinum-based drugs are normally used in ovarian cancer treatment. These include carboplatin and cisplatin. Carboplatin has reduced toxicity when compared to cisplatin [38]. Resistance to chemotherapy is one of the biggest challenges for using platinum compounds; gemcitabine, doxorubicin, and bevacizumab are mainly used when resistance is detected [38]. Extensive research is continuing to find alternative treatments, like targeted therapies and immunotherapies [38].

1.2. Lipids

Lipids are a diverse group of hydrophobic or amphipathic molecules which play essential biological functions as those played by other major biomolecules, like proteins, carbohydrates and nucleic acids. Their main biological functions include energy storing, serving as structural components of cell membranes and cell signalling. Lipids are a diverse group of molecules which have been classified into eight main categories that encompass both eukaryotic and prokaryotic sources [39, 40]; fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids and polyketides.

Fatty acyls, which includes fatty acids and conjugates, represent the major lipid building block of complex lipids. Fatty acids consist of a carboxylic acid group linked to a hydrocarbon chain of variable length and degrees of unsaturation (double

bonds). Glycerolipids and sterol lipids are involved in energy storage, mainly as triacylglycerol (TG) and sterol esters respectively, in lipid droplets [41]. These lipid droplets not only function as an energy reservoir but also as a reserve of fatty acids and sterol components for membrane biogenesis. Glycerolipids (e.g. diacylglycerol (DG) and TG), refers to lipids containing glycerol. These lipids have at least one hydrophobic chain linked to a glycerol backbone. Glycerophospholipids also contain glycerol, but they have been grouped separately as they also contain a phosphate group as well as two fatty acid chains of variable carbon group length. Glycerophospholipids are the major structural lipids of biological membranes. This group includes, amongst others (Figure 1.2.1), phosphatidic acid (PA), phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylglycerol (PG) and phosphatidylinositol (PI). The lysophospholipid forms of these glycerophospholipids have only one of the two fatty acid chains, lysophosphatidic acid (LPA), lysophosphatidylcholine (LPC), lysophosphatidylethanolamine (LPE), lysophosphatidylserine (LPS), lysophosphatidylglycerol (LPG) and phosphatidylinositol (LPI).

Sphingolipids and sterol lipids are also an important component of biological membranes along with glycerophospholipids [39]). Sphingolipids are lipids that contain a sphingoid base (such as sphingosine) instead of the glycerol that is found in glycerolipids (e.g. ceramide (Cer) and sphingomyelins (SM)) (Figure 1.2.2). Prenol lipids include, amongst others, isoprenoids (e.g., carotenoids) and quinones and hydroquinones (e.g. vitamin E and vitamin K). Saccharolipids refers to lipids in which fatty acids are linked directly to a sugar backbone. This sugar substitutes for the glycerol present in glycerolipids and glycerophospholipids [39]. Polyketides are mainly cyclic molecules with great structural diversity, which include secondary metabolites and natural products [39].

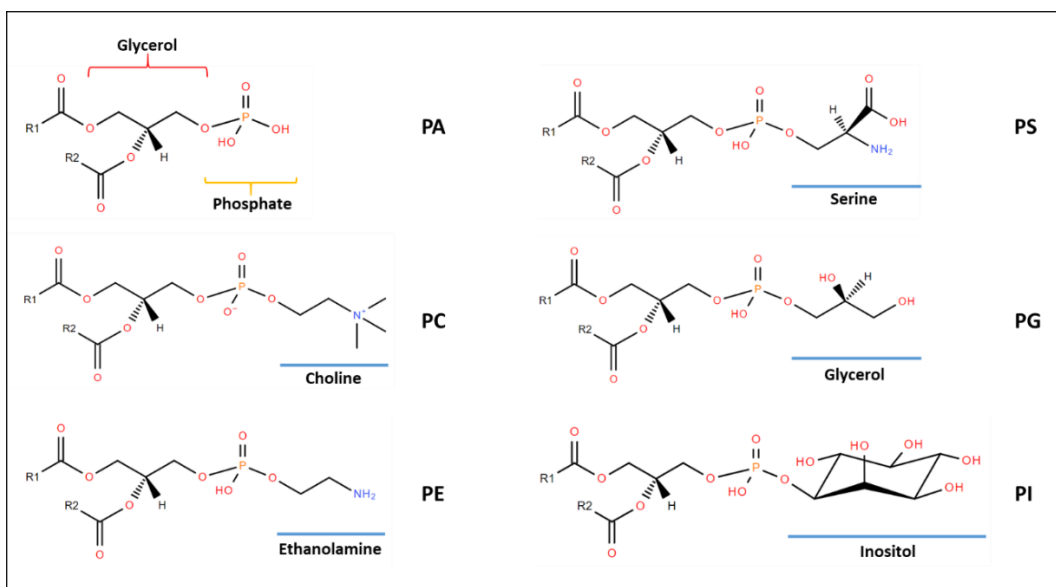


Figure 1.2.1. Major glycerophospholipids. The general structure consist of two fatty acids (R1 an R2) ester- (or ether-) linked to glycerol, and a phosphate group linked to a polar head group (indicated by the blue lines in the figure).

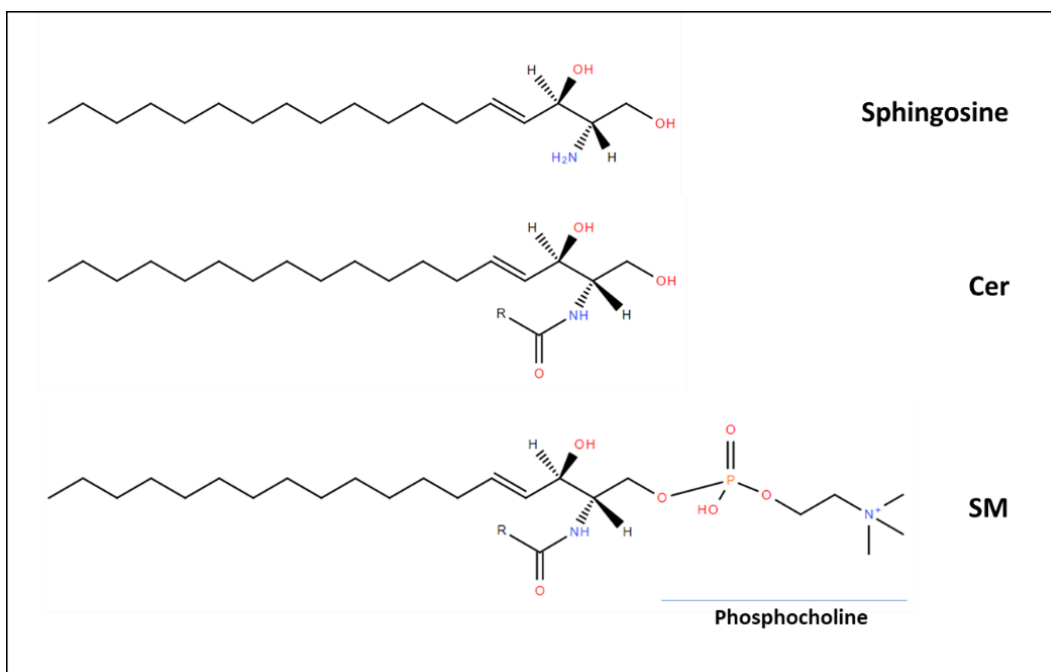


Figure 1.2.2. Sphingosine and major sphingolipids. The general structure of sphingolipids consist of a sphingoid base (mainly sphingosine in mammals),

which is an amide linked to a fatty acid. Cer are composed of a sphingoid base and a fatty acid. SM have a Cer backbone and a phosphocoline.

Cellular membranes are mainly formed by polar structural lipids with a hydrophobic tail and a hydrophilic head (Figure 1.2.3) [41]. The hydrophobic part of these structural glycerophospholipids is basically a DG and the hydrophilic part consists of a polar head group linked to a phosphate group. More than 50% of phospholipids in most eukaryotic bilayer membranes are PC, this is followed by PE which enforces a curvature stress onto the membrane that facilitates cells budding, fission and fusion [41]. The asymmetric distribution of lipids between the bilayer leaflets also contributes to the curvature stress in the membranes. The lipid composition of early endosomes is similar to plasma membranes. In contrast, late endosomes show an increase of bis(monoacylglycero)phosphate (BMP) in their inner membranes, and a decrease of PS and sterols. BMP is involved in multivesicular body production, fusion and sphingolipid hydrolysis. The endoplasmatic reticulum is the main site for the synthesis of major glycerophospholipids (PC, PE, PI, PS and PA), Cer, cholesterol and TG. Galactosylceramide is produced in the endoplasmatic reticulum of epithelial cells. The Golgi lumen is where mainly SM and complex glycosphingolipids are assembled. Mitochondria synthesise PE, PA and PG. PA and PG are used to produce cardiolipins (CL).

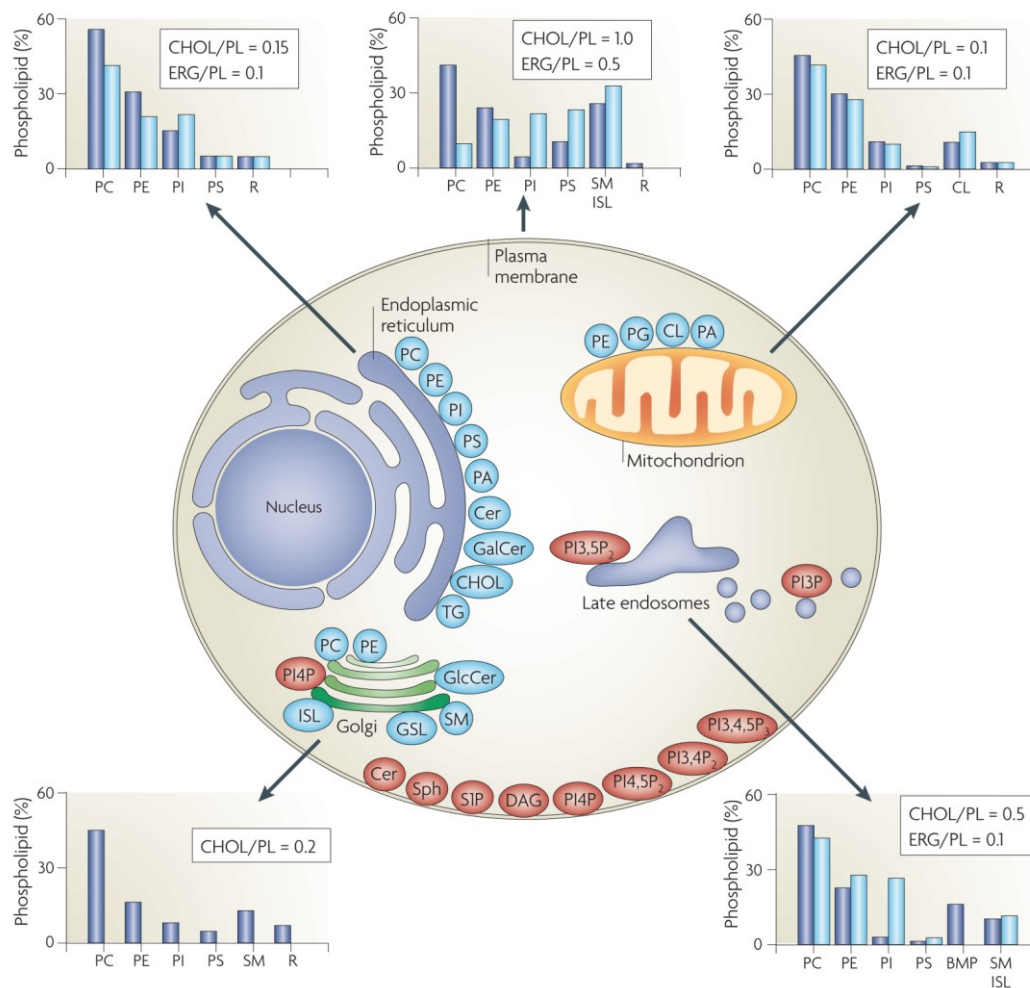


Figure 1.2.3. Lipid synthesis and composition of cell membranes. Lipid composition is indicated as a percentage of the total phospholipid (PL) in mammals (blue) and yeast (light blue). Sterol content is indicated as the molar ratio of cholesterol (CHOL; in mammals) and ergosterol (ERG; in yeast) to phospholipid. Lipids in red circles are involved in signalling and organelle recognition pathways. Taken from van Meer *et al.* (2008) [41].

Lipids can function as signalling molecules. They can serve as first and second messengers in signal transduction and molecular recognition processes. Phosphorylated derivatives of PI are involved in signalling and recognition. They are important in defining organelle identity as well as in recruiting both soluble and membrane proteins to specific membranes [41]. Similarly, the hydrolysis of

glycerolipids, glycerophospholipids, and sphingolipids produce messenger lipids which are DG, PA, LPC and LPA, sphingosylphosphorylcholine (SPC), sphingosine, sphingosine-1-phosphate (S1P), ceramide-1-phosphate (C1P) and Cer [41] (van Meer 2008). These lipid messengers (LPC, LPA, SPC, S1P and sphingosine) can leave membranes and signal through related membrane receptors or remain in the membrane (PA, DG, C1P and Cer) and recruit cytosolic proteins [41]. S1P can function as an intra- and extracellular signal and second messenger. In addition, sphingolipids and cholesterol function to define membrane domains called lipid rafts which recruit proteins from the cytosol. These lipid rafts are involved in membrane signalling and trafficking [42].

1.2.1. Lipids metabolism and cancer

The metabolism of cancer cells is reprogrammed to fulfil their uncontrolled cell proliferation which requires an efficient generation of energy and cellular components. Cancer cells proliferation is supported by an increase in glucose consumption and enhancement of glycolysis [7], but they also rely on a high rate of glutamine consumption to produce energy and to synthesise proteins, nucleotides and fatty acids [43]. Metabolic changes are driven by oncogene activation and loss of tumour suppressors, as well as the tumour microenvironment [43]. Proliferating cancer cells also require a continuous supply of lipids, which is facilitated by increasing the uptake of exogenous lipids or the activation of fatty acids and cholesterol synthesis in a low-lipid environment [44]. Increased expression of lipogenic enzymes is observed in cancer cells; acetyl-CoA carboxylase, fatty acid synthase and ATP citrate lyase. Increasing *de novo* lipogenesis in tumour cells is associated with cancer development and progression as well as poorer prognosis [45]. Lipogenic tumours show an increase in the degree of lipid saturation compared with nonlipogenic tumours which protects cells from free radicals and chemotherapeutics [45]. This provides a rationale for the use of inhibitors of the lipogenic pathway [44, 45]. Free radicals can extract electrons from lipids and lead to the formation of oxidized lipid species which can degrade into reactive products

(including malondialdehydes and 4-hydroxyalkenals) that can cause cell damage when expressed at high levels [45].

Signalling lipids are also involved in cancer, from cell growth to metastasis. Lipid second messengers such as PA participate in several signalling cascades and regulates signalling proteins, including protein kinases and phosphatases [46] (Fang 2001). PA, a product of phospholipase D, is involved in the mammalian target of rapamycin (mTOR) signalling that governs cell growth and proliferation [46]. PA can stimulate the ribosomal subunit S6 kinase 1 and the eukaryotic initiation factor 4E binding protein 1 which are downstream effectors of mTOR [46]. In addition, LPA stimulates cell proliferation, migration and survival [47]. PI3K, promotes the formation of signalling lipid phosphatidylinositol-3,4,5-trisphosphate (PIP3) that modulates cell growth, proliferation and motility [48]. Cer and sphingosines have pro-apoptotic and antiproliferative actions. However, they can be phosphorylated by lipid kinases. Cer is converted to C1P by Cer kinase and sphingosine is converted into S1P by sphingosine kinases. C1P and S1P promotes cell growth and proliferation [48]. Cer can be generated by *de novo* in the cell or by hydrolysis of SM by sphingomyelinase. In addition, lipid raft which are rich in cholesterol and sphingolipids have also been associated with cancer progression and associated with the PI3K/protein kinase B (AKT) pathway [49].

1.3. Extracellular vesicles

Membrane-enclosed vesicles secreted by cells into the extracellular space were first observed in solid tissues and biological fluids around four decades ago, but it is only during the last two decades that interest in these extracellular vesicles (EVs) has increased significantly [50]. Exosomes and microvesicles (also known as microparticles or ectosomes) are the most studied EVs. Exosomes originate by exocytosis of multivesicular bodies while microvesicles are derived by the outward budding and fission of the plasma membrane (Figure 1.3.1) [51]. Other EVs include apoptotic bodies which are generated during programmed. Exosomes

are believed to have a diameter of between 50 and 100 nm, microvesicles between 100 and 1,000 nm, and apoptotic bodies 100 and 5,000 nm approximately [51]. However, these measurements can vary between the methods used for the EV's quantification. Currently, there is not a well-defined classification of EVs by size. Moreover, to date, there are not specific markers of EV subtypes to be able to assign a specific biogenesis pathway when studying EVs. Therefore, it is recommended to refer to them as EVs [52]. The term EV has been suggested to define non-replicating particles delimited by a lipid bilayer which are released by cells [52].

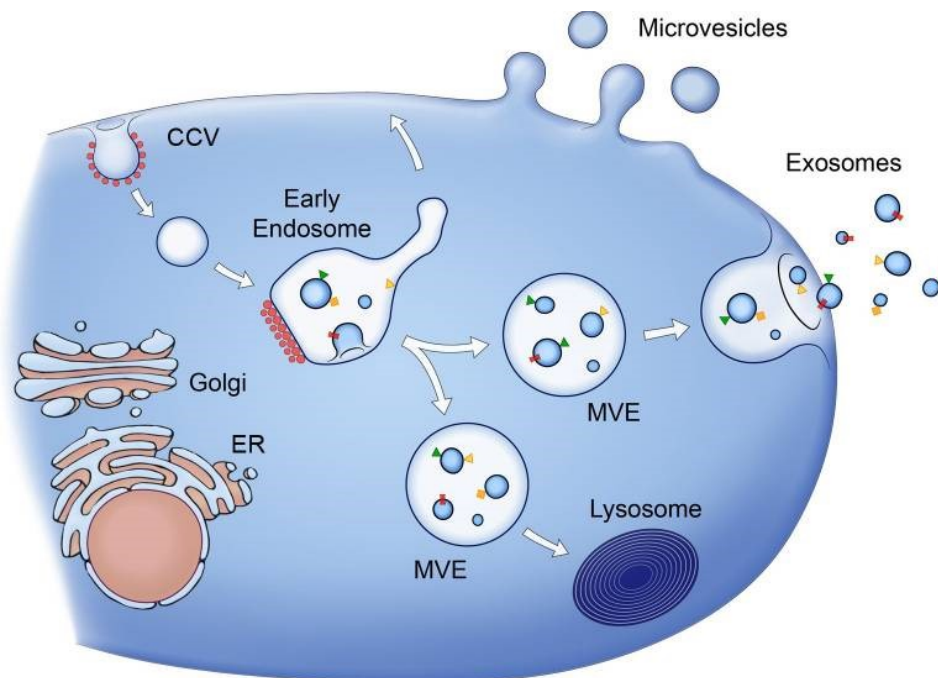


Figure 1.3.1. Extracellular vesicles biogenesis. Figure taken from Raposo and Stoorvogel 2013 [53].

All cells can produce EVs, including tumour cells. EVs have been studied in a wide variety of cells, including reticulocytes, platelets, T cell, B cells, neurons, mast cells, antigen presenting cells and cancer cells, amongst others [50, 53, 54]. Similarly, they have been found in body fluids including blood, saliva, urine, cerebrospinal fluid, amniotic fluid, ascites fluid and semen [50, 53]. These EVs not

only consist of a lipid bilayer but also proteins, nucleic acids and small metabolites. Several proteins have been associated with EVs. These include, amongst others, tetraspanins (e.g. CD9, CD63, CD81, CD82), annexins, G-proteins, antigen presentation (MHC class-I and class-II proteins), as well as proteins associated with the endosomal sorting complex required for transport (e.g. ALIX, TSG101, Flotillins-1 and 2), chaperons (e.g. Hsp70 and Hsp90) and cytoskeletal proteins (actin and tubulin) [50, 53, 54]. Those proteins have been suggested as protein EV markers and are used for protein characterisation of EVs. Currently, there are two databases for EVs; Vesiclepedia and ExoCarta. These databases show a high number of entries for proteins associated with EVs, followed by RNA (mRNA and miRNA), with the lowest number entries for lipids.

EVs play an important role in intercellular communication processes. They can transfer their molecular content into local or distant recipient cells and alter their cell signalling. Different ways of exosome internalisation into recipient cells have been proposed. These include; endocytosis [55], direct fusion with the recipient cell membrane [56], phagocytosis [57] and micropinocytosis [58]. Consequently, EVs capability to transport and transfer bioactive molecules to target cells make them potential therapeutic vehicles. EVs as therapeutic vehicles have been explored because they are biocompatible, biodegradable and produced by most cells. In addition, they have great potential as biomarkers for different diseases and even as therapeutic targets [59]. But they can also mediate drug resistance [59].

1.3.1. EVs' isolation methods

Several methods have been used for the extraction of EVs from different biological matrices and conditioned medium. A few methods have been proposed for a quick and simple way for the extraction of EVs. Most of these methods are based on polymer precipitation, for example, by using polyethylene glycol (PEG). Particles can be precipitated from a sample of interest by hydrophilic polymers (such as PEG) which decrease the solubility of the particles in the solution. This strategy has been

used for the development of commercial kits for EVs extraction, for example ExoQuick (System Biosciences), Total Exosome Isolation Reagent (Invitrogen) and ExoPrep (HansaBioMe). Unfortunately, in addition to EVs contaminant proteins, lipoproteins are also co-precipitated. Moreover, EVs aggregation and alteration in EV's structure has been observed [60]. Furthermore, an additional step is needed to remove the polymer from the samples for downstream analysis. Immuno-isolation methods are also considered a simple and rapid approach as precipitation methods [61]. They are based on the antibodies affinity to EV's proteins which allow the extraction of pure EVs. However, it has the disadvantage of the isolation of EV subpopulations. In addition, it is an expensive method and it is useful mainly for a small volume of sample.

The most used approach for the extraction of EVs is ultracentrifugation at 100,000 x g to pellet EVs (but it also pellet contaminants), followed by a second round to remove the contaminating proteins [62]. Most methods are accompanied by previous differential centrifugation steps to remove cell debris (300 – 2,000 x g) and large vesicles (10,000 – 20,000 x g). This is a useful technique for the extraction of EVs from *in vitro* cultured cells when a large amount of sample is processed. But, it is not recommended as a stand-alone method for the extraction of EVs from complex matrices such as blood plasma that contains components similar to EVs, since both EVs and non-EVs particles with similar sedimentation rate will pellet together. Another method that uses ultracentrifugation is density gradient ultracentrifugation. This method is considered the best method to obtain pure EVs as it is needed for downstream omics analysis and outperforms ultracentrifugation and precipitation methods [63].

Size exclusion chromatography (SEC) is a useful technique for the extraction of EVs from conditioned cell culture media. This method allows to separate EVs from soluble proteins. This is an easy and fast method; however, it needs a previous sample concentration step for large volume samples. Most SEC-based methods use sepharose (CL-2B) [60, 64], but Capto™ Core 700 has also been used [65].

The extraction of EVs from human blood plasma is a difficult task because of the co-isolation of contaminants which affect EV's associated analysis. Soluble proteins and lipoproteins contained in blood plasma are problematic contaminants for the study of EVs. For instance, soluble proteins such as albumin could mask proteomics analysis. Lipoproteins contain lipids, proteins and RNA which could interfere in the interpretation of lipidomics, proteomics and transcriptomics studies. Methods for extraction of EVs from blood plasma are mainly based on a combination of approaches to separate blood plasma components, based on their density and size. Density gradient ultracentrifugation has become a useful technique for the separation of particles in blood plasma based on their densities, but this requires additional steps for the separation of particles with similar density to EVs. SEC in combination with density gradient ultracentrifugation, is a useful strategy for the separation of EVs from lipoproteins and non-EV associated proteins [66, 67].

Sucrose and iodixanol are used in density gradient ultracentrifugation methods, but iodixanol has advantages over sucrose-based methods. Iodixanol is not viscose like sucrose. It does not affect EVs structure and functionality and it has the advantage to generate iso-osmotic gradients that preserve EVs properties [63]. The disadvantages are that it requires expensive equipment as mentioned above for ultracentrifugation, and the current methods are laborious and long. They are difficult to reproduce as they require specific rotors and need to follow precise steps. However, to date, density gradient ultracentrifugation in combination with SEC-based methods are the best methods for the isolation of EVs from blood plasma [66, 67].

The choice of a specific EVs' extraction method depends mainly on the downstream analyses, the starting material and the sample volume. Advances in this subject will allow the EVs field to progress at a faster pace. So far, pure EVs are much more easily obtained from conditioned cell culture media, but the extraction of EVs from body fluids still require improvement.

1.3.2. EVs' lipid composition

The lipid composition of EVs has been studied less than other molecules found in these nanovesicles, like proteins. The lipid content of EVs secreted by *in vitro* maturation of guinea pig reticulocytes was found to be similar to erythrocyte membrane composition but with a lower content of PE [68]. EVs derived from B-cells showed a high content of cholesterol, SM, and ganglioside GM3 [69]. Similarly, the enrichment of SM and cholesterol has also been observed in the lipid content of EVs secreted by T lymphocyte cells (Jurkat cells) and T cell blasts when compared to their parent cells [70]. In contrast, the lipid composition of exosomes released by rat mast cells (RBL-2H3 cells) or human dendritic cells did not show enrichment of cholesterol but of SM when compared to parental cells [71]. The same study found that the EVs lipid bilayer was similar to the plasma membrane containing cholesterol and major phospholipids, but exhibiting a random distribution of PE and a rapid lipid translocation (flip-flop) between the two leaflets of the bilayer. Enrichment of Cer and cholesterol was found in a mouse oligodendroglial cell line (Oli-neu cells) containing a large number of multivesicular endosomes [72]. EVs secreted by Oli-neu cell lines had a lower content of PC but a higher content of sphingolipids (SM and hexosylceramides (HexCer)) compared to plasma membranes [72]. EVs in seminal fluid, called prostasomes had a high content of SM, cholesterol, and glycosphingolipids [73].

For the analysis of EVs secreted by cancer cells, a study of one non-cancerous and two prostate cancer cell lines found that EVs secreted by cancer cell lines were highly abundant in sphingolipids, glycerophospholipids and sterol lipids [74]. Another group studied the cells and EVs from the prostate cancer cell line PC-3 and found an enrichment of glycosphingolipids, SM, cholesterol, and PS in the EVs compared to cells [75]. Likewise, urinary exosomes from prostate cancer patients also had a high content of cholesterol and SM [76]. In comparison with the EVs released by the PC-3 cell line, urinary exosomes had a higher content of cholesterol and a lower phospholipid content [76]. EVs secreted from other types of cancer cells have been less explored and limited to the study of only one or two cancer cell

lines. For instance, a group studied a glioblastoma (U87), a hepatocellular carcinoma (Huh7) and also bone marrow-derived mesenchymal stem cells (MSC) [77]. They found that Huh7 and MSC derived EVs were enriched in cardiolipins, while U87 derived EVs were enriched in SM. More recently, in 2020, a study of EVs from two cell lines derived from one metastatic TN breast cancer cell line found that cholesterol and SM were enriched in EVs compared to cells [78]. Another study, found cholesterol ester and zymosterol to be abundant in EVs from one endometrioid ovarian cancer cell line compared to non-cancerous ovarian surface epithelial cells, but the EVs lipid composition was not compared to their parental cells and therefore this study does not provide information about the ovarian cancer-derived EVs lipid composition [79]. More comprehensive studies, including the study of a higher number of cell lines from different cancer subtypes and/or EVs found in body fluids, will produce a more representative understanding of the lipid composition of cancer-derived EVs and a more concrete understanding of its potential for the identification of cancer biomarkers.

1.3.3. EVs and cancer

Cancer cells produce EVs and they can be found in cancer patient's body fluids, such as urine and blood plasma [76, 80]. They participate actively in the modulation of the cellular microenvironment facilitating tumorigenesis, cancer migration and establishment of metastatic niches [81-83]. Cancer-derived EVs are able to interact with immune cells to avoid immune response. For example, cancer EVs can induce apoptosis in cytotoxic T-cells, differentiation of T-helper cells to regulatory T-cells or suppression of T-cell signalling molecules. Similarly, Natural Killers cells can be inactivated or their proliferation reduced by tumour-derived EVs [81]. In addition, microRNAs found in EVs can participate in modulation of the tumour-immune response by inducing the secretion of inflammatory cytokines in macrophages [82]. MicroRNA-21 participates in tumour growth, migration and invasion [82]. However, the number of microRNA copies in EVs have been found

to be low, and therefore the function of EVs as vehicles of microRNA in intercellular communication processes has been questioned [84].

Cancer-derived EVs have significant potential as a source of cancer biomarkers, for the development of non-invasive or minimally invasive diagnostic methods based on body fluids. For instance, a method for the detection and quantification of blood plasma EVs in melanoma patients based on the detection of a tumour-associated marker (caveolin-1) and CD63/Rab-5b proteins has been proposed [80]. In addition, proteins associated with EVs (PDCD6IP, FASN, XPO1 and ENO1 proteins) have been identified as candidate biomarkers for prostate cancer [85]. MicroRNA found in EVs has also been studied for several types of cancer, for instance, prostate [83], colorectal [87] and lung [88]. Furthermore, a panel of 8 microRNAs detected in EVs found in serum were suggested for ovarian cancer diagnosis [89]. Moreover, lipids found in urinary EVs from prostate cancer patients (PS(18:1/18:1), PS(18:0_18:2) and LacCer(d18:1/16:0)) were identified as potential biomarkers for prostate cancer diagnosis [76]. Although these and other EV-based biomarkers have been proposed for cancer diagnosis, their clinical application is still far from being implemented.

1.4. Lipidomics and Mass spectrometry

The large-scale study of lipids in a biological system, lipidomics, has experienced significant growth in the last decades due to the recognition of lipids as fundamental biomolecules contributing to pathogenesis of several diseases, such as cancer, diabetes, heart diseases, amongst many others. Equally, mass spectrometry (MS) has also facilitated and massively contributed to the development of the lipidomics field. Lipidomic analysis can be focused on the analysis of a defined number of specific lipids (targeted lipidomics) or on the global analysis of the lipids in a sample (untargeted lipidomics). Targeted lipidomic analysis is useful for the quantification of specific lipids whereas untargeted lipidomic analysis is semi-quantitative. A targeted approach requires previous knowledge of the lipid

composition of the sample to be studied, and to be able to select effectively the specific lipids of interest. In contrast, untargeted or discovery lipidomics does not require previous knowledge of the lipids content in the sample. This approach allows an in-depth global analysis of the lipids content in a sample and the discovery of trends in the samples studied, and consequently the discovery of relevant lipids. The disadvantage of this approach is the complexity of the data processing and analysis, as well as the annotation of the lipids identified. The selection of a lipidomics approach depends on the research questions.

1.4.1. Mass spectrometer

Mass spectrometers consist of a sample inlet, an ionisation source, one or several mass analyser, a detector and a data processing system (Figure 1.4.1) [90]. The sample to be analysed is introduced into the mass spectrometer by the sample inlet and ions are produced from the sample in the ionisation source. Those ions are separated according to their mass-to-charge ratio (m/z) in the mass analyser. In tandem mass spectrometry (MS/MS), selected ions are fragmented and analysed in a second analyser. The abundance of the ions emerging from the analyser is measured in the detector that converts the ions into electrical signals and then a mass spectrum is produced by the data processing system [90].

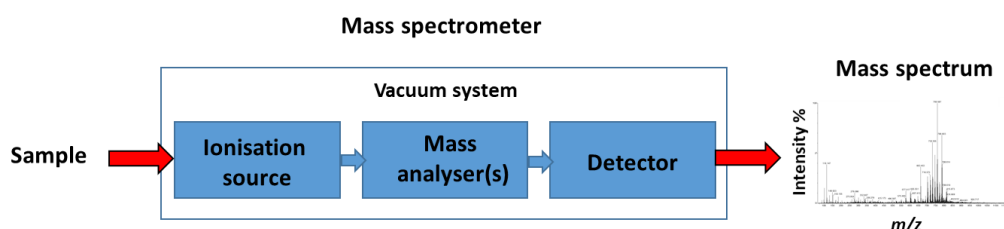


Figure 1.4.1. Mass spectrometer main components.

Ionisation sources

A large number of ionisation techniques are available, and these can be divided into several categories. The choice of the ionisation method used is determined by the type of sample to be evaluated and the type of information required. The ionisation techniques include, amongst others, electron ionisation (EI), chemical ionisation (CI), secondary ion mass spectrometry (SIMS), matrix-assisted laser desorption/ionisation (MALDI) and electrospray ionisation (ESI). EI is useful for gas-phase molecules, but induces fragmentation of the molecular ion which can prevent its detection [90]. CI produces a spectrum with less fragmentation, facilitating the detection of molecular species, thus it is complementary to EI. In SIMS, a surface is irradiated with a primary ion beam and the emitted secondary ions are analysed.

The most common ionisation techniques are MALDI and ESI, which are considered as “soft” ionisation techniques because, both minimise the sample degradation due to the low internal energy used. These techniques allow the analysis of small and large molecules with high sensitivity. MALDI is based on the desorption and ionisation of an analyte co-crystallised in a light-absorbing matrix (of small organic molecules) on a special surface by using pulsed laser radiation. This technique is very sensitive and reasonably tolerant to the presence of impurities; in fact, more tolerant than ESI [91]. ESI produces highly charged electrospray droplets by the application of a strong electric field to the liquid (sample) passing through a capillary tube with a weak flux (Figure 1.4.2) [90, 92, 93]. The solvent evaporates from the highly charged droplets, decreasing the size of the droplets until it reaches the point (called Rayleigh limit) where the surface tension can no longer sustain the Coulomb force of repulsion. Then, the droplet undergoes Coulomb fission generating smaller but more stable droplets [93]. Positively charged aerosol is formed under positive ion mode and the positive charges are mostly contributed by protons. While negatively charged aerosol is formed in negative ion mode, these charges are contributed by some negative ions (e.g. OH^-) [93]. ESI is more

frequently used in lipidomics analysis that use liquid chromatography to separate lipids.

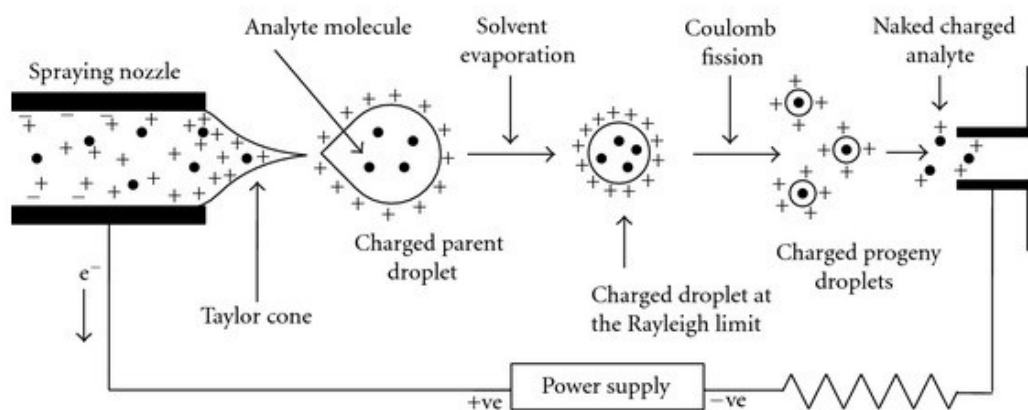


Figure 1.4.2. Ion formation in ESI. Initially, the analyte solution is pumped through the high-voltage capillary (emitter), then the accumulated charges (positive or negative) of the analyte would be repelled and a meniscus would be drawn out and deformed into a cone (called Taylor cone) under a very high electric field. Then, the analyte undergoes repeatedly solvent evaporation and Coulomb fission producing smaller droplets and finally nanodroplets. Figure taken from Banerjee and Mazumdar 2012 [93].

Mass analysers

There are different types of mass spectrometers. These can consist of a single mass analyser or, of the coupling of two or more mass analysers in the same system, called hybrid mass instruments. Spectrometers can show MS/MS capabilities which mean that a second analysis after a molecular dissociative process can be undertaken within the instrument.

Time-of-flight (TOF) analysers are based on the measurement of the m/z of an ionised analyte (after their initial acceleration by an electric field) according to the time that ions take to move through a field-free region from the ion source to the ion detector (TOF). MALDI is normally coupled to a TOF analyser [91]. In quadrupole (Q) analysers, the ionised analyte moves with vibration through four parallel cylindrical rods subjected to an electric field. This analyser separates ions according to their m/z using the stability of the trajectories in oscillating electric fields [90]. Q spectrometers can be operated as a triple quadrupole Q-Q-Q and as a hybrid Q-Q-TOF. Tandem TOF instruments (TOF-TOF) and Q-Q-TOF instruments show high mass accuracy, resolution and sensitivity. Both can be operated in MS or MS/MS modes.

Ion traps (IT) use an oscillating electric field to store ions of different masses which are expelled according to their masses. This is contrary to quadrupole analysers where ions with a selected mass go through the rods. IT mass analysers exhibit a low resolution and low accuracy but a high sensitivity. These analysers are usually coupled to ESI. IT instruments can be combined with TOF analysers (IT-TOF). Orbitrap and fourier transform-ion cyclotron resonance (FT-ICR) analysers are both based on the fourier transform (FT) principle. Signal intensity, as a function of time is transformed (through a FT) into intensity as a function of frequency [90]. This generates an intensity to m/z association; mass spectra. Orbitrap is an electrostatic ion trap that uses FT [90]. It consists of mainly two electrodes, an outer electrode (barrel shaped) and a central electrode (spindle-shaped) where ions are radially trapped and their harmonic oscillations frequencies along the electric field (ions with different masses) allow the m/z measurements [94]. A FT-ICR spectrometer captures the ions under high vacuum in a static magnetic field and the m/z is measured by the frequency of circular cyclotron motion of the ions [91]. Orbitrap and FT-ICR have high accuracy and high resolution.

Detectors

Once the ions pass through the mass analyser, the detectors generate an electric current proportional to the ions abundances. In the first mass spectrometers, a photographic plate was used as a detector. In this type of detector, ions with the same m/z reach the plate at the same place and their abundances were measured based on the darkness of the spots. A Faraday cup (a metal cylinder) was also used in the beginning of mass spectrometry but it had low sensitivity.

Electron multipliers are the most widely used ion detectors. Ions from the mass analyser are accelerated to high velocity generating secondary particles which are converted to electrons and then amplified by a cascade effect in the electron multiplier to produce a signal [90]. This signal is processed by the computer and converted into a mass spectrum (m/z vs. relative abundances). The detection in Orbitrap and FT-ICR is incorporated in the mass analysers.

1.4.2. Liquid chromatography mass spectrometry and lipidomics

Liquid chromatography mass spectrometry (LC-MS) is an analytical chemistry technique used to separate compounds based on their interactions with the stationary and mobile phases, and subsequently their analysis by a mass spectrometer. In untargeted lipidomics, reverse phase LC is the most common type of chromatography used, as it allows the separation of individual lipids species and it is reproducible. This technique allows the separation of lipids based on their hydrophobicity. In reverse phase LC, the stationary phase is non-polar (hydrophobic) and the mobile phase is polar (aqueous). The aqueous mobile phase is mainly mixed with organic solvents, such as acetonitrile (ACN) or isopropanol (IPA). In this type of chromatography, the compounds which interact less with the hydrophobic stationary phase (but with high interaction with the mobile phase) will elute first, and the compounds with more affinity with the stationary phase (hydrophobic) will be strongly retained and will elute later. The elution times vary

with the length of the fatty acyl chains (number of C atoms) and degree of unsaturation (number and position of double bonds). Therefore, reverse phase LC-MS is a useful approach to study lipids content in a sample, allowing the study of a wide range of lipid classes as well as separation of lipid species (Figure 1.4.3.) [95]. The two main LC techniques used in lipidomics are high-performance LC (HPLC) and ultra-performance LC (UPLC ®). UPLC ® was introduced by Waters as an improvement of HPLC with higher analytical resolution and sensitivity.

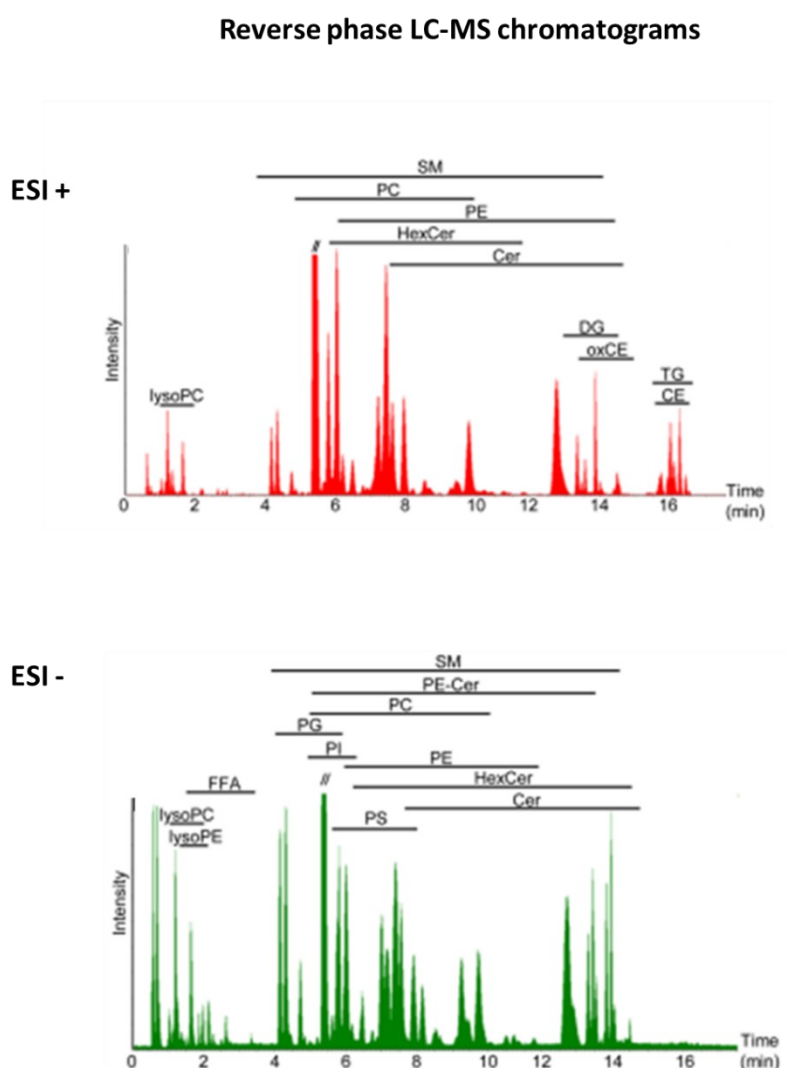


Figure 1.4.3. Lipids separation by reverse phase LC-MS. Figure adapted from Vorkas et al. 2015 [95].

In lipidomics, HPLC and UPLC systems are coupled to mass spectrometers with an ESI source. This allows the study of a wide range of lipid classes in both positive ESI mode (ESI+) or negative ESI mode (ESI-) (Figure 1.4.3). Cer and SM can be ionised in both ESI+ and ESI- modes, but SM is favoured in ESI+. Similarly, PC, PE, PS, and their lysophospholipid forms can be detected in both ESI+ and ESI-, producing mainly $[M+H]^+$ ions. Free fatty acids, PA, PI, PS, PG are mainly detected in ESI-, producing $[M+H]^-$ ions. Glycerolipids and sterol lipids (e.g. DG, TG, Cholesterol esters (CE)) are detected in ESI+.

1.4.3. Lipidomics and cancer

Lipidomics has contributed to advances in cancer research. Alterations in cancer lipid metabolism have been established as well as the potential of lipids identification for both breast and ovarian cancer diagnosis. The analysis of breast cancer cell lines by LC-MS has revealed differences in the lipid profile between breast cancer subtypes [96]. This study found an increase of TG in ER/PR+ cell lines as well as PC in TN breast cancer. While TG, PC and PE were found in a cell line HER2+. Analysis of breast cancer tissue samples has also demonstrated the potential of lipidomics analysis for cancer diagnosis using LC-MS and MS imaging (MSI) platforms. High levels of sphingolipids have been detected in breast cancer tissue when compared to normal breast tissue using LC-MS [97]. MALDI-MSI revealed differences in the lipid profile between breast cancer and adjacent normal tissue as well as between breast cancer subtypes [98]. Desorption electrospray ionisation (DESI)-MSI has also revealed changes in the fatty acid and phospholipid composition of tumour tissue and tumour-associated stroma when compared with normal glandular and stromal tissue [99]. Similar results were obtained by a multicentre study, confirming the reproducibility of this MS approach [100]. Rapid evaporative ionization mass spectrometry (REIMS) has also revealed changes in phospholipids and TG species between breast and normal cancer samples [101].

Similarly, LC-MS based approaches have been used for the study of lipid alterations in ovarian cancer. Lysophospholipids were identified as potential biomarkers for ovarian cancer diagnosis based on the analysis of blood plasma samples from patients with ovarian epithelial cancer [102]. Another study found that a combination of six metabolites between them two lysophosphatidylcholines allowed to discriminate between ovarian epithelial cancer and benign ovarian tumours [103]. Alterations in phospholipids and TG have also been observed in blood plasma from ovarian cancer patients [104]. Lipid alterations have also been revealed in ovarian cancer tissues by DESI-MSI [105]. This approach allowed the classification of samples in different epithelial ovarian cancer subtypes, as well as to distinguish between cancerous tumour and normal ovarian stroma.

Together these demonstrate the importance of lipodomic MS-based analysis, not only for the understanding of cancer biology, but also the potential of this type of analysis for the discovery of biomarkers for cancer diagnostics. MSI is a promising tool for cancer diagnosis and overcomes limitations in classical histopathology. However, access to tumour tissue is necessary. Cancer associated biomolecules can be found circulating in body fluids from cancer patients, such as blood and urine. Liquid biopsies will overcome the need to obtain tumour tissue for cancer diagnosis. But, human body fluids are complex matrices that contain numerous biomolecules and lipoproteins which vary between individuals and can mask the identification of cancer biomarkers. This could explain why numerous studies based on blood plasma analysis are difficult to be validated, as many identified lipids can be from lipoproteins, for example. Analysis of specific components of cancer patients' body fluids, like circulating cancerous cells or EVs, are a much more promising target for biomarker discovery. EVs are mainly composed of lipids, but an understanding of their lipid composition is limited, especially for those EVs produced by cancer cells, including breast and ovarian cancer-derived EVs. EVs' lipid composition will contribute massively to the understanding of the role of EVs in different pathologies as well as to understand their potential for the identification of biomarkers for cancer diagnosis and prognosis as well as their use as drug delivery vehicles.

1.5. Bioinformatics and statistical analysis

Untargeted UPLC-MS analysis produces complex data sets that require pre-processing to extract the information from the mass spectral data before statistical analysis can be carried out. Pre-processing of UPLC/MS data is mainly based on the use of algorithms for peak detection, peak alignment and area extraction as well as characterisation of the peaks in their mass, retention time and intensity. To translate MS data into a matrix of features, there are available copyrighted software from companies such as Waters Corp, or open-source tools such as XCMS [106], MZmine2 [107] and OpenMS [108].

Another key step in MS data pre-processing is the elimination of features corresponding to noise signal and the evaluation of the feature quality, as well as data normalisation, scaling and/or transformation. Data normalisation allows to remove systematic variation between samples due to factors that can mask the biological meaning of the datasets, like sample collection, concentration and instrument variability. Total ion count or total intensity normalisation, quantile normalisation and median fold change normalisation, also called probabilistic quotient normalisation (PQN) are among the most used normalisation methods in metabolomics and lipidomics analysis. PQN normalisation has been applied in metabolomics studies and has shown to be the method that is least compromised by biologically relevant changes in the mixture components [109, 110]. Scaling methods divide each variable by a scaling factor, like standard deviation or mean, which is different for each variable [111]. Auto-scaling (use standard variation) and pareto (the square root of the standard deviation) scaling are the most common methods in metabolomics. Auto-scaling has demonstrated itself to be a useful method in metabolomics analysis [111]. Data transformation, corrects for data heteroscedasticity which is common in metabolomics and lipidomics analysis. Log transformation is the most used method in metabolomics. Variance-stabilisation with respect to peak intensity can be achieved with log transformation in metabolomics studies [109]. However, the selection of a data normalisation, scaling

and/or transformation method is mainly data dependent, and it depends also on the data analysis to be used to answer the relevant biological questions.

In order to extract the biological information from the MS datasets and answer the research questions, univariate or multivariate statistical analysis can be performed. Descriptive analysis, such as mean and median can be calculated to measure the central tendency across the data set, and standard deviation and variance to measure the dispersion in the data set. For graphical representations of these analyses, for example, volcano plots and box plots can be generated. Volcano plots allow to identify changes in data sets based on the magnitude of the changes (fold changes) and the statistical significance. Boxplots show the distribution of the data indicating their quartiles, median, maximum and minimum values. This can be useful to find patterns between groups but at univariate level. In addition, the statistical significance of a variation of a specific feature in a data set can be tested by a *t*-test and analysis of variance (ANOVA), these tests determine whether there are statistically differences between groups or classes.

In addition, chemometrics techniques allow the extraction of relevant information from very large and complex datasets containing hundreds or thousands of variables, which is the case of the data sets generated by MS analysis. Multivariate statistical analyses are usually used in lipidomic analysis after MS data pre-processing. This type of analysis can be unsupervised or supervised analysis. In unsupervised analysis, there is not a previous knowledge of the sample groups studied. Principal component analysis (PCA) is an unsupervised analysis commonly used as a starting point in chemometric analysis. PCA is used to reduce the dimensionality of the dataset composed of multiple variables. PCA transforms the dataset to a coordinate system that exhibits in the first component, the greatest variation in the data, the second greatest variation in the second component, and so on [112]. This analysis shows an overview of the distribution and variation of the data, and allows to identify trends and outliers.

The dataset can also be analysed, taking into account the sample groups studied, using supervised analyses in order to study the relationship between groups. Partial least-squares (PLS) regression is widely used in high dimensional data sets for constructing predictive models. A variation of PLS, is PLS discriminant analysis (PLS-DA) used to study categorical sample groups. PLS-DA is capable of distinguishing between groups and within groups' variability. Therefore, when dimension reduction and discrimination is the main objective, PLS based analysis is preferred to PCA [113]. A variation to PLS-DA, is OPLS-DA, which is also used for discrimination, but it has the ability to separate predictive from non-predictive (orthogonal) variation, and the data interpretation is straightforward [114]. Machine learning algorithms can also be used for recursive feature elimination (RFE) with logistic regression (LR) analysis for the selection of a subset of relevant variables. This works by ranking features in the data set by importance and removing the less important predictors from the model [115]. The choice of a method for the analysis of data sets with multiple variables depends on the properties of the data but, most importantly on the biological questions to be answered.

Chapter 2

Hypotheses and Aims

2. Hypotheses and Aim

Cancer cells produce EVs which play an important role in cancer development and progression. They can be found in cancer patients' body fluids and be a source of lipid biomarkers for the development of non-invasive methods for cancer diagnosis.

Based on the current body of knowledge in this field as discussed in Chapter 1, the following hypotheses will be evaluated:

- Sphingolipids and glycerophospholipids are within the main lipids found in breast cancer and ovarian cancer-derived EVs.
- EVs secreted by breast and ovarian cancer cells exhibit a different lipid composition to EVs secreted by non-cancerous cells.
- EVs' lipid composition differs between breast cancer subtypes, as well as between ovarian cancer subtypes.

Aim

The aim of this PhD thesis was to study the lipid composition of breast and ovarian cancer-derived EVs and whether their lipid composition has potential for the identification of lipid biomarkers for breast and ovarian cancer diagnosis. In order to achieve this aim, the objectives of this project are:

1. To analyse the lipid composition of cancerous and non-cancerous mammary cells and their secreted EVs, by untargeted lipidomic analysis.

2. To analyse the lipid composition of cancerous and non-cancerous ovarian cells and their secreted EVs, by untargeted lipidomics analysis.
3. To investigate whether there are differences in the lipid composition between cancerous cells and EVs, and the non-cancerous cells and EVs.
4. To investigate whether there are differences in the lipid composition of EVs secreted by cancer subtypes.
5. To isolate EVs from human blood plasma for lipidomic analysis.
6. To analyse by untargeted lipidomic analysis, EVs found in blood plasma from breast cancer patients and healthy volunteers, and whether those EVs exhibit lipid biomarkers for breast cancer diagnosis.

This PhD thesis required the optimisation of state-of-the-art technologies for isolation of EVs from human blood plasma. I, Erika Dorado, developed a method for isolation of EVs from human blood plasma for effective application in lipidomic analysis. This required in depth research and testing of state-of-the-art technologies for isolation of EVs (including commercial kits and protocols developed by the Stevens group for the extraction of EVs from conditioned media) as well as the evaluation of their application for lipidomic analysis. The method developed for isolation of EVs from human blood plasma for lipidomic analysis is described in Chapter 5 and its application for the analysis of clinical samples is shown in Chapter 6.

Contribution by other people:

The transmission electron microscopy (TEM) analyses in this PhD thesis were carried out by trained members of the Stevens group. Cryogenic transmission electron microscopy (CryoTEM) was carried out for the analysis of the morphology of EVs isolated from conditioned media from mammary epithelial cells (Chapter 3) and TEM for the analysis of the morphology of EVs isolated from conditioned media from ovarian epithelial cells (Chapter 4) and EVs isolated from blood plasma (Chapter 5). Both techniques, TEM and CryoTEM, are effective for the analysis of EVs. These analyses were carried out at different times, and therefore the use of either TEM or CryoTEM was only dependent on the availability of the trained personnel at the time of the experiments.

In Chapter 3, only the isolation and protein analysis of EVs from conditioned media from mammary epithelial cells was carried out by the Stevens group, and therefore are included in the Appendix as supplementary Figure 1 and supplementary methods. Note: the Stevens group support was limited to only these specific points in Chapter 3 and not to the development of the method for the isolation of EVs from human blood plasma described in Chapter 5.

Chapter 3

Lipidomic analysis of mammary epithelial cells and their secreted EVs

3. Lipidomic analysis of mammary epithelial cells and their secreted EVs

3.1. Background

Breast cancer continues to be one of the most frequently occurring types of cancer among women [1]. A breast lump is the most common symptom of breast cancer, but patients can also experience other symptoms than a breast lump [13]. Self-examinations, clinical breast exams and imaging are the first steps for a breast cancer diagnosis, but in most cases a tissue biopsy using invasive methods is required if malignancy is suspected [21]. The development of effective and non- or minimally invasive diagnostic methods for breast cancer will bring significant benefits in the detection at an early stage and subsequently to an early access to treatment.

Cancer cells can produce EVs that play an important role in cancer development, progression and metastasis [81-83]. They have been found in cancer patients' body fluids, such as blood plasma and urine, and could be a source of biomarkers for cancer diagnosis [76, 80]. Most studies have focused on the study of proteins associated with cancer-derived EVs, but the study of cancer-derived EVs' lipid composition is still underexplored. To date, the lipid content of cancer-derived EVs has been studied mainly in prostate cancer by analysing cell lines and urinary EVs [74-76]. In breast cancer, only EVs from two cell lines derived from one metastatic cell line have been studied [78]. The study of urinary EVs from prostate cancer patients have suggested the potential of a combination of PS(18:1/18:1), PS(18:0_18:2) and LacCer(d18:1/16:0) to distinguish between healthy individuals and prostate cancer patients [76]. This demonstrates the importance of lipidomics analysis of cancer-derived EVs and its potential for the identification of lipid biomarkers for cancer diagnosis.

The aim of this chapter (Chapter 3) is to study the lipid composition of EVs obtained from different breast cancer subtypes (luminal, HER2+ and TN breast cancer), and its potential as a source of lipid biomarkers for breast cancer diagnosis.

3.2. Methods

Note: As stated in “contribution by other people” (page 57), in this Chapter 3 the only analyses carried out by the Stevens group were the EV’s isolation (point 3.2.2) and their characterisation (point 3.2.3) (the results are shown in Supplementary Figure 1 and Supplementary methods). The analyses were undertaken following the methods already published by the group [64, 116, 117].

3.2.1. Mammary cells evaluated

Nine breast cancer cell lines were evaluated (three biological replicates): two ER+/PR+ (MCF-7 and T47D), two HER2+ (HCC1954 and JIMT-1) and five TN breast cancer cell lines (MDA-MB-436, MDA-MB-231, MDA-MB-468, HCC1937 and Hs578T). In addition, normal mammary epithelial cells (MCF-10A and HuMEC, three biological replicates) were evaluated. All cell lines were obtained from LGC standards, except JIMT-1 (Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures). HuMEC was obtained from Thermo Fisher Scientific.

3.2.2. EVs’ isolation

A description of the mammary cells’ *in vitro* culture is shown in supplementary methods (Supplementary methods). For production of EVs, cells were expanded in 10 T225 cell culture flasks. Once flasks reached 80% confluency, cells were washed with DPBS and then maintained in a serum-free medium for 72 h, after which the

conditioned medium was harvested, filtered with a 0.45 µm bottle top filter and stored at – 80 °C until further processing. The conditioned medium was concentrated using 100 kDa MWCO Amicon® Ultra-15 centrifugal filter devices and centrifugation at 5,000 x g at 4 °C to a final volume of 500 µL. Subsequently, EVs were isolated by SEC using a Sepharose CL-2B (Sigma-Aldrich) column (1 cm x 30 cm) packed to approximately 28 cm and equilibrated using DPBS (Thermo Fisher Scientific). The concentrated conditioned medium was loaded onto the column and DPBS was used as the mobile phase. Peak fractions of 1 mL were eluted and collected for EVs analysis.

3.2.3. EVs' characterisation

The number and size distribution of EVs were measured by nanoparticle tracking analysis using a Nanosight 300 (Malvern). Five 60 sec videos were analysed for each sample. The EVs morphology was analysed by CryoTEM (Supplementary methods). EV associated protein markers were analysed by dot blotting. Dot blots were performed using the BioDot apparatus (BioRad) according to the manufacturer's instructions. To summarise, 0.45 µm nitrocellulose membranes incubated in TBS were loaded into the BioDot apparatus and hydrated with TBS. EV fractions were added and allowed to drain under gravity. Wells were washed three times with TBS. The BioDot apparatus was disassembled, and membranes blocked in 5% non-fat dry milk (BioRad) in TBS-T for 1 h at room temperature, followed by three 10 min washes in TBS-T. Membranes were incubated overnight at 4 °C with primary antibodies; mouse-anti-CD9, CD63 and CD81 (all Thermo Fisher Scientific) in a 1:1000 dilution in 5% BSA in TBS-T. Subsequently, membranes were washed three times in TBS-T and incubated for 1 h at room temperature with secondary antibody goat-anti-mouse (Thermo Fisher Scientific).

3.2.4. Lipid extraction

Lipids were extracted from EVs and cells based on the Blight and Dyer method [118]. To summarise, 3.75 mL of chloroform:methanol (1:2, v/v) were added to 1

mL of the sample homogenized in LC-MS grade water and vortexed. Samples were incubated for 30 min for the extraction of the lipids into the organic phase. After the incubation time, 1.25 mL of chloroform was added and the samples were vortexed. Then, 1.25 mL of LC-MS grade water was added and the mixture vortexed again. The samples were centrifuged at 1 000 x g for 10 min and two layers were formed; the chloroform layer (bottom layer) containing the lipids, and the aqueous or methanolic layer (top layer) containing non-lipids. In parallel “blank samples” were prepared by replacing the sample by 1 mL of LC-MS grade water. A washing step was included to obtain a much cleaner sample of lipids. The washing step consisted of interchanging the organic and aqueous layers between the “blank samples” and the samples, and then centrifuging at 1,000 x g for 10 min. After the final centrifugation, the organic layer was collected and dried using nitrogen. The lipids extracted from EVs and cells were reconstituted in IPA/ACN (1:1, v/v).

3.2.5. LC-MS analysis

Samples were analysed by UPLC-MS using an ACQUITY UPLC System coupled to a Waters XEVO G2 Q-TOF mass spectrometer (Waters). The liquid chromatographic separation of the compounds was carried out using an ACQUITY UPLC ®BEH C8 Column, 1.7 µm, 2.1 x 100 mm (Waters), at 55 °C. Mobile phase A consisted of H₂O:IPA:ACN (2:1:1, v/v/v), 5 mM ammonium acetate, 0.05% acetic acid and 20 µM phosphoric acid. Mobile B consisted of IPA:ACN (1:1, v/v), 5 mM ammonium acetate and 0.5% acetic acid. The binary gradient profile (mobile phase A and B) was 99.9% mobile phase A (0.0- 2.0 min), 70% A (2.0-11.5 min), 10% A (11.5-12.0 min), 0.1% A (12.0-12.50 min) and 99.9% A (12.50-14.6 min). The MS and MS/MS data were acquired for 0.1 s in the centroid mode in both ESI+ and ESI-, mass range 50 – 2000 Da. Mass spectrometer source parameters were as follow: capillary voltage 2 kV (ESI+) or 1.5 kV (ESI-), sampling cone voltage 25 V, source temperature 120 °C, desolvation temperature 600 °C, cone gas flow 150 (L/h) and desolvation flow 1000 (L/h). Quality Control (QC) samples containing a pool of all the samples evaluated and lipid internal standards (Supplementary Table

1) were included in the set of samples analysed, in order to assess the quality of the data throughout the whole running time. Data dependent acquisition (DDA) was achieved for QC samples to perform MS/MS experiments in order to identify lipid species. A collision energy ramp of 22 - 37 V was applied for the fragmentation of the parent ions for both electrospray ionisation modes. In addition, individual MS/MS experiments were also carried out for the most relevant features.

3.2.6. Data analysis

A workflow for the data analysis of this PhD thesis was developed as indicated below and included the use of in-house and open-source scripts based on R, MATLAB and Python programming languages which were adapted for the data analysis of this PhD thesis. MS data pre-processing was undertaken using XCMS open-source software [106]. The XCMS outputs included m/z values, retention time and intensities for each m/z value identified. Sample intensities were normalised by probabilistic quotient normalisation for univariate analysis. For multivariate analysis, the intensities were also log transformed. Compounds with coefficients of variation (CV%) in QC samples higher than 30% (non-reproducible measurements) were removed from the outputs. Univariate analyses were carried out to identify the lipid compounds significantly enriched in the breast cancer cells and their EVs. Log2 fold change was calculated based on the mean of the normalised abundance for each group studied (i.e. cells and EVs). One-way ANOVA was performed to generate a p -value for each detected compound (m/z values) and a threshold of statistical significance was set at lower than 0.05. Spearman rank-order correlation coefficient was calculated for paired cells and their EVs to identify lipids correlated between these two groups (i.e. cells and EVs). RFE-LR analysis was performed for the selection of a subset of the most relevant features for the analysis between groups (i.e. cancer and no-cancer, or between the three types of breast cancer subtypes studied). To evaluate the performance of the RFE-LR model a leave-one-group cross-validation analysis (biological replicates were defined as a group) was performed. P -values calculated for the most relevant features identified were adjusted by Benjamini-Hochberg correction (q -values) and a threshold of statistical

significance was set at lower than 0.05 (q -value < 0.05). The Area Under the Receiver Operating Characteristic (AUROC) curve was measured to test the accuracy of the most relevant features identified for the discrimination between cancer and non-cancerous samples (EVs and cells). The LC-MS datasets obtained by ESI+ and ESI- modes were fused by concatenation of the variables and auto-scaled for RFE-LR analysis. Box plots of the fused data are based on the auto-scaled data. Univariate and multivariate analyses were carried out using Python programming language and MATLAB. Lipid annotation were performed using in-house databases, matching with online lipid databases (LIPID MAPS, HMDB and METLIN) and by the manual verification of MS/MS spectrum. The lipids annotated for EVs and cells in this study (Chapter 3) are indicated in results and discussion section (3.3) and contributed to the creation of an in-house lipid database for EVs.

3.3. Results and Discussion

Lipidomic analysis of cancerous and non-cancerous human mammary epithelial cells and their EVs were carried out by untargeted LC-MS, in both ESI+ and ESI- modes (Figure 3.3.1 and Supplementary Figure 2). This methodology allowed to analyse a wide range of lipid classes in order to understand the lipid composition of breast cancer-derived EVs. A group of nine breast cancer cell lines were studied. This included ER+/PR+, HER2+ and TN breast cancer cell lines. EVs isolated from the conditioned media of the mammary epithelial cells by SEC, were positive for the EV protein markers CD9, CD63 and CD81 and had a vesicular morphology (Supplementary Figure 1).

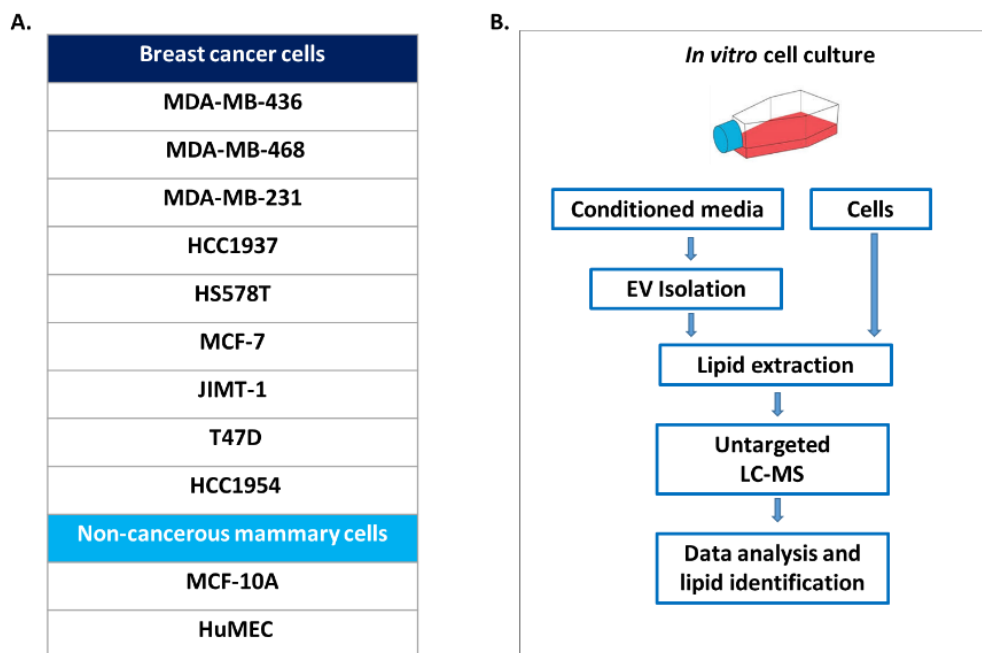


Figure 3.3.1. Workflow for the lipidomic analysis of mammary cells and their secreted EVs. **A.** List of cancerous and non-cancerous human mammary epithelial cells studied. **B.** Workflow used for the lipidomic analysis of the mammary epithelial cells and their EVs.

Lipidomic analysis of breast cancer cells and their secreted EVs

Initially, PCA were carried out to understand the distribution of the LC-MS data in both ESI+ and ESI- (Figure 3.3.2A). This unsupervised analysis included all the cancerous and non-cancerous cells and EVs. The PCA showed a separation between EVs and cells based on their lipid compositions. A volcano plot was performed to display the lipid species with the largest changes between breast cancer cells and their secreted EVs. This graph also shows significantly differences between these two groups (i.e. cells and EVs) (Figure 3.3.2B). Forty three lipid species were found to be significantly (p -value < 0.05) enriched ($\log_2$ fold change > 2) in breast cancer-derived EVs when compared to their parental cells (Figure 3.3.2B and Table 3.3.1).

lipid enrichment between EVs and cells, respectively. Statistically significant differences between the means of EVs and cells were calculated by one-way ANOVA with a p -value less than 0.05 as statistically significant.

Table 3.3.1. List of lipids species enriched in EVs when compared to their parental cells. The table shows the list of the 43 lipids and their m/z values identified as significantly (p -value < 0.05) enriched ($\log_2$ fold change > 2) in breast cancer-derived EVs when compared to their parental cells.

m/z	Theoretical m/z	Lipid ID	EV/Cells	ANOVA p -value	Log ₂ FC EV/Cells
881.5175	881.5186	PI(38:6)	EV	3.10E-05	2.474
582.3770	582.3777	LPC(18:0)	EV	3.57E-05	2.748
630.4479	630.4469	LPC(24:0)	EV	9.57E-06	3.276
508.3395	508.3409	LPE(20:0)	EV	2.19E-07	3.369
782.5727	782.5694	PC(36:4)	EV	8.88E-07	2.215
812.6105	812.6164	PC(38:3)	EV	1.27E-06	2.427
904.7028	904.7012	PC(40:0)	EV	9.27E-08	3.897
840.6406	840.6477	PC(40:3)	EV	2.11E-07	3.041
930.7168	930.7169	PC(42:1)	EV	8.78E-06	2.601
868.6721	868.679	PC(42:3)	EV	2.28E-09	5.251
688.4995	688.4923	PE(32:1)	EV	9.79E-08	2.135
824.6136	824.6175	PE(42:3)	EV	1.29E-09	2.592
962.7571	962.7583	PE(52:4)	EV	9.84E-07	2.865
702.5405	702.5443	PE(O-34:1)	EV	1.03E-02	2.575
820.6432	820.6437	PE(O-38:0)	EV	1.33E-06	2.627
510.4882	510.4892	Cer(32:0)	EV	8.42E-05	2.768
508.4726	508.4735	Cer(32:1)	EV	1.91E-04	2.034

522.4883	522.4892	Cer(33:1)	EV	1.67E-05	2.701
548.5047	548.5048	Cer(35:2)	EV	6.30E-08	2.238
564.5358	564.5361	Cer(36:1)	EV	1.31E-04	2.158
562.5200	562.5205	Cer(36:2)	EV	8.66E-05	2.075
594.5830	594.5831	Cer(38:0)	EV	1.19E-04	2.073
592.5693	592.5674	Cer(38:1)	EV	1.70E-06	2.953
590.5514	590.5518	Cer(38:2)	EV	2.23E-05	3.337
622.6128	622.6144	Cer(40:0)	EV	1.26E-10	2.611
620.5987	620.5987	Cer(40:1)	EV	9.77E-12	2.248
618.5831	618.5831	Cer(40:2)	EV	7.79E-10	2.348
650.6454	650.6457	Cer(42:0)	EV	1.65E-12	2.171
664.6642	664.6613	Cer(43:0)	EV	3.21E-05	3.256
662.6456	662.6457	Cer(43:1)	EV	5.89E-09	2.456
678.6767	678.677	Cer(44:0)	EV	5.68E-04	2.638
692.6923	692.6926	Cer(45:0)	EV	1.54E-05	2.011
720.7299	720.7239	Cer(47:0)	EV	2.30E-03	2.286
659.5127	659.5134	SM(32:1)	EV	1.54E-11	3.303
673.5272	673.529	SM(33:1)	EV	1.33E-16	2.003
687.5440	687.5447	SM(34:1)	EV	2.10E-08	3.905
685.5280	685.529	SM(34:2)	EV	3.85E-11	2.557
741.5908	741.5916	SM(38:2)	EV	3.41E-07	3.169
759.6301	759.6386	SM(39:0)	EV	1.14E-05	2.286
903.7483	903.7536	SM(44:0)	EV	2.17E-02	2.289
901.7356	901.7379	SM(44:1)	EV	7.84E-03	2.031
756.6335	756.6359	HexCer(38:0)	EV	1.72E-03	2.019
794.5997	794.5907	HexCer(38:1)	EV	4.71E-04	2.689

Table 3.3.2. Lipids enriched in breast cancer cells when compared to their secreted EVs. This table shows the list of lipids species enriched in cells when compared to their secreted EVs. The table shows the list of the 46 lipids and their *m/z* values identified as significantly (*p*-value < 0.05) enriched ($\log_2$ fold change > 2) in breast cancer cells when compared to their secreted EVs.

<i>m/z</i>	Theoretical <i>m/z</i>	Lipid ID	EV/Cells	ANOVA <i>p</i> -value	$\log_2$ FC EV/Cells
307.2643	307.2632	FA(20:2)	Cells	1.62E-07	-3.115
303.2330	303.2330	FA(20:4)	Cells	1.13E-06	-3.041
329.2486	329.2475	FA(22:5)	Cells	3.82E-11	-2.799
407.2956	407.2953	FA(28:8)	Cells	7.04E-03	-2.374
305.2486	305.2476	FA(20:3)	Cells	1.34E-13	-2.759
809.5186	809.5112	PI(32:0)	Cells	3.60E-04	-2.037
860.6164	860.6130	PC(42:7)	Cells	2.50E-02	-2.320
938.6975	938.7059	PC(44:1)	Cells	1.67E-05	-2.948
560.2994	560.2971	LPE(20:4)	Cells	1.09E-06	-2.739
525.3198	525.3107	LPG(19:0)	Cells	3.48E-12	-2.086
770.5694	770.5740	PE(38:3)	Cells	6.16E-04	-2.159
824.5447	824.5421	PE(38:5)	Cells	4.41E-08	-2.099
822.5291	822.5353	PE(38:6)	Cells	8.29E-05	-2.571
788.6175	788.6244	PE(39:0)	Cells	7.15E-05	-2.151
802.6331	802.6390	PE(40:0)	Cells	1.48E-08	-2.184
792.5538	792.5406	PE(40:6)	Cells	2.66E-02	-2.013
741.4712	741.4694	PG(34:4)	Cells	3.62E-09	-2.085
833.4729	833.4652	PG(38:6)	Cells	4.39E-03	-4.618
863.5199	863.5121	PG(40:5)	Cells	6.17E-03	-4.641
907.6399	907.6416	PG(44:3)	Cells	9.63E-04	-2.660
853.5719	853.5728	PG(O-40:3)	Cells	2.46E-04	-3.203

772.4525	772.4453	PS(32:1)	Cells	4.73E-03	-2.099
788.5447	788.5425	PS(36:1)	Cells	6.54E-07	-2.080
810.5291	810.5260	PS(38:4)	Cells	3.14E-08	-2.346
716.7279	716.7350	Cer(48:1)	Cells	2.39E-02	-2.018
730.6191	730.6156	HexCer(36:0)	Cells	5.82E-04	-2.843
766.6919	766.6920	TG(44:1)	Cells	1.64E-02	-3.629
822.7545	822.7554	TG(48:1)	Cells	1.70E-02	-3.742
812.6763	812.6857	TG(48:6)	Cells	4.25E-03	-2.628
850.7858	850.7869	TG(50:1)	Cells	1.38E-02	-3.582
848.7702	848.7705	TG(50:2)	Cells	4.55E-04	-2.948
876.8015	876.8025	TG(52:2)	Cells	2.23E-04	-2.852
874.7858	874.7863	TG(52:3)	Cells	9.18E-06	-2.319
868.7389	868.7326	TG(52:6)	Cells	3.43E-03	-2.994
904.8328	904.8335	TG(54:2)	Cells	3.51E-05	-2.356
926.8171	926.8174	TG(56:5)	Cells	1.68E-05	-2.431
924.8015	924.8001	TG(56:6)	Cells	1.01E-04	-2.206
958.8797	958.8782	TG(58:3)	Cells	1.54E-05	-2.156
956.8641	956.8638	TG(58:4)	Cells	2.45E-06	-2.660
952.8328	952.8334	TG(58:6)	Cells	1.78E-05	-2.476
950.8171	950.8172	TG(58:7)	Cells	8.91E-04	-2.859
986.9110	986.9099	TG(60:3)	Cells	6.45E-05	-2.113
984.8954	984.8931	TG(60:4)	Cells	5.05E-07	-2.395
980.8641	980.8669	TG(60:6)	Cells	7.81E-04	-2.545
978.8484	978.8483	TG(60:7)	Cells	1.33E-02	-3.639
976.8328	976.8326	TG(60:8)	Cells	1.92E-03	-3.487

Sphingolipids (Cer, SM and HexCer) and glycerophospholipids (LPC, LPE PC, PE, PI and PS) have also been found enriched in EVs secreted by prostate cancer cell lines [74]. Enrichment of sphingolipids (Cer, SM and HexCer) have also been observed in other cancer cells-derived EVs studies [74-76, 119], including EVs derived from TN breast cancer cells [78]. Similarly, high levels of sphingolipids have been detected in the MS analysis of breast cancer tissues [97]. SM have been found enriched in EVs released by a wide variety of cells, including B-cells [69], prostate epithelial cells [73], T lymphocyte cells (Jurkat cells) [70], T cell blasts [70], human dendritic cells [71], TN breast cancer cells [78], prostate cancer cells [74, 75] and urinary EVs from prostate cancer patients [76]. The study of one prostate cancer cell line [75], and the study of urinary EVs from prostate cancer patients [76], also found sphingolipids enriched in EVs compared to cells, and the main glycerophospholipid found in EVs was PS [75]. Contrary to that study, in this study (Chapter 3) PS were identified in both EVs and cells but not within the most enriched lipids in EVs. A study of EVs found in blood plasma using both MS and nuclear magnetic resonance analyses did not find PS in EVs [120]. This suggests the idea that although PS can be part of the lipid composition of EVs, they are not necessarily enriched in all EVs secreted by cancer cells. Sphingolipids as indicated above are the most characteristic and enriched lipids found in EVs, to date.

Specific species of PC and PE and their lyso-forms were also found within the lipids enriched in EVs when compared to cells. Highly abundant PC and PE have also been observed in breast cancer tissues [121, 122], and EVs as mentioned above. PC and PE are essential structural lipids but also as a source of signalling molecules. PE are important in cells curvature facilitating cells budding, fission and fusion [41]. Lysophospholipids can act as signalling molecules, but they can also play a role in the EVs' curvature [123]. Lysophospholipids have also been identified in prostate cancer-cells derived EVs and EVs found in human blood plasma [74, 75, 120].

Enrichment of sphingolipids in EVs has been associated with the presence of lipid rafts in EVs [73, 124]. Lipid rafts which are rich in sphingolipids, such as Cer and

SM, are involved in cell signalling process [42], and play an essential role in cancer progression and the PI3K/AKT pathway that regulates cell proliferation and survival [49, 125]. C1P and S1P which can be produced from Cer and sphingosine, respectively, promote cell growth and proliferation [48]. The role of S1P in the interaction between cancer cells and the tumour microenvironment has already been suggested by the analysis of breast cancer tissues [97]. Lipid rafts in EVs could play a key role in EV's participation in intercellular communication as well as tumorigenesis and the establishment of metastatic niches. In addition, participation of Cer in the release of EVs has been suggested [126]. However, the same results were not observed in EVs release by PC-3 cells suggesting that this mechanism could be cell specific [127]. SM enrichment could indicate a close relationship between the Golgi apparatus and the EV populations as SM are exclusively synthesized in the Golgi (in contrast to other membrane lipids, which are generally produced in the ER). Ether PC and ether PE were also identified in EVs which together with SM could play an important role in the stability and organisation of the lipid rafts in EVs as well as in cell signalling processes [128]. Ether phospholipids have also been identified in EVs released by prostate cancer cell lines [75] and urinary EVs from prostate cancer patients [76].

PC, PE, LPE, Cer and HexCer species were found in both cells and EVs. In comparison to EVs, cells were enriched in TG when compared to EVs (Figure 3.3.2 and Table 3.3.2). This latter finding is not surprising as these lipids are not building blocks for biological membranes, hence not expected to contribute to the lipid composition of EVs. In addition, FA and PG species were found with higher fold changes in cells when compared to EVs. Cells and their secreted EVs can differ in their lipid composition due to the fact that EVs are largely produced by multivesicular bodies, plasma membrane and the Golgi apparatus, and consequently do not represent the entirety of cellular components. The contribution of the typical lipids in organelles' membranes such as mitochondrial or nuclear membranes to EVs lipidome is negligible. Breast cancer cells with a high content of TG [96] and FA have already reported [99]. Active cell proliferation is fundamental in cancer development and progression which require the continuous supply of FA [44, 129].

Arachidonic acid (AA) was one of the FA found to be enriched in EVs when compared to cells. AA is important in cell proliferation and cancer by activation of PI3K/AKT/mTOR signalling pathway [131, 132]. In addition, AA can be produced from PC by phospholipases (A2, C and D).

As shown above, there are significant differences in the lipid composition of EVs and cells, and specific lipids can be enriched in EVs compared to their parental cells. But also, there are lipid species associated between EVs and cells. In order to identify lipids highly correlated between EVs and cells, cells and their respective secreted EVs were paired and Spearman rank-order correlation coefficients calculated (Figure 3.3.3 and Table 3.3.3). Ether phospholipids showed a very strong (Spearman $r_s > 0.8$) correlation between EVs and cells (Figure 3.3.3 and Table 3.3.3). As mentioned before, ether lipids could be involved in the stability of lipid raft in EVs, but also in cellular membranes. In addition, a high level of ether lipids has been observed in cancer cells and their participation in cell differentiation and signalling pathways have been suggested [128, 130]. In addition, SM, HexCer, PC and PE species were also identified within those lipids which had a very strong relationship between cells and EVs. This suggests that cells and their secreted EVs' lipid composition is related to some extent and that the lipid arrangement in EVs could be connected to their cellular origin, but also to the cancerous origin.

Correlation analysis EVs-Cells

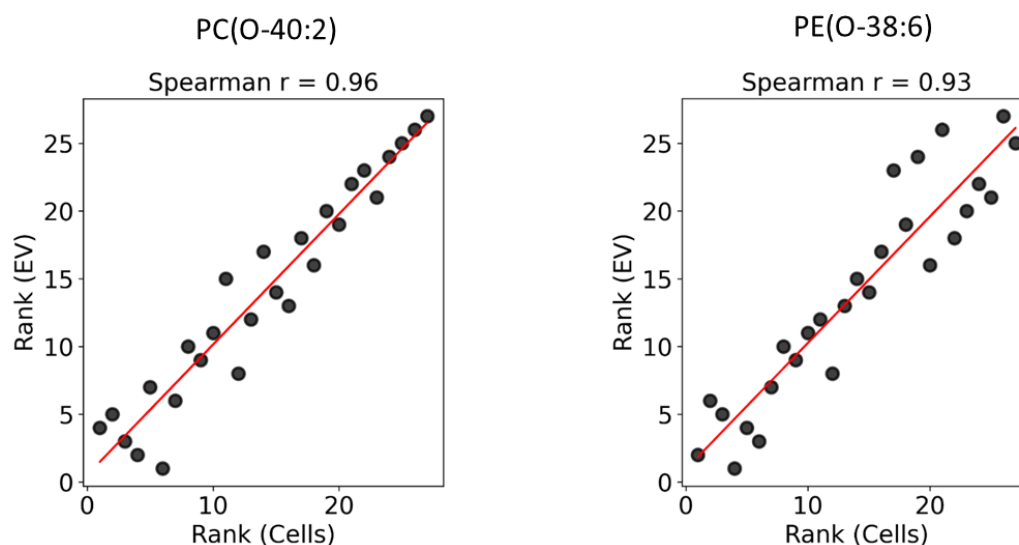


Figure 3.3.3. Lipid species highly correlated between EVs and cells. Spearman rank-order correlation coefficient calculated for paired cells and their EVs identified two ether phospholipid species as the main two lipids highly correlated between EVs and their parental cells, PC(O-40:2) and PE(O-38:6).

Table 3.3.3. Lipids with very strong correlation between EVs and their parental cells. List of lipids species with a Spearman rank-order correlation coefficient $r > 0.8$ indicating very strong correlation between EVs and their parental cells, with a p -value < 0.05 indicating that the correlation is statistically significant.

<i>m/z</i>	Theoretical <i>m/z</i>	Lipid ID	Spearman r_s	Spearman p -value
886.6907	886.6993	PC(O-40:2)	0.9627	1.03E-15
808.5498	808.5468	PE(O-38:6)	0.9334	1.24E-12
686.5130	686.5128	PE(O-33:2)	0.9273	3.60E-12
780.6359	780.6357	HexCer(40:2)	0.9242	5.92E-12
678.5432	678.5414	PE(O-32:0)	0.9145	2.56E-11

810.6828	810.6826	HexCer(42:1)	0.9053	8.68E-11
859.6910	859.6909	SM(41:1)	0.9053	8.68E-11
782.6515	782.6502	HexCer(40:1)	0.9047	9.37E-11
742.5745	742.5688	PC(O-34:3)	0.9010	1.47E-10
672.4974	672.4971	PE(O-32:2)	0.9004	1.58E-10
698.5576	698.5570	HexCer(34:1)	0.8888	5.89E-10
884.6750	884.6827	PC(O-40:3)	0.8888	5.89E-10
782.5342	782.5319	PE(O-36:5)	0.8870	7.15E-10
674.5130	674.5128	PE(O-32:1)	0.8864	7.62E-10
844.6801	844.6862	PE(43:0)	0.8835	1.03E-09
646.4817	646.4813	PE(O-30:1)	0.8546	1.40E-08
834.5655	834.5639	PE(O-40:7)	0.8539	1.48E-08
772.5287	772.5284	PE(O-40:8)	0.8504	1.95E-08
734.6058	734.6434	PE(O-36:0)	0.8461	2.70E-08
676.4923	676.4916	PE(31:0)	0.8424	3.55E-08
724.5276	724.5207	PE(O-36:5)	0.8388	4.64E-08
662.4766	662.4761	PC(28:0)	0.8357	5.76E-08
722.4978	722.4971	PE(30:0)	0.8321	7.43E-08
768.4821	768.4867	PE(34:5)	0.8321	7.43E-08
854.6644	854.6726	PE(44:2)	0.8321	7.43E-08
780.5913	780.5906	PE(O-40:4)	0.8247	1.22E-07
706.5745	706.5701	PE(O-34:0)	0.8247	1.22E-07
856.6801	856.6895	PE(44:1)	0.8235	1.32E-07
645.4501	645.4531	PA(32:1)	0.8229	1.37E-07
756.5902	756.5845	PE(O-38:3)	0.8186	1.80E-07
750.5443	750.5439	PE(O-38:5)	0.8174	1.94E-07
790.5968	790.5960	PE(O-36:1)	0.8144	2.35E-07
920.6386	920.6306	PC(42:6)	0.8119	2.73E-07

736.5498	736.5483	PE(O-32:0)	0.8105	2.97E-07
696.4974	696.4970	PE(O-34:4)	0.8095	3.16E-07
750.5655	750.5640	PC(O-30:0)	0.8034	4.53E-07
804.6124	804.6115	PC(O-34:1)	0.8028	4.69E-07

Cancerous mammary cells and their secreted EVs can be distinguished from the non-cancerous mammary cells and their secreted EVs based on their lipid composition

The analysis of the lipid composition of non-cancerous mammary epithelial cells and their EVs was also included in this study. This was in order to understand whether there are differences in the lipid composition between cancerous and non-cancerous mammary epithelial cells and EVs. Figure 3.3.4A showed the same PCA as in figure 3.3.2A for the ESI+ and ESI- modes, but in figure 3.3.4A the non-cancerous and cancerous EV-Cells are coloured yellow and purple respectively to highlight their LC-MS data distribution. However, as mentioned before, the PCA of the LC-MS data shows a main separation between cells and EVs.

For the analysis of cancerous and non-cancerous mammary epithelial cells and EVs, the MS data from the ESI+ and ESI- modes were fused and auto scaled to ensure equal contribution of features in both datasets. For this fused MS datasets, a RFE-LR analysis was performed to select a subset of the most relevant lipid species between cancerous and non-cancerous samples. This model allowed the identification of six statistically significant (q -value < 0.05) lipid species which distinguish cancerous from non-cancerous EVs and Cells (Figure 3.3.4B). These relevant lipid species included PE(O-18:2_22:6), PE(18:2_22:3), PE(20:3_24:1), SM(40:0), PC(O-16:0_16:1) and PC(18:1_18:0). To evaluate the performance of the generated model a leave-one-group cross-validation analysis (biological replicates were defined as a group) was performed. This showed a perfect classification (100% accuracy) of EVs and Cells based on their non-cancerous or

cancerous origin (Figure 3.3.4C). Moreover, a ROC analysis based on the combination of these relevant lipids showed an AUROC equal to 1, indicating the usefulness of these lipid species to distinguish between cancerous or non-cancerous samples (Figure 3.3.4D).

These findings show the potential of the EVs' lipid composition for the identification of lipid biomarkers for breast cancer diagnosis. This reflects the fact that alterations in lipid metabolism are often found in breast cancer cells and tissues when compared to normal cells and healthy tissues, respectively [96-101]. PE and SM were within the relevant lipids identified. High levels of PE have also been observed in breast cancer tissues [122]. PE can be precursors of PC, which in turn are precursors of second messengers including PA, DG and AA that participate in activation of signalling pathways. These include the PI3K/AKT/mTOR pathway whose over-activation contribute to cell proliferation, growth, and cell migration in tumour cells [131, 132]. Additionally, dysregulation of sphingolipid metabolism has been observed in breast cancer [97, 133]. SM is broken down by sphingomyelinases to generate Cer that subsequently is deacylated by ceramidase to generate sphingosine whose phosphorylation by sphingosine kinase form S1P, that contributes to cancer growth and tumour metastasis [134-136]. Cer is converted to C1P by Cer kinase, and as mentioned before C1P and S1P are involved in cell growth and proliferation [48]. Moreover, extracellular S1P can regulate the tumour microenvironment [97, 135, 136].

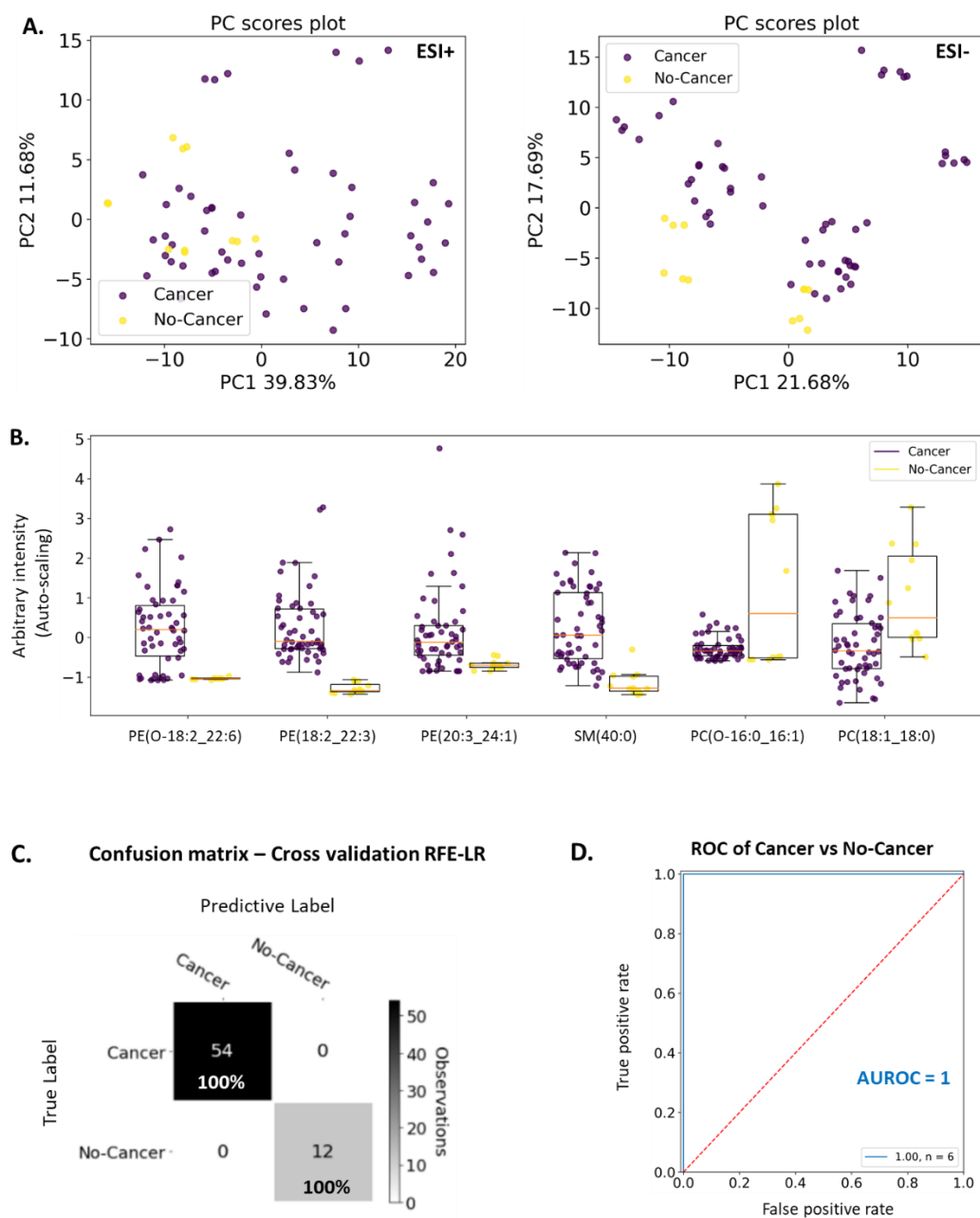


Figure 3.3.4. Statistical analysis of cancerous and non-cancerous EVs and cells. **A.** PCA showing the LC-MS data distribution of cancerous (Cancer) and non-cancerous (No-Cancer) EVs and cells. This is the same PCA as in figure 3.3.2A for the ESI+ and ESI- modes, but in this PCA the non-cancerous (yellow) and cancerous (purple) EVs and cells are indicated to highlight their data distribution. **B.** Box plots for the six lipid species identified by the statistical model obtained by RFE-LR that distinguish breast cancer cells and their derived

EVs from non-breast cancer cells and their derived EVs. The statistical model and box plots were based on the fused and auto scaled ESI+ and ESI- datasets. **C.**

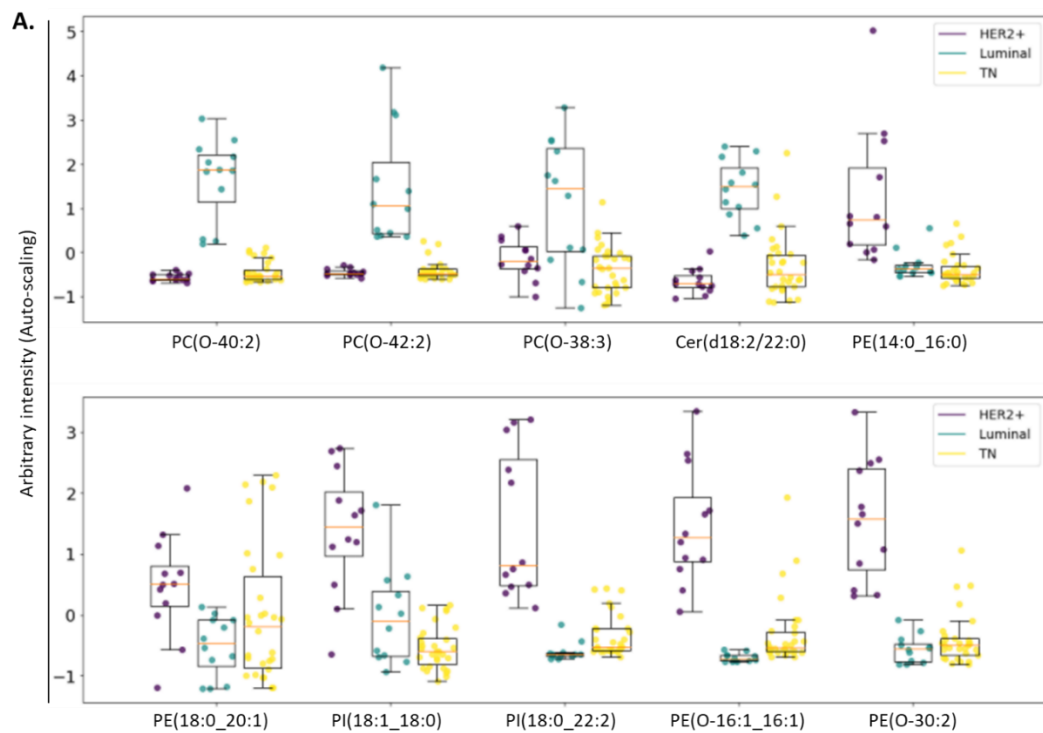
Leave-one-group cross-validation analysis was performed to evaluate the classification model. This analysis shows that EVs and cells can be perfectly classified based on their cancerous (Cancer) or non-cancerous (No-cancer) origin with an overall accuracy of 100%. **D.** AUROC curve for the combination of the relevant lipid species was equal to one indicating their usefulness for the discrimination between non-cancerous and cancerous samples.

Breast cancer subtypes can be distinguished between them based on the lipid composition of the cells and EVs

The analysis of breast cancer subtypes was also based on LC-MS data fused for the ESI+ and ESI- modes. For these fused LC-MS datasets a RFE-LR analysis was also performed to select a subset of the most relevant lipid species between the three breast cancers subtypes studied. The model for the analysis of breast cancer subtypes allowed the identification of ten statistically significant (q -value<0.05) lipid species contributing to the classification of the cancerous cells and EVs in the three breast cancer subtypes: luminal (ER+/PR+), HER2+ and TN breast cancer (Figure 3.3.5A). These relevant lipid species included PC(O-40:2), PC(O-42:2), PC(O-38:3), Cer(d18:2/22:0), PE(14:0_16:0), PE(18:0_20:11), PI(18:1_18:0), PI(18:0_22:2), PE(O-16:1_16:1) and PE(O-30:2). To evaluate the performance of this classification, a leave-one-group cross-validation analysis was performed. This showed an overall accuracy > 90% for classification of the EVs and Cells according to their breast cancer subtype (Figure 3.3.5B).

This is in agreement with other studies that have suggested a differentiation of breast cancer subtypes based on their lipid profile [98, 100]. Although the number of cell lines per breast cancer subtype is low and the test can be underpowered because of this, the results suggest that cancer-derived EVs carry phenotype information from the cancer subtype from which they were originated. We found

that EVs from ER+/PR+ cell lines in comparison with HER2+ and TN breast cancer cell lines showed an increase of ether phospholipids and Cer. HER2+ and TN breast cancer cell lines which are ER-/PR- showed an increase of PE and PI species, but also ether phospholipids. This could suggest that an increase or decrease of specific phospholipid species could be related to the expression or not of the ER and/or PR hormone receptors in cells and therefore to breast cancer aggressiveness. Dysregulation of the PI3K pathway is frequently found in breast cancer with mutations being identified in the different subtypes of breast cancer [137-139]. PI can be phosphorylated to form phosphoinositides, including PIP3 which play a key role in PI3K/AKT/mTOR pathway activation.



B.

Confusion matrix – Cross validation RFE-LR

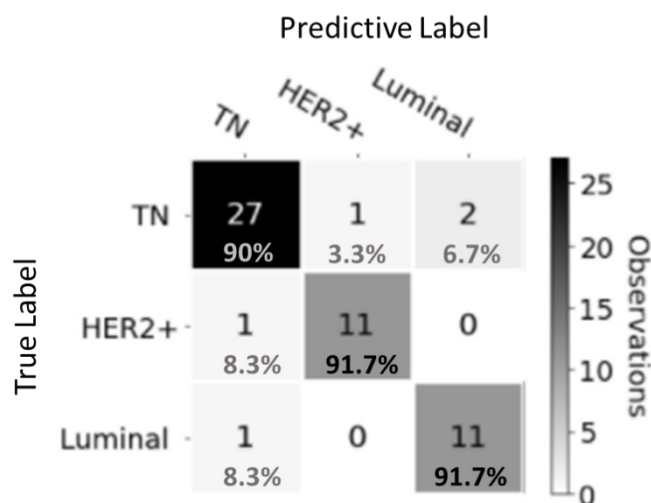


Figure 3.3.5. Statistical analysis of cells and EVs obtained from different breast cancer subtypes. A. Box plots for the ten lipid species identified by the statistical model obtained by RFE-LR that classify breast cancer cells and their derived EVs into their respective breast cancer subtypes based on their lipid profile. The statistical model and box plots were based on the fused and auto scaled ESI+ and ESI- datasets. **B.** Confusion matrix of the leave-one-group cross-validation analysis carried out to evaluate the performance of the classification model for breast cancer subtypes. This analysis shows that cancerous EVs and cells can be classified in their respective breast cancer subtype (luminal (ER+/PR+), HER2+ or TN breast cancer subtype) with an overall accuracy > 90%.

In conclusion, this comprehensive study of cells and EVs from different breast cancer subtypes show that breast cancer-derived EVs are mainly enriched in sphingolipids and that glycerophospholipids are also one of the main molecular lipids found in EVs. More importantly, the findings in this study demonstrate that there is potential in the lipid composition of breast cancer-derived EVs for the identification of lipid biomarkers for breast cancer diagnosis and that it is possible

to classify EVs into the different breast cancer subtypes based on their lipid profile, which have a great impact in breast prognosis.

Chapter 4

Lipidomic analysis of ovarian epithelial cells and their secreted EVs

4. Lipidomic analysis of ovarian epithelial cells and their secreted EVs

4.1. Background

Ovarian cancer is one of the leading causes of cancer deaths in women. It is considered to be the most lethal gynaecological cancer as it is mostly diagnosed at a late stage. Symptoms are ambiguous and commonly confused with other pathologies different to cancer. In addition, early-stage ovarian cancer patients are generally asymptomatic. CA125 detection in blood serum is usually recommended to patients with suspect ovarian cancer. However, the use of this approach is limited by the fact that CA125 is not expressed by all ovarian cancers and it is not specific for this disease as it can also be detected in pregnant women or those with gynaecological pathologies such as endometriosis and fibrosis. Additionally, transvaginal sonography is a common diagnostic imaging tool used mainly when patients have symptoms which are in the late stage of the disease. Patient survival depends on early detection of ovarian cancer and therefore, identification of molecular biomarkers for this heterogeneous type of cancer are needed.

In the previous study (Chapter 3), it was found that the lipid composition of breast cancer derived EVs has potential for the identification of lipid biomarkers for cancer diagnosis but also for the classification of breast cancer subtypes. Breast cancer and its subtypes have been explored more deeply than ovarian cancer, which is still a developing field. Breast cancer diagnosis studies can serve as a platform for the study other cancer types.

As mentioned in Chapter 3, the lipid composition of EV's secreted by cancer cells is an underexplored topic in the EV's field. To date, there is not a comprehensive study of the lipid composition of ovarian cancer-derived EVs, when compared to their parental cells. There is only a study of the lipid composition of one endometrioid ovarian cancer cell line when compared to non-cancerous ovarian

surface epithelial cells, which found mainly sterol lipids in EVs [79]. The aim of this chapter is to study the lipid composition of ovarian cancer- derived EVs and its potential as a source of lipid biomarkers for ovarian cancer diagnosis. High-grade serous ovarian cancer cells (the most common subtype of ovarian cancer in humans), endometrioid ovarian cancer cells and ovarian surface epithelial cells were studied.

4.2. Methods

4.2.1. Cell culture of epithelial ovarian cells

Two high grade serous (PEO4 and OVCAR3) and two endometrioid human ovarian cancer cell lines (A2780 and SKOV3) were maintained in RPMI media 1640 (ThermoFisher, Gibco) supplemented with 10% foetal bovine serum and 1 x penicillin/streptomycin. In addition, human ovarian surface epithelial cells (HOSEpiC) were also evaluated (ScienCell Research Laboratories, Inc.). All cells lines were incubated at 37 °C in 5% CO₂. Three biological replicates were analysed.

4.2.2. EVs' isolation and characterisation

EV's production, isolation and protein marker characterisation were carried out following the methods described in Chapter 3 [64, 116, 117]. The size and number of EVs obtained by SEC were determined by nanoparticle tracking analysis using a Nanosight 300 (Malvern).

4.2.3. TEM

Vesicular morphology confirmation of the EVs isolated using SEC was carried out by TEM. Samples for TEM imaging were prepared by simple drop casting. In one typical preparation, 1 µL of sample was drop casted directly onto a copper grid

(Carbon support film square grid, 400 mesh, 5-6 nm, copper, EMS) and allowed to dry at room temperature overnight. Dried samples were then imaged with a JEOL 2100Plus transmission electron microscope at 200 kV. Note: As stated in “contribution by other people” (page 57), the TEM images were captured by the Stevens group.

4.2.4. Lipid extraction

Lipids were extracted from EVs and cells based on the Blight and Dyer method [118], as described in Chapter 3. The only change was that the extracted lipids were reconstituted in IPA/LC-MS grade H₂O (4:1, v/v).

4.2.5. LC-MS analysis

LC-MS analysis was carried out using the facilities and methodologies defined by the National Phenome Centre from Imperial College London. The lipids extracted per sample were analysed by UPLC-MS using an ACQUITY UPLC System (ACQUITY UPLC® BEH C8 Column, 1.7 µm, 2.1 x 100 mm, at 55 °C) coupled to a Waters XEVO G2 Q-TOF mass spectrometer (Waters). Mobile phase A consisted of H₂O:IPA:ACN (2:1:1, v/v/v), 5 mM ammonium acetate, 0.05% acetic acid and 20 µM phosphoric acid. Mobile B consisted of IPA:ACN (1:1, v/v), 5 mM ammonium acetate and 0.5 % acetic acid. The binary gradient profile (mobile phase A and B) was 99% mobile phase A (0.0-2.0 min), 70% A (2.0-11.5 min), 10% A (11.5-12.0 min), 0.1% A (12.0-12.50 min), 35% A (12.50-12.55 min), 70% A (12.55-12.65 min), 99% A (12.65-12.75 min), and 99% A (12.75-13.25 min). The MS and MS/MS data were acquired for 0.1 s in the centroid mode in both ESI⁺ and ESI⁻, mass range 50 – 2000 Da. Mass spectrometer source parameters were as follow: capillary voltage 2 kV (ESI⁺)/1.5 kV (ESI⁻), sampling cone voltage 25 V, source temperature 120 °C, desolvation temperature 600 °C, cone gas flow 150 (L/h) and desolvation flow 1000 (L/h). QC samples and lipid internal standards were included in the set of samples analysed (Supplementary Table 1), in order to assess the quality of the data throughout the whole running time. DDA was achieved

for QC samples to perform MS/MS experiments in order to identify lipid species. MS/MS experiments were also carried out for the most relevant features.

4.2.6. Data analysis

MS data pre-processing was undertaken using XCMS open-source software [106], and the data analysis was carried out as described in Chapter 3. The lipids annotated for EVs and cells in this study (Chapter 4) are indicated in results and discussion section (4.3) and contributed to the creation of an in-house lipid database for EVs.

4.3. Results and Discussion

An untargeted LC-MS based lipidomics analysis was carried out for cancerous and non-cancerous ovarian epithelial cells and their secreted EVs (Figure 4.3.1 and Supplementary Figure 3). The cancerous cells included two high grade serous ovarian cancer cell lines, the most common subtype of human ovarian cancer, and two endometrioid ovarian cancer cell lines (Figure 4.3.1A). EVs isolated from the conditioned media of the ovarian epithelial cells by SEC, were positive for the EV protein marker CD9, CD63 and CD81 (Figure 4.3.2A) and had a vesicular morphology by TEM (Figure 4.3.2B).

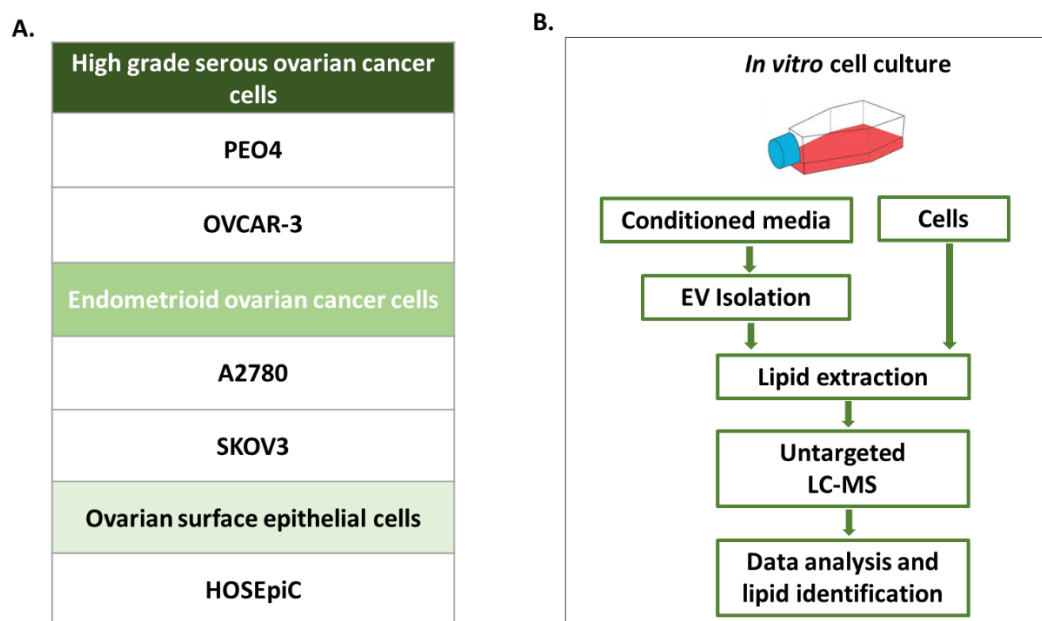
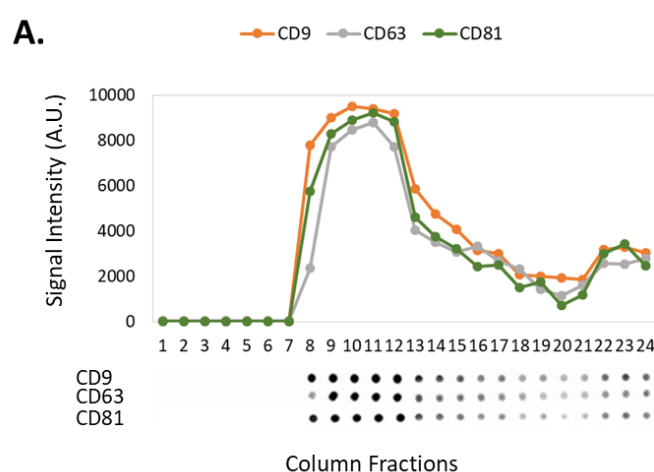


Figure 4.3.1. Workflow for the lipidomic analysis of the cancerous and non-cancerous ovarian epithelial cells and their secreted EVs. **A.** List of the human ovarian epithelial cancer cells lines (PEO4, OVCAR-3, A2780 and SKOV3) and the non-cancerous human ovarian epithelial cells (HOSEpiC) studied. **B.** Workflow used for the lipidomic analysis of the cells and their secreted EVs.



B. TEM

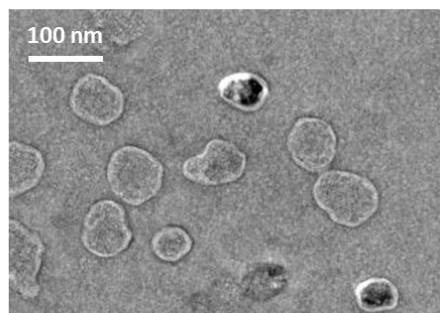


Figure 4.3.2. Protein and morphology analyses of the EVs secreted by ovarian epithelial cells. **A.** EV associated protein markers; CD9, CD81 and CD63, were detected in the EVs secreted by the ovarian epithelial cells and isolated by SEC from conditioned media. **B.** The EVs' showed a vesicular morphology by TEM analysis.

Lipidomic analysis of ovarian epithelial cancer cells and their EVs

PCA was performed for the LC-MS datasets in order to understand the distribution of the data (Figure 4.3.3A). In contrast to LC-MS data distribution observed for the mammary cells and their secreted EVs (Chapter 3), the LC-MS data for ovarian cells and EVs does not show a separation between cells and EVs. It does, however, show a separation of the endometrioid ovarian cells and their EVs, from the other two groups (the high grade serous ovarian cancer and the normal ovarian surface epithelial cells and their EVs). This separation can be observed in the first principal component of the PCA for the ESI+ mode and in the second principal component in the PCA for the ESI- mode. It has been suggested that ovarian epithelial cancer can originate in Müllerian-derived epithelial tissue (this includes the cervix, endometrium, and the fallopian tubes epithelium). High serous ovarian cancer could have an origin in the fallopian tube epithelium that embeds on the ovary, while

endometrioid ovarian cancer could have an origin in the endometrium [36]. This could explain the variability observed in the PCA.

A volcano plot was performed to display the lipid species with the largest changes between ovarian cancer cells and their secreted EVs, which were also significantly different between these two groups (i.e. cells and EVs) (Figure 4.3.3B). Thirty seven lipid species were found to be significantly (p -value < 0.05) enriched ($\log_2$ fold change > 2) in ovarian cancer-derived EVs when compared to their parental cells (Table 4.3.1). Eighty lipid species were found to be significantly (p -value < 0.05) enriched ($\log_2$ fold change > 2) in ovarian cancer cells when compared to their secreted EVs (Table 4.3.2). Similarly, to breast cancer-derived EVs (Chapter 3), the volcano plot in figure 4.3.3B shows sphingolipids and glycerophospholipids significantly enriched in EVs when compared to their parental cells. HexCer and Cer showed the highest fold changes in EVs, when compared to cells. This was followed by LPC and PC species. SM species were also found within the most significant and enriched lipids in EVs when compared to cells. In addition, DG, second messenger signalling lipid, was also found within the lipids enriched in EVs compared to cells. Signalling lipids, such as PA and PI, and structural lipids, such as PC and PE species, were found in both cells and EVs with high fold changes. PC and PE can also be a source of signalling molecules. LPE and LPC were only found enriched in EVs compared to cells. They can act as signalling molecules in EVs, but they can also be involved in EVs curvature [123]. Lyso-forms of PE and PC were also identified in breast-cancer derived EVs (Chapter 3), and previously in EVs found in human blood plasma and EVs secreted by prostate cancer cells [74, 75, 120].

Similarly to the breast cancer analysis, epithelial ovarian cancer cells were enriched in TG when compared to EVs (Figure 4.3.3B and Table 4.3.2). TG are not structural lipids of biological membranes and they were not expected to be enriched in EVs compared to cells. As mentioned in chapter 3, due to the origin of EVs in specific parts of the cells, they do not represent the entirety of cellular components of their parental cells. PG species were also found enriched in cells when compared to EVs,

as in breast cancer cells (Chapter 3). FA were also found enriched in cells compared to EVs, AA was found in ovarian cells as well as breast cancer cells (Chapter 3). This is an interesting finding, as AA plays an essential role in the activation of the PI3K/AKT/mTOR signalling pathway and therefore in cancer development and progression [131, 132]. Cell proliferation requires continuous production of FA [44, 129].

To date, this is the first study of the lipid composition of ovarian cancer-derived EVs. There is a study of the lipid composition of one endometrioid ovarian cancer cell line, but in that study the lipid composition of EVs and their parental cells was not compared [79]. Only a comparison between EVs from endometrioid ovarian cancer cells and ovarian surface epithelial cells was carried out, which found mainly sterol lipids in EVs [79]. In this study, Chapter 4, sphingolipids as well as glycerophospholipids were found enriched in EVs when compared to their parental cells. Similar results were found in the study of breast cancer-derived EVs (Chapter 3). Moreover, the study of EVs from prostate cancer cells [74, 75], urinary EVs from prostate cancer patients [76] and two TN breast cancer cell lines [78], showed similar results. Sphingolipids could be organised in lipid domains in EVs, like lipid rafts [73, 124], playing an important role in cell proliferation and survival [49, 125]. Ether phospholipids were also found in ovarian cancer derived EVs, which could be involved in lipid raft stability and signalling processes [128]. Ether phospholipids were also found in breast cancer derived EVs (Chapter 3) and prostate cancer derived EVs [75, 76].

In Chapter 3, it was discussed that the study of one prostate cancer cell line; PC-3 [75], found PS as the most enriched glycerophospholipid in EVs when compared to cells, but in the study of breast cancer derived-EVs (Chapter 3) PS was not identified as one of the most enriched lipids in EVs. In this study of ovarian cancer-derived EVs (Chapter 4), one PS species and one ether PS species were identified within the lipids enriched in EVs compared to cells, but also PS were found in cells. This supports the idea that PS is not necessarily one of the most enriched phospholipids in cancer-derived EVs. Moreover, PS has not been identified in EVs

found in blood plasma [120], but they have been found in urinary EVs [76]. To date, sphingolipids as discussed in Chapter 3 continue to be the most enriched lipids found in EVs compared to cells, from a wide variety of cell types.

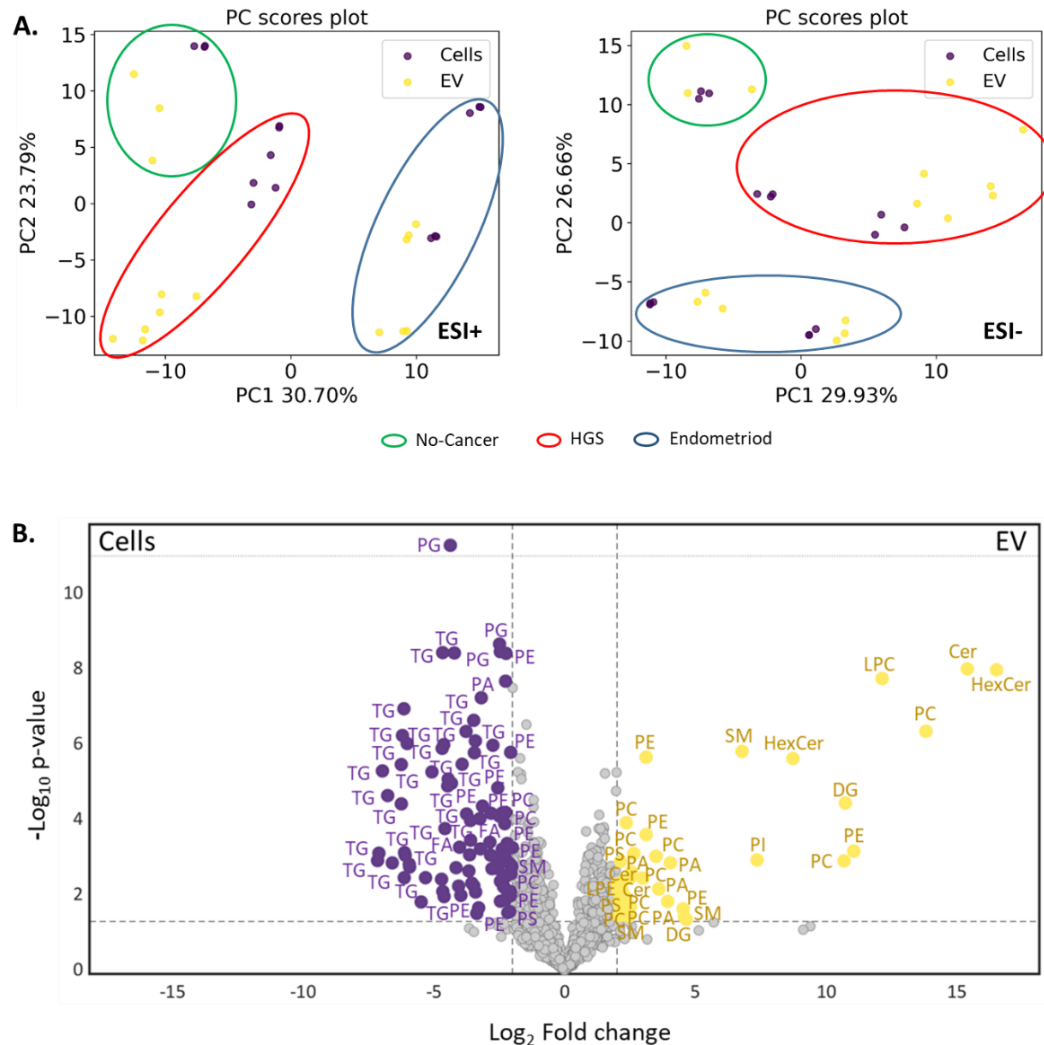


Figure 4.3.3. Lipids enriched in EVs when compared to their parental cells.

A. PCA showing the LC-MS data distribution for both cells and their derived EVs. This includes both cancerous and non-cancerous cells and their EVs. The data was acquired in both ESI+ and ESI-. **B.** Volcano plot showing the lipids enriched in ovarian cancer cells (purple) when compared to their secreted EVs (yellow). Mean fold changes between EVs and cells were calculated based on the relative abundances for each variable in the datasets. A threshold of $\log_2$ fold > 2

or < -2 was defined for the analysis of the lipid enrichment between EVs and cells, respectively. At the same time, statistically significant differences between the means of EVs and cells were calculated by one-way ANOVA with a p -value less than 0.05 as statistically significant.

Table 4.3.1. List of lipids species enriched in EVs when compared to their parental cells. The table shows the list of the 37 lipids and their m/z values identified as significantly (p -value < 0.05) enriched ($\log_2$ fold change > 2) in ovarian cancer-derived EVs when compared to their parental cells.

m/z	Theoretical m/z	Lipid ID	EV/Cells	ANOVA p -value	Log ₂ FC EV/Cells
590.4544	590.4544	LPC(24:0)	EV	1.79E-08	12.142
500.2783	500.2773	LPE(20:4)	EV	8.28E-03	2.059
701.5127	701.5139	PA(36:1)	EV	2.39E-03	2.547
725.5127	725.5120	PA(38:3)	EV	6.87E-03	3.597
751.5283	751.5307	PA(40:4)	EV	1.38E-03	4.041
749.5127	749.5133	PA(40:5)	EV	1.46E-02	3.941
915.5968	915.5899	PI(40:3)	EV	1.16E-03	7.360
760.5851	760.5936	PC(34:1)	EV	4.50E-07	13.819
824.5566	824.5598	PC(36:2)	EV	1.46E-02	2.070
782.5694	782.5675	PC(36:4)	EV	3.45E-02	2.068
838.5604	838.5608	PC(36:5)	EV	5.76E-03	2.028
812.6164	812.6091	PC(38:3)	EV	3.51E-03	2.994
834.6007	834.6019	PC(40:6)	EV	1.12E-02	2.318
892.6790	892.6735	PC(44:5)	EV	1.23E-03	10.686
938.7583	938.7572	PC(48:2)	EV	7.87E-04	2.642
772.5498	772.5495	PC(O-32:3)	EV	1.71E-02	2.227
826.5968	826.5952	PC(O-36:4)	EV	3.43E-02	2.155

846.6347	846.6357	PC(O-40:4)	EV	1.20E-04	2.354
876.6124	876.6169	PC(O-40:7)	EV	2.04E-02	2.514
656.4297	656.4246	PE(30:3)	EV	2.48E-04	3.113
740.5225	740.5278	PE(36:4)	EV	2.31E-02	4.530
796.5134	796.5354	PE(36:5)	EV	5.06E-03	2.010
764.5236	764.5250	PE(38:5)	EV	1.32E-02	2.286
648.4963	648.4891	PE(O-30:1)	EV	2.18E-06	3.111
702.5432	702.5504	PE(O-34:2)	EV	6.80E-04	11.056
828.5151	828.5154	PS(36:1)	EV	1.27E-03	2.136
794.5342	794.5349	PS(O-38:5)	EV	9.12E-03	2.103
520.5088	520.5096	Cer(34:1)	EV	3.46E-03	2.510
548.5401	548.5403	Cer(36:1)	EV	7.48E-03	2.194
628.5065	628.5149	Cer(38:3)	EV	9.88E-09	15.403
655.5173	655.5137	SM(32:2)	EV	1.53E-06	6.783
739.5151	739.5250	SM(34:2)	EV	3.81E-02	4.573
713.5956	713.5942	SM(36:1)	EV	3.67E-02	2.443
672.5409	672.5412	HexCer(32:1)	EV	1.05E-08	16.518
692.5072	692.5144	HexCer(32:2)	EV	2.36E-06	8.729
619.5296	619.5274	DG(36:3)	EV	4.41E-02	4.654
643.5296	643.5329	DG(38:5)	EV	3.60E-05	10.737

Table 4.3.2. Lipids enriched in cells when compared to their secreted EVs.

The table shows the list of the 80 lipids and their m/z values identified as significantly (p -value < 0.05) enriched ($\log_2$ fold change > 2) in ovarian cancer cells when compared to their secreted EVs.

m/z	Theoretical m/z	Lipid ID	EV/Cells	ANOVA p -value	Log ₂ FC EV/Cells
-------	----------------------	----------	----------	---------------------	---------------------------------

305.2486	305.2471	FA(20:3)	Cells	5.35E-04	-4.023
303.2330	303.2317	FA(20:4)	Cells	3.94E-04	-2.863
751.5283	751.5381	PA(40:4)	Cells	1.11E-03	-2.559
807.5898	807.5855	PA(44:5)	Cells	2.09E-08	-2.263
704.5225	704.5226	PC(30:1)	Cells	1.74E-03	-2.023
730.5381	730.5378	PC(32:2)	Cells	2.17E-02	-3.294
814.6320	814.6347	PC(38:2)	Cells	6.48E-03	-2.293
900.7416	900.7435	PC(44:1)	Cells	1.04E-02	-2.077
690.5432	690.5424	PC(O-30:1)	Cells	6.41E-05	-2.258
754.5756	754.5753	PC(O-36:3)	Cells	1.33E-02	-2.274
858.7310	858.7320	PC(O-42:1)	Cells	1.51E-03	-2.089
856.7154	856.7166	PC(O-42:2)	Cells	1.27E-04	-2.296
744.5549	744.5544	PE(36:1)	Cells	7.99E-05	-2.494
742.5392	742.5394	PE(36:2)	Cells	6.00E-03	-2.342
740.5236	740.5241	PE(36:3)	Cells	1.45E-02	-2.458
738.5068	738.5092	PE(36:5)	Cells	3.85E-09	-2.247
772.5862	772.5863	PE(38:1)	Cells	1.99E-03	-2.566
770.5705	770.5701	PE(38:2)	Cells	8.68E-04	-2.836
768.5549	768.5552	PE(38:3)	Cells	5.47E-04	-2.004
792.5514	792.5516	PE(38:3)	Cells	9.69E-04	-2.284
798.6018	798.6029	PE(40:2)	Cells	1.98E-03	-2.783
826.6331	826.6342	PE(42:2)	Cells	1.42E-05	-2.562
822.6018	822.6016	PE(42:4)	Cells	5.98E-04	-2.389
850.6331	850.6329	PE(44:4)	Cells	6.72E-05	-2.816
848.6175	848.6184	PE(44:5)	Cells	6.71E-05	-2.402
650.5119	650.5254	PE(O-30:0)	Cells	8.08E-03	-2.076
700.5252	700.5269	PE(O-32:0)	Cells	7.93E-03	-2.151
758.6058	758.6058	PE(O-38:2)	Cells	1.63E-06	-2.067

748.5287	748.5209	PE(O-38:6)	Cells	9.34E-05	-3.246
782.6069	782.6065	PE(O-40:3)	Cells	4.29E-03	-2.458
810.6382	810.6378	PE(O-42:3)	Cells	3.06E-02	-3.365
693.4712	693.4702	PG(30:0)	Cells	1.87E-03	-4.156
769.4990	769.4986	PG(34:2)	Cells	3.79E-03	-4.711
745.5025	745.5025	PG(34:2)	Cells	5.65E-03	-3.479
765.4712	765.4708	PG(36:6)	Cells	5.21E-12	-4.377
783.5182	783.5166	PG(37:4)	Cells	1.01E-02	-3.974
791.4869	791.4861	PG(38:7)	Cells	1.36E-03	-2.146
821.5338	821.5327	PG(40:6)	Cells	2.18E-09	-2.494
821.5327	821.5332	PG(38:4)	Cells	3.50E-04	-3.600
843.5171	843.5156	PG(40:7)	Cells	5.93E-04	-3.225
762.5291	762.5286	PS(34:0)	Cells	2.73E-02	-2.106
760.5134	760.5179	PS(34:1)	Cells	3.48E-03	-2.302
632.5378	632.5696	Cer(38:1)	Cells	2.69E-03	-2.115
631.5173	631.5664	SM(30:0)	Cells	2.71E-03	-2.055
768.7076	768.7076	TG(44:0)	Cells	8.00E-03	-4.665
766.6919	766.6926	TG(44:1)	Cells	3.45E-03	-5.327
794.7232	794.7242	TG(46:1)	Cells	1.52E-03	-5.948
773.6654	773.6626	TG(46:3)	Cells	1.09E-02	-4.615
822.7545	822.7557	TG(48:1)	Cells	1.07E-05	-4.335
820.7389	820.7403	TG(48:2)	Cells	5.80E-07	-6.209
801.6967	801.6949	TG(48:3)	Cells	1.66E-06	-3.456
799.6810	799.6793	TG(48:4)	Cells	1.40E-03	-6.597
850.7858	850.7877	TG(50:1)	Cells	8.05E-07	-3.418
848.7702	848.7719	TG(50:2)	Cells	1.02E-06	-4.629
827.7123	827.7113	TG(50:4)	Cells	1.27E-06	-4.687
825.6967	825.6956	TG(50:5)	Cells	3.41E-06	-6.254

878.8171	878.8186	TG(52:1)	Cells	4.53E-07	-3.774
876.8015	876.8038	TG(52:2)	Cells	7.02E-05	-3.754
857.7593	857.7592	TG(52:3)	Cells	5.37E-06	-5.081
855.7436	855.7431	TG(52:4)	Cells	1.05E-06	-2.748
853.7280	853.7273	TG(52:5)	Cells	3.67E-09	-4.669
851.7123	851.7109	TG(52:6)	Cells	1.13E-07	-6.153
904.8328	904.8351	TG(54:2)	Cells	1.72E-04	-4.583
902.8171	902.8189	TG(54:3)	Cells	4.83E-03	-3.554
883.7749	883.7745	TG(54:4)	Cells	5.79E-08	-3.192
898.7858	898.7878	TG(54:5)	Cells	8.29E-03	-3.410
879.7436	879.7431	TG(54:6)	Cells	3.74E-09	-4.221
894.7545	894.7561	TG(54:7)	Cells	1.52E-02	-5.493
932.8641	932.8661	TG(56:2)	Cells	3.79E-05	-6.253
928.8328	928.8348	TG(56:4)	Cells	1.79E-03	-5.934
909.7906	909.7906	TG(56:5)	Cells	8.33E-06	-4.456
907.7749	907.7746	TG(56:6)	Cells	8.49E-04	-3.638
922.7858	922.7884	TG(56:7)	Cells	5.78E-03	-4.059
960.8954	960.8980	TG(58:2)	Cells	5.05E-06	-6.975
954.8484	954.8511	TG(58:5)	Cells	8.75E-04	-6.101
935.8062	935.8058	TG(58:6)	Cells	7.45E-04	-6.128
931.7749	931.7737	TG(58:8)	Cells	1.26E-05	-4.475
986.9110	986.9132	TG(60:3)	Cells	7.73E-04	-7.110
963.8375	963.8372	TG(60:6)	Cells	1.21E-03	-7.160
993.8845	993.8847	TG(62:5)	Cells	2.28E-05	-6.773

Spearman rank correlation coefficients showed glycerophospholipids species (Figure 4.3.4 and Table 4.3.3), including ether phospholipids, and sphingolipids (Cer, SM and HexCer) within the lipids with a very strong correlation (Spearman r_s

> 0.9) between cells and their paired EVs (Table 4.3.3). This suggests that although EV and cells' lipid composition is significantly different, there are lipid species that are highly associated between EVs and their parental cells, and that this includes not only structural lipids, but also signalling lipids which play an important role in cancer development and progression.

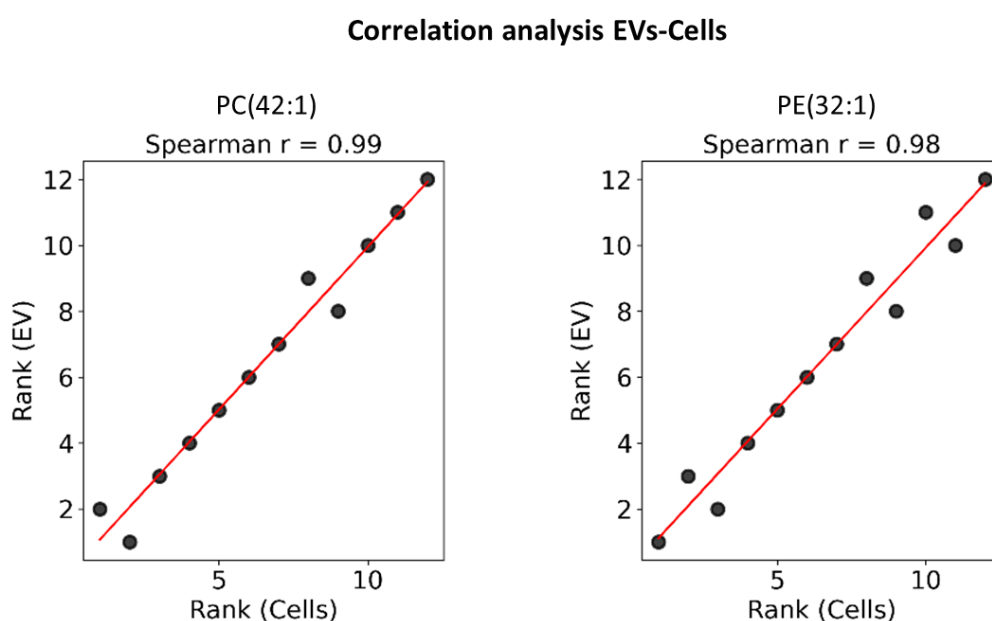


Figure 4.3.4. Lipid species highly correlated between EVs and their parental cells. Spearman rank-order correlation coefficient calculated for paired cells and their EVs identified a PC and a PE as the top 2 lipid species highly correlated between EVs and cells; PC(42:1) and PE(32:1).

Table 4.3.3. Lipids with a very strong correlation between EVs and their parental cells. List of lipids species with a Spearman rank-order correlation coefficient $r_s > 0.9$ indicating a very strong correlation between EVs and their parental cells, with a p -value < 0.05 indicating that the correlation is statistically significant.

<i>m/z</i>	Theoretical <i>m/z</i>	Lipid ID	Spearman <i>r_s</i>	Spearman <i>p</i> -value
872.7103	872.7122	PC(42:1)	0.986	4.12E-09
690.5068	690.5111	PE(32:1)	0.979	3.09E-08
871.6910	871.6891	SM(42:2)	0.979	3.09E-08
941.6125	941.6129	PI(42:4)	0.972	1.29E-07
920.6386	920.6308	PC(42:6)	0.972	1.29E-07
674.5130	674.5123	PC(O-30:1)	0.965	3.88E-07
658.6496	658.6489	Cer(44:2)	0.965	3.88E-07
654.5303	654.5304	HexCer(32:1)	0.965	3.88E-07
940.7131	940.7094	PC(44:0)	0.958	9.54E-07
900.7416	900.7435	PC(44:1)	0.951	2.04E-06
776.5589	776.5597	PE(O-40:7)	0.951	2.04E-06
632.6340	632.6342	Cer(42:1)	0.951	2.04E-06
860.6175	860.6104	PC(43:6)	0.9441	3.93E-06
796.6215	796.6205	PC(O-38:4)	0.9441	3.93E-06
809.6507	809.6507	SM(40:1)	0.9441	3.93E-06
725.5127	725.512	PA(38:3)	0.9371	6.99E-06
891.5968	891.5971	PI(38:1)	0.9371	6.99E-06
834.6007	834.6019	PC(40:6)	0.9371	6.99E-06
718.5745	718.5742	PC(O-32:1)	0.9371	6.99E-06
746.6058	746.6057	PC(O-34:1)	0.9371	6.99E-06
772.6215	772.6215	PC(O-36:2)	0.9371	6.99E-06
676.5276	676.5258	PE(O-32:1)	0.9371	6.99E-06
750.5432	750.5435	PE(O-38:6)	0.9301	1.17E-05
804.5913	804.5909	PE(O-42:6)	0.9301	1.17E-05
788.5447	788.5446	PS(36:1)	0.9301	1.17E-05
518.4931	518.4928	Cer(34:2)	0.9301	1.17E-05
548.5401	548.5403	Cer(36:1)	0.9301	1.17E-05

815.7000	815.7014	SM(42:1)	0.9301	1.17E-05
510.2979	510.2973	LPE(22:5)	0.9231	1.86E-05
764.5236	764.524	PC(36:5)	0.9231	1.86E-05
704.5600	704.5571	PE(O-34:0)	0.9231	1.86E-05
704.5589	704.5588	PE(O-34:1)	0.9231	1.86E-05
821.6507	821.6426	SM(41:2)	0.9231	1.86E-05
636.4963	636.4952	LPC(26:0)	0.9161	2.84E-05
786.6007	786.6014	PC(36:2)	0.9161	2.84E-05
714.5079	714.5083	PE(34:2)	0.9161	2.84E-05
912.6699	912.6663	PE(44:3)	0.9161	2.84E-05
703.5748	703.5723	SM(34:1)	0.9161	2.84E-05
764.5447	764.5465	PC(30:0)	0.9091	4.19E-05
776.6528	776.6532	PC(O-36:0)	0.9091	4.19E-05
774.5443	774.5444	PC(O-38:7)	0.9091	4.19E-05
844.6191	844.6137	PC(O-40:5)	0.9091	4.19E-05
690.5079	690.5093	PE(32:0)	0.9091	4.19E-05
780.5902	780.5904	PE(O-40:5)	0.9091	4.19E-05
691.4556	691.4547	PG(30:1)	0.9091	4.19E-05
799.5495	799.5473	PG(38:3)	0.9091	4.19E-05
723.497	723.4974	PA(38:4)	0.9021	6.00E-05
716.5589	716.5583	PC(O-32:2)	0.9021	6.00E-05

Ovarian cancer cells and their EVs can be distinguished from the non-cancerous ovarian cells and their EVs based on their lipid composition

To understand whether there are differences in the lipid composition between cancerous and non-cancerous ovarian cells and EVs, normal ovarian surface epithelial cells and their EVs were analysed together with the ovarian cancer cells

and their EVs. Figure 4.3.5A showed the same PCA as in figure 4.3.3A for the ESI+ and ESI- modes, but in figure 4.3.5A the non-cancerous EVs and cells are coloured yellow and the cancerous EVs and cells are coloured purple. The PCA shows that the non-cancerous EVs and cells are separated from the cancerous EVs and cells. The LC-MS data from the ESI+ and ESI- modes were fused and auto scaled and a RFE-LR analysis was performed. This analysis selected a subset of the most relevant lipid species which allowed the identification of seven statistically significant (q -value<0.05) lipid species which distinguish cancerous from non-cancerous EVs and Cells (Figure 4.3.5B). These seven lipid species included LPE(20:4), PE(18:0_20:4), Cer(d18:0/16:0), Cer(d18:0/16:1), PC(16:0_20:2), SM(d18:1_26:0) and SM(d18:1_26:1) (Figure 4.3.5B). The leave-one-group-out cross-validation analysis showed that EVs and cells can be perfectly classified based on their cancerous or non-cancerous origin with an overall accuracy of 100% (Figure 4.3.5C). Furthermore, a ROC analysis based on the combination of these seven relevant lipids showed an AUROC equal to 1. This demonstrates the usefulness of this combination of lipid species to distinguish between cancerous or non-cancerous EVs (Figure 4.3.5D). Again, as found in Chapter 3, lipidomic analysis of cancer-derived EVs shows that the EVs' lipid composition has potential for the identification of lipid biomarkers for cancer diagnosis, even for ovarian cancer diagnosis.

LPE(20:4), PE(18:0_20:4), Cer(d18:0/16:0) and Cer(d18:0/16:1) had higher relative intensities in the ovarian cancer cells and their EVs. High levels of PE were also observed in breast cancer derived EVs (Chapter 3). Lipid alterations have also been detected in ovarian cancer tissues by DESI-MSI [105]. LPE can act as signalling lipids, and PE can be precursors of second signalling messengers. Lysophospholipids have already shown potential as biomarkers for ovarian cancer diagnosis [102, 103]. Alterations in phospholipids have also been observed in blood plasma samples from ovarian cancer patients [104]. Sphingolipids can have a pro- or anti-proliferation role. C1P and S1P derived from Cer are involved in cell growth and proliferation [48, 134-136]. This could explain why Cer had high relative

intensities in cancer but also SM were found with high intensities in non-cancerous EV and cells.

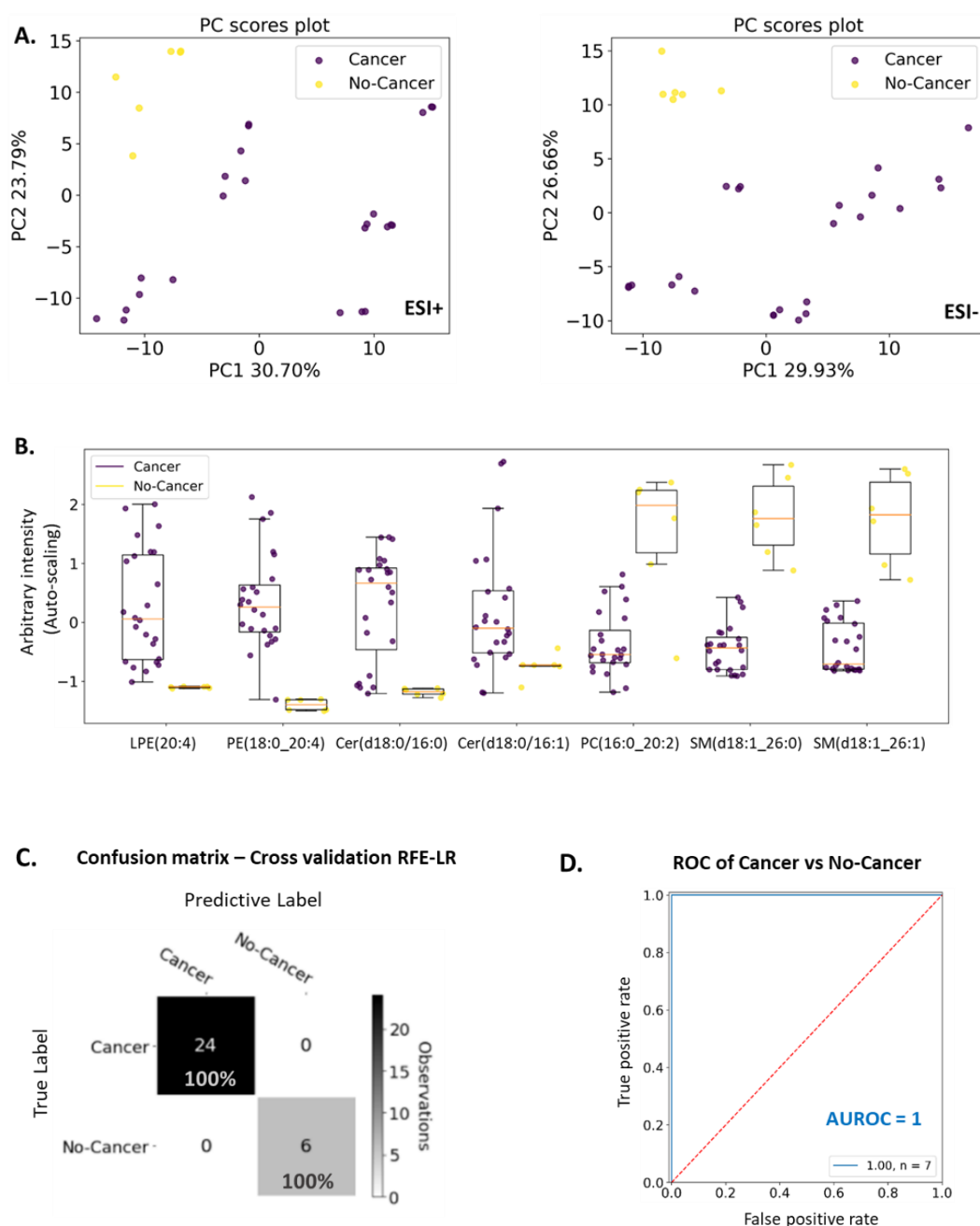


Figure 4.3.5. Statistical analysis of cancerous and non-cancerous EVs and cells. A. PCA showing the LC-MS data distribution of cancerous (Cancer) and

non-cancerous (No-Cancer) EVs and cells. This is the same PCA as in figure 4.3.3A for the ESI+ and ESI- modes, but in this PCA the non-cancerous (yellow) and cancerous (purple) EVs and cells are indicated to highlight their data distribution. **B.** Box plots for the seven lipid species identified by the statistical model obtained by RFE-LR that distinguish ovarian cancer cells and their derived EVs from non-ovarian cancer cells and their derived EVs. The statistical model and box plots are based on the fused and auto scaled ESI+ and ESI- datasets. **C.** Leave-one-group-out cross-validation was performed to evaluate the classification model. This analysis showed that EVs and cells can be perfectly classified based on their cancerous (Cancer) or non-cancerous (No-cancer) origin with an overall accuracy of 100%. **D.** AUROC curve for the combination of the relevant lipid species was equal to 1. This indicates their usefulness for the discrimination between non-cancerous and cancerous samples.

Endometrioid and high grade serous ovarian cancer subtypes have a different lipid composition

When analysing only the ovarian cancer cells and EVs, the PCA for the ESI+ and ESI- datasets shows a clear separation between the two ovarian cancer subtypes, endometrioid and high grade serous ovarian cancer (Figure 4.3.6A). In order to identify the lipid species contributing to this separation, a RFE-LR analysis was performed. For this analysis the MS data obtained by ESI+ and ESI- were fused and the fused data was also auto scaled. The model generated from this analysis allowed the identification of six statistically significant (q -value<0.05) lipid species. This contributes to the classification of the cancerous cells and EVs in the endometrioid and high grade serous ovarian cancer subtypes (Figure 4.3.6B). The relevant lipids included a LPE(22:5) and Cer(d18:0/16:0) which had higher relative intensities in high grade serous ovarian cancer samples as well as PE(16:0/16:0), PC(O-16:1_18:2), PC(O-18:1/20:3) and PE(O-18:1_20:3) which had higher relative intensities in endometrioid samples. Leave-one-group-out cross-validation showed a perfect classification of the cells and EVs (100% accuracy) in endometrioid or

high grade serous ovarian cancer (Figure 4.3.6C). DESI-MSI analysis of ovarian cancer tissues have also demonstrated that samples can be classified into ovarian epithelial cancer subtypes based on their lipid profile [105].

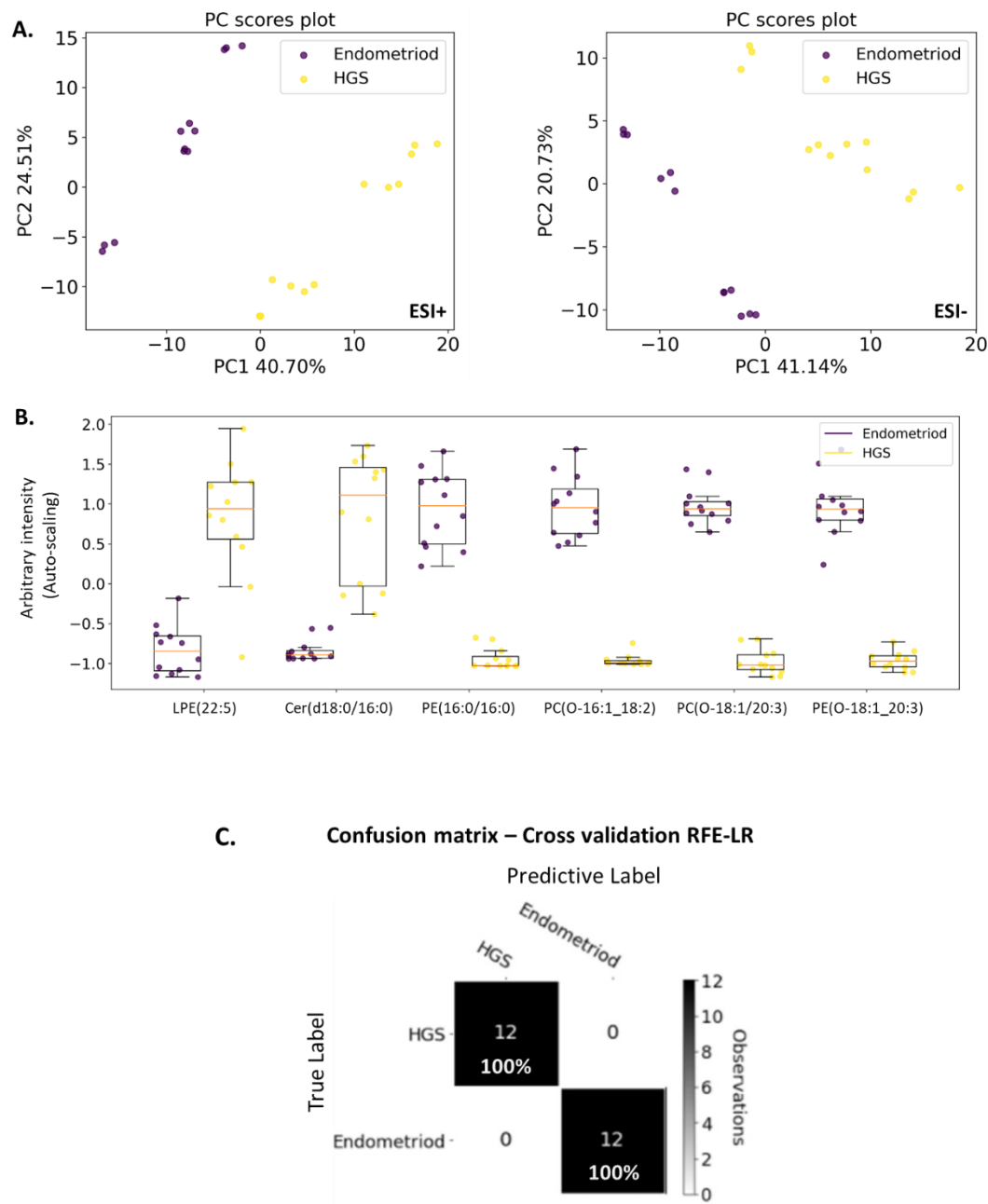


Figure 4.3.6. Statistical analysis of cells and EVs obtained from endometrioid and high grade serous ovarian cancer subtypes. A. Box plots for the six lipid

species identified by the RFE-LR model that classify ovarian cancer cells and their derived EVs based on their lipid profile into their respective ovarian cancer subtypes. Box plots are based on the fused and auto-scaled ESI+ and ESI- datasets. **B.** Confusion matrix of the leave-one-group-out cross-validation analysis carried out to evaluate the performance of the classification model for ovarian cancer subtypes. This analysis shows that cancerous EVs and cells can be perfectly classified into endometrioid or high grade serous ovarian cancer with 100% accuracy.

In conclusion, this study shows that ovarian cancer-derived EVs are mainly enriched in sphingolipids and that glycerophospholipids are also significant lipids in EVs, as breast cancer-derived EVs. This study demonstrates that there is potential in the lipid composition of ovarian cancer-derived EVs for the identification of lipid biomarkers for ovarian cancer diagnosis and for the classification ovarian cancers in their respective ovarian cancer subtype; endometrioid and high grade serous ovarian cancer.

Chapter 5

Isolation of EVs from human blood plasma

5. Isolation of EVs from human blood plasma

5.1. Background

One of the biggest challenges of studying extracellular vesicles found in human body fluids, is the complexity to separate EVs from other body fluids components. As discussed in Chapter 1, there are several methods that can be used for the extraction of EVs from different biological samples. However, there is not a fast and simple method for the isolation of pure EVs from complex matrices such as human blood plasma. Human blood plasma contains amongst others soluble proteins and lipoproteins. This affects MS-based analysis of EVs, such as proteomics and lipidomics. Both lipoproteins and EVs contain lipids which could affect analysis focused on the EVs' lipid composition. The main lipid classes in EVs and lipoproteins which are common between them are sterol lipids and glycerophospholipids. Human blood plasma lipoproteins include chylomicrons, very-low-density lipoproteins (VLDL), intermediate-density lipoproteins (IDL), low-density lipoproteins (LDL) and high-density lipoproteins (HDL). Lipoproteins not only match the size of EVs but also their density, more specifically HDL. Two methods have demonstrated the depletion of co-isolated contaminants in EVs extracted from blood plasma, using a combination of density gradient ultracentrifugation and SEC [66, 67]. One study used human blood plasma [66] and the other rat blood plasma [67]. Unfortunately, these methods are not easy to reproduce due to the use of specific ultracentrifugation rotors [66, 67], lack of details in their methodology ([66] and use of specific apparatus for SEC [67]. They may also require long centrifugation times (24 h) [67], which is not compatible when analysing a high number of samples.

The aim of this chapter was to develop an effective method for the isolation of EVs from human blood plasma, with the highest possible purity useful for lipidomic analysis. The method described in this chapter was developed based on EVs'

extraction methods which have been tested in conditioned cell culture media [65], human blood plasma [66] and rat blood plasma [97].

5.2. Methods

5.2.1. Sample collection

Whole blood samples from ten healthy volunteers were obtained under the Imperial College Healthcare NHS Trust Tissue Bank project Ref R17023-3A. Whole blood samples were collected by venipuncture using EDTA (K2E) vacutainer tubes. Blood plasma was obtained by 10 min centrifugation at 1,000 g at 4 °C, followed by the collection of the supernatant and its centrifugation in a new tube for 10 min at 2,000 g at 4 °C. The blood plasma obtained was stored at -80 °C until further analyses. A pool of samples representative of the ten blood plasma samples from the healthy volunteers was produced to carry out the analysis described in this Chapter 5. Note: The whole blood samples were collected by Professor Charles Coombes team from Imperial College London, Department of Surgery and Cancer. The same team obtained the blood plasma from the blood samples analysed.

5.2.2. Sample preparation

The pool of samples obtained from the healthy volunteers' blood plasma was centrifuged at 5,000 x g for 15 min at 4 °C in order to remove cell debris if present. The supernatant was transferred into a new tube and centrifuged at 16,000 x g for 30 min at 4 °C to remove larger microvesicles. The supernatant was transferred into a new tube and it was kept at 4 °C to be used the same day.

5.2.3. Density gradient ultracentrifugation

Iodixanol density gradients were prepared using OptiPrepTM (60%, w/v). Initially a working solution of 50% OptiPrepTM was obtained by mixing five volumes of 60% OptiPrepTM with one volume of a homogenisation solution consisting of 0.25 M sucrose, 6 mM EDTA and 6 mM Tris-HCl at pH 7.4. The OptiPrepTM working solution was diluted to 30% and 6%, in a buffer solution consisting of 0.25 M sucrose, 1 mM EDTA and 1 mM Tris-HCl at pH 7.4. One mL of 30% and 1 mL of 6% OptiPrepTM were layered in ultra-clearTM centrifuge tubes (Beckman Coulter) from bottom to top, followed by 3 mL of the blood plasma which was clear of debris and large microvesicles. The samples were centrifuged in a Beckman Coulter's SW 55 Ti swinging-bucket rotor at 4 °C for 2 h at 120,000 x g_{avg} , setting up the acceleration to maximum and deceleration to zero. The centrifugation tubes were carefully removed from the rotor and ten 500 μ L fractions were collected from top to bottom.

5.2.4. Density gradient measurement

The absorbance of 6%, 10%, 20% and 30% iodixanol solutions were measured at 340 nm by spectrophotometry (as suggested by OptiPrepTM manufactures) and a standard curve was created. The absorbance of each fraction collected was also measured and their density calculated, based on the iodixanol standard curve created.

5.2.5. Size exclusion and bind-elute chromatography

After EVs found in human blood plasma were separated based on their density, by density gradient ultracentrifugation, the fractions nine and ten containing the EVs (but also lipoproteins with the same density) were added on top of a chromatography column containing 2 mL of CaptoTM Core 700. Once the mix of the fractions containing EVs entered the stationary phase completely, 3 mL of DPBS were used as the mobile phase and 1 mL fractions were collected by gravity.

5.2.6. EVs' characterisation

EV associated protein markers were analysed by dot blotting as described in Chapter 3. In addition to the primary antibodies used in Chapter 3, mouse anti-human apolipoprotein A1 (Apo A1) monoclonal antibody (Thermo Fisher Scientific) was evaluated. NTA was performed to measure the number and size distribution of the EVs, using a Nanosight 300 (Malvern). Five 60 sec videos were analysed for each sample. TEM was carried out as indicated in Chapter 4, to confirm the vesicular morphology of the EVs. Note: the TEM images were captured by the Stevens group as stated in the “contribution by others” in page 57.

5.3. Results and Discussion

Human blood plasma is a complex biological matrix that contains amongst others soluble proteins and lipoproteins. More specifically, lipoproteins can interfere in lipidomic analysis, as they contain similar lipids as those found in EVs. The best approaches for the depletion of lipoproteins in EV samples extracted from blood plasma consist of the separation of lipoproteins and EVs based on their densities and size. In this chapter 5, a reproducible and effective method for isolation of EVs found in human blood plasma was developed (Figure 5.3.1).

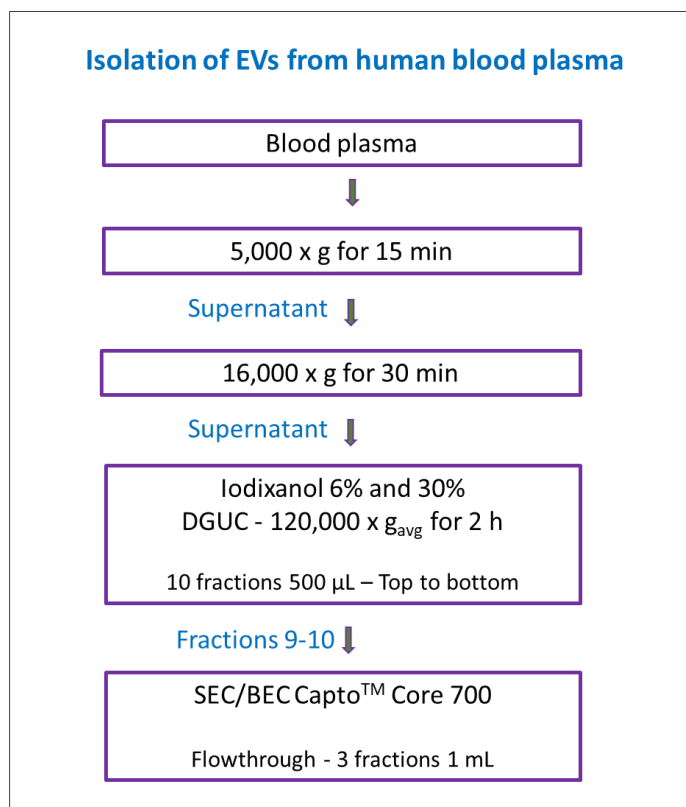


Figure 5.3.1. Workflow of the methodology used for isolation of EVs from human blood plasma.

A pool sample was obtained from the blood plasma from ten healthy individuals. The pool sample of blood plasma was centrifuged at 5,000 x g for 15 min to remove any remaining debris. This was followed by centrifugation at 16,000 x g for 30 min to remove large microvesicles. In parallel to this process, iodixanol gradients were prepared. Only two gradients of iodixanol with contrasting densities were used; 6% iodixanol and 30%. The 30% iodixanol gradient was added at the bottom of the ultracentrifugation tube and then the 6% iodixanol gradient was added on top. This allowed particles with the lowest density to flow in the top of the ultracentrifugation tube and particles with the highest density to flow at the bottom of the tube. Once clearly defined density gradients were formed, 3 mL of the blood plasma cleared of debris and large microvesicles were added on top of the 6% iodixanol gradient. Then, the samples were centrifuged for 2 h at 120,000 x g_{avg}. After completing the

ultracentrifugation process, the particles found in the blood plasma are separated, based on their density. Ten fractions of 500 μL were collected. Apo A1 was detected in all 10 fractions, with higher intensity in fractions 1 to 8, while EV protein markers were only detected in two fractions; fraction 9 and 10 (Figure 5.3.2). This shows that EVs are effectively separated from those lipoproteins that have a lower density than EVs. Those lipoproteins were found in the fractions 1 to 8. However, lipoproteins with a similar density to EVs are separated in the same fractions together with EVs; 9 and 10. HDL have a similar density to EVs, therefore it was expected that they could be detected in the same fractions together with EVs. This suggests that a second step is required to separate the EVs (found in fraction 9 and 10) from the lipoproteins with a similar density to EVs. Those fractions containing EVs had a density of approximately 1.18 g/ml (fraction 9) and 1.23 g/ml (fraction 10). The density of EVs isolated from human blood plasma has been reported to be approximately 1.11-1.19 g/ml, which is similar to HDL (1.06-1.21 g/ml) [66]. The other lipoproteins (chylomicron, VLDL, IDL and LDL) have a density < 1.06 g/ml [66].

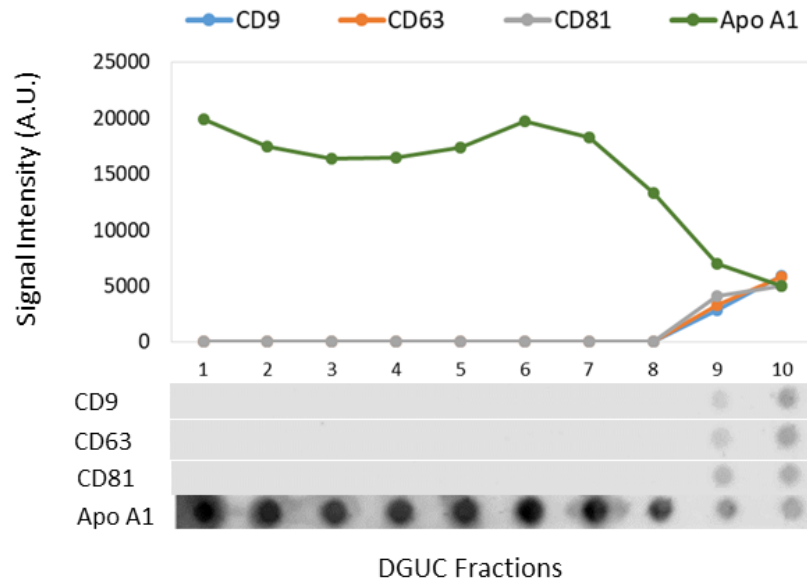


Figure 5.3.2. Detection of EV protein markers in the fractions obtained by density gradient ultracentrifugation of human blood plasma. The presence of

EV protein markers (CD9, CD6 and CD81) in the ten fractions obtained by density gradient ultracentrifugation of human blood plasma was evaluated by dot blotting. In addition Apo A1, which is detected in lipoproteins, was also evaluated. Apo A1 was detected in all ten fractions collected while EV protein markers were detected only in fractions 9 and 10. DGUC = density gradient ultracentrifugation.

After density gradient ultracentrifugation, fractions containing EVs were transferred into a chromatography column containing 2 mL of CaptoTM Core 700. This resin combines both size exclusion and bind-elute chromatography, also known as flowthrough chromatography. The smallest molecules (< 700 kDa) bind the core bead containing octyl amine ligands. They are retained in the column, while the larger molecules (> 700 kDa) pass through the core bead and are eluted with the mobile phase. EV protein markers (CD9, CD63 and CD81) were detected in the flowthrough fractions, whereas Apo A1 (which is found mainly in HDL) was not detected (Figure 5.3.3A). This suggests that HDL are retained in the core bead. This confirms the usefulness of the method developed in this study (Chapter 5), for the isolation of EVs from blood plasma for lipidomic analysis. The EVs isolated from human blood plasma this method had an average diameter of 140.6 nm and show a characteristic vesicular morphology by TEM (Figure 5.3.3B and C, respectively). The number of EVs obtained in the flowthrough fractions was 1.9E+09 particles/mL. However, this is an estimated value from the pool of blood plasma samples from healthy individuals. CaptoTM Core 700 has previously been used for extraction of EVs from conditioned media [65], and in combination with density gradient ultracentrifugation for extraction of EVs from rat blood plasma [67]. Those studies demonstrated the usefulness of CaptoTM Core 700 in EVs isolation methods. However, those studies did not demonstrate the usefulness of their isolation methods for MS-based analysis of EVs.

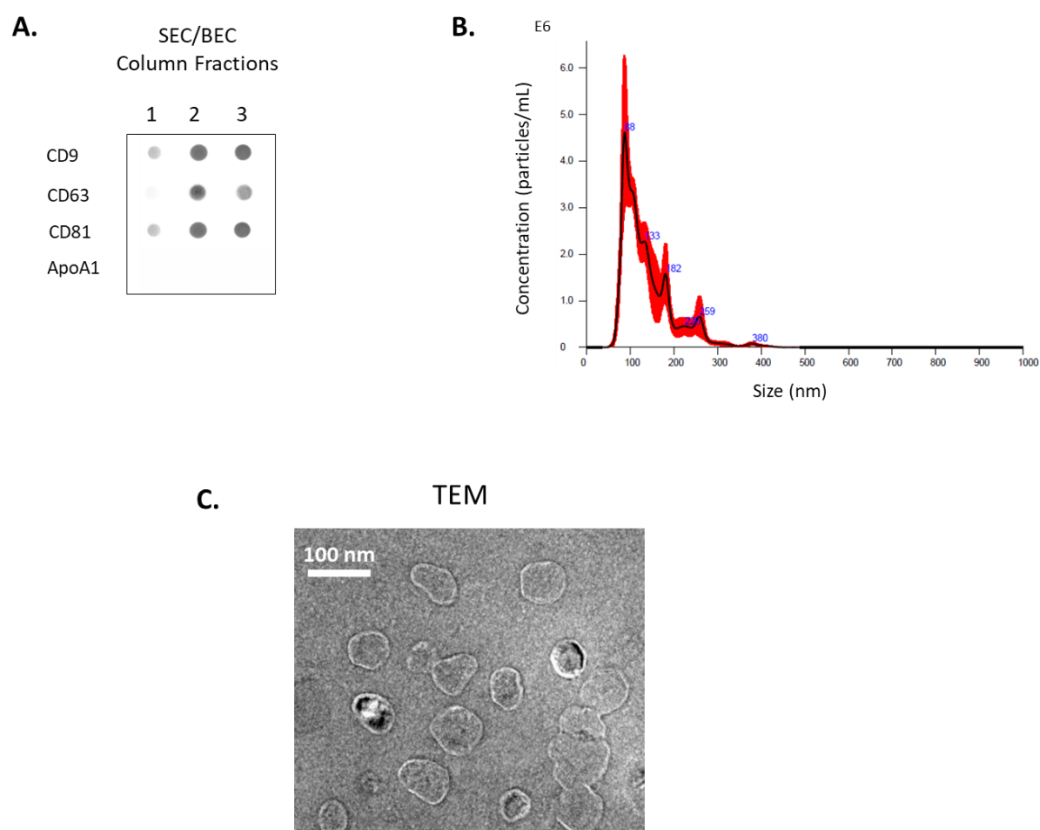


Figure 5.3.3. Characterisation of EVs isolated from human blood plasma by combination of density gradient ultracentrifugation and size exclusion/bind-elute chromatography. **A.** Dot blotting was used to assess the presence of EV protein markers (CD9, CD63 and CD81) in the flowthrough fractions obtained by size exclusion and bind-elute chromatography (SEC-BEC) from fractions 9 and 10 obtained by density gradient ultracentrifugation of blood plasma. All EV protein markers (CD9, CD63 and CD81) evaluated were detected in the flowthrough, but Apo A1, which is mainly detected in HDL, was not detected. **B.** Example of the size distribution by NTA for the EVs found in human blood plasma. NTA records the scattered light from the EVs. **C.** EVs morphology analysed by TEM.

One of the advantages of the method described in this Chapter 5, over other methods based on density gradient ultracentrifugation, is that it uses only two density

gradients and they have contrasting densities, 6% and 30%. Using exactly 1 mL of two iodixanol gradients with contrasting densities, make the formation of gradients a simple and reproducible process. The formation of iodixanol gradients is a very time consuming and complex task, especially when the gradients have similar densities. The gradients can start diffusing if they are not added very carefully and it could result in being a continuous gradient instead of a discontinuous one as expected. Therefore, this method is less laborious and more reproducible than other methods which use several gradients with similar densities. The other methods require the formation of a discontinuous gradient by layering three different iodixanol gradients, in addition to adding on top, a specific volume of blood plasma [66, 67]. Reproducibility is a key factor of an optimal method and therefore using only two 1 ml density gradients, make the production of the gradients a much easier process when compared to the layering of more than two density gradients. Another advantage of the density gradient used in this study (Chapter 5) is that the EVs are separated and detected in only two fractions at the bottom of the tube (fractions 9 and 10), which simplifies the collection and direct transference of the fractions containing EVs into the CaptoTM Core 700 column. But, more importantly, there is not a separation between EVs' subpopulations, as all EVs found in the sample are only found in two fractions (9 and 10). This means that all EVs found in the sample are transferred to the CaptoTM Core 700 column and lipidomic analysis can be done, based on the whole range of EVs found in human blood plasma. Another advantage of this method is that it allows the processing of six samples in around four hours, because of the density gradients' design and ultracentrifugation for only 2 h.

In conclusion, a reproducible and effective method for isolation of EVs found in human blood plasma was developed which can be used for lipidomic analysis. The combination of density gradient ultracentrifugation and size exclusion/bind-elute chromatography allows the purification of EVs from human blood plasma. Although this is a useful method for the isolation of EVs from human blood plasma and it has several advantages over other methods based on density gradient ultracentrifugation, it is still a laborious method. This method is very useful for the scientific research of EVs and especially for proof-of-concept studies, but not for

use in clinical settings. However, to date, there is not a method that does not require specialised equipment and it is not laborious to isolate EVs from blood plasma without compromising the purity of the EVs obtained. Therefore, development of simple methods for isolation of EVs from human blood plasma is still needed.

Chapter 6

Lipidomic analysis of EVs found in blood plasma from breast cancer patients and healthy individuals

6. Lipidomic analysis of EVs found in blood plasma from breast cancer patients and healthy individuals

6.1. Background

Breast cancer is one of the most common types of cancer in women and it is also one of the most studied types of cancer. However, to date, there is not an effective diagnostic method for breast cancer based on patients' body fluids, which could improve the chances to diagnose this type of cancer at an early stage. As demonstrated in Chapter 3, the lipid composition of EVs produced by breast cancer cells has potential for the identification of lipid biomarkers for breast cancer diagnosis. The same can be observed for ovarian cancer diagnosis (Chapter 4). Cancer-derived EVs can be found in cancer patients' body fluids such as blood plasma and urine. Lipidomic analysis of urinary EVs has elucidated the potential of the analysis of EVs' associated lipids for prostate cancer diagnosis [76]. Based on the findings explained above, it could be hypothesised that EVs found in blood plasma from breast cancer patients could harbour lipid species that differentiate them from the lipids in EVs found in individuals with no cancer. To date, there is not a study of the EVs' lipid composition found in breast cancer patients.

After developing a method for effective isolation of EVs from blood plasma useful for lipidomic analysis (Chapter 5), EVs were isolated from blood plasma collected from breast cancer patients and healthy individuals. These were analysed by untargeted lipidomics. The aim of this study was to evaluate the potential of the lipid composition of EVs found in blood plasma from breast cancer patients for the identification of lipid biomarkers for breast cancer diagnosis.

6.2. Methods

6.2.1. Sample collection

Whole blood samples were obtained under the Imperial College Healthcare NHS Trust Tissue Bank project Ref R17023-3A. All women participated voluntarily, and their personal data was kept anonymous. Before surgery was undertaken, ten blood samples were collected from women with primary breast cancer (range of age at diagnosis: 36-68 years, mean age: 57.6 years). A further nine blood samples were obtained from patients with progressive metastatic breast cancer (range of age at initial diagnosis, primary cancer: 47-67 years, mean age: 55.3 years). In addition, ten samples were obtained from healthy women who neither had pre-existing medical conditions and infections at the time the samples were collected or in the previous weeks. The blood plasma samples from healthy women are the same samples used in Chapter 5. As described in Chapter 5, whole blood samples were collected by venipuncture using EDTA (K2E) vacutainer tubes. Blood plasma was obtained from the 29 whole blood samples by 10 min centrifugation at 1,000 g at 4 °C, followed by the collection of the supernatant and its centrifugation in a new tube for 10 min at 2,000 g at 4 °C. The blood plasma obtained was stored at -80 °C until further analyses. Note: The whole blood samples were collected by Professor Charles Coombes' team from Imperial College London, Department of Surgery and Cancer. The same team obtained the blood plasma from the blood samples.

6.2.2. Isolation of EVs from blood plasma

EVs found in the blood plasma from cancer patients and healthy individuals were isolated following the methods described in Chapter 5. In summary, the blood plasma collected was initially centrifuged to remove cell debris and larger microvesicles. The blood plasma clear of debris and microvesicles was then layered on top of a 6% and 30% iodoxianol density gradients and centrifuged at 120,000 $\times g_{avg}$ for 2h at 4 °C. After ultracentrifugation, ten fractions of 500 μ l were collected

from top to bottom from the ultracentrifugation tubes. Fractions 9 and 10 were mixed and added on top of a CaptoTM core 700 chromatography column. Fractions containing EVs were collected and stored at -80 °C until further analysis.

6.2.3. EVs' characterisation

EV associated protein markers were analysed by dot blotting as described in Chapter 3. NTA was performed to measure the number of the EVs, using a Nanosight 300 (Malvern). Five 60 sec videos were analysed for each sample.

6.2.4. Lipid extraction

The lipids of the EVs isolated from blood plasma were extracted based on the Bligh and Dyer method [118], as described in Chapter 3. The lipids extracted were stored at -80 °C until the MS analyses were performed.

6.2.5. LC-MS and data analysis

LC-MS analysis was carried using the facilities and methodologies defined by the NPC from Imperial College London, for untargeted lipidomics analysis. The same methodology indicated in Chapter 4 for the LC-MS analysis and data analysis was used in this study (Chapter 6). The lipids annotated for EVs in this study (Chapter 6) are indicated in results and discussion section (6.3) and contributed to the creation of an in-house lipid database for EVs.

6.3. Results and Discussion

An untargeted lipidomic analysis was carried out to explore the potential of the lipid composition of EVs found in human blood plasma for breast cancer diagnosis. For this analysis, blood plasma samples were obtained from primary breast cancer patients, patients with progressive metastatic breast cancer and healthy women.

Figure 6.3.1 shows the workflow of this study and table 6.3.1 indicates the pathological characteristics of the breast cancer samples evaluated in this study (Chapter 6). The breast cancer samples evaluated included samples from patients with IDC, ILC, DCIS and LCIS, with histological grade from 1 to 3. These samples included hormone receptor-positive breast cancer, HER2+ breast cancer as well as triple negative breast cancer.

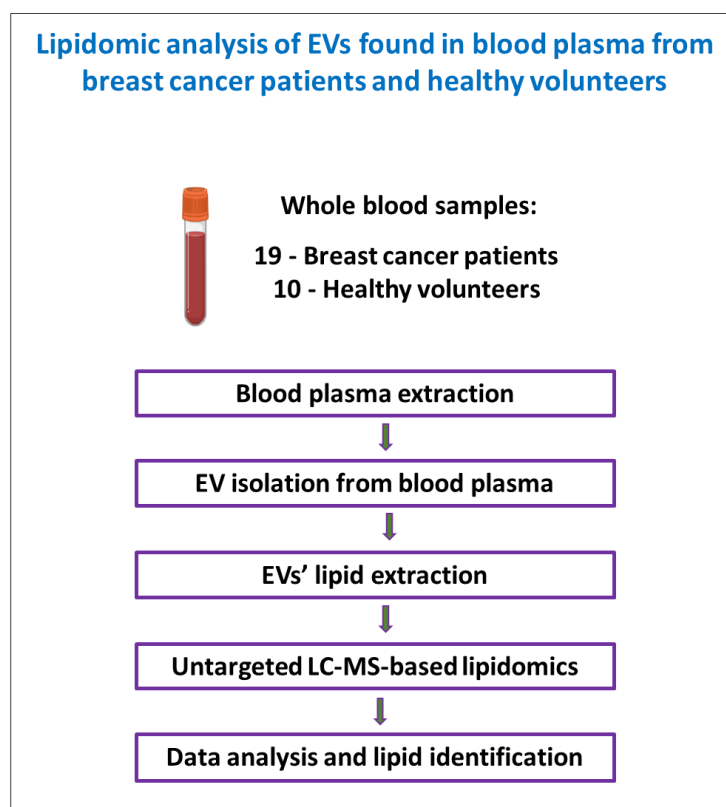


Figure 6.3.1. Workflow of the lipidomic analysis of EVs found in blood plasma from breast cancer patients and healthy women.

Table 6.3.1. Pathological characteristics of breast cancer samples evaluated.

Ten samples from patients with primary breast cancer were evaluated, as well as nine samples from patients with progressive metastatic breast cancer. The information in the table for progressive metastatic breast cancer samples

corresponds to the information reported for the initial breast cancer diagnosis (primary cancer).

Cancer	Sample	Histological grade	Histopathology	ER/PR/HER2
Primary	1	2	IDC + DCIS	+/+/-
Primary	2	2	ILC	+/+/-
Primary	3	2	IDC	+/+/-
Primary	4	1	IDC	+/+/-
Primary	5	3	IDC	+/+/+
Primary	6	2	ILC + LCIS	n.s
Primary	7	1+2	IDC	+/+/-
Primary	8	2	ILC	+/+/-
Primary	9	2	IDC	-/-/-
Primary	10	1	DCIS	n.s
Metastatic	1	2	IDC	+/+/-
Metastatic	2	3	IDC	+/+/-
Metastatic	3	3	IDC	+/+/-
Metastatic	4	3	IDC	-/-/-
Metastatic	5	2	ILC	+/+/-
Metastatic	6	3	IDC	-/-/+
Metastatic	7	2	IDC	+/+/-
Metastatic	8	n.s	n.s	n.s
Metastatic	9	n.s	n.s	n.s

* n.s. = not specified.

EVs were isolated from the blood plasma samples by a combination of density gradient centrifugation and SEC-BEC (methodology described in Chapter 5), in order to obtain EVs with the highest purity possible. This means, for lipidomic analysis, no presence or depletion of unrelated EV's lipids such as those contained in lipoproteins, which are found in human blood plasma. The EVs isolated from blood plasma from breast cancer patients and healthy volunteers, were positive for

the EV protein marker CD9, CD63 and CD81 (Figure 6.3.2A). The vesicular morphology and average diameter of EVs obtained using the methodology developed and described in Chapter 5 were shown in Chapter 5. A higher number of EVs per mL were obtained from blood plasma samples from primary breast cancer patients (mean = $1.0\text{E}+10$ particles/mL), followed by progressive metastatic samples (mean = $5.0\text{E}+9$ particles/mL) and samples from healthy volunteers (mean = $2.3\text{E}+09$ particles/mL) (Figure 6.3.2B). These results could be attributed to the fact that cancer cells produce EVs, which contributes to a higher number of EVs circulating in blood plasma from cancer patients. Similarly, the number of EVs found in the blood plasma of patients with progressive metastasis was higher than the number of EVs found in healthy volunteers. However, a lower number of EVs were found in the metastatic samples compared to the number of EVs found in blood plasma from primary breast cancer patients. This could be due to the fact that the metastatic samples are from breast cancer patients who are being treated but not responding to the cancer treatment. It is not clear how this can affect the EVs production. However, these findings should be taken cautiously as it is known that, to date, there is not a device that can measure the number of EVs with precision.

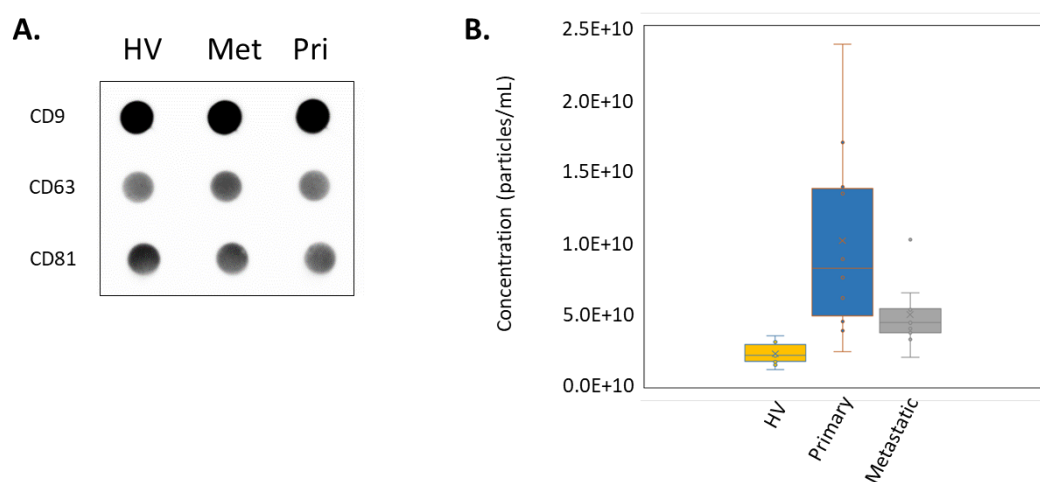
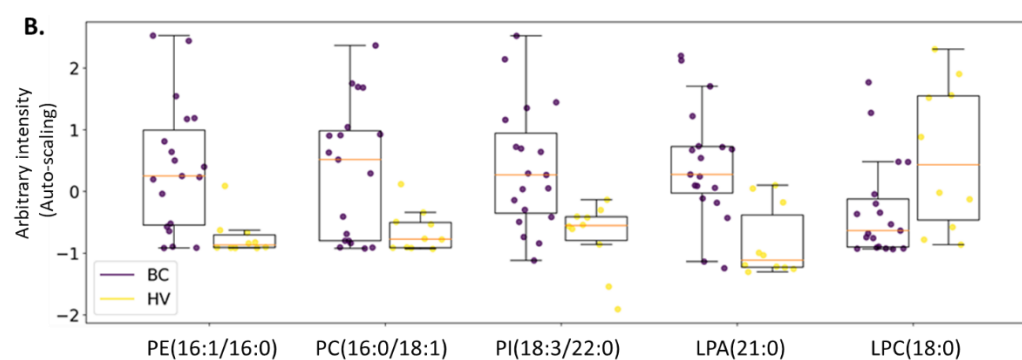
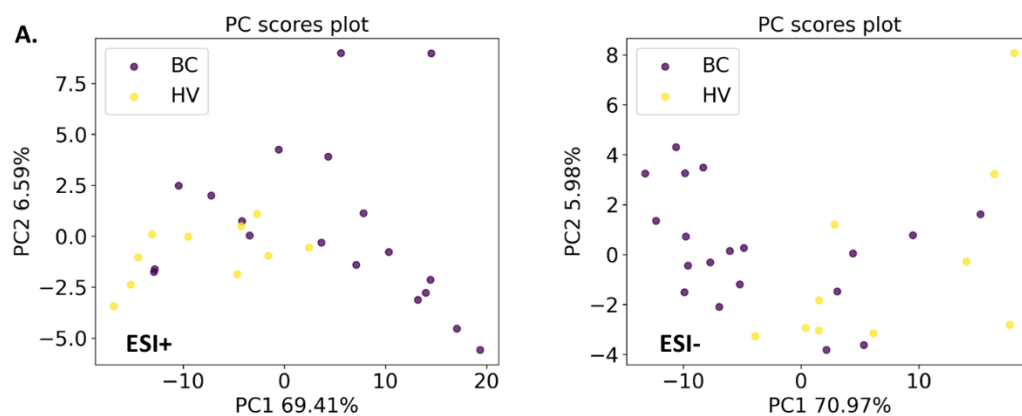


Figure 6.3.2. Protein characterisation and number of EVs isolated from blood plasma from breast cancer patients and healthy volunteers. CD9, CD81 and CD63 proteins were detected in EVs found in blood plasma from breast

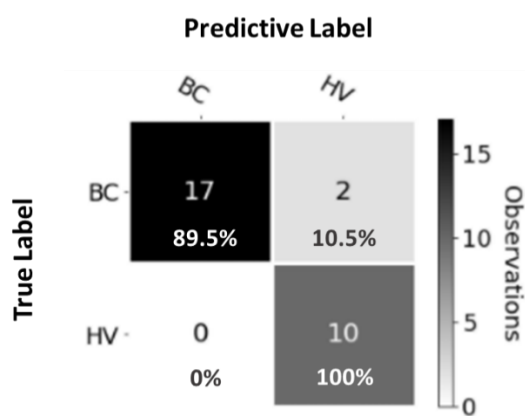
cancer and healthy volunteers, using the method for isolation of EVs from blood plasma described in Chapter 5 that combines combination of density gradient ultracentrifugation and SEC-BEC. **B.** EVs' NTA analysis.

Breast cancer samples can be distinguished from non-breast cancer samples based on the lipid composition of EVs found in blood plasma

Initially, PCA was performed for the LC-MS datasets in order to understand the distribution of the data (Figure 6.3.3A). PCA for both ESI+ and ESI- modes shows mainly separation between breast cancer samples (primary and progressive metastatic breast cancer samples) and samples from healthy volunteers, for most of the samples. The LC-MS data from the ESI+ and ESI- modes were fused and auto scaled, and then a RFE-LR analysis was performed. This analysis selected a subset of the most relevant lipid species in the datasets which allowed the identification of five statistically significant (q -value<0.05) lipid species which distinguish breast cancer samples from non-breast cancer samples (Figure 6.3.3B). These relevant lipid species included PE(16:1/16:0), PC(16:0/18:1), PI(18:3/22:0), LPA(21:0) and LPC(18:0) (Figure 6.3.3B). A leave-patient-out cross-validation analysis performed to validate the model, showed that EVs from healthy volunteers can be perfectly classified (100% accuracy) based on their lipid composition, and that EVs from breast cancer patients can be classified with an overall accuracy of approximately 90%, based on their lipid profile (Figure 6.3.3C). A ROC analysis based on the combination of the relevant lipids showed an AUROC equal to 0.94. This analysis demonstrates the discriminatory power of those lipids to differentiate between breast cancer and non-breast cancer samples and their diagnostic value for breast cancer diagnosis (Figure 6.3.3D).



C. Confusion matrix – Cross validation RFE-LR



D.

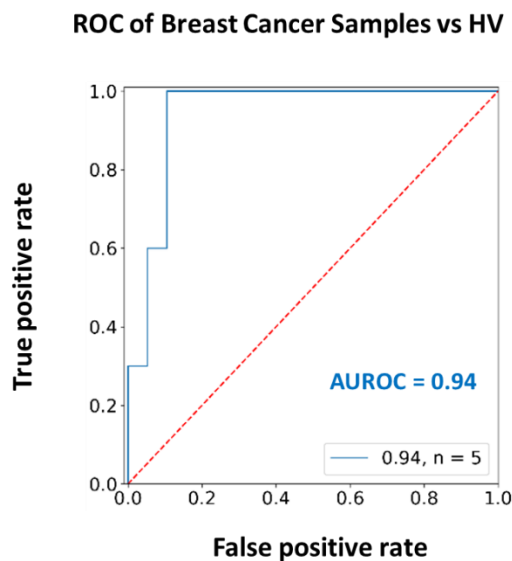


Figure 6.3.3. Statistical analysis of the LC-MS data for EVs found in blood plasma from breast cancer patients and healthy volunteers. **A.** PCA showing the LC-MS data distribution of EVs isolated from blood plasma samples from breast cancer patients (BC) and healthy volunteers (HV), for both ESI+ and ESI- modes. The BC samples include both primary and metastatic breast cancer samples. **B.** Box plots for the lipid species identified by the statistical model obtained by RFE-LR which distinguish BC samples from non-breast cancer samples. The statistical model was based on the fused and auto scaled ESI+ and ESI- datasets. Box plots are based on the auto-scaled datasets. **C.** Leave-patient-out cross-validation was performed to evaluate the classification model based on RFE-LR analysis. The confusion matrix shows that EVs isolated from HV samples can be perfectly classified (100% accuracy) based on their lipid composition, and that EVs isolated from BC samples can be classified with an overall accuracy of approximately 90%. **D.** AUROC curve for the combination of the relevant lipid species was equal to 0.94. This indicates the usefulness of the relevant lipids to distinguish between breast cancer and non-breast cancer samples.

To date, this is the first lipidomic analysis of EVs found in blood plasma from breast cancer patients. This untargeted lipidomic analysis, Chapter 6, demonstrates that the lipid composition of EVs found in blood plasma from breast cancer patients has potential for the identification of lipid biomarkers for breast cancer diagnosis. PE(16:1/16:0), PC(16:0/18:1), PI(18:3/22:0), LPA(21:0) and LPC(18:0) were identified as lipids species with potential in breast cancer diagnosis.

Specific PE and PI species were also identified as relevant lipids for breast cancer diagnosis in the study of EVs from mammary epithelial cells (Chapter 3). In addition, PC and PE have also been detected in breast cancer tissues as high abundant lipids [121, 122]. PC and PE are fundamental lipids in biological membranes which includes EVs membranes. But, they are also an important source of signalling molecules. PE, PI and PC could act as precursors of signalling molecules like PA and DG. PA is not only the precursor of many other lipids but it is also an important lipid involved in cell proliferation and cancer cell survival by regulating mTOR cell signalling. [140]. PI metabolites like phosphatidylinositol phosphates are key lipids in the activation of PI3K/AKT. This pathway is also involved in cell proliferation, growth, motility and survival, and it is altered in cancer, including breast cancer [48, 141].

PC(16:0/18:1) was identified as one of the relevant lipid species in this study, Chapter 6. This lipid species has also been identified in another study of EVs found in blood plasma from healthy volunteers [120]. In that study, PC(16:0/18:1) was the most abundant PC identified in EVs [120]. This could suggest that PC(16:0/18:1) is an abundant lipid specie in EVs found in blood plasma, and that it could act as an important source of important lipids as mentioned above, but also other metabolites. Phosphocholine can be produced from PC but also it can be a precursor of PC. High levels of phosphocholine and total choline is characteristic of cancer cells, including breast cancer cells [142]. Moreover, overexpression of choline kinase α in breast cancer has been associated with high levels of PC and contributing to cancer progression [143].

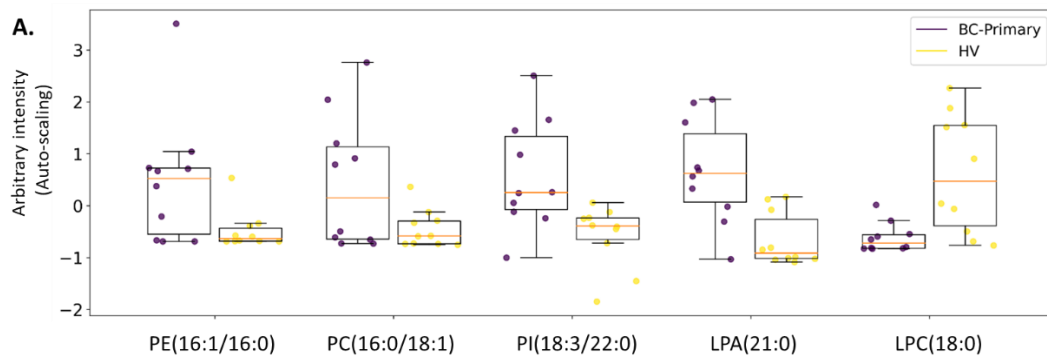
LPA(21:0) was identified between the relevant lipids in breast cancer diagnosis. LPA has also been identified in EVs found in human blood plasma [120]. The LPA identified in this study contains an odd-chain fatty acid (C21:00). Odd-carbon lipid species, including glycerophospholipids and sphingolipids species, have also been detected in both human blood plasma [144] and EVs found in blood plasma [120]. Therefore, it is not surprising that odd-chain lipid species were identified in this study and this indicates that these types of lipids are transported in EVs but are not specific to EVs found in blood plasma from cancer patients. Odd-chain fatty acids could be produced endogenously or could have a dietary origin [145]. More interesting, is the association of LPA with breast cancer metastasis, breast cancer treatment failure and postmenopausal breast cancers [146]. LPA can bind G-protein-coupled receptors and participate in the activation of Ras and Rho GTPases involved in cell proliferation, survival and migration [147, 148]. LPA receptors are abundant in breast cancer. LPA receptors have been associated with activation and upregulation of PI3K/AKT, p38-MAPK and ERK/MAPK signalling pathways, leading to cancer development and progression [146]. It has also been suggested that LPA receptors and autotaxin receptors contribute to breast cancer development and progression [146].

In addition, LPC(18:0) was found to be more abundant in healthy volunteers. This is in agreement with the findings reported by the authors of the study of EVs found in blood plasma from healthy volunteers [123]. They identified LPC(18:0) as the second more abundant LPC, after LPC(16:0). LPC(18:0) was also found enriched in EVs isolated from breast cancer cells lines (Chapter 3) when compared to their parental cells. This could suggest that LPC(18:0) is an important component of EVs. LPC could play a role in the EVs' curvature [123], and not directly in cancer cell signalling. However, LPC could be a precursor of LPA via lysophospholipase D activity of autotaxin, and as mentioned above, LPA plays an important role in cancer [149].

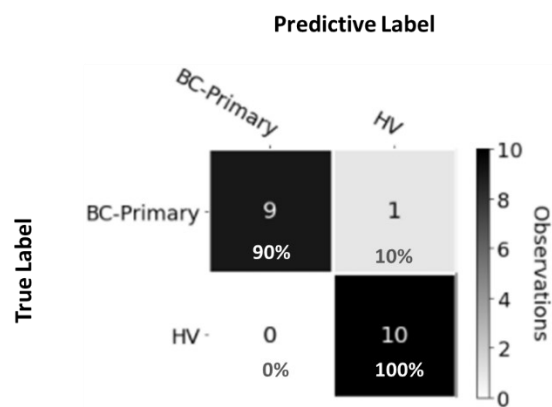
EVs' lipid composition has potential for the identification of biomarkers for early diagnosis of breast cancer

Interestingly, it was found that the same relevant lipids that allow the discrimination between breast cancer samples (primary and progressive metastatic breast cancer samples) and samples from healthy volunteers, also distinguish primary breast cancer samples from the samples from healthy volunteers (Figure 6.3.4A). A leave-patient-out cross-validation analysis was performed to evaluate this model. It shows once again that EVs from healthy volunteers can be perfectly classified (100% accuracy) based on their lipid composition, and that EVs from primary breast cancer patients cancer can be classified with an overall accuracy of 90%, based on their lipid profile (Figure 6.3.4B). Similarly, a ROC analysis shows an AUROC equal to 0.97 for the combination of the relevant lipids. This demonstrates the potential of the combination of the relevant lipids as biomarkers for primary breast cancer diagnosis (Figure 6.3.4C).

As discussed previously in this Chapter 6, as well as Chapter 3, alterations in lipid metabolism have already been documented in cancer, including breast cancer. Lipid alterations could be associated to different cancer stages but also to different subtypes of cancer. Based on the findings obtained in this study (Chapter 6) and those showed in Chapter 3, specific species of lipids in EVs have potential for breast cancer diagnosis. As found in this study (Chapter 6), a combination of PE(16:1/16:0), PC(16:0/18:1), PI(18:3/22:0), LPA(21:0) and LPC(18:0) can be used, not only to discriminate between breast cancer and non-breast cancer samples, but also to discriminate between primary breast cancer and non-breast cancer samples. This has a great impact for breast cancer diagnosis, but more importantly for the detection of breast cancer at an early stage. This is vital in cancer diagnosis as an early diagnosis is translated into earlier cancer treatment which can contribute notably to breast cancer patients' survival.



B. Confusion matrix – Cross validation RFE-LR



C. ROC of Primary Breast Cancer Samples vs HV

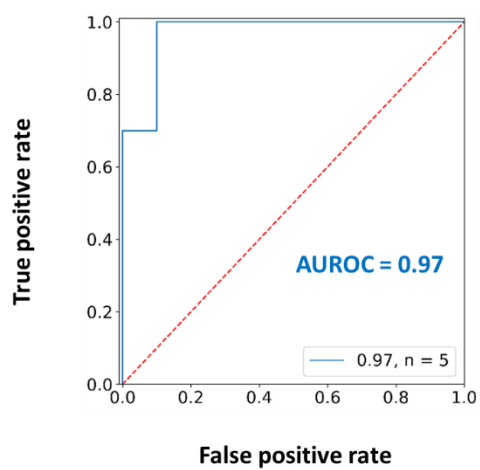


Figure 6.3.4. Statistical analysis of the LC-MS data for EVs found in blood plasma from primary breast cancer patients and healthy volunteers. **A.** Box plots for the relevant lipid species identified by the statistical model obtained by RFE- LR that distinguish breast cancer patients' samples from HV samples, but also primary breast cancer patients' samples (BC-primary) from healthy volunteer samples (HV). The statistical model was based on the fused and auto scaled ESI+ and ESI- datasets. The box plots are based on the auto-scaled datasets. **B.** Leave-patient-out cross-validation was performed to evaluate the classification model.

This analysis shows that EVs isolated from primary breast cancer patients' samples can be classified with an overall accuracy of 90%, and that EVs isolated from HV samples can be perfectly classified (100% accuracy). **C.** AUROC curve for the combination of the relevant lipid species was equal to 0.97. This indicates the usefulness of the relevant lipids to distinguish between primary breast cancer patients' samples and non-breast cancer samples.

In conclusion, the findings of this study (Chapter 6) confirm that EVs are a valuable source of lipid biomarkers for diagnosis of breast cancers with diverse histopathology and from different breast cancer subtypes. This is the first study to date of the lipid composition of EVs found in blood plasma from breast cancer patients. The findings of this study reinforce the findings of the lipidomic analysis of EVs secreted by breast cancer cell lines (Chapter 3). Furthermore, this study demonstrated the usefulness of LC-MS-based lipidomic analysis for the identification of lipid biomarkers for cancer diagnosis. The methodology used in this study demonstrated to be a valuable platform for the identification of lipid biomarkers in EVs, and can be applied to the study of other types of cancer and even other diseases. Although the results of this study are statistically significant and relevant, this is a proof-of-concept study and future studies of a higher number of samples are recommended.

Chapter 7

General discussion and Conclusions

7. General discussion and Conclusions

Lipid metabolism alterations are associated with numerous pathologies including cancer. Advances in lipidomics have contributed to the understanding of the lipid alterations associated with cancer. In addition, the potential of lipidomic analysis for cancer diagnosis and prognosis, using MS-based approaches such as LC-MS and MSI, has already been demonstrated. However, most studies have been carried out based on the analysis of tumour tissue from cancer patients. Limitations associated with obtaining tumour tissue from cancer patients could be overcome by EV-based liquid biopsies. Lipidomic analysis of EVs found in cancer patient body fluids combines the robustness and specificity of lipidomics for cancer diagnosis with the clinical advantage of using non- or minimally invasive methods for sample collection. This approach has great potential not only for cancer diagnosis, but also for cancer prognosis, screening, monitoring and stratified medicine.

EVs play an important role in intercellular communication processes and may alter the cell signalling of neighbouring or distant cells. Cancer-derived EVs can participate in the modulation of the cellular microenvironment contributing to tumorigenesis, migration and establishment of metastatic niches. Furthermore, EVs can be a source of lipid biomarkers for cancer diagnosis. Therefore, a deep understanding of the lipid composition of EVs will bring new insights not only about biological processes in cancer in which EVs can be involved, but also in the discovery of biomarkers for cancer diagnosis and prognosis.

The analysis of human mammary and ovarian epithelial cells and their secreted EVs evidenced that breast cancer-derived EVs are mainly enriched in sphingolipids (Cer and SM) and that glycerophospholipids, including lysophospholipids, are also important lipids in their lipid composition (Figure 7.1). These lipid classes were also identified in EVs isolated from human blood plasma. This is in agreement with other lipidomic studies of cancer-derived EVs from EVs isolated from conditioned media and human body fluids. The lipids identified can play a structural role in the EVs or could participate in fundamental signalling processes in cancer. For

instance, PC and PE are mainly structural lipids of biological membranes including EVs membranes, but those lipids could also act as precursors of signalling lipids. The enrichment of sphingolipids supports the idea that EVs contain lipid rafts which can act as a platform for signalling transduction. Overall, the findings suggest that EVs could actively participate in intercellular processes and especially in cancer cell signalling. Progress in the study of cancer-derived EVs will contribute to a better understanding of the biological processes in cancer development, progression and metastasis.

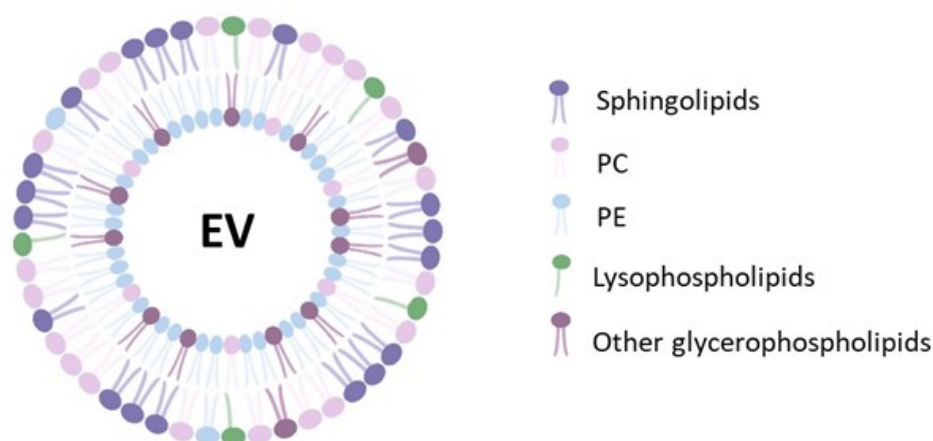


Figure 7.1. Illustration of lipids identified in EVs. Sphingolipids, including Cer and SM, were found to be abundant in EVs. Structural lipids such as PC and PE are important structural lipids of the EVs lipid composition along with lysophospholipids such as LPC, LPE and LPA. This illustration shows the most important types of lipids present in an EV, but it does not make reference to the specific proportion of the thousands of lipids contained in an EV.

The most important finding of the study of EVs secreted by cancerous and non-cancerous mammary and ovarian epithelial cells was to find that there is potential in the EVs' lipid composition for the identification of lipid biomarkers for breast and ovarian cancer diagnosis, respectively. Furthermore, it was found that it is possible to classify EVs into different cancer subtypes based on their lipid profile

which have a great impact in breast and ovarian cancer diagnosis and prognosis. Especially, this is a promising finding for ovarian cancer diagnosis. As discussed previously, ovarian cancer patients' survival is very low, as this type of cancer is mainly diagnosed at an advanced stage. This is because the early-stage symptoms are ambiguous and commonly confused with non-cancer related diseases.

The findings discussed above, obtained from the study of EVs secreted by *in vitro* cultured cells, were confirmed by lipidomic analysis of EVs found in human blood plasma. It was found that there are differences in the lipid composition of EVs found in blood plasma from breast cancer patients and healthy individuals. A combination of five specific lipid species had the diagnostic power to distinguish breast cancer samples from non-breast cancer samples. The most noteworthy finding was that the same combination of relevant lipids can distinguish between samples from patients with primary breast cancer and samples from healthy individuals. One of the limitations of the study of blood plasma samples (Chapter 6) was the number of samples analysed. However, this is a proof-of-concept study and the first study, to date, of EVs found in blood plasma from breast cancer patients. The aim of this PhD thesis was to evaluate the potential of the EVs' lipid composition for the identification of biomarkers for breast cancer diagnosis. A future study with a higher number of samples from both breast cancer patients and healthy individuals will be needed for an accurate identification and validation of lipid biomarkers for breast cancer. Future work will involve the analysis of a higher number of blood plasma samples from breast cancer patients before and after surgery. Furthermore, the lipid composition of EVs isolated from blood plasma from healthy individuals will be included and used as control samples. In addition, the distribution of lipid species in tumour tissues from breast cancer patients could be analysed by DESI-MSI and compared with the lipid content of EVs circulating in the blood plasma from the same patients analysed by LC-MS. This will allow us to evaluate the relationship between the lipid content of tumour tissues and EVs found in blood plasma. With respect to the study of ovarian cancer, the next step is to analyse the lipid composition of EVs found in body fluids such as urine and blood plasma from both ovarian cancer patients and healthy volunteers. Future analysis of potential

lipid biomarkers for breast cancer or ovarian cancer diagnosis could be carried out by designing a targeted LC-MS method for the relevant lipid biomarkers. In addition, technologies such as REIMS, which are sample preparation-free, could also be evaluated.

One of the biggest challenges in the field of EVs is the isolation of EVs from body fluids. This was also one of the biggest challenges in this PhD thesis. A method was developed for the isolation of EVs found in human blood plasma, based on the combination of density gradient ultracentrifugation and SEC-BEC. The method was reproducible and effective for the isolation of EVs from human blood plasma for lipidomic analysis. This method is useful for scientific research which requires analysis of pure EVs but not for future applications in clinical settings. The development of simple methods for the isolation of EVs from body fluids are still required. However, it is important to highlight that the experimental workflow used here for the lipidomic analysis of EVs found in blood plasma (from the isolation of EVs from human blood plasma to the identification of potential lipid biomarkers) is a useful platform not only for the study of cancer but also other diseases. Future studies could include the evaluation of simpler methods for isolation of EVs from human body fluids and their use in lipidomic analysis, with potential application in clinical settings.

In conclusion, the findings of this PhD thesis demonstrate that there is potential in the lipid composition of EVs found in cancer patients' blood plasma for the identification of lipid biomarkers for breast cancer diagnosis, and most remarkably for the detection of lipid biomarkers for early diagnosis of breast cancer. Furthermore, the findings of this PhD thesis also demonstrated that ovarian cancer-derived EVs are a promising source of lipid biomarkers for ovarian cancer diagnosis, to date, the most lethal gynaecologic cancer. These findings together support the idea that EV-based liquid biopsies could be the future for an early diagnosis of cancer and it could become a fundamental tool in cancer screening programmes.

Bibliography

1. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Abate D, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* 2019;5(12):1749-1768.
2. Dangoor A, Seddon B, Gerrand C, Grimer R, Whelan J, Judson I. UK guidelines for the management of soft tissue sarcomas. *Clin Sarcoma Res.* 2016;6:20.
3. Mawardi H, Cutler C, Treister N. Medical management update: Non-Hodgkin lymphoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod.* 2009;107(1):e19-e33.
4. Davis AS, Viera AJ, Mead MD. Leukemia: an overview for primary care. *Am Fam Physician.* 2014 May 1;89(9):731-8.
5. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-674.
6. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science.* 2017;357(6348):eaal2380.
7. WARBURG O. On the origin of cancer cells. *Science.* 1956;123(3191):309-314.
8. Koppenol WH, Bounds PL, Dang CV. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer.* 2011;11(5):325-337.

9. DeBerardinis RJ, Chandel NS. Fundamentals of cancer metabolism. *Sci Adv.* 2016;2(5):e1600200.
10. Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. *Nat Med.* 2006;12(8):895-904.
11. Lewis CE, Pollard JW. Distinct role of macrophages in different tumor microenvironments. *Cancer Res.* 2006;66(2):605-612.
12. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. *Oncogene.* 2008;27(45):5904-5912.
13. Koo MM, von Wagner C, Abel GA, McPhail S, Rubin GP, Lyratzopoulos G. Typical and atypical presenting symptoms of breast cancer and their associations with diagnostic intervals: Evidence from a national audit of cancer diagnosis. *Cancer Epidemiol.* 2017 Jun;48:140-146.
14. Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology.* 2004;233(3):830-849.
15. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology.* 2002;225(1):165-175.
16. Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening [published correction appears in *Radiology.* 2012 Jul;264(1):306]. *Radiology.* 2011;258(1):98-105.

17. Evans A, Trimboli RM, Athanasiou A, et al. Breast ultrasound: recommendations for information to women and referring physicians by the European Society of Breast Imaging. *Insights Imaging*. 2018;9(4):449-461.
18. Warner E, Plewes DB, Hill KA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA*. 2004;292(11):1317-1325.
19. Facius M, Renz DM, Neubauer H, Böttcher J, Gajda M, Camara O, Kaiser WA. Characteristics of ductal carcinoma in situ in magnetic resonance imaging. *Clin Imaging*. 2007;31:394-400.
20. Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology*. 2004;233(3):830-849.
21. Newell MS, Mahoney MC. Ultrasound-guided percutaneous breast biopsy. *Tech Vasc Interv Radiol*. 2014;17(1):23-31.
22. Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. *J Natl Cancer Inst*. 2002;94(20):1546-1554.
23. Li CI, Anderson BO, Daling JR, Moe RE. Trends in incidence rates of invasive lobular and ductal breast carcinoma. *JAMA*. 2003;289(11):1421-1424.
24. Barroso-Sousa R, Metzger-Filho O. Differences between invasive lobular and invasive ductal carcinoma of the breast: results and therapeutic implications. *Ther Adv Med Oncol*. 2016;8(4):261-266.

25. Wen HY, Brogi E. Lobular Carcinoma In Situ. *Surg Pathol Clin*. 2018;11(1):123-145. doi:10.1016/j.path.2017.09.009
26. O'Shaughnessy J. Extending survival with chemotherapy in metastatic breast cancer. *Oncologist*. 2005;10 Suppl 3:20-29.
27. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res*. 2010;12(4):207.
28. Ali S, Coombes RC. Endocrine-responsive breast cancer and strategies for combating resistance. *Nat Rev Cancer*. 2002;2(2):101-112.
29. Langlands FE, Horgan K, Dodwell DD, Smith L. Breast cancer subtypes: response to radiotherapy and potential radiosensitisation. *Br J Radiol*. 2013;86(1023):20120601.
30. Chavez KJ, Garimella SV, Lipkowitz S. Triple negative breast cancer cell lines: one tool in the search for better treatment of triple negative breast cancer. *Breast Dis*. 2010;32(1-2):35-48.
31. Sauter ER. Reliable Biomarkers to Identify New and Recurrent Cancer. *Eur J Breast Health*. 2017;13(4):162-167.
32. Polyak K. Breast cancer: origins and evolution. *J Clin Invest*. 2007;117(11):3155-3163.
33. Bast RC Jr, Hennessy B, Mills GB. The biology of ovarian cancer: new opportunities for translation. *Nat Rev Cancer*. 2009;9(6):415-428.

34. Prat J. Ovarian carcinomas: five distinct diseases with different origins, genetic alterations, and clinicopathological features. *Virchows Arch.* 2012;460(3):237-249.
35. Prat J. New insights into ovarian cancer pathology. *Ann Oncol.* 2012;23 Suppl 10:x111-x117.
36. Kurman RJ, Shih IeM. Molecular pathogenesis and extraovarian origin of epithelial ovarian cancer--shifting the paradigm. *Hum Pathol.* 2011;42(7):918-931.
37. Muyldermans M, Cornillie FJ, Koninckx PR. CA125 and endometriosis. *Hum Reprod Update.* 1995;1(2):173-187.
38. Chandra A, Pius C, Nabeel M, et al. Ovarian cancer: Current status and strategies for improving therapeutic outcomes. *Cancer Med.* 2019;8(16):7018-7031.
39. Fahy E, Subramaniam S, Brown HA, et al. A comprehensive classification system for lipids. *J Lipid Res.* 2010 Jun;51(6):1618].
40. Fahy E, Subramaniam S, Murphy RC, et al. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res.* 2009;50 Suppl(Suppl):S9-S14.
41. van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol.* 2008;9(2):112-124.
42. Lingwood D, Simons K. Lipid rafts as a membrane-organizing principle. *Science.* 2010;327(5961):46-50.

43. DeBerardinis RJ, Mancuso A, Daikhin E, et al. Beyond aerobic glycolysis: transformed cells can engage in glutamine metabolism that exceeds the requirement for protein and nucleotide synthesis. *Proc Natl Acad Sci U S A*. 2007;104(49):19345-19350.
44. Daniëls VW, Smans K, Royaux I, Chypre M, Swinnen JV, Zaidi N. Cancer cells differentially activate and thrive on de novo lipid synthesis pathways in a low-lipid environment. *PLoS One*. 2014;9(9):e106913.
45. Rysman E, Brusselmans K, Scheys K, et al. De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Res*. 2010;70(20):8117-8126.
46. Fang Y, Vilella-Bach M, Bachmann R, Flanigan A, Chen J. Phosphatidic acid-mediated mitogenic activation of mTOR signaling. *Science*. 2001;294(5548):1942-1945.
47. Mills GB, Moolenaar WH. The emerging role of lysophosphatidic acid in cancer. *Nat Rev Cancer*. 2003;3(8):582-591.
48. Wymann MP, Schneider R. Lipid signalling in disease. *Nat Rev Mol Cell Biol*. 2008;9(2):162-176.
49. Mollinedo F, Gajate C. Lipid rafts as major platforms for signaling regulation in cancer. *Adv Biol Regul*. 2015;57:130-146.
50. Colombo M, Raposo G, Théry C. Biogenesis, secretion, and intercellular interactions of exosomes and other extracellular vesicles. *Annu Rev Cell Dev Biol*. 2014;30:255-289.

51. Buzas EI, György B, Nagy G, Falus A, Gay S. Emerging role of extracellular vesicles in inflammatory diseases. *Nat Rev Rheumatol*. 2014;10(6):356-364.
52. Théry C, Witwer KW, Aikawa E, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J Extracell Vesicles*. 2018;7(1):1535750.
53. Raposo G, Stoorvogel W. Extracellular vesicles: exosomes, microvesicles, and friends. *J Cell Biol*. 2013;200(4):373-383.
54. Keller S, Sanderson MP, Stoeck A, Altevogt P. Exosomes: from biogenesis and secretion to biological function. *Immunol Lett*. 2006;107(2):102-108.
55. Svensson KJ, Christianson HC, Wittrup A, et al. Exosome uptake depends on ERK1/2-heat shock protein 27 signaling and lipid Raft-mediated endocytosis negatively regulated by caveolin-1. *J Biol Chem*. 2013;288(24):17713-17724.
56. Parolini I, Federici C, Raggi C, et al. Microenvironmental pH is a key factor for exosome traffic in tumor cells. *J Biol Chem*. 2009;284(49):34211-34222.
57. Feng D, Zhao WL, Ye YY, et al. Cellular internalization of exosomes occurs through phagocytosis. *Traffic*. 2010;11(5):675-687.
58. Fitzner D, Schnaars M, van Rossum D, et al. Selective transfer of exosomes from oligodendrocytes to microglia by macropinocytosis. *J Cell Sci*. 2011;124(Pt 3):447-458.

59. Wang J, Zheng Y, Zhao M. Exosome-Based Cancer Therapy: Implication for Targeting Cancer Stem Cells. *Front Pharmacol*. 2017;7:533.
60. Gámez-Valero A, Monguió-Tortajada M, Carreras-Planella L, Franquesa Ml, Beyer K, Borràs FE. Size-Exclusion Chromatography-based isolation minimally alters Extracellular Vesicles' characteristics compared to precipitating agents. *Sci Rep*. 2016;6:33641.
61. Clayton A, Court J, Navabi H, et al. Analysis of antigen presenting cell derived exosomes, based on immuno-magnetic isolation and flow cytometry. *J Immunol Methods*. 2001;247(1-2):163-174.
62. Théry C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr Protoc Cell Biol*. 2006;Chapter 3.
63. Van Deun J, Mestdagh P, Sormunen R, et al. The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. *J Extracell Vesicles*. 2014;3:10.3402/jev.v3.24858.
64. Whittaker TE, Nagelkerke A, Nele V, Kauscher U, Stevens MM. Experimental artefacts can lead to misattribution of bioactivity from soluble mesenchymal stem cell paracrine factors to extracellular vesicles. *J Extracell Vesicles*. 2020;9(1):1807674.
65. Corso G, Mäger I, Lee Y, et al. Reproducible and scalable purification of extracellular vesicles using combined bind-elute and size exclusion chromatography. *Sci Rep*. 2017;7(1):11561.
66. Karimi N, Cvjetkovic A, Jang SC, et al. Detailed analysis of the plasma extracellular vesicle proteome after separation from lipoproteins. *Cell Mol Life Sci*. 2018;75(15):2873-2886.

67. Onódi Z, Pelyhe C, Terézia Nagy C, et al. Isolation of High-Purity Extracellular Vesicles by the Combination of Iodixanol Density Gradient Ultracentrifugation and Bind-Elute Chromatography From Blood Plasma. *Front Physiol.* 2018;9:1479.
68. Vidal M, Sainte-Marie J, Philippot JR, Bienvenue A. Asymmetric distribution of phospholipids in the membrane of vesicles released during in vitro maturation of guinea pig reticulocytes: evidence precluding a role for "aminophospholipid translocase". *J Cell Physiol.* 1989;140(3):455-462.
69. Wubbolts R, Leckie RS, Veenhuizen PT, et al. Proteomic and biochemical analyses of human B cell-derived exosomes. Potential implications for their function and multivesicular body formation. *J Biol Chem.* 2003;278(13):10963-10972.
70. Ventimiglia LN, Alonso MA. Biogenesis and Function of T Cell-Derived Exosomes. *Front Cell Dev Biol.* 2016;4:84.
71. Laulagnier K, Motta C, Hamdi S, et al. Mast cell- and dendritic cell-derived exosomes display a specific lipid composition and an unusual membrane organization. *Biochem J.* 2004;380(Pt 1):161-171.
72. Trajkovic K, Hsu C, Chiantia S, et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science.* 2008;319(5867):1244-1247.
73. Brouwers JF, Aalberts M, Jansen JW, et al. Distinct lipid compositions of two types of human prostasomes. *Proteomics.* 2013;13(10-11):1660-1666.

74. Brzozowski JS, Jankowski H, Bond DR, et al. Lipidomic profiling of extracellular vesicles derived from prostate and prostate cancer cell lines. *Lipids Health Dis.* 2018;17(1):211.
75. Llorente A, Skotland T, Sylvänne T, et al. Molecular lipidomics of exosomes released by PC-3 prostate cancer cells. *Biochim Biophys Acta.* 2013;1831(7):1302-1309.
76. Skotland T, Ekroos K, Kauhanen D, et al. Molecular lipid species in urinary exosomes as potential prostate cancer biomarkers. *Eur J Cancer.* 2017;70:122-132.
77. Haraszti RA, Didiot MC, Sapp E, et al. High-resolution proteomic and lipidomic analysis of exosomes and microvesicles from different cell sources. *J Extracell Vesicles.* 2016;5:32570.
78. Nishida-Aoki N, Izumi Y, Takeda H, Takahashi M, Ochiya T, Bamba T. Lipidomic Analysis of Cells and Extracellular Vesicles from High- and Low-Metastatic Triple-Negative Breast Cancer. *Metabolites.* 2020;10(2):67.
79. Cheng L, Zhang K, Qing Y, et al. Proteomic and lipidomic analysis of exosomes derived from ovarian cancer cells and ovarian surface epithelial cells. *J Ovarian Res.* 2020;13(1):9.
80. Logozzi M, De Milito A, Lugini L, et al. High levels of exosomes expressing CD63 and caveolin-1 in plasma of melanoma patients. *PLoS One.* 2009;4(4):e5219.
81. Tickner JA, Urquhart AJ, Stephenson SA, Richard DJ, O'Byrne KJ. Functions and therapeutic roles of exosomes in cancer. *Front Oncol.* 2014;4:127.

82. Bell E, Taylor MA. Functional Roles for Exosomal MicroRNAs in the Tumour Microenvironment. *Comput Struct Biotechnol J*. 2016;15:8-13.
83. Boelens MC, Wu TJ, Nabet BY, et al. Exosome transfer from stromal to breast cancer cells regulates therapy resistance pathways. *Cell*. 2014;159(3):499-513.
84. Chevillet JR, Kang Q, Ruf IK, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc Natl Acad Sci U S A*. 2014;111(41):14888-14893.
85. Duijvesz D, Burnum-Johnson KE, Gritsenko MA, et al. Proteomic profiling of exosomes leads to the identification of novel biomarkers for prostate cancer. *PLoS One*. 2013;8(12):e82589.
86. Rodríguez M, Bajo-Santos C, Hessvik NP, et al. Identification of non-invasive miRNAs biomarkers for prostate cancer by deep sequencing analysis of urinary exosomes. *Mol Cancer*. 2017;16(1):156.
87. Yan S, Han B, Gao S, et al. Exosome-encapsulated microRNAs as circulating biomarkers for colorectal cancer. *Oncotarget*. 2017;8(36):60149-60158.
88. Cazzoli R, Buttitta F, Di Nicola M, et al. microRNAs derived from circulating exosomes as noninvasive biomarkers for screening and diagnosing lung cancer. *J Thorac Oncol*. 2013;8(9):1156-1162.
89. Taylor DD, Gercel-Taylor C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol Oncol*. 2010 Jan;116(1):153].

90. De Hoffmann E, Stroobant V. Mass spectrometry: principles and applications. 3rd Edition. 2007;John Wiley & Sons, Ltd.
91. Domon B, Aebersold R. Mass spectrometry and protein analysis. *Science*. 2006;312(5771):212-217.
92. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM. Electrospray ionization for mass spectrometry of large biomolecules. *Science*. 1989;246(4926):64-71.
93. Banerjee S, Mazumdar S. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem*. 2012;2012:282574.
94. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new mass spectrometer. *J Mass Spectrom*. 2005;40(4):430-443.
95. Vorkas PA, Isaac G, Anwar MA, et al. Untargeted UPLC-MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease. *Anal Chem*. 2015;87(8):4184-4193.
96. Eiriksson FF, Nøhr MK, Costa M, Bödvarsdóttir SK, Ögmundsdóttir HM, Thorsteinsdóttir M. Lipidomic study of cell lines reveals differences between breast cancer subtypes. *PLoS One*. 2020;15(4):e0231289.
97. Nagahashi M, Tsuchida J, Moro K, et al. High levels of sphingolipids in human breast cancer. *J Surg Res*. 2016;204(2):435-444.
98. Kang HS, Lee SC, Park YS, et al. Protein and lipid MALDI profiles classify breast cancers according to the intrinsic subtype. *BMC Cancer*. 2011;11:465.

99. Guenther S, Muirhead LJ, Speller AV, et al. Spatially resolved metabolic phenotyping of breast cancer by desorption electrospray ionization mass spectrometry. *Cancer Res.* 2015;75(9):1828-1837.
100. Porcari AM, Zhang J, Garza KY, Rodrigues-Peres RM, Lin JQ, Young JH, Tibshirani R, Nagi C, Paiva GR, Carter SA, Sarian LO, Eberlin MN, Eberlin LS. Multicenter Study Using Desorption-Electrospray-Ionization-Mass-Spectrometry Imaging for Breast-Cancer Diagnosis. *Anal Chem.* 2018;90(19):11324-11332.
101. St John ER, Balog J, McKenzie JS, et al. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. *Breast Cancer Res.* 2017;19(1):59.
102. Sutphen R, Xu Y, Wilbanks GD, et al. Lysophospholipids are potential biomarkers of ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2004;13(7):1185-1191.
103. Zhang T, Wu X, et al. Discrimination between malignant and benign ovarian tumors by plasma metabolomic profiling using ultra performance liquid chromatography/mass spectrometry. *Clin Chim Acta.* 2012;413(9-10):861-8.
104. Buas MF, Gu H, Djukovic D, et al. Identification of novel candidate plasma metabolite biomarkers for distinguishing serous ovarian carcinoma and benign serous ovarian tumors. *Gynecol Oncol.* 2016;140(1):138-144.
105. Dória ML, McKenzie JS, Mroz A, et al. Epithelial ovarian carcinoma diagnosis by desorption electrospray ionization mass spectrometry imaging. *Sci Rep.* 2016;6:39219.

106. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem.* 2006;78(3):779-87.
107. Pluskal T, Castillo S, Villar-Briones A, Oresic M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics.* 2010 Jul 23;11:395.
108. Röst HL, Sachsenberg T, Aiche S, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods.* 2016;13(9):741-748.
109. Veselkov KA, Vingara LK, Masson P, et al. Optimized preprocessing of ultra-performance liquid chromatography/mass spectrometry urinary metabolic profiles for improved information recovery. *Anal Chem.* 2011;83(15):5864-5872.
110. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ¹H NMR metabonomics. *Anal Chem.* 2006;78(13):4281-4290.
111. van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics.* 2006;7:142.
112. Trygg J, Holmes E, Lundstedt T. Chemometrics in metabonomics. *J Proteome Res.* 2007;6(2):469-479.

113. Barker, M. and W. Rayens, Partial least squares for discrimination. *J Chemometr.* 2003;17(3):166-173.
114. Bylesjo M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemometr.* 2006; 20: 341–351.
115. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* 2002;46:89–422.
116. Horgan CC, Nagelkerke A, Whittaker TE, et al. Molecular imaging of extracellular vesicles in vitro via Raman metabolic labelling. *J Mater Chem B.* 2020;8(20):4447-4459.
117. Kauscher U, Penders J, Nagelkerke A, et al. Gold Nanocluster Extracellular Vesicle Supraparticles: Self-Assembled Nanostructures for Three-Dimensional Uptake Visualization. *Langmuir.* 2020;36(14):3912-3923.
118. BLIGH EG, DYER WJ. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol.* 1959;37(8):911-917.
119. Hosseini-Beheshti E, Pham S, Adomat H, Li N, Tomlinson Guns ES. Exosomes as biomarker enriched microvesicles: characterization of exosomal proteins derived from a panel of prostate cell lines with distinct AR phenotypes. *Mol Cell Proteomics.* 2012;11(10):863-885.
120. Jakubec M, Maple-Grødem J, Akbari S, Nesse S, Halskau Ø, Mork-Jansson AE. Plasma-derived exosome-like vesicles are enriched in lysophospholipids and pass the blood-brain barrier. *PLoS One.* 2020;15(9):e0232442.

121. Ide Y, Waki M, Hayasaka T, et al. Human breast cancer tissues contain abundant phosphatidylcholine(36:1) with high stearoyl-CoA desaturase-1 expression [published correction appears in PLoS One. 2013;8(9).
122. Punnonen K, Hietanen E, Auvinen O, Punnonen R. Phospholipids and fatty acids in breast cancer tissue. *J Cancer Res Clin Oncol*. 1989;115(6):575-578.
123. Fuller N, Rand RP. The influence of lysolipids on the spontaneous curvature and bending elasticity of phospholipid membranes. *Biophys J*. 2001;81(1):243-254.
124. de Gassart A, Geminard C, Fevrier B, Raposo G, Vidal M. Lipid raft-associated protein sorting in exosomes. *Blood*. 2003;102(13):4336-4344.
125. Gao X, Lowry PR, Zhou X, et al. PI3K/Akt signaling requires spatial compartmentalization in plasma membrane microdomains. *Proc Natl Acad Sci U S A*. 2011;108(35):14509-14514.
126. Trajkovic K, Hsu C, Chiantia S, et al. Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science*. 2008;319(5867):1244-1247.
127. Phuyal S, Hessvik NP, Skotland T, Sandvig K, Llorente A. Regulation of exosome release by glycosphingolipids and flotillins. *FEBS J*. 2014;281(9):2214-2227.
128. Dean JM, Lodhi IJ. Structural and functional roles of ether lipids. *Protein Cell*. 2018;9(2):196-206.
129. Koundouros N, Poulogiannis G. Reprogramming of fatty acid metabolism in cancer. *Br J Cancer*. 2020;122(1):4-22.

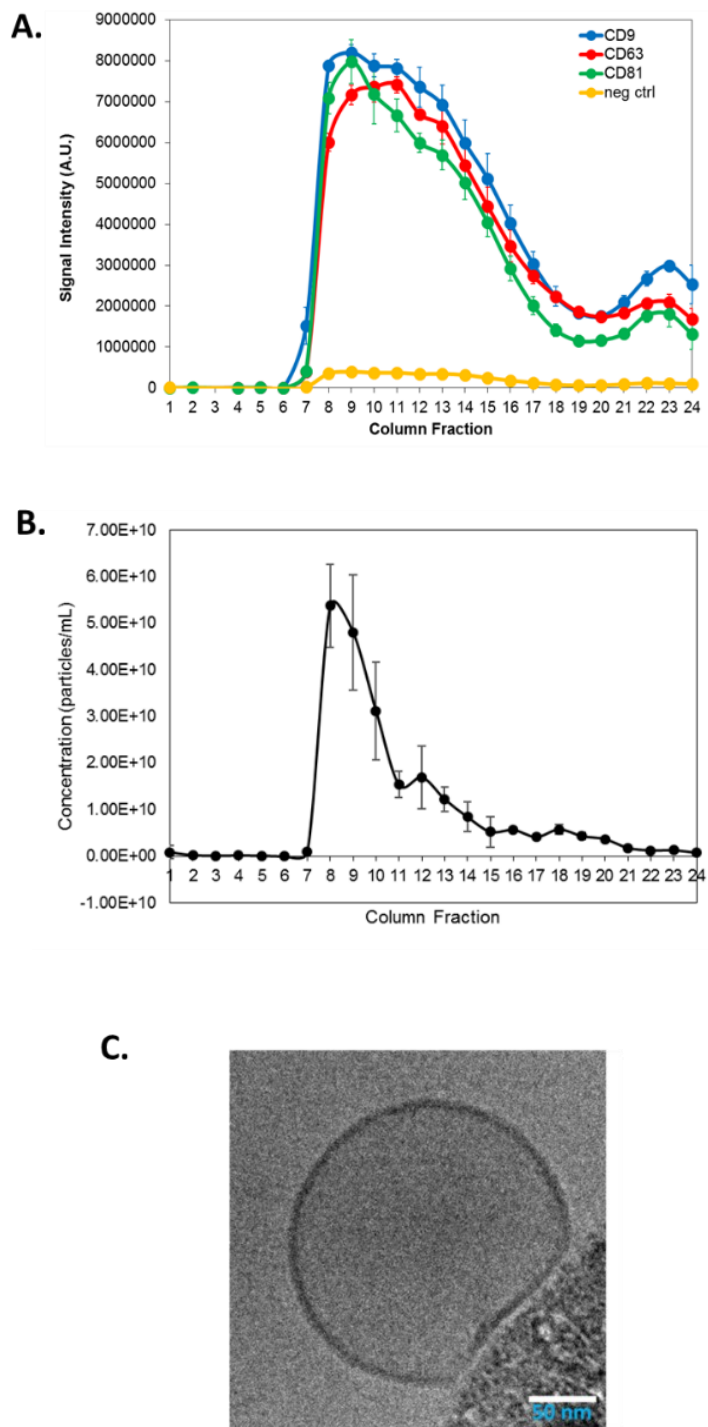
130. Benjamin DI, Cozzo A, Ji X, et al. Ether lipid generating enzyme AGPS alters the balance of structural and signaling lipids to fuel cancer pathogenicity. *Proc Natl Acad Sci U S A*. 2013;110(37):14912-14917.
131. Janku F, Yap TA, Meric-Bernstam F. Targeting the PI3K pathway in cancer: are we making headway?. *Nat Rev Clin Oncol*. 2018;15(5):273-291.
132. Fruman DA, Chiu H, Hopkins BD, Bagrodia S, Cantley LC, Abraham RT. The PI3K Pathway in Human Disease. *Cell*. 2017;170(4):605-635.
133. Ryland LK, Fox TE, Liu X, Loughran TP, Kester M. Dysregulation of sphingolipid metabolism in cancer. *Cancer Biol Ther*. 2011;11(2):138-149.
134. Ogretmen B. Sphingolipid metabolism in cancer signalling and therapy. *Nat Rev Cancer*. 2018;18(1):33-50.
135. Takabe K, Paugh SW, Milstien S, Spiegel S. "Inside-out" signaling of sphingosine-1-phosphate: therapeutic targets. *Pharmacol Rev*. 2008;60(2):181-195.
136. Nakajima M, Nagahashi M, Rashid OM, Takabe K, Wakai T. The role of sphingosine-1-phosphate in the tumor microenvironment and its clinical implications. *Tumour Biol*. 2017;39(4):1010428317699133.
137. Stemke-Hale K, Gonzalez-Angulo AM, Lluch A, et al. An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res*. 2008;68(15):6084-6091.
138. Baselga J. Targeting the phosphoinositide-3 (PI3) kinase pathway in breast cancer. *Oncologist*. 2011;16 Suppl 1:12-19.

139. Hernandez-Aya LF, Gonzalez-Angulo AM. Targeting the phosphatidylinositol 3-kinase signaling pathway in breast cancer. *Oncologist*. 2011;16(4):404-414.
140. Foster DA. Phosphatidic acid signaling to mTOR: signals for the survival of human cancer cells. *Biochim Biophys Acta*. 2009;1791(9):949-955.
141. Luo J, Manning BD, Cantley LC. Targeting the PI3K-Akt pathway in human cancer: rationale and promise. *Cancer Cell*. 2003;4(4):257-262.
142. Ackerstaff E, Glunde K, Bhujwalla ZM. Choline phospholipid metabolism: a target in cancer cells?. *J Cell Biochem*. 2003;90(3):525-533.
143. Sanchez-Lopez E, Zimmerman T, Gomez del Pulgar T, Moyer MP, Lacal Sanjuan JC, Cebrian A. Choline kinase inhibition induces exacerbated endoplasmic reticulum stress and triggers apoptosis via CHOP in cancer cells. *Cell Death Dis*. 2013;4(11):e933.
144. Huynh K, Barlow CK, Jayawardana KS, et al. High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with Cardiometabolic Risk Factors. *Cell Chem Biol*. 2019;26(1):71-84.e4.
145. Jenkins BJ, Seyssel K, Chiu S, et al. Odd Chain Fatty Acids; New Insights of the Relationship Between the Gut Microbiota, Dietary Intake, Biosynthesis and Glucose Intolerance. *Sci Rep*. 2017;7:44845.
146. Liu S, Umezu-Goto M, Murph M, et al. Expression of autotaxin and lysophosphatidic acid receptors increases mammary tumorigenesis, invasion, and metastases [published correction appears in *Cancer Cell*. 2009 Aug 4;16(2):172]. *Cancer Cell*. 2009;15(6):539-550.

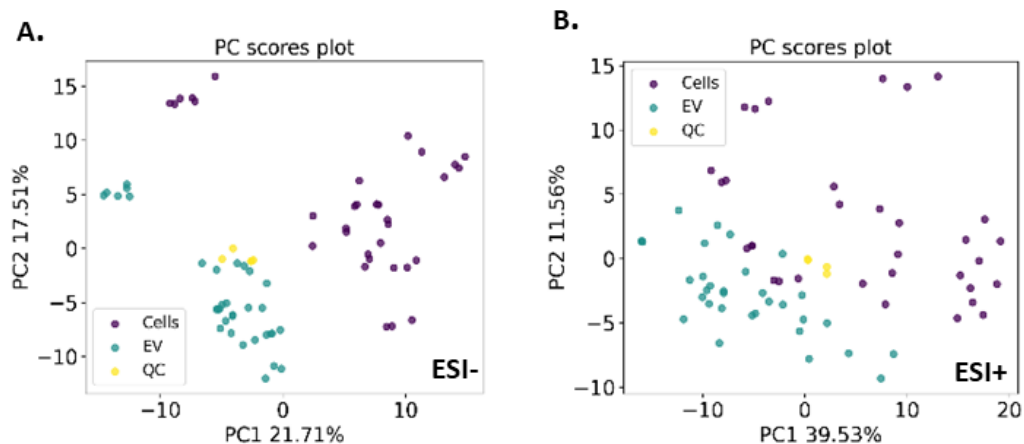
147. Moolenaar WH, van Meeteren LA, Giepmans BN. The ins and outs of lysophosphatidic acid signaling. *Bioessays*. 2004;26(8):870-881. Bar-Sagi D, Hall A. Ras and Rho GTPases: a family reunion. *Cell*. 2000;103(2):227-238.
148. Bar-Sagi D, Hall A. Ras and Rho GTPases: a family reunion. *Cell*. 2000;103(2):227-238.
149. Umezū-Goto M, Kishi Y, Taira A, et al. Autotaxin has lysophospholipase D activity leading to tumor cell growth and motility by lysophosphatidic acid production. *J Cell Biol*. 2002;158(2):227-233.

Appendix I.

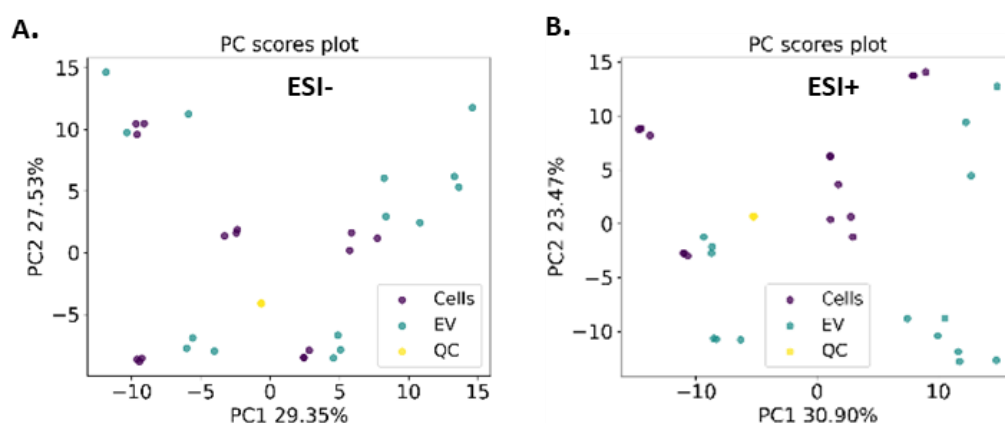
Supplementary Figures



Supplementary Figure 1. Characterisation of EVs isolated from mammary epithelial cells. EV markers (CD9, CD63 and CD81) evaluated for MDA-MB-231, measured using dot blot analysis. **B.** Particle concentration, measured with nanoparticle tracking analysis. **C.** CryoTEM image from MDA-MB-231 EVs, confirming their vesicular morphology. These analyses and figures were carried by the Stevens group.



Supplementary Figure 2. PCA of the LC-MS data acquired for the human mammary epithelial cells and their EVs. The data was acquired in both **A.** ESI- and **B.** ESI+. These PCA show the LC-MS data distribution for cells (purple) and their derived EVs (green) as well as for the QC samples (yellow). QC samples (pool of all the samples) are clustered together indicating a good quality of the LC-MS data in both ESI- and ESI+.



Supplementary Figure 3. PCA of the LC-MS data acquired for the human ovarian epithelial cells and their EVs. The data was acquired in both **A. ESI-** and **B. ESI+**. These PCA show the LC-MS data distribution for cells (purple) and their derived EVs (green) as well as for the QC samples (yellow). QC samples (pool of all the samples) are clustered together indicating a good quality of the LC-MS data in both ESI- and ESI+.

Supplementary Table 1. List of lipid standards analysed to test the quality of the LC-MS data and mass spectrometers.

Internal standard	Reference number
LPC(9:0)	855276P (Avanti)
PC(11:0/11:0)	850330P (Avanti)
FA(17:0)	H3500 (Sigma)
PG(15:0/15:0)	840446P (Avanti)
PE(15:0/15:0)	850704P (Avanti)
PS(17:0/17:0)	840028P (Avanti)
PA(17:0/17:0)	830856P (Avanti)
Cer(d18:1/17:0)	860517P (Avanti)
DG(19:0/19:0)	68633 (Sigma)
PC(23:0/23:0)	850372P (Avanti)
TG(15:0/15:0/15:0)	T4257 (Sigma)
TG(17:0/17:0/17:0)	T2151 (Sigma)

Supplementary methods.

Human mammary epithelial cells *in vitro* culture

All cell lines except MCF-10A and HuMEC were maintained in high glucose Dulbecco's Modified Eagle's medium (DMEM) (ThermoFisher, Gibco) supplemented with 10% foetal bovine serum, 1 x penicillin/streptomycin, 10 mM HEPES and 1 x non-essential amino acids. MCF-10A cells were maintained in supplemented DMEM/F12 medium as previously described³⁶. HuMEC cells were maintained in HuMEC medium with HuMEC supplement (Thermo Fisher). All cells lines were incubated at 37 °C in 5% CO₂.

CryoTEM

Samples for CryoTEM were prepared using an automatic plunge freezer (Leica EM GP). To summarise, a small aliquot of sample (4 µL) was injected on plasma-treated (O₂/H₂ for 15 seconds) HoleyCarbon copper grids in an environmental chamber (relative humidity: 90%, temperature: 20 °C). Excess suspension was blotted on filter paper and the obtained film was vitrified in liquid ethane. Samples were stored in liquid nitrogen and imaged at -170 °C (Gatan cryo-holder) in a JEOL 2100Plus transmission electron microscope. Micrographs were taken at 200 kV in low electron dose mode.

Appendix II.

ORAL PRESENTATIONS

- Title: Lipidomic profiling of cancer-derived extracellular vesicles and its potential for the identification of body fluid-based biomarkers for breast cancer diagnosis.
UKEV Forum 2019. UK Society for Extracellular Vesicles. London, UK.
- Title: Lipidomic profiling of extracellular vesicles released by breast cancer cells and their potential as body-fluid based biomarkers for breast cancer detection.
MSACL 2019 US. The Association for Mass Spectrometry and Advances in the Clinical Lab. Palm Springs, USA.
 - Award - Young Investigator Travel Grant

POSTER PRESENTATIONS

- Title: Lipidomic analysis of extracellular vesicles and its potential for the identification of body fluid-based biomarkers for breast cancer diagnosis.
Authors: **Erika Dorado**, M Luisa Doria, Anika Nagelkerke, James S McKenzie, Stefania Maneta-Stavarakaki, Charlotte Ion, Thomas Whittaker, Jeremy Nicholson, Molly M Stevens, R Charles Coombes, Zoltan Takats.
AACR Annual Meeting 2021. American Association for Cancer Research. Virtual meeting.
- Title: Lipid composition of cancer-derived extracellular vesicles and its potential for the identification of body fluid-based biomarkers for breast cancer diagnosis.
Authors: **Erika Dorado**, M Luisa Doria, Anika Nagelkerke, James S McKenzie, Stefania Maneta-Stavarakaki, Ulrike Kauscher, R Charles Coombes, Jeremy Nicholson, Molly M Stevens, Zoltan Takats.
MSACL annual conference. Poster presentation. Salzburg, Austria. 2019.
- Title: Lipidomic profiling of extracellular vesicles secreted by breast cancer cells. Authors: **Erika Dorado**, M Luisa Doria, Anika Nagelkerke, Thomas Whittaker, Ulrike Kauscher, R Charles Coombes, Jeremy Nicholson, Molly M Stevens, Zoltan Takats.
38th BMSS Annual Meeting. British Mass Spectrometry Society. Cambridge, UK. 2018.

- Title: Lipidomic profiling of breast cancer exosomes and their parental cell lines. Authors: **Erika Dorado**, M Luisa Doria, Anika Nagelkerke, Thomas Whittaker, Ulrike Kauscher, R Charles Coombes, Jeremy Nicholson, Molly M Stevens, Zoltan Takats. 66th ASMS conference. American Society for Mass Spectrometry. San Diego, USA. 2018.

PUBLICATIONS

Abstract 459: Lipidomic analysis of extracellular vesicles and its potential for the identification of body fluid-based biomarkers for breast cancer diagnosis. **Erika Dorado**, M Luisa Doria, Anika Nagelkerke, James S McKenzie, Stefania Maneta-Stavarakaki, Charlotte Ion. Thomas Whittaker, Jeremy Nicholson, Molly M Stevens, R Charles Coombes and Zoltan Takats. Cancer Res July 1 2021 (81) (13 Supplement 469; DOI: 10.1158/1538-7445.AM2021-459.

Erika Dorado, M Luisa Doria, Anika Nagelkerke, James McKenzie, Stefania Maneta-Stavarakaki, Charlotte Ion, Brian Cheng, Ulrike Kausher, Jeremy Nicholson, R Charles Coombes, Molly M Stevens, Zoltan Takats. **Extracellular vesicles as a promising source of lipid biomarkers for early detection of breast cancer.** In preparation.

Erika Dorado, James McKenzie, Stefania Maneta-Stavarakaki, Brian Cheng, Jeremy Nicholson, Molly M Stevens, Zoltan Takats. **Lipid composition of cancer-derived extracellular vesicles and its potential for identification of biomarkers for ovarian cancer diagnosis.** In preparation.