Electronic Thesis and Dissertation Repository

6-27-2018 11:00 AM

Proteomic Characterization of Ovarian and Breast Cancer Microenvironments for Improved Diagnostics and Therapeutic Targeting

Dylan Z. Dieters-Castator
The University of Western Ontario

Supervisor
Postovit, Lynne-Marie
The University of Western Ontario Co-Supervisor
Lajoie, Gilles A
The University of Western Ontario

Follow this and additional works at: https://ir.lib.uwo.ca/etd

🌀 Part of the Cancer Biology Commons

Abstract

Cancers exist within complex microenvironments formed by heterogeneous cell types. This diversity creates significant challenges for detection, diagnosis and treatment. Mass spectrometry-based proteomics is a powerful approach capable of characterizing complex biological systems which are characteristic of cancer biology. In this thesis, proteomics was utilized to answer several questions related to ovarian cancer diagnosis and detection, and the effects of NODAL, an embryonic morphogen, on the breast cancer secretome and stromal cell recruitment. First, I compared multiple sample preparation techniques and found high-pH/low-pH fractionation to yield the greatest proteome coverage over commonly used approaches. Second, I compared the proteomes from two ovarian cancer subtypes (high-grade serous and endometrioid) for which histological discrimination remains difficult in a proportion of cases. I documented several unknown proteins, including KIAA1324, which were validated and confirmed to improve the differential diagnosis of endometrial ovarian cancer. Third, I extensively characterized extracellular vesicle proteomes from biological fluids (conditioned media, plasma and ascites) to catalogue potential biomarkers associated with malignant ovarian cancer. I detected many factors associated with advanced stage, high-grade serous ovarian cancer including CFHR4, MUC1, APCS and PZP that may be useful for early detection. Last, I characterized the global effects of the Transforming Growth Factor-β superfamily member NODAL on the breast cancer secretome and stromal cell recruitment *in vitro*. I found a previously unknown role for NODAL in modulating pro-inflammatory factors, including CXCL1 and IL6 that were correlated with multipotent stromal cell recruitment. In summary, this work represents a significant contribution to the histological assessment and detection of ovarian cancer and our understanding of the malignant properties of NODAL within the breast cancer microenvironment.

Keywords

Co-Authorship Statement

Each chapter constitutes a manuscript that has been published or in preparation to be submitted to a peer-reviewed journal. Contributions from all authors for each chapter are as follows:

Chapter 2: Comparison of sample preparation techniques for large scale proteomics

Chapter 2 was accepted for publication in Proteomics as a Research Article. In order, the authors include Kuljanin M, Dieters-Castator DZ, Hess DA, Postovit LM, and Lajoie GA. MK and DZDC contributed equally to study as co-first authors and performed experiments, analyzed and interpreted the results, and wrote the manuscript. DAH, LMP and GAL contributed to experimental design, interpretation of results, and writing the manuscript.

Chapter 3: Proteomic-based discovery of putative biomarkers for improved classification of endometrioid and high grade serous ovarian cancer subtypes

Chapter 3 constitutes a manuscript in preparation. In order, the authors include Dieters-Castator DZ, Rambau P, Kelemen LE, Lajoie GA, Köbel M, and Postovit LM. DZDC and PR contributed equally to the study. DZDC performed, analyzed and interpreted proteomics experiments and wrote the manuscript. PR and MK performed IHC optimization, staining, tissue scoring and analysis. GAL, LMP and MK contributed to the experimental design, interpretation of results, and writing the manuscript.

Chapter 4: Proteomic profiling of ovarian cancer extracellular vesicles for biomarker discovery

Chapter 4 constitutes a manuscript in preparation. In order, the authors include Dieters-Castator DZ, Liu J, Lajoie GA, and Postovit LM. DZDC and JL contributed equally to the study. JL optimized and performed EV isolation from condition media, plasma and ascites. DZDC performed, analyzed and interpreted proteomics experiments and wrote the manuscript. JL, GAL and LMP contributed to the experimental design, interpretation of results, and writing the manuscript.

Chapter 5: Embryonic Protein NODAL broadly regulates the breast tumour microenvironment by reprogramming cancer-derived secretomes

Chapter 5 constitutes a manuscript in preparation. In order, the authors include Dieters-Castator DZ, Piaseczny M, Zhang G, Liu J, Kuljanin M, Sherman S, Jewer M, Katherine Q, Siegers G, Leask A, Hess DA, Lajoie GA, and Postovit LM. DZDC performed, analyzed and interpreted proteomics and in *vitro* MSC experiments. DZDC, MK, SS and GS performed flow cytometry experiments and analysis. MP and GZ performed HFF experiments. KQ and LA performed HDF experiments. GZ, JL and MJ generated transgenic cell lines. DZDC, GS, DAH, GAL and LMP contributed to the experimental design, interpretation of results, and writing the manuscript.

Acknowledgments

Thank you everyone who has taken the time to help me in one way or another during my studies at Western University. Without your support, this thesis would not be possible.

Thank you all past and current members of the Postovit lab. In particular, thank you Dr. Scott Findlay Michael Jewer for the awesome company and critical feedback. Thank you Dr. Daniela Quail and Meghan Taylor for showing me the ropes during my start in the lab as an undergraduate. Thank you Dr. Gabrielle Siegers for always finding time to help with flow cytometry. Thank you Dr. Krista Vincent for your help with bioinformatics analysis. Thank Guihua Zhang and Jiahui Liu for making everything run smoothly and treating me like a family member.

Thank you all past and current members of the Lajoie lab. In particular, thank you Dr. Chris Hughes for introducing me to field of proteomics and helping with sample preparation and data analysis. Thank you Amelia Nuhn and Kevin Kania for being great lab mates. Thank you Paula Pittock for having the patience and experience to troubleshoot the seemingly infinite number of instrument problems we encountered over the years. Without you, none of the proteomics would be possible. Thank you Dr. Miljan Kuljanin for being a close friend and lab mate. You're an outstanding role model and have tremendously helped me achieve my goals inside and outside the lab.

Thank you members of the Hess lab for all the support. Stephen Sherman, thank you for help with flow cytometry and mouse experiments. Dr. David Hess and Gillian Bell, you have been an outstanding resource for all things MSC related.

Thank you Dr. David Hess, Dr. Alison Allan and Dr. Paul Walton for being excellent committee members and taking the time provide feedback. Thank you Dr. Dean Betts, Dr. John Di Guglielmo, and Dr. Dwayne Jackson for providing access to your facilities.

Importantly, thank you Dr. Lynne Postovit and Dr. Gilles Lajoie for the support, patience and the opportunity to be a member of your research groups. Gilles, thank you for all the thoughtful but critical feedback and unparalleled expertise in chemistry and mass spectrometry. Lynne, your enthusiasm, optimism and brilliance has helped

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

α-SMA ....α-Smooth Muscle Actin
ABC .... Ammonium Bicarbonate
ActRIIB .... ACTIVIN type IIB receptor
ADAM .... A Disintegrin and Metalloproteinase
AJCC .... American Joint Committee on Cancer
ALDH.. Aldehyde Dehydrogenase
ALK4/7.. ACTIVIN Receptor-Like-Kinase
ANGPT ....Angiopoietin
AP..Anterior Posterior
APCS..Serum Amyloid P-component Precursor
AUC ....Area Under the Curve
AVE ....Anterior Visceral Endoderm
BM.. Bone Marrow
BMDC..Bone Marrow Derived Cell
BME.. Basement Membrane Extract
BRCA1/2 ....Breast Cancer susceptibility type 1 and/or 2
CA-125 ....Cancer Antigen 125
CAF .... Carcinoma Associated Fibroblast
CAM ....Chorioallantoic membrane
CCC..ovarian Clear Cell Carcinoma
CD9AP ....CD9 Affinity Purification
CFHR4 .... Complement Factor H Related protein 4
CIC .... Cortical Inclusion Cyst
CM..Conditioned Media
CPL ....Combinatorial Peptide Libraries
CRP.. C-Reactive Protein
CTAP...................................Cell Type specific labeling using Amino acid Precursors
CTGF.. Connective Tissue Growth Factor
CTS .... Cathepsin
DCIS..Ductal Carcinoma In Situ
DTT ....Dithiothreitol
DVE .... Dorsal Visceral Endoderm
EC .... Endometrioid ovarian Cancer
ELISA ....Enzyme Linked Immunosorbent Assay
EMT.. Epithelial-Mesenchymal Transition
EOC.. Epithelial Ovarian Cancer
EPC .... Endothelial Progenitor Cells
ER ....Estrogen Receptor
ESM.. Electronic Supplementary Material
EV.. Extracellular Vesicle
FA..Formic Acid
FAP .... Fibroblast Activation Protein
FASP ....Filter Aided Sample Preparation
FGF .... Fibroblast growth factor
FIGO.....................................International Federation of Gynecology and Obstetrics
fMaSC ....fetal MaSC

xvii

Chapter 1

Introduction

1.1 General introduction

Cancer is a complex and heterogeneous group of diseases characterized by dysregulated cell growth, metastatic dissemination and the disruption of healthy tissue and organ function. The lifetime risk of developing cancer in the US is 38.4% with an average 5-year survival of 66.9% [1]. In Canada, this risk is even greater with 49% of males and 45% of females expected to develop cancer in their lifetime [2]. Cancer initiation and development is linked to genomic alterations incurred by normal cells during replication and exposure to environmental stresses. Prerequisites or "hallmarks" underlying malignant transformation include perturbations in carefully regulated cellular processes involving replicative immortality, Epithelial-Mesenchymal Transition (EMT), and sustained angiogenesis [3]. While these intrinsic alterations are critical for cancer initiation, extrinsic factors are similarly essential for tumour progression by mediating communication between cancerous and non-transformed host cells within the tumour microenvironment (TME) [4]. Characterizing the TME may lead to improved diagnostics and therapeutics, however this task is not trivial given the complexity. High resolution mass spectrometry (MS)-based proteomics is a powerful technique which can achieve quantitative information on thousands of proteins and provide insight into biological systems. In this thesis, I employed MS-based proteomics to improve the differential diagnosis of endometrioid and high grade serous ovarian cancer subtypes, to discover biomarkers for detecting high grade serous ovarian cancer in bio-fluids, and to characterize the role of NODAL in multipotent stromal cell (MSC) recruitment in breast cancer.

1.2 Ovarian cancer

Collectively, ovarian cancer (OC) is a broad and diverse group of diseases with disparate origins traced to sex organs of the female reproductive system [5]. In the United States (US), OC is the 5th most lethal female cancer and most lethal

gynecological disease. Although the lifetime risk of developing OC is low (1.27%), the 5-year survival is relatively poor (46.5%) [1]. Endometrial cancer (cancer of the uterine corpus), for comparison, is the most common gynecological cancer in the US with a lifetime risk developing of 2.86% but 5-year overall survival of 81.1% [1]. Cervical cancer, however, comprises the greatest disease burden worldwide (incidence and total deaths) of all gynecological cancers [6]. This is largely attributed to the limited resources of low- to middle-income countries for prevention and early detection that is routinely practiced in health care systems of developed regions. Unfortunately, regardless of region, there is currently an unmet need for better detection and treatment of OC globally.

1.2.1   *Basic anatomy of the female reproductive system*

The female reproductive system is comprised of several internal and external sex organs which function in concert to regulate oocyte production, maturation and fertilization, and support embryonic (fetal) development [7]. The internal sex organs, located within the lower abdomen (intraperitoneal cavity), are comprised of two ovaries (left and right sides) connected to the uterus by separate fallopian tubes (Figure 1.1) [7]. The ovaries are surround by an epithelium and contain an inner mass made up of stromal cells and follicles [7]. Within each ovary, there are a finite number of follicles made up of a centrally located oocyte (egg) surrounded by cumulus cells, granulosa cells and an outer layer of theca cells [8]. Approximately once every month, until menopause, a mature follicle will rupture from the ovarian surface epithelium (OSE) and release (ovulate) an oocyte [7]. At the end of the fallopian tube, finger like projections (fimbriae) grasp the ovary and capture ovulated oocytes. Once captured, oocytes are transported to the uterus through the fallopian tube which is comprised of several cell types with specialized functions. The inner fallopian tube epithelium is lined with ciliated and secretory (glandular) cells which support egg transport and tubal fluid synthesis, respectively [5]. Stromal and contractile smooth muscle cells make up the inner region and outer layer of the fallopian tube, respectively. If an egg becomes fertilized (zygote)

during the journey through the fallopian tube over several days, it will implant into uterine wall (endometrium) and begin developing into a fetus [7]. The endometrium is a specialized organ comprised of epithelial cells, stromal cells and blood vessels and is part of the maternal-fetal interface (placenta) which is critical for exchange of nutrients and waste to and from developing fetus [9]. While these organs are integral for reproduction, perturbations within various cells types can lead to the development of abnormal lesions and ovarian malignancies [5,7].

a

b

1st → 2nd

Primordial follicle   OSE   3rd follicle

c

Endometrial gland

Endometrial cavity

Myometrium

Fallopian tube

Uterus

Ovary

Basal or progenitor cell	Ciliated cell	EME cell	EMS cell	Fallopian tube stromal cell	
Immune cell	Oocyte	Ovarian stromal cell	Secretory cell	Granulosa cell	Theca cell
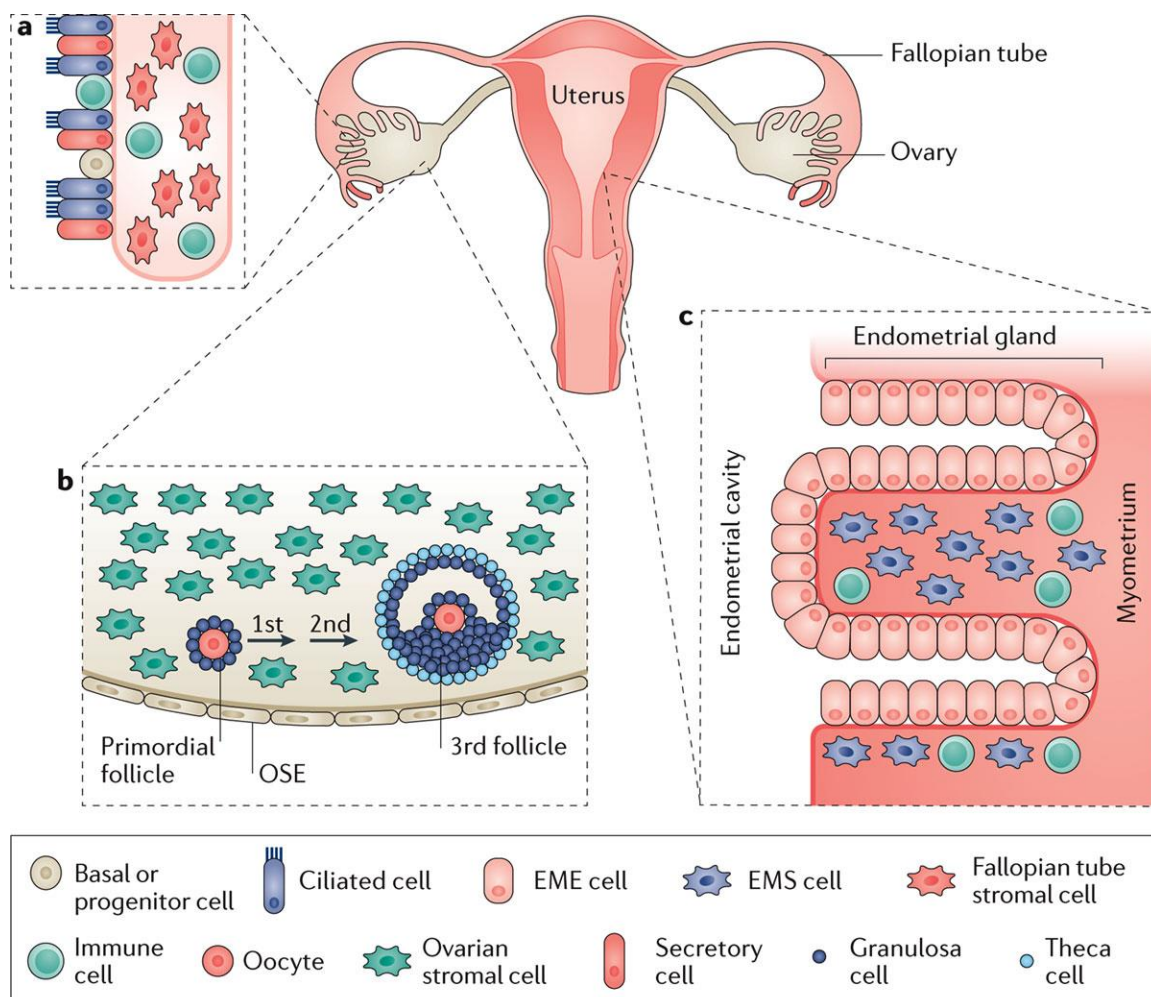
Figure 1.1 Anatomy of female reproductive system

The female sex organs are comprised of the ovaries (b), fallopian tubes (a) and uterus (c). Each organ contains specialized cell types which facilitate reproduction. (a and b) During ovulation, a follicle will rupture form the ovarian surface epithelium (OSE) and be captured by fimbrae connected to the fallopian tube. Ciliated and secretory epithelial cells lining the fallopian tube are important for egg transportation and fertilization. (c) If an egg becomes fertilized during this journey, the zygote will implant into the uterine wall or endometrium. EME endometrial epithelium, EMS endometrial stroma. Adapted by permission from RightsLink Permissions Springer Customer Service Centre GmbH: Springer Nature. Nature Reviews Cancer. The disparate origins of ovarian cancers: pathogenesis and prevention strategies, Anthony N. Karnezis, Kathleen R. Cho, C. Blake Gilks, Celeste Leigh Pearce, David G. Huntsman (2016) 1:65-74.

1.2.2 Staging

Epithelial ovarian cancer (EOC) is the most common type of ovarian cancer and accounts for approximately 90% of all diagnoses in developed countries [16]. During ovarian cancer progression, tumours may spread to neighboring organs through direct contact or passively disseminate as single cells or spheroids via fluid known as ascites within the intraperitoneal cavity [10]. The intraperitoneal wall (peritoneum) and omentum are the most common sites affected by metastatic ovarian cancer followed by the colon, liver, and kidneys. In rare cases, ovarian cancer metastasizes to the lungs, brain, skin, spleen and lymph nodes [11].

Staging is critical part of all cancer care and involves documenting the extent of cancer progression to better predict patient prognosis and guide treatment decisions [12]. For ovarian cancers, two systems are used for staging; the International Federation of Gynecology and Obstetrics (FIGO) and the American Joint Committee on Cancer (AJCC) Tumour Node Metastasis (TNM) [13]. The FIGO staging system is more widely adopted and recognized worldwide but shares a number of similarities with TNM staging.

Individual components monitored by the TNM scoring system include tumour (T) size and spread, the presence of cancer cells in the lymph nodes (N), and metastasis (M) to distant organs. A cancer patient's overall stage is comprised of the number and/or letter assigned to each TNM component. Higher TNM values/letters are associated with later stages and therefore greater disease burden. In patients with stage I ovarian cancer, tumours are confined to the ovary(ies) and/or fallopian tube(s). During stage II disease, tumour growth has spread or invaded into adjacent organs such as the bladder, uterus, colon and/or formed primary peritoneal cancer. At stage III, small cancer growths (<2cm) are present on the surface of organs beyond the pelvic region such as the spleen or liver and cancer cells may also be found in the lymph nodes. In patients with stage IV cancer, the last and most deadly stage, cancer cells can be detected in fluid

surrounding the lungs (pleural effusions) and/or tumours have formed in distant organs such as the liver, lungs, skin, brain, spleen, lymph nodes and in rare cases, bone [11]. Patients with early stage invasive EOC (I and II) have much better prognosis with 5-year overall survival rates ranging from 70-90% [14]. Unfortunately, most invasive EOCs are detected during later stages for which the 5-year overall survival drops significantly (39-59% for stage III and 17% for stage IV). While lifetime risk of developing ovarian cancer is low, the risk of dying, in absolute terms, is quite high at 0.93% [1].

1.2.3   Subtypes

In addition to staging, subtype classification based on morphological and molecular features is another important factor for diagnosing and treating cancer [15]. There are 5 major histotypes of invasive EOC and two histotypes of borderline disease (serous and mucinous) [17]. The 5 invasive EOC subtypes are high grade serous, low grade serous, endometrioid, clear cell and mucinous (Table 1.1) [18]. Up to 30% of all EOC cases are diagnosed as endometrioid, clear cell or mucinous. These cancers are generally benign or slow growing with intermediate to favourable prognosis and are also referred to as type I tumours [15]. Unfortunately, the majority of patients (~70%) are diagnosed with high grade serous ovarian cancer (HGSC) or type II tumours which comprise the most aggressive and deadliest form(s) of EOC. Notably, mutations in *TP53* are prevalent in 95% of HGSC given the central role in dictating cell cycle arrest and apoptosis [19–22]. Indeed, *TP53* is the most frequently mutated gene in cancer. Diagnosis and subtype identification of tumour samples is initially determined by a pathologist however, in ~10% of cases, EOC may be misclassified due to similarities in histology [23]. Of note, this is particularly problematic when differentiating (high-grade) endometrioid ovarian cancer (EC) from HGSC and low grade serous ovarian cancer (LGSC) from HSGC [23,24].

1.2.4   Pathogenesis

In terms of pathogenesis, histopathological and genetic profiling has recently identified distinct precursor lesions for type I and type II tumours [15]. Until recently, EOC was predominately believed to arise through invagination and neoplastic transformation of ovarian surface epithelial (OSE) cells following repeated ovulation. However, the incessant-ovulation hypothesis is less favoured with new evidence tracing the origins of EOC to the fallopian tube and endometrial epithelium [5]. Indeed, the majority of HGSC are associated with Serous Tubal Intraepithelial Carcinomas (STICs) primarily located within fimbriae (Figure 1.2) [25]. STICs are derived from fallopian tubal epithelial cells (FTECs) and importantly, frequently share identical *TP53* mutations with their HGSC counterparts [26,27]. Moreover, HGSC may arise from ectopic ciliated and/or secretory fallopian tube epithelium (endosalpingiosis) or cortical inclusion cysts (CICs) of the ovary. Endosalpingiosis and CICs are believed to arise due to invagination of the OSE or tubal-type epithelium following ovulation [28]. The rarer and less aggressive LGSC and serous borderline tumours may also originate from CICs. However, unlike HGSC, these tumours are characterized by intact and wild-type *TP53* expression and patchy *CDKN2A* staining [29]. EC and clear cell carcinoma (CCC), in contrast, are associated with atypical endometriosis (endometriomas) of the ovary [30–33]. Endometriomas, like endosalpingiosis, are ecotopic endometrial tissue derived from ciliated and/or secretory cells. Retrograde menstruation, and possibly endometrial remnants, are thought to be the primary source of endometriomas. Moreover, recent work by Cochrane *et al.* suggests EC and CCC may be derived from secretory and ciliated endometrial epithelial cells, respectively [33]. Interestingly, whereas STICs give rise to HGSC, fallopian tube cancers are extremely rare. Hence, genotoxicity of the fallopian tube is theorized to prevent ectopic implantation of embryos and coincidentally, tumour growth [5]. Conversely, the ovary has been suggested to contain supportive stromal cells and factors which can promote tumour growth [5]. Indeed, metastasis

to the ovary has been linked to breast, colon, stomach and appendiceal cancers [34]. How EOC preferentially co-opts the ovaries during progression is not understood and improved characterization of this close association may reveal targets for inhibiting tumour growth.

Table 1.1 Features of main epithelial ovarian cancer (EOC) subtypes

Subtype[a]	Incidence[a]	Tissue (site) or origin[b]	Cell of origin[b]	Precursor lesion[b]	Histotypes[c]	Mutations, pathways and defects[b,d]	Treatment[a]	Prognosis[a]
High-grade serous (HGSC)	68-71%	Fallopian tube fimbria OR Ovary cortical inclusion cyst	FTEC	STIC	WT1 positive AND TP53 overexpression OR WT1 positive AND TP53 negative/ diffuse p16	TP53, BRCA1/2, NF1, CDK12, HR defects, PI3K/Ras/ Notch/FoxM1	Surgery plus chemo, PARP inhibitors, Anti-angiogenic therapy	Poor
Low-grade serous (LGSC)	<5%	Endosalpingosis OR Papillary tube hyperplasia	Fallopian tube epithelial cell	Borderline serous tumour	WT1 positive AND TP53 wild type pattern/p16 patchy	BRAF, KRAS, NRAS, ERBB2	Surgery plus chemo	Intermediate
Endometrioid (EC)	9-11%	Endometrium	Endometrial epithelial cell	Atypical endometriomiosis	WT1 negative AND HNF1Bnegative/ PR positive	ARID1A, PI3KCA, PTEN, PPP2R1A, MMR defects	Surgery plus chemo	Favourable
Clear cell (CCC)	12-13%	Endometrium	Endometrial epithelial cell	Atypical endometriomiosis	WT1 negative AND PR negative/ HNF1B positive/ ARID1A negative	ARID1A, PI3KCA, PTEN, CTNNB1, PP2R1A	Surgery plus chemo, Anti-angiogenic therapy	Intermediate
Mucinous	3%	Unknown	Unknown	Mucinous borderline tumour, Brenner tumour, Teratoma or Endometriosis	WT1 negative AND PR negative	KRAS, ERBB2 amplification	Surgery plus chemo	Favourable

[a]Rojas *et al*. 2017. Molecular characterization of epithelial ovarian cancer: Implications for diagnosis and treatment.
[b]Karnezis *et al*. 2017. The disparate origins of ovarian cancers: pathogenesis and prevention strategies.
[c]Köbel *et al*. 2014. Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry.
[d]Jayson *et al*. 2014. Ovarian Cancer.
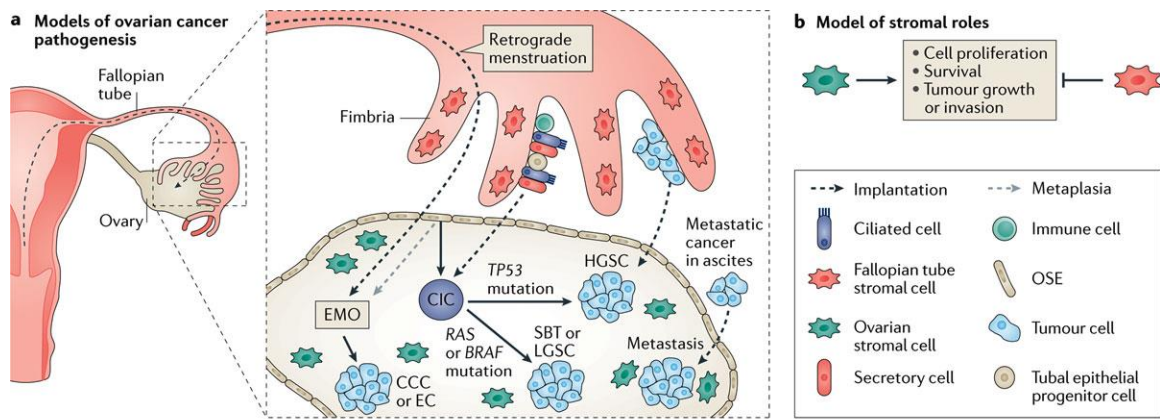
Figure 1.2 Anatomical overview of cell types and precursor lesions which give rise to EOC subtypes.

(a) Cells from different origins of the female reproductive system give rise to ovarian cancer. Precursor lesions and signalling pathways underlying progression of individual EOC subtypes are shown. (b) Proposed roles of ovarian and fallopian tube stromal cells on ovarian cancer progression. Adapted by permission from RightsLink Permissions Springer Customer Service Centre GmbH: Springer Nature. Nature Reviews Cancer. The disparate origins of ovarian cancers: pathogenesis and prevention strategies, Anthony N. Karnezis, Kathleen R. Cho, C. Blake Gilks, Celeste Leigh Pearce, David G. Huntsman (2016) 1:65-74.

1.2.5   Treatments

HGSC accounts for ~70% of all EOC cases and ~90% of advanced (stage III/IV) disease [35]. First line treatments for advanced ovarian cancer remain limited and require aggressive surgery combined with chemotherapy [36]. Surgery should be performed by a gynecological oncologist and generally involves tumour debulking, hysterectomy (removal of the uterus), bilateral salpingo-oophorectomy (removal of the ovaries and fallopian tubes), and omentectomy [37]. When tumour debulking is not feasible due to risk of patient morbidity associated with extensive surgery, neoadjuvant (preoperative) chemotherapy may be administered followed by interval debulking [37]. Taxanes (i.e. paclitaxel, docetaxol) and platinum based compounds (i.e. carboplatin, cisplatin) are standard chemotherapeutics administered following surgery. These drugs inhibit cellular proliferation by stabilizing microtubules (taxanes) and crosslinking DNA to inhibit repair and replication (platinum based compounds) [38,39]. While short term responses are often achieved with aggressive treatment, tumours often develop resistance [40].

Accurate molecular and histotype classification is becoming increasingly relevant for the management and treatment of EOC [15]. For example, germline mutations in breast cancer susceptibility type 1 and/or 2 (*BRCA1/2*) are currently the single best predictors of developing ovarian and breast cancer [41]. Indeed, *BRCA1* and *BRCA2* mutations increase the likelihood of developing ovarian cancer by 44% and 17% respectively. In breast cancer, the risk is even greater at 72% for *BRCA1* and 69% for *BRCA2* carriers [42]. Therefore, females bearing germline *BRCA1/2* mutations may undergo risk reducing prophylactic salpingo-oophorectomy and/or mastectomy (removal of the breasts) to prevent ovarian and breast cancer, respectively. Moreover, up to 33% of HGSC cases are estimated to exhibit alterations in *BRCA1/2* due to germline and somatic mutations or epigenetic silencing [37]. Notably, both BRCA1/2 and Poly (ADP-ribose) polymerase (PARP) PARP carry out homologous recombination mediated DNA repair and maintain

genome integrity [43,44]. Together, these findings have led to the clinical use of PARP inhibitors in BRCA1/2 defective cancers like HGSC by exploiting the concept of synthetic lethality [45]. In essence, the combination of BRCA1/2 defects and PARP inhibition impairs cancer cells ability to repair DNA damage and results in cell death. In contrast, healthy cells without DNA damage are largely spared. Indeed, clinical trials investigating PARP inhibitors in HGSC have demonstrated impressive increases in progression free survival (PFS) with few significant side effects [46]. However, significant increases in PFS with PARP inhibitors such as Olaparib did not translate into overall survival (OS) but combination therapies, dosing and stratification of patient subgroups are still be explored to realize their full potential.

Alternatively, the PI3K/AKT/mTOR pathway is frequently activated in EC and CCC and may be potential target when cancers do not respond to chemotherapy [44]. Furthermore, estrogen receptor (ER) and/or progesterone receptor (PR) expression has been reported in EOC, in particular EC and HGSC. High ER and PR expression was highest in EC and associated with increased disease free survival [47]. Unfortunately, clinical studies targeting PR in EC are lacking, in part due to side effects and poor specificity of progesterone receptor modulators [48]. Given the limited treatment options for EOC, additional studies focused on hormonal therapy may improve patient outcomes.

1.2.6 *Detection and diagnosis*

According to the National Institutes of Health Biomarkers Definitions Working Group, a biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [49]. In terms of biomarkers used for detecting or diagnosing diseases like cancer, sensitivity and specificity are standard metrics used to assess their performance. Sensitivity corresponds to the true positive rate or percentage of individuals within the

diseased population which test positive for a given biomarker. Alternatively, the specificity of biomarker indicates the percentage individuals which correctly test negative for the disease in question.

In the context of ovarian cancer, early stage disease is often asymptomatic and therefore difficult to detect. Moreover, symptoms associated with disease progression are generally non-specific and primarily include abdominal pain and discomfort due to accumulation of ascites within the intraperitoneal cavity and/or malignant transformation of the fat pad (omentum) covering the bowel and abdominal cavity [10].

Currently, there are no FDA approved biomarkers for ovarian cancer screening; however cancer antigen 125 (CA-125) levels in the blood are widely used to aid in diagnosis, monitor disease progression and treatment response [50]. CA-125 is derived from the transmembrane glycoprotein Mucin 16 (MUC16), and is typically elevated in ~83% of EOC cases [50]. Unfortunately, several issues primarily related to sensitivity and specificity make CA-125 unsuitable as a single agent for early screening [50]. Firstly, CA-125 is only elevated in ~67% of patients with stage I disease when using a less stringent cut-off of 30 U/mL [51]. Secondly, non-malignant/benign gynecological conditions can increase CA-125 levels necessitating the need for surgical follow-up [51]. Lastly, CA-125 levels fluctuate between post-menopausal women depending on race/ethnicity, age, hysterectomy, smoking history and obesity and thus confound the use of a set cut-off to establish disease status [52]. For example, CA-125 significantly increases with age and is higher in Caucasians compared to African Americans and other minorities.

In addition to CA-125, highly sensitive imaging methods like transvaginal ultrasound (TVUS) can detect abnormal ovarian masses/features. However, invasive follow-up surgeries/biopsies associated with significant patient discomfort remain essential for accurate diagnosis by a pathologist. Moreover, the majority of

these surgeries do not reveal malignant disease with the ratio of surgeries to the number of cancers detected by TVUS being ~19:1 [53]. For reference, two large scale trials, the Prostate, Lung, Colorectal, Ovarian cancer screening trial (PLCO) in the US and UK Collaborative Trial of Ovarian Cancer Screening trial (UKCTOCS) evaluated the role CA-125 and TVUS screening on reducing ovarian cancer deaths [54,55]. Both trials monitored tens of thousands of females over a number of years. Unfortunately, results from these trials did not achieve a significant reduction in overall mortality based on their initial analyses. In the PLCO trial, a substantial number of false-positive screening results were reported in the intervention group (CA-125 plus TVUS) that were later associated with follow-up surgery related complications [55]. Surprisingly, the authors did not observe a stage shift in detecting early ovarian cancer in the intervention group compared to usual care. The UKCTOCS trial, on the other hand, noted a significantly higher proportion of cases with low volume disease (stage I, II, IIIa) in the multimodal screening group (CA-125 plus TVUS) compared to no screening (40% versus 26%) [54]. Initially, the authors reported a small but insignificant mortality reduction in the screening groups (CA-125 plus TVUS or TVUS only) compared to standard care (no screening) [54]. However, an overall mortality reduction of 20% (between years 0 to 14 after screening) in favour of multimodal screening (CA-125 plus TVUS) was reached when subgroup analysis was performed excluding prevalent cases (females likely harbouring ovarian cancer prior to enrolment). Together, these studies highlight the limitations associated with CA-125 and TVUS screening modalities and suggest that longer follow-up studies may be needed to achieve significant reductions in mortality.

Consequently, the lack of clinically available screening strategies remains a major bottleneck for improving survival of ovarian cancer patients. Given the absence or limited number of biomarkers for EOC screening and diagnosis, a substantial amount of research effort is focused on identifying/developing new or improved markers. For example, human epididymis protein 4 (HE4) is another clinically

approved marker with sensitivity and specificity similar to CA-125 but has not been widely implemented [56]. Moreover, Cramer *et al.* surveyed 49 markers in a large cohort of malignant and non-malignant plasma samples from the PLCO trial [57]. When using a sensitivity cut-off of 95%, CA-125 followed by HE4 were the best predictors of disease status with specificities of 0.73 and 0.54, respectively. In addition to single biomarkers, several assays have been developed which monitor a panel of plasma proteins and achieve greater performance than CA-125 or HE4 alone. These products include OvaSure, OVA1 and OvPlex and can aid in the diagnosis of EOC such that patients may not have to undergo surgeries for benign conditions [56]. Very recently, a study published in the journal Science reported a multi-analyte blood test capable of achieving very high detection rates for 8 different cancers [58]. This test, called CancerSEEK, monitors ~60 amplicons in cell-free DNA and several proteins in the blood. Impressively, CancerSEEK achieved sensitivities ranging from 69-98% (98% for ovarian cancer) at a specificity >99% and could also predict the anatomical location of cancers. Studies like this highlight the significant challenges associated with low signal to noise in detecting tumour-derived products that are present in minute quantities. However, sensitive assays which monitor multiple biomarkers may hold the future of screening tools.

1.3   Breast cancer

Breast cancer is second leading cause of cancer related deaths in females with a lifetime risk of developing breast cancer (invasive and ductal carcinoma in situ) of 14.73% (12.41% for invasive disease only) [1]. Although the incidence of breast cancer is high in the US, 89.5% of patients are alive after 5-years. In absolute terms, the lifetime risk of dying from breast cancer for females in the US is 2.62%. Screening strategies have significantly improved early detection and treatment, however metastatic disease remains difficult to cure. In terms of risk factors, both breast and ovarian cancers have several in common even though they originate in

anatomically different locations. For example, early age of first menarche (menstruation), non-parous (non-child bearing) females, and delayed menopause are all associated with increased risk of developing breast and ovarian cancer. Conversely, early age of first childbirth and breastfeeding are protective against breast cancer and long term oral contraceptive use are protective against developing EOC [59–61]. Intriguingly, the most aggressive forms of breast and ovarian cancer share similar genetic profiles [19].

1.3.1 Basic anatomy of the breast

The breasts contain a branching of network glandular structures (mammary glands) within a vascularized layer of fatty tissue that is externally supported by connective tissue (Figure 1.3) [62]. The primary function of the breasts (mammary glands) are milk production (lactation) to support newborns. These organs are unique to mammals and undergo extensive morphogenesis/remodelling throughout embryonic development, puberty, and pregnancy [63]. Within each breast, the mammary gland forms a hollow tree like network of ~15-20 lobes or terminal end buds (TEBs) which radiate outward from the nipple. These tubular structures (mammary ducts) are made up of a bilayer of inner luminal and outer basal epithelial cells that arise through proliferation and elongation into the fat pad throughout puberty [63,64]. During pregnancy, the TEB luminal epithelium proliferates and differentiates into secretory (milk producing) cells comprised of 10-100 alveoli. While breastfeeding, basal myoepithelial cells surrounding alveoli contract to expel milk into the ducts. After breastfeeding is no longer required, alveoli undergo massive apoptosis to remove up to 80% of the epithelium [64]. Strikingly, this dynamic cycle of proliferation, differentiation and apoptosis is tightly regulated by a number of extracellular factors and can be repeated many times over several decades [64].

Breast anatomy and histology · Eric Wong

Clin Obstet Gynecol. 2011 Mar;54(1):91-5.

The breast is composed of glandular and stromal tissue. Glandular tissue includes the ducts and lobules. **Stroma** comprises area between lobes.

Breast

Pectoralis muscle

Chest wall and ribs

Nipple

Ducts

Stroma

Lumen

Each lobe drains into a major lactiferous duct that dilates into a lactiferous sinus beneath the areola opening onto the nipple

Each **terminal duct lobular unit** is composed of branched tubuloalveolar glands organized into lobes

Inner layer of **myoepithelial cells** provide structural support to the lobules and assist in milk ejection during lactation

Outer layer of lobular **luminal epithelial cells** produce milk during lactation

Figure 1.3 Anatomy of the breast.

The breast is comprised a branching network of ducts which terminate into lobular units surrounded by stromal and adipose tissue. Each lobular unit (mammary gland) contains an inner and outer layer of luminal epithelial and myoepithelial cells, respectively. Adapted from the McMaster Pathophysiology Review (MPR) website: http://www.pathophys.org/breast-cancer/.

1.3.2   Staging

Cancer staging is a critical component of patient management. Similar to ovarian cancer, breast cancer staging is also based upon the AJCC TNM system with TNM components assigned a value/letter corresponding to disease progression at diagnosis [60]. For example, T1 corresponds to the presence of a tumour <2 cm across while a score of T4 includes tumours of any size that have grown into the chest wall or skin. In addition, the AJCC TNM staging system was recently updated to include tumour grade and estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor 2 (HER2) expression status [65]. Tumour grade is based on how normal and or undifferentiated the cancer cells appear with less differentiated cells often associated with advanced disease.

1.3.3   Subtypes

Breast cancer is primarily comprised of 4 intrinsic subtypes: luminal A, luminal B, HER2-enriched and triple negative breast cancer (TNBC) (Table 1.2) [66]. Luminal breast cancers express ER and PR and accordingly, are classified as hormone receptor (HR) positive. Of note, luminal A breast cancers are HER2- while luminal B can be either HER2- or HER2+. For reference, the Surveillance, Epidemiology, and End Results (SEER) breast cancer registry in the US population revealed 83% of all breast cancers diagnosed are HR+ [67]. When accounting for HER2 expression, 72.7% and 10.3% of HR+ breast cancers were HER2- and HER2+, respectively. A similar report by the North American Association of Central Cancer Registries found 72.1% and 10.3% of all HR+ breast cancer to be HER2- and HER2+, respectively [68]. From these findings, the luminal A subtype is the most predominate and fortunately, exhibits the highest OS of all invasive breast cancers. Luminal B cancers, in contrast, are less frequent, characterized by high levels of proliferating Ki67+ cells and have a worse prognosis [69–71]. Compared to luminal cancers, HER2-enriched tumours are rarer and account for ~4.6% of cases [67].

This subtype typically lacks ER and PR, is highly aggressive and associated with low OS. The last main subtype, TNBC, lacks ER, PR, and HER2 and comprises the most heterogeneous form of breast cancer. TNBC accounts for ~12.2% of all cases and based on molecular profiling, can be divided into two subgroups, basal A and basal B (or claudin-low) [67,72]. Like HGSC, TP53 mutations are found in up to 84% of all TNBCs [73]. Of note, inflammatory breast cancer is another form that is extremely rare but highly proliferative and aggressive [74]. It is characterized by swelling and redness of breasts and associated with a higher rate of recurrence and worse OS compared to non-inflammatory locally advanced breast cancer (Stages IIB, IIIB,IIIA) [75].

Table 1.2 Features of main invasive breast cancer subtypes

Subtype	Incidence[a,b]	Tissue (site) or origin	Cell of origin	Precursor lesion	Histotypes[b]	Mutations, pathways and defects[c]	Treatment[d]	Prognosis[d]
Luminal A	73%	Mammary duct	Luminal epithelium	DCIS	HR+/HER2- Ki67 low	PIK3CA, PTEN	Hormone therapy (Tamoxifen)	Good
Luminal B	10%	Mammary duct	Luminal epithelium	DCIS	HR+/HER2-/+ Ki67 high	PIK3CA, PTEN	Radiation plus Chemotherapy AND/OR surgery	Intermediate
HER2-enriched	5%	Mammary duct	Luminal epithelium	DCIS	HR-/HER2+	PIK3CA, PTEN	HER2 mAb (Trastuzumab)	Poor
Triple negative breast cancer* (basal-like and claudin low)	12%	Mammary duct	Luminal epithelium	DCIS	HR-/HER2-	TP53, CCNE1, RB1, MYC, AKT	Surgery plus chemo, PARP inhibitors	Poor
Inflammatory	2.5%	Mammary duct	Luminal epithelium	DCIS	Mixed?	TP53, PIK3CA, ERBB2	Surgery plus chemo	Poor

[a]Howlader *et al.* 2014. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status.
[b]Anderson *et al.* 2014. How many etiological subtypes of breast cancer: Two, three, four, or more?
[c]Koboldt *et al.* 2012. Comprehensive molecular portraits of human breast tumours.
[d]Shah *et al.* 2014. Pathogenesis, prevention, diagnosis and treatment of breast cancer.

1.3.4 Pathogenesis

Ductal Carcinoma *In Situ* (DCIS) accounts for about 1 in 5 of all breast cancers diagnosed and is the precursor to invasive breast cancer (IBC) [82]. DCIS is characterized by confined hyperplasia of ductal epithelial cells and when detected, 5-year survival rates are nearly 100%. Alternatively, Lobular Carcinoma *In Situ* (LCIS) is a more rare form of breast cancer that involves neoplasia of the secretory cells lining the mammary glands. DCIS which proceeds to invade through the myoepithelium and breach the basement membrane is termed Invasive Ductal Carcinoma (IDC). IDC accounts for the majority of breast cancer cases and is typically detected when localized to the breast and axillary lymph nodes, chest wall and skin of the breast [2]. Metastatic breast cancer (mBC) is the rarest and most aggressive form which occurs when cancer cells travel through the bloodstream or lymphatic system and form secondary tumours at distant sites. Bone, liver, lungs and brain are the main organs of breast cancer metastasis [82]. Patients with metastasis to the bone have better prognosis compared to those with lung and liver and brain metastases are the least frequent but most lethal [76]. In terms of origin, gene expression patterns for luminal and TNBC subtypes more closely resemble those of ductal and myoepithelial cells, respectively, however EpCAM+ luminal epithelial cells are main progenitors from which most breast carcinomas arise [77–79].

1.3.5 Detection and diagnostics

In contrast to ovarian cancer, breast cancer screening has had a significant impact on patient mortality and is performed via routine mammograms in females aged 50-55 or older [80]. Even more so, the advent of 3D mammography (breast tomosynthesis) has further increased detection rates while reducing the false positives associated with 2D mammography [80]. Self-examination of the breast for masses, irritation and discoloration can also aid in early detection. Taken

together, breast cancer screening has significantly reduced breast cancer mortality and rates of stage II disease or higher [81]. A number of gene-expression assays are also clinically available to aid predicting recurrence and response to therapy for breast cancer. For example, OncotypeDx™ – a genomic test which measures the expression of a panel of 21 genes – prognosticates the likelihood of recurrence and predicts the likelihood of benefit from chemotherapy in patients with node-negative, early-stage, ER positive breast cancer treated with surgery and/or hormonal therapy [66].

1.3.6   *Treatments*

In cases of DCIS, females may opt for breast conserving surgery or mastectomy depending on the tumour size and spread [82]. Lymph node removal may be required depending on sentinel lymph node biopsy results and radiation and/or adjuvant hormone therapy may be administered. Although luminal breast cancers are HR positive, hormone therapy including selective estrogen receptor modulators (SERMs, i.e. tamoxifen) and aromatase inhibitors are largely ineffective in Luminal B cancers for which chemotherapy may be appropriate. Surgery plus chemotherapy is standard treatment for TNBC, inflammatory breast cancer and HER2-enriched tumours. However, the monoclonal antibody (trastuzumAb/Herceptin™) which targets the HER2 receptor is therapeutically effective in HER2-enriched subtypes when combined with chemotherapy but relapse rates are high [83]. TNBC lack HR and HER2 expression and consequently are the most difficult to target. Coincidentally, TNBCs are frequently BRCA1/2 defective and therefore susceptible to PARP inhibitors like HGSC. Indeed, ongoing and recently completed trials exploring the potential of PARP inhibitors as a monotherapy or combination therapy have shown positive increases in PFS [84].

1.4 Breast cancer plasticity

Cellular plasticity exists throughout embryonic development and refers to the ability of cells to adapt to and modulate their surrounding microenvironment [85]. Cellular plasticity is also characterized by the capacity for stem cells to undergo self-renewal and multi-lineage differentiation and remains important for tissue repair/homeostasis. However, during development and in adult tissues, cellular plasticity becomes increasingly restricted [86]. For example, during mammary gland development, mammary stem cells (MaSCs) within differentiate into luminal and basal epithelial cells which subsequently undergo rapid expansion and apoptosis throughout and following pregnancy, respectively [63]. Although cellular plasticity is tightly regulated, during (breast) cancer progression, terminally differentiated cells can acquire plastic, stem-like phenotypes through genetic, epigenetic and microenvironmental mechanisms which in turn confer selective advantages to intrinsic and extrinsic stresses such as TP53 inactivation and low oxygen (hypoxia) [86]. For instance, in transgenic mouse models, oncogenic PIK3CA^{H1047R} expression in luminal cells permitted multi-lineage luminal-to-basal differentiation and recapitulation of all main breast cancer subtypes (luminal, HER2 and basal-like) in mouse xenografts [87]. In basal cells, although oncogenic PIK3CA^{H1047R} expression promoted basal-to-luminal differentiation, oncogenic PIK3CA^{H1047R} expression combined with *TP53* deletion was required to form breast tumours with characteristics of the most aggressive subtypes (basal-like TNBC).

Currently, it is not clear whether stem-like breast cancer cells arise from tissue resident progenitors or differentiated ductal cells [85,88]. Regardless, stem-like breast cancer cells have been implicated in cancer initiation, metastasis and secondary tumour formation and are often demarcated based on CD44+/CD24- and/or ALDH1 expression [85,89]. Remarkably, stem-like breast cancer cells constitute a small fraction of the overall tumour mass but when purified/sorted, efficiently recapitulate entire tumours in limiting dilution assays with as little as

1000 cells [90]. Moreover, CD44+/ALDH^{hi} breast cancer cells were also found to be more resistant to chemotherapy and radiation [91]. However, the association between CD44+/CD24- breast cancers, patient survival and prognosis is not entirely clear [92,93]. For instance, a meta-analysis of 16 studies found CD24 expression to be a prognostic indicator of worse prognosis and no significant association was found for CD44+ and CD44+/CD24- breast cancers [93]. In this meta-analysis, subtype specific expression of CD44/CD24 was not available which may influence prognostic value. Other properties of stem-like cancer cells include anchorage independent growth and expression of embryonic signalling pathways [94]. For example, human embryonic stem cell transcription factors OCT-4 (POU5F1), NANOG, SOX2, MYC and their target genes, are enriched in poorly differentiated cancers, in particular, ER-/basal-like breast tumours [95]. However, embryonic signatures were not associated with CD44+/CD24- expression in this study, suggesting that different markers may be used to identify populations of stem-like breast cancer cells. In another study by Spike *et al.*, HER2-enriched and basal breast cancer subtypes were found to exhibit similar gene expression patterns characteristic of mouse fetal MaSC (fMaSC) [96]. Ultimately, a number of embryonic signalling pathways have been implicated promoting plasticity in breast cancer including WNT, NOTCH and recently, NODAL [95,97–99].

1.5   The embryonic morphogen NODAL

NODAL is an embryonic morphogen which has been shown to regulate plasticity in a variety embryonic and tumourigenic settings [100]. NODAL belongs to the Transforming Growth Factor-beta (TGF-β) superfamily and is widely regarded for its roles in patterning left-right asymmetry during embryonic development, meso-endoderm induction and maintaining human embryonic stem cell (hESC) pluripotency [101]. NODAL is synthesized as a homodimeric pro-protein which requires cleavage by PACE4 and/or FURIN to generate mature NODAL (Figure 4) [99]. Mature NODAL ligands signal through a receptor complex involving the type

I ACTIVIN receptor-like-kinase (ALK4/7), the ACTIVIN type IIB receptor (ActRIIB) and the co-receptor CRIPTO. Phosphorylation of downstream SMAD2/3 signalling proteins by the ALK4/7-ActRIIB receptor complex leads to the formation and translocation of SMAD2/3 dimers with SMAD4 into the nucleus. SMAD2/3-SMAD4 complexes associate with several transcription factors such as FOXH1 to regulate gene expression for a number of targets including NODAL and the endogenous inhibitor Lefty1/2. NODAL has also been shown to induce ERK1/2 activation in non-canonical manner in breast cancer cell lines [102].

During early murine embryonic development, NODAL expression is proximally concentrated and regulated at the extraembryonic ectoderm-epiblast interface [86]. Diffusion of NODAL towards the distal region induces expression of its endogenous inhibitor Lefty at dorsal visceral endoderm (DVE) to generate a proximal distal (PD) gradient. The anterior posterior (AP) axis is formed when the DVE rotates anteriorly to become the anterior visceral endoderm (AVE). NODAL expression subsequently localizes posteriorly during formation of the germ layers and primitive streak (early gastrulation). NODAL expression eventually reaches the Node located distally where its activity becomes restricted to the left hemisphere of the developing embryo due to inhibition by Lefty in the right hemisphere. In hESCs, the addition of recombinant ACTIVIN and/or NODAL with FGF can maintain expression of pluripotency markers (POU5F1, SSEA3, SSEA4, Tra-1-60[PODXL]) over prolonged culture in feeder free conditions [103]. ACTIVIN/NODAL signalling drives *NANOG* transcription in hESCs via SMAD2/3 binding within the *NANOG* promoter [104]. Nanog expression in turn prevents neuroectoderm differentiation by inhibiting FGF signalling, which is also required for pluripotency, and blocking transcription of ACTIVIN/NODAL in a negative feedback loop [104]. Conversely, targeting ALK4/5/7 with the small molecule inhibitor SB431542 accelerates hESC neuroectoderm differentiation [104]. Thus, NODAL contributes to a complex network of signalling pathways which can regulate pluripotency and differentiation.
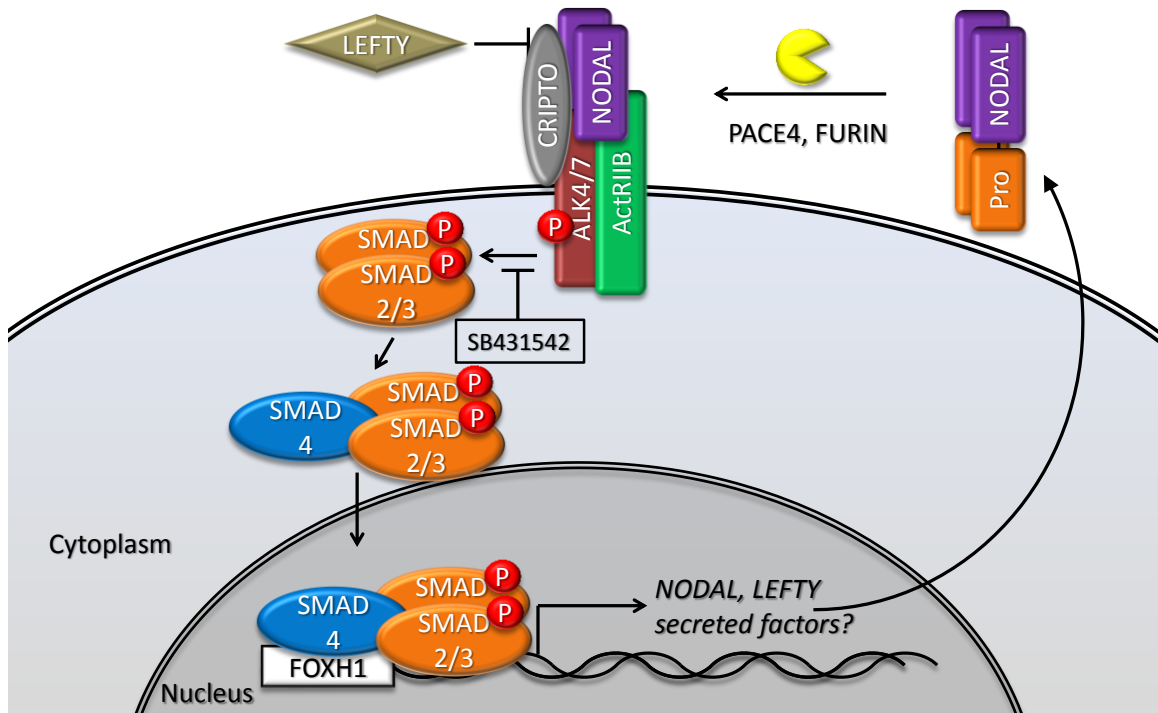
Figure 1.4 NODAL signalling pathway

NODAL is a secreted pro-protein that is cleaved by the convertases PACE4 and Furin to generate mature NODAL ligands. Mature NODAL homodimers induce phosphoryation of SMAD2/3 proteins via binding to the receptor complex invloving the type I ACTIVIN Like Kinase (ALK4/7), ACTIVIN type IIB receptor (ActRIIB) and co-repector CRIPTO. Phosphorylated SMAD2/3 dimers associate with SMAD4 which together, translocate into the nucleus to regulate target gene expression by associating with transcription factors including FOXH1. NODAL signalling can promote expression of itself as well as the endogenous inhibitor *LEFTY*.

Intriguingly, NODAL is silenced in most adult tissues but aberrantly expressed and promotes malignant phenotypes in a variety of cancers including melanoma, breast, prostate, pancreatic and glioma [86]. Fascinatingly, early work by Topczewska *et al.* found aggressive C8161 melanoma cells transplanted into zebrafish embryos could induce ectopic outgrowths and body axis duplication which is a unique phenotype of NODAL signalling [100]. NODAL expression was further found to be positively correlated with metastatic melanoma in patient samples [100]. In a subsequent study, hESC conditioned Matrigel™ was shown to suppress NODAL signalling in aggressive melanoma cells via hESC derived Lefty. Importantly, inhibition of NODAL signalling could differentiate aggressive melanoma cells into a pigment producing melanocytic phenotype [105]. In breast cancer models, NODAL was recently found to positively regulate stem cell markers and promote stem cell phenotypes in claudin-low TNBC (MDA-MB-231) and luminal breast cancer (MCF-7) cell lines [98]. In addition to NODAL, its co-receptor Cripto-1 and family members TGF-β and ACTIVIN are frequently dysregulated in cancer [106–108]. Together, these studies and others highlight the dominant role NODAL signalling plays in promoting embryonic and plastic phenotypes in cancer.

Recent work by our lab and others has also demonstrated that NODAL is a potent regulator of tumour growth, vascularization and invasion in breast cancer [102,109,110]. Of note, inhibition of NODAL signalling in the MDA-MB-231 significantly reduces tumour growth *in vivo* and vascularization of angioreactors. Although Vascular Endothelial Growth Factor (VEGFA) and Platelet Derived Growth Factor (PDGFA) expression are positively associated with NODAL signalling, how NODAL regulates breast cancer vascularization and growth is not fully clear. For instance, the addition of exogenous recombinant human NODAL only partially rescues vascularization deficiencies following NODAL knockdown in breast cancer cells. Furthermore, VEGF is not able to rescue tumour growth of MDA-MB-231 breast cancer cells lacking NODAL in a chick CAM assay. Together, these findings suggested NODAL may regulate tumour vascularization through

VEGF-dependent and independent mechanisms by altering the recruitment of accessory cells including bone marrow derived cells. Moreover, aggressive and undifferentiated cancers which are difficult to treat by current means may be sensitive to NODAL-targeted therapies.

1.6 The tumour microenvironment – a reservoir of biomarkers and therapeutic targets

Current research efforts aimed at tackling cancer are not limited to targeting tumour cells alone but the tumour microenvironment (TME) as a whole [3,111]. The TME is integral to all areas of cancer progression and is comprised of all cellular, biochemical and biophysical components in which cancer develops [4]. It harbours a wealth of information pertaining to tumour biology and is a rich source of biomarkers and potential drug targets [3]. Moreover, modulating interactions between auxiliary cell types within the TME is a complementary approach to reduce tumour burden in the case of anti-angiogenic and immune therapies [4]. Therefore, improved characterization and modeling of the TME can yield better diagnostics and treatments.

1.7 Non-transformed cells of the TME

In many cancers, non-transformed cells form a substantial proportion of the overall tumour mass. These cells exhibit distinct yet overlapping behaviours and can be broadly classified as endothelial, stromal (fibroblasts) and immune.

1.7.1 Endothelial cells

Endothelial cells form the main building block of all vessels in the body which are essential for the transportation and exchange of oxygen, nutrients and cellular (by-)products needed for cell survival [112]. New blood vessel formation (neovascularization) occurs during embryonic development and wound repair but

also in diseases like cancer. In adult tissues, neovascularization primarily occurs through a process known as angiogenesis which entails endothelial cell sprouting and branching from pre-existing vasculature [113,114]. Importantly, tumours must also become vascularized in order to grow beyond 1-2mm (the diffusion limit of oxygen) [115]. This was first recognized by the late Judah Folkman who proposed targeting tumour vascularization as a means to starve growth [112,116]. Given that tumours utilize angiogenesis heavily, neutralizing the highly potent angiogenic factor VEGFA significantly impaired rhabdomyosarcoma, glioblastoma multiforme (GBM) or leiomyosarcoma tumour growth *in vivo* [117]. This study and work by other groups ultimately led to the development and clinical approval of the VEGFA monoclonal Antibody (mAb) Bevacizumab (Avastin™) [118]. While promising results were achieved *in vivo*, Bevacizumab in combination with chemotherapy only modestly improved PFS and OS in renal cell carcinoma (RCC), clear cell carcinoma (CCC), metastatic breast cancer (mBC), glioblastoma multiforme (GBM), and lung cancer [119]. Unfortunately, most cancers develop resistance to anti-angiogenic therapies or exhibit limited responses, in part due to the emergence of VEGF-independent vascularization [112,120]. Indeed, tumour vasculature is often described as tortuous and comprised of leaky, immature vessels, indicating processes other than angiogenesis. Moreover, anti-angiogenic therapies may even reduce the efficacy of chemotherapy and inadvertently promote metastasis or selection of aggressive cancer cells [121].

Vasculogenesis or *de novo* blood vessel formation in the absence of pre-existing vasculature precedes angiogenesis during early embryonic development. Hemangioblasts are multipotent precursor cells within the yolk sac which give rise to haematopoietic and endothelial lineages and form primitive vessel like networks termed blood islands [122]. Although vasculogenesis was initially thought to be restricted to embryonic development, Asahara *et al.* identified a small population of Bone Marrow Derived Cells (BMDCs) referred to as Endothelial Progenitor Cells (EPCs) that could migrate to sites of ischemia and wounds to support postnatal

neovascularization [123]. Green fluorescent protein (GFP) or β-galactosidase (LacZ+) expressing EPCs from mouse bone marrow also augmented tumour angiogenesis by incorporating into a small but significant percentage vessels [124]. For reference, VEGF, Id1/3 and Tie-2 receptors have been shown to mediate the vasculogenic effects of EPCs [124–126]. For example, inhibiting VEGF receptor 1/2 (VEGFR1/2) or deleting VEGFR2 in EPCs dramatically reduces tumour frequency, growth rate and size *in vivo* [127]. Alternatively, tumour cells can generate channel-like networks through a process referred to as "vasculogenic mimicry" [128]. In light of this, multi-pronged approaches may be necessary to overcome resistance associated with anti-angiogenic therapy.

1.7.2   *Stromal cells*

Stromal cells, primarily Carcinoma Associated Fibroblasts (CAFs), often comprise the largest mass of non-transformed cells within the TME [129,130]. CAFs are mostly derived from fibroblasts within connective tissue which have become activated/transformed following exposure to signals from the TME [131]. Normally, during wound repair and tissue/ECM remodeling, fibroblasts become temporarily activated and express myofibroblast markers including α-smooth muscle actin (α-SMA) and fibroblast activation protein (FAP). In cancer however, activated fibroblasts persist as CAFs to promote tumourigenic behaviours including growth, vascularization, metastasis and resistance to therapy. Indeed, tumours often resemble a chronic state of wound healing and hence are referred to as "Wounds that do not heal" [132,133]. Since early observations of desmoplastic stroma (fibrous connective tissue) in cancer, a substantial amount of work has focused on uncovering the molecular mechanisms associated with CAF phenotypes. TGF-β, connective tissue growth factor (CTGF), stromal derived factor-1 (SDF-1 also known as CXCL12), collagens and pro-inflammatory cytokines are just some of the proteins expressed by CAFs [134]. Targeting these signalling pathways is one method to prevent stromal cell recruitment and cancer progression

Multipotent Stromal Cells (MSC) are highly plastic stromal cells also found within the TME [135,136]. MSC were first discovered within the bone marrow but also reside in many other vascularized niches including adipose tissue and often promote tumour progression *in vivo* [137]. *In vitro*, MSC form cartilage, bone, fat and muscle, and express adult stem cell-associated markers (CD133) and can form vessel-like networks [138]. Mounting evidence suggests MSCs are important for tumour neovascularization although their contributions may be restricted to perivascular roles and paracrine roles [139,140]. Moreover, MSCs can also differentiate into CAFs or myofibroblasts to remodel the extracellular environment by secreting ECM components and proteases linked to cancer aggressiveness [141]. Furthermore, up to 20-25% of CAFs have been found to originate from BM-MSC [142,143]. A number of chemokine/cytokine signalling networks have been implicated in MSC recruitment in breast, prostate and gastric cancer mouse models which involve CXCL12/CXCR4, migration inhibitory factor (MIF)/CXCR4, CXCL16/CXCR6 [144–146]. Of note, the contributions and effects of MSC in the TME appear to vary depending on their origin and cancer model [137]. In one study, adipose MSCs were shown to preferentially localize to sites of vascularization while bone marrow MSCs tended to remain at the periphery of tumours [147].

1.7.3   Immune cells

Although not investigated in this this, immune cells or tumour infiltrating lymphocytes (TILs) are the last main subset of non-transformed cells within the TME. TILs comprise a diverse repertoire of cell types which can enhance or abrogate tumour progression [148]. For example, cytotoxic CD8+ T-cells mediate cancer cell death while Treg CD4+ T-cells suppress CD8+ T-cell cytotoxicity [149]. Gamma delta T-cells (γδTc) are another subset of TILs with both cytotoxic and regulatory roles depending on the type of cancer [150]. In a mouse model of PDGF-driven GBM, CSF1R inhibition resulted in tumour regression via re-education of

tumour associated macrophages (TAMs) from a pro-tumourigenic M2 state to an anti-tumourigenic phenotype [151]. However, a follow-up study by the same group found >50% of GBM tumours acquired resistance to prolonged CSF1R inhibition [152]. Intriguingly, resistance was linked to PIK3CA pathway activation in response to macrophage-derived IGF1 which, when blocked in combination with CSF1R, could reduce tumour recurrence. In a mouse model of pancreatic ductal adenocarcinoma (PDAC), a sub-population BM macrophages we required to establish a pre-metastatic niche in the liver, and subsequently PDAC metastasis [153]. A more comprehensive overview of TIL subset within the TME is provided by Gajewksi *et al.* [148]. In summary, many cell types within the TME are regulated by numerous extracellular factors that may be useful targets for cancer therapy.

1.8 The extracellular proteome

The complement of proteins and vesicular cargo that are secreted, shed or released by cells into the extracellular space is collectively known as the secretome [154]. Proteins secreted within the TME or tumour secretome harbour a diverse array of functions and can be broadly categorized into 4 groups: 1) signalling factors 2) extracellular matrix (ECM) proteins 3) proteases (enzymes) and 4) vesicular cargo.

1.8.1 Signalling factors

Growth factors, cytokines, chemokines and developmental morphogens all belong to a group of signalling factors which mediate autocrine, paracrine, and endocrine signalling events between cells in TME. Many of these factors regulate immune responses and wound healing but are often associated with worse patient prognosis and malignant progression [155]. For instance, CXCL12 and TGF-β potently stimulate fibroblast migration and invasion and their expression is often dysregulated in the tumour stroma [156]. Moreover, the pleiotropic cytokine interleukin-6 (IL6) is perhaps one of the most frequently elevated cytokines in

cancer [157]. IL6 is expressed by multiple cells within the TME and carries out a number of pro-inflammatory and tumour promoting functions. For example, in breast cancer, IL6 promotes monocyte and macrophage recruitment and high levels are correlated with worse overall survival [158]. Additionally, IL6 expression is associated with basal-like and stem-like CD44+/CD24- breast cancer cells [159]. Of note, the effects of IL6 are largely mediated through downstream activation of the JAK/STAT pathway which is consequently hyperactive in most cancers. IL6 signalling involves binding to and activation of the IL6 receptor/gp130 complex to recruit and phosphorylate Janus kinases (JAKs). JAKs subsequently phosphorylate and activate downstream signal transducer and activator of transcription (STAT) proteins, particularly STAT3, which in turn regulate transcription of target genes such as VEGF and matrix metalloproteainses (MMPs) [157]. Thus, given the link between inflammation and tumour progression, a number of drugs have been developed to target the IL6/JAK/STAT pathway for not only treating chronic inflammatory diseases but also cancer [157].

In addition to inflammatory cytokines, a number of proteins have been shown to exhibit pro-angiogenic functions like VEGFA. In particular, PDGFA, Placental Growth Factor (PlGF), Angiopoietins (ANGPTs) and FGFs may be important during acquired resistance to VEGF targeted therapies like Bevacizumab™ [160]. More recently, the extracellular factor microseminoprotein (MSMP) was found to harbour neovascular properties in ovarian tumour xenografts resistant to anti-VEGF therapy [161]. In this study, ovarian cancer conditioned media (CM) was found to induce endothelial tube formation and ERK activation in an MSMP/CCR2 dependent manner. Notably, combined inhibition of MSMP and VEGF *in vivo* profoundly reduced tumour burden and vessel density in xenografts previously resistant to anti-VEGF therapy [161]. Thus, multiple modes of anti-angiogenic interference may be required to prevent or overcome resistance.

1.8.2   Extracellular matrix proteins

The 3D network of scaffolding proteins in which tissues exist is termed the extracellular matrix (ECM) [162]. The ECM is deposited by cells to support tissue integrity and structure and to regulate cellular phenotypes and behaviours. For example, stemness, EMT, apical-basal polarity, migration, invasion and angiogenesis are all regulated by components of the ECM [162,163]. The ECM is a dynamic structure that is continuously remodeled during wound repair and cancer progression [163]. Although many proteins form the ECM, only ~300 within the mammalian genome are annotated as core matrisomal proteins which exhibit distinct, and often repeating, domains and post translational modifications (PTMs) [164]. The remaining matrisome-associated or 'ECM-affiliated' proteins are more diverse and varied. Core matrisomal proteins are comprised of collagens, proteoglycans and glycoproteins and a number of these proteins are cross-linked via transglutaminase or contain inter-chain disulfide bonds which make their extraction and solubilisation difficult [165,166]. Collagens form homo- and hetero-trimers that are characterized by repeating Gly-X-Y triplets with proline and 4-hydroxyprolines frequently located in the X and Y positions, respectively [167]. Collagen VI, in particular, has been shown to regulate apoptosis, proliferation, angiogenesis and inflammation and is often highly abundant in tumours relative to benign tissues [167]. Moreover, the COL6A3 fragment, Endotrophin, is upregulated in mammary tumours during breast cancer progression in mice expressing the mammary tumour virus–polyoma middle T antigen (MMTV-PyMT) promoter [168]. Proteoglycans are heavily modified with glycosaminoglycans (GAGs) or polymeric repeating units of disaccharides with carboxyl and sulfate groups [166]. GAGs sequester water and divalent cations due to their negative charge and therefore confer lubricating and space filling properties. The largest and most complex subset of core matrisomal proteins are glycoproteins. They may include basement membrane components like laminins, CCN-family members such as CTGF and growth factor binding proteins like latent transforming growth

factor binding protein (LTBP). In the case of TGF-β, LTBPs facilitate folding in the ER, proteolytic processing in the trans-Golgi network, secretion, matrix association and activation [169]. Insulin-like growth factor binding proteins (IGFBPs) are also glycoproteins which regulate the activity and stability of IGFs which are potent growth factors that are often highly expressed in cancer [170]. Growth factor binding proteins such as LTBPs and IGFBPs often have multiple binding partners and functions which complicate our understanding of their roles in normal physiology and cancer. For instance, LTBPs may also signal independently of TGF-β to modulate the organization of fibrillin microfibrils [169]. There are numerous other glycoproteins within the ECM however a full catalogue of their functions is lacking.

1.8.3 Proteases

Protease expression and activity is frequently dysregulated in cancer [171]. Next to kinases, proteases form the second largest group of enzymes in mammalian cells and provide exciting opportunities for targeting. This large family of proteins includes matrix metalloproteinases (MMPs), a disintegrin and metalloproteinases (ADAMs), convertases, cathepsins, kallikreins and other endoproteases which are important for regulating tissue homeostasis, ECM remodelling and the bio-activity and availability of signalling ligands. Proprotein convertases (PC) belong to a 9 membered family of mammalian Ca^{2+}-dependent serine proteases related to bacterial subtilisin that proteolytically process hormones, growth factors, receptors, cell adhesion molecules and enzymes [172]. PC function both intra and extracellularly and typically cleave the pro-domains of inactive precursors to generate mature ligands. FURIN is a ubiquitously expressed PC which becomes upregulated in a number of cancers and is also involved in NODAL signalling [173]. For instance, in the TNBC cell line MDA-MB-231, Lapierre *et al.* observed a decrease in the maturation of pro-PDGFA and pro-IGF1R when the PCs furin, and to lesser extent, PACE4, were inhibited [174]. However, inhibition of FURIN and

PACE4 differentially decreased and increased MDA-MB-231 cell motility, migration and invasion, respectively. Accordingly, FURIN and PACE4 activity were positively and negatively associated with MMP9 activity, respectively. For reference, MMP9 is a matrix metalloproteinase that is frequently upregulated in cancers including breast cancer and is important for ECM remodelling and cancer cell intravasation and extravasation [175]. Tissue kallikreins and kallikrein-related peptidases form a 15 membered group (KLK1-15) of serine proteases with tryptic and/or chymotryptic-like activity which regulate substrate activation, inactivation or degradation [176]. KLK expression profiles vary between tissues and their physiological roles include maintaining proper kidney, brain and skin-barrier function. Prostate specific antigen (PSA; also known as KLK3) is perhaps the most widely known KLK. The main function of KLK3 is seminal liquefaction via cleavage of fibronectin (FN) and seminogelin-1 and 2. However, during prostate cancer progression, KLK4 overexpression and low Zn^{2+} levels lead to increased KLK3 levels and in turn, cleavage of tumour-derived ECM proteins [176]. For example, cleavage of IGFBP3 by KLK3 promotes IGF1 dissociation and increased proliferation of prostatic stromal cells [177]. Other members of the kallikrein family have also been implicated in cancer progression and disease [178]. Cathepsins (CTS) are an 11 membered family of cysteine proteases with most members containing endopeptidases activity [179]. Like most proteases, cathepsin are synthesized as inactive zymogens with inhibitory pro-peptides that are cleaved by autocatalysis at low pH or other proteases. Cathepsins generally function within endosomal/lysosomal compartments but can be secreted into the extracellular space upon alterations in trafficking which occur in cancer and activated stromal cells. Of note, TAMs express CTS more robustly than any other cell type in the TME which is relevant given that Cathepsin S and L promote shedding of cell adhesion molecules (ALCAM, CD44 and neuropilin-1) and migration of MDA-MB-231 breast cancer cells [180]. Moreover, similar to KLKs, CTS may be useful diagnostic markers for cancer. For example, high CSTB expression is associated with worse overall survival in many cancers including breast and ovarian

[179,181,182]. Lastly, a recent large scale proteomic screen employing terminal amine isotopic labeling of substrates (TAILS) identified over 1110 substrates of the cathepsin family in prostate cancer [183]. This report highlights the widespread substrates of CTS (and proteases in general) in the TME and how proteomic technologies can aid in their detection.

1.8.4   Vesicular cargo

Intracellular proteins, counterintuitively, often comprise the largest fraction of species identified in secretome studies employing MS-based proteomics [184]. Indeed, highly abundant metabolic, transcriptional and translational proteins are liberated upon cell death and consequently, easily contaminate secretomes [184]. However, when cell death is carefully minimized, many intracellular proteins still remain suggesting the presence of non-classical secretion pathways for proteins lacking signal peptides [184]. Additionally, many intracellular proteins may be shed and located within extracellular vesicles (EVs) [185]. Several types of EVs are released by cells which vary in size, protein content and density. Exosomes, for example, are EVs ~30-120 nm in diameter which can transmit proteins and RNA (mRNA and miRNA) between cells. They are generated through endosomal trafficking and often express the surface markers CD9, CD63 and CD81 [186]. Little is known regarding role of EVs in intercellular protein transfer. Intriguingly, in 2008, Khalid Al-Nedawi *et al.* demonstrated the capacity of exosomes to transfer truncated EGFRvIII between expressing and non-expressing U373 glioma cells [187]. Remarkably, EGFRvIII containing exosomes were found to promote tumour growth *in vivo* and induce ERK1/2 and Akt activation *in vitro* which was correlated with VEGF and Bcl-x production. These findings have significant implications for cancer detection and treatment and also show that oncogenic transmission can be reversed upon blocking exosomal mediated protein transfer. In addition to exosomes, cells also release/shed vesicles and apoptotic blebs ranging from 50-1000 nm in diameter comprised of various waste and cellular (by-)products. A

recent study by Harel *et al.* characterized plasma microparticles by LC-MS and identified a signature of protein biomarkers (PPTN1, SFXN3 and LPP) which could identify prostate cancer with a receiver operating characteristic area-under-the-curve (ROC-AUC) of ~0.84 [188]. Taken together, the complement of EV associated proteins may reveal useful biomarkers and insights in biological processes exploited by cancers.

1.9   *Modelling the complexity of TME*

The TME is complex and dynamic and a number of approaches have been used to investigate its components individually or as a whole [189]. Properties intrinsic to the sample(s) or the model in question can limit the applicability of experimental findings and therefore complementary approaches may help generate a more comprehensive picture of the TME. Monolayer cultures of established cell lines are the most widely utilized model for examining cancer related questions [190,191]. This can be primarily attributed to ease of propagation and manipulation afforded by tissue culture plastic formats. Moreover, commonly used cancer lines have well defined genetic and molecular profiles which make experimental outcomes and findings relatable to previously published work [192,193]. Unfortunately, monolayer cultures and cancer lines do not accurately recapitulate the TME due to several reasons [194]. For one, cancer lines at their point of isolation, reflect a state of disease progression which over time in culture, may shift as a result of clonal selection and genomic instability. For instance, an analysis of the Cancer Cell Line Encyclopedia (CCLE) found a higher median mutation frequency in ovarian cell lines compared with tumours samples (4.3/Mb versus 1.6/Mb) [20]. Another shortcoming of monolayer cultures is the absence of accessory cell types frequently found in the TME. To examine these variables, Vincent *et al.* compared RNA sequencing data from breast cancer cell lines and breast tumours to measure differences in immune and stromal scores based on the 'Estimation of STromal and Immune cells in MAlignant Tumours using Expression data' (ESTIMATE)

algorithm [195,196]. Strikingly, 134 of the top 163 proteins differentially expressed between breast cancer cells and tumours were positively correlated with stromal score and likely attributed to the lack of stromal cells in monolayer cultures. This work and earlier studies also found proliferative, metabolic and transcriptional pathways upregulated in cell lines compared to tumours [191,196]. Conversely, pathways involved cell adhesion and communication were downregulated in cell lines. Regardless of their shortcomings, cell lines have been essential for understanding cell biology. For example, cell lines in isolation afford the ability determine the origin of extracellular factors involved in bi-directional cellular communication within the TME. Moreover, the NCI panel of 60 cell line resource has been exhaustively studied and is an import cornerstone for drug screening [192].

However, several modifications to existing culture systems can be employed to better mimic the TME and limit the deficiencies highlighted above. First, primary patient derived samples which have undergone less selective pressure may resemble *in vivo* phenotypes more closely [193]. Moreover, tissue/tumour pieces can be cultured *ex vivo* for short periods of time to monitor invasion in response to inhibitors or drugs [197]. Co-cultures enable reciprocal signalling events between cancer cells and other cell types although tracing cell-type derived products require specific labelling techniques [198]. To represent the 3D structure of tissues *in vivo*, cell lines can be cultured in bioreactors or in ultra-low attachment conditions to promote spheroid formation [189,199]. These spheroids can develop regions of hypoxia with anoxia and necrotic cores that more accurately recapitulate tumour physiology.

The ECM is another critical component of tumour biology often absent in monolayer cultures. Reconstituted basement membrane extracts (BME) such as Matrigel™ and GelTrex™ can be utilized to substitute the ECM *in vitro* and *in vivo* [200]. For example, Matrigel™ which is derived from the Engelbreth-Holm-Swarm

mouse sarcoma, is widely used to support feeder-free hESC cultures. Although predominantly comprised of laminin and entactin, mass spectrometry-based proteomic characterization of Matrigel identified over 1800 proteins including growth factors, regulatory proteins, intracellular contents and membrane bound proteins [200]. Furthermore, matrices preconditioned by hESC can maintain pluripotency over multiple passages and thus highlights the importance of the ECM for modeling complex biological systems like stem cells and cancer [201].

Finally, patient derived xenografts (PDXs) afford the most complex models of tumour biology which can be serially propagated *in vivo* [202,203]. While PDXs may be an attractive model of cancer pathology, a recent analysis of over 1100 PDXs found divergence and loss of copy number alterations following serial propagation in mice [204]. Thus, there are a number of models to consider when studying the TME, each with their own benefits and drawbacks.

1.10 Mass spectrometry-based proteomics

As discussed above, the tumour secretome plays an integral role in dictating the activities and types of accessory cells within the TME [3]. The complexity of the TME, however, poses significant challenges for characterization. Commonly used 'omic' approaches such as microarrays, RNA sequencing and whole exome sequencing can infer pathways involved in cancer progression [205]. Of caution, transcripts are not the functional machinery of cellular behaviours and their levels do not always correlate with protein expression [206]. Proteomics involves the study of proteomes or the protein complement of a biological sample, and is an alternative means of investigating cellular activities. Western blotting and Enzyme Linked Immunosorbent Assays (ELISAs) are perhaps the simplest proteomic tools available. Both are robust techniques with several limitations. ELISAs are widely used in clinical settings due to high sensitivity and accurate quantitation [207]. ELISAs are capable of processing large numbers of samples but are restricted to one or several targets and require access to high-quality antibodies [208]. Multiplex

cytokine arrays and Reverse Phase Protein Arrays (RPPA) are an alternative approach with limited throughput but the capacity to measure a larger number of factors (dozens to hundreds of targets) [208]. Like ELISAs and Western blotting, protein arrays also require validated antibodies and targets must be selected a priori.

Although ELISAs and protein arrays are compatible with complex biological samples such as serum, they remain limited in their capacity for comprehensive tumour secretome characterization which is comprised of thousands of proteins. Bottom-up mass spectrometry-based approaches, however, identify peptide sequence information which can be matched to thousands of proteins in an unbiased, medium-to-high throughput manner [209]. Moreover, with some modifications, MS-based approaches are amenable to a wide array of biological formats including cell lines, fluids (i.e. blood, urine, saliva, effusions, ascites), and tissues. In effect, MS-based proteomics holds enormous potential for identifying drug targets, biomarkers and understanding cancer biology [207].

1.10.1   Sample preparation and fractionation strategies

Sample preparation is critical task that should be carefully planned and optimized and is vital to the success of any proteomics study. The nature of experimental question(s) dictate the means in which samples are collected, processed and analyzed. A standard bottom-up proteomics workflow involves sample collection, protein extraction and solubilisation, reduction and alkylation, digestion, clean-up, chromatographic separation by HPLC, acquisition of MS and MS/MS spectra, and database searching.

Depending on the source of material to be analyzed, samples may require enrichment or pre-processing. Harsh lysis buffers utilizing chaotropes, detergents and reducing agents combined with mechanical dissociation are often an effective means of protein extraction and amenable to most materials. For instance, urea

and disulfide reduction can improve the extraction and detection of ECM proteins that are otherwise insoluble. Unlike tissues/cell pellets, the protein concentrations of fluids are more dilute and may require concentration or enrichment. Molecular weight cut-off (MWCO) filter units combined with centrifucation are a simple yet effective apparatus for concentrating extracellular proteins. One caveat when working with biological fluids or cell culture media is the presence of high abundance proteins (HAP) such as Albumin. FBS or serum supplements can often be omitted for short periods of time *in vitro*. However, this is not possible with patient fluids like plasma. Depletion strategies are one method to overcome the presence of HAP which often exceed the dynamic range of MS instruments [210]. Alternatively, proteins with specific post-translational modifications can be affinity purified. For example, secreted proteins may be enriched using glyco-capture and phospho-peptide enrichment with TiO2 beads for measuring kinase signalling [211,212]. Moreover, extracellular vesicles can be isolated using affinity purification or ultracentrifugation [213]. With regard to the ECM proteins, ammonium sulfate precipitation has been used to crudely precipitate less soluble moderate-to-high molecular weight proteins (generally present in high abundance) from more soluble, low molecular weight proteins which often contain low abundance growth factors [200]. Lastly, fractionation strategies can improve the detection of low abundance proteins by analyzing more material across multiple LC-MS runs. A large number of strategies exist for sample fractionation with SDS-PAGE traditionally being the most widely used. Recent innovations and newer techniques have significantly simplified fractionation workflows or achieved greater results. These include Stop-and-go-extraction Tip (StageTip) fractionation, offline HPLC basic reversed phase separation, and selective elution from beads via a novel SP3 approach [214–217]. Impressively, a recent study combined multiple enzyme digestions with high-pH fractionation and state-of-the-art mass-spectrometry instrumentation to achieve unprecedented HeLa proteome coverage to a depth of 14K proteins [209]. Continuous advances in sample preparation

modalities and MS instrumentation are narrowing the gap between genomic sequencing and proteomic technologies.

1.10.2 Quantification strategies

Quantification is a key aspect of modern day MS-based proteomics [218]. Robust, accurate and flexible quantification strategies are needed to enable comparisons between as little as two samples to potentially hundreds. Mass spectrometry techniques can utilize a number methods for relative and absolute protein (peptide) quantification and like sample preparation, the method employed will vary depending on the experimental questions being asked.

Unlike most quantification techniques, label free quantification (LFQ) does not require special considerations and can be implemented into most proteomic workflows. Although LFQ strategies have been available for decades, only recently have software and instrumentation improvements increased the accuracy and robustness [219]. Spectral counting and intensity based absolute quantification (iBAQ) have been largely utilized as surrogates or proxies for protein abundance [206,220]. Interference due to co-eluting peptides with highly similar m/z values is a primary concern for all extracted ion chromatogram (XIC)-based quantification strategies such as LFQ. Of note, high resolution mass analyzers combined with narrow m/z search tolerances can alleviate most interference issues. Furthermore, feature detection algorithms utilizing retention time (RT) windows and isotopic distribution patterns improve quantification across adjacent runs and fractions where peak splitting and technical variation are encountered [219]. Stable Isotopic Labelling in Cell Culture (SILAC) is a metabolic labelling technique that also employs XIC-based quantification like LFQ [221]. SILAC uniquely enables protein samples from different cell lines or treatments to be mixed which eliminates technical variability associated with sample handling and LC-MS. SILAC, however, requires isotopically labelled cells in order to compare relative changes in protein expression/PTMs between two or more conditions. In a typical SILAC workflow,

cell lines are grown in media containing either light or heavy isotopes of Arginine and Lysine. Thus, newly synthesized proteins incorporate light or heavy amino acids and after several population doublings entire proteomes become isotopically labelled. Peak intensities between pairs of light and heavy peptides (m/z values) from pooled SILAC samples are used to compare relative protein (peptide) levels. SILAC provides more accurate quantification and can achieve more than twice the throughput than that of LFQ. However, sample complexity is effectively doubled with SILAC which therefore limits the number of unique peptides that can be sequenced in a given LC-MS run due to redundant MS/MS scans for light and heavy pairs. Of note, several variations of SILAC have increased its utility beyond comparing two cell types or conditions. Super SILAC, for example, employs a heavy standard comprised of multiple heavy labelled cell lines. When spiked into light (unlabeled) samples, the super SILAC standard acts as an internal reference which protein levels can be normalized to and compared across many samples. For reference, Pozniak *et al.* utilized super SILAC to compare proteomes from healthy breast ductal epithelia, lymph node negative, lymph node positive and metastatic breast tumours [222]. This study found key that DNA repair proteins involved in non-homologous end joining, mismatch repair, and the single-strand DNA break repair complex were downregulated in tumours while several enzymes involved in oxidative phosphorylation and glycolysis were significantly increased and decreased, respectively. Interestingly, in tumour versus normal tissues, these authors noted an inverse relationship between The Cancer Genome Atlas (TCGA) mRNA expression and protein levels for DNA repair, ribosomal and glycolytic enzymes. Cell Type specific labeling using Amino acid Precursors (CTAP) is another elegant metabolic labelling technique that enables proteins to be traced to cell of origin in co-culture conditions [198]. In this technique, transgenic cell lines express enzymes (lysine racemase or diaminopimelate decarboxylase) which utilize different L-lysine precursors. Given that lysine precursors are only utilized by one cell type, transgenic line can be studied in the same culture dish. Using this strategy, Tape *et al.* investigated the reciprocal interactions between PDAC and

pancreatic stromal cells [223]. CyToF is another MS-based approach which employs antibodies labelled with metals to multiple surface protein expression on single cells to a greater extent than afforded by flow cytometry [224]. Using this technology, Bendall *et al.* were able to measure 34 parameters including 31 proteins on single cells to generate a complex map of hematopoietic lineages.

To overcome limitations associated with LFQ, chemical labelling strategies including isotope coded affinity tags (ICAT) and tandem mass tags (TMT) were developed [225]. These chemical labelling technique are implemented following sample digestion and do not induced mass differences between MS1 precursors like SILAC. Rather, quantification is performed on "tags" from different channels which are produced during MS/MS fragmentation. Up to 10 or more samples can be multiplexed with this technique allowing for high throughput proteome analysis [226]. Of note, noise associated with the low m/z regions where tags are quantified can interfere with quantification [227]. MS3 scans can significantly improve specificity and the cost of sensitivity and speed [228]. However, an elegant solution to these issues was the development of Notched MS3 scans (trademarked as synchronous precursor selection) by Coon's group [227]. This feature is only available on state-of-the-art Orbitrap Fusion instruments.

Selective reaction monitoring (SRM) and more recently, parallel reaction monitoring (PRM), are the last quantification strategies which harbour the greatest potential for diagnostic and clinical utility [229]. These targeted methods require peptide retention time (RT) information and are typically limited to small number of transitions. Synthetic stable isotope labeled (SIL) peptides can be utilized for absolute quantification. They afford highly accurate and sensitive MS2 quantification that is less susceptible to interference from all sources and can be rapidly performed using shorter gradients.

1.11 Rationale and thesis objectives

A variety of approaches and model systems are required to answer complex biological questions involving the TME. MS-based proteomics is a powerful technology that can improve our understanding the TME. In this thesis, I focused on better understanding the complexities of the TME by utilizing mass spectrometry to study several aspects of cancer biology important for: **1) the classification of ovarian cancer subtypes, 2) the identification of biomarkers for ovarian cancer detection, and 3) understanding the global effects of NODAL in the breast cancer microenvironment.**

1.12 References:

[1]   Howlader, N., Noone, A., Krapcho, M., Miller, D., et al., SEER Cancer Statistics Review, 1975-2014, National Cancer Institute 2017.

[2]   Canadian Cancer Society, Canadian Cancer Statistics 2017. *Can. Cancer Soc.* 2017, 2017, 1–132.

[3]   Hanahan, D., Weinberg, R. a., Hallmarks of cancer: the next generation. *Cell* 2011, 144, 646–74.

[4]   Balkwill, F.R., Capasso, M., Hagemann, T., The tumor microenvironment at a glance. *J. Cell Sci.* 2012, 125, 5591–6.

[5]   Karnezis, A.N., Cho, K.R., Gilks, C.B., Pearce, C.L., Huntsman, D.G., The disparate origins of ovarian cancers: pathogenesis and prevention strategies. *Nat. Rev. Cancer* 2017, 17, 65–74.

[6]   Torre, L.A., Islami, F., Siegel, R.L., Ward, E.M., Jemal, A., Global cancer in women: Burden and trends. *Cancer Epidemiol. Biomarkers Prev.* 2017, 26, 444–457.

[7]   Roy, A., Matzuk, M.M., Reproductive tract function and dysfunction in women. *Nat. Rev. Endocrinol.* 2011, 7, 517–525.

[8]   Bremnes, R.M., Dønnem, T., Al-Saad, S., Al-Shibli, K., et al., The Role of Tumor Stroma in Cancer Progression and Prognosis: Emphasis on Carcinoma-Associated Fibroblasts and Non-small Cell Lung Cancer. *J. Thorac. Oncol.* 2011, 6, 209–217.

[9]   Cross, J., Werb, Z., Fisher, S., Implantation and the placenta: key pieces of the development puzzle. *Science (80-.).* 1994, 266, 1508–1518.

[10]   Lengyel, E., Ovarian Cancer Development and Metastasis. *Am. J. Pathol.* 2010, 177, 1053–1064.

[11]   Cormio, G., Rossi, C., Cazzolla, A., Resta, L., et al., Distant metastases in ovarian carcinoma. *Int. J. Gynecol. Cancer* 2003, 13, 125–129.

[12]   Ludwig, J. a, Weinstein, J.N., Biomarkers in cancer staging, prognosis and treatment selection. *Nat. Rev. Cancer* 2005, 5, 845–856.

[13]   American Cancer Society, Ovarian Cancer Stages 2018.

[14]   American Cancer Society, Survival Rates for Ovarian Cancer, by Stage 2018.

[15]   Kurman, R., Shih, I., The Origin and pathogenesis of epithelial ovarian cancer-a proposed unifying theory. *Am. J. Surg. Pathol.* 2010, 34, 433–443.

[16]   Sankaranarayanan, R., Ferlay, J., Worldwide burden of gynaecological cancer: The size of the problem. *Best Pract. Res. Clin. Obstet. Gynaecol.* 2006, 20, 207–225.

[17]   Phelan, C.M., Kuchenbaecker, K.B., Tyrer, J.P., Kar, S.P., et al., Identification of 12 new

susceptibility loci for different histotypes of epithelial ovarian cancer. *Nat. Genet.* 2017, 49, 4–5.

[18]   Köbel, M., Kalloger, S.E., Boyd, N., McKinney, S., et al., Ovarian carcinoma subtypes are different diseases: Implications for biomarker studies. *PLoS Med.* 2008, 5, 1749–1760.

[19]   Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma. *Nature* 2011, 474, 609–15.

[20]   Domcke, S., Sinha, R., Levine, D. a, Sander, C., Schultz, N., Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* 2013, 4, 2126.

[21]   Olivier, M., Hollstein, M., Hainaut, P., TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb. Perspect. Biol.* 2010, 2, 1–17.

[22]   Harris, S.L., Levine, A.J., The p53 pathway: Positive and negative feedback loops. *Oncogene* 2005, 24, 2899–2908.

[23]   Köbel, M., Bak, J., Bertelsen, B.I., Carpen, O., et al., Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry. *Histopathology* 2014, 64, 1004–1013.

[24]   Gilks, C.B., Ionescu, D.N., Kalloger, S.E., Köbel, M., et al., Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. *Hum. Pathol.* 2008, 39, 1239–1251.

[25]   Kindelberger, D.W., Lee, Y., Miron, A., Hirsch, M.S., et al., Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. *Am. J. Surg. Pathol.* 2007, 31, 161–169.

[26]   Lee, Y., Miron, A., Drapkin, R., Nucci, M., et al., A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *J. Pathol.* 2007, 211, 26–35.

[27]   Labidi-Galy, S.I., Papp, E., Hallberg, D., Niknafs, N., et al., High grade serous ovarian carcinomas originate in the fallopian tube. *Nat. Commun.* 2017, 8, 1–10.

[28]   Banet, N., Kurman, R.J., Two types of ovarian cortical inclusion cysts: Proposed origin and possible role in ovarian serous carcinogenesis. *Int. J. Gynecol. Pathol.* 2015, 34, 3–8.

[29]   Altman, A.D., Nelson, G.S., Ghatage, P., McIntyre, J.B., et al., The diagnostic utility of TP53 and CDKN2A to distinguish ovarian high-grade serous carcinoma from low-grade serous ovarian tumors. *Mod. Pathol.* 2013, 26, 1255–1263.

[30]   Fukunaga, M., Nomura, K., Ishikawa, E., Ushigome, S., Ovarian atypical endometriosis: its close association with malignant epithelial tumours. *Histopathology* 1997, 30, 249–255.

[31]   Ogawa, S., Kaku, T., Amada, S., Kobayashi, H., et al., Ovarian endometriosis associated with ovarian carcinoma: A clinicopathological and immunohistochemical study. *Gynecol. Oncol.* 2000, 77, 298–304.

[32]   Prowse, A.H., Manek, S., Varma, R., Liu, J., et al., Molecular genetic evidence that endometriosis is a precursor of ovarian cancer. *Int. J. Cancer* 2006, 119, 556–562.

[33] Cochrane, D.R., Tessier-Cloutier, B., Lawrence, K.M., Nazeran, T., et al., Clear Cell and Endometrioid Carcinomas: are their differences attributable to distinct cells of origin? *J. Pathol.* 2017, 26–36.

[34] Kubeček, O., Laco, J., Špaček, J., Petera, J., et al., The pathogenesis, diagnosis, and management of metastatic tumors to the ovary: a comprehensive review. *Clin. Exp. Metastasis* 2017, 34, 295–307.

[35] Wang, Y.K., Bashashati, A., Anglesio, M.S., Cochrane, D.R., et al., Genomic consequences of aberrant DNA repair mechanisms stratify ovarian cancer histotypes. *Nat. Genet.* 2017, 49, 856–864.

[36] Raja, F.A., Chopra, N., Ledermann, J.A., Optimal first-line treatment in ovarian cancer. *Ann. Oncol.* 2012, 23, x118–x127.

[37] Jayson, G.C., Kohn, E.C., Kitchener, H.C., Ledermann, J.A., Ovarian cancer. *Lancet* 2014, 384, 1376–1388.

[38] McGrogan, B.T., Gilmartin, B., Carney, D.N., McCann, A., Taxanes, microtubules and chemoresistant breast cancer. *Biochim. Biophys. Acta - Rev. Cancer* 2008, 1785, 96–132.

[39] Dasari, S., Bernard Tchounwou, P., Cisplatin in cancer therapy: Molecular mechanisms of action. *Eur. J. Pharmacol.* 2014, 740, 364–378.

[40] Rojas, V., Hirshfield, K.M., Ganesan, S., Rodriguez-Rodriguez, L., Molecular characterization of epithelial ovarian cancer: Implications for diagnosis and treatment. *Int. J. Mol. Sci.* 2016, 17.

[41] Institute, N.C., Genetics of Breast and Gynecologic Cancers (PDQ®)-Health Professional Version 2018.

[42] Kuchenbaecker, K.B., Hopper, J.L., Barnes, D.R., Phillips, K.-A., et al., Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers. *Jama* 2017, 317, 2402.

[43] Roy, R., Chun, J., Powell, S.N., BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nat. Rev. Cancer* 2012, 12, 68–78.

[44] Banerjee, S., Kaye, S.B., New strategies in the treatment of ovarian cancer: Current clinical perspectives and future potential. *Clin. Cancer Res.* 2013, 19, 961–968.

[45] Lord, C.J., Ashworth, A., PARP inhibitors: Synthetic lethality in the clinic. *Science* 2017, 355, 1152–1158.

[46] Ledermann, J.A., PARP inhibitors in ovarian cancer. *Ann. Oncol.* 2016, 27, i40–i44.

[47] Sieh, W., Köbel, M., Longacre, T.A., Bowtell, D.D., et al., Hormone-receptor expression and ovarian cancer survival: An Ovarian Tumor Tissue Analysis consortium study. *Lancet Oncol.* 2013, 14, 853–862.

[48] Lange, C.A., Yee, D., Progesterone and Breast Cancer. *Women's Heal.* 2008, 4, 151–162.

[49]   Strimbu, K., Tavel, J. a, What are Biomarkers? *Curr Opin HIV AIDS* 2011, 5, 463–466.

[50]   Nolen, B., Lokshin, A., Protein biomarkers of ovarian cancer : the forest and the trees. *Futur. Oncol.* 2012, 8, 55–71.

[51]   Woolas, R.P., Xu, F.J., Jacobs, I.J., Yu, Y.H., et al., Elevation of multiple serum markers in patients with stage I ovarian cancer. *J. Natl. Cancer Inst.* 1993, 85, 1748–1751.

[52]   Johnson, C.C., Kessel, B., Riley, T.L., Ragard, L.R., et al., The epidemiology of CA-125 in women without evidence of ovarian cancer in the Prostate, Lung, Colorectal and Ovarian Cancer (PLCO) Screening Trial. *Gynecol. Oncol.* 2008, 110, 383–389.

[53]   van Nagell, J.R., Hoff, J.T., Transvaginal ultrasonography in ovarian cancer screening: current perspectives. *Int. J. Womens. Health* 2013, 6, 25–33.

[54]   Jacobs, I.J., Menon, U., Ryan, A., Gentry-Maharaj, A., et al., Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* 2016, 387, 945–956.

[55]   Buys, S.S., Partridge, E., Black, A., Johnson, C.C., et al., Effect of Screening on Ovarian Cancer Mortality. *JAMA* 2011, 305, 2295–303.

[56]   Yip, P., Chen, T.H., Seshaiah, P., Stephen, L.L., et al., Comprehensive serum profiling for the discovery of epithelial ovarian cancer biomarkers. *PLoS One* 2011, 6, 1–10.

[57]   Cramer, D.W., Bast, R.C., Berg, C.D., Diamandis, E.P., et al., Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev. Res.* 2011, 4, 365–374.

[58]   Cohen, J.D., Li, L., Wang, Y., Thoburn, C., et al., Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science (80-.).* 2018, 3247, eaar3247.

[59]   Moorman, P.G., Calingaert, B., Palmieri, R.T., Iversen, E.S., et al., Hormonal risk factors for ovarian cancer in premenopausal and postmenopausal women. *Am. J. Epidemiol.* 2008, 167, 1059–1069.

[60]   Shah, R., Pathogenesis, prevention, diagnosis and treatment of breast cancer. *World J. Clin. Oncol.* 2014, 5, 283.

[61]   Beral, V., Doll, R., Hermon, C., Peto, R., Reeves, G., Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet* 2008, 371, 303–14.

[62]   Hassiotou, F., Geddes, D., Anatomy of the human mammary gland: Current status of knowledge. *Clin. Anat.* 2013, 26, 29–48.

[63]   Macias, H., Hinck, L., Mammary gland development. *Wiley Interdiscip. Rev. Dev. Biol.* 2013, 1, 533–57.

[64]   Gjorevski, N., Nelson, C.M., Integrated morphodynamic signalling of the mammary gland. *Nat. Rev. Mol. Cell Biol.* 2011, 12, 581–593.

[65] Hortobagyi, G.N., Connolly, J.L., D'Orsi, C.J., Edge, S.B., et al., *AJCC Cancer Staging Manual, Eighth Edition: Breast*, 2017.

[66] Yersal, O., Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J. Clin. Oncol.* 2014, 5, 412.

[67] Howlader, N., Altekruse, S.F., Li, C.I., Chen, V.W., et al., US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. *J. Natl. Cancer Inst.* 2014, 106.

[68] Kohler, B.A., Sherman, R.L., Howlader, N., Jemal, A., et al., Annual Report to the Nation on the Status of Cancer, 1975-2011, Featuring Incidence of Breast Cancer Subtypes by Race/Ethnicity, Poverty, and State. *J. Natl. Cancer Inst.* 2015, 107, djv048.

[69] Spitale, A., Mazzola, P., Soldini, D., Mazzucchelli, L., Bordoni, A., Breast cancer classification according to immunohistochemical markers: Clinicopathologic features and short-term survival analysis in a population-based study from the South of Switzerland. *Ann. Oncol.* 2009, 20, 628–635.

[70] Cheang, M.C.U., Chia, S.K., Voduc, D., Gao, D., et al., Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J. Natl. Cancer Inst.* 2009, 101, 736–750.

[71] Dai, X., Li, T., Bai, Z., Yang, Y., et al., Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res* 2015, 5, 2929–2943.

[72] Prat, A., Parker, J.S., Karginova, O., Fan, C., et al., Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res.* 2010, 12, R68.

[73] Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., et al., Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490, 61–70.

[74] Fredika M. Robertson, S.K., Sanford H. Barsky, Melissa Bondy, Wei Yang, Savitri Krishnamurthy, Wendy A. Woodward, Hideko Yamauchi, Huong Le-Petross, T.B., Cristofanilli., M., Inflammatory Breast Cancer. *CA Cancer J Clin* 2010, 60, 351–375.

[75] Cristofanilli, M., Valero, V., Buzdar, A.U., Kau, S.W., et al., Inflammatory breast cancer (IBC) and patterns of recurrence: Understanding the biology of a unique disease. *Cancer* 2007, 110, 1436–1444.

[76] Gong, Y., Liu, Y.-R., Ji, P., Hu, X., Shao, Z.-M., Impact of molecular subtypes on metastatic breast cancer patients: a SEER population-based study. *Sci. Rep.* 2017, 7, 45411.

[77] Molyneux, G., Geyer, F.C., Magnay, F.A., McCarthy, A., et al., BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 2010, 7, 403–417.

[78] Keller, P.J., Arendt, L.M., Skibinski, A., Logvinenko, T., et al., Defining the cellular precursors to human breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* 2012, 109, 2772–7.

[79] Skibinski, A., Kuperwasser, C., The origin of breast tumor heterogeneity 2015, 34, 5309–5316.

[80] Hodgson, R., Heywang-Köbrunner, S.H., Harvey, S.C., Edwards, M., et al., Systematic review of 3D mammography for breast cancer screening. *Breast* 2016, 27, 52–61.

[81] Tabár, L., Gad, A., Holmberg, L.., Ljungquist, U., et al., Reduction in Mortality From Breast Cancer After Mass Screening With Mammography. *Lancet* 1985, 325, 829–832.

[82] American Cancer Society, Understanding a Breast Cancer Diagnosis 2018.

[83] Pohlmann, P.R., Mayer, I.A., Mernaugh, R., Resistance to trastuzumab in breast cancer. *Clin. Cancer Res.* 2009, 15, 7479–7491.

[84] Geenen, J.J.J., Linn, S.C., Beijnen, J.H., Schellens, J.H.M., PARP Inhibitors in the Treatment of Triple-Negative Breast Cancer. *Clin. Pharmacokinet.* 2017, 57, 427–437.

[85] Wahl, G.M., Spike, B.T., Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. *npj Breast Cancer* 2017, 3, 14.

[86] Quail, D.F., Siegers, G.M., Jewer, M., Postovit, L.M., Nodal signalling in embryogenesis and tumourigenesis. *Int. J. Biochem. Cell Biol.* 2013, 45, 885–898.

[87] Van Keymeulen, A., Lee, M.Y., Ousset, M., Brohée, S., et al., Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* 2015, 525, 119–123.

[88] Nguyen, L. V, Vanner, R., Dirks, P., Eaves, C.J., Cancer stem cells: an evolving concept. *Nat. Rev. cancer* 2012, 12, 133–143.

[89] Croker, A.K., Goodale, D., Chu, J., Postenka, C., et al., High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. *J. Cell. Mol. Med.* 2009, 13, 2236–2252.

[90] Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., Clarke, M.F., Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci.* 2003, 100, 3983–3988.

[91] Croker, A.K., Allan, A.L., Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44+ human breast cancer cells. *Breast Cancer Res. Treat.* 2012, 133, 75–87.

[92] Louderbough, J.M. V, Schroeder, J.A., Understanding the Dual Nature of CD44 in Breast Cancer Progression. *Mol. Cancer Res.* 2011, 9, 1573–1586.

[93] Sacks, J.D., Barbolina, M. V., Expression and function of CD44 in epithelial ovarian carcinoma. *Biomolecules* 2015, 5, 3051–3066.

[94] Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., et al., Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res.* 2005, 65, 5506–5511.

[95] Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., et al., An embryonic stem cell-like

gene expression signature in poorly differentiated aggressive human tumors. *Nat. Genet.* 2008, 40, 499–507.

[96] Spike, B.T., Engle, D.D., Lin, J.C., Cheung, S.K., et al., A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* 2012, 10, 183–197.

[97] Anastas, J.N., Moon, R.T., WNT signalling pathways as therapeutic targets in cancer. *Nat. Rev. Cancer* 2013, 13, 11–26.

[98] Gong, W., Sun, B., Sun, H., Zhao, X., et al., Nodal signaling activates the Smad2/3 pathway to regulate stem cell-like properties in breast cancer cells. *Am. J. Cancer Res.* 2017, 7, 503–517.

[99] Bodenstine, T.M., Chandler, G.S., Seftor, R.E.B., Seftor, E.A., Hendrix, M.J.C., Plasticity underlies tumor progression: role of Nodal signaling. *Cancer Metastasis Rev.* 2016, 35, 21–39.

[100] Topczewska, J.M., Postovit, L.-M., Margaryan, N. V, Sam, A., et al., Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat. Med.* 2006, 12, 925–932.

[101] Schier, A.F., Nodal morphogens. *Cold Spring Harb. Perspect. Biol.* 2009, 1.

[102] Quail, D.F., Zhang, G., Findlay, S.D., Hess, D. a, Postovit, L.-M., Nodal promotes invasive phenotypes via a mitogen-activated protein kinase-dependent pathway. *Oncogene* 2014, 33, 461–73.

[103] Vallier, L., Alexander, M., Pedersen, R.A., Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell Sci.* 2005, 118, 4495–509.

[104] Vallier, L., Mendjan, S., Brown, S., Chng, Z., et al., Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. *Development* 2009, 136, 1339–1349.

[105] Postovit, L.-M., Margaryan, N. V, Seftor, E.A., Kirschmann, D. a, et al., Human embryonic stem cell microenvironment suppresses the tumorigenic phenotype of aggressive cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 4329–34.

[106] Bianco, C., Rangel, M.C., Castro, N.P., Nagaoka, T., et al., Role of Cripto-1 in stem cell maintenance and malignant progression. *Am. J. Pathol.* 2010, 177, 532–540.

[107] Massagué, J., TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* 2012, 13, 616–630.

[108] Antsiferova, M., Werner, S., The bright and the dark sides of activin in wound healing and cancer. *J. Cell Sci.* 2012, 125, 3929–3937.

[109] Quail, D.F., Zhang, G., Walsh, L. a, Siegers, G.M., et al., Embryonic Morphogen Nodal Promotes Breast Cancer Growth and Progression. *PLoS One* 2012, 7, 1–12.

[110] Quail, D.F., Walsh, L. a., Zhang, G., Findlay, S.D., et al., Embryonic protein nodal promotes breast cancer vascularization. *Cancer Res.* 2012, 72, 3851–3863.

[111] Hanahan, D., Coussens, L.M., Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. *Cancer Cell* 2012, 21, 309–322.

[112] Carmeliet, P., Jain, R.K., Molecular mechanisms and clinical applications of angiogenesis. *Nature* 2011, 473, 298–307.

[113] Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., et al., Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* 1996, 380, 439–442.

[114] Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., et al., Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* 1996, 380, 435–439.

[115] Gilkes, D.M., Semenza, G.L., Wirtz, D., Hypoxia and the extracellular matrix: drivers of tumour metastasis. *Nat. Rev. Cancer* 2014, 14, 430–9.

[116] Folkman, J., Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.* 1971, 285, 1182–6.

[117] Kim, K.J., Li, B., Winer, J., Armanini, M., et al., Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo. *Nature* 1993, 362, 841–4.

[118] Ferrara, N., Hillan, K.J., Gerber, H.-P., Novotny, W., Case history: Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat. Rev. Drug Discov.* 2004, 3, 391–400.

[119] Keating, G.M., Bevacizumab: A review of its use in advanced cancer. *Drugs* 2014, 74, 1891–1925.

[120] Bergers, G., Hanahan, D., Modes of resistance to anti-angiogenic therapy. *Nat. Rev. Cancer* 2008, 8, 592–603.

[121] Wang, D., Tan, C., Xiao, F., Zou, L., et al., The "inherent vice" in the anti-Angiogenic theory may cause the highly metastatic cancer to spread more aggressively. *Sci. Rep.* 2017, 7, 1–14.

[122] Patenaude, A., Parker, J., Karsan, A., Involvement of endothelial progenitor cells in tumor vascularization. *Microvasc. Res.* 2010, 79, 217–223.

[123] Asahara, T., Murohara, T., Sullivan, a, Silver, M., et al., Isolation of putative progenitor endothelial cells for angiogenesis. *Science* 1997, 275, 964–967.

[124] Asahara, T., Masuda, H., Takahashi, T., Kalka, C., et al., Bone marrow origin of endothelial progenitor cells responsible for postnatal vasculogenesis in physiological and pathological neovascularization. *Circ. Res.* 1999, 85, 221–228.

[125] Asahara, T., Takahashi, T., Masuda, H., Kalka, C., et al., VEGF contributes to postnatal neovascularization by mobilizing bone marrow-derived endothelial progenitor cells. *EMBO J.* 1999, 18, 3964–3972.

[126] Nolan, D.J., Ciarrocchi, A., Mellick, A.S., Jaggi, J.S., et al., Bone marrow-derived endothelial progenitor cells are a major determinant of nascent tumor neovascularization.

Genes Dev. 2007, 21, 1546–1558.

[127] Lyden, D., Hattori, K., Dias, S., Costa, C., et al., Impaired recruitment of bone-marrow-derived endothelial and hematopoietic precursor cells blocks tumor angiogenesis and growth. *Nat. Med.* 2001, 7, 1194–1201.

[128] Cao, Z., Shang, B., Zhang, G., Miele, L., et al., Tumor cell-mediated neovascularization and lymphangiogenesis contrive tumor progression and cancer metastasis. *Biochim. Biophys. Acta* 2013, 1836, 273–86.

[129] Orimo, A., Weinberg, R. a., Stromal fibroblasts in cancer: A novel tumor-promoting cell type. *Cell Cycle* 2006, 5, 1597–1601.

[130] Madar, S., Goldstein, I., Rotter, V., "Cancer associated fibroblasts" - more than meets the eye. *Trends Mol. Med.* 2013, 19, 447–453.

[131] Olumi, A.F., Grossfeld, G.D., Hayward, S.W., Carroll, P.R., et al., Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* 1999, 59, 5002–11.

[132] Dvorak, H.F., Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N. Engl. J. Med.* 1986, 315, 1650–9.

[133] Dvorak, H.F., Tumors: Wounds That Do Not Heal--Redux. *Cancer Immunol. Res.* 2015, 3, 1–11.

[134] Bhowmick, N. a, Neilson, E.G., Moses, H.L., Stromal fibroblasts in cancer initiation and progression. *Nature* 2004, 432, 332–7.

[135] Wang, Y., Chen, X., Cao, W., Shi, Y., Plasticity of mesenchymal stem cells in immunomodulation: pathological and therapeutic implications. *Nat. Immunol.* 2014, 15, 1009–1016.

[136] Bergfeld, S. a, DeClerck, Y. a, Bone marrow-derived mesenchymal stem cells and the tumor microenvironment. *Cancer Metastasis Rev.* 2010, 29, 249–61.

[137] Ridge, S.M., Sullivan, F.J., Glynn, S.A., Mesenchymal stem cells: Key players in cancer progression. *Mol. Cancer* 2017, 16, 1–10.

[138] Jiang, Y., Jahagirdar, B.N., Reinhardt, R.L., Schwartz, R.E., et al., Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* 2002, 418, 41–49.

[139] Crisan, M., Yap, S., Casteilla, L., Chen, C.-W., et al., A perivascular origin for mesenchymal stem cells in multiple human organs. *Cell Stem Cell* 2008, 3, 301–313.

[140] Rajantie, I., Ilmonen, M., Alminaite, A., Ozerdem, U., et al., Adult bone marrow-derived cells recruited during angiogenesis comprise precursors for periendothelial vascular mural cells. *Blood* 2004, 104, 2084–2086.

[141] Spaeth, E.L., Dembinski, J.L., Sasser, a. K., Watson, K., et al., Mesenchymal stem cell transition to tumor-associated fibroblasts contributes to fibrovascular network expansion and tumor progression. *PLoS One* 2009, 4.

[142] Direkze, N.C., Hodivala-Dilke, K., Jeffery, R., Hunt, T., et al., Bone marrow contribution to tumor-associated myofibroblasts and fibroblasts. *Cancer Res.* 2004, 64, 8492–8495.

[143] Quante, M., Tu, S.P., Tomita, H., Gonda, T., et al., Bone Marrow-Derived Myofibroblasts Contribute to the Mesenchymal Stem Cell Niche and Promote Tumor Growth. *Cancer Cell* 2011, 19, 257–272.

[144] Spaeth, E., Klopp, A., Dembinski, J., Andreeff, M., Marini, F., Inflammation and tumor microenvironments: defining the migratory itinerary of mesenchymal stem cells. *Gene Ther.* 2008, 15, 730–738.

[145] Lourenco, S., Teixeira, V.H., Kalber, T., Jose, R.J., et al., Macrophage migration inhibitory factor-CXCR4 is the dominant chemotactic axis in human mesenchymal stem cell recruitment to tumors. *J. Immunol.* 2015, 194, 3463–74.

[146] Jung, Y., Kim, J.K., Shiozawa, Y., Wang, J., et al., Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nat. Commun.* 2013, 4, 1795.

[147] Kidd, S., Spaeth, E., Watson, K., Burks, J., et al., Origins of the tumor microenvironment: Quantitative assessment of adipose-derived and bone marrow-derived stroma. *PLoS One* 2012, 7.

[148] Gajewski, T.F., Schreiber, H., Fu, Y.-X., Innate and adaptive immune cells in the tumor microenvironment. *Nat. Immunol.* 2013, 14, 1014–1022.

[149] Quail, D.F., Joyce, J.A., The Microenvironmental Landscape of Brain Tumors. *Cancer Cell* 2017, 31, 326–341.

[150] Wesch, D., Peters, C., Siegers, G.M., Human gamma delta T regulatory cells in cancer: Fact or fiction? *Front. Immunol.* 2014, 5, 1–7.

[151] Pyonteck, S.M., Akkari, L., Schuhmacher, A.J., Bowman, R.L., et al., CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat. Med.* 2013, 19, 1264–1272.

[152] Quail, D.F., Bowman, R.L., Akkari, L., Quick, M.L., et al., The tumor microenvironment underlies acquired resistance to CSF-1R inhibition in gliomas. *Science (80-.).* 2016, 352.

[153] Costa-Silva, B., Aiello, N.M., Ocean, A.J., Singh, S., et al., Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. *Nat Cell Biol* 2015, 17, 816–826.

[154] Schaaij-Visser, T.B.M., De Wit, M., Lam, S.W., Jiménez, C.R., The cancer secretome, current status and opportunities in the lung, breast and colorectal cancer context. *Biochim. Biophys. Acta - Proteins Proteomics* 2013, 1834, 2242–2258.

[155] Raman, D., Baugher, P.J., Thu, Y.M., Richmond, A., Role of chemokines in tumor growth. *Cancer Lett.* 2007, 256, 137–165.

[156] Kojima, Y., Acar, A., Eaton, E.N., Mellody, K.T., et al., Autocrine TGF- and stromal cell-derived factor-1 (SDF-1) signaling drives the evolution of tumor-promoting mammary stromal myofibroblasts. *Proc. Natl. Acad. Sci.* 2010, 107, 20009–20014.

[157] Johnson, D.E., O'Keefe, R.A., Grandis, J.R., Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat. Rev. Clin. Oncol.* 2018, 15, 234–248.

[158] Salgado, R., Junius, S., Benoy, I., Van Dam, P., et al., Circulating interleukin-6 predicts survival in patients with metastatic breast cancer. *Int. J. Cancer* 2003, 103, 642–6.

[159] Marotta, L.L.C., Almendro, V., Marusyk, A., Shipitsin, M., et al., The JAK2 / STAT3 signaling pathway is required for growth of CD44 + CD24 – stem cell – like breast cancer cells in human tumors. *J. Clin. Invest.* 2011, 121, 2723–2735.

[160] Folkman, J., Angiogenesis: an organizing principle for drug discovery? *Nat. Rev. Drug Discov.* 2007, 6, 273–86.

[161] Mitamura, T., Pradeep, S., McGuire, M., Wu, S.Y., et al., Induction of anti-VEGF therapy resistance by upregulated expression of microseminoprotein (MSMP). *Oncogene* 2017, 1–10.

[162] Bonnans, C., Chou, J., Werb, Z., Remodelling the extracellular matrix in development and disease. *Nat. Rev. Mol. Cell Biol.* 2014, 15, 786–801.

[163] Lu, P., Weaver, V.M., Werb, Z., The extracellular matrix: A dynamic niche in cancer progression. *J. Cell Biol.* 2012, 196, 395–406.

[164] Naba, A., Clauser, K.R., Hoersch, S., Liu, H., et al., The matrisome: in silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Mol. Cell. Proteomics* 2012, 11, M111.014647.

[165] Naba, A., Pearce, O.M.T., Del Rosario, A., Ma, D., et al., Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics. *J. Proteome Res.* 2017, 16, 3083–3091.

[166] Hynes, R.O., Naba, A., Overview of the matrisome-An inventory of extracellular matrix constituents and functions. *Cold Spring Harb. Perspect. Biol.* 2012, 4.

[167] Chen, P., Cescon, M., Bonaldo, P., Collagen VI in cancer and its biological mechanisms. *Trends Mol. Med.* 2013, 19, 410–417.

[168] Park, J., Scherer, P.E., Adipocyte-derived endotrophin promotes malignant tumor progression. *J. Clin. Invest.* 2012, 122, 4243–4256.

[169] Robertson, I.B., Horiguchi, M., Zilberberg, L., Dabovic, B., et al., Latent TGF-β-binding proteins. *Matrix Biol.* 2015, 47, 44–53.

[170] Baxter, R.C., IGF binding proteins in cancer: Mechanistic and clinical insights. *Nat. Rev. Cancer* 2014, 14, 329–341.

[171] Kappelhoff, R., Puente, X.S., Wilson, C.H., Seth, A., et al., Overview of transcriptomic analysis of all human proteases, non-proteolytic homologs and inhibitors: Organ, tissue and ovarian cancer cell line expression profiling of the human protease degradome by the CLIP-CHIP™ DNA microarray. *Biochim. Biophys. Acta - Mol. Cell Res.* 2017, 1864, 2210–2219.

[172] Seidah, N.G., Prat, A., The biology and therapeutic targeting of the proprotein convertases. *Nat. Rev. Drug Discov.* 2012, 11, 367–383.

[173] Jaaks, P., Bernasconi, M., The proprotein convertase furin in tumour progression. *Int. J. cancer* 2017, 141, 654–663.

[174] Lapierre, M., Siegfried, G., Scamuffa, N., Bontemps, Y., et al., Opposing function of the proprotein convertases furin and PACE4 on breast cancer cells' malignant phenotypes: Role of tissue inhibitors of metalloproteinase-1. *Cancer Res.* 2007, 67, 9030–9034.

[175] Kessenbrock, K., Plaks, V., Werb, Z., Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell* 2010, 141, 52–67.

[176] Prassas, I., Eissa, A., Poda, G., Diamandis, E.P., Unleashing the therapeutic potential of human kallikrein-related serine proteases. *Nat. Rev. Drug Discov.* 2015, 14, 183–202.

[177] Sutkowski, D.M., Goode, R.L., Baniel, J., Teater, C., et al., Growth regulation of prostatic stromal cells by prostate-specific antigen. *J. Natl. Cancer Inst.* 1999, 91, 1663–1669.

[178] Filippou, P.S., Karagiannis, G.S., Musrap, N., Diamandis, E.P., Kallikrein-related peptidases (KLKs) and the hallmarks of cancer. *Crit. Rev. Clin. Lab. Sci.* 2016, 53, 277–291.

[179] Olson, O.C., Joyce, J.A., Cysteine cathepsin proteases : regulators of cancer progression and therapeutic response. *Nat. Publ. Gr.* 2015, 15, 712–729.

[180] Sobotič, B., Vizovišek, M., Vidmar, R., Van Damme, P., et al., Proteomic Identification of Cysteine Cathepsin Substrates Shed from the Surface of Cancer Cells. *Mol. Cell. Proteomics* 2015, 14, 2213–2228.

[181] Scorilas, A., Fotiou, S., Tsiambas, E., Yotis, J., et al., Determination of cathepsin B expression may offer additional prognostic information for ovarian cancer patients. *Biol. Chem.* 2002, 383, 1297–1303.

[182] Foekens, J.A., Kos, J., Peters, H.A., Krasovec, M., et al., Prognostic significance of cathepsins B and L in primary human breast cancer. *J. Clin. Oncol.* 1998, 16, 1013–21.

[183] Prudova, A., Gocheva, V., auf dem Keller, U., Eckhard, U., et al., TAILS N-Terminomics and Proteomics Show Protein Degradation Dominates over Proteolytic Processing by Cathepsins in Pancreatic Tumors. *Cell Rep.* 2016, 16, 1762–1773.

[184] Villarreal, L., Méndez, O., Salvans, C., Gregori, J., et al., Unconventional secretion is a major contributor of cancer cell line secretomes. *Mol. Cell. Proteomics* 2013, 12, 1046–60.

[185] Raposo, G., Stoorvogel, W., Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell Biol.* 2013, 200, 373–383.

[186] Kreimer, S., Belov, A.M., Ghiran, I., Murthy, S.K., et al., Mass-spectrometry-based molecular characterization of extracellular vesicles: Lipidomics and proteomics. *J. Proteome Res.* 2015, 14, 2367–2384.

[187] Al-Nedawi, K., Meehan, B., Micallef, J., Lhotak, V., et al., Intercellular transfer of the

oncogenic receptor EGFRvIII by microvesicles derived from tumour cells. *Nat. Cell Biol.* 2008, 10, 619–624.

[188] Harel, M., Oren-Giladi, P., Kaidar-Person, O., Shaked, Y., Geiger, T., Proteomics of microparticles with SILAC Quantification (PROMIS-Quan): a novel proteomic method for plasma biomarker quantification. *Mol. Cell. Proteomics* 2015, 14, 1127–36.

[189] Yamada, K.M., Cukierman, E., Modeling Tissue Morphogenesis and Cancer in 3D. *Cell* 2007, 130, 601–610.

[190] Sandberg, R., Ernberg, I., Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). *Proc. Natl. Acad. Sci.* 2005, 102, 2052–2057.

[191] Ertel, A., Verghese, A., Byers, S.W., Ochs, M., Tozeren, A., Pathway-specific differences between tumor cell lines and normal and tumor tissue cells. *Mol. Cancer* 2006, 5.

[192] Shoemaker, R.H., The NCI60 human tumour cell line anticancer drug screen. *Nat. Rev.* 2006, 6, 813–823.

[193] Gillet, J., Maria, A., Varma, S., Marino, M., et al., Redefining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. *Proc. Natl. Acad. Sci. U. S. A.* 2011, 108, 18708–18713.

[194] van Staveren, W.C.G., Solis, D.Y., Hebrant, A., Detours, V., et al., Human cancer cell lines: Experimental models for cancer cells in situ? For cancer stem cells? *Biochim. Biophys. Acta* 2009, 1795, 92–103.

[195] Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., et al., Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 2013, 4.

[196] Vincent, K.M., Findlay, S.D., Postovit, L.M., Assessing breast cancer cell lines as tumour models by comparison of mRNA expression profiles. *Breast Cancer Res.* 2015, 17, 1–12.

[197] Quail, D.F., Maciel, T.J., Rogers, K., Postovit, L.M., A unique 3D in vitro cellular invasion assay. *J. Biomol. Screen.* 2012, 17, 1088–1095.

[198] Gauthier, N.P., Soufi, B., Walkowicz, W.E., Pedicord, V.A., et al., Cell-selective labeling using amino acid precursors for proteomic studies of multicellular environments. *Nat. Methods* 2013, 10, 768–773.

[199] Ivascu, A., Kubbies, M., Rapid generation of single-tumor spheroids for high-throughput cell function and toxicity analysis. *J. Biomol. Screen. Off. J. Soc. Biomol. Screen.* 2006, 11, 922–932.

[200] Hughes, C.S., Postovit, L.M., Lajoie, G.A., Matrigel: a complex protein mixture required for optimal growth of cell culture. *Proteomics* 2010, 10, 1886–1890.

[201] Hughes, C., Radan, L., Chang, W.Y., Stanford, W.L., et al., Mass spectrometry-based proteomic analysis of the matrix microenvironment in pluripotent stem cell culture. *Mol. Cell. Proteomics* 2012, 11, 1924–36.

[202] Tentler, J.J., Tan, A.C., Weekes, C.D., Jimeno, A., et al., Patient-derived tumour xenografts as models for oncology drug development. *Nat. Rev. Clin. Oncol.* 2012, 9, 338–350.

[203] Cassidy, J.W., Caldas, C., Bruna, A., Maintaining tumor heterogeneity in patient-derived tumor xenografts. *Cancer Res.* 2015, 75, 2963–2968.

[204] Ben-David, U., Ha, G., Tseng, Y.Y., Greenwald, N.F., et al., Patient-derived xenografts undergo mouse-specific tumor evolution. *Nat. Genet.* 2017, 49, 1567–1575.

[205] Ciriello, G., Cerami, E., Sander, C., Schultz, N., Mutual exclusivity analysis identifies oncogenic network modules. *Genome Res.* 2012, 22, 398–406.

[206] Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., et al., Global quantification of mammalian gene expression control. *Nature* 2011, 473, 337–342.

[207] Geyer, P.E., Holdt, L.M., Teupser, D., Mann, M., Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* 2017, 13, 942.

[208] Tighe, P.J., Ryder, R.R., Todd, I., Fairclough, L.C., ELISA in the multiplex era: Potentials and pitfalls. *Proteomics - Clin. Appl.* 2015, 9, 406–422.

[209] Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., et al., An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* 2017, 4, 587–599.e4.

[210] Keshishian, H., Burgess, M.W., Gillette, M.A., Mertins, P., et al., Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol. Cell. Proteomics* 2015, 14, 2375–2393.

[211] Kuhlmann, L., Cummins, E., Samudio, I., Kislinger, T., Cell-surface proteomics for the identification of novel therapeutic targets in cancer. *Expert Rev. Proteomics* 2018, 0, 14789450.2018.1429924.

[212] Humphrey, S.J., Azimifar, S.B., Mann, M., High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat. Biotechnol.* 2015, 33, 990–995.

[213] Li, P., Kaslan, M., Lee, S.H., Yao, J., Gao, Z., Progress in Exosome Isolation Techniques 2017, 7, 1–35.

[214] Ishihama, Y., Rappsilber, J., Mann, M., Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J. Proteome Res.* 2006, 5, 988–94.

[215] Kulak, N. a, Pichler, G., Paron, I., Nagaraj, N., Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimationa in eukaryotic cells. *Nat. Methods* 2014, 11, 319–24.

[216] Batth, T.S., Francavilla, C., Olsen, J. V, Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* 2014, 13, 6176–6186.

[217] Hughes, C.S., Foehr, S., Garfield, D.A., Furlong, E.E., et al., Ultrasensitive proteome

analysis using paramagnetic bead technology. *Mol. Syst. Biol.* 2014, 10, 757.

[218] Hughes, C., Krijgsveld, J., Developments in quantitative mass spectrometry for the analysis of proteome dynamics. *Trends Biotechnol.* 2012, 30, 668–676.

[219] Cox, J., Hein, M.Y., Luber, C. a, Paron, I., et al., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 2014, 13, 2513–26.

[220] Al Shweiki, M.H.D.R., Mönchgesang, S., Majovsky, P., Thieme, D., et al., Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. *J. Proteome Res.* 2017, 16, 1410–1424.

[221] Mann, M., Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* 2006, 7, 952–958.

[222] Pozniak, Y., Balint-Lahat, N., Rudolph, J.D., Lindskog, C., et al., System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. *Cell Syst.* 2016, 2, 172–184.

[223] Tape, C.J., Ling, S., Dimitriadi, M., McMahon, K.M., et al., Oncogenic KRAS Regulates Tumor Cell Signaling via Stromal Reciprocation. *Cell* 2015, 165, 910–920.

[224] Bendall, S.C., Simonds, E.F., Qiu, P., Amir, E.D., et al., Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 2011, 332, 687–96.

[225] Rauniyar, N., Yates, J.R., Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.* 2014, 13, 5293–309.

[226] Frost, D.C., Greer, T., Li, L., High-resolution enabled 12-plex DiLeu isobaric tags for quantitative proteomics. *Anal. Chem.* 2015, 87, 1646–1654.

[227] McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., et al., MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* 2014, 86, 7150–8.

[228] Ting, L., Rad, R., Gygi, S.P., Haas, W., MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat. Methods* 2011, 8, 937–940.

[229] Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., Coon, J.J., Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* 2012, 11, 1475–1488.

Chapter 2

Comparison of sample preparation techniques for large scale proteomics

2.1 Abstract

Numerous workflows exist for large scale bottom-up proteomics, many of which achieve exceptional proteome depth. Herein, we evaluated the performance of several commonly used sample preparation techniques for proteomic characterization of HeLa lysates (unfractionated in-solution digests, SDS-PAGE coupled with in-gel digestion, gel-eluted liquid fraction entrapment electrophoresis (GELFrEE) technology, SCX StageTips and High-/Low-pH reversed phase fractionation (HpH)). HpH fractionation was found to be superior in terms of proteome depth (>8400 proteins detected) and fractionation efficiency compared to other techniques. SCX StageTip fractionation required minimal sample handling and was also a substantial improvement over SDS-PAGE separation and GELFrEE technology. Sequence coverage of the HeLa proteome increased to 38% when combining all workflows however total proteins detected improved only slightly to 8710. In summary, HpH fractionation and SCX StageTips are robust techniques and highly suited for complex proteome analysis.

2.2   Introduction

Sample preparation strategies employed in bottom-up proteomics can be broadly categorized into workflows which omit or implement fractionation at the protein and/or peptide level prior to LC-MS [1]. Prior limitations in MS instrumentation and HPLC technology rendered unfractionated approaches insufficient when characterizing complex proteomes and thus necessitated the need for extensive fractionation [2]. Advances in instrumentation speed and sensitivity coupled with UPLC systems utilizing longer columns and smaller particles sizes has substantially improved proteome depth of unfractionated samples. This is exemplified by Nagaraj and colleagues and Pirmoradian *et al.*, who detected over 3900 yeast and 4800 HeLa proteins, respectively, by employing long gradients and 50 cm reversed phase columns coupled to a high resolution Q Exactive mass spectrometer [3,4]. In fact, it is now possible to detect ~4000 yeast and ~4400 HeLa proteins in approximately 1h with current state-of-the-art Orbitrap-based mass spectrometers (Orbitrap Fusion and Q Exactive HF) [5,6]. In contrast, fractionation approaches based on molecular weight (MW), charge, pI, or hydrophobicity require substantially more acquisition time and sample handling [7]. Nonetheless, these workflows contribute to our understanding of biological systems by characterizing PTMs, enriching for low abundance species and quantifying expression for thousands of proteins [8–11].

SDS-PAGE protein separation coupled with in-gel digestion has been widely used in mass spectrometry-based proteomic studies [12,13]. Its robustness, low cost, high resolution and ability to handle detergent containing samples make it amenable to many workflows. In fact, two recently published drafts of the human proteome utilized SDS-PAGE to obtain unprecedented protein expression profiles for multiple tissues (peptide evidence from 84-92% of protein coding genes in the human proteome) [14,15]. However, SDS-PAGE fractionation is time consuming, manually intensive and subject to variable peptide extraction efficiencies [12].

Newer fractionation strategies, like Gel Eluted Liquid Fractionation Entrapment Electrophoresis (GELFrEE) technology, can overcome most limitations associated with SDS-PAGE [16]. Exceptional proteome coverage has been reported with peptide based fractionation using MudPIT [17], IEF [18,19], and High-/Low-pH reversed phase chromatography (HpH) [6,20]. Alternatively, small scale peptide fractionation (C18, SCX, SAX) can be readily performed in StageTips without requiring dedicated fractionation equipment [21–23].

Many large scale proteomic studies employ methodologies such as those listed but few have compared their performance relative to one another [24–26]. In light of this, we compared HeLa proteomes obtained from unfractionated in-solution, SDS-PAGE, GELFrEE, SCX StageTip and HpH sample preparations across several parameters. When controlling for sample loading and LC-MS time, most workflows performed well, however, HpH fractionation provided the greatest proteome coverage.

2.3   Results

2.3.1   Proteome coverage of different workflows on a Q Exactive mass spectrometer

HeLa proteomes obtained from unfractionated in-solution, SDS-PAGE, GELFrEE, SCX StageTip and High-/Low-pH pH reversed phase (HpH) sample preparations were systematically compared (Figure 2.1). Where possible, experimental parameters were kept constant between workflows. To minimize sample handling and processing, all techniques, with the exception of SDS-PAGE, utilized chloroform methanol precipitation followed by in-solution digestion. In total, 10 fractions were analyzed by LC-MS (~1µg per fraction) for each technique from 3 biological replicates (3 different passages). Unfractionated in-solution samples were injected once per replicate. In our hands, long LC-MS gradients generate more protein identifications compared to shorter gradients when analyzing cellular

lysates (data not shown). Therefore, all fractions were run on a 4h gradient plus washing and re-equilibration.

Replicates were searched both individually and grouped using the match between runs feature in MaxQuant. Where applicable, protein identifications containing ≥1 unique peptide(s) in 2 out of 3 biological replicates were used for analysis [27]. In general, each method tested received more instrument time compared to previous reports and when necessary, utilized chloroform/methanol precipitation to remove SDS coupled with on-pellet, in solution digestion.

Within each replicate from different workflows, ~7.1% of the HeLa proteins identified were "matched-between-runs" and >95% of the proteins were present in all biological replicates suggesting high reproducibility for each technique (data not shown). Moreover, only a small fraction of proteins were identified by one unique peptide (Figure 2.2A). In total, 5189, 6959, 5919, 7655, 8470 proteins were detected with in-solution, SDS-PAGE, GELFrEE, SCX and HpH workflows, respectively (Table 2.1).

In terms of proteomic depth, nearly 5200 proteins were detected with unfractionated in-solution HeLa digests, which is on par with previous reports [4,27]. Recently, Beck *et al.* detected ~5200 HeLa proteins from triplicate, 90 minute runs on a Bruker Impact II™ Q-TOF [28]. For gel-based techniques, proteome coverage with SDS-PAGE was also comparable to reports by other groups [29,30]. While Botelho *et al.* previously demonstrated similar performance between GELFrEE and SDS-PAGE using an LTQ ion trap, we detected ~1,000 more proteins with SDS-PAGE fractionation compared to GELFrEE separation [26].

Regarding peptide fractionation approaches, we identified a respectable number of proteins using SCX StageTips but our numbers were less than that obtained by Kulak *et al.* using a similar technique (7655 versus 9667) [22]. Differences in sample preparation (lysis buffer and digestion) as well as column size may be contributing factors. Protein identifications with HpH was similar to *Kelstrup et al.* (8470 versus 8400 protein IDs) and the highest out of all workflows tested [6]. Even greater proteome depth has been achieved with HpH fractionation by Beck *et al.* who reported >11,000 proteins from a more complex mouse cerebral tissue [28].

Several proteomic studies have documented increased proteome depth by combining data from multiple workflows and/or instrumentation [31–33]. Combining all our datasets yielded >8700 unique proteins. This represents a small increase of ~3% over HpH fractionation, which had nearly 700 exclusive proteins alone (Figure 2.2B).
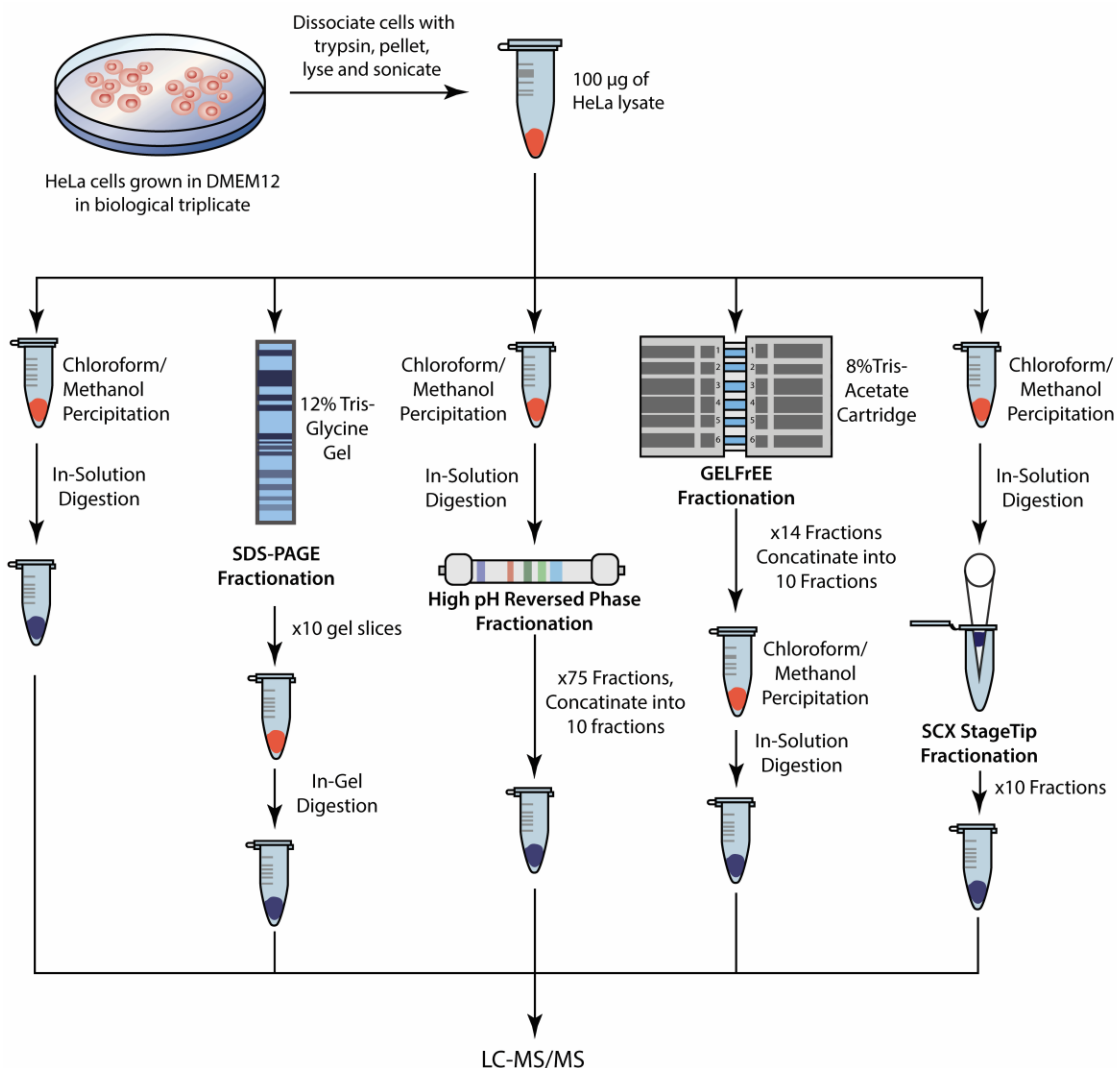
Figure 2.1 Schematic of workflows tested.

Hela lysates (100µg) were separated by SDS-PAGE and GELFrEE or subjected chloroform/methanol precipitation and on-pellet, in-solution digestion prior to High pH reversed phase and SCX StageTip fractionation. GELFrEE fractions were also chloroform/methanol precipitated to remove SDS. Three independent biological replicates were analyzed by LC-MS on the Q Exactive.

Table 2.1 Total HeLa proteins detected on a Q Exactive for each technique

Biological Rep.	In-Solution	SDS-PAGE	GELFrEE	SCX	HpH
1	5144	6835	5868	7571	8394
2	5150	6934	5877	7595	8439
3	5127	6967	5894	7665	8444
[a] Total	5189	6959	5919	7655	8470

[a] Total number of proteins that were identified in 2 out of 3 biological replicates, with minimum one unique peptide, after removal of common contaminants.
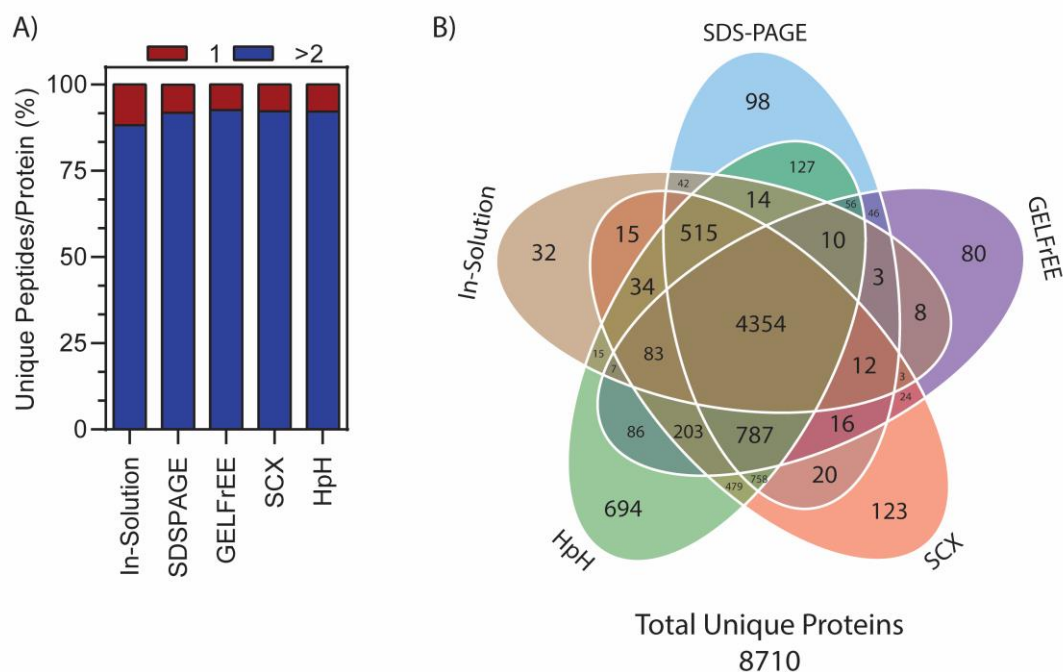
Figure 2.2 High confidence identifications and large overlap between proteomes from different techniques.

(A) Majority of proteins (~95%) were identified with high confidence (2 or more unique peptides; blue) for each technique. Only a small number (~5%) were identified by a single peptide (red). (B) Total proteins exclusive (11%) and common (50%) to 5 different preparative techniques analyzed on a Q Exactive. Gene symbols were used for analysis and proteins exclusive to one biological replicate were omitted.

2.3.2   *Comparison of fractionation efficiency*

The capacity of pre-fractionation to resolve unique proteins or peptides into discrete packets, reduces sample complexity and improves peptide detection and identification by MS [34]. To assess the fractionation efficiency for each technique, we examined how many unique peptides, and proteins where applicable, were exclusive to 1, 2 or ≥3 fractions. For this analysis, biological replicates were searched individually without the "match-between-runs" feature in MaxQuant.

In principle, SDS-PAGE displays good protein separation and resolution over a wide range of MWs, generally within a few kDa. Surprisingly, we found with SDS-PAGE that only 27.7% and 59.5% of proteins identified were exclusive to 1 or 2 fractions, respectively (Figure 3A). However, at the peptide level, 58.2% and 82.8% were exclusive to 1 or 2 fractions, respectively (Figure 2.3C). Of note, the GELFrEE protocol was less efficient at separating proteins than SDS-PAGE (Figure 3B, D). Silver stained GELFrEE fractions ran on 1D SDS-PAGE revealed moderate overlap between adjacent lanes (Figure 4A). Moreover, Box-and-Whisker plots of median fraction MW further illustrate limited separation with GELFrEE compared to SDS-PAGE (Figure 2.4C). Peptide fractionation efficiency with SCX StageTips was similar to Kulak *et al.* with 53.8% and 77.4% exclusive to 1 or 2 SCX fractions, respectively (Figure 2E) [22]. However, the HpH method gave the best fractionation performance with 80.1% and 94.5% of all peptides exclusive to 1 or 2 fractions, respectively (Figure 2.3F).
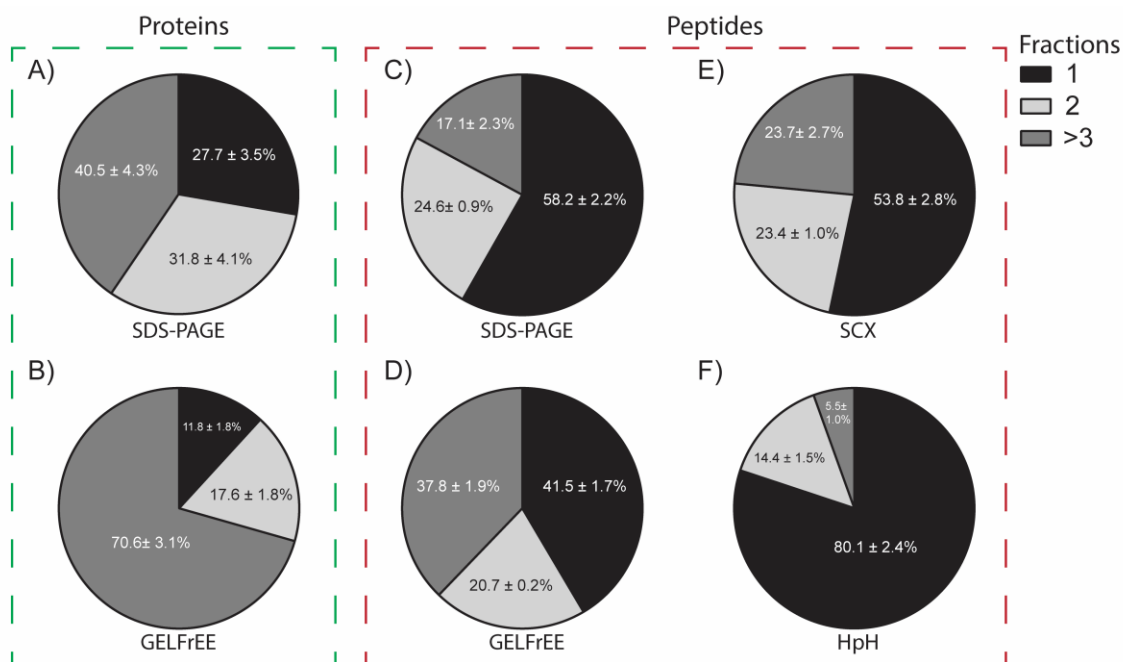
Figure 2.3 Fractionation efficiency varies between protein and peptide separation techniques.

Pie chart displaying percentage of unique proteins and peptides (mean +/- SD) exclusive to one (black), two (light grey) and three or more fractions (grey) for SDS-PAGE (A,C), GELFrEE (B,D), SCX (E) and HpH (F). SDS-PAGE and HpH exhibited the greatest fractionation efficiency for protein and peptide based separation techniques, respectively.
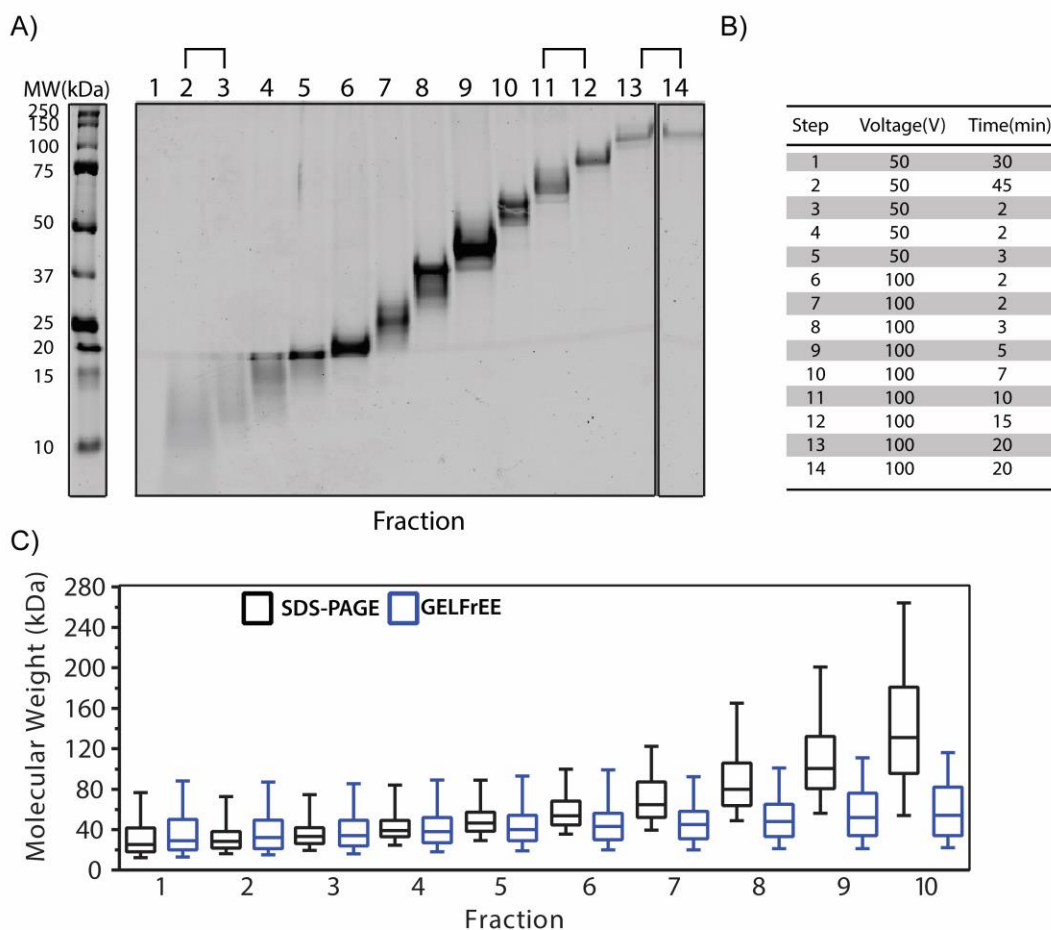
Figure 2.4 Overview of GELFrEE separation

(A) Silver stained 1D SDS-PAGE of fractions collected on an 8% Tris-acetate GELFrEE cartridge starting with 100ug of HeLa lysate. Fraction 14 was run on the same gel. (B) Gradient was slightly modified to allow leading dye front to be eluted in the first fraction. Fraction 1 was discarded and fractions 2 and 3, 11 and 12, and 13 and 14 were combined to produce a total of 10 fractions. (C) Box-and-Whisker plot of median protein MW detected in each fraction for SDS-PAGE (black) and GELFrEE (blue). Boxes represent 75% and 25% percentiles and Whiskers indicate 90% and 10% percentiles.

2.3.3   Distribution of proteins and peptides

Fractionation techniques which exhibit orthogonal separation should be more efficient at maximizing MS/MS time across the entire gradient space [34]. For each method, the distribution of peptides and proteins across all fractions varied. We did not observe any trends between gel- and peptide-based fractionation approaches, however, changes in unique peptides and proteins per fraction consistently mirrored each other within each technique. For example, total peptides and proteins increased slightly with MW for SDS-PAGE separation before declining in later fractions (Figure 2.5A). SCX displayed a sharp increase in peptides/proteins detected in early fractions before both plateaued (Figure 2.5C). Changes between fractions with GELFrEE and HpH fractionation were relatively smaller (Figure 5B and D). Peptide density versus retention time plots also illustrate a similar trend (Figure 2.6). For example, SDS-PAGE exhibited a higher peptide density with increasing fraction number (MW) while HpH remained even throughout, most likely due to its concatenation scheme.

We next examined peptides:protein ratios for each technique and found gel-based methods had approximately 1 less peptide identified per protein compared to peptide fractionation even though gel-based approaches generated roughly 300K more MS2 scans (Table 2.2). As expected, in-solution digests had the lowest peptides:protein ratio (10.1:1) but identified the greatest proteins/hour (1298/hour). While total unique peptides was indicative of proteins identified for each technique, this was not the case with PSMs. For example, GELFrEE had the greatest number of PSMs but the fewest unique peptides and proteins.
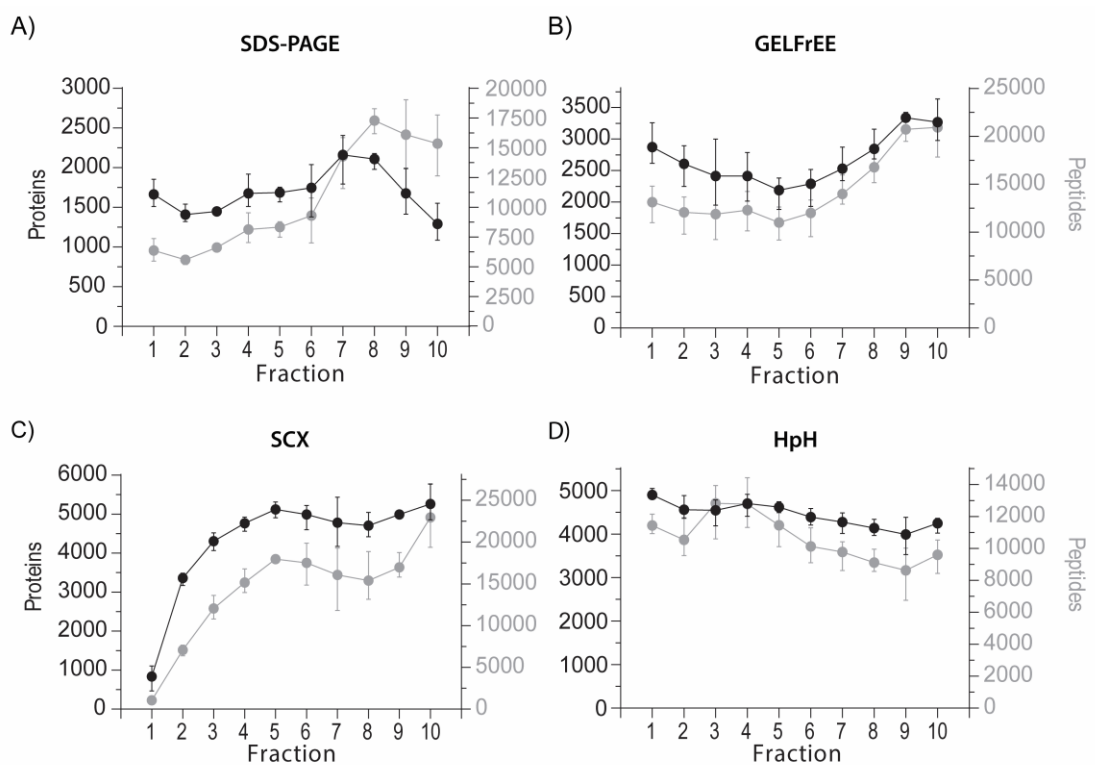
Figure 2.5 Peptide and protein distribution profiles deviate for each technique.

(A-D) Distributions of unique peptides (right y-axis, grey) and proteins (left y-axis, black) identified per fraction for SDS-PAGE, GELFrEE, SCXand HpH, respectively.
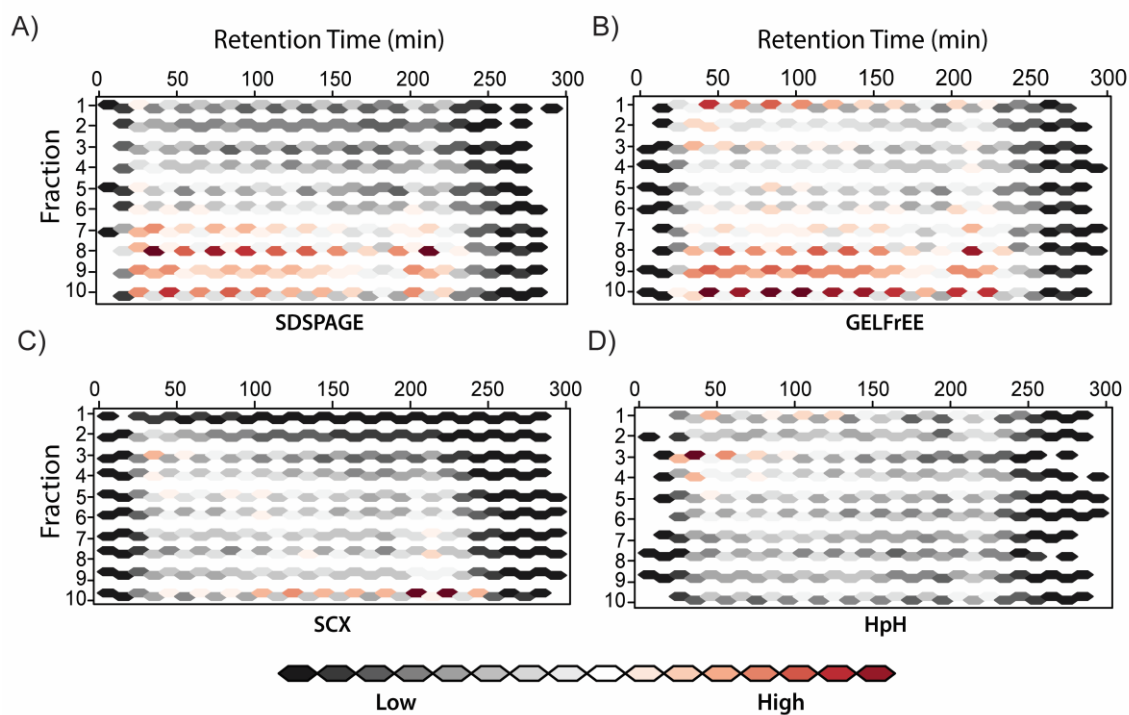
Figure 2.6 HpH fractionation is highly orthogonal compared to other methods.

Peptide density distribution was assessed using the hexbin package in R, each hexagon represents 500 peptides with red indicating the highest density. HpH fractionation yields a more even distribution of peptides throughout the entire gradient and across all fractions compared with gel- and SCX-based fractionation which exhibit more extreme (low and high) peptide densities.

Table 2.2 Total MS/MS scans, peptides and proteins detected with each technique

Method	In-Solution	SDS-PAGE	GELFrEE	SCX	HpH	Combined
MS/MS Scans	318K	2.96M	2.90M	2.55M	2.68M	11.4M
PSMs	182K	557K	830K	769K	653K	2.97M
Unique Peptides	52.5K	85.9K	71.7K	104K	111K	165K
Proteins	5189	6959	5919	7655	8470	8710
Pep./Pro. Ratio	10.1	12.3	12.1	13.5	13.1	18.9
Gradient Time [a]	4.0h	40h	40h	40h	40h	492h
Avg. Proteins/hrs [b]	1298	173	148	191	212	n/a

[a] Total amount of time per LC MS gradient excluding washing and equilibration.
[b] Average proteins per hour identified for one biological replicate,.

2.3.4   Evaluating peptide characteristics

Examining the median sequence coverage achieved for each method revealed a peak in peptide density between 10% and 15% before tailing off (Figure 2.7A). For gel-based workflows, median sequence coverage with GELFrEE (~27%) was higher than SDS-PAGE (22.6%) (Figure 2.7B). SCX and HpH peptide fractionation improved median sequence coverage to ~27% and 24.4%, respectively, over the unfractionated in-solution digest (22.8%) (Figure 2.7B). Combining sequence information from all methods improved median sequence coverage of all HeLa proteins identified to 38.0%. This can be attributed to a 48.6% increase in total unique peptides (165K) over HpH, which had the second highest number of unique peptides (111K) (Table 2.3). For comparison, Kelstrup *et al.* achieved a median sequence coverage of >40% with HeLa digests fractionated by HpH (14 fractions) using the latest generation Q Exactive HF [6].

Next, we calculated GRAVY scores for unique peptides detected by each workflow to determine whether any bias towards hydrophobic or hydrophilic species existed (Figure 2.9A) [35]. All methods displayed a propensity to enrich for hydrophilic peptides as indicated by negative GRAVY scores. Dunn's multiple comparison, post hoc analysis revealed a significantly higher (P<0.001) median GRAVY score with unfractionated in-solution digests compared to all other techniques (Figure 2.9B). These findings are in line with previous groups which found cellular digests to be primarily hydrophilic and also suggests a proportion of hydrophobic peptides are lost during sample handling [36,37].

Lastly, analysis of missed cleavages revealed that a large number were present in most sample preparations although many (~70%) were restricted to 1 site (Table 2.3). As expected, SDS-PAGE was the highest (44.5%) which may be explained by poor absorption and diffusion of trypsin into the gel pieces. Missed cleavages with GELFrEE was relatively high (34.8%) compared to unfractionated in-solution

samples (25.6%) even though digestion was performed essentially the same for both techniques. SCX fractionation exhibited the least missed cleavages (15.9%) followed by HpH (22.1%).
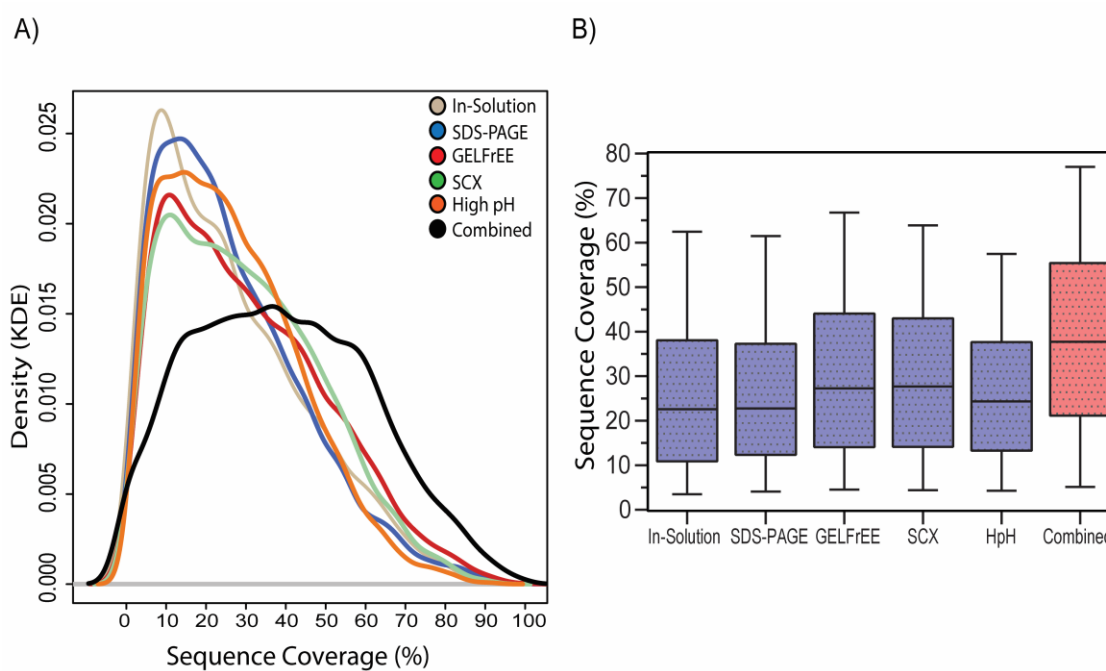
Figure 2.7 Different fractionation techniques provide complementary sequence coverage

(A) Kernel density estimation comparing percent sequence coverage for each technique. A slight maxima is observed near 40% (black line) when combining sequence information from all methods. (B) Box-and-Whisker plots displaying median percent sequence coverage for individual techniques (blue). Box and whiskers indicate 75% and 25% percentiles, and 95% and 5% percentiles, respectively. One-way ANOVA was performed using the Kruskal-Wallis test to assess differences in mean sequence coverage distribution between methods. Combined data set is shown in red. All methods were significantly different from each other ($p < 0.05$), with the exception of GELFrEE and SCX.
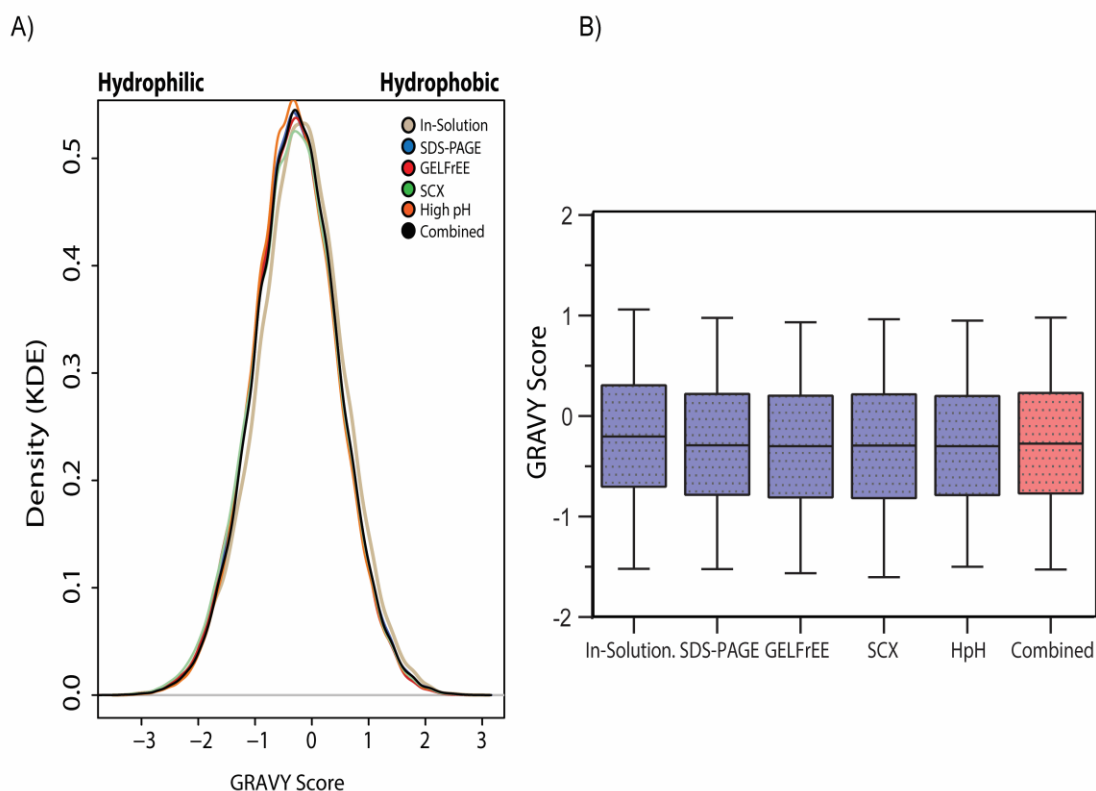
Figure 2.8 All workflows preferentially enrich for hydrophilic peptides.

(A) Kernel density estimation was performed using GRAVY scores from each method. GRAVY scores <0 indicate the presence and relative abundance of hydrophilic species. (B) Box-and-Whisker plot displaying mean GRAVY scores. Boxes represent 75% and 25% percentiles and whiskers indicate 95% and 5% percentiles. One-way ANOVA was performed using the Kruskal-Wallis test to assess differences in mean GRAVY score distribution between methods. Combined dataset is shown in red. In-solution was significantly from different from all other methods (p-value <0.0001).

Table 2.3 Total missed cleavages present in all preparative techniques

Method	1 Site (%)	2 Sites (%)	3 Sites (%)	Total (%)
In- Solution	21.9	3.59	0.11	25.6
SDS-PAGE	33.1	9.48	1.92	44.5
GELFrEE	27.9	5.82	1.08	34.8
SCX	14.2	1.51	0.19	15.9
HpH	19.6	2.25	0.25	22.1

2.4 Discussion

In this study, we compared the performance of several commonly used sample preparative techniques for bottom-up proteomics. As expected, fractionation yielded more protein identifications, and in most cases, greater sequence coverage, than unfractionated in-solution digests. Peptide-based fractionation outperformed gel-based workflows in terms of protein IDs and fractionation efficiency but not necessarily sequence coverage.

Interestingly, in our hands, we did not achieve similar proteome depth with GELFrEE compared to SDS-PAGE. GELFrEE appears to suffer from poor resolution which may be inherent to the low 8% tris-acetate cartridges and short resolving gel (1cm) required for eluting high MW proteins within a reasonable time frame. In addition, the GELFrEE collection chamber was not rinsed between cycles. Hence, carry over between fractions from residual sample in the GELFrEE collection chamber could have led to an under representation of separation. However, both GELFrEE and SDS-PAGE exhibited poor protein separation in comparison to other studies which found 64-67% of proteins identified were exclusive to a single fraction [26,29]. At the peptide level, our numbers more closely resemble these values. Therefore, we believe that proteoforms (isoforms, PTMs and cleaved/fragmented proteins) migrating at different MWs are recorded as single entries during database searching, thereby underestimating the true fractionation efficiency of both SDS-PAGE and GELFrEE. For example, Titin, a 3.6 MDa protein was detected in low, intermediate and high MW SDS-PAGE and GELFrEE fractions. Nonetheless, GELFrEE remains an invaluable tool for top-down proteomics [38].

Differences between digestion efficiency with in-gel and in-solution preparations is another factor which likely affected proteome coverage and warrants further investigation. For reference, we investigated peptides with missed cleavages from

HpH preparations which contained internal lysine and/or arginine residues. Notably, we found the frequency of internal K residues to be ~2 fold higher than R even though their abundance in the human proteome (Uniprot) is approximately even (~5.8% for K and ~5.6% for R) (data not shown). This difference can be explained by trypsin's higher affinity for arginine compared to lysine [39]. Hence, future sample preparations may benefit from utilizing Trypsin/LysC to minimize the number of missed cleavages occurring at lysine [40].

Reproducibility, feasibility/cost and throughput are important parameters to consider when choosing a sample preparation to employ in bottom-up proteomics. Although it is difficult to objectively quantify these parameters for each technique, SCX StageTip fractionation was by far the most efficient and straightforward method due to the capacity to process samples in parallel and short elution times. HpH fraction collection was automated but is limited to processing one sample at a time. In addition, HpH requires a dedicated fractionation system as well as additional time for concatenation, drying and column cleaning between replicates. GELFrEE can multiplex up to 8 samples but needs ~3 hours to run plus chloroform/methanol precipitation of each fraction. It also requires a dedicated unit and custom cartridges. SDS-PAGE, as expected, was the most labour intensive technique and required an additional day for destaining. However, SDS-PAGE as well as SCX StageTips, were the most cost-effective and accessible methods.

Although combining multiple techniques improved protein identifications and sequence coverage, the additional acquisition time needed is not feasible for the majority of medium to large scale proteomic studies (≥1 proteome/day of instrument time). It is doubtful that faster mass spectrometers with increased sensitivity and dynamic range will bypass the need for some form of sample fractionation to achieve maximum proteome coverage. Utilizing multiple enzyme digestion strategies or iterative exclusion in tandem with techniques like HpH fractionation may be more appropriate for achieving optimal sequence coverage

[10,41–43]. Additional improvements to protein extraction/handling, column technology and instrumentation could also yield increased proteome depth. In summary, the findings reported here illustrate the benefits and limitations of different techniques for analyzing a complex cellular proteome and should help aid in the design of future bottom-up proteomics studies.

2.5   Materials and Methods

2.5.1   Cell culture and protein extraction

HeLa cells (obtained from the ATCC) were maintained in DMEM F12 media supplemented with 10% FBS (Life Technologies, Grand Island, NY). Confluent 15cm plates of HeLa cells were rinsed with PBS, trypsinized and then centrifuged at 400 xg for 5 minutes to pellet cells. Cell pellets were re-suspended in PBS, pelleted again and stored at -80°C. To prepare lysates for LC-MS, frozen cell pellets were incubated in 8M Urea, 50mM ammonium bicarbonate (ABC), 10mM DTT, 2% SDS and sonicated with a probe sonicator (20 X 0.5 second pulses; Level 1) (Fisher Scientific, Waltham, MA) to shear DNA. Lysates were quantified using a Pierce™ 660nm Protein Assay (ThermoFisher Scientific) and stored at -80°C until future use.

2.5.2   Chloroform/methanol protein precipitation

HeLa lysates were reduced in 10mM DTT for 30 minutes and alkylated in 100mM Iodoacetamide (IAA) for 30 minutes at room temperature in the dark. Next, lysates were precipitated in chloroform/methanol in 1.5mL microfuge tubes according to Wessel and Flügge [44]. Briefly, 100µg aliquots of HeLa lysates were topped up to 150µL with 50mM ABC. To each sample, 600µL of cold methanol was added followed by 150µL of chloroform and thorough vortexing. A volume of 450µL of water was added before additional vortexing and centrifugation at 14,000 xg for 5 min. The upper aqueous/methanol phase was carefully removed to avoid

disturbing the precipitated protein interphase. A second 450µL volume of cold methanol was added to each sample followed by vigorous vortexing and centrifugation at 14, 000 xg for 5 min. Remaining chloroform/methanol was discarded and the precipitated protein pellet air dried in a fume hood.

2.5.3   Unfractionated on-pellet in-solution digestion

On-pellet protein digestion was performed using a modified protocol described by Duan *et al.* [45]. Briefly, 150µL of 50mM ABC (pH 8) trypsin solution was added to precipitated protein pellets (1:50 ratio) and incubated overnight at 37°C in a water bath shaker. An additional aliquot of trypsin was added the next day (1:100 ratio) for ~4 hours before acidifying (pH 3-4) with 10% formic acid (FA). Digests were centrifuged at 14,000 xg to pellet insoluble material before LC-MS or peptide fractionation.

2.5.4   SDS-PAGE followed by in-gel digestion

HeLa lysates were fractionated by SDS-PAGE as previously described [46]. Briefly, 100µg of lysate was separated on a 12% acrylamide tris-glycine gel followed by fixing, staining with Coomassie blue and destaining overnight on a horizontal shaker. Each lane was divided into 10 equal gel fractions which were manually processed into ~1x1 mm^3 cubes using a razor blade. Gel pieces were reduced in 10mM DTT for 30 minutes and alkylated in 100mM IAA for 30 minutes at room temperature in the dark. After dehydration with ACN, gel pieces were swelled in 100µL of 50mM ABC (pH 8) trypsin solution (1:25 ratio distributed evenly across 10 fractions) and incubated overnight in a water bath shaker at 37°C. Peptides were extracted from gel pieces in the presence of a water bath sonicator by adding a small volume of 10% FA followed by dehydration in 300µL ACN for 10 minutes, two times. Samples were dried in a SpeedVac (Thermo Scientific, Waltham, MA) and re-suspended in 0.1% FA prior to LC-MS.

2.5.5   GELFrEE fractionation followed by in-solution digestion

HeLa lysates (100µg/chamber) were fractionated on an 8% tris-acetate cartridge using the GELFrEE system according to the manufacturer (Expedeon, San Diego, CA). The voltage gradient used to operate the GELFrEE station is outlined in Supplementary Figure 1B. Sample collection was not started until blue loading dye was visible in the collection chamber after which 150-200µL of liquid was removed and replaced following each time interval. Running buffer was changed every hour or half hour when using 50 or 100V, respectively. Fractions 2 and 3, 11 and 12, and 13 and 14 were concatenated to generate a total of 10 fractions that were processed using chloroform/methanol and in-solution digestion as described above.

2.5.6   SCX peptide fractionation

Tryptic peptides recovered from chloroform/methanol precipitated, in-solution digests of HeLa lysate (100µg) were fractionated using SCX StageTips similarly to Kulak *et al.*[22]. Approximately 100µg of peptides, acidified with 1% TFA, were distributed evenly between four 12-plug SCX StageTips. In total, 10 SCX fractions were collected by eluting in 75, 100, 125, 150, 175, 200, 225, 250 and 300mM ammonium acetate/20% ACN solutions followed by a final elution with 5%mM ammonium hydroxide/80% ACN. Fractions eluted with identical buffers from quadruplicate StageTips were combined, dried in a SpeedVac, resuspended in ddH$_2$O and dried again to evaporate residual ammonium acetate. All samples were resuspended in 0.1% FA prior to LC-MS analysis.

2.5.7   High pH reversed phase peptide fractionation

Proteins (100µg) obtained from chloroform/methanol precipitation were digested in-solution with trypsin as described above. Next, tryptic peptides were fractionated on a Waters XBridge BEH130 C18 5µm 4.6mm x 250mm column connected to an

Agilent 1100 HPLC system at a flow rate of 1 mL/min at 20°C. Buffer A (100% water) and buffer B (10% water/90% ACN) were maintained at pH 10.0 by the addition of ammonium hydroxide immediately prior to fractionation. The gradient consisted of 5% to 35% B over 55 minutes, 70% B over 8 min, hold at 70% B for 2 minutes, return to 5% B over 5 min and then hold for 15 minutes. A total of 50 fractions were collected during the first 75 minutes of the gradient (1.5 mL per fraction) using an automated fraction collector. The volume of each fraction was reduced using a SpeedVac and every 10th fraction was concatenated. The final 10 fractions were dried completely using a SpeedVac and resuspended 0.1% FA prior to LC-MS.

2.5.8   LC-MS

All fractions/digests were analyzed using an M-class nanoAquity UHPLC system (Waters) connected to a Q Exactive mass spectrometer (Thermo Scientific). Buffer A consisted of Water/0.1% FA and Buffer B consisted of ACN/0.1%FA. Peptides (~1µg measured by BCA) were initially loaded onto an ACQUITY UPLC M-Class Symmetry C18 Trap Column, 5 µm, 180 µm x 20 mm and trapped for 4 minutes at a flow rate of 10 µl/min at 99% A/1% B. Peptides were separated on an ACQUITY UPLC M-Class Peptide BEH C18 Column, 130Å, 1.7µm, 75µm X 250mm operating at a flow rate of 300 nL/min at 35°C using a non-linear gradient consisting of 1-7% B over 7 minutes, 7-19% B over 173 minutes and 19-30% B over 60 minutes before increasing to 95% B and washing. Settings for data acquisition on the Q Exactive are outlined in Table 2.4.

2.5.9   Data Analysis

All raw MS files were searched in MaxQuant version 1.5.2.8 using the Human Uniprot database (reviewed only; updated May 2014 with 40,550 entries) [47,48]. Missed cleavages were set to 3 and I=L. Cysteine carbamidomethylation was set as a fixed modification. Oxidation (M), N-terminal acetylation (protein), and

deamidation (NQ) were set as a variable modifications (max. number of modifications per peptide = 5) and all other setting were left as default. Precursor mass deviation was left at 20 ppm and 4.5 ppm for first and main search, respectively. Fragment mass deviation was left at 20 ppm. Protein and peptide FDR was set to 0.01 (1%) and the decoy database was set to revert. Match-between-runs was enabled where specified in the main text in order to transfers missed proteins identifications between replicate LC-MS/MS runs due to limitations in instrument speed during data dependent acquisition of complex samples. Bioinformatics analysis was performed using Perseus version 1.5.5.3. Briefly, protein lists obtained from MaxQuant search results were loaded into Perseus and entries (proteins) indicated as identified by site, reverse or potential contaminant were removed [49]. When using the match-between-runs feature, datasets were filtered for proteins containing a minimum of one unique peptide in at least 2 out of 3 biological replicates. Kernel density estimation was performed using R statistical software version 3.2.3. Graphpad Prism version 6.01 was used to conduct nonparametric Kruskal-Wallis test coupled with Dunn's multiple comparison, along with the Mann-Whitney test to assess significance.

Table 2.4 Overview of parameters used for data acquisition on a Q Exactive.

Parameters	Q Exactive
Mass Range (m/z)	400-1500
Isolation Window (m/z)	1.2
MS Resolution	70K @ 200m/z
MSMS Resolution	17.5K
MS Injection Time (ms)	250
MSn Injection Time (ms)	64
AGC Target (MS)	3E6
AGC Target (MSn)	2E5
Preview Scan	n/a
Threshold (counts)	3.1E4
Minimum AGC Target	2.0E3
Data Dependent Aquisition	Top 12
Dynamic Exclusion (s)	30
Exclusion Mass Width (m/z)	n/a
Exlcude Isotopes/ Monoisotopic precursor Selection	enabled
Fragmentation Type	HCD
Normalized Collision Energy	25
Lock Mass (445.120025m/z)	best
Charge State Rejection	unassigned ,+1, 7, >8
Default Charge State	+2

2.6 References

[1]   Ly, L., Wasinger, V.C., Protein and peptide fractionation, enrichment and depletion: Tools for the complex proteome. *Proteomics* 2011, 11, 513–534.

[2]   Mann, M., Kulak, N. a., Nagaraj, N., Cox, J., The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. *Mol. Cell* 2013, 49, 583–590.

[3]   Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., et al., System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* 2012, 11, M111.013722.

[4]   Pirmoradian, M., Budamgunta, H., Chingin, K., Zhang, B., et al., Rapid and deep human proteome analysis by single-dimension shotgun proteomics. *Mol. Cell. Proteomics* 2013, 12, 3330–8.

[5]   Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., et al., The one hour yeast proteome. *Mol. Cell. Proteomics* 2014, 13, 339–47.

[6]   Kelstrup, C.D., Jersie-Christensen, R.R., Batth, T.S., Arrey, T.N., et al., Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* 2014, 13, 6187–95.

[7]   Mayne, J., Starr, A.E., Ning, Z., Chen, R., et al., Fine tuning of proteomic technologies to improve biological findings: advancements in 2011-2013. *Anal. Chem.* 2014, 86, 176–95.

[8]   Batth, T.S., Francavilla, C., Olsen, J. V, Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* 2014, 13, 6176–6186.

[9]   Pozniak, Y., Balint-Lahat, N., Rudolph, J.D., Lindskog, C., et al., System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. *Cell Syst.* 2016, 2, 172–184.

[10]   Bendall, S.C., Hughes, C., Campbell, J.L., Stewart, M.H., et al., An enhanced mass spectrometry approach reveals human embryonic stem cell growth factors in culture. *Mol. Cell. Proteomics* 2009, 8, 421–32.

[11]   Wang, H., Chang-Wong, T., Tang, H.Y., Speicher, D.W., Comparison of extensive protein fractionation and repetitive LC-MS/MS analyses on depth of analysis for complex proteomes. *J. Proteome Res.* 2010, 9, 1032–1040.

[12]   Speicher, K.D., Kolbas, O., Harper, S., Speicher, D.W., Systematic analysis of peptide recoveries from in-gel digestions for protein identifications in proteome studies. *J. Biomol. Tech.* 2000, 11, 74–86.

[13]   Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V, Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 2006, 1, 2856–60.

[14]   Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., et al., A draft map of the human proteome. *Nature* 2014, 509, 575–81.

[15] Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., et al., Mass-spectrometry-based draft of the human proteome. *Nature* 2014, 509, 582–7.

[16] Tran, J.C., Doucette, A. a, Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Anal. Chem.* 2008, 80, 1568–73.

[17] Washburn, M.P., Wolters, D., Yates, J.R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001, 19, 242–7.

[18] Stein, D.R., Hu, X., Mccorrister, S.J., Westmacott, G.R., et al., High pH reversed-phase chromatography as a superior fractionation scheme compared to off-gel isoelectric focusing for complex proteome analysis. *Proteomics* 2013, 13, 2956–2966.

[19] Branca, R.M.M., Orre, L.M., Johansson, H.J., Granholm, V., et al., HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat. Methods* 2014, 11, 59–62.

[20] McQueen, P., Krokhin, O., Optimal selection of 2D reversed-phase–reversed-phase HPLC separation techniques in bottom-up proteomics. *Expert Rev. Proteomics* 2012, 9, 125–8.

[21] Ishihama, Y., Rappsilber, J., Mann, M., Modular stop and go extraction tips with stacked disks for parallel and multidimensional Peptide fractionation in proteomics. *J. Proteome Res.* 2006, 5, 988–94.

[22] Kulak, N. a, Pichler, G., Paron, I., Nagaraj, N., Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimationa in eukaryotic cells. *Nat. Methods* 2014, 11, 319–24.

[23] Rappsilber, J., Ishihama, Y., Mann, M., Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Anal. Chem.* 2003, 75, 663–670.

[24] Wang, H., Sun, S., Zhang, Y., Chen, S., et al., An off-line high pH reversed-phase fractionation and nano-liquid chromatography–mass spectrometry method for global proteomic profiling of cell lines. *J. Chromatogr. B* 2015, 974, 90–95.

[25] Yin, X., Zhang, Y., Liu, X., Chen, C., et al., Systematic comparison between SDS-PAGE/RPLC and high-/low-pH RPLC coupled tandem mass spectrometry strategies in a whole proteome analysis. *Analyst* 2015, 140, 1314–1322.

[26] Botelho, D., Wall, M.J., Vieira, D.B., Fitzsimmons, S., et al., Top-down and bottom-up proteomics of SDS-containing solutions following mass-based separation. *J. Proteome Res.* 2010, 9, 2863–70.

[27] Thakur, S.S., Geiger, T., Chatterjee, B., Bandilla, P., et al., Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. *Mol. Cell. Proteomics* 2011, 10, M110.003699-M110.003699.

[28] Beck, S., Michalski, A., Raether, O., Lubeck, M., et al., The impact II, a very high resolution quadrupole time-of-flight instrument for deep shotgun proteomics. *Mol. Cell. Proteomics* 2015, 2014–2029.

[29]   Weston, L.A., Bauer, K.M., Hummon, A.B., Comparison of bottom-up proteomic approaches for LC-MS analysis of complex proteomes. *Anal. Methods* 2013, 5, 1–15.

[30]   Heroux, M.S., Chesnik, M. a, Halligan, B.D., Al-Gizawiy, M., et al., Comprehensive characterization of glioblastoma tumor tissues for biomarker identification using mass spectrometry-based label-free quantitative proteomics. *Physiol. Genomics* 2014, 46, 467– 481.

[31]   Darville, L.N.F., Sokolowski, B.H. a, In-depth proteomic analysis of mouse cochlear sensory epithelium by mass spectrometry. *J. Proteome Res.* 2013, 12, 3620–30.

[32]   Jones, K.A., Kim, P.D., Patel, B.B., Kelsen, S.G., et al., Immunodepletion plasma proteomics by tripleTOF 5600 and Orbitrap elite/LTQ-Orbitrap Velos/Q exactive mass spectrometers. *J. Proteome Res.* 2013, 12, 4351–65.

[33]   Mostovenko, E., Hassan, C., Rattke, J., Deelder, A.M., et al., Comparison of peptide and protein fractionation methods in proteomics. *EuPA Open Proteomics* 2013, 1, 30–37.

[34]   Wang, Y., Yang, F., Gritsenko, M. a., Wang, Y., et al., Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 2011, 11, 2019–2026.

[35]   Kyte, J., Doolittle, R.F., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982, 157, 105–32.

[36]   Manadas, B., English, J. a, Wynne, K.J., Cotter, D.R., Dunn, M.J., Comparative analysis of OFFGel, strong cation exchange with pH gradient, and RP at high pH for first-dimensional separation of peptides from a membrane-enriched protein fraction. *Proteomics* 2009, 9, 5194–8.

[37]   Magdeldin, S., Yamamoto, K., Yoshida, Y., Xu, B., et al., Deep proteome mapping of mouse kidney based on OFFGel prefractionation reveals remarkable protein post-translational modifications. *J. Proteome Res.* 2014, 13, 1636–46.

[38]   Catherman, A.D., Durbin, K.R., Ahlf, D.R., Early, B.P., et al., Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol. Cell. Proteomics* 2013, 12, 3465–73.

[39]   Giansanti, P., Tsiatsiani, L., Low, T.Y., Heck, A.J.R., Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat. Protoc.* 2016, 11, 993–1006.

[40]   Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., et al., Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* 2012, 11, 5145–5156.

[41]   Wang, N., Li, L., Exploring the precursor ion exclusion feature of liquid chromatography-electrospray ionization quadrupole time-of-flight mass spectrometry for improving protein identification in shotgun proteome analysis. *Anal. Chem.* 2008, 80, 4696–4710.

[42]   Meyer, J.G., In Silico Proteome Cleavage Reveals Iterative Digestion Strategy for High Sequence Coverage. *ISRN Comput. Biol.* 2014, 2014, 1–7.

[43]   Guo, X., Trudgian, D.C., Lemoff, A., Yadavalli, S., Mirzaei, H., Confetti: A Multi-protease Map of the HeLa Proteome for Comprehensive Proteomics. *Mol. Cell. Proteomics* 2014, 13, 1573–1584.

[44]   Wessel, D., Flügge, U.I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984, 138, 141–143.

[45]   Duan, X., Young, R., Straubinger, R.M., Page, B., et al., A Straightforward and Highly Efficient Precipitation/On-Pellet Digestion Procedure Coupled with a Long Gradient Nano-LC Separation and Orbitrap Mass Spectrometry for Label-Free Expression Profiling of the Swine Heart Mitochondrial Proteome. *J. Proteome Res.* 2009, 8, 2838–2850.

[46]   Bendall, S.C., Booy, A.T., Lajoie, G., in:, *Curr. Protoc. Stem Cell Biol.*, vol. Chapter 1, John Wiley & Sons, Inc., Hoboken, NJ, USA 2007, p. Unit 1B.1.

[47]   Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 2008, 26, 1367–1372.

[48]   Consortium, T.U., UniProt: a hub for protein information. *Nucleic Acids Res.* 2014, 43, D204–D212.

[49]   Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., et al., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 2016, 13, 731–40.

Chapter 3

Proteomic-based discovery of putative biomarkers for improved classification of endometrioid and high grade serous ovarian cancer subtypes

3.1 *Abstract*

Epithelial ovarian cancer (EOC) is a heterogeneous disease consisting of 5 main subtypes. Overall survival for EOC remains low and treatment modalities do not differ significantly between subtypes. However, patients may benefit from alternate therapies targeting pathways associated with EOC subtypes, such as Poly (ADP-ribose) polymerase (PARP) inhibitors for high grade serous ovarian cancer (HGSC). Thus accurate subtype classification will become increasingly important for patient management and outcomes. HGSC and endometrioid ovarian cancer (EC) are associated with poor and good patient prognosis, respectively. However, in a subset of cases, the differential diagnosis of HGSC from EC (primarily high-grade) based on pathological assessment is challenging. Although histotype specific markers for HGSC exist in the clinic, positive markers for EC are lacking. Therefore, we undertook a label free quantitative proteomics approach to characterize differences between EC and HGSC tumours that may reveal markers specific to EC. Our findings highlight differences between HGSC and EC biology that relate to integrin, estrogen and interferon signalling pathways. Although a subset of EC and HGSC tumours exhibited similar protein expression profiles, we identified a number of proteins consistently enriched in EC. Accordingly, several candidates, including progesterone receptor (PR), were validated by immunohistochemistry on a cohort of over 300 (EC and HGSC) tumour sections. KIAA1324 was identified as a novel marker for EC with diagnostic performance similar to PR, the current best marker of EC, and may aid in pathological assessment of difficult to discriminate tumours.

3.2   Introduction

Ovarian cancer affects 1.27% of females and remains a difficult disease to treat with a 5-year overall survival of only 46.5% [1]. Approximately 90% of all ovarian cancer cases are classified as invasive epithelial ovarian cancer (EOC) which can be divided into 5 main subtypes: high grade serous (HGSC), low grade serous (LGSC), clear cell (CCC), endometrioid (EC) and mucinous. Although EOC is comprised of heterogeneous entities, cytoreductive surgery combined with platinum/taxol-based chemotherapy remains the standard first line treatment regardless of subtype [2]. For example, CCC, mucinous, and LGSC respond poorly to chemotherapy but are treated identically to HGSC and EC [3–5]. In light of advances in genetic and molecular profiling, only recently have targeted therapies for managing EOC begun to emerge [6]. For example, clinically approved poly (ADP-ribose) polymerase (PARP) inhibitors are now a promising second line treatment option for BRCA1/2 defective, chemoresistant or recurrent HGSC [7]. Immunotherapy might be an alternative treatment option for EC which is associated with defects in DNA mismatch repair genes (Lynch syndrome) [8,9]. Moreover, the anti-angiogenic therapy bevacizumab (Avastin™) significantly increases median progression free survival (PFS) and may preferentially improve overall survival (OS) in EOC with mesenchymal and proliferative molecular subtypes [10]. Therefore, accurate characterization and discrimination of EOC subtypes is becoming increasingly pertinent for making informed treatment decisions and developing targeted therapies.

Pathological assessment of EOC can be supported by immunohistochemistry (IHC) in a reproducible manner. For example, TP53 combined with WT1 or p16/CDKN2A staining can be used to discern HGSC and LGSC [11,12]. Alternatively, hepatocyte nuclear factor-1 beta (HNF-1β) is a highly specific marker of CCC [13]. In a proportion of EOC cases, discrimination between HGSC and EC tumours remains challenging, in particular high-grade (2 and 3) EC [9,11,14,15].

Although WT1 can help confirm HGSC cases misclassified as EC, positive markers for EC exhibit limited specificity. For instance, progesterone receptor (PR) is expressed by ~81% of EC but present in 56% of HGSC [16]. Thus, identification and validation of EC specific markers are needed to improve clinicopathologic assessment and treatment selection for difficult diagnosis.

High-resolution mass spectrometry-based proteomics is a powerful and unbiased technology which is being increasingly applied to characterize complex biological systems which are inherent in cancer. A recent study by Hughes *et al.* exemplified the translational capabilities of mass spectrometry by performing high-throughput quantitative proteomic characterization of formalin fixed paraffin embedded (FFPE) tumour samples from HGSC, CCC and EC. One of the findings reported in this study was the identification of cystathionine γ-lyase (CTH), a highly specific and novel marker for CCC [17]. In another large scale proteomic study, ovarian cancer cell lines were stratified based on clustering with mesenchymal, clear cell and epithelial expression profiles to identify their possible site of origin [18]. In an attempt to elucidate histotype specific markers of EC, we undertook a mass spectrometry-based proteomics approach using fresh frozen tumour samples from HGSC and EC patients. In addition to previously reported markers, we observed several proteins by proteomics that may offer improved diagnostic performance over PR alone for detecting EC. Lastly, IHC was performed on a cohort of up to 311 (HGSC and EC) ovarian tumour sections to determine the specificity and sensitivity for a subset of proteins (PLCB1, PAM, KIAA1324, PR, CTNNB1, MUC5B, PIGR, SCGB2A1 and PIGR) for detecting EC.

3.3   Results

3.3.1   Global proteomic analysis of serous and endometrioid ovarian cancer

To identify markers enriched in or exclusive to endometrioid (EC) versus high-grade serous ovarian cancer (HGSC) subtypes, we performed label free quantitative proteomics on 20 unfractionated EC and HGSC tumour samples (10 samples from each subtype) and 4 normal ovarian tissues (Figure 3.1). Patient characteristics for each subtype are listed in Table 3.1. On average, ~4500 proteins were identified in each sample of which ~97% contained LFQ intensity values (Figure 3.2A). As anticipated, most proteins detected were of cytoplasmic origin (cell part) or belonging to macromolecular protein complexes and organelles according to PantherDB (Figure 3.2B) [19]. Approximately 78% of all proteins identified in the dataset were present in at least one sample from each group (Normal, HGSC or EC) and 13.5% (870 proteins) were exclusively shared between HGSC and EC groups (Figure 3.2C). Entries with LFQ values in ≤2 samples/group (Normal, EC or HGSC) were subsequently removed to minimize the number of proteins detected in a limited number of samples. Missing LFQ values were imputed in Perseus and Principal Component Analysis (PCA) was performed on the remaining proteins (~5800) [20]. EC samples clustered relatively well as illustrated by PCA however HGSC clustering was more dispersed (Figure 3.2D). In addition, only 3 of 4 normal samples were distinct from malignant samples. Interestingly, several tumour samples clustered more tightly with the opposite subtype suggesting similarities in protein expression profiles and/or intra-subtype heterogeneity. Indeed, ovarian cancers are known to be highly complex and heterogeneous. One HGSC sample contained substantially less protein IDs (~3100) which likely impacted its clustering with other HGSC samples.

Frozen tumour biopsy	Protein extraction and digestion	LC-MS
x4 Normal ovarian tissue x10 HGSC tumours x10 EC tumours	Tissue disruption, Urea lysis buffer, sonication, Chloroform/Methanol precipitation, Trypsin/LysC digestion	Unfractionated tumour digests analyzed on an Orbitrap Elite mass spectrometer

Database searching and LFQ	Differential expression and annotation	Candidate selection and validation
Protein identification and label free quantification (LFQ) performed in MaxQuant	'Perseus' 'EnrichR' Gene set enrichment analysis (GSEA)	'geNetClassifier' ROC-AUC IHC validation on ~300 tumours

Figure 3.1 Proteomics-based workflow for detecting EC enriched markers
Proteins were initially extracted, precipitated and digested from normal ovarian tissue (Normal), high-grade serous ovarian cancer (HGSC) and endometrioid ovarian cancer (EC) tumours. Unfractionated digests (peptides) were analyzed by liquid chromatography-mass spectrometry (LC-MS) on an Orbitrap Elite followed by database searching and label free quantification (LFQ) in MaxQuant with match-between-runs enabled and a minimum ratio count of 1 to improve proteome coverage and quantification. Protein LFQ intensities were used to identify differentially expressed proteins and perform gene ontology (GO) and pathway analysis. A subset of protein candidates specific to or highly expressed by EC tumours were validated by immunohistochemistry (IHC) on tumour microarrays consisting of EC and HGSC tumour sections to assess their performance.

Table 3.1 Patient characteristics

	HGSC	EC
Number of samples	10	10
Age at diagnosis		
Mean	64.3	55.8
Median	66	58.5
Range	41-78	36-82
FIGO stage		
IA	0	4
IC	5	3
IIC	0	2
IIIC	4	1
IV	1	0
Tumour grade		
1	0	8
2	0	2
3	10	0
Alive at last contact	4	9
Deceased	6	1
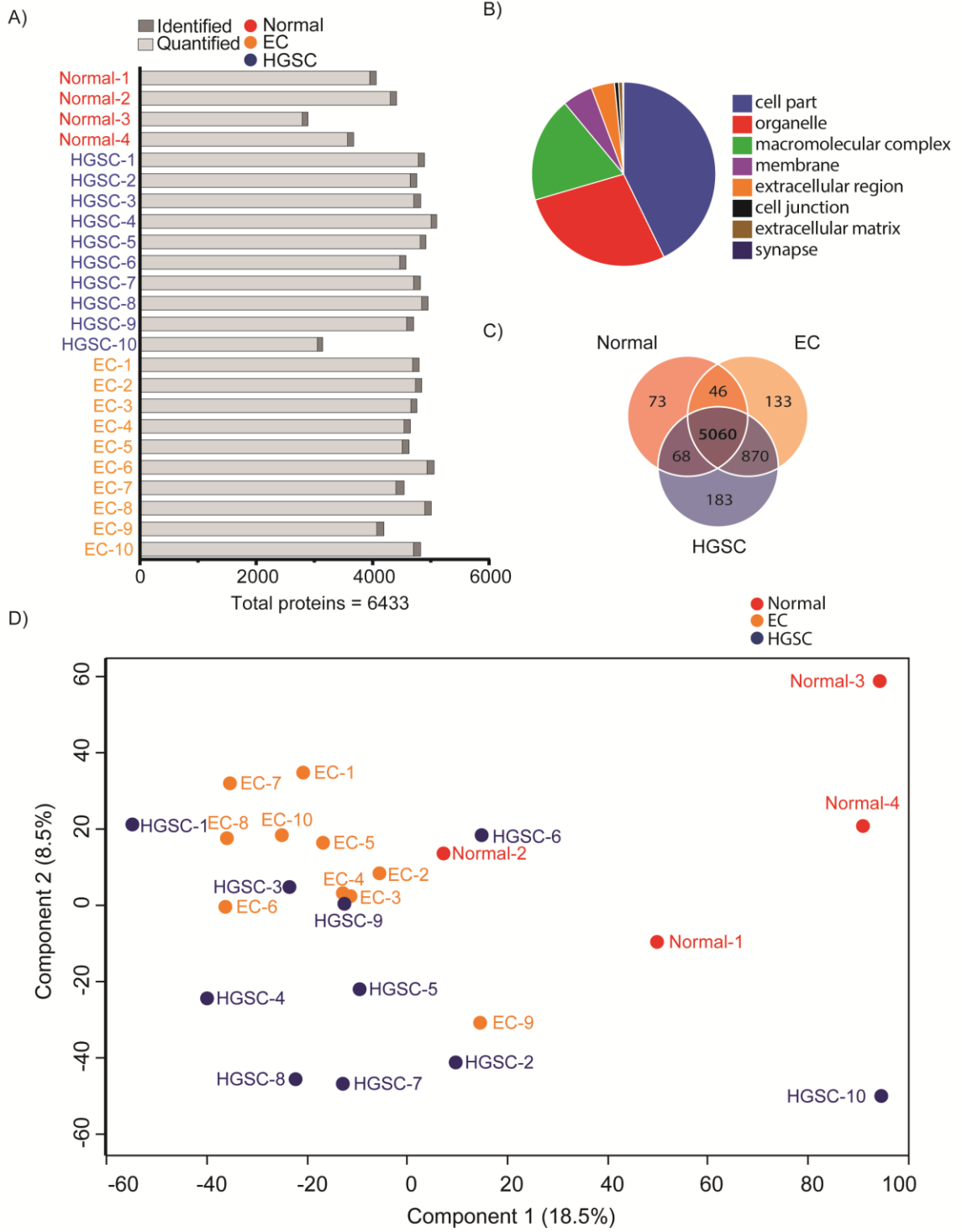
Figure 3.2 Proteome coverage and clustering ovarian (tumour) samples.

(A) Stacked bar plot showing total proteins identified (light and dark grey) and quantified (light grey) in normal ovarian tissue samples (Normal), endometrioid ovarian cancer (EC) and high-grade serous ovarian cancer (HGSC) tumour samples. (B) Distribution of GO cellular components (GOCCs) for all proteins identified. (C) Venn diagram showing overlap between Normal, HGSC and EC datasets. (D) Principal Component Analysis (PCA) using protein entries containing ≥3 LFQ intensities in at least one group (Normal, HGSC or EC). Missing LFQ values were imputed in Perseus using a down shift of 1.8 and width of 0.3.

We next determined which proteins were differentially expressed (two sample t-test, p<0.05) between ovarian tumour samples (EC or HGSC) and normal ovarian tissue (Normal). Substantially more proteins were significantly elevated in EC and HGSC tumours compared to normal tissue (Figure 3.3A and B). Interestingly, $\log_2$ fold-changes in protein expression between EC versus Normal and HGSC versus Normal correlated relatively well (Pearson correlation coefficient of ~0.8) suggesting a fair degree of similarity between these two subtypes (Figure 3.3C). Indeed, regardless of subtype (EC or HGSC), multiple proteins were highly elevated in ovarian tumours, such as phosphoserine aminotransferase (PSAT1) and L-amino-acid oxidase (IL4I1), that may warrant further investigation.

To better interrogate differences between EC and HGSC, we omitted the normal ovarian samples from further analyses. Proteins shared between HGSC and EC samples were relatively unchanged however clustering within each subtype appeared to improve slightly (Figure 3.4A and B). Moreover, numerous proteins were found to be differentially expressed between the two subtypes (two-sample t-test, p<0.05) (Figure 3.4C, ESM3.1). Accordingly, cellular tumour antigen p53 (TP53) and PR expression were significantly lower (-2.06 $\log_2$ fold-change, $-\log_{10}$ p-value = 2.03) and higher (4.20 $\log_2$ fold-change, $\log_{10}$ p-value = 4.46) in EC tumours relative to HGSC tumours, respectively. We did not detect Wilms tumour protein (WT1), a known positive marker of HGSC, in our proteomics data however CDKN2A/p16 was significantly higher in HGSC [11]. Additional proteins previously found to be differentially expressed between EC versus HGSC included MSLN (-2.67 $\log_2$ fold-change, $-\log_{10}$ p-value = 1.50), IGF2 (-1.62 $\log_2$ fold-change, $-\log_{10}$ p-value = 1.51), MMP7 (1.45 $\log_2$ fold-change, $-\log_{10}$ p-value = 2.23) and CTNNB1 (0.86 $\log_2$ fold-change, $-\log_{10}$ p-value = 2.15) [21].
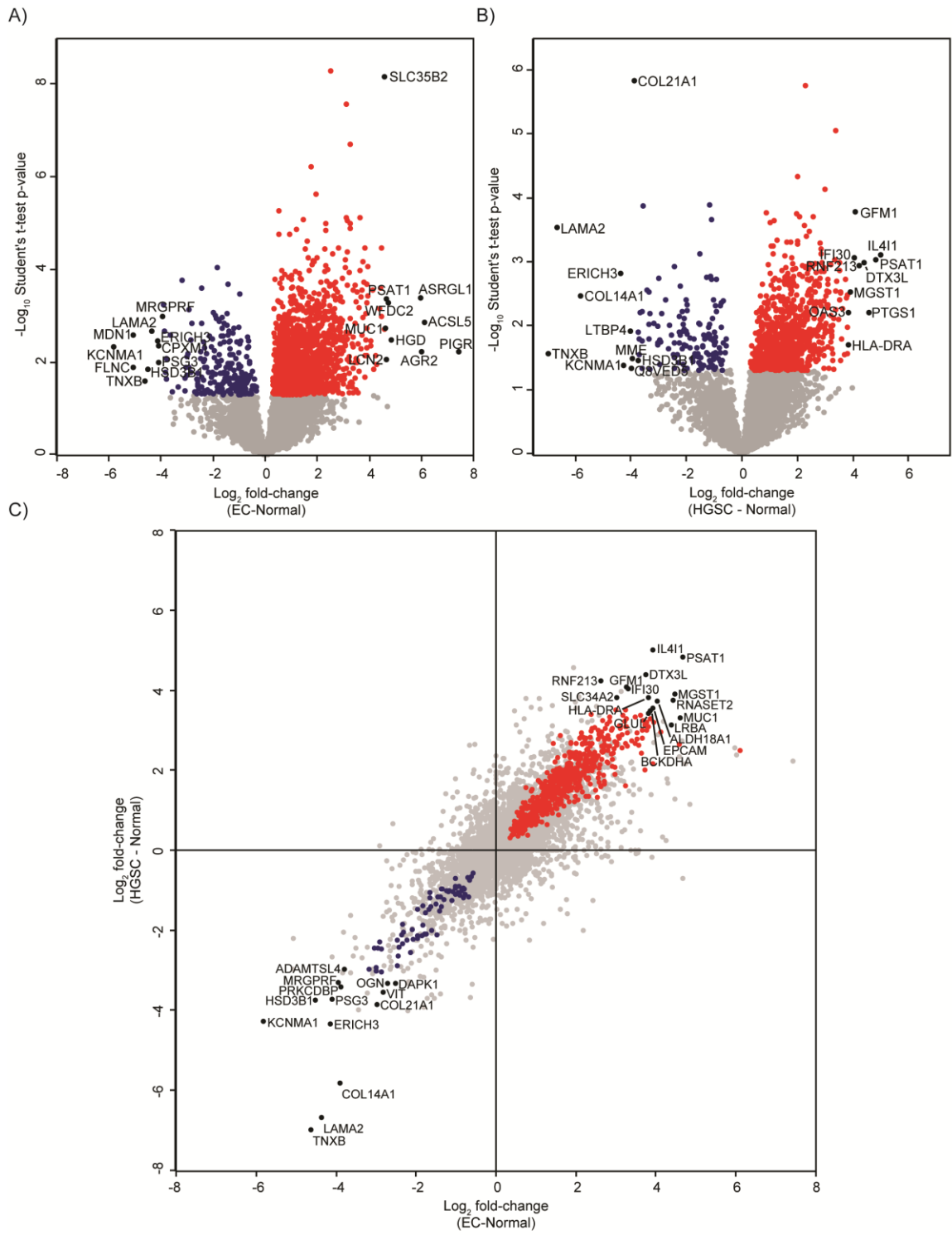
Figure 3.3 Utilizing normal ovarian tissues to assess commonalities between EC and HGSC proteomes.

(A and B) Volcano plot of $\log_2$ fold-changes in LFQ protein intensities between EC tumours and normal ovarian tissue or HGSC tumours and normal ovarian tissue. Proteins significantly elevated (p-value<0.05) in tumour samples or normal ovarian tissue are highlighted in red and blue, respectively. (C) Scatter of plot of $\log_2$ fold-changes versus $\log_2$ fold-changes from A and B. Proteins differentially expressed in both datasets (EC versus Normal and HGSC versus Normal) are highlighted in red and blue. A subset of significant proteins highly expressed in ovarian tumours regardless of subtype (EC and HGSC) or normal ovarian tissue are labelled in black.

A)

EC   HGSC

41   5559   48

Total proteins=5648

B)

C)
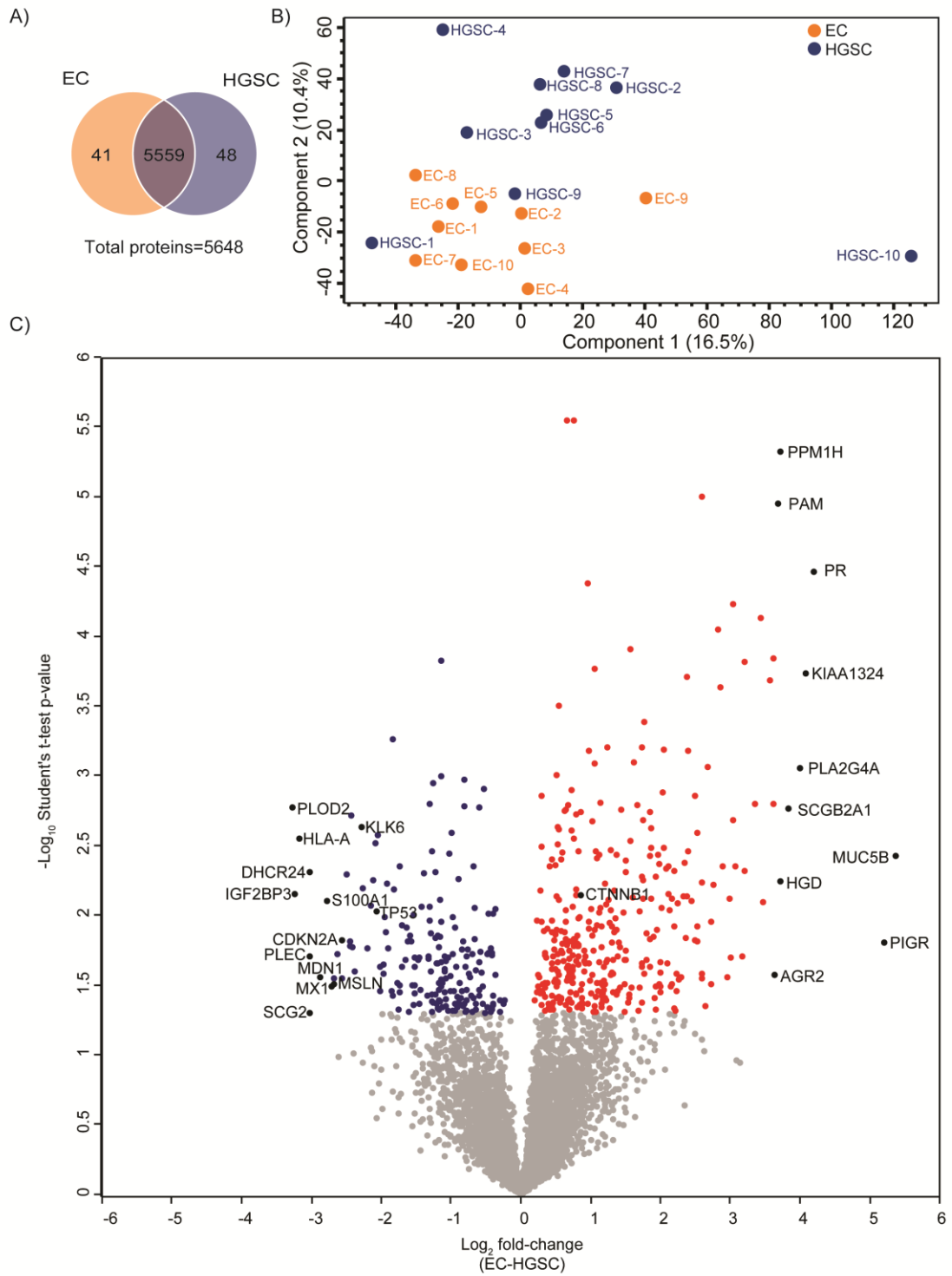
Figure 3.4 Interrogating differences between HGSC and EC proteomes.

(A) Venn diagram reveals high overlap between proteins expressed in HGSC and EC subtypes. (B) Principal Component Analysis (PCA) illustrating moderate clustering of tumour samples within each subtype. (C) Volcano plot of $\log_2$ fold-changes in LFQ intensities (EC versus HGSC) reveal a large number of differentially expressed proteins. Proteins significantly elevated in EC or HGSC are coloured in red and blue, respectively. Several of the top differentially expressed proteins are labelled in black. Only protein entries containing LFQ intensities in at least 3 out of 10 samples for either subtype (HGSC or EC) were retained for downstream comparisons.

3.3.2   *Pathway annotation reveals differences between HGSC and EC biology*

To document pathways and processes which may differ between EC and HGSC, we utilized two approaches. First, proteins significantly elevated in EC or HGSC proteomes were analyzed using 'EnrichR' to identify over-represented pathways (Reactome, BioCarta, NCI-Nature, KEGG) or GO biological processes (GOBPs) [22]. Second, using the entire proteomic expression dataset, we performed gene set enrichment analysis (GSEA) which is sensitive to cumulative changes in the expression of groups of multiple proteins.

In terms of over-representation, a wide variety of signalling pathways, metabolic/enzymatic processes, immune associated responses and transcriptional/translational activities were significantly associated with each subtype (Figure 3.5, ESM3.2 and 3.3). For example, "Neutrophil mediated immunity", "Lysosome" and "Alpha6 beta4 integrin ligand interactions" were significant in EC (Figure 3.5A). Alternatively, "Interferon signaling", "Mismatch repair" and "Integrin family cell surface interactions" were over-represented in HSGC (Figure 3.5B). Accordingly, significant differences in integrin signalling were supported by higher ITGA6, ITGB4 and ITGA1 expression in EC and elevated ITGA3 and ITGA5 levels in HGSC.

GSEA was performed using Hallmark and canonical pathways (v6.1) from the Molecular Signatures Database (MSigDB) and identified a moderate number of gene sets with nominal p-values≤0.05 albeit a subset were within the recommended FDR cut-off of 0.25 (Table 3.2) [23,24]. Substantially more gene sets were enriched in EC compared to HGSC. Indeed, melanoma, bladder cancer, and DNA replication (KEGG) were among the limited gene sets potentially enriched (negative normalized enrichment score (NES)) in HGSC. Conversely, numerous gene sets were highly enriched (positive NES) in EC with estrogen

response early (Hallmark; NES ~1.6) among the most significant (p-value<0.001 and FDR q-value <0.25).

To further illustrate differences between EC and HSGC biology, we compared normalized (z-scored) expression values of core proteins enriched in estrogen response early (Hallmark) and interferon alpha beta signalling (Reactome) gene sets (Figure 3.6A). In general, EC samples expressed high levels of proteins associated with estrogen signalling while proteins associated with interferon signalling were elevated in HGSC samples. Interestingly, HGSC-1 and EC-9 appeared to exhibit protein expression profiles characteristic of the opposing subtype.

As a complement, we performed 1D annotation enrichment in Perseus on the combined list of ~5600 proteins which revealed modest but significant alterations in GOBPs (BH FDR<0.02) (Figure 3.6B and ESM3.4) [20]. Proteins associated with defense response to fungus (1.24 $\log_2$ fold-change), O-glycan processing (0.99 $\log_2$ fold-change), and fatty acid metabolic processes (0.44 $\log_2$ fold-change) were enriched in EC. High mucin expression (MUC1, MUC5AC, MUC5B, MUC6 and MUC16) was primarily associated with elevated O-glycan processing in EC. Conversely, processes enriched in HGSC included response to type 1 interferon (0.90 $\log_2$ fold-change), DNA strand elongation (0.61 $\log_2$ fold-change), positive regulation of adaptive immune response (0.51 $\log_2$ fold-change) and mitochondrial translation (0.36 $\log_2$ fold-change). High expression of HLA-A histocompatibility antigens, and interferon induced and regulatory proteins were primarily responsible for increased interferon signalling in HGSC.

A)
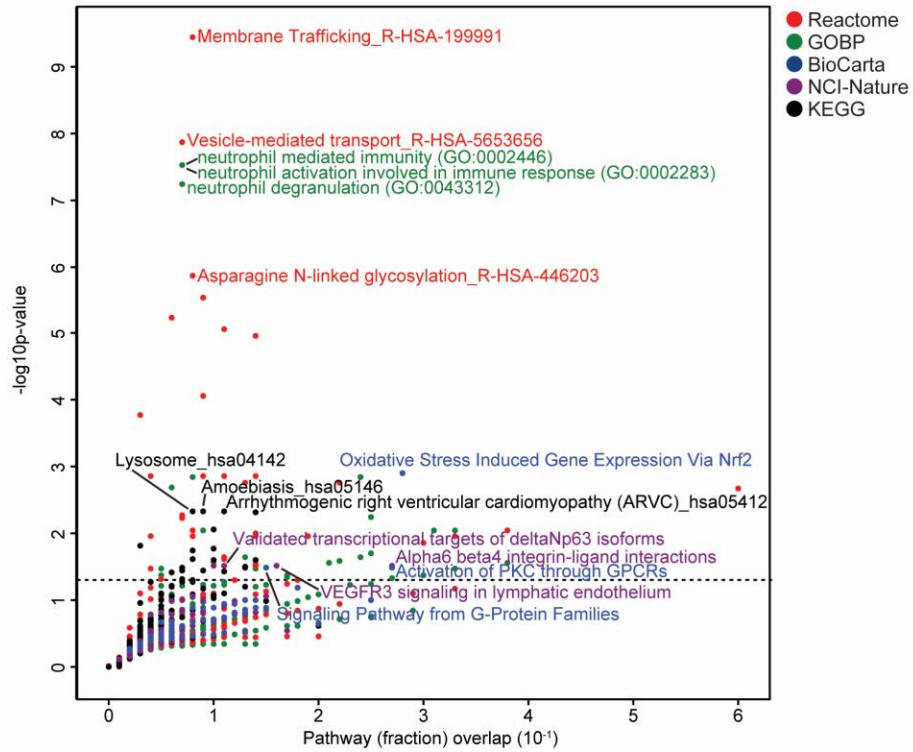
B)
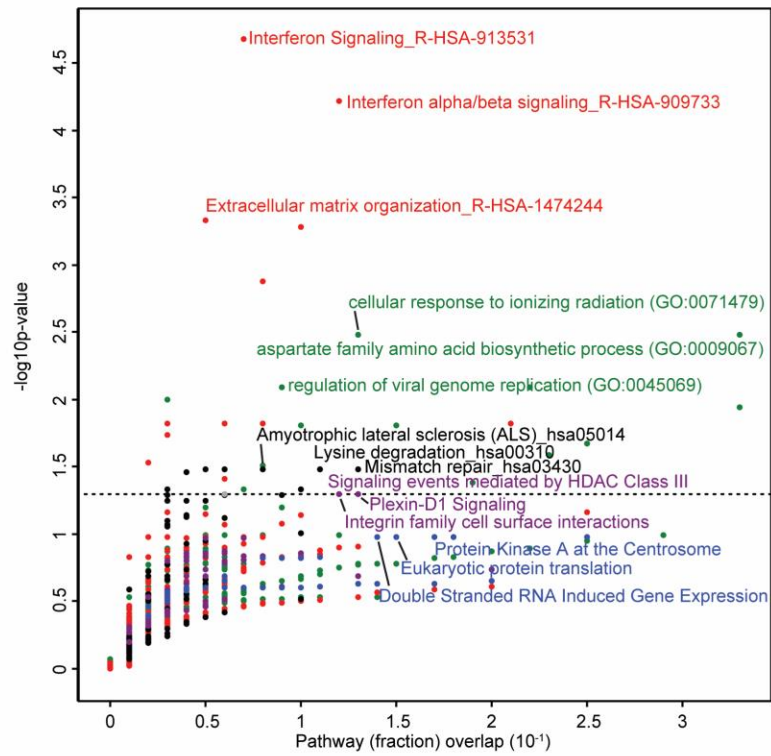
Figure 3.5 Pathway annotation of proteins differentially expressed between EC and HGSC tumours

Proteins significantly elevated in EC or HGSC tumour samples were analyzed in EnrichR using Reactome (red), BioCarta (blue), KEGG (black) and NCI-Nature (purple) pathways and GO biological processes (GOBPs; green). (A and B) The top 3 most significant (Benjamini-Hochberg (BH) adjusted p-value<0.05) pathways and GOBPs in (A) EC or (B) HGSC tumours are labelled. Horizontal dotted line indicates an adjusted p-value cut-off of 0.05 and no significant BioCarta pathways were observed in HGSC tumours.

Table 3.2 Gene sets enriched in EC and HGSC tumours

Enriched in EC (EC versus HGSC)

MSigDB	Gene hits	NES	p-value	FDR q-value	FWER p-value
HALLMARK_ESTROGEN_RESPONSE_EARLY	81	1.574	0.000	0.243	0.274
HALLMARK_ESTROGEN_RESPONSE_LATE	97	1.596	0.002	0.381	0.226
HALLMARK_TGF_BETA_SIGNALING	19	1.509	0.036	0.306	0.424
ST_P38_MAPK_PATHWAY	17	1.773	0.000	0.230	0.379
PID_HIV_NEF_PATHWAY	16	1.812	0.002	0.546	0.277
PID_ENDOTHELIN_PATHWAY	25	1.664	0.006	0.559	0.718
PID_PDGFRB_PATHWAY	85	1.612	0.008	0.417	0.829
PID_IL8_CXCR1_PATHWAY	15	1.598	0.014	0.395	0.865
PID_VEGFR1_2_PATHWAY	49	1.610	0.018	0.395	0.834
PID_PLK1_PATHWAY	19	1.547	0.019	0.431	0.932
PID_CDC42_PATHWAY	44	1.542	0.022	0.417	0.939
KEGG_LONG_TERM_DEPRESSION	30	1.769	0.002	0.192	0.387
KEGG_VASOPRESSIN_REGULATED_WATER_REABSORPTION	26	1.798	0.004	0.335	0.311
KEGG_GLYCEROPHOSPHOLIPID_METABOLISM	26	1.641	0.010	0.470	0.77
KEGG_GNRH_SIGNALING_PATHWAY	37	1.560	0.014	0.414	0.92
KEGG_GLYCEROLIPID_METABOLISM	22	1.560	0.027	0.433	0.92
KEGG_MELANOGENESIS	27	1.497	0.035	0.455	0.973
KEGG_ALDOSTERONE_REGULATED_SODIUM_REABSORPTION	15	1.526	0.041	0.445	0.956
KEGG_LONG_TERM_POTENTIATION	33	1.561	0.047	0.452	0.92
KEGG_ARRHYTHMOGENIC_RIGHT_VENTRICULAR _CARDIOMYOPATHY_ARVC	38	1.382	0.049	0.539	0.998
BIOCARTA_PAR1_PATHWAY	20	1.775	0.002	0.294	0.371
BIOCARTA_FCER1_PATHWAY	19	1.640	0.008	0.430	0.772
BIOCARTA_FAS_PATHWAY	22	1.581	0.012	0.419	0.896
BIOCARTA_PROTEASOME_PATHWAY	27	1.351	0.016	0.549	0.999
BIOCARTA_HIVNEF_PATHWAY	34	1.619	0.022	0.416	0.809
BIOCARTA_CXCR4_PATHWAY	17	1.532	0.025	0.439	0.952
BIOCARTA_TOLL_PATHWAY	16	1.571	0.027	0.434	0.907
BIOCARTA_P38MAPK_PATHWAY	15	1.515	0.041	0.473	0.964
BIOCARTA_FMLP_PATHWAY	16	1.599	0.041	0.415	0.86
BIOCARTA_SPPA_PATHWAY	17	1.468	0.044	0.486	0.982
BIOCARTA_MYOSIN_PATHWAY	18	1.514	0.044	0.460	0.964
REACTOME_PLC_BETA_MEDIATED_EVENTS	20	1.640	0.006	0.359	0.773
REACTOME_SIGNALING_BY_GPCR	99	1.433	0.008	0.506	0.994
REACTOME_GLYCEROPHOSPHOLIPID_BIOSYNTHESIS	33	1.513	0.014	0.449	0.965
REACTOME_OPIOID_SIGNALLING	37	1.585	0.020	0.425	0.889
REACTOME_REGULATION_OF_APOPTOSIS	47	1.643	0.023	0.519	0.767
REACTOME_CYCLIN_E_ASSOCIATED_EVENTS_DURING_G1_S _TRANSITION	46	1.656	0.024	0.519	0.737
REACTOME_GPCR_DOWNSTREAM_SIGNALING	65	1.443	0.025	0.502	0.991
REACTOME_ACTIVATION_OF_NF_KAPPAB_IN_B_CELLS	50	1.640	0.029	0.391	0.772
REACTOME_TRANSPORT_OF_INORGANIC_CATIONS_ANIONS _AND_AMINO_ACIDS_OLIGOPEPTIDES	19	1.438	0.030	0.509	0.992
REACTOME_G_ALPHA_Q_SIGNALLING_EVENTS	27	1.470	0.031	0.491	0.982
REACTOME_NFKB_AND_MAP_KINASES_ACTIVATION_MEDIATED _BY_TLR4_SIGNALING_REPERTOIRE	35	1.454	0.038	0.483	0.988
REACTOME_NEUROTRANSMITTER_RECEPTOR_BINDING_AND _DOWNSTREAM_TRANSMISSION_IN_THE_POSTSYNAPTIC_CELL	32	1.548	0.046	0.445	0.93
REACTOME_SCFSKP2_MEDIATED_DEGRADATION_OF_P27_P21	45	1.543	0.048	0.427	0.938
REACTOME_POST_TRANSLATIONAL_PROTEIN_MODIFICATION	83	1.510	0.049	0.445	0.966
REACTOME_MYD88_MAL_CASCADE_INITIATED_ON_PLASMA _MEMBRANE	37	1.380	0.050	0.536	0.998

Enriched in HGSC (EC versus HGSC)

MSigDB	Gene hits	NES	p-value	FDR q-value	FWER p-value
PID_BARD1_PATHWAY	16	-1.518	0.044	1.000	0.969
KEGG_MELANOMA	17	-1.517	0.014	1.000	0.971
KEGG_BLADDER_CANCER	19	-1.411	0.039	1.000	0.995
KEGG_DNA_REPLICATION	21	-1.443	0.044	1.000	0.991
REACTOME_INTERFERON_ALPHA_BETA_SIGNALING	30	-1.454	0.028	1.000	0.990
REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED _GENES	58	-1.547	0.030	1.000	0.943

Abbreviations: MSigDB Molecular signatures database, NES Normalized enrichment score, FDR False discovery rate, FWER Family-wise error rate

A)

B)
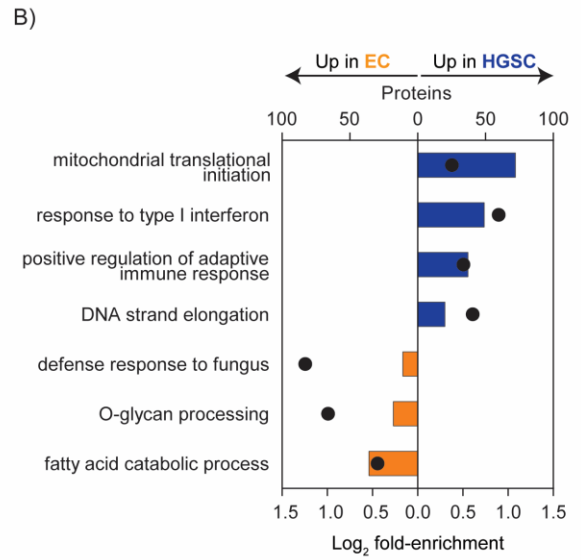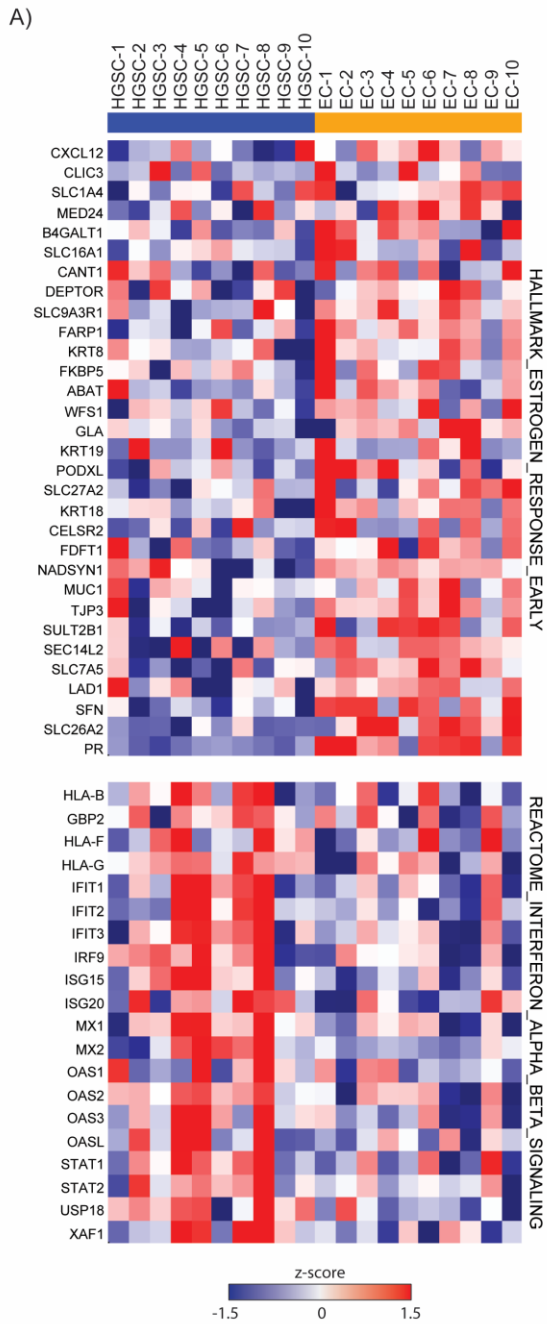
z-score
-1.5   0   1.5

Figure 3.6 Selected pathways and processes characteristic of EC and HGSC biology.

(A) Heat map showing normalized (z-score) expression of core proteins enriched in estrogen response early (top panel) or interferon alpha/beta signalling (bottom panel). (B) Selected GOBPs enriched in HGSC and EC subtypes following 1D annotation enrichment in Perseus. Bars and black dots represent total proteins and $\log_2$ fold-enrichment for each GOBP listed, respectively.

3.3.3   Selecting EC enriched proteins for IHC validation

An ideal diagnostic marker is one that achieves high sensitivity (detects all or most true positives) and high specificity (reports no or few false positives). To filter candidates of EC for validation by IHC, we first tabulated proteins with peptide evidence in ≥80% of tumour samples from one subtype and ≤20% of samples from the other subtype (Table 3.3). This approach identified 15 and 7 proteins that were largely exclusive to EC and HGSC tumours, respectively. Of note, PR was among the list of proteins enriched in EC with a $\log_2$ fold-change (EC versus HGSC) of 4.20. While this strategy may detect proteins with high specificity, it may not capture all proteins with high differential expression in EC. Therefore, we analyzed our proteomics data using the R package 'geNetClassifier' (GNC) to rank proteins with the greatest classification power in an unbiased fashion [25]. In total, 106 proteins passed the posterior probability cut-off of 0.95 and were used in training the support vector machine (Figure 3.7A and ESM3.4). The lowest error rate achieved by GNC was 0.1 (10%) and corresponded to a set of 69 proteins (Figure 3.7B). Interestingly, the top 2 ranked proteins (MUC5B and PIGR) were not identified based on our initial filtering criteria. We subsequently performed unsupervised hierarchical clustering utilizing Pearson correlation coefficients calculated from LFQ expression data restricted to the top 106 ranked proteins. While this analysis segregated HGSC and EC samples relatively well, hierarchical clustering revealed a third central cluster comprised of 3 EC and 2 HGSC samples (Figure 3.7C). This finding was in agreement with our earlier PCA which indicated some tumour samples cluster more closely with opposite subtype, such as HGSC-1 and EC-9.

3.3.4   Validation of EC enriched proteins by IHC

Based on our proteomic analysis, we opted to select several candidates with large $\log_2$ fold-changes (MUC5B and PIGR) and/or high specificity (PLCB1, PAM,

KIAA1324 and SCGB2A1) to validate using IHC on a cohort of up to 311 tumour sections (174-176 EC and 134-136 HGSC). For comparison, we included known markers of EC (PR and CTNNB1 [beta-catenin]) in our IHC panel. Receiver Operating Characteristic (ROC) Area Under the Curves (AUCs) ranged from 0.82-0.99 for these candidates when calculated from LFQ expression values (Figure 3.4D). PPM1H exhibited the highest ROC-AUC but was not validated in our initial cohort.

IHC staining for each of the candidates were scored using a 4-tier system; absent = 0, 1-50% focal = 1, 50-95% diffuse = 2, ≥95% block pattern = 3 (Figure 3.5). PR was scored as absent (0) or present (1) and CTNNB1 as membranous (0) or nuclear (1). IHC images from tumour sections demonstrating positive staining in EC (score ≥2) and negative staining in HGSC (score = 0) for each marker are shown (Figure 3.8). Importantly, staining scores ≥1 were observed in a significantly greater proportion of EC tumour sections than HGSC (Figure 3.9A). Of note, KIAA1324 was the only marker which exhibited performance similar to that of PR with a sensitivity of 88.5% and specificity of ~53% (Figure 3.9B). Increasing the cut-off score to ≥2 reduced sensitivity to 65% but increased specificity to 84% for KIAA1324 (data not shown). Although PAM yielded a very high ROC-AUC, it was the second least sensitive marker next to CTNNB1 (Figure 3.4D and Figure 3.9B). SCGB2A1 (mammaglobin-B), conversely, was the least specific marker which was unexpected given the expression, like KIAA1324, was relatively exclusive to EC versus HGSC samples in our proteomics analyses (Table 3.3 and Figure 3.9B).

Table 3.3 High confidence proteins frequently detected in HGSC and EC tumours.

Enriched in EC

Gene name	Protein name	HGSC samples (/10)	EC samples (/10)	Log$_2$ fold-change	Log$_{10}$ p-value
AKR1D1	3-oxo-5-beta-steroid 4-dehydrogenase	2	9	1.18	0.99
CHDH	Choline dehydrogenase, mitochondrial	2	9	2.98	2.11
ENTPD3	Ectonucleoside triphosphate diphosphohydrolase 3	2	8	1.73	3.20
FREM2	FRAS1-related extracellular matrix protein 2	1	8	3.48	2.10
KIAA1324	UPF0577 protein KIAA1324	2	8	4.08	3.73
NAAA	N-acylethanolamine-hydrolyzing acid amidase; N-acylethanolamine-hydrolyzing acid amidase subunit alpha; N-acylethanolamine-hydrolyzing acid amidase subunit beta	2	8	1.34	2.17
PAM	Peptidyl-glycine alpha-amidating monooxygenase; Peptidylglycine alpha-hydroxylating monooxygenase; Peptidyl-alpha-hydroxyglycine alpha-amidating lyase	0	9	3.69	4.95
PR	Progesterone receptor	1	8	4.20	4.46
PLCB1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1	2	10	2.60	5.00
PPAP2C	Lipid phosphate phosphohydrolase 2	2	9	2.82	4.05
PPM1H	Protein phosphatase 1H	1	9	3.72	5.32
SCGB2A1	Mammaglobin-B	2	8	3.84	2.76
STEAP4	Metalloreductase STEAP4	2	8	1.32	1.01
TMEM132A	Transmembrane protein 132A	2	8	0.92	0.67
TTC19	Tetratricopeptide repeat protein 19, mitochondrial	1	8	1.58	3.91

Enriched in HGSC

Gene name	Protein name	HGSC samples (/10)	EC samples (/10)	Log$_2$ fold-change	Log$_{10}$ p-value
ANGPTL2	Angiopoietin-related protein 2	9	1	-2.43	2.72
L1RE1	LINE-1 retrotransposable element ORF1 protein	9	1	-1.74	1.44
MYL6	Myosin light polypeptide 6	9	2	-1.54	2.00
PDXP	Pyridoxal phosphate phosphatase	8	1	-0.41	0.50
S100A1	Protein S100-A1	8	2	-2.78	2.10
STX16	Syntaxin-16	8	2	-1.12	1.34
TSPYL5	Testis-specific Y-encoded-like protein 5	9	2	-2.49	2.30
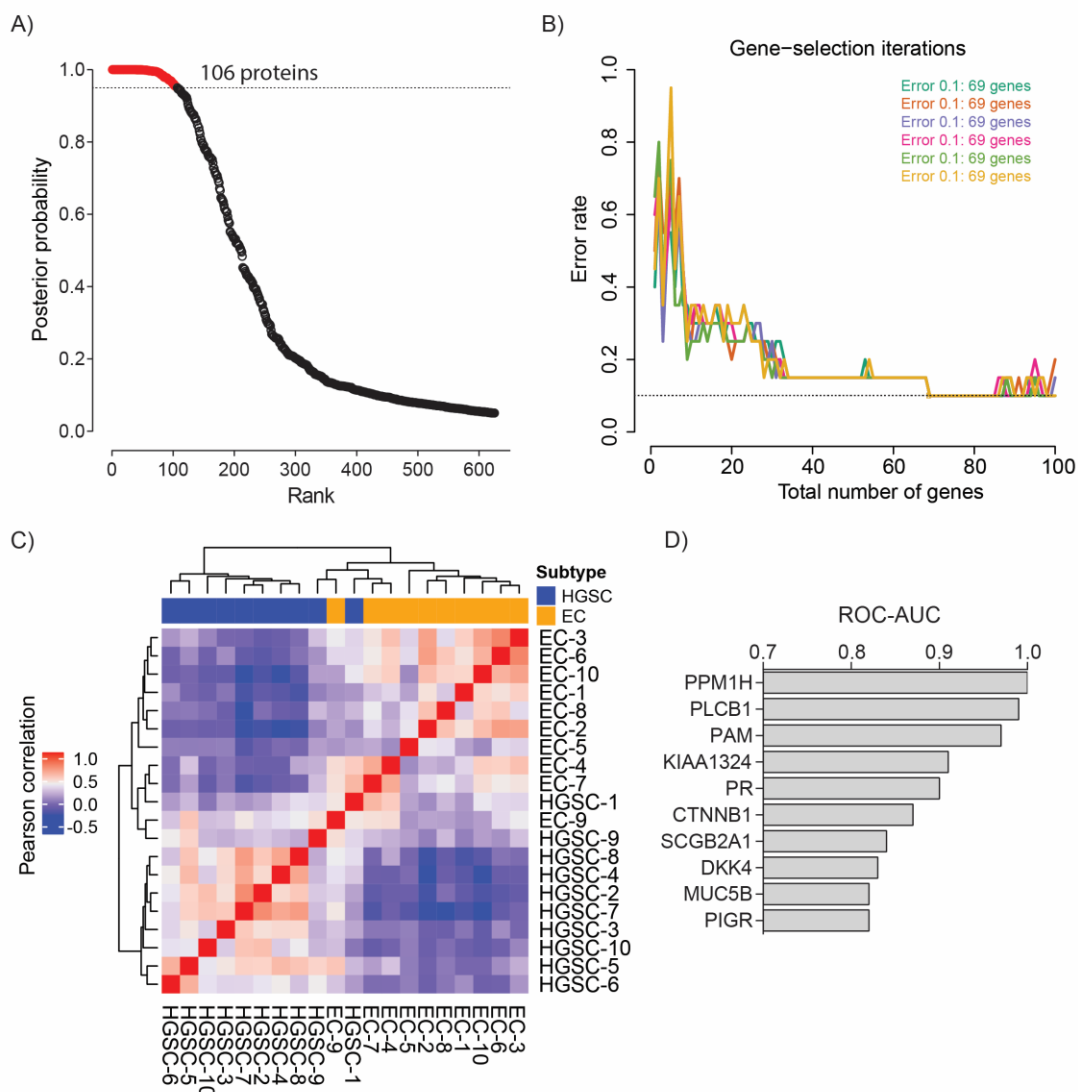
Figure 3.7 Protein selection for downstream validation of EC markers.

(A) Proteomes with LFQ intensities from each tumour sample were analyzed using the Bioconductor package 'geNetClassifier' (GNC) in R. (A and B) Total number of proteins assigned a posterior probability ≥0.95 by GNC. A minimum set of 69 genes was found to discriminate between HGSC and EC at an error rate of 10% after being trained. Unsupervised hierarchical clustering of tumour samples according to Pearson correlation coefficients using the 106 proteins with posterior probability ≥0.95. (D) Receiver-operating-characteristic (ROC) area under the curve (AUC) of several EC targets selected for validation by IHC.
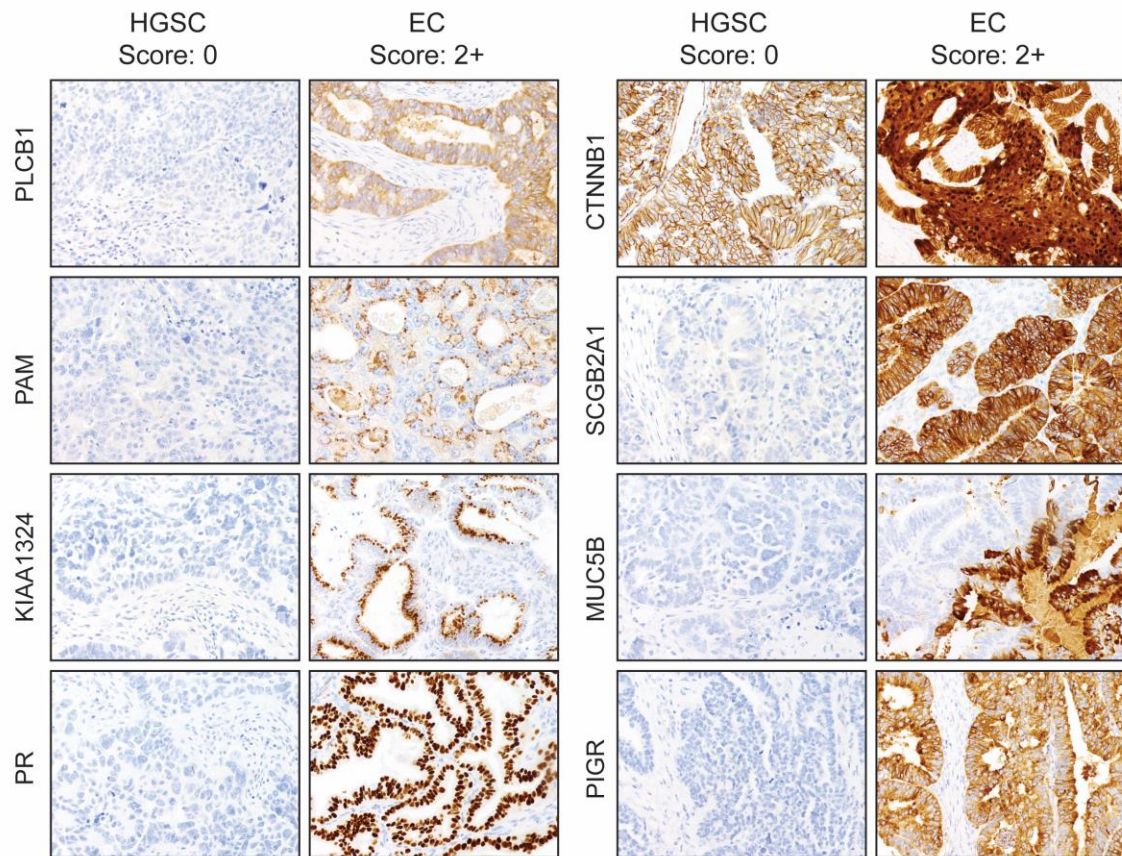
Figure 3.8 Immunohistochemical staining for markers of EC.

Images taken from EC and HGSC tumour sections showing positive (score ≥ 2) and absent (score = 0) staining for each marker (PLCB1, PAM, KIAA1324, PR, CTNNB1, SCGB2A1, MUC5B and PIGR).

A)
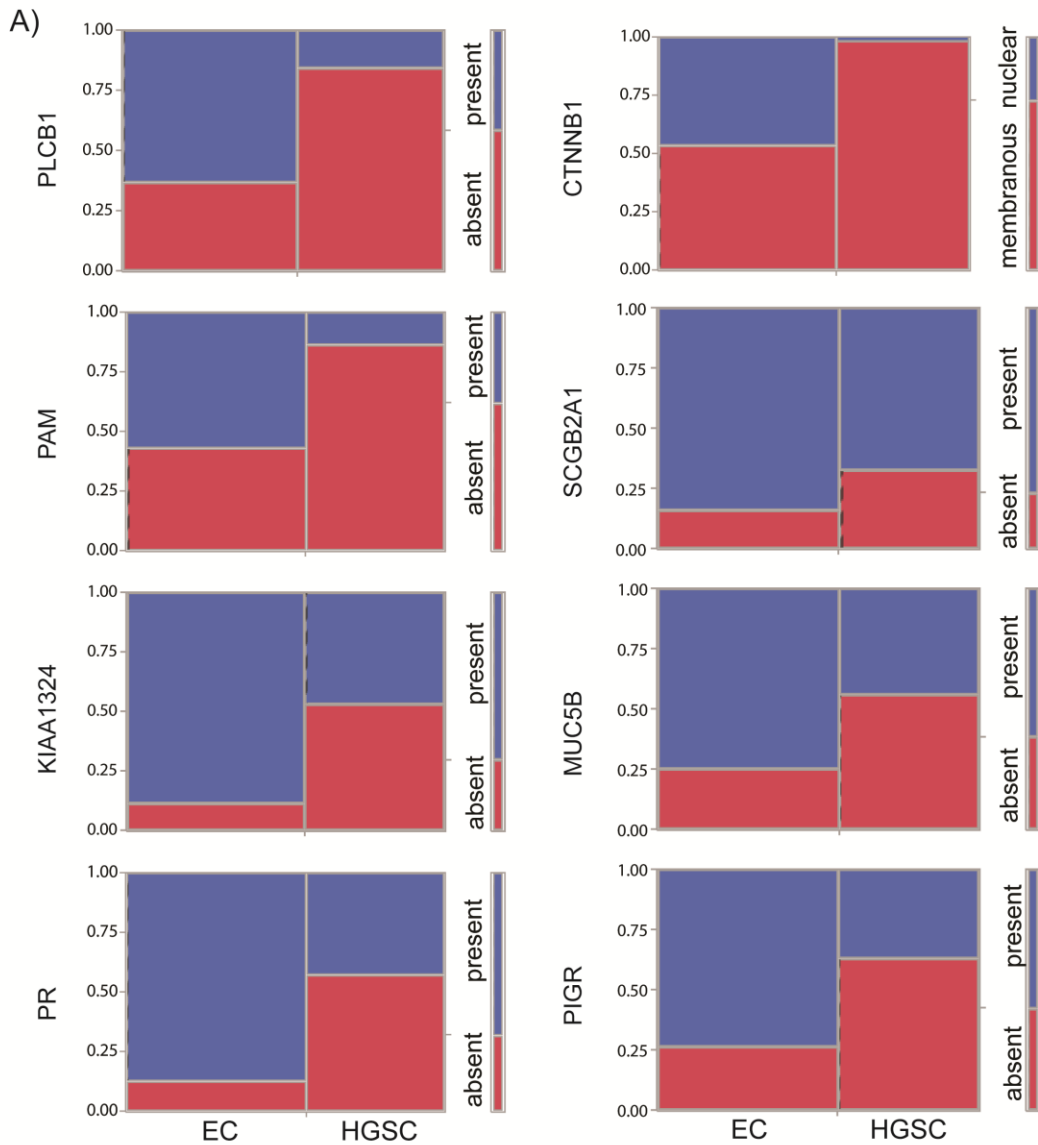
B)

Gene name	Sensitivity (%)	Specificity (%)	Significance
PLCB1	62.94	84.40	<0.001
PAM	56.90	86.47	<0.001
KIAA1324	88.51	52.94	<0.001
PR	87.29	57.25	<0.001
CTNNB1	46.45	98.53	<0.001
SCGB2A1	84.09	32.59	<0.001
MUC5B	75.00	55.67	<0.001
PIGR	73.71	63.24	<0.001

Figure 3.9 Proportion of tumours staining positive for EC markers.

(A) Mosaic plots showing the proportion of EC and HSGC sections with positive (blue region) or negative (red region) staining for PLCB1, PAM, KIAA1324, PR, CTNNB1, SCGB2A1, MUC5B and PIGR. Nuclear and membranous staining for CTNNB1 was assigned a score of 1 and 0, respectively. The proportion of tumour sections staining positive (score ≥1) was significantly higher for all markers in EC compared to HGSC (Fisher's exact test, p<0.001). (B) Corresponding sensitivity and specificity values of each target for detecting EC using a cut-off score ≥1. PR and KIAA1324 were the best performing markers with relatively good sensitivity but intermediate specificity.

3.4   Discussion

Accurate subtype classification is becoming increasingly important for the management of patients with EOC. However, differential diagnosis between EC and HGSC tumours remains troublesome in a small proportion of cases [2,15]. In light of this, we performed label free proteomic analysis on fresh frozen tumours to identify markers specific to the EC histotype. We did not employ fractionation or multiplexed-based quantification strategies but achieved relatively good proteomic coverage (~4500 protein identifications/sample) and detected a significant number of proteins that may aid in the discrimination of EC from HGSC.

While not the main focus of this study, pathway analysis revealed several interesting differences between EC and HGSC biology. For instance, members of integrin signalling pathways which have been implicated in a wide range of diseases including cancer were differentially expressed between EC and HGSC [26]. More specifically, integrin α6/β4 were elevated in EC and commonly expressed by epithelial cancers [26]. Integrin α5 (ITGA5) expression, in contrast, was higher in HGSC and has been previously associated with worse patient prognosis compared to tumours with low expression [27]. Proteins linked to interferon and estrogen signalling were also elevated in HGSC and EC, respectively, and may provide opportunities for therapeutic intervention. Interestingly, mismatch repair was significantly over-represented in HGSC compared to EC. Our group and others have previously found mismatch repair to be defective in a subset EC which supports this finding [9,28,29]. Given the limited treatment options for EOC, fractionation techniques which yield greater proteome coverage and in-depth pathway analysis are warranted [30].

Importantly, our proteomics analyses identified several proteins highly enriched in EC. All markers validated by IHC were expressed in significantly higher proportion of EC tumours however KIAA1324 was the only marker which exhibited sensitivity and specificity comparable to PR. According to the Human Protein Atlas

(www.proteinatlas.org), KIAA1324 is a prognostic (favourable) indicator for endometrial, lung and ovarian cancer [31]. Little is known regarding the role of KIAA1324 in ovarian cancer although it was initially identified as a novel-estrogen induced gene (EIG121) in endometrial cancer [32]. More recently, KIAA1324 was reported in HGSC with high ERα and KIAA1324 gene expression correlated with worse OS [33]. In this study, the authors note KIAA1324, compared to ERα alone, may better stratify HGSC patients for which hormonal therapy is beneficial. For reference, up to 88% of HGSC tumours express ERα but the number of patients estimated to respond to hormonal to therapy ranges from 13-26%. We did not correlate ERα with KIAA1324 in this study however if KIAA1324 can stratify hormone sensitive EC or HGSC, retrospective analysis of tumour samples from patients treated with estrogen antagonists are warranted.

Similar to KIAA1324, the human protein atlas lists PR, PAM, SCGB2A1, and PIGR but not PLCB1 or MUC5B as prognostic indicators of endometrial cancer [31]. Of note, TCGA studies have revealed a number of similarities between EC and the copy-number low (endometrioid) endometrial cancer subtype [34,35]. For instance, PR is upregulated in both EC and copy-number low endometrial cancer. Given the prognostic value of these markers in endometrial cancer and their high expression in EC, one or more of these factors may exhibit prognostic potential if detected in EC or HGSC. However, this needs to be confirmed by comparing PFS and OS data. Beyond the markers we validated by IHC, it would be useful to investigate additional targets with potentially greater specificity for EC including PPMH1 and PPAP2C.

Several limitations of our proteomic analysis that may mask or exaggerate differences between EC and HGSC is the unknown contribution of stromal proteins and potential differences in protein localization or activity. Stromal cell contributions remains an issue for most 'omic' studies but have been approximated with the 'Estimation of STromal and Immune cells in MAlignant Tumours using

Expression data' (ESTIMATE) score [36]. However, like CTNNB1, putative diagnostic proteins may not exhibit high differential expression but rather differ in their compartmentalization and activity and are therefore likely to be overlooked during candidate selection. Techniques such as laser capture microdissection, PTM enrichment and/or subcellular fractionation may reveal additional subtype specific proteins missed by our approach.

In summary, the lack of identifying a strong candidate for discriminating EC from HGSC (>90% sensitivity and specificity) resonates with the challenges reported by previous histopathological studies [2,11]. This was evident following PCA and hierarchical clustering whereby several EC and HGSC proteomes clustered together and appeared to share similar expression profiles. A similar finding was also documented in the quantitative proteomic study of FFPE ovarian cancer tumour sections by Hughes *et al.* where HGSC and EC samples clustered more closely at the proteome and transcriptome level compared to CCC samples [17]. In light of prior and current findings, it may be difficult to establish a single marker which discriminates EC from HGSC. Alternatively, multiple markers may be required to aid in accurate diagnosis. For example, ward's algorithm was utilized to determine which combination of TP53 and CDKN2A staining features could best discriminate HGSC from LGSC [12]. Moving forward, multivariate analysis of our IHC data may reveal subtype specific expression patterns for improved diagnosis of EC in challenging cases.

3.5   Materials and methods

3.5.1   Protein extraction from fresh frozen tumours

Fresh frozen tumour biopsy cores where provided by Dr. Köbel at the University of Calgary and confirmed to be HGSC or EC based on pathological assessment. To prepare samples for LC-MS, tumours cores were partially thawed on ice and a section corresponding to ~100mg was removed with a razor blade. Tumour

sections were immediately wrapped in tin foil and submerged in liquid nitrogen for ~10 minutes. Cryopreserved tumour pieces wrapped in tin foil were hit with a mallet 3-5 times to pulverize the tissue into a fine powder. One mL of 8M Urea, 50mM ammonium bicarbonate (ABC), 10mM dithiothreitol (DTT), 2% SDS lysis buffer was added directly to the dissociated tumour sample on tin foil and carefully transferred into a 1.5mL microfuge tube. Tumour samples were sonicated with a probe sonicator (~20 X 0.5s pulses; Level 1) (Fisher Scientific, Waltham, MA) on ice to break up residual tissue chunks and reduce viscosity. Lysates were quantified using a Pierce™ 660nm Protein Assay (Thermo Scientific™) and stored at -80°C until future use.

3.5.2   *Chloroform/Methanol protein precipitation*

A 100µg aliquot of tumour lysate was reduced in 10mM DTT for 30 minutes and alkylated in 100mM iodoacetamide (IAA) for 30 minutes at room temperature in the dark. Proteins were precipitated in chloroform/methanol in 1.5mL microfuge tubes according to Wessel and Flügge [37]. Briefly, samples in lysis buffer were topped up to 150µL with 50mM ABC then mixed with ice cold methanol (600µL) followed by adding ice cold chloroform (150µL) and vortexed thoroughly. An additional volume (450µL) of 4°C water was added followed by vortexing and centrifugation at 14, 000 xg for 5 min. The upper aqueous/methanol phase was carefully removed to avoid disturbing the precipitated protein interphase. A second 450µL volume of cold methanol was added to each sample followed by vigorous vortexing and centrifugation at 14, 000 xg for 5 min. The remaining chloroform/methanol supernatant was discarded and the precipitated protein pellet was left to air dry in a fume hood.

3.5.3   *On-pellet in-solution digestion*

On-pellet in-solution protein digestion was performed similarly to Duan *et al.* [38]. Briefly, precipitated tumour proteins were reconstituted in 100µL of 50mM ABC

(pH 8) and sonicated for 1-3 x 0.5s pulses to the break up the pellet. LysC (Wako Chemicals, USA) and mass spec grade trypsin/LysC mix (Promega, Madison, WI, USA) were added to protein samples at a 1:100 and 1:50 ratios, respectively. Protein digestion was carried out 37°C on a ThermoMixer C (Eppendorf) held at 300 rpm overnight (~18h). The next day an additional volume of trypsin/LysC mix (1:100 ratio) was added to each sample and mixed at 1400 rpm. After 4h, digests were acidified to pH 3-4 with 10% FA and centrifuged at 14,000 xg to pellet insoluble material prior to LC-MS.

3.5.4   LC-MS

Digests were analyzed using an M-class nanoAquity UHPLC system (Waters) connected to an Orbitrap Elite mass spectrometer (Thermo Scientific). Buffer A consisted of Water/0.1% FA and Buffer B consisted of ACN/0.1%FA. Peptides (~1μg measured by BCA) were initially loaded onto an ACQUITY UPLC M-Class Symmetry C18 Trap Column, (5 μm, 180 μm x 20 mm) and trapped for 6 minutes at a flow rate of 5 μl/min at 99% A/1% B. Peptides were separated on an ACQUITY UPLC M-Class Peptide BEH C18 Column (130Å, 1.7μm, 75μm X 250mm) operating at a flow rate of 300 nL/min at 35°C using a non-linear gradient consisting of 1-7% B over 1 minute, 7-23% B over 173 minutes and 23-35% B over 60 minutes before increasing to 95% B and washing. Settings for data acquisition on the Orbitrap Elite are outlined in Table 3.4.

3.5.5   Data analysis

MS files were searched in MaxQuant (1.5.8.3) using the Human Uniprot database (reviewed only; updated May 2017 with 42, 183 entries) [39,40]. Missed cleavages were set to 3 and cysteine carbamidomethylation was set as a fixed modification. Oxidation (M), N-terminal acetylation (protein), and deamidation (NQ) were set as a variable modifications (max. number of modifications per peptide = 5) and all other settings were left default. Protein and peptide FDR was set to 0.01 (1%) and

the decoy database was set to revert. The match-between-runs feature was utilized to maximize proteome coverage and quantitation by LFQ [41]. Datasets were loaded into Perseus (version 1.5.5.3) and proteins containing peptides only identified by site (modified peptides exceeding 1% FDR) or matched to reverse (decoy) database hits were removed [20]. Protein identifications with quantitative values in ≥2 samples in a least one tumour subtype (HGSC or EC) were retained for downstream analysis unless specified elsewhere. Missing values were imputed using a width of 0.3 and down shift of 1.8. Gene ontology cellular component (GOCC) analysis was performed using PantherDB. Pathway annotation was performed in using EnrichR [19,22]. Gene set enrichment analysis was carried out using GenePattern or GSEA 3.0 with all settings left as default [24]. Support Vector Machine (SVM) analysis and Pearson correlation heatmaps were produced using the Bioconductor packages 'geNetClassifier' and 'complexHeatmap', respectively [25,42]. ROC-AUC was performed in GraphPad Prism (Version 6.01).

3.5.6   Immunohistochemistry

Immunohistochemistry (IHC) staining was performed by Calgary Laboratory Services in the Department of Pathology and Laboratory Medicine at the University of Calgary using a DAKO Omnis platform on two 0.6mm representative cores (4 micron sections) from previously constructed tissue microarrays (174-176 EC and 134-136 HGSC tumour sections). Antibody information is provided in ESM3.5. All markers were scored using a 4-tier system by pathologists Dr. Martin Köbel and/or Dr. Peter Rambau. Absent staining was assigned a score of 0. Tumour cells staining as focal (1-50%), diffuse (>50-95%), or block pattern (≥95%) were assigned scores 1, 2, or 3, respectively. For CTNNB1, nuclear and membranous staining were assigned a score of 1 or 0, respectively. Fisher's exact test was used to determine statistically significant differences (p<0.05) between the proportion of EC and HGSC tumour sections staining positive and negative for each marker.

Table 3.4 Orbitrap Elite instrument parameters for data acquisition

Parameter	Setting
Mass range	400-1500m/z
MS1 resolution (Orbitrap)	120K
MS1 AGC target	1E+06
MS1 Injection time	200ms
Lock mass	445.120025
MS2 detection	IT
MS2 scan rate	Rapid
MS2 AGC target	1E+04
MS2 injection time	50ms
Top N	20
Isolation width	2.0
MS2 Activation	CID
Normalized Collision Energy	35
Dynamic exclusion	enabled
Minimum signal required	1E+03
Exclusion duration	30s
Exclusion mass width low	0.5
Exclusion mass width high	1.5
Charge Exclusion	unassigned, 1, >8

3.6   References

[1]   Howlader, N., Noone, A., Krapcho, M., Miller, D., et al., SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. 2016, 2016.

[2]   Gilks, C.B., Ionescu, D.N., Kalloger, S.E., Köbel, M., et al., Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. *Hum. Pathol.* 2008, 39, 1239–1251.

[3]   Sugiyama, T., Kamura, T., Kigawa, J., Terakawa, N., et al., Clinical characteristics of clear cell carcinoma of the ovary: A distinct histologic type with poor prognosis and resistance to platinum-based chemotherapy. *Cancer* 2000, 88, 2584–2589.

[4]   Pisano, C., Greggi, S., Tambaro, R., Losito, S., et al., Activity of chemotherapy in mucinous epithelial ovarian cancer: A retrospective study. *Anticancer Res.* 2005, 25, 3501–3505.

[5]   Schmeler, K.M., Sun, C.C., Bodurka, D.C., T. Deavers, M., et al., Neoadjuvant chemotherapy for low-grade serous carcinoma of the ovary or peritoneum. *Gynecol. Oncol.* 2008, 108, 510–514.

[6]   Banerjee, S., Kaye, S.B., New strategies in the treatment of ovarian cancer: Current clinical perspectives and future potential. *Clin. Cancer Res.* 2013, 19, 961–968.

[7]   Vaughan, S., Coward, J.I., Bast, R.C., Berchuck, A., et al., Rethinking ovarian cancer: recommendations for improving outcomes. *Nat. Rev. Cancer* 2011, 11, 719–725.

[8]   Westdorp, H., Fennemann, F.L., Weren, R.D.A., Bisseling, T.M., et al., Opportunities for immunotherapy in microsatellite instable colorectal cancer. *Cancer Immunol. Immunother.* 2016, 65, 1249–59.

[9]   Assem, H., Rambau, P.F., Lee, S., Ogilvie, T., et al., High-grade Endometrioid Carcinoma of the Ovary: A Clinicopathologic Study of 30 Cases. *Am. J. Surg. Pathol.* 2018, 42, 534–544.

[10]   Kommoss, S., Winterhoff, B., Oberg, A.L., Konecny, G.E., et al., Bevacizumab may differentially improve ovarian cancer outcome in patients with proliferative and mesenchymal molecular subtypes. *Clin. Cancer Res.* 2017, 23, 3794–3801.

[11]   Köbel, M., Bak, J., Bertelsen, B.I., Carpen, O., et al., Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry. *Histopathology* 2014, 64, 1004–1013.

[12]   Altman, A.D., Nelson, G.S., Ghatage, P., McIntyre, J.B., et al., The diagnostic utility of TP53 and CDKN2A to distinguish ovarian high-grade serous carcinoma from low-grade serous ovarian tumors. *Mod. Pathol.* 2013, 26, 1255–1263.

[13]   Kato, N., Sasou, S.I., Motoyama, T., Expression of hepatocyte nuclear factor-1beta (HNF-1beta) in clear cell tumors and endometriosis of the ovary. *Mod. Pathol.* 2006, 19, 83–89.

[14]   Schwartz, D.R., Kardia, S.L.R., Shedden, K. a, Kuick, R., et al., Gene expression in

ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. *Cancer Res.* 2002, 62, 4722–9.

[15] Madore, J., Ren, F., Filali-Mouhim, A., Sanchez, L., et al., Characterization of the molecular differences between ovarian endometrioid carcinoma and ovarian serous carcinoma. *J. Pathol.* 2010, 220, 392–400.

[16] Chen, S., Dai, X., Gao, Y., Shen, F., et al., The positivity of estrogen receptor and progesterone receptor may not be associated with metastasis and recurrence in epithelial ovarian cancer. *Sci. Rep.* 2017, 7, 1–7.

[17] Hughes, C.S., McConechy, M.K., Cochrane, D.R., Nazeran, T., et al., Quantitative Profiling of Single Formalin Fixed Tumour Sections: proteomics for translational research. *Sci. Rep.* 2016, 6, 34949.

[18] Coscia, F., Watters, K.M., Curtis, M., Eckert, M.A., et al., Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 2016, 7, 12645.

[19] Mi, H., Muruganujan, A., Thomas, P.D., PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013, 41, 377–386.

[20] Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., et al., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 2016, 13, 731–40.

[21] Köbel, M., Kalloger, S.E., Boyd, N., McKinney, S., et al., Ovarian carcinoma subtypes are different diseases: Implications for biomarker studies. *PLoS Med.* 2008, 5, 1749–1760.

[22] Kuleshov, M. V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016, 44, W90–W97.

[23] Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., et al., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 2005, 102, 15545–15550.

[24] Reich, M., Liefeld, T., Gould, J., Lerner, J., et al., GenePattern 2.0. *Nat. Genet.* 2006, 38, 500–501.

[25] Aibar, S., Fontanillo, C., Droste, C., Rivas, J.D. Las, geNetClassifier classify multiple diseases and build associated gene networks using gene expression profiles Introduction to geNetClassifier 2013.

[26] Kobayashi, M., Sawada, K., Kimura, T., Potential of integrin inhibitors for treating ovarian cancer: A literature review. *Cancers (Basel).* 2017, 9.

[27] Sawada, K., Mitra, A.K., Radjabi, A.R., Bhaskar, V., et al., Loss of E-cadherin promotes ovarian cancer metastasis via alpha 5-integrin, which is a therapeutic target. *Cancer Res.* 2008, 68, 2329–39.

[28]   Rambau, P.F., Duggan, M.A., Ghatage, P., Warfa, K., et al., Significant frequency of MSH2/MSH6 abnormality in ovarian endometrioid carcinoma supports histotype-specific Lynch syndrome screening in ovarian carcinomas. *Histopathology* 2016, 69, 288–297.

[29]   Xiao, X., Dong, D., He, W., Song, L., et al., Mismatch repair deficiency is associated with MSI phenotype, increased tumor-infiltrating lymphocytes and PD-L1 expression in immune cells in ovarian cancer. *Gynecol. Oncol.* 2018.

[30]   Kuljanin, M., Dieters-Castator, D.Z., Hess, D.A., Postovit, L.-M., Lajoie, G.A., Comparison of sample preparation techniques for large-scale proteomics. *Proteomics* 2017, 17, 1600337.

[31]   Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., et al., Tissue-based map of the human proteome. *Science* 2015, 347, 1260419–1260419.

[32]   Deng, L., Broaddus, R.R., McCampbell, A., Shipley, G.L., et al., Identification of a novel estrogen-regulated gene, EIG121, induced by hormone replacement therapy and differentially expressed in type I and type II endometrial cancer. *Clin. Cancer Res.* 2005, 11, 8258–8264.

[33]   Schlumbrecht, M.P., Xie, S.S., Shipley, G.L., Urbauer, D.L., Broaddus, R.R., Molecular clustering based on ERα and EIG121 predicts survival in high-grade serous carcinoma of the ovary/peritoneum. *Mod. Pathol.* 2011, 24, 453–462.

[34]   Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma. *Nature* 2011, 474, 609–15.

[35]   Getz, G., Gabriel, S.B., Cibulskis, K., Lander, E., et al., Integrated genomic characterization of endometrial carcinoma. *Nature* 2013, 497, 67–73.

[36]   Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., et al., Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 2013, 4.

[37]   Wessel, D., Flügge, U.I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984, 138, 141–143.

[38]   Duan, X., Young, R., Straubinger, R.M., Page, B., et al., A Straightforward and Highly Efficient Precipitation/On-Pellet Digestion Procedure Coupled with a Long Gradient Nano-LC Separation and Orbitrap Mass Spectrometry for Label-Free Expression Profiling of the Swine Heart Mitochondrial Proteome. *J. Proteome Res.* 2009, 8, 2838–2850.

[39]   Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 2008, 26, 1367–1372.

[40]   Consortium, T.U., UniProt: a hub for protein information. *Nucleic Acids Res.* 2014, 43, D204–D212.

[41]   Cox, J., Hein, M.Y., Luber, C. a, Paron, I., et al., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 2014, 13, 2513–26.

[42]   Gu, Z., Eils, R., Schlesner, M., Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016, 32, 2847–2849.

Chapter 4

Proteomic profiling of ovarian cancer extracellular vesicles for biomarker discovery

4.1 Abstract

Epithelial ovarian cancer is often detected at later stages and is thus associated with poor survival rates. Combined with transvaginal ultrasound, CA-125 is the most widely used biomarker for diagnosing and monitoring ovarian cancer progression and recurrence. However, other common gynecological conditions increase CA-125 levels and not all ovarian cancer patients exhibit elevated CA-125 levels, necessitating the discovery of additional markers or refined detection methods. Extracellular vesicles (EVs) such as exosomes and microparticles contain cell-of-origin specific cargo and may be reservoirs for cancer-associated biomarkers. Moreover, EVs can be fractionated from high abundance proteins including albumin, enabling the *de novo* detection of low abundance proteins in plasma using mass spectrometry-based proteomics. Herein we describe the systemic analysis of EVs derived from established and primary ovarian cancer cell lines, from patient ascites and from plasma. We found that samples taken directly from patients are more appropriate for biomarker discovery, as cell lines whether normal or malignant, cluster similarly in terms of EV cargo. In addition, we discovered that crude EV collection via ultracentrifugation isolates more cancer-associated cargo than does CD9-immunopreciptation. Using liquid chromatography-mass spectrometry, we uncovered proteins present in EVs derived from patient ascites that were absent or extremely low in EVs derived from the plasma of healthy donors. Several of these proteins included GPRC5A, SLC34A2, ACTBL2, and MUC16 (CA-125). We used this information to create an EV-specific parallel reaction monitoring method for plasma EVs that was able to differentiate between patients with ovarian cancer compared to non-malignant controls. Collectively, these results suggest that when paired with EV-isolation, several analytes, including CA-125, may be used in combination to detect ovarian cancer in plasma samples.

4.2 Introduction

Cancer biomarker discovery and validation has profound clinical implications for screening, diagnosis, personalized medicine and monitoring response to therapy [1]. Importantly, cancer biomarkers may improve patient survival by detecting preclinical or early stage disease. Currently, most protein based FDA approved cancer biomarkers are restricted to differential diagnosis and monitoring disease progression and recurrence but not screening or detection [2,3].

In developed countries, epithelial ovarian cancer (EOC) is the second most common gynecological malignancy but most lethal with a 5-year overall survival of about 46.2% [4,5]. Early detection is hampered by the lack of symptoms and therefore, patients are frequently diagnosed during later stages for which prognosis is poor [6]. Women experiencing abdominal and/or pelvic discomfort may undergo a transvaginal ultrasound (TVUS) to check for pelvic masses however this procedure cannot accurately discriminate between benign or malignant masses [7]. Therefore, if ovarian cancer is suspected, invasive surgical staging and debulking is required [8].

A biomarker includes any measurable entity (molecule and/or physiological process) which indicates a biological state or phenotype [9–11]. In addition to TVUS, CA-125 is the most widely used biomarker to assist in diagnosing and monitoring ovarian cancer progression and response to therapy [12]. CA-125 is an extracellular epitope consisting of repeating domains generated through cleavage of the transmembrane glycoprotein MUC16 [12]. Approximately 83% patients with advanced epithelial ovarian exhibit elevated CA-125 levels (>35 U/mL) in the blood. However, other common gynecological conditions increase CA-125 levels and not all ovarian cancer patients have elevated CA-125 levels making it problematic for screening and diagnosis [12,13]. To address this, several algorithms have been developed to improve its diagnostic performance. For example, the Risk of Malignancy Index (RMI) incorporates menopausal status, CA-

125 levels and ultrasound characteristics [14]. Alternatively, the Risk of Ovarian Cancer Algorithm (ROCA) monitors CA-125 levels over time, rather than using a set cut-off, to assess the risk of developing ovarian cancer [15]. In addition to these diagnostic assays, three large randomized control trials were recently completed which sought to determine whether screening asymptomatic women for ovarian cancer could improve patient survival [16]. Unfortunately, results from the US Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) found no survival advantage for screening by CA-125 and TVUS compared to standard care [17]. While the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) found a mortality reduction of 15% and 11% in the multimodal (CA-125+TVUS) and TVUS screening groups compared to standard care, respectively, this difference was not statistically significant based on primary analyses [18].

Given the limited sensitivity and specificity of CA-125, there is an unmet need for discovering and/or validating alternative biomarkers for the early detection and diagnosis of ovarian cancer. However, due to the low incidence of EOC, it has been suggested that any biomarker used to screen postmenopausal women will require a sensitivity of 99.6% and specificity of >75% in order to achieve a positive predictive value of 10% or higher (1 correct diagnosis for every 10 positive test results) [19,20]. Plasma is a non-invasive resource which contains numerous tissue derived biomarkers beyond CA-125 that can potentially aid in diagnosis and monitoring progression. For example, the risk of ovarian malignancy algorithm (ROMA) monitors human epididymis protein 4 (HE4) and CA-125 [21]. The FDA approved OVA1 *in vitro* diagnostic multivariate index assay measures 5 biomarkers (CA-125-II, transferrin [TF], transthyretin (prealbumin), apolipoprotein A1 [APOA1], and beta-2 microglobulin [B2M]) and was recently shown to predict the malignancy of pelvic masses better than a physician's pre-operative assessment or CA-125 alone [22]. Moreover, Yip *et al.* screened 259 serum biomarkers from nearly 500 patients with ovarian cancer or benign disease and found a panel of 9 biomarkers with greater specificity than OVA1 (88.9 versus

63.4%) at a sensitivity threshold of 90% [23]. Høgdall *et al.* screened serum from 150 cancer patients and found B2M, TF, and inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4) robustly predicted overall survival and progression free survival [24]. Unfortunately, despite the best efforts of academia and industry, no biomarkers for ovarian cancer detection have been approved. The difficulties associated with this task are reflected by the fact that on average, only 1.5 protein based biomarker tests (plasma or blood) were introduced into the clinic each year between 1993 and 2008 for any disease or condition [25].

While plasma is comprised of over 10,000 proteins and is the primary biofluid utilized for biomarker discovery, plasma proteins span a dynamic range greater than 10 orders in magnitude and therefore low abundance species are often underrepresented which could potentially provide valuable biomarkers [26–29]. Conventional, immunoaffinity-based assays (ELISAs and antibody arrays) afford extreme sensitivity (pg/mL) and specificity but require high-quality antibody pairs and are limited to predefined analytes [26]. Alternatively, mass spectrometry (MS)-based proteomics can detect and quantify thousands of proteins across several orders of magnitude in an unbiased manner. Unfortunately, high abundance proteins (HAPs) which comprise ~99% of the total plasma protein content (i.e. albumin, immunoglobulins, and transferrin) hinder MS-based plasma profiling [25–27]. HAP depletion techniques can improve proteome coverage however they are often laborious, incomplete and expensive [30]. For example, Keshishian *et al.* combined protein depletion (14 most abundant and ~50 moderately abundant plasma proteins) with extensive high-pH reversed phase fractionation to detect ~5300 plasma proteins from 4 patient samples [31,32]. In practice, most plasma profiling strategies are limited in comparison to cellular lysates in which ~10,000 proteins are being reported [33–36] .

Extracellular vesicles (EVs) include exosomes and membrane bound microparticles that are released or shed from cells which contain cellular (by-

)products and range between 40-1000 nm in diameter [37]. Unlike soluble plasma proteins, MS-based profiling of EVs purified from biological fluids (plasma and urine) are less susceptible to severe dynamic range issues associated with HAPs. While EVs may contain valuable biomarkers for clinical use, only a handful of studies have characterized their proteomes in prostate, ovarian, and colorectal cancer cell lines or biofluids [38–41].

To circumvent the limitations of traditional plasma-based biomarker discovery and foster the development of targeted proteomics assays for ovarian cancer detection and diagnosis, we characterized EV proteomes obtained from cell lines, plasma and ascites. Our proteomics approach identified a large number of previously annotated and unknown biomarkers associated with malignant EV samples. While many proteins were highly enriched in ascites, the majority remained difficult to detect or undetectable in plasma EVs. As a proof-of-principle, we performed targeted proteomics using parallel reaction monitoring (PRM) to measure ~471 peptides (240 proteins) in a cohort of non-malignant and malignant plasma EVs. In this cohort, ROC-AUC analysis revealed several peptides with similar or better performance than CA-125 which warrant further validation.

4.3   Results

4.3.1   MS-based workflow for interrogating ovarian cancer EV proteomes

Cell lines remain an integral tool for studying disease however they may not fully reflect *in vivo* biology [42]. Patient samples (tissues and biofluids) can provide uncompromised insight into human physiology but are complex in nature and limited in quantity. In light of these factors, we undertook a multi-pronged MS-based approach to characterize EV proteomes from cancer cell lines, healthy donor plasma, and ascites for biomarker discovery (Figure 4.1). Bulk EVs were primarily obtained by differential ultracentrifugation (UC) however CD9-affinity

purification (CD9AP) was also performed on healthy donor plasma and ascites to preferentially capture exosomes (endosome derived EVs <150nm in size which express the transmembrane protein CD9) [43]. LC-MS analyses were performed on Q Exactive mass spectrometer and data were searched in MaxQuant with the match-between-runs enabled across all sample types (cell lines, plasma and ascites) to improve coverage and label free quantification (LFQ) [44].

Two established (OV-90, OVCAR3) and two primary (EOC6 and EOC18) ovarian cancer cell lines and one non-malignant human immortalized ovarian surface epithelium cell line (hIOSE) were initially chosen for comparison. EOC6 and EOC18 were derived from patient ascites and characterized to be high-grade serous and low-grade serous, respectively. In cellular EV preparations, ~6230 proteins were identified in each sample on average with 76% of the total proteins detected shared amongst 4 out of 5 cell lines (Table 4.1, Figure 4.2A). We next profiled EV proteomes from healthy donor plasma and ascites (3 individuals each) which yielded substantially fewer proteins but was comparable in size to a previous study characterizing prostate cancer plasma microparticles by LC-MS [41]. Between 61-75% proteins detected within plasma or ascites were present in 2 out of 3 samples (Figure 4.2B and C). Within each purification technique (UC or CD9AP), 60-66% of the proteins identified were shared between plasma and ascites. Unexpectedly, a large number of proteins were exclusive to CD9AP preparations that were not found with UC (Figure 4.2D). This may be due to enrichment of proteins specifically associated with CD9-labelled vesicles and/or non-specific absorption of plasma proteins to beads during purification [43].

Regarding differences between plasma and ascites, we noticed two interesting results. Firstly, ascites contained more exclusive proteins than plasma in the UC group (1546 versus 503) (Figure 4.2B). Secondly, UC yielded substantially more exclusive proteins than CD9AP in ascites samples (2075 versus 424) (Figure 4.2D). We reasoned that aberrant secretion/shedding of EVs by either cancer cells

and cell types within the intraperitoneal cavity, as well as accumulation or reduced clearance, significantly expanded the repertoire of EV enriched proteins detected in ascites compared to plasma [37]. Furthermore, UC crudely isolates small to large membrane bound cargo irrespective of surface marker expression and therefore yields greater more heterogeneous EVs compared to selective purification techniques [37,43]. In total, ~7700 proteins were identified in at least one sample with ~2400 proteins common to all groups (Figure 4.2E). Few proteins were unique to plasma or ascites with cellular EVs containing the majority of exclusive proteins.
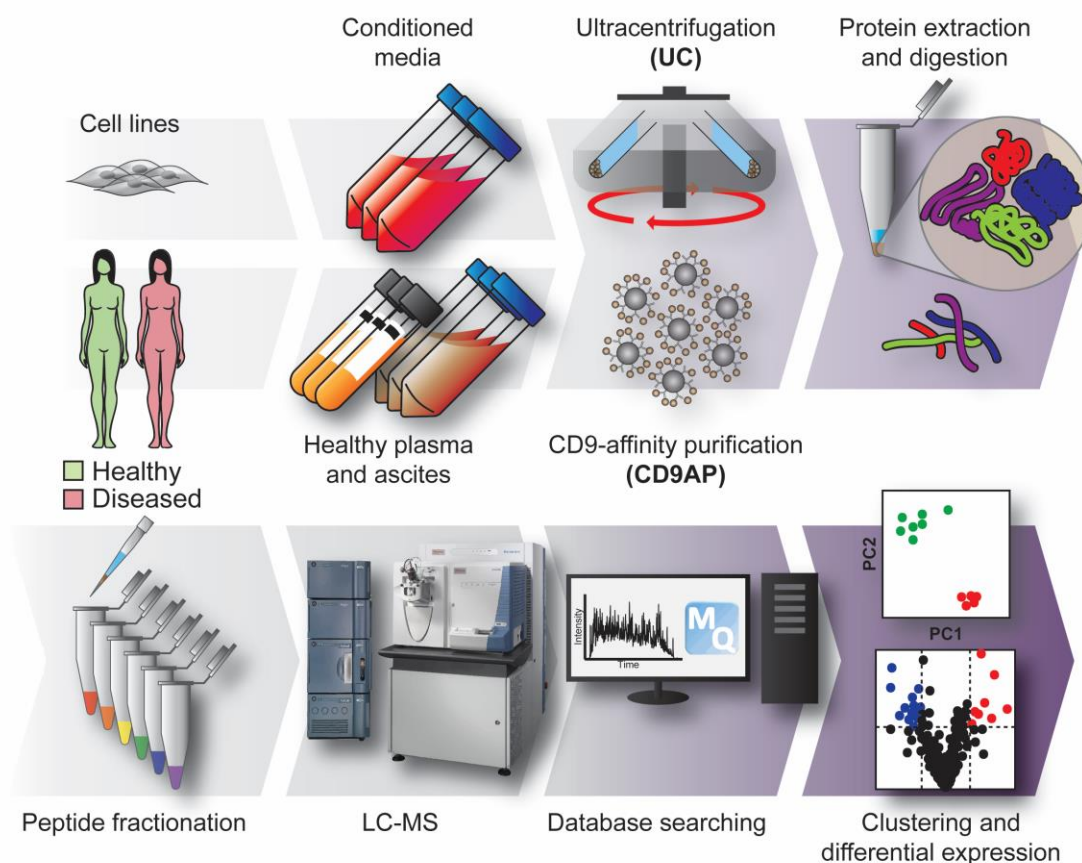
Figure 4.1 Workflow for extracellular vesicle sample preparation.

Extracellular vesicles (EVs) were purified from cell lines, healthy donor plasma and ascites by differential ultracentrifugation (UC) or CD9-affinity purification (CD9AP). Following purification, EV proteins were extracted, precipitated in chloroform/methanol and digested with trypsin/LysC overnight. Peptides were fractionated using SCX StageTips and analyzed by LC-MS on a Q Exactive mass spectrometer. *CD9AP was performed on healthy donor plasma and ascites only.

Table 4.1 Total proteins identified and quantified

Sample name	Sample type	Purification technique	Total Proteins	Proteins (≥2 peptides)	Fraction (≥2 peptides)	Proteins quantified	Fraction quantified
hIOSE	Cell line	UC	6601	5663	0.86	5498	0.83
OV-90	Cell line	UC	6527	5581	0.86	5448	0.83
OVCAR3	Cell line	UC	6652	5739	0.86	5719	0.86
EOC6	Cell line	UC	5988	4731	0.79	4522	0.76
EOC18	Cell line	UC	5382	3878	0.72	3706	0.69
Plasma-10	Plasma	UC	2992	1550	0.52	1454	0.49
Plasma-14	Plasma	UC	3277	1663	0.51	1627	0.50
Plasma-17	Plasma	UC	3111	1604	0.52	1552	0.50
EOC24	Ascites	UC	4308	2865	0.67	2714	0.63
EOC26	Ascites	UC	4417	2953	0.67	2828	0.64
EOC29	Ascites	UC	4092	2652	0.65	2550	0.62
Plasma-6	Plasma	CD9AP	3189	1852	0.58	1762	0.55
Plasma-7	Plasma	CD9AP	2508	1212	0.48	1177	0.47
Plasma-9	Plasma	CD9AP	2728	1373	0.50	1333	0.49
EOC24	Ascites	CD9AP	2797	1445	0.52	1394	0.50
EOC26	Ascites	CD9AP	2667	1286	0.48	1257	0.47
EOC29	Ascites	CD9AP	2634	1333	0.51	1287	0.49

*Healthy donor plasma and ascites samples were obtained from 3 different individuals within each purification technique (UC, Ultracentrifugation; CD9AP, CD9-affinity purification).

A) Ultracentrifugation (UC)

B) Ultracentrifugation (UC)

C) CD9-affinity purification (CD9AP)

D)

E)

Figure 4.2 Effect of sample type and purification technique on EV proteome coverage.

(A) Occurrence of proteins identified in 1 or more cell lines and overlap between non-malignant hIOSE cells and 4 ovarian cancer cell lines. (B and C) Variability in proteins detected within individual plasma and ascites samples from UC and CD9AP preparations. (D) Number of proteins common and exclusive to each sample preparation technique for plasma and ascites. (E) 5-way Venn diagram showing exclusive and common proteins for each group of samples.

4.3.2 Annotation of EV proteomes reveals similarities and differences between sample types and preparations

EVs may form via budding of ectosomes from the plasma membrane or through exocytosis of multivesicular bodies (MVB) containing intraluminal vesicles (ILVs) [45]. While differential UC can yield a heterogeneous population EVs, immunoaffinity purification approaches like CD9AP may preferentially capture small ILVs ~50-100nm in diameter (exosomes) that may be important for cellular communication [43]. To analyze the composition of EV preparations, we first performed GO cellular component (GOCC) analysis on proteomes from each group using the statistical over-representation test in PantherDB [46]. Importantly, all proteomes were significantly enriched for GOCC terms associated with EVs including SNARE complex, vesicle coat, integral to membrane and endosome (Figure 4.3). Several intracellular terms were also enriched in EV proteomes so we next compared LFQ intensities for a subset of subcellular markers to more precisely investigate their purity (Figure 4.4). Unsupervised hierarchical clustering revealed high LFQ intensities for many EV/exosomal markers suggesting good enrichment [47]. CD9, HSPA8 and GAPDH are among the most frequently detected exosomal markers and in our hands, exhibited the least variability and highest LFQ intensities across all samples (Figure 4.4 and Figure 4.5) [47]. However, many additional EV markers including CD63, CD81, CFL1, ANAX2 displayed variable and decreased levels in CD9AP samples (Figure 4.4 and Figure 4.5). Plasma and ascites samples contained abundant levels of the serum proteins ALB, TF and APOA1 and thereby suggesting incomplete removal following UC or nonspecific binding during CD9AP. Nonetheless, most golgi (TGOLN2), autophagosomal (ATG12), mitochondrial (ACO2, VDAC1) and nuclear (NUP98, LMNA) markers were extremely low in all samples. Low to high LFQ values were observed for several cytoskeletal and endoplasmic reticulum proteins which may associate with EVs naturally.
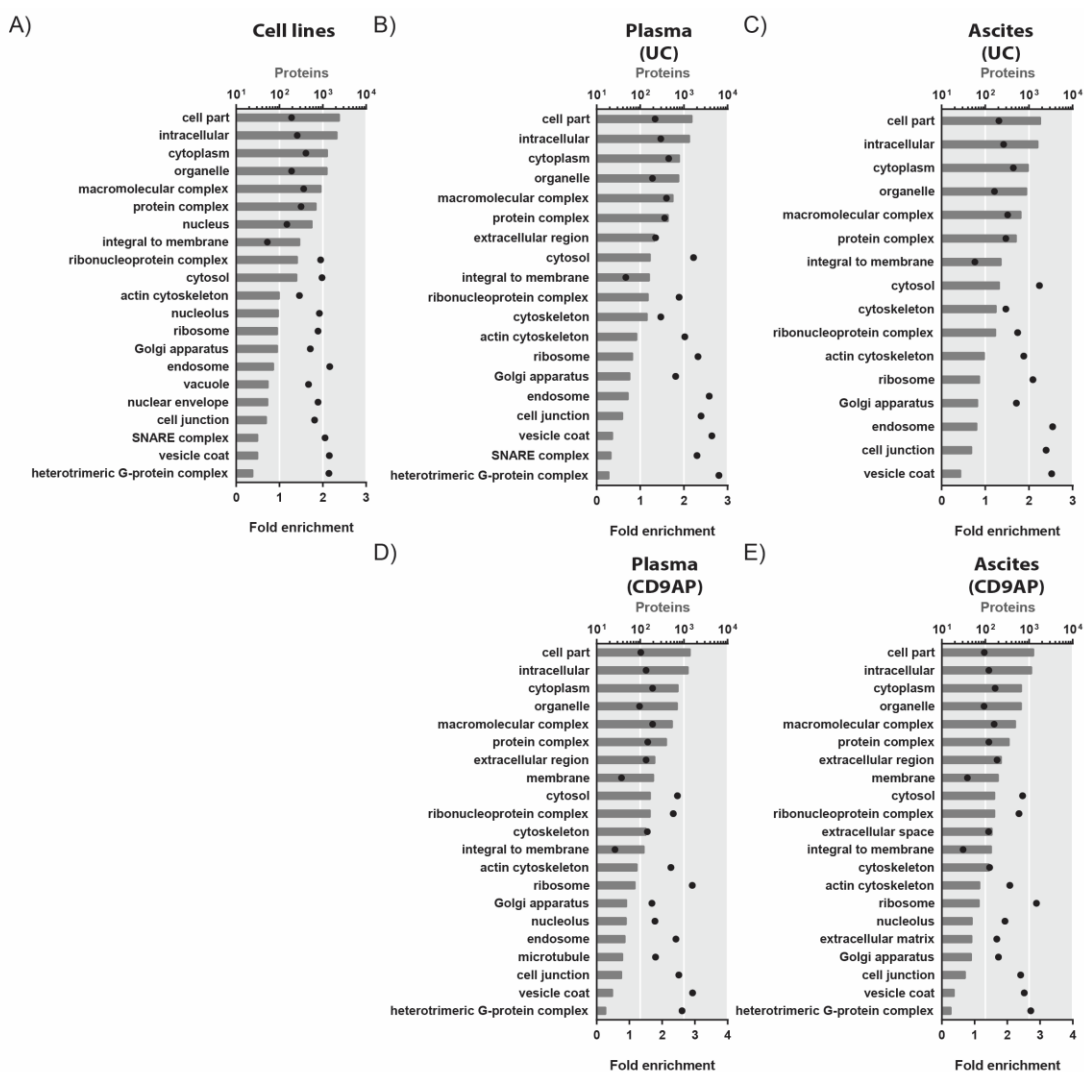
Figure 4.3 GO cellular component analysis of EV proteomes

(A-E) Fold enrichment (black dots) and total proteins (bars) associated with specified GOCC terms. Integral to membrane, vesicle coat, endosome, organelle and Golgi apparatus GOCC terms were in enriched sample preparations (cell lines, plasma and ascites) indicating the presence of vesicular proteins from cell lines, plasma and ascites by either UC or CD9AP. The presence of intracellular proteins within EVs may account for GOCC terms associated with cytoplasmic proteins.
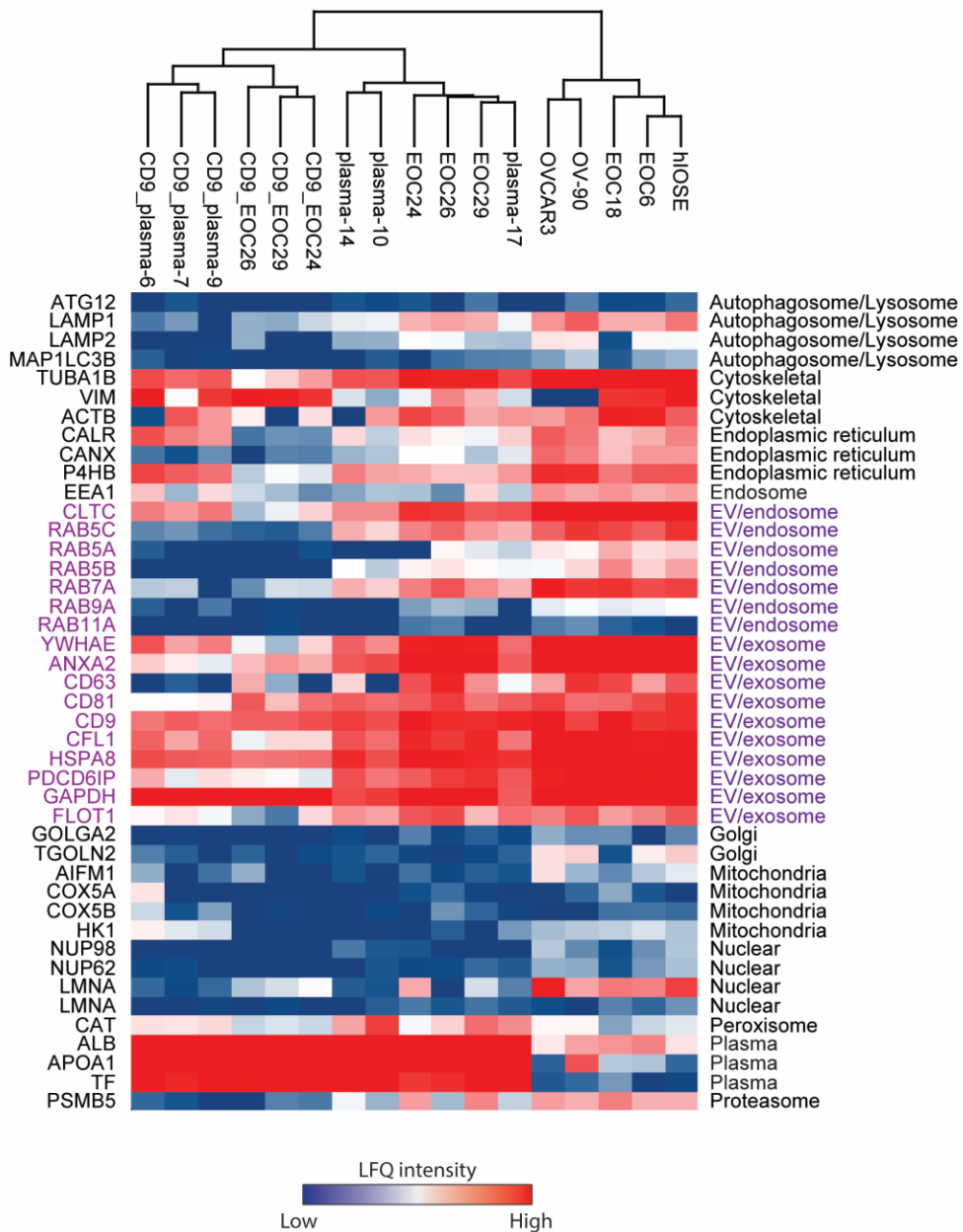
Figure 4.4 High but variable detection of markers associated with EVs in UC and CD9AP preparations.

Heat map of LFQ intensities from common EV and subcellular markers across all samples reveals high levels of EV/exosomal proteins (purple text) with minimal (intra) cellular contamination. Samples prepared by UC clustered more closely and plasma proteins were highly abundant in patient samples but not cell lines.
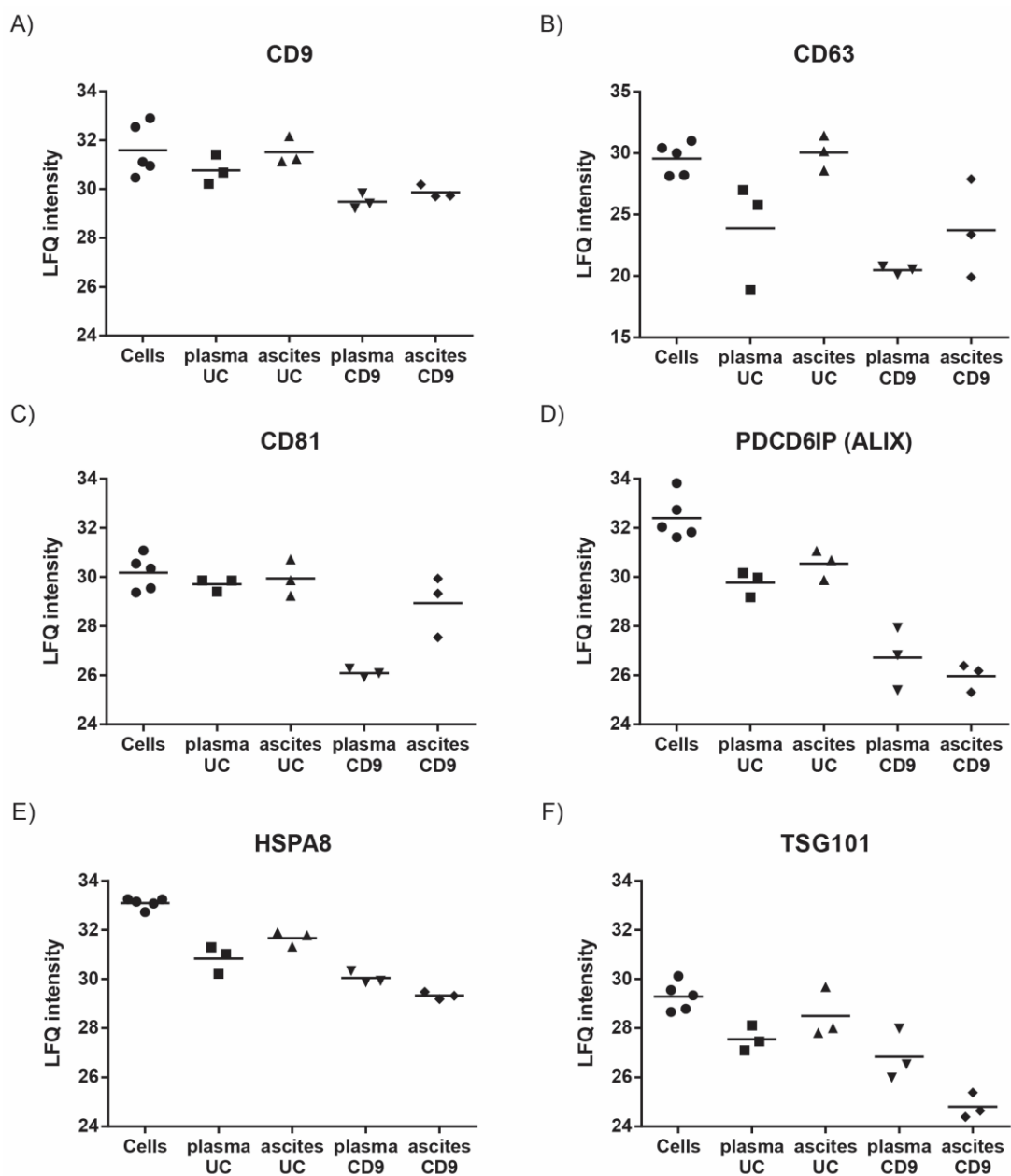
Figure 4.5 Effect of sample type and preparation technique on EV/exosome markers.

(A-F) LFQ intensities for frequently detected EV/exosomal markers in each sample reveals differences between purification techniques. (A and E) CD9 and HSPA8 were the most reliable markers across all conditions. CD9AP was associated with decreased marker intensities.

4.3.3   Integrated   proteomic   analyses   reveals   putative biomarkers associated with malignant EVs

To more stringently filter our proteomics dataset, we removed entries exclusive to a single sample within each group and imputed missing LFQ intensities as described in the materials and methods. Cell lines correlated moderately well with Pearson coefficients ranging from 0.65 to 0.81. Similarly, Pearson correlation coefficients within ascites or plasma samples were relatively good (0.72-0.84) but low between these sample types (0.43-0.61). Interestingly, Principal Component Analysis (PCA) revealed little to no clustering amongst cancer lines and no clear distinction from the non-malignant hIOSE cells (Figure 4.6A). Independent plasma or ascites samples, in contrast, clustered together regardless of the purification strategy employed (Figure 4.6B and C). As a complement, following analysis of the main dataset, we acquired LC-MS data using EVs derived from immortalized fallopian tube epithelial cells (FTEC) which have been documented as the cell of origin for most HGSC. However, a secondary PCA comparing the additional cellular EV proteomes had a negligible effect on clustering (Figure 4.7).

To identify proteins enriched in EVs from ovarian cancer cells, we subtracted LFQ intensity values of the non-malignant hIOSE sample from the proteomes of each cancer line. A one sample t-test revealed 385 proteins significantly (p<0.05) elevated in ovarian cancer EVs with EPCAM, MUC16 and SLC34A2 comprising the most differentially elevated proteins (Figure 4.6D and ESM4.1). Of note, $\log_2$ fold-changes (cancer line-hIOSE cells) for most targets varied in 1 out of 4 cancer lines suggesting a high degree heterogeneity between malignant proteomes (Figure 4.8). MUC16 expression, for example, was substantially lower in OV-90 EVs compared to OVCAR3, EOC6 and EOC18 EVs (Figure 4.8A). Like MUC16, EPCAM overexpression is common in EOC tumours and associated with decreased overall survival [48]. Interestingly, less work has been published on the Sodium-Dependent Phosphate Transport Protein 2B (SLC34A2) which is elevated

in ovarian and lung cancer and the target of antibody drug conjugate in clinical trials [49,50]. E-cadherin (CDH1) was also abundant in ovarian cancer EVs which is consistent with previous reports documenting elevated expression in EOC [51,52].

To confirm biomarkers with bona fide expression in malignant human samples, we next examined differences between plasma and ascites EV proteomes from UC and CD9AP preparations. In agreement with our earlier findings, a disproportionate number of proteins (1162) were significantly elevated (two-sample t-test, $p<0.05$) in ascites EVs compared to plasma in the UC group (Figure 4.6E and ESM4.2). Similar to our cancer cell line comparison, MUC16 and SLC34A2 were among the most differentially expressed proteins in ascites EVs from the UC group. Macrophage migration inhibitory factor (MIF), cellular retinoic acid-binding protein 2 (CRABP2) and claudin-3 (CLDN3) were abundant in ascites EVs and are known to be frequently elevated in EOC [13,35]. Retinoic acid-induced protein 3 (GPRC5A) was highly elevated in ascites EVs and may be relevant for predicting cellular responses to all-trans-retinoic acid (ATRA) in which some ovarian cancer cell lines, but not others, are sensitive to its the growth inhibitory effects [35,53,54].

Regarding CD9AP samples, the number differentially expressed proteins (two-sample t-test, $p<0.05$) between plasma and ascites was substantially smaller and more evenly distributed than UC preparations. Specifically, 259 proteins were significantly elevated in ascites EVs (Figure 4.6F and ESM4.3). Surprisingly, MUC16 and EPCAM levels were not different between plasma and ascites although a number of proteins previously implicated in EOC progression were highly enriched such as CRP, FBLN1 and MXRA5 [13,55]. For example, increased levels of C-reactive protein (CRP) are associated with worse overall survival and fibulin-1 (FBLN1) was initially identified as an estrogen responsive extracellular matrix protein in EOC [56,57]. Furthermore, Bukanovich *et al.* found Aldican or the matrix-remodeling-associated protein 5 (MXRA5) expression 10-350 fold higher in

ovarian cancer vasculature compared to normal ovarian tissues [58]. Intriguingly, we did not detect HE4 in any EV proteomes which suggested it may not associate with EVs. Indeed, Zhao et al. detected EPCAM and CA-125 but not HE4 on CD9+ exosomes purified from ovarian cancer plasma using a microfluidic ExoSearch chip [59]. In light of these comparisons, we believe CD9-negative vesicles harbour potentially valuable biomarkers that may be lost during selective enrichment strategies like immunoaffinity purification.

While EVs may be important for EOC detection and diagnosis, they can also provide insight into disease biology. Therefore, we selected protein entries from each comparison with Log_2 fold-changes >0 and searched them in EnrichR against the NCI-Nature pathway database (Figure 4.6G-I). PDGFRβ signalling was among the top significantly enriched pathway in the cancer lines and ascites (UC group). Proteins involved in CXCR4, ErbB1 (EGFR), VEGFR1/2, E-cadherin (CDH1) and hepatocyte growth factor (c-Met) signalling pathways were also significantly enriched (over-represented) in these conditions. The CD9AP ascites proteome, in contrast, was enriched in proteins primarily associated with integrin and syndecan signalling. Taken together, these observations indicate a number of proliferative and angiogenic processes likely to be utilized by HGSC.

As a complement, we examined whether EV proteomes from each comparison (cancer lines versus hIOSE, ascites versus plasma (UC group), and ascites versus plasma (CD9AP group)) were potentially correlated. Scatter plots of log_2 fold-changes versus log_2 fold-changes revealed little to no similarities between any of our three comparisons (Figure 4.9). Only a small fraction of proteins including MUC16, EPCAM, CRP, MIF were highly correlated between two out of three malignant sample types. UC and CD9AP groups were the only comparison which appeared marginally correlated. Lastly, we reduced the dataset to ~1300 entries comprised of proteins with LFQ values shared between ≥2 cell lines, plasma and ascites samples (UC and CD9AP groups). PCA illustrated tight clustering within

each sample type (Figure 4.10A). Ascites samples from the UC group clustered in-between the cell lines and plasma samples. CD9AP samples were loosely clustered although each sample type (plasma or ascites) could be distinguished from one another. In addition, we performed unsupervised hierarchical clustering after z-scoring (Figure 4.10B). Similar to PCA, ascites samples (UC group) clustered more closely to the cell lines. Taken together, our integrated proteomics approach revealed a number of similarities and differences between ovarian samples and purification strategies which may significantly impact downstream biomarker validation.

A) Ultracenrifugation
B) Ultracenrifugation
C) CD9-affinity purification

D)
E)
F)

G)
H)
I)
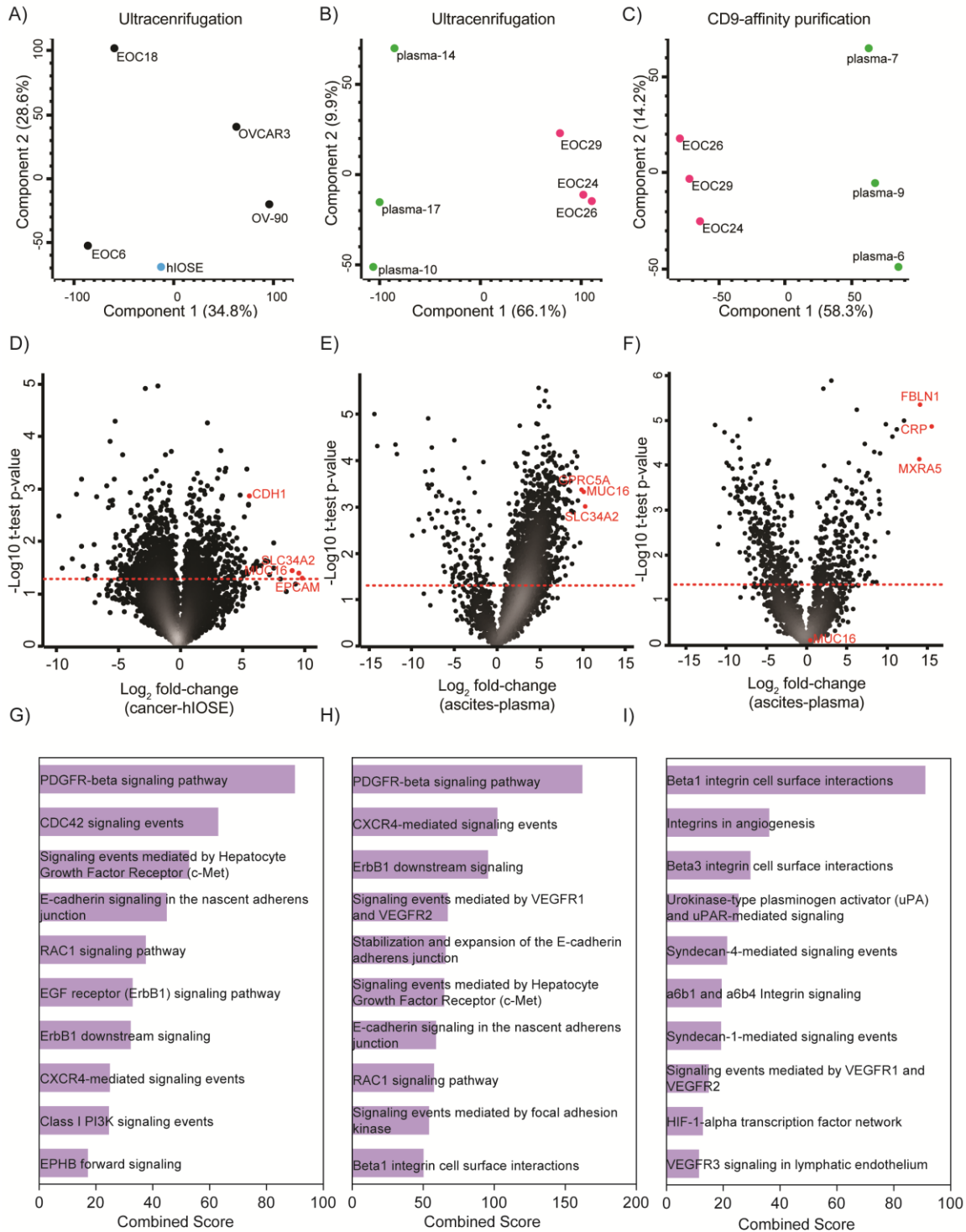
Figure 4.6 Label free proteomics identifies highly abundant proteins associated with ovarian cancer EVs.

(A-C) Principle Component Analysis (PCA) of EV proteomes reveals heterogeneity between ovarian (cancer) cell lines but high similarity within plasma and ascites samples from each preparation strategy (UC and CD9AP). (D-F) Volcano plots of $\log_2$ fold-changes LFQ intensities between ovarian cancer and hIOSE cells, UC ascites and UC plasma samples, and CD9AP ascites and CD9AP plasma samples. Highly abundant EV proteins associated with ovarian cancer samples and ascites are highlighted. (G-I) Corresponding NCI-Nature Pathways significantly enriched in EVs from cancer cells or ascites using protein lists with $\log_2$ fold-changes >0 from volcano plots in (D-F). Pathways were ranked based on the combined score ($\log_{10}$ p-value multiplied by the z-score of the deviation from the expected rank) calculated by EnrichR. Notably, pathways involved in proliferation, angiogenesis and cell adhesion were highly enriched.
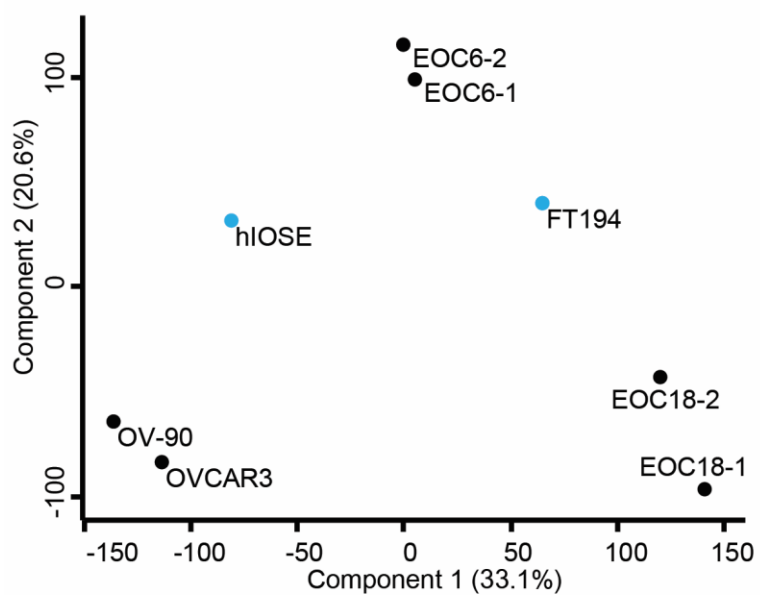
Figure 4.7 Principle component analysis of cellular EVs including a fallopian tube epithelial cell line.

PCA reveals low similarity between EV proteomes derived from high grade serous ovarian cancer cell lines. The primary EOC lines cluster more closely to the immortalized fallopian tube epithelial line FT194 while established cancer lines OV-90 and OVCAR3 formed their own group. Biological duplicates for EOC6 and EOC18 are shown.
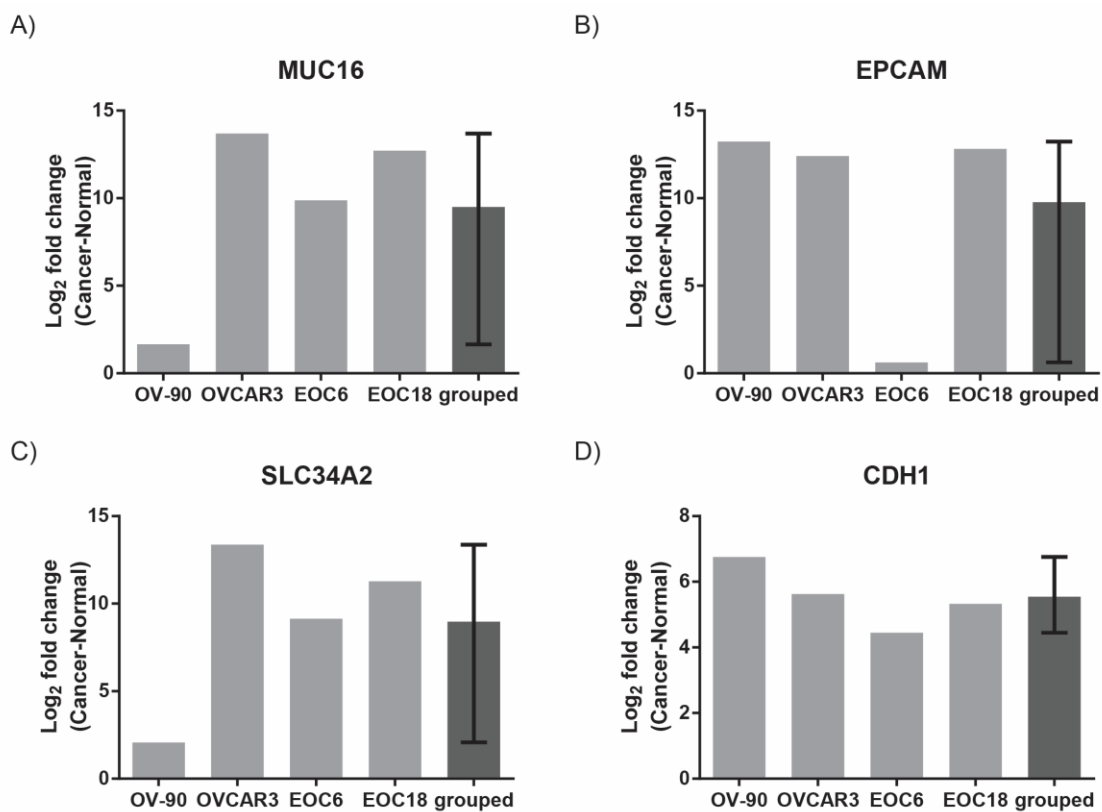
Figure 4.8 Variable levels of EOC markers detected in cellular EVs.

LFQ intensities for differentially expressed proteins were varied in 1 out 4 cell lines revealing heterogeneity within ovarian cancer lines.
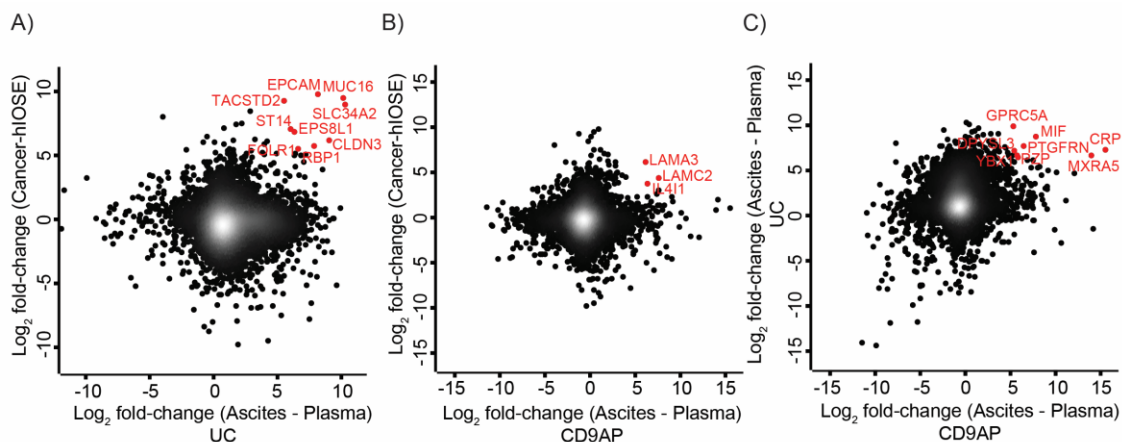
Figure 4.9 Proteins elevated in malignant EVs are weakly correlated between different sample types and preparation techniques.

(A-C) Scatter plots of log$_2$ fold-changes versus log$_2$ fold-changes from each comparison in Figure 4D-F indicate most differences in EV protein levels between malignant and non-malignant samples are not correlated. Only a small number of proteins elevated in malignant EVs were highly correlated between each comparison and are highlighted in red. (C) Log$_2$ fold-changes in EV proteins appear most correlated in the Ascites-Plasma (UC) versus Ascites-Plasma (CD9AP) comparison.
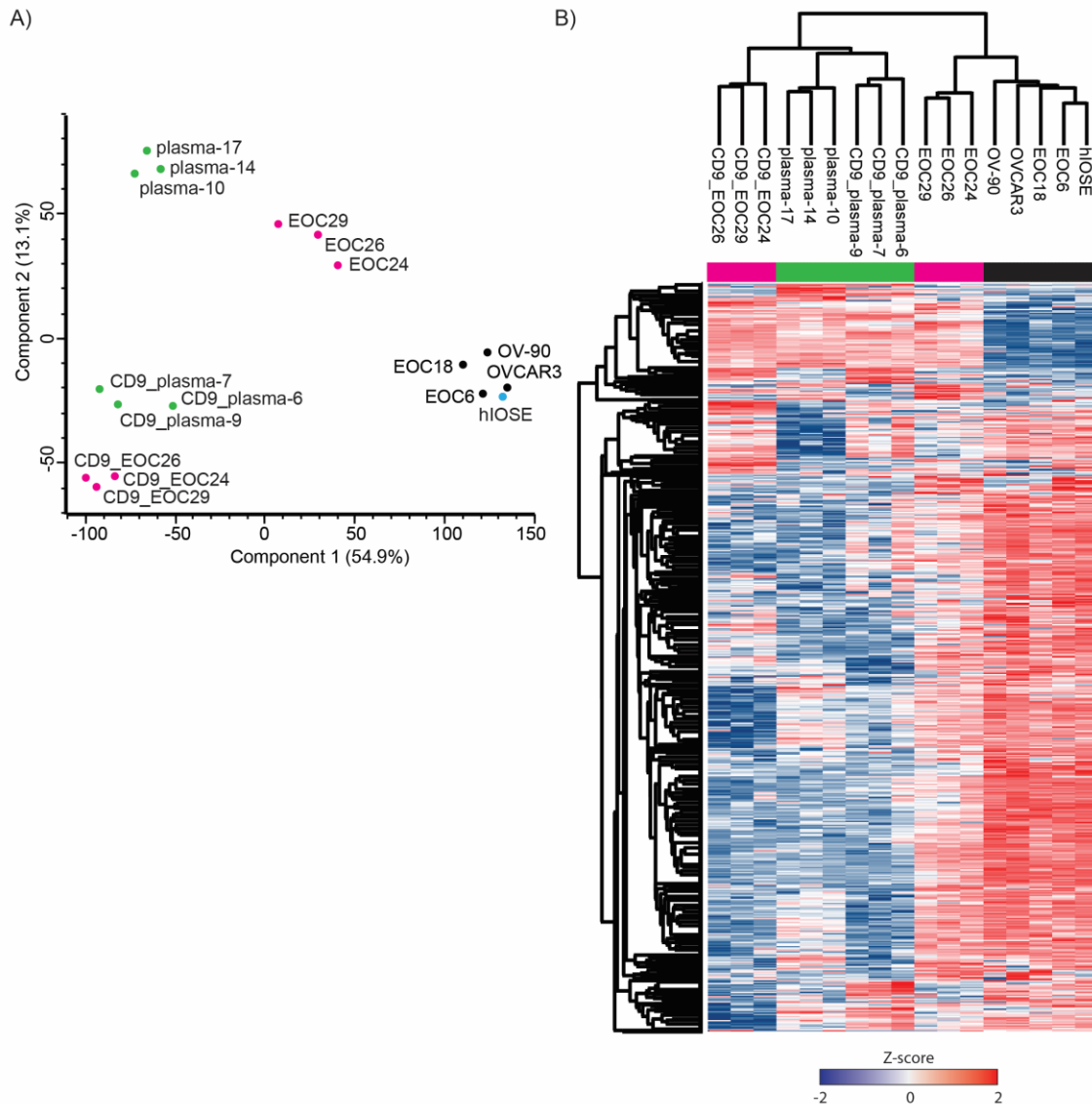
Figure 4.10 Clustering of EV proteomes.

(A) PCA showing the effect sample type and preparation method on clustering. CD9AP samples clustered more tightly compared to UC samples. Ascites (UC group) exhibited protein expression profiles characteristic of both cell lines and plasma samples (UC group). (B) Unsupervised hierarchical clustering of ~1300 proteins with quantitative values present ≥2 samples in each comparison from Figure 4D-F. Similar to PCA, sample type and preparation technique were clustered with ascites (UC group) between cell lines and plasma samples.

4.3.4   *Monitoring ovarian cancer biomarkers in patient plasma*

Based on the aforementioned comparisons, we sought to determine which proteins enriched in ascites, were the strongest predictors of disease status. For this, proteomes from the three plasma and ascites samples belonging to UC group were used to train the support vector machine (SVM) 'geNetClassifier' (GNC) in R [60]. As a consequence of the small training set size and limited sample heterogeneity, over 1300 proteins exceeded the posterior probability cut-off of 0.95 set by GNC (ESM4.4). However, GPRC5A, SLC34A2, ACTBL2, and MUC16 were amongst the top proteins identified by GNC for classifying ascites samples from healthy donor plasma (Table 4.2). Additional proteins frequently elevated in ovarian malignancies with high classification power included MIF, CLDN3, and CRABP2.

Given the global differences between EV proteomes from healthy donor plasma and ascites, we next investigated if any proteins significantly elevated in ascites EVs could be identified in an independent cohort of patient plasma samples. For this, EVs were purified by UC from 9 non-malignant and 8 malignant plasma samples (~0.5mL of plasma/sample). Of the 1435 proteins significantly enriched in either UC or CD9AP ascites EVs, we detected 354 in one or more unfractionated plasma EV samples (Figure 4.11A). Using this list, plus several housekeeping EV markers, a spectral library was built in Skyline to develop a targeted LC-MS method using Parallel Reaction Monitoring (PRM) [61,62]. A refined list of 271 proteins corresponding to 470 peptides were subsequently targeted using a PRM method in the same cohort of plasma EVs for more accurate label free quantification (Figure 5.11A). Peak areas were normalized to heavy glufibrinopeptide (hGFP) which was equally spiked into each sample. While median CV values demonstrated good reproducibility between technical replicates and low inter-sample variability, we observed heterogeneity between peptides for common EV housekeeping markers such as CD9, CD81, and HSPA8 (Figure

4.11B; data not shown). Therefore, peak areas were additionally normalized to the CD9 peptide EVQEFYK (extracellular region, AAs 120-126) which was used a surrogate marker of EV recovery and purity. We calculated Receiver Operating Characteristic Area Under the Curves (ROC-AUC) for the most differentially expressed targets to determine peptides with the greatest performance (Figure 4.11C and D). Of note, only one peptide derived from the extracellular domain of MUC16 (ELGPYTLDR) met our selection criteria with an AUC value of 0.75 (95% CI of 0.5-1.0) (Figure 4.5E). Although this peptide was 2.92 median fold higher in malignant EVs, this difference was not quite significant (Mann Whitney test, p=0.0894) (Figure 4.11E). In contrast, several peptides belonging to complement factor H related protein 4 (CFHR4), MUC1, and Serum amyloid P-component (APCS) were significantly elevated in malignant EVs (Mann Whitney test, p<0.05) (Figure 4.11E and F). CFHR4 and APCS were exclusively enriched in the CD9AP dataset while MUC1 was specific to UC comparison. Pregnancy zone protein (PZP) was significantly higher in malignant plasma EVs and enriched in both UC and CD9AP groups (log_2 fold-changes of 6.63 and 5.74, respectively). MXRA5 and GPRC5A were also significant but not SLC34A2 and CLDN3 which had lower AUCs of 0.75 and 0.722, respectively. In general, most candidates were slightly elevated in malignant plasma but not significant. In our hands, proteins enriched in ascites EVs were challenging to detect in unfractionated samples and we were unable to monitor all peptides due to the large number of targets. While most peptides were well correlated, further work is needed to refine this list and confirm the best predictors of disease status.

Table 4.2 Top ranking proteins of ascites EVs identified by 'geNetClassifier'

Rank	Gene Symbol	Gene Name	Log$_2$ fold-change (Ascites-Plasma)
1	GPRC5A	Retinoic acid-induced protein 3	9.91
2	SLC34A2	Sodium-dependent phosphate transport protein 2B	10.30
3	ACTBL2	Beta-actin-like protein 2	9.44
4	MUC16	Mucin-16	10.13
5	CALB2	Calretinin	9.3
6	CRABP2	Cellular retinoic acid-binding protein 2	9.33
7	CLDN3	Claudin-3	9.09
8	THY1	Thy-1 membrane glycoprotein	9.68
9	GLUL	Glutamine synthetase	8.92
10	MIF	Macrophage migration inhibitory factor	8.78

A) Build spectral library → Select peptides → Scheduled PRM method → Filter results in Skyline → Normalize to hGFP → %CV → Normalize to CD9 avg. technical replicates → ROC-AUC

B) %CV plot: Non-malignant (C5529, C5535, C5554, C5598, C5627, C5634, C5635, C5640, C5668) and Malignant (C3859, C5501, C5530, C5733, C5669, C5721, C5760, C5528)

C) ROC-AUC bar chart (peptides)

D) ROC curves: Sensitivity % vs 100%-Specificity%

E)

Gene	Protein name	Peptide	AUC	Fold-change	p-value
CFHR4	Complement factor H-related protein 4	FCDMPVFENSR	0.8750	2.97	0.0079
MUC1	Mucin-1	YVPPSSTDR	0.8611	10.48	0.0111
APCS	Serum amyloid P-component	AYSLFSYNTQGR	0.8611	1.93	0.0101
PZP	Pregnancy zone protein	IISIMDEK	0.8542	2.97	0.0119
MUC16	Mucin-16	ELGPYTLDR	0.7500	2.92	0.0894

F) Area scatter plots: FCDMPVFENSR_CFHR4, YVPPSSTDR_MUC1, AYSLFSYNTQGR_APCS, IISIMDEK_PZP, ELGPYTLDR_MUC16
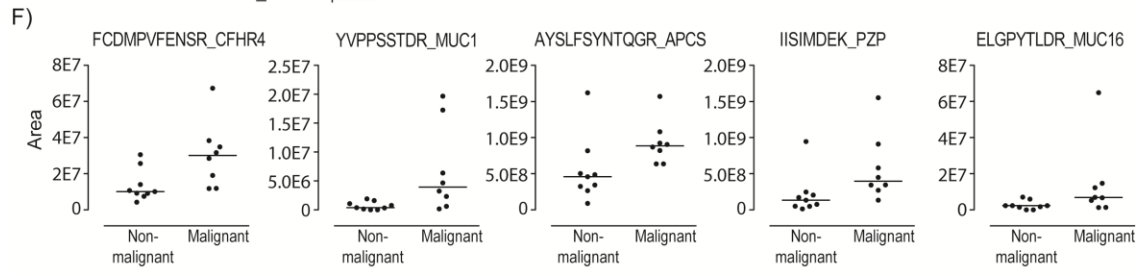
Figure 4.11 Monitoring ascites enriched biomarkers in plasma EVs

EVs were purified by ultracentrifugation from 9 non-malignant and 8 malignant patient plasma samples (~0.5mL/sample). (A) Unfractionated plasma EV preparations were initially analyzed by LC-MS and searched in Peaks®. Protein identifications common to plasma EVs and those significantly enriched in ascites EVs (UC and CD9AP datasets) were retained and used to build a spectral library in Skyline and a scheduled PRM method. A minimum of 3 transitions were required to measure peak areas with most containing 4 or 5. Peak areas were subsequently normalized to hGFP to correct for technical variability between LC-MS runs. (B) Box and whisker plots showing median CV values (horizontal line) between technical duplicates within the acceptable cut-off of 20%. Boxes indicate 75% and 25% quartiles and whiskers represent 90% and 10% ranges and mean CVs are indicated by the '+' sign. Peak areas were further normalized to the CD9 peptide EVQEFYK to account for differences in EV recovery and purity following ultracentrifugation. Peak boundaries for hGFP and CD9 were manually selected to ensure accuracy prior to normalization and non-transformed peaks were used to measure peak areas. (C) Bar plot of ROC-AUC values from peptides with the largest differences between non-malignant and malignant plasma EVs. (D) ROC curves for the top 4 peptides with the largest AUCs and MUC16 (CA-125) are shown. (E) ROC-AUC values, (median) fold-changes and p-values for the top 4 peptides and MUC16. (F) Dot plots comparing peak areas between non-malignant and malignant samples for the top 4 peptides and MUC16. Horizontal lines indicate median peak areas. The top 4 peptides were significantly elevated (p<0.05) in malignant samples however MUC16 was not (p=0.0894).

4.4   Discussion

In this study, we characterized EV proteomes derived from three different sources (cell lines, plasma and ascites) using two different techniques (UC and CD9AP) to identify proteins which may be used as biomarkers for detecting ovarian cancer. Importantly, our findings complement and expand upon previous work by several other groups that also utilized mass spectrometry to characterize ascites or EVs derived from ascites or ovarian cancer cell lines [39,40,55,59,63–67].

Our comparison of EV proteomes from ovarian (cancer) cells suggests the presence intercellular heterogeneity although OVCAR3, OV-90 and EOC6 are likely HGSC, albeit OVCAR3 was not specified in the original publication [68]. Of caution, this heterogeneity may reflect differences in tissue of origin [35,68]. For example, three distinct proteomic expression profiles (epithelial, clear cell and mesenchymal) were identified during a recent large scale proteomic analysis of ovarian cancer lines, HGSC tumours, and FTEC and hIOSE cell lines [35]. From our PCA, it is difficult to ascertain whether any cellular EV proteomes can be stratified according to this criteria. For example, OVCAR3 and FTEC were categorized as epithelial in origin by Coscia *et al.* but in our hands they were dissimilar. Although intra-subtype profiling may have downstream implications for treatment decisions, we opted to focus our efforts on interrogating plasma and ascites EVs which likely harbour high confidence biomarkers.

In total, ~6000 proteins were identified in at least one or more EV preparation from ascites samples obtained by UC or CD9AP purification (4878 proteins in two or more samples). To our knowledge, this is the most in depth ascites EV proteome reported to date. Several factors which enabled us to achieve this level of coverage include the 'match-between-runs' feature in MaxQuant, the recently developed minimal encapsulated SCX StageTip fractionation technology and the use of high resolution Orbitrap-based instrumentation [44,69,70]. Indeed, over 10 years have passed since the first global proteomic analysis of ovarian cancer ascites was

reported by Gortzak-Uzan *et al.* whom identified 229 high confidence proteins in the soluble fraction of ascites from a single patient [64]. Although these authors employed multi-dimensional protein identification technology (MuDPIT) and gel-enhanced fractionation, the presence of HAPs and complexity were noted as potential reasons for their limited detection of low abundance factors such as VEGF and TGF-β [64,71]. To reduce interference associated with HAPs, Kuk *et al.* preferentially isolated low molecular proteins (≤30kDa or ≤100kDa) by size-exclusion chromatography (SEC) or filtration and established a set 52 biomarker candidates not previously reported in ascites [55]. We cross-referenced this dataset with our results for comparison and found 48 out of 52 proteins in ascites EVs; 28 of which were significantly enriched (data not shown). In a follow-up study, this group utilized combinatorial peptide libraries (CPL) on beads containing ~2.4×10^7 unique peptides to deplete HAPs in ascites and enrich for low abundant species [72]. This strategy revealed a number of new ascites biomarkers and enabled the quantification of 30 low abundant proteins by multiple reaction monitoring (MRM). While these studies elucidated numerous biomarkers, it remained to be determined which were specific to malignant versus non-malignant ascites or plasma samples. To better delineate proteins exclusive to or upregulated in EOC, Shender *et al.* compared ascites from patients with ovarian cancer to those with alcohol induced cirrhosis [66]. These authors combined CPL depletion with gel-based fractionation similar to Drabovich *et al.* and identified 424 proteins associated with malignant ascites. We also cross-referenced this dataset with our list of proteins significantly enriched in ascites EVs and observed ~43% overlap (data not shown). The difference in overlap may be attributed to the fact that our proteome analysis was predominately restricted to proteins associated with EVs (and residual HAPs) and may be lacking extracellular signalling factors and ECM components. For example, KLK6 was identified by Shender *et al.* but not present in our dataset [66]. These authors also documented a large number of factors involved in RNA splicing which appeared to be missing from our dataset. To date,

we are not aware of any study which has compared EV depleted ascites to determine which factors are shared or exclusive to each fraction.

In addition to profiling EVs from cell lines, plasma and ascites, we performed PRM analysis on a cohort of patient plasma EVs as a proof-of-concept to verify potential biomarkers for detecting EOC. Although we performed two rounds of ultracentrifugation to dilute HAPs from EVs, we experienced a significant degree of contamination. This combined with the absence of fractionation was a substantial obstacle in the detection of low abundance cancer derived factors. Nonetheless, we identified a subset of peptides associated with malignant (ovarian cancer) EVs from HGSC patients compared to individuals with non-malignant conditions. While MUC16 was higher in malignant samples, it was not significant and among the best performing peptides. Interestingly, CFHR4, one of the top performing candidates, was specifically enriched in the CD9AP, but not UC, comparison. Two isoforms of CFHR4 have been identified and both have been shown to bind and recruit C-reactive protein (CRP) to necrotic cells and tumour tissue [73]. CRP is important for initiating complement activation and opsonisation of dead cells for phagocytosis and clearance. In effect, necrotic tissue produced during EOC progression is a likely culprit for elevated CFHR4 and CRP. CRP was also significantly elevated in malignant plasma EVs and previously reported in EOC, however, to our knowledge, CFHR4 has not been documented [57].

Taken together, we identified several peptides with high confidence that are likely elevated during EOC progression. In tandem with CA-125, these markers may enhance the specificity and sensitivity of blood based assays for detecting EOC. Additional work is required to refine our list of targets and employ heavy synthetic peptides for absolute quantification in a larger cohort patients to monitor levels across multiple stages of EOC

4.5   Materials and Methods

4.5.1   Cell culture

OV-90 (ATCC® CRL-11732) and NIH:OVCAR3 (ATCC® HTB-161) were obtained from the ATCC. hIOSE (OSE364) and immortalized FTEC (FT194) were kindly provided by Dr. Ronny Drapkin (Department of Obstetrics and Gynecology, University of Pennsylvania). Primary cell lines EOC6 and EOC18 were isolated from the ascites of patients with high-grade and low-grade serous ovarian cancer, respectively. All cell lines, except OVCAR3, were maintained in M199+MCDB105 supplemented with 5-15% FBS. NIH:OVCAR3 cells were cultured in RPMI-1640 supplemented with 20% FBS and 5ug/mL insulin. Media was exchanged with serum free media for 20-30 hours to generate conditioned media (CM) for EV purification. All work involving the use of patient samples (cell lines, plasma and ascites) was approved by the Health Research Ethics Board of Alberta-Cancer Committee (Appendix D).

4.5.2   Ultracentrifugation (UC)

CM, plasma and ascites samples were first centrifuged at 200-300 xg at 4°C to pellet cells. Supernatants were transferred to clean tubes, diluted 1:10 in PBS (except CM) and centrifuged at 3,000 xg for 20 minutes at 4°C to remove cell debris. To remove large membrane fragments, supernatants were spun at 10,000 xg for an additional 20 minutes at 4°C. Lastly, supernatants were transferred into clean tubes and ultracentrifuged at 120,000 to 140,000 xg (SW-28 rotor) for 2 hours at 4°C to pellet EVs on an OptimaTM L-100 XP ultracentrifuge (Beckman Coulter). The supernatant was removed and EVs were resuspended in 100-300µL of PBS and stored at -20°C until further use

4.5.3   CD9-affinity purification (CD9AP)

Hydrophilic streptavidin magnetic beads (120mg) were washed 3x with PBS on a magnetic rack then resuspended in 5mL of PBS (New England Biosystems, S1421S, 20mg/5ml). The bead slurry was mixed with 650µg of biotin conjugated anti-CD9 antibody (Abcam ab28094) at room temperature for 30 minutes and then washed 2x with PBS to remove unbound antibody. Beads were resuspended in 6mL of PBS and 1mL (~20mg) was added to 10mL of plasma or ascites (diluted 1:1 in PBS). Samples were placed on a rotary mixer overnight at 4°C and then carefully rinsed 3x with PBS next day. Exosomes were eluted from beads with three, 500 µl glycine-HCl (0.1M, pH 2.39) washes. A small volume (75µL) of Tris-HCl (1.8M, pH 8.54) was used to neutralize each eluent.

4.5.4   EV protein extraction

To prepare EVs for LC-MS/MS, ~20-25µg of protein quantified by micro BCA was lyophilized to dryness and reconstituted in 8M Urea, 50mM ammonium bicarbonate (ABC), 10mM dithiothreitol (DTT), 2% SDS lysis buffer. EV protein samples were sonicated with a probe sonicator (3 X 0.5s pulses; Level 1) (Fisher Scientific, Waltham, MA), reduced in 10mM DTT for 30 minutes and alkylated in 100mM IAA for 30 minutes at room temperature in the dark. Proteins were precipitated in chloroform/methanol in 1.5mL microfuge tubes according to Wessel and Flügge [74]. Briefly, EV samples in lysis buffer were topped up to 150µL with 50mM ABC then mixed with ice cold methanol (600µL) followed by adding ice cold chloroform (150µL) and vortexed thoroughly. An additional volume (450µL) of 4°C water was added followed by vortexing and centrifugation at 14, 000 xg for 5 min. The upper aqueous/methanol phase was carefully removed to avoid disturbing the precipitated protein interphase. A second 450µL volume of cold methanol was added to each sample followed by vigorous vortexing and centrifugation at 14, 000

xg for 5 min. The remaining chloroform/methanol supernatant was discarded and the precipitated protein pellet was left to air dry in a fume hood.

4.5.5   EV protein digestion

On-pellet in-solution protein digestion was performed similarly to Duan *et al.* [75]. Briefly, 100µL of 50mM ABC (pH 8) plus LysC/Trypsin (1:50 ratio) was added to precipitated EV proteins and vortexed vigorously. EV samples were incubated at 37°C overnight (~18h) in a water bath shaker (Polyscience) at 190 rpm or ThermoMixer C (eppendorf) at 300 rpm. An additional volume of trypsin (1:100 ratio) was added the next day for ~4 hours before acidifying to pH 3-4 with 10% FA. EV digests were briefly centrifuged at 14,000 xg to pellet insoluble material prior to LC-MS/MS or SCX peptide fractionation.

4.5.6   SCX peptide fractionation

Tryptic peptides recovered from EV digests were fractionated using SCX StageTips similarly to Kulak *et al.* [69]. Briefly, peptides were acidified with 1% TFA and loaded onto a pre-rinsed 12-plug SCX StageTips (Empore™ Supelco, Bellefonte, PA, USA). In total, 6 SCX fractions were collected by eluting in 75, 125, 200, 250, 300 mM ammonium acetate/20% ACN followed by a final elution in 5% ammonium hydroxide/80% ACN. SCX fractions were dried in a SpeedVac (Thermo Fisher), re-suspended in ddH$_2$O, and dried again to evaporate residual ammonium acetate. All samples were re-suspended in 0.1% FA prior to LC-MS analysis.

4.5.7   LC-MS

SCX fractions were analyzed using an nanoAquity UHPLC M-class system (Waters) connected to a Q Exactive mass spectrometer (Thermo Scientific) using a nonlinear gradient. Buffer A consisted of Water/0.1% FA and Buffer B consisted of ACN/0.1%FA. Peptides (~1µg estimated by BCA) were initially loaded onto an ACQUITY UPLC M-Class Symmetry C18 Trap Column, 5 µm, 180 µm x 20 mm

and trapped for 4 minutes at a flow rate of 10 µl/min at 99% A/1% B. Peptides were separated on an ACQUITY UPLC M-Class Peptide BEH C18 Column (130Å, 1.7µm, 75µm X 250mm) operating at a flow rate of 300 nL/min at 35°C using a non-linear gradient consisting of 1-7% B over 3.5 minutes, 7-19% B over 86.5 minutes and 19-30% B over 30 minutes before increasing to 95% B and washing. Settings for data acquisition on the Q Exactive and Q Exactive Plus are outlined in Table 4.3.

4.5.8   Data Analysis

MS raw files were searched in MaxQuant (1.5.2.8) using the Human Uniprot database (reviewed only; updated May 2014 with 40,550 entries). Missed cleavages were set to 3 and I=L. Cysteine carbamidomethylation was set as a fixed modification. Oxidation (M), N-terminal acetylation (protein), and deamidation (NQ) were set as a variable modifications (max. number of modifications per peptide = 5) and all other setting were left as default. Precursor mass deviation was left at 20 ppm and 4.5 ppm for first and main search, respectively. Fragment mass deviation was left at 20 ppm. Protein and peptide FDR was set to 0.01 (1%) and the decoy database was set to revert. The match-between-runs feature was utilized across all sample types to maximize proteome coverage and quantitation. Datasets were loaded into Perseus (1.5.5.3) and proteins identified by site, reverse and potential contaminants were removed [76].   Protein identifications with quantitative values in ≥2 samples in a least one group (cells, plasma or ascites) were retained for downstream analysis unless specified elsewhere. Missing values were imputed using a width of 0.3 and down shift of 1.8.

4.5.9   Parallel Reaction Monitoring (PRM) assay development

Plasma EVs were obtained and processed as described above with slight modifications. Plasma samples were first diluted ~1:20 in PBS and EVs were pelleted at 120, 000 xg (SW-41 rotor) for 2 hours at 4°C on an OptimaTM L-100

XP ultracentrifuge (Beckman Coulter). To reduce serum contamination, plasma EVs were resuspended in 5mL of PBS and ultracentrifugated a second time at 111,000 xg (MLA-80 rotor) for 2 hours at 4°C on an Optima™ MAX ultracentrifuge (Beckman Coulter). To facilitate digestion, protein pellets were sonicated 1-3 times following precipitation and LysC (Wako) was added (1:100) during overnight digestion. Digests were transferred to pre-rinsed (100µL of 25mM ABC/50% ACN) 10 kDa MWCO microcon YM-10 centrifugal filter units (Millipore) and spun at 14,000xg for 20 min to recover peptides. Centrifugal filter units were washed with an additional 50µL of 25 mM ABC/50% ACN for 15 min at 14,000xg to collect residual peptides with high binding capacity. Filtered samples were dried in a SpeedVac, reconstituted in 0.1%FA and quantified by BCA. Unfractionated plasma EV digests (~1µg/sample) were analyzed on a Q Exactive Plus using a non-linear 2.5h gradient consisting of 1-7% B over 1 minute, 7-23% B over 134 minutes and 23-35% B over 45 minutes before increasing to 95% B and washing. Plasma EV raw files were searched against the human Uniprot databased (20, 274 entries) using the *de novo* search engine Peaks™ (version 8) [61,77]. Parent and fragment mass error tolerances were set to 20 ppm and 0.05 Da, respectively. Max. missed cleavages were set to 3 and 1 non-specific cleavage was allowed. Carbamidomethylation was set as a fixed modification and deamidation, oxidation and acetylation (protein N-term) were included as variable modifications with a maximum of 3 PTMs per peptide allowed. pepXML peptide information and mzXML spectral data were exported from Peaks® prior to building a spectral library in Skyline. Peptides with missed cleavages or containing tryptophan were removed and up to 3 peptides/protein, 7-18 amino acids in length, were chosen for monitoring. In Skyline, peptides with low dotp scores or lacking peak areas were removed prior to exporting isolation lists for PRM. An 8 minute window was chosen to account for deviations in chromatography and minimize the chance of truncation while maximizing the number of MS/MS scans. Heavy glufibribrinopeptide (hGFP) was spiked into each plasma sample which were run in a randomized order in technical duplicate. PRM results were imported into Skyline and transitions (b and

y ions) with interference were removed. A minimum of 3 transitions were used to measure peak areas and targets with dotp scores <0.8 were assumed to be noise and assigned a peak area of 0.

4.5.10 Statistical analysis

Differential protein expression between conditions were determined using a two-tailed Student's t-test (p<0.05) in Perseus (version 1.5.5.3) [76]. ROC-AUCs and Mann-Whitney statistical tests were calculated in in GraphPad Prism version 6.01 (GraphPad Software, San Diego, CA).

Table 4.3 Q Exactive (Plus) instrument parameters for data acquisition

Parameter	Q Exactive	Q Exactive Plus	Q Exactive Plus (PRM settings)
Orbitrap resolution (MS1)	70K	70K	70K
Mass range	400-1500m/z	400-1500m/z	395-1500m/z
MS1 injection time	250ms	250ms	250ms
MS1 AGC taget	3E+06	3E+06	3E+06
Lock mass	445.120025	445.120025	445.120025
MS2 detection	FT	FT	FT
MS2 resolution	17.5K	17.5K	35K
MS2 AGC target	2E+05	2E+05	1E+06
MS2 injection time	64ms	64ms	120ms
Loop count	12	12	30
Isolation width	1.2	1.2	1.2
Isolation offset	0.5m/z	0.5m/z	0.5m/z
MS2 Activation	HCD	HCD	HCD
Normalized Collision Energy	25	25	25
Dynamic exclusion	enabled	enabled	n/a
Minimum AGC target	2.0E+03	2.0E+03	n/a
MS2 intensity threshold	3.1E+04	3.1E+04	n/a
Exclusion duration	30s	30s	n/a
Charge exclusion	unassigned, 1, 7, >8	unassigned, 1, 7, >8	n/a

4.6 References

[1]   Crutchfield, C.A., Thomas, S.N., Sokoll, L.J., Chan, D.W., Advances in mass
        spectrometry-based clinical biomarker discovery. *Clin. Proteomics* 2016, 13, 1.

[2]   Füzéry, A.K., Levin, J., Chan, M.M., Chan, D.W., Translation of proteomic biomarkers into
        FDA approved cancer diagnostics: issues and challenges. *Clin. Proteomics* 2013, 10, 13.

[3]   Diamandis, E.P., Cancer biomarkers: Can we turn recent failures into success? *J. Natl.
        Cancer Inst.* 2010, 102, 1462–1467.

[4]   Howlader, N., Noone, A., Krapcho, M., Miller, D., et al., SEER Cancer Statistics Review,
        1975-2014, National Cancer Institute 2017.

[5]   Torre, L.A., Islami, F., Siegel, R.L., Ward, E.M., Jemal, A., Global cancer in women:
        Burden and trends. *Cancer Epidemiol. Biomarkers Prev.* 2017, 26, 444–457.

[6]   American Cancer Society, Cancer Facts & Figures 2018. *Am. Cancer Soc.* 2018.

[7]   van Nagell, J.R., Hoff, J.T., Transvaginal ultrasonography in ovarian cancer screening:
        current perspectives. *Int. J. Womens. Health* 2013, 6, 25–33.

[8]   Hennessy, B.T., Coleman, R.L., Markman, M., Ovarian cancer. *Lancet* 2009, 374, 1371–
        1382.

[9]   Hayes, D.F., Bast, R.C., Desh, C.E., Fritsche Jr, H., et al., Tumor marker utility grading
        system: a framework to evaluate clinical utility of tumo markers. *J. Natl. Cancer Inst.* 1996,
        88, 1456–1466.

[10]   Atkinson A.J., J., Colburn, W.A., DeGruttola, V.G., DeMets, D.L., et al., Biomarkers and
        surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol.
        Ther.* 2001, 69, 89–95.

[11]   Rifai, N., Gillette, M. a, Carr, S. a, Protein biomarker discovery and validation: the long
        and uncertain path to clinical utility. *Nat. Biotechnol.* 2006, 24, 971–983.

[12]   Felder, M., Kapur, A., Gonzalez-Bosquet, J., Horibata, S., et al., MUC16 (CA125): Tumor
        biomarker to cancer therapy, a work in progress. *Mol. Cancer* 2014, 13, 1–15.

[13]   Nolen, B., Lokshin, A., Protein biomarkers of ovarian cancer : the forest and the trees.
        Futur. Oncol. 2012, 8, 55–71.

[14]   Jacobs, I., Oram, D., Fairbanks, J., Turner, J., et al., A risk of malignancy index
        incorporating CA 125, ultrasound and menopausal status for the accurate preoperative
        diagnosis of ovarian cancer. *Br. J. Obstet. Gynaecol.* 1990, 97, 922–929.

[15]   Skates, S.J., Xu, F.-J.J., Yu, Y.-H.H., Sjövall, K., et al., Toward an optimal algorithm for
        ovarian cancer screening with longitudinal tumor markers. *Cancer* 1995, 76, 2004–10.

[16]   Leung, F., Bernardini, M.Q., Brown, M.D., Zheng, Y., et al., Validation of a novel
        biomarker panel for the detection of ovarian cancer. *Cancer Epidemiol. Biomarkers Prev.*

2016, 25, 1333–1340.

[17] Cramer, D.W., Bast, R.C., Berg, C.D., Diamandis, E.P., et al., Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer Prev. Res.* 2011, 4, 365–374.

[18] Jacobs, I.J., Menon, U., Ryan, A., Gentry-Maharaj, A., et al., Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* 2016, 387, 945–956.

[19] Jacobs, I.J., Menon, U., Progress and challenges in screening for early detection of ovarian cancer. *Mol. Cell. Proteomics* 2004, 3, 355–366.

[20] Moore, R.G., MacLaughlan, S., Bast, R.C., Current state of biomarker development for clinical application in epithelial ovarian cancer. *Gynecol. Oncol.* 2010, 116, 240–245.

[21] Karlsen, M.A., Sandhu, N., Høgdall, C., Christensen, I.J., et al., Evaluation of HE4, CA125, risk of ovarian malignancy algorithm (ROMA) and risk of malignancy index (RMI) as diagnostic tools of epithelial ovarian cancer in patients with a pelvic mass. *Gynecol. Oncol.* 2012, 127, 379–383.

[22] Zhang, Z., Chan, D.W., The road from discovery to clinical diagnostics: Lessons learned from the first FDA-cleared in vitro diagnostic multivariate index assay of proteomic biomarkers. *Cancer Epidemiol. Biomarkers Prev.* 2010, 19, 2995–2999.

[23] Yip, P., Chen, T.H., Seshaiah, P., Stephen, L.L., et al., Comprehensive serum profiling for the discovery of epithelial ovarian cancer biomarkers. *PLoS One* 2011, 6, 1–10.

[24] Høgdall, E., Fung, E.T., Christensen, I.J., Yip, C., et al., Proteomic biomarkers for overall and progression-free survival in ovarian cancer patients. *Proteomics - Clin. Appl.* 2010, 4, 940–952.

[25] Anderson, N.L., The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clin. Chem.* 2010, 56, 177–185.

[26] Anderson, N.L., Anderson, N.G., The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* 2002, 1, 845–867.

[27] Nanjappa, V., Thomas, J.K., Marimuthu, A., Muthusamy, B., et al., Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res.* 2014, 42, D959–D965.

[28] Schwenk, J.M., Omenn, G.S., Sun, Z., Campbell, D.S., et al., The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* 2017, 16, 4299–4310.

[29] Geyer, P.E., Holdt, L.M., Teupser, D., Mann, M., Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* 2017, 13, 942.

[30] Jones, K.A., Kim, P.D., Patel, B.B., Kelsen, S.G., et al., Immunodepletion plasma proteomics by tripleTOF 5600 and Orbitrap elite/LTQ-Orbitrap Velos/Q exactive mass spectrometers. *J. Proteome Res.* 2013, 12, 4351–65.

[31]   Keshishian, H., Burgess, M.W., Gillette, M.A., Mertins, P., et al., Multiplexed, Quantitative Workflow for Sensitive Biomarker Discovery in Plasma Yields Novel Candidates for Early Myocardial Injury. *Mol. Cell. Proteomics* 2015, 14, 2375–2393.

[32]   Keshishian, H., Burgess, M.W., Specht, H., Wallace, L., et al., Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat. Protoc.* 2017, 12, 1683–1701.

[33]   Hughes, C.S., McConechy, M.K., Cochrane, D.R., Nazeran, T., et al., Quantitative Profiling of Single Formalin Fixed Tumour Sections: proteomics for translational research. *Sci. Rep.* 2016, 6, 34949.

[34]   Nagaraj, N., Kulak, N.A., Cox, J., Neuhauser, N., et al., System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* 2012, 11, M111.013722.

[35]   Coscia, F., Watters, K.M., Curtis, M., Eckert, M.A., et al., Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 2016, 7, 12645.

[36]   Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., et al., An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* 2017, 4, 587–599.e4.

[37]   Raposo, G., Stoorvogel, W., Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell Biol.* 2013, 200, 373–383.

[38]   Choi, D.-S., Park, J.O., Jang, S.C., Yoon, Y.J., et al., Proteomic analysis of microvesicles derived from human colorectal cancer ascites. *Proteomics* 2011, 11, 2745–2751.

[39]   Liang, B., Peng, P., Chen, S., Li, L., et al., Characterization and proteomic analysis of ovarian cancer-derived exosomes. *J. Proteomics* 2013, 80, 171–182.

[40]   Sinha, A., Ignatchenko, V., Ignatchenko, A., Mejia-Guerrero, S., Kislinger, T., In-depth proteomic analyses of ovarian cancer cell line exosomes reveals differential enrichment of functional categories compared to the NCI 60 proteome. *Biochem. Biophys. Res. Commun.* 2014, 445, 694–701.

[41]   Harel, M., Oren-Giladi, P., Kaidar-Person, O., Shaked, Y., Geiger, T., Proteomics of microparticles with SILAC Quantification (PROMIS-Quan): a novel proteomic method for plasma biomarker quantification. *Mol. Cell. Proteomics* 2015, 14, 1127–36.

[42]   Goodspeed, A., Heiser, L.M., Gray, J.W., Costello, J.C., Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Mol. Cancer Res.* 2016, 14, 3–13.

[43]   Kowal, J., Arras, G., Colombo, M., Jouve, M., et al., Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl. Acad. Sci. U. S. A.* 2016, 113, E968-77.

[44]   Cox, J., Hein, M.Y., Luber, C. a, Paron, I., et al., Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 2014, 13, 2513–26.

[45]   Cocucci, E., Meldolesi, J., Ectosomes and exosomes: Shedding the confusion between extracellular vesicles. *Trends Cell Biol.* 2015, 25, 364–372.

[46]   Mi, H., Muruganujan, A., Thomas, P.D, PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013, 41, 377–386.

[47]   Mathivanan, S., Simpson, R.J., ExoCarta: A compendium of exosomal proteins and RNA. *Proteomics* 2009, 9, 4997–5000.

[48]   Spizzo, G., Went, P., Dirnhofer, S., Obrist, P., et al., Overexpression of epithelial cell adhesion molecule (Ep-CAM) is an independent prognostic marker for reduced survival of patients with epithelial ovarian cancer. *Gynecol. Oncol.* 2006, 103, 483–488.

[49]   Lin, K., Rubinfeld, B., Zhang, C., Firestein, R., et al., Preclinical development of an anti-NaPi2b (SLC34A2) antibody-drug conjugate as a therapeutic for non-small cell lung and ovarian cancers. *Clin. Cancer Res.* 2015, 21, 5139–5150.

[50]   Banerjee, S., Oza, A.M., Birrer, M.J., Hamilton, E.P., et al., Anti-NaPi2b antibody-drug conjugate lifastuzumab vedotin (DNIB0600A) compared with pegylated liposomal doxorubicin in patients with platinum-resistant ovarian cancer in a randomized, open-label, phase II study. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* 2018, 29, 917–923.

[51]   Sundfeldt, K., Piontkewitz, Y., Ivarsson, K., Nilsson, O., et al., E-cadherin expression in human epithelial ovarian cancer and normal ovary. *Int. J. cancer* 1997, 74, 275–80.

[52]   Auersperg, N., Pan, J., Grove, B.D., Peterson, T., et al., E-cadherin induces mesenchymal-to-epithelial transition in human ovarian surface epithelium. *Proc. Natl. Acad. Sci. U. S. A.* 1999, 96, 6249–54.

[53]   Wu, S., Zhang, D., Zhang, Z.P., Soprano, D.R., Soprano, K.J., Critical role of both retinoid nuclear receptors and retinoid-X-receptors in mediating growth inhibition of ovarian cancer cells by all-trans retinoic acid. *Oncogene* 1998, 17, 2839–2849.

[54]   Schug, T.T., Berry, D.C., Shaw, N.S., Travis, S.N., Noy, N., Opposing Effects of Retinoic Acid on Cell Growth Result from Alternate Activation of Two Different Nuclear Receptors. *Cell* 2007, 129, 723–733.

[55]   Kuk, C., Kulasingam, V., Gunawardana, C.G., Smith, C.R., et al., Mining the ovarian cancer ascites proteome for potential ovarian cancer biomarkers. *Mol. Cell. Proteomics* 2009, 8, 661–669.

[56]   Clinton, G.M., Rougeot, C., Derancourt, J., Roger, P., et al., Estrogens increase the expression of fibulin-1, an extracellular matrix protein secreted by human ovarian cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 1996, 93, 316–320.

[57]   Hefler, L.A., Concin, N., Hofstetter, G., Marth, C., et al., Serum C-reactive protein as independent prognostic variable in patients with ovarian cancer. *Clin Cancer Res* 2008, 14, 710–714.

[58]   Buckanovich, R.J., Sasaroli, D., O'Brien-Jenkins, A., Botbyl, J., et al., Tumor vascular proteins as biomarkers in ovarian cancer. *J. Clin. Oncol.* 2007, 25, 852–861.

[59]   Zhao, Z., Yang, Y., Zeng, Y., He, M., A microfluidic ExoSearch chip for multiplexed exosome detection towards blood-based ovarian cancer diagnosis. *Lab Chip* 2016, 16, 489–496.

[60]   Aibar, S., Fontanillo, C., Droste, C., Rivas, J.D. Las, geNetClassifier classify multiple diseases and build associated gene networks using gene expression profiles Introduction to geNetClassifier 2013.

[61]   MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., et al., Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 2010, 26, 966–968.

[62]   Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., Coon, J.J., Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* 2012, 11, 1475–1488.

[63]   Hanash, S.M., Pitteri, S.J., Faca, V.M., Mining the plasma proteome for cancer biomarkers. *Nature* 2008, 452, 571–579.

[64]   Gortzak-Uzan, L., Ignatchenko, A., Evangelou, A.I., Agochiya, M., et al., A proteome resource of ovarian cancer ascites: Integrated proteomic and bioinformatic analyses to identify putative biomarkers. *J. Proteome Res.* 2008, 7, 339–351.

[65]   Elschenbroich, S., Ignatchenko, V., Clarke, B., Kalloger, S.E., et al., In-depth proteomics of ovarian cancer ascites: Combining shotgun proteomics and selected reaction monitoring mass spectrometry. *J. Proteome Res.* 2011, 10, 2286–2299.

[66]   Shender, V.O., Pavlyukov, M.S., Ziganshin, R.H., Arapidi, G.P., et al., Proteome–Metabolome Profiling of Ovarian Cancer Ascites Reveals Novel Components Involved in Intercellular Communication. *Mol. Cell. Proteomics* 2014, 13, 3558–3571.

[67]   Shapira, I., Oswald, M., Lovecchio, J., Khalili, H., et al., Circulating biomarkers for detection of ovarian cancer and predicting cancer outcomes. *Br. J. Cancer* 2014, 110, 976–83.

[68]   Domcke, S., Sinha, R., Levine, D. a, Sander, C., Schultz, N., Evaluating cell lines as tumour models by comparison of genomic profiles. *Nat. Commun.* 2013, 4, 2126.

[69]   Kulak, N. a, Pichler, G., Paron, I., Nagaraj, N., Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimationa in eukaryotic cells. *Nat. Methods* 2014, 11, 319–24.

[70]   Kelstrup, C.D., Jersie-Christensen, R.R., Batth, T.S., Arrey, T.N., et al., Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* 2014, 13, 6187–95.

[71]   Washburn, M.P., Wolters, D., Yates, J.R., Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001, 19, 242–7.

[72]   Drabovich, A.P., Diamandis, E.P., Combinatorial Peptide Libraries Facilitate Development of Multiple Reaction Monitoring Assays for Low-Abundance Proteins research articles 2010, 1236–1245.

[73] Mihlan, M., Hebecker, M., Dahse, H.M., Hälbich, S., et al., Human complement factor H-related protein 4 binds and recruits native pentameric C-reactive protein to necrotic cells. *Mol. Immunol.* 2009, 46, 335–344.

[74] Wessel, D., Flügge, U.I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984, 138, 141–143.

[75] Duan, X., Young, R., Straubinger, R.M., Page, B., et al., A Straightforward and Highly Efficient Precipitation/On-Pellet Digestion Procedure Coupled with a Long Gradient Nano-LC Separation and Orbitrap Mass Spectrometry for Label-Free Expression Profiling of the Swine Heart Mitochondrial Proteome. *J. Proteome Res.* 2009, 8, 2838–2850.

[76] Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., et al., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 2016, 13, 731–40.

[77] Ma, B., Zhang, K., Hendrie, C., Liang, C., et al., PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003, 17, 2337–2342.

Chapter 5

Embryonic protein NODAL mediates stromal cell chemotaxis to breast cancer cells and broadly regulates secretome composition

5.1 Abstract

The tumour microenvironment (TME), consisting of several stromal cell types including fibroblasts, endothelial cells, multipotent stromal cells (MSC) and immune cells, is an important mediator of breast cancer progression. Breast cancers regulate the composition of the TME by secreting a myriad of factors. One such factor is NODAL, an embryonic morphogen belonging to the Transforming Growth Factor-beta (TGF-□) superfamily. NODAL has been shown to act directly on breast cancer cells and macrophages to induce tumour-promoting phenotypes. NODAL has also been shown to indirectly affect endothelial cell behavior, by supporting the secretion of angiogenic proteins by breast cancer cells. However, the global effects of NODAL on cellular secretomes, have not yet been described. Moreover, the effects of NODAL on other components of the breast cancer microenvironment, including fibroblasts and MSC, have not been explored. Herein, we report that NODAL acts directly on fibroblasts to induce an activated phenotype but was unable to directly signal to MSC. Instead, NODAL caused broad alterations in breast cancer secretome components such as IL6, concomitant with changes in MSC chemotaxis. These results demonstrate the ability of NODAL to impact the breast cancer TME.

5.2   Introduction

Non-transformed stromal cells within the tumour microenvironment (TME) undergo a dynamic reciprocity with cancer cells to drive tumour progression [1]. For example, breast cancers contain a significant proportion of auxiliary cells, including endothelial cells, pericytes, cancer-associated fibroblasts (CAFs), tumour-associated macrophages (TAMs) and additional immune and progenitor cell types, which co-operate to promote processes such as metastasis [2–4]. Extracellular factors, such as cytokines and matrix proteins, mediate the pro-tumorigenic behaviours of stromal cells. For instance, CAF-derived CXCL12/stromal derived factor (SDF-1) can mobilize endothelial progenitor cells (EPCs) to increase vascularization of MCF-7 xenografts [5]. Furthermore, secretion of both CXCL12 and Transforming Growth Factor-beta (TGF-β) by CAFs drives malignant progression by directly affecting Ras-transformed MCF-7 breast cancer xenografts [6].

Mesenchymal stromal cells (MSC) are another cell type frequently associated with neoplastic development [7,8]. MSC primarily arise from the bone marrow (BM) but also reside in most connective tissues. MSC can form bone, cartilage and fat *in vitro* and home to sites of ischemia, injury and inflammation *in vivo* [9]. MSC also exhibit pro-tumorigenic properties. For example, MSC increase tumour size and metastatic potential of MDA-MB-231 breast cancer xenografts via CCL5 secretion and CCR5 activation on breast cancer cells [10]. MSC also acquire CAF-like phenotypes when cultured in tumour conditioned media or mixed with cancer cells in mouse xenografts [11–13]. The mechanisms underlying MSC recruitment are not fully understood; however, up to 20% of CAFs in a mouse model of gastric cancer were derived from BM-MSC in a CXCL6/CXCR6 dependent manner [14]. Moreover, stereotactic body radiation therapy was found to promote MSC recruitment in Lewis Lung Carcinoma (LLC) and B16F10 xenografts via SDF-1/CXCR4 and Platelet-Derived Growth Factor B (PDGFB)/PDGF Receptor β

(PDGFR-β) signalling pathways [15]. MSC engraftment was associated with tumour recurrence and increased pericyte coverage of endothelial cells, hence supporting their roles in neovascularization and vasculogenesis [15]. Therefore, improved characterization of factors involved in MSC homing may provide additional avenues for therapeutic intervention.

Several recent studies have uncovered tumour promoting roles for the TGF-β superfamily member and embryonic morphogen NODAL [16,17]. NODAL expression, while primarily restricted to embryonic development and human embryonic stem cells (hESCs), has been observed in melanoma, glioblastoma, breast, pancreatic and hepatocellular cancers [16,18–22]. In breast cancer, NODAL clinically correlates with stage and vascularization, and has been shown to promote blood vessel formation [17]. Accordingly, NODAL inhibition reduces breast cancer-induced neovascularization and mitigates tumour growth in Nude mice, in part through decreased PDGF and Vascular Endothelial Growth Factor (VEGF) expression [17,23]. Notably, however, in a chick chorioallantoic membrane model, the effects of NODAL knockdown on blood vessel formation could not be rescued by VEGF, suggesting that alternative pro-angiogenic secreted factors are affected when NODAL expression is altered [17].

In this study, we utilized mass spectrometry-based proteomics to analyze the NODAL-regulated secretomes of a claudin-low triple negative breast cancer cell line (MDA-MB-231) and a triple negative inflammatory breast cancer cell line (SUM149) [24,25]. Our analyses revealed cancer cell-type specific alterations in several undocumented NODAL-regulated factors including CXCL1, CXCL8, Interleukin 6 (IL6) and colony-stimulating factor-1 (CSF1); suggesting that NODAL may impact the ability of breast cancer cells to recruit a variety of stromal cell types. As a corollary, we found that NODAL-regulated alterations (knockdown and overexpression) negatively affected the ability of breast cancer conditioned medium to attract MSC. This effect was not due to NODAL itself, but to alterations

in other factors such as IL6. Conversely, we demonstrate that NODAL itself can signal to fibroblasts, increasing their invasion and inducing activation toward a CAF-like phenotype. Collectively, these data reveal a previously unknown role for NODAL in the regulation of breast cancer TME composition.

5.3 Results

5.3.1 Proteomics reveals alterations in the NODAL-regulated breast cancer secretome

Mass spectrometry is a powerful approach for proteomic characterization of cancer cell lines and tissues [26,27]. In this study, we employed high resolution mass spectrometry to identify NODAL-regulated factors in serum-free conditioned media (CM) from breast cancer cells that may act on stromal cells (Figure 5.1). Stable Isotopic Labelling of Amino Acids in Culture (SILAC) was combined with SDS-PAGE fractionation to determine relative changes in secreted proteins from MDA-MB-231 breast cancer cells stably expressing scrambled (shControl) or NODAL knockdown (shNODAL) shRNA (Figure 5.2a). In total, this approach identified over 3200 proteins, which were reduced to ~1300 entries after filtering for proteins annotated with Gene Ontology Cellular Component (GOCC) terms containing "extracellular" and quantified in ≥2 out of 3 biological replicates (Figure 5.2b, ESM5.1). A one-sample, two-sided t-test revealed 122 proteins that were significantly different (p<0.05) between shControl and shNODAL CM (Figure 5.2b, ESM5.1). From this list, 1D annotation enrichment in Perseus revealed a significant decrease (Benjamini Hochberg (BH) FDR threshold<0.02) in proteins involved in GO Biological Processes (GOBPs) associated with cell migration, inflammation and cytokine signalling following NODAL knockdown (Figure 5.2c, ESM5.2) [28]. Alternatively, proteins matching to GOBP terms mRNA processes, protein localization and macromolecular complex disassembly were significantly increased (BH FDR threshold<0.02). This was attributed to higher levels of

ribosomal proteins (RPS and RPL members) shed by shNODAL MDA-MB-231 cells. We plotted Heavy/Light ratios (shNODAL/shControl) and their corresponding $-\log_{10}$ p-values for the ~1300 filtered extracellular proteins found in MDA-MB-231 CM (Figure 5.2d). All proteins annotated with the aforementioned GOBP terms were highlighted in blue (depleted) or red (enriched); there was a clear trend towards a reduction in secretion of inflammatory and chemotactic proteins following NODAL knockdown and an opposing increase in transcriptional and translational proteins. CXCL chemokines (CXCL1/3/8), IL6 and CSF1 were significantly lower in shNODAL CM (p<0.05). Interleukin 11 (IL11), on other the hand, was significantly higher (~1.85 fold, p<0.05). These factors have been associated with malignant phenotypes and may contribute to MSC chemotaxis given that they can promote chemotaxis of various immune cells and, in some cases, MSC [29–31]. Similar to previous findings, PDGFA was significantly lower in shNODAL CM (-2.31 fold) [17].

As a corollary, we performed label-free quantitative proteomics on Strong Cation Exchange (SCX)-fractionated CM digests obtained from Green Fluorescent Protein (GFP) and NODAL overexpressing SUM149 cells (Figure 5.3a). Approximately 1500 proteins were annotated as "extracellular" and quantified in ≥2 out of 3 biological replicates, and 344 proteins were significantly different (two-sided, two-sample t-test, p<0.05) between NODAL and GFP expressing SUM149 cells (Figure 5.3b, ESM5.3). GOBPs that were significantly enriched/depleted (BH FDR threshold<0.02) included terms associated with inflammation, cell migration/locomotion, translation and transcription (Figure 5.3c, ESM5.4). Unexpectedly, GOBPs depleted in shNODAL MDA-MB-231 samples were also depleted in NODAL overexpressing SUM149 CM. For example, proteins matching to the cytokine- mediated signalling pathway had a mean $\log_2$ fold-change of -2.28 and -2.44 following NODAL knockdown and overexpression, respectively. Conversely, proteins matching to "mRNA metabolic process" were increased significantly by NODAL knock down in MDA-MB-231 and NODAL over-expression

in SUM149 with mean $\log_2$ fold-changes of 1.51 and 1.88, respectively. We also plotted $\log_2$ protein fold-changes for SUM149 secretomes (NODAL-GFP) versus – $\log_{10}$ p-values and highlighted all proteins annotated with the aforementioned GOBPs (Figure 5.3d). Several inflammatory and migratory factors decreased following NODAL overexpression while translational and transcriptional proteins were elevated. For instance, CXCL1 and 3, IL6 and CSF1 levels decreased following NODAL knockdown in MDA-MB-231 cells and NODAL overexpression in SUM149 cells. Although PDGFA was not detected in SUM149 CM, the angiogenic factors Angiopoietin-1 (ANGPT1) and Angiogenin (ANG) were significantly elevated in CM from NODAL overexpressing SUM149 cells [32,33]. Highly similar proteomic results were also observed when comparing CM from NODAL overexpressing SUM149 cells to cells expressing an empty vector (EV) (Figure 5.4, ESM5.5 and ESM5.6). In total, 56 proteins were significantly altered by NODAL in both MDA-MB-231 and SUM149 datasets; however, only a handful were associated with NODAL expression in a positive (CLU and CLSTN3) and negative (Leukemia Inhibitory Factor [LIF] and Neuropillin-2 [NRP2]) manner in both cell lines. To verify the proteomic findings, ELISAs were performed with CM from MDA-MB-231 cells for CXCL1, CXCL8, IL6 and CSF1 and CM from SUM149 cells for CXCL1 and IL6 (Figure 5.5a). For reference, CXCL1 and IL6 levels were substantially higher in GFP expressing SUM149 cells compared to MDA-MB-231 cell lines (Figure 5.5b and c). In effect, NODAL appears to influence a subset of cellular processes involved in inflammation, motility, transcription and translation, albeit differentially based on the breast cancer line.
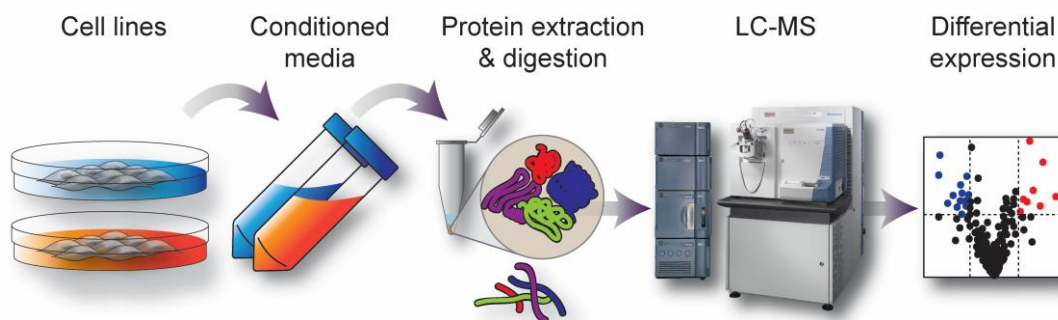
Figure 5.1 MS-based proteomics workflow for interrogating the NODAL regulated secretome in breast cancer.

Serum free conditioned media (CM) were isolated from MDA-MB-231 (shControl and shNODAL) or SUM149 (EV, GFP and NODAL overexpressing) breast cancer cell lines. Extracellular proteins from MDA-MB-231 cell lines were concentrated from CM, fractionated by SDS-PAGE and digested into peptides. For SUM149 cell lines, extracellular proteins were concentrated, digested into peptides, and fractionated with SCX StageTips. Fractions were analyzed by liquid chromatography-mass spectrometry to detect secreted proteins altered by NODAL.

a

0.5% | 10%
SFM | BSA | FBS

shControl | shNODAL | shControl | shNODAL | shCotrol | shNODAL

kDa
39 — NODAL
42 — ACTIN

b

3273 — Total protein identifications
1299 — "Extracellular" and quantified proteins
122 — Significantly different proteins

c

Significant proteins
0   5   10

positive regulation of cell migration
inflammatory response
cytokine-mediated signaling pathway

0   -1   -2   -3
Mean log$_2$ fold-change

Significant proteins
0   10   20   30

establishment of protein localization to organelle
mRNA metabolic process
cellular macromolecular complex disassembly

0   1   2
Mean log$_2$ fold-change

d

-Log$_{10}$ T-test p-value vs Log$_2$ Heavy/Light Ratio (shNODAL/shControl)

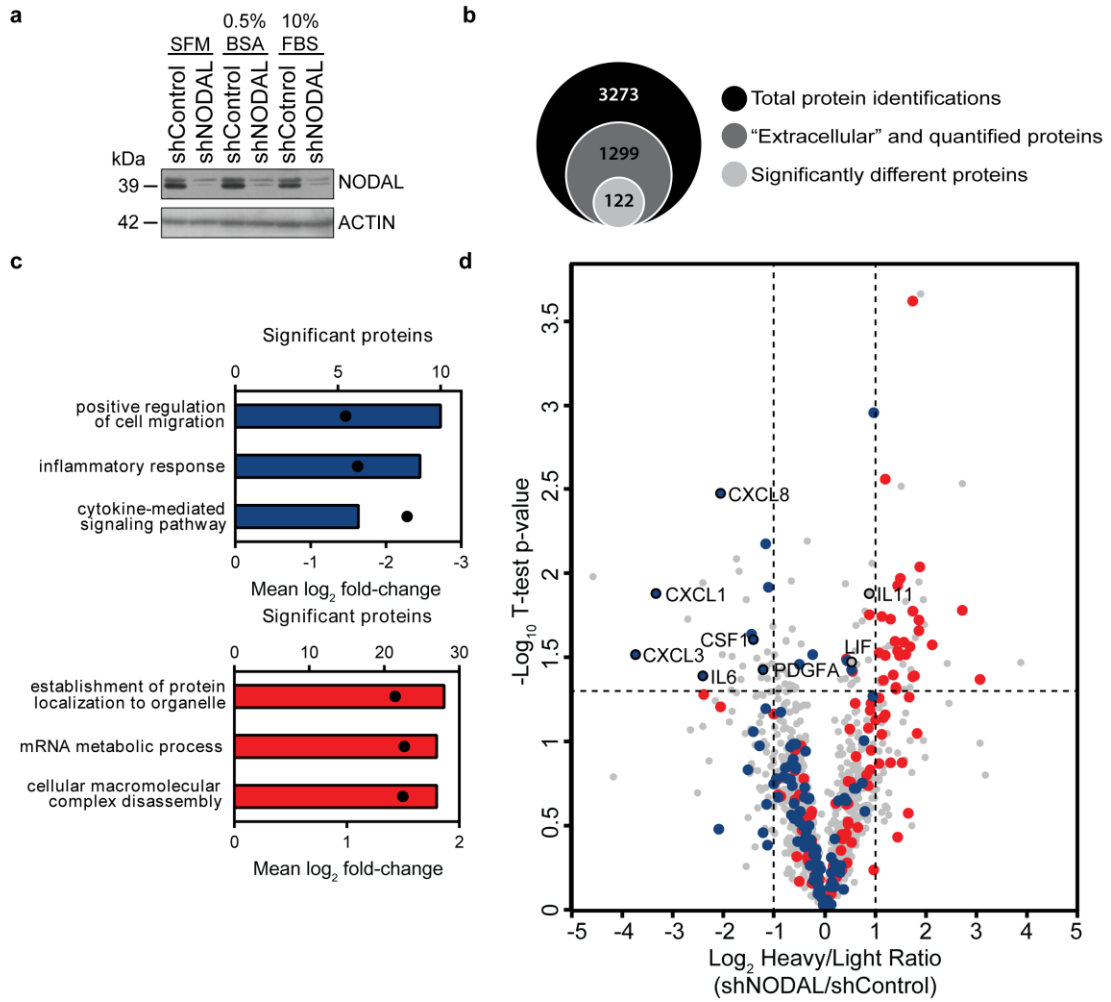CXCL8, CXCL1, CXCL3, CSF1, IL6, PDGFA, IL11, LIF

Figure 5.2 NODAL knockdown alters the MDA-MB-231 secretome.

(a) NODAL expression in lysates from MDA-MB-231 breast cancer cells stably expressing scrambled (shControl) or NODAL knockdown (shNODAL) shRNA cultured in serum free media (SFM), SFM+0.5%BSA (0.5%BSA), or complete media (10% FBS). Extracellular proteins from serum free, isotopically (SILAC) labelled shControl and shNODAL conditioned media (CM) were concentrated using 3kDa MWCO filter units, fractionated and in-gel digested using SDS-PAGE, and analyzed on a Q Exactive mass spectrometer. Raw files were searched in MaxQuant and protein lists were filtered and annotated in Perseus. **(b)** Venn diagram highlighting total protein identifications, number of "extracellular" and quantified proteins, and significantly different proteins between shControl and shNODAL CM. A two-tailed, one sample t-test was used to identify differentially expressed proteins ($p<0.05$). **(c)** Number of significant proteins (bars) matching to a subset of significantly enriched GO biological processes (GOBPs). Mean $\log_2$ fold-changes in GOBPs are indicated by black dots. Blue and red bars highlight GOBPs decreased and increased in MDA-MB-231 CM following NODAL knockdown, respectively. **(d)** Volcano plot of quantified "extracellular" proteins. Negative and positive $\log_2$ Heavy/Light ratios indicate proteins decreased and increased in MDA-MB-231 CM following NODAL knockdown, respectively ($n=3$). All proteins matching to corresponding GOBPs mentioned are highlighted in blue and red. Several cytokines and chemokines altered by NODAL are labelled in black. Vertical and horizontal dotted lines indicate $\log_2$ fold-changes ≥ 2 and the $-\log_{10}$ p-value cut-off corresponding to $p<0.05$, respectively.
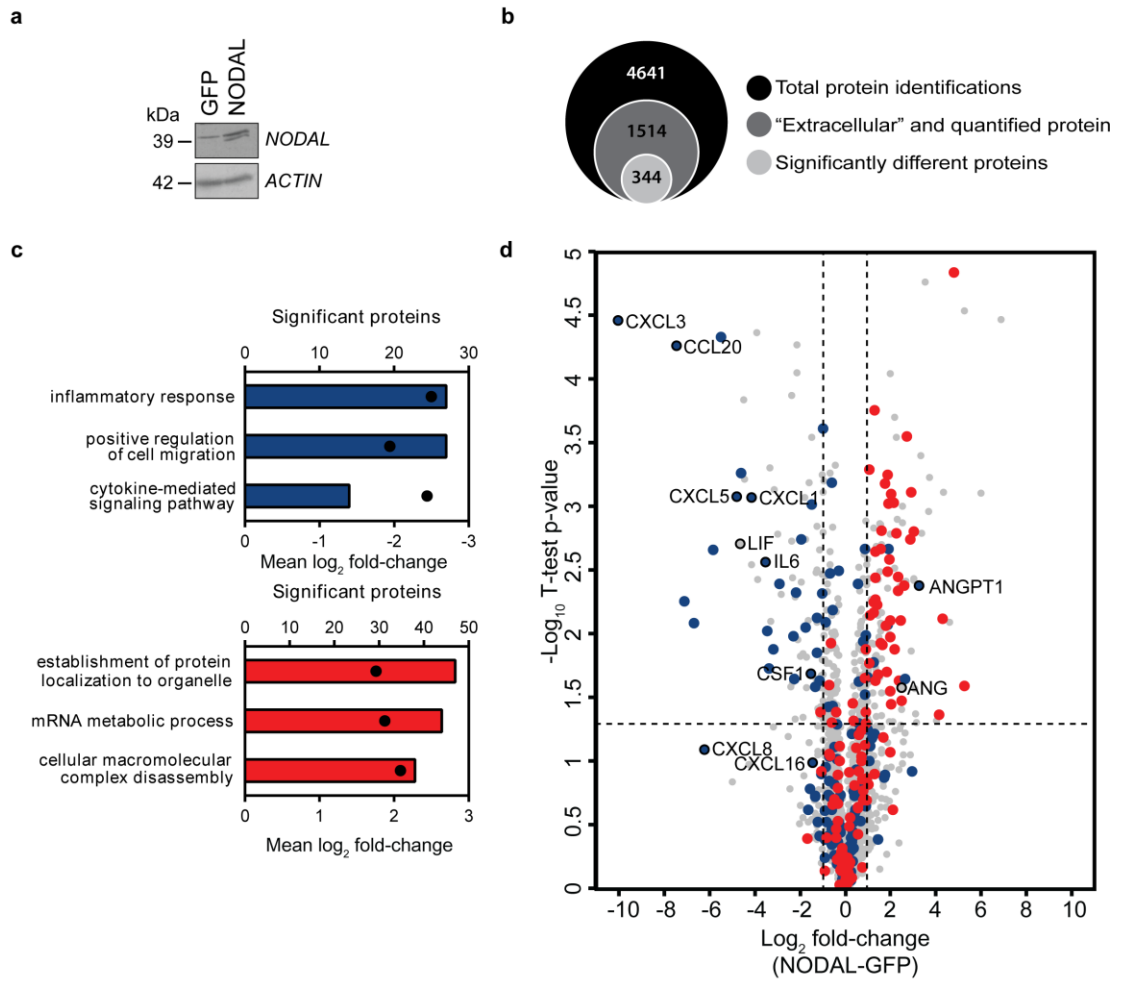
a

b

c

d

Figure 5.3 NODAL overexpression alters the SUM149 secretome.

(a) NODAL expression in lysates from GFP or NODAL over-expressing SUM149 breast cancer cells. Extracellular proteins from serum free CM (GFP or NODAL) were concentrated using 3kDa MWCO filter units, digested with trypsin, fractionated using SCX StageTips and analyzed on an Orbitrap Elite mass spectrometer. Raw files were searched with MaxQuant and protein lists were filtered and annotated in Perseus. **(b)** Venn diagram highlighting total protein identifications, "extracellular" and quantified proteins, and significantly different proteins between GFP and NODAL CM. A two-tailed, two sample t-test was used to identify differentially expressed proteins ($p<0.05$).   **(c)** Number of significant proteins (bars) matching to subset of significantly enriched GOBPs. Mean $\log_2$ fold-changes in GOBPs are indicated by black dots.   Blue and red bars highlight GOBPs decreased and increased in SUM149 CM following NODAL overexpression. **(d)**   Volcano plot of quantified "extracellular" proteins. Negative and positive $\log_2$ fold-changes indicate proteins decreased and increased in SUM149 CM following NODAL overexpression, respectively (n=3). All proteins matching to corresponding GOBPs mentioned are highlighted in blue and red. Several cytokines, chemokines and growth factors altered by NODAL are labelled in black. Vertical and horizontal dotted lines indicate $\log_2$ fold-changes ≥ 2 and the $-\log_{10}$ p-value cut-off corresponding to $p<0.05$, respectively.
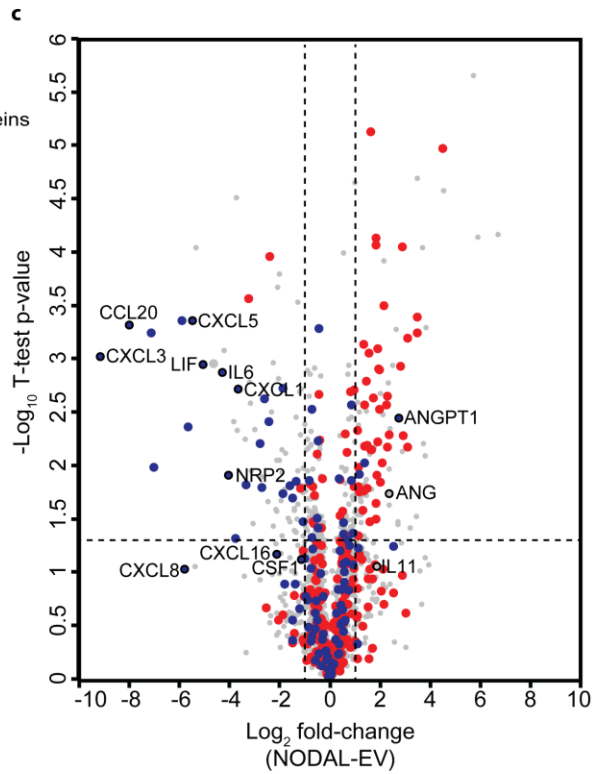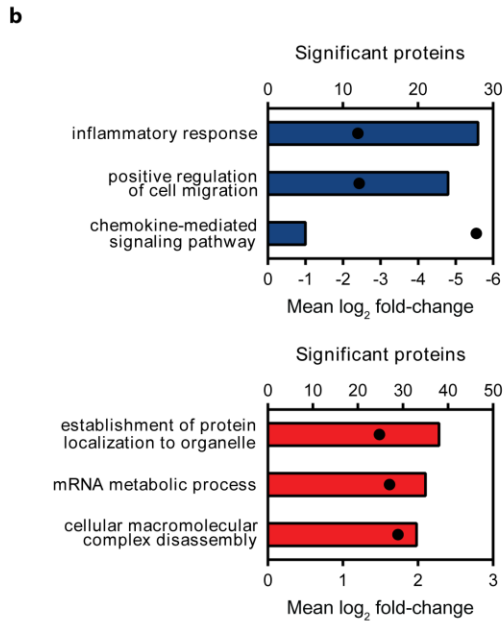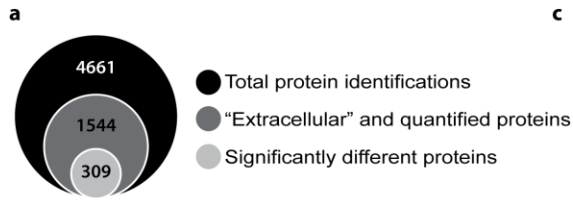
Figure 5.4 NODAL overexpression alters the SUM149 secretome compared to an empty vector control.

Extracellular proteins (CM) from empty vector (EV) or Nodal overexpressing SUM149 cell lines were concentrated using 3kDa MWCO filter units, digested with trypsin, fractionated using SCX StageTips and analyzed on an Orbitrap Elite mass spectrometer. Raw files were searched with MaxQuant and protein lists were filtered and annotated in Perseus. **(a)** Venn diagram highlighting total protein identifications, "extracellular" and quantified proteins, and significantly different proteins between GFP and NODAL CM. A two-tailed, two sample t-test was used to identify differentially expressed proteins (p<0.05). **(c)** Number of significant proteins (bars) matching to significantly enriched GO biological processes (GOBPs). Mean $\log_2$ fold-changes in GOBPs are indicated by black dots. Blue and red bars highlight GOBPs decreased and increased in SUM149 CM following NODAL overexpression. **(d)** Volcano plot of quantified "extracellular" proteins. Negative and positive $\log_2$ fold-changes indicate proteins decreased and increased in SUM149 CM following NODAL overexpression, respectively (n=3). All proteins matching to corresponding GOBPs mentioned are highlighted in blue and red. Several cytokines, chemokines and growth factors altered by NODAL are labelled in black. Vertical and horizontal dotted lines indicate $\log_2$ fold-changes ≥2 and the $-\log_{10}$ p-value cut-off corresponding to p<0.05, respectively.

Figure 5.5 Cytokine and chemokine levels in breast cancer CM.

Cytokine and chemokine levels in breast cancer CM. **(a)** SILAC quantification in MaxQuant and ELISAs report similar fold-changes in secreted factors between shControl and shNODAL MDA-MB-231 cells. **(b and c)** ELISAs reveal substantially higher levels of CXCL1 and IL6 in CM derived from GFP expressing SUM149 cells compared to MDA-MB-231 cell lines. Black and grey dots indicate replicate values.

5.3.2 *The NODAL-regulated breast cancer secretome impacts MSC chemotaxis and NODAL induces fibroblast activation*

The breadth of changes in the NODAL-regulated breast cancer secretome indicate that this morphogen may affect components of the TME *via* both direct and indirect mechanisms. Given the involvement of MSC in tumour growth and neovascularization, we first examined how NODAL affects the capacity of breast cancer cells to promote MSC chemotaxis. We first compared the ability of CM from shControl and shNODAL knockdown MDA-MB-231 breast cancer cells to influence MSC chemotaxis (Fig. 5.6a-d). Several primary human bone marrow (BM)-derived MSC lines were utilized herein, some of which have been previously shown to form tubes *in vitro* and stimulate islet regeneration and revascularization *in vivo* [34,35]. Compared to shControl CM, in 3 out 4 MSC lines, chemotaxis was significantly decreased (~1.8 to 3.5 fold, p<0.0001, Dunnett's multiple comparison test) towards shNODAL CM. We did not observe appreciable differences in proliferation or viability of MSC cultured in CM for 24h, suggesting that the effects observed were not due to alterations in cell numbers, but rather a result of altered chemotaxis (Figure 5.7a; data not shown). As a corollary, we investigated whether CM derived from empty vector (EV), GFP and NODAL overexpressing SUM149 breast cancer cells could also affect MSC chemotaxis (Figure 5.6e). In accordance with our proteomics results, CM from NODAL overexpressing SUM149 cells induced less chemotaxis in MSC2 cells compared to the GFP expressing control. Again, we confirmed by flow cytometry that differences in chemotaxis were not due to altered proliferation or viability (Figure 5.7b; data not shown).

The reduction in MSC chemotaxis observed when NODAL was knocked down could not be rescued by the addition of 100ng/mL of recombinant human NODAL

(rhNODAL) (Figure 5.6 a,b and d) suggesting that MSC are unable to sense this morphogen, perhaps due to an absence of receptor components. Hence, we performed Real time-PCR and Western blotting for NODAL, its receptor (ALK4) and co-receptor (CRIPTO) on two MSC lines (Figure 5.84a and b). MSC expressed moderate levels of NODAL and high levels of ALK4 at the transcript and protein level (Figure 5.8a and b). CRIPTO mRNA expression approached the reliable limit of detection by quantitative real-time PCR (35 cycles). Hence, while MSC appear to make NODAL and to express NODAL receptors, they may not express enough CRIPTO to sense NODAL. Indeed, stimulation with 10 and 100 ng/mL rhNODAL had no effect on canonical or non-canonical signalling through SMAD2 or ERK1/2 phosphorylation, respectively (Figure 5.8c).

Fibroblasts are critical components of the TME and studies indicate CAFs can originate from BM-derived populations including MSC [14,36]. We examined whether NODAL affects breast cancer-induced fibroblast phenotypes by performing chemotaxis assays using primary Human Foreskin Fibroblasts (HFFs) (Figure 5.9a and b). We did not detect differences in HFF chemotaxis towards CM from shControl and shNODAL MDA-MB-231 cells; however, CM from NODAL overexpressing SUM149 cells significantly increased HFF chemotaxis compared to the GFP expressing control (two-sample t-test, p<0.001). We therefore asked if NODAL could directly promote fibroblast activation. Indeed, rhNODAL (10 and 100 ng/mL) increased HFF chemotaxis (Figure 5.9c), invasion (Figure 5.9d) (Dunnett's multiple comparison test, p<0.05) and proliferation (Figure 5.9e). In contrast to MSC, which did not respond to NODAL, we found that rhNODAL (10 and 100ng/mL) caused an increase in both SMAD2 and ERK1/2 activation in fibroblast cells (Figure 5.9f). In addition, Real time RT-PCR revealed that rhNODAL (10 and 100ng/mL) induced expression of α-Smooth Muscle Actin (α-SMA), Desmin and Connective Tissue Growth Factor (CTGF) (Figure 5.9g). We performed gene

expression profiling on human dermal fibroblasts (HDFs) treated with 10ng/mL rhNODAL for 6h. Transcripts upregulated by at least 1.7 fold were analyzed in DAVID and gene clusters associated with the GO terms "wound healing", "cell motion", "extracellular matrix" and "growth factor" were significantly enriched (Figure 5.9h; ESM5.7) [37]. We overlapped proteins differentially expressed from the proteomics and microarray datasets, and found several factors to be consistently altered by NODAL, albeit some inversely correlated with NODAL levels (Figure 5.9i). IL6, LIF and NRP2 were shared amongst all three datasets; however, CXCL1/3 appeared be exclusive to breast cancer cells. Hence, while NODAL indirectly affects MSC chemotaxis by altering the breast cancer secretome, NODAL can directly induce fibroblast activation. Moreover, certain key factors, such as IL6 and LIF, are commonly affected by NODAL in all cell types investigated here.
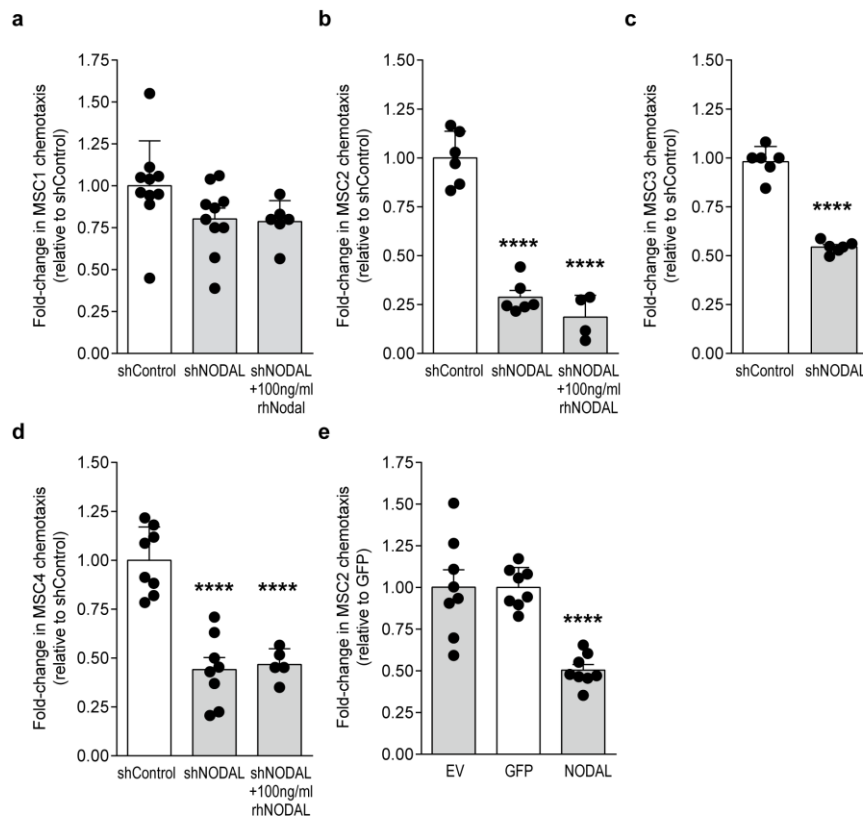
Figure 5.6 CM from NODAL expressing breast cancer cells indirectly modulates MSC migration.

Human bone-marrow derived MSC lines (MSC1-4) were plated onto fibronectin-coated transwells in the presence CM (shControl or shNODAL +/-100ng/mL recombinant human NODAL; rhNODAL). **(a-d)** MSC chemotaxis was quantified after ~24h and was significantly lower ($p<0.05$) towards shNODAL CM compared to shControl CM. Moreover, rhNODAL could not rescue MSC chemotaxis. **(e)** Paradoxically, CM from NODAL overexpressing SUM149 cells was less chemotactic compared to empty vector (EV) and GFP controls. Data are presented as mean fold-changes relative to controls from a minimum of 3 biological replicates ± standard deviation (SD). Black dots indicate replicate values and asterisks indicate significance differences (one way ANOVA, Dunnett's multiple comparison test) in MSC chemotaxis compared to shControl or GFP conditions (** $p<0.01$, **** $p<0.0001$).
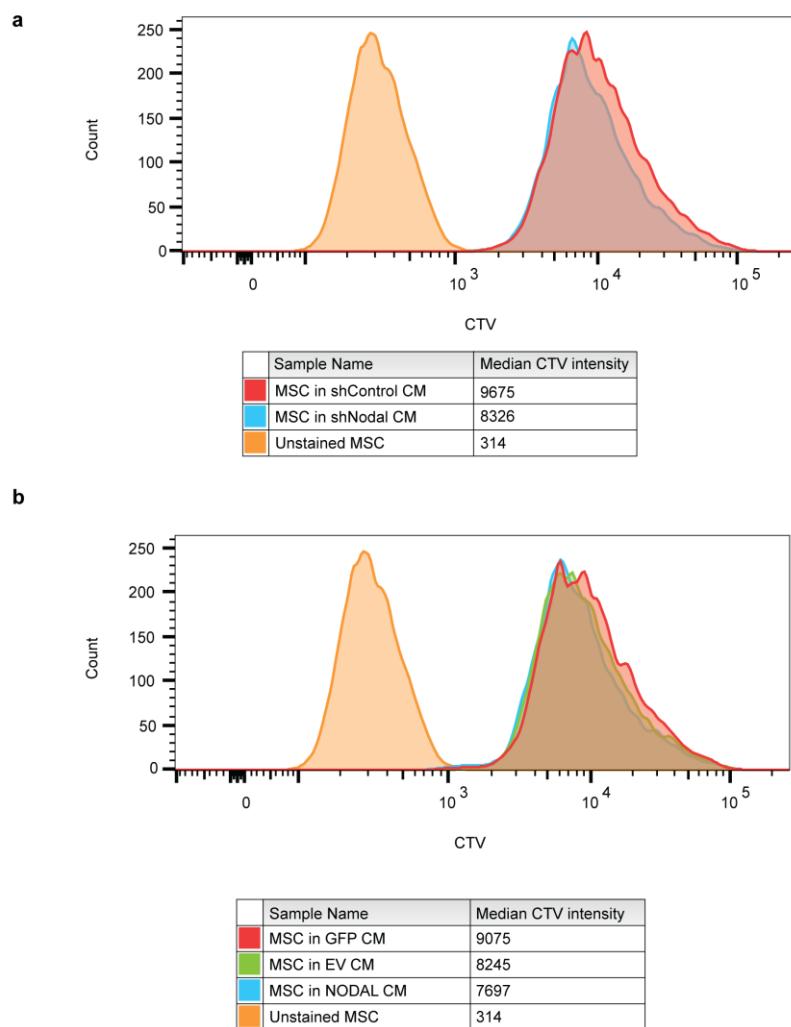
Figure 5.7 MDA-MB-231 and SUM149 CM have a negligible effect on MSC proliferation.

MSC were labelled with cell trace violet (CTV) in suspension and plated onto tissue culture plates. After ~48h, MSC were rinsed in PBS and cultured for additional 24h in CM from MDA-MB-231 (shControl and shNODAL) or SUM149 (EV, GFP and NODAL) cells. MSC were then trypsinized, pelleted and resuspended in 5% FBS/PBS. **(a and b)** Histograms and tables showing median CTV intensity of MSC after 24h in CM from MDA-MB-231 and SUM149 conditions. Histograms are representative images from 3 biological replicates.
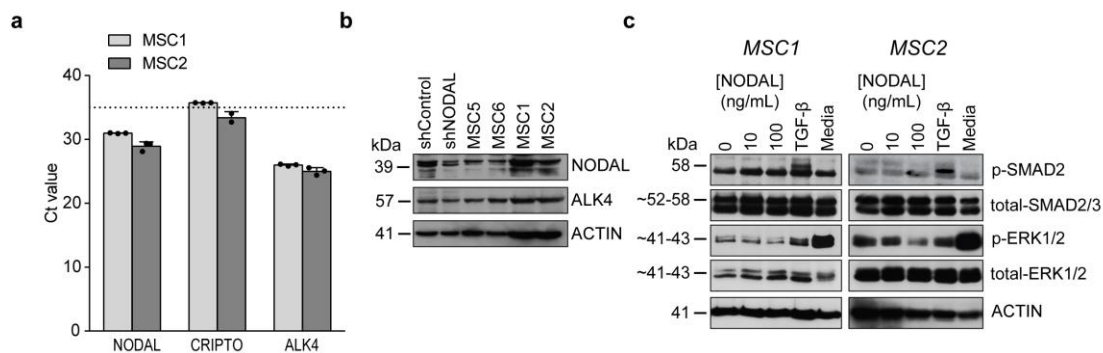
Figure 5.8 NODAL signalling in MSC

(a) Real time PCR cycle threshold (Ct) values for NODAL, ALK4 and Cripto in MSC. **(b)** Western blots showing expression of NODAL and ALK4 (receptor) in lysates from four MSC lines. Lysates from shControl and shNODAL MDA-MB-231 breast cancer cells were used as positive controls. **(c)** Serum starved MSC treated with varying concentrations (ng/mL) of recombinant human NODAL (rhNODAL) for 1h had no effect on downstream SMAD2 (p-SMAD2) or ERK1/2 (p-ERK1/2) activation. TGF- β (10ng/mL for 30mins) and cell culture media were used as positive controls for SMAD2 and ERK1/2 activation, respectively. Data are presented as mean Ct values ± SD from 3 biological replicates except for Cripto (n=2 for MSC2). High Ct values indicate low transcript expression with the horizontal dotted line corresponding to a Ct value of 35 or the reliable limit of detection. Western blots are representative images taken from 3 biological replicates.
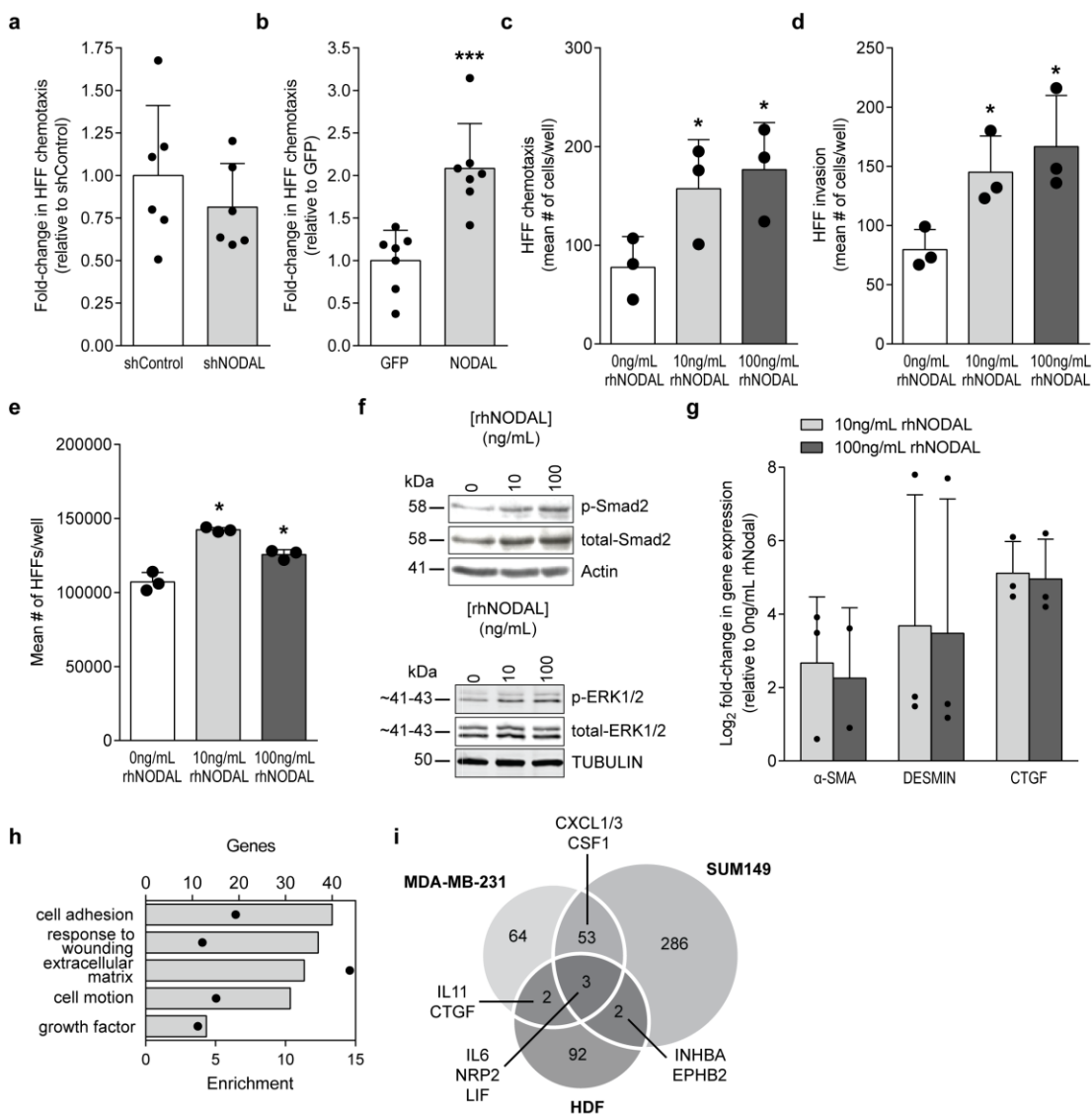
Figure 5.9 NODAL directly promotes phenotypes associated with activated fibroblasts in HFFs.

(a and b) HFF chemotaxis towards CM from MDA-MB-231 and NODAL SUM149 breast cancer cells (n=6). **(c to e)** Exposure to rhNODAL (10 and 100ng/mL) for 24h significantly increased HFF chemotaxis, invasion and proliferation (n=3). **(f)** Stimulation with rhNODAL (10 and 100 ng/mL) for 30mins dose dependently activates SMAD2 and ERK1/2 phosphorylation in HFFs. **(g)** HFFs upregulate transcripts (a-SMA, DESMIN and CTGF) associated with activated fibroblasts following treatment with rhNODAL (10 and 100ng/mL) for 72h (n=3, n=2 for α-SMA from 100ng/mL treatment). **(h)** Total genes (bars) upregulated by NODAL (10ng/mL) more than 1.7 fold in human dermal fibroblasts (HDFs) after 6h treatment and their corresponding enrichment (black dots) following GO analysis in DAVID. **(i)** Overlap in proteins differentially expressed (increased or decreased) in MDA-MB-231 (shControl versus shNODAL), SUM149 (NODAL versus GFP) and HDF (treated versus untreated) datasets. Data are presented as mean fold-changes relative to controls or mean values ± SD. Black dots indicate replicate values and asterisks indicate significance differences (one way ANOVA, Dunnett's multiple comparison test) in MSC chemotaxis compared to controls (* $p<0.05$, *** $p<0.001$). Conditions stimulated with 10ng/mL and 100ng/mL rhNODAL were significantly different when compared using a paired two-tailed, two sample t-test ($p<0.05$).

5.3.3   IL6 promotes MSC migration

Given that NODAL consistently altered CXCL1 and IL6 levels in breast cancer CM, concomitant with differential MSC chemotaxis, we sought to determine whether receptors for these ligands were expressed by MSC. While three MSC lines were highly positive for IL6R based on flow cytometry (Figure 5.10a, Figure 5.11a and b), surface CXCR1 and CXCR2 expression could not be detected on 4 MSC lines by real-time PCR or flow cytometry (Figure 5.11c; data not shown). Accordingly, treatment with 10 and 25ng/mL recombinant human IL6 (rhIL6) induced STAT3 phosphorylation in MSC2 cells, which could be blocked by the addition of an IL6 neutralizing monoclonal antibody (mAb, Figure 5.10b). Moreover, low doses of rhIL6 (1 and 10ng/mL) significantly increased MSC2 chemotaxis by ~1.6 fold (Dunnett's multiple comparison test, $p < 0.05$) although higher concentrations had no effect (Figure 5.10c). Neutralizing IL6 in shControl CM or supplementing shNODAL CM with rhIL6 (1ng/mL) resulted in a small, but significant, reduction and increase in MSC2 chemotaxis, respectively (Figure 5.10d and 6e). These findings suggest that IL6 may be involved in promoting MSC recruitment to breast cancers. Notably, IL6 levels in CM from shControl MDA-MB-231 cells were far lower than in CM from GFP expressing SUM149 cells (~40pg/mL versus ~800pg/mL), suggesting different regulatory mechanisms between these cell lines, which could explain the differential effects of NODAL.
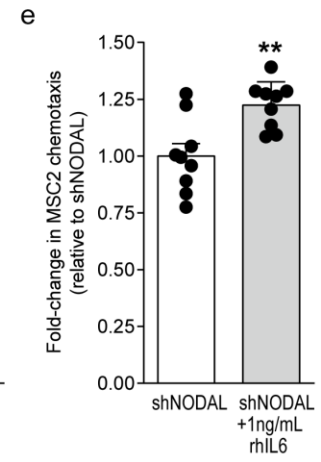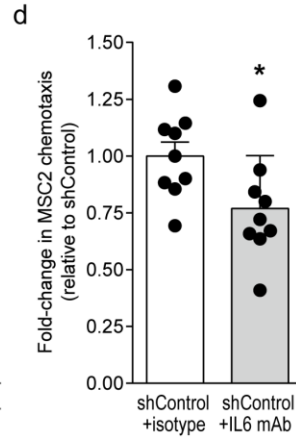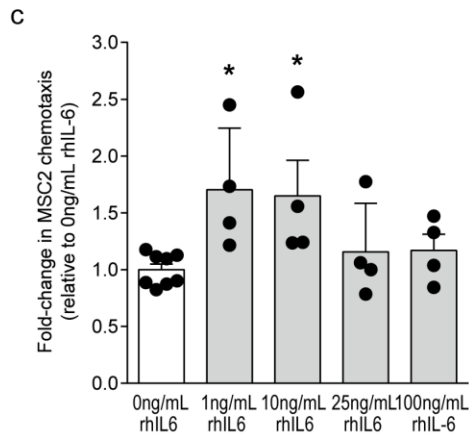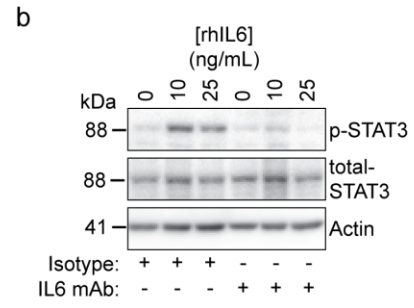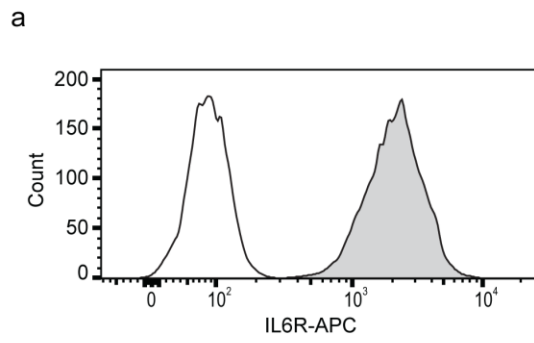
Figure 5.10 IL6 contributes to MDA-MB-231 mediated MSC chemotaxis.

(a) Flow cytometry showing nearly homogenous expression of the IL6 receptor (IL6R) by MSC. **(b)** Stimulation with rhIL6 (10 and 25ng/mL) for 30 minutes induced phosphorylation of STAT3 in MSC2 which could be blocked by a 5 minute pre-incubation with an IL6 neutralizing mAb (2.5µg/mL). **(c)** MSC chemotaxis towards 0, 1, 10, 25 or 100ng/mL recombinant human IL6 (rhIL6) after 24h (n=4-8). Low concentrations (1 and 10ng/mL) of rhIL6 significantly induced MSC chemotaxis. **(d)** Neutralizing endogenous IL6 in MDA-MB-231 CM with an IL6 mAb (2.5µg/mL) significantly attenuates MSC chemotaxis. (e) Exogenous rhIL6 (1ng/mL) significantly increased MSC chemotaxis towards shNODAL CM. Flow histogram and Western blots are representative images from 3 biological replicates. Data are presented as mean fold-changes relative to controls ± SD. Black dots indicate replicate values and asterisks indicate significance differences (one way ANOVA, Dunnett's multiple comparison test for IL6 dose response and two tailed, two sample t-test for MDA-MB-231 treatments) in MSC chemotaxis compared to controls (* $p<0.05$, ** $p<0.01$).
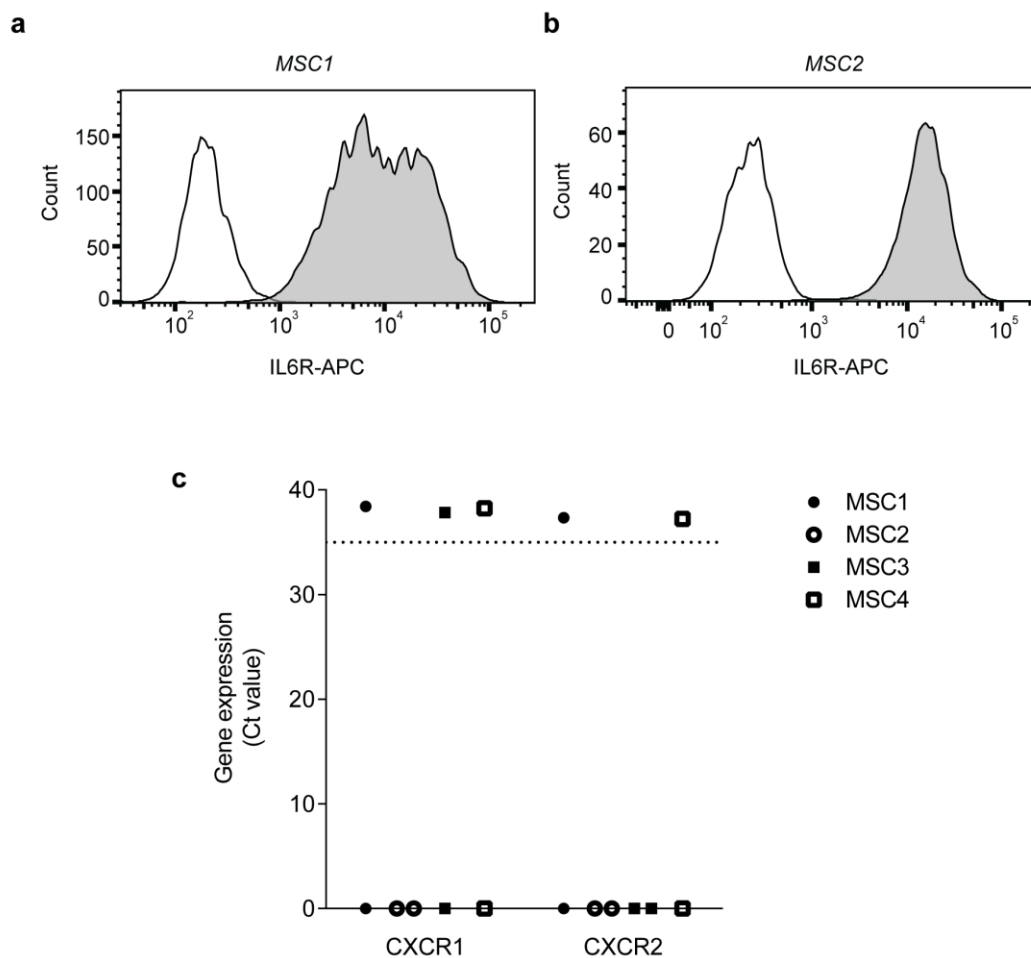
Figure 5.11 MSC express IL6R but lack *CXCR1/2* gene expression.

(a and b) Flow cytometry showing high surface expression of IL6R (grey region) on two additional MSC lines compared to unstained MSC (white region). **(c)** Real-time PCR results reveal no detectable transcript or reliable gene expression for *CXCR1* or *CXCR2* in four primary MSC lines. Individual data points correspond to real-time PCR readings from 2 biological replicates. A Ct value of 0 indicates no transcript was detected. The dotted line indicates a Ct value of 35 or the reliable limit of detection. Histograms are representative images from 3 biological replicates.

5.3.4 Differential signalling pathways may dictate cell-type dependent effects of NODAL

NODAL/ACTIVIN regulates cell fate specification and phenotype by activating signal transduction pathways that directly affect transcription and/or mediate epigenetic modifications [38]. The ability of NODAL to broadly affect gene expression is context-dependent. These differential responses may be due, in part, to which signal transduction pathways are induced by NODAL. Canonically, NODAL triggers phosphorylation of SMAD2/3 via binding to its receptors ActRIIB/ALK(4/7) and co-receptor CRIPTO [39]. Phospho-SMAD2/3-SMAD4 heterodimers subsequently translocate into the nucleus to regulate the epigenetic status and transcription of target genes. NODAL can also signal non-canonically to activate ERK1/2, which is required for the induction of Epithelial-to-Mesenchymal Transition (EMT) and invasion [40].

Given the disparate effects of NODAL on cytokine secretion in MDA-MB-231 versus SUM149 cells, we hypothesized that NODAL may activate different signalling mediators in a cell-type dependent manner like TGF-β. Accordingly, the activation of two documented mediators of NODAL signalling (SMAD2/3 and ERK1/2) were measured by Western blotting in breast cancer cells wherein NODAL levels were modified and then cells were cultured under serum free conditions for 24 hours (Figure 5.12). NODAL knockdown in MDA-MB-231 resulted in an expected and previously described reduction in both SMAD2/3 and ERK1/2 phosphorylation [40]. While overexpression of NODAL in SUM149 increased SMAD2 phosphorylation, a small reduction in ERK1/2 phosphorylation was observed. Moreover, constitutive SMAD2 and ERK1/2 activation was respectively higher and lower in SUM149 cells as compared to MDA-MB-231 cells. We also probed for p38 activation, which is regulated by TGF-β. NODAL expression was associated with decreased p38 activation and increased STAT3 signalling.

Hence, NODAL appears to modulate several signalling pathways both positively and inversely correlated with the secretion of factors such as IL6 and CXCL1 in a cell-type dependent manner (Figure 5.13). One notable difference that may explain some of the cell-type-specific effects of NODAL relates to the levels of SMAD2/3 and ERK1/2 activation; MDA-MB-231 cells have higher levels of ERK activation and lower levels of SMAD2/3 activation as compared to SUM149 cells. Hence NODAL may preferentially signal through SMAD2/3 or ERK1/2 in SUM149 and MDA-MB-231 cells, respectively.
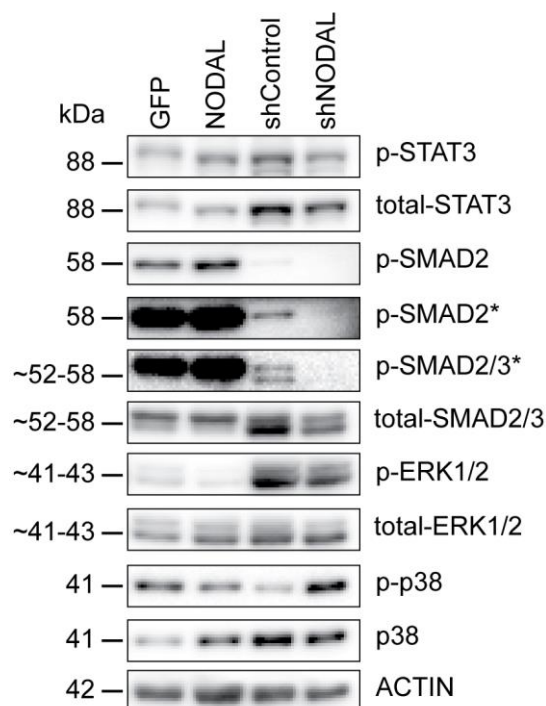
Figure 5.12 Effects of NODAL manipulation on signalling pathways in MDA-MB-231 and SUM149 cells.

Transgenic MDA-MB-231 and SUM149 cells were serum starved and then cultured in serum free media for an additional 24h. Western blotting revealed similarities and differences in activation of downstream pathways. NODAL expression (NODAL and shControl cell lines) was associated with increased phosphorylation of STAT3, SMAD2 and SMAD3 and decreased phosphorylation of p38. Basal levels of p-SMAD2 and p-ERK1/2 were substantially higher in SUM149 and MDA-MB-231 cell lines, respectively. p-ERK1/2 decreased slightly following NODAL overexpression in SUM149 cells, respectively. Western blots are representative images taken from 3 biological replicates and asterisks denote high contrast image settings.
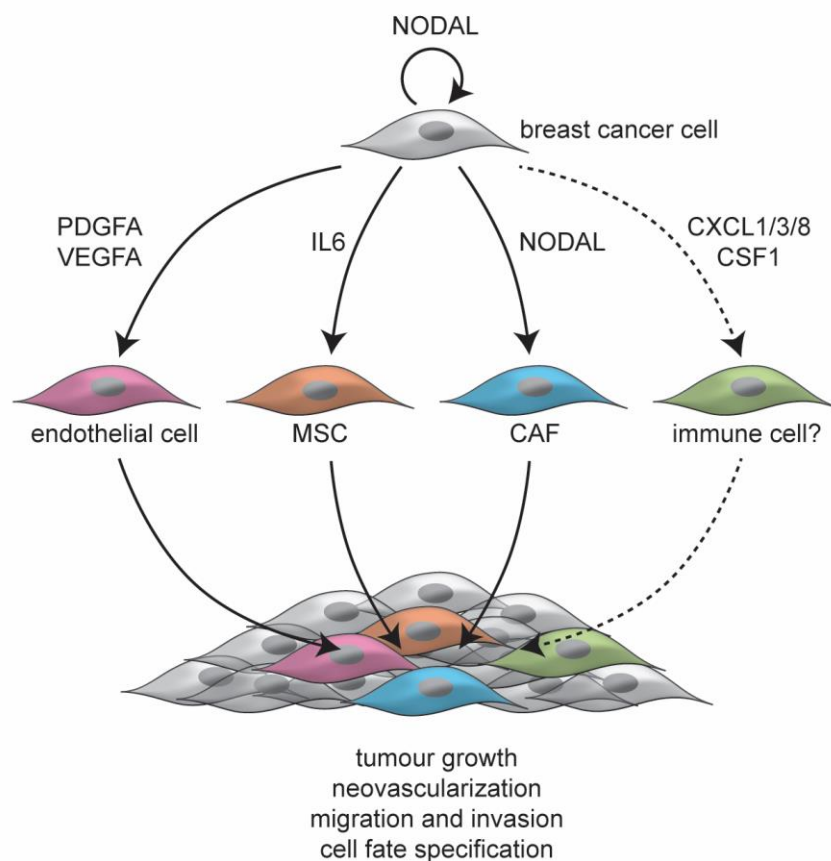
Figure 5.13 Proposed model for NODAL signalling in the breast cancer microenvironment.

NODAL directly signals to breast cancer cells and CAFs. NODAL indirectly regulates inflammatory, chemotactic and angiogenic factors which act on endothelial cells and MSC and possibly immune cell types. Collectively, NODAL promotes tumorigenic phenotypes including tumour growth, neovascularization and cell migration.

5.4 Discussion

We demonstrate for the first time that NODAL signals directly on fibroblasts to induce an activated phenotype and that it can modulate the chemotaxis of MSC cells indirectly, by altering the breast cancer secretome (Figure 5.13). NODAL has been shown to directly induce migration and/or invasion of breast, pancreatic and hepatocellular cancer cell lines *in vitro* [22,40,41]. Moreover, ectopic overexpression of NODAL in breast cancer cells indirectly promotes endothelial tube formation by increasing the expression of pro-angiogenic proteins such as PDGFA [17]. We build upon these studies by showing that NODAL broadly regulates the breast cancer secretome, which may affect TME composition.

Our robust proteomics approach allowed us to uncover dozens of secreted proteins that are affected by NODAL expression in breast cancer cells. For these studies, we took two approaches: We knocked down NODAL in MDA-MB-231 that are claudin-low and basally express NODAL, and we overexpressed NODAL in SUM149, which represent inflammatory breast cancer cells and express low levels of NODAL. Consistent with the effects of NODAL *in vitro* and *in vivo*, the levels of several pro-angiogenic factors (PDGFA, ANGPT1, and ANG) in breast cancer CM were positively correlated with its expression [17]. However, we also made the seemingly paradoxical discovery that the expression of NODAL in MDA-MB-231 and SUM149 breast cancer cells oppositely regulates cytokines involved in chemotaxis. This may be coincident with the models chosen: MDA-MB-231 express relatively low levels of pro-inflammatory cytokines as compared to SUM149 and thus the epigenetic regulation of the genes encoding these proteins may vary dramatically. Genes regulated by NODAL appear to be dictated, at least in part, by accessibility of genomic regions, and NODAL induces histone modifications to affect gene expression [42]. Hence the differential effects of NODAL in MDA-MB-231 versus SUM149 cells may be due to differences in chromatin accessibility in the areas surrounding chemotactic and inflammatory

cytokines. The differences observed may be also due to the ability of NODAL to activate ERK signaling in MDA-MB-231 cells but not in SUM149 cells. Several studies have demonstrated the role of ERK signaling in the up-regulation of inflammatory cytokines such as IL6 [43,44]. Hence the effects of NODAL knock down in MDA-MB-231 cells may be due to reduced ERK signalling.

Our discordant results are not uncommon for studies involving members of the TGF-β family, which function in a context dependent manner. TGF-β1, for example, induces IL6 production in PC3 and DU145 prostate cancer cells via SMAD2/TGFBRII and p38 MAPK [45]. Moreover, in MDA-MB-231 and MDA-MB-468 breast cancer cells, TGF-β1 stimulates IL8(CXCL8) and IL11 secretion via SMAD3/TGFBRI and p38 MAPK [46]. However, in Polyoma virus middle T antigen transformed mouse mammary carcinoma, loss of TGF-β signalling results in an up-regulation of CXCL1, CXCL5 and CCL20 [47]. Remarkably, these factors decreased substantially in SUM149 CM following NODAL overexpression, thus suggesting negative regulatory roles for both NODAL and TGF-β. We did not observe significant differences in the levels of TGF-β1/2 between breast cancer lines; hence the effects of NODAL were not likely mediated *via* alterations in TGF-β1/2. Taken together, both NODAL and TGF-β may differentially regulate chemokine and cytokine expression in cancer, depending on context. This should be considered as treatment modalities designed to target these pathways evolve [48].

While IL6R was detected on MSC, CXCR1 and CXCR2 were not. Heterogeneity in MSC receptor expression has been reported among multiple studies and may be a product of culture conditions and donor heterogeneity [9,49]. For reference, Ponte *et al.* observed CXCR4 and CXCR5 but not CXCR1 or CXCR2 on human BM-MSC [50]. Chamberlain *et al.* also reported high expression for CXCR4 and CXCR5 but low to intermediate expression of CXCR1 and CXCR2, respectively [51]. Conversely, Ringe *et al.* extensively profiled chemokine receptors on human

BM-MSC and detected CXCR1 and CXCR2 but noted loss of expression following 10 passages [52]. While these pathways may play a role in MSC recruitment to tumours in breast cancer patients, we were unable to test this possibility.

In our hands, MDA-MB-231 cells produced less IL6 and CXCL1 compared to work by Hartman *et al*. who investigated the role of cytokines in triple negative breast cancer cell growth [29]. Notwithstanding, neutralizing IL6 in MDA-MB-231 CM was sufficient to attenuate MSC chemotaxis [12,53,54]. We did not neutralize IL6 in SUM149 CM, however CM from either SUM149 and/or SUM159 breast cancer cells was previously shown to promote migration of aldehyde dehydrogenase-high MSC or macrophage-educated MSC in an IL6 dependent manner [31,54].

Although CXCR1/2 was not detected on MSC, differences in CXCL1 and CXCL8 levels following NODAL knockdown/overexpression remain important for cancer progression and trafficking of additional cell types and justifies additional interrogation. For instance, CXCL1 mediated recruitment of CD11b+Gr1+ myeloid cells enhanced breast cancer cell survival, chemoresistance and metastasis [30]. Moreover, obesity-associated CXCL1 expression in prostate tumours was linked to adipose derived stromal cell migration *in vitro* and tumour engraftment *in vivo* [55]. Given the importance of NODAL-regulated cytokines in the TME, future studies interrogating the extent to which NODAL may modulate TME composition are warranted.

In summary, our findings demonstrate the capacity of NODAL to bi-directionally regulate a number pro-tumorigenic factors in the breast cancer secretome. Moreover, NODAL regulates stromal cell recruitment and activation through direct and indirect means. Characterizing the effects of NODAL on cytokines and chemokine expression in additional cell lines and cancers may further improve our understanding of its complex roles during development and cancer progression.

5.5   Materials and methods

5.5.1   Cell culture

MDA-MB-231 cells stably expressing scrambled (shControl) or NODAL targeting (shNODAL) short hairpin RNAs as previously described and validated[17,23,40] were maintained in DMEM/F12 (Gibco) supplemented with 10% FBS (Gibco) and 500ng/mL puromycin. To generate SUM149 cells stably expressing an empty vector (EV), green fluorescent protein (GFP) or NODAL, cells were transduced with lentiviral particles (GeneCopia) overnight then selected and maintained in HAM's F10 (Gibco) supplemented with 10% FBS, 5µg/mL insulin (Santa Cruz Biotechnology, Dallas TX, USA), 1µg/mL hydrocortisone (Sigma-Aldrich, St. Louis, MO, USA) and 100 ng/mL puromycin. Human BM-MSC lines were maintained in Amniomax with C100 supplement (Life Technologies, Carlsbad, CA, USA) and previously confirmed to express characteristic stromal markers (>95% CD90+, CD105+, and CD73+) and exhibit multipotent differentiation.[35,56] HFFs were maintained in DMEM/F12 supplemented with 10% FBS. For SILAC labelling, shControl and shNODAL MDA-MB-231 cells were cultured in DMEM F12 supplemented with dialyzed FBS (Life Technologies) containing light (Advanced ChemTech, Louisville, KY, USA) or heavy (Cambridge Isotope Laboratories, Tewksbury, MA, USA and Silantes GmbH, Germany) isotopes of arginine (0.398mM) and lysine (0.274mM) for at least 9 days to achieve >90% label incorporation. SILAC media was additionally supplemented with 400 mg/L of proline (Sigma-Aldrich) to limit arginine to proline conversion.[57] CM was prepared by plating equal cell numbers onto flasks in culture media (Corning, NY, USA). After 24h (MDA-MB-231 cells) or 48h (SUM149 cells), media was removed and cells were thoroughly rinsed 3 times in PBS (with Ca2+ and Mg2+) to remove serum components. Cells were incubated in serum free media (SFM) with 0.5% BSA for an additional 24h to generate CM (BSA was omitted for LC-MS samples).

Conditions used to stimulate cells with rhNODAL and rhIL6 are specified in main text.

5.5.2 Sample preparation for liquid chromatography-mass spectrometry (LC-MS)

CM (without BSA) were concentrated using 3 kDa molecular weight cut-off (MWCO) Amicon ultracentrifugal units (Millipore) and lyophilized overnight. The following day, CM was reconstituted in lysis buffer (8M urea, 50mM ammonium bicarbonate, 10mM dithiothreitol and 2% SDS), sonicated (3 X 0.5s pulses) with a probe sonicator (Level 1; Fisher Scientific, Waltham, MA) and quantified using a Pierce™ 660 nm assay (Thermo Scientific™) with ionic detergent compatibility reagent. For SILAC samples, light shControl and heavy shNODAL CM was pooled based on equal cell numbers and ~100µg of protein was fractionated using SDS-PAGE on 12% acrylamide tris-glycine gels. In-gel digestion with trypsin (1:25 enzyme:protein ratio) was performed on 16-17 slices (fractions) from each lane in biological triplicate as previously described [58]. For label free samples, ~50µg of protein from SUM149 CM was precipitated in chloroform/methanol, digested overnight with trypsin (1:50 ratio) on a water bath shaker and fractionated on SCX StageTips as previously described [58–60]. Peptides were dried in a SpeedVac, reconstituted in 0.1% formic acid (FA; Fisher Scientific) and a volume corresponding to 1/10th of the total material recovered or 1µg as determined by BCA (Pierce™) was injected for each in-gel and SCX fraction, respectively.

5.5.3 LC-MS

In-gel and SCX fractions were analyzed using a Q Exactive or Orbitrap Elite mass spectrometer (Thermo Scientific™), respectively. Samples were injected using a nanoAcquity HPLC system (Waters) and initially trapped on a Symmetry C18 Trap Column (5 µm, 180 µm x 20 mm) for 4 or 5 minutes in 99% Solvent A (Water/0.1% FA)/1% Solvent B (acetonitrile/0.1% FA) at a flow rate of 10 µl/min. Peptides were

separated on an ACQUITY Peptide BEH C18 Column (130Å, 1.7µm, 75µm X 250mm) at a flow rate of 300 nL/min maintained at 35°C. The LC-MS gradient for in-gel digests consisted of 1-7% B over 1 minute and 7-37.5% B over 79 minutes. SCX fractions were separated using gradient consisting of 7.5% B over 1 minute, 25% B over 179 minutes, 32.5% B over 40 minutes and 60% B over 20 minutes. Column washing and re-equilibration was performed following each run and settings for data acquisition are outlined in ESM5.8.

5.5.4   *Data analysis and statistics*

MS files were searched in MaxQuant (1.5.2.8) with the Human Uniprot database (reviewed only; updated May 2014 with 40,550 entries) [61]. Missed cleavages were set to 3 and I=L. Cysteine carbamidomethylation was set as a fixed modification. Oxidation (M), n-terminal acetylation (protein), and deamidation (NQ) were used as variable modifications (max. number of modifications per peptide = 5) and min ratio count was set to 1. All other settings were left default. The match-between-runs feature was utilized to maximize proteome coverage and quantitation between samples. Datasets were loaded into Perseus (version 1.5.5.3) and proteins identified by site, reverse and potential contaminants were removed [28]. Protein identifications with quantitative values in ≥2 biological replicates were retained for downstream analysis unless specified elsewhere. Missing values were imputed using a width of 0.3 and down shift of 1.8 for label free datasets. Statistical analysis was performed in Perseus or GraphPad Prism version 6.01 (San Diego, CA). All experiments were carried in at least 3 biological replicates unless specified otherwise. Where specified, replicate treatment values were normalised to the control group and relative fold-changes were reported. Two-tailed, one sample and two-sample t-tests ($p < 0.05$) were performed to determine statistical differences unless more than 2 conditions were being compared and a one way ANOVA using Dunnett's multiple comparison test ($p < 0.05$) was performed instead.

5.5.5 Chemotaxis and invasion assays

MSC were rinsed in warm PBS (with Ca2+ and Mg2+) and serum starved for ~2h in Amniomax prior to dissociation with trypsin for chemotaxis assays. In parallel, 8μM transwells (Falcon) were coated with 10μg/cm^2 of bovine fibronectin (Sigma-Aldrich) in 100μL of PBS for 2h. After coating, excess solution was aspirated and 40K MSC in 0.5mL of DMEM F12+0.5% BSA were plated in each transwell. HFFs were serum starved 24h prior to dissociation and plated at a density of 50K cells/transwell. For HFF chemotaxis and invasion assays, fibronectin and Matrigel™ were omitted and included, respectively. To the bottom chamber, 1mL of DMEM/F12+0.5%BSA or CM was added +/- rhNODAL (R&D systems), rhIL6 (eBioscience), isotype or IL6 neutralizing monoclonal antibodies (R&D systems). After ~24h, transwells were rinsed in warm PBS and placed in cold methanol for 20 minutes to fix migrating cells. After fixing, transwells were rinsed in PBS and the inside membrane was thoroughly wiped with a cotton swab to remove non-migrated cells. Membranes were excised and mounted onto glass slides with ProLong™ Gold Antifade Mountant with DAPI (Invitrogen™). Migrated cells were counted from at least 5-10 high power fields uniformly distributed across the entire membrane for each condition.

5.5.6 Western blotting

Cells were thoroughly washed with PBS (with Ca2+ and Mg2+) and directly lysed on tissue culture plates in lysis buffer. Lysates recovered by pipetting were sonicated with a probe sonicator (20 X 0.5s pulses) to shear DNA and reduce viscosity. Equal protein amounts (15-25μg) were separated on hand cast 8-20% acrylamide Tris-glycine gels then transferred to Immobilon-P® PVDF membranes (Millipore™, Billerica, MA, USA). Membranes were stained with amido black and rinsed in ddH$_2$O for 5 minutes followed by blocking for 1h on rocker in 5% non-fat dry milk in TBST (Tris-buffered saline, 0.1% Tween 20) and overnight incubation

in primary antibody at 4°C. Chemiluminescent detection was performed using film or a VersaDoc CCD camera with Clarity™ Western ECL Substrate and horseradish peroxidase-conjugated secondary antibodies (Bio-Rad) the next day. Antibody information is available in ESM5.9 and ACTIN and TUBULIN were used as loading control. PVDF membranes were stippled in 0.2 M NaOH and reprobed when possible, otherwise Western blots were run in duplicate.

5.5.7 Real-time PCR

RNA was isolated from cells and treated with DNAse using a Perfect Pure RNA cultured cell kit (5 PRIME). RNA was quantified by NanoDrop™ (Thermo Scientific™) and 2µg was reverse transcribed with a High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Carlsbad, CA, USA). Real-time PCR was performed with TaqMan™ Universal PCR mastermix on a Bio-Rad CFX96/384 thermocycler. HPRT1 or RPLPO were used as housekeeping to genes monitor variations between biological replicates. TaqMan™ primer probes were purchased from Applied Biosystems and are listed in ESM5.10.

5.5.8 Flow cytometry

MSC dissociated in 10mM EDTA/PBS solution for 5-10 minutes were resuspended in 5% FBS/PBS, counted and pelleted at 450 xg. Excess buffer was aspirated and MSC were divided 50-100K cell aliquots in 100µL of 5% FBS/PBS. Isotype controls and primary antibodies were added to cell suspensions and incubated for ~45 minutes in the dark on ice (ESM5.9). Cell suspensions were washed in excess 5% FBS/PBS and pelleted to remove unbound antibody. Flow cytometry data was acquired on an LSR II (Becton Dickinson, NJ, USA) using FACSDiva at the London Regional Flow Cytometry Facility and analyzed with FlowJo (Treestar, Ashland, OR, Version 10.0.8r1). Gating strategy for live singlets was based on forward and side-scatter and is illustrated in figure 5.14.

5.5.9   ELISAs

ELISA kits were purchased from eBioscience (IL6) or R&D systems (CXCL1, CXCL8 and CSF1) and performed according to the manufacturer's specifications using CM derived from MDA-MB-231 and SUM149 cell lines.

5.5.10   Gene expression profiling

Human dermal fibroblasts (ATCC) were cultured in DMEM supplemented with 10% FBS until ~40-60% confluence, washed twice with PBS and incubated overnight in DMEM+0.5%FBS. The following day, cells were treated +/- rhNODAL (10 ng/mL) for 6h and RNA was harvested using TRIzol™ (Invitrogen). RNA was subjected to expression profiling at the London Regional Genomics Centre essentially as previously described [62,63]. RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies Inc., Palo Alto, CA) prior to preparing single stranded complimentary DNA (sscDNA) from 200ng of total RNA (Ambion WT Expression Kit for Affymetrix GeneChip Whole Transcript WT Expression Arrays; Applied Biosystems, Carlsbad, CA) according to the Affymetrix User Manual (Affymetrix, Santa Clara, CA). In total, 5.5µg of sscDNA was synthesized, converted into cRNA, end labeled and hybridized (16h at 45°C) to Human Gene 1.0 ST arrays. Liquid handling steps were performed by a GeneChip Fluidics Station 450 and GeneChips were scanned (GeneChip Scanner 3000 7G; Affymetrix, Santa Clara, CA) using Command Console v1.1 to generate Probe level (.CEL file) data. Gene level data was generated using the RMA algorithm [64]. Partek Genomics Suite v6.5 (St. Louis, MO) was used to determine gene level ANOVA p-values and fold-changes. Fold-changes were obtained by averaging data from two experiments (GeneSpring). Fold-changes exceeding 1.7 in response to rhNODAL were required to identify a transcript as being altered ($p<0.05$). Altered genes were annotated using DAVID (version 6.7) and lists enriched >3.5 fold and comprised of >10 genes were reported (ESM5.10).
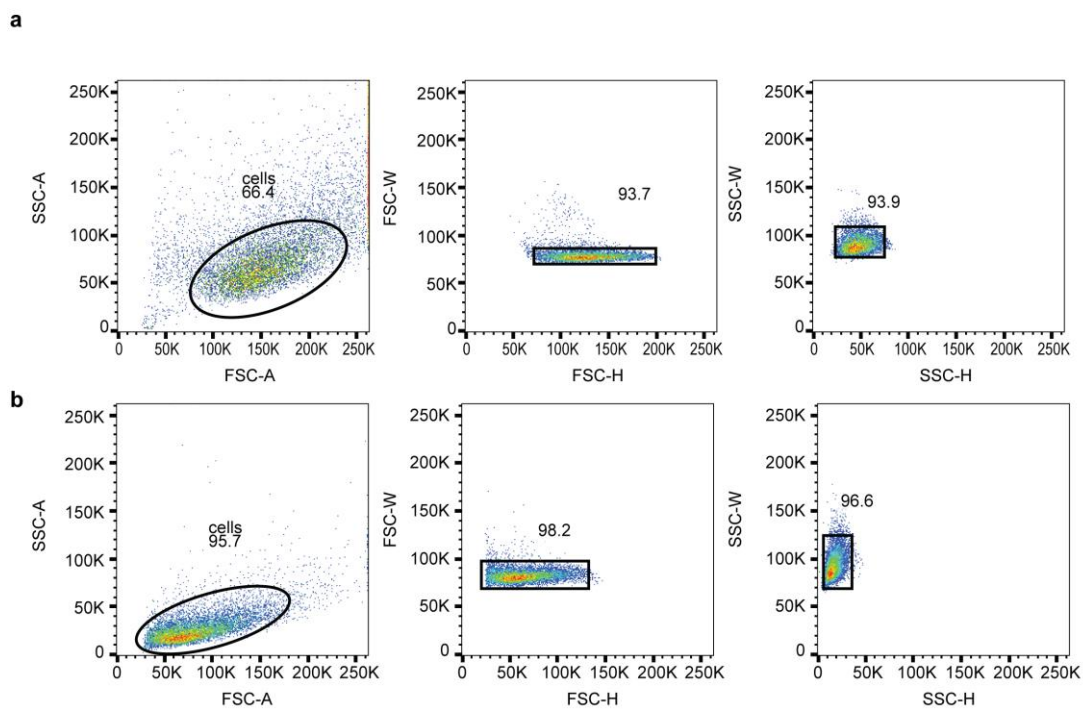
5.5.11   Acknowledgements

Figure 5.14 Gating strategy for flow cytometry analysis.

(a and b) Forward- and side-scatter gating method used to select live singlets for surface marker expression and CTV analyses.

5.6   References:

[1]   Balkwill, F.R., Capasso, M., Hagemann, T., The tumor microenvironment at a glance. *J. Cell Sci.* 2012, 125, 5591–6.

[2]   Rønnov-Jessen, L., Petersen, O.W., Bissell, M.J., Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol. Rev.* 1996, 76, 69–125.

[3]   Olumi, A.F., Grossfeld, G.D., Hayward, S.W., Carroll, P.R., et al., Carcinoma-associated fibroblasts direct tumor progression of initiated human prostatic epithelium. *Cancer Res.* 1999, 59, 5002–11.

[4]   Hanahan, D., Coussens, L.M., Accessories to the Crime: Functions of Cells Recruited to the Tumor Microenvironment. *Cancer Cell* 2012, 21, 309–322.

[5]   Orimo, A., Gupta, P.B., Sgroi, D.C., Arenzana-Seisdedos, F., et al., Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* 2005, 121, 335–348.

[6]   Kojima, Y., Acar, A., Eaton, E.N., Mellody, K.T., et al., Autocrine TGF- and stromal cell-derived factor-1 (SDF-1) signaling drives the evolution of tumor-promoting mammary stromal myofibroblasts. *Proc. Natl. Acad. Sci.* 2010, 107, 20009–20014.

[7]   Wu, Y., Zhao, R.C.H., Tredget, E.E., Concise review: Bone marrow-derived stem/progenitor cells in cutaneous repair and regeneration. *Stem Cells* 2010, 28, 905–915.

[8]   Mezawa, Y., Orimo, A., The roles of tumor- and metastasis-promoting carcinoma-associated fibroblasts in human carcinomas. *Cell Tissue Res.* 2016, 365, 675–689.

[9]   Karp, J.M., Leng Teo, G.S., Mesenchymal Stem Cell Homing: The Devil Is in the Details. *Cell Stem Cell* 2009, 4, 206–216.

[10]   Karnoub, A.E., Dash, A.B., Vo, A.P., Sullivan, A., et al., Mesenchymal stem cells within tumour stroma promote breast cancer metastasis. *Nature* 2007, 449, 557–563.

[11]   Mishra, P.J.P.J.P.J., Mishra, P.J.P.J.P.J., Humeniuk, R., Medina, D.J., et al., Carcinoma-associated fibroblast-like differentiation of human mesenchymal stem cells. *Cancer Res.* 2008, 68, 4331–4339.

[12]   Shangguan, L., Ti, X., Krause, U., Hai, B., et al., Inhibition of TGF-β/Smad signaling by BAMBI blocks differentiation of human mesenchymal stem cells to carcinoma-associated fibroblasts and abolishes their protumor effects. *Stem Cells* 2012, 30, 2810–2819.

[13]   Spaeth, E., Klopp, A., Dembinski, J., Andreeff, M., Marini, F., Inflammation and tumor microenvironments: defining the migratory itinerary of mesenchymal stem cells. *Gene Ther.* 2008, 15, 730–738.

[14]   Quante, M., Tu, S.P., Tomita, H., Gonda, T., et al., Bone Marrow-Derived Myofibroblasts Contribute to the Mesenchymal Stem Cell Niche and Promote Tumor Growth. *Cancer Cell* 2011, 19, 257–272.

[15]   Wang, H.-H., Cui, Y.-L., Zaorsky, N.G., Lan, J., et al., Mesenchymal stem cells generate pericytes to promote tumor recurrence via vasculogenesis after stereotactic body radiation therapy. *Cancer Lett.* 2016, 375, 349–359.

[16]   Topczewska, J.M., Postovit, L.-M., Margaryan, N. V, Sam, A., et al., Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat. Med.* 2006, 12, 925–932.

[17]   Quail, D.F., Walsh, L. a., Zhang, G., Findlay, S.D., et al., Embryonic protein nodal promotes breast cancer vascularization. *Cancer Res.* 2012, 72, 3851–3863.

[18]   Pauklin, S., Vallier, L., Activin/Nodal signalling in stem cells. *Development* 2015, 142, 607–19.

[19]   Postovit, L.-M., Margaryan, N. V, Seftor, E.A., Kirschmann, D. a, et al., Human embryonic stem cell microenvironment suppresses the tumorigenic phenotype of aggressive cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 4329–34.

[20]   Lonardo, E., Hermann, P.C., Mueller, M.T., Huber, S., et al., Nodal/activin signaling drives self-renewal and tumorigenicity of pancreatic cancer stem cells and provides a target for combined drug therapy. *Cell Stem Cell* 2011, 9, 433–446.

[21]   Lee, C.-C., Jan, H.-J., Lai, J.-H., Ma, H.-I., et al., Nodal promotes growth and invasion in human gliomas. *Oncogene* 2010, 29, 3110–3123.

[22]   Chen, J., Liu, W. Bin, Jia, W.D., Xu, G.L., et al., Embryonic morphogen nodal is associated with progression and poor prognosis of hepatocellular carcinoma. *PLoS One* 2014, 9, 1–10.

[23]   Quail, D.F., Zhang, G., Walsh, L. a., Siegers, G.M., et al., Embryonic Morphogen Nodal Promotes Breast Cancer Growth and Progression. *PLoS One* 2012, 7, 1–12.

[24]   Cailleau, R., Young, R., Olivé, M., Reeves, W.J.J., Breast tumor cell lines from pleural effusions. *J. Natl. Cancer Inst.* 1974, 53, 661–74.

[25]   Van Golen, K.L., Davies, S., Wu, Z.F., Wang, Y.F., et al., A novel putative low-affinity insulin-like growth factor-binding protein, LIBC (lost in inflammatory breast cancer), and RhoC GTPase correlate with the inflammatory breast cancer phenotype. *Clin. Cancer Res.* 1999, 5, 2511–2519.

[26]   Geiger, T., Wehner, A., Schaab, C., Cox, J., Mann, M., Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* 2012, 11, M111.014050.

[27]   Pozniak, Y., Balint-Lahat, N., Rudolph, J.D., Lindskog, C., et al., System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. *Cell Syst.* 2016, 2, 172–184.

[28]   Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., et al., The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* 2016, 13, 731–40.

[29]   Hartman, Z.C.C., Poage, G.M.M., den Hollander, P., Tsimelzon, A., et al., Growth of triple-negative breast cancer cells relies upon coordinate autocrine expression of the proinflammatory cytokines IL-6 and IL-8. *Cancer Res.* 2013, 73, 3470–80.

[30]   Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., et al., A CXCL1 paracrine network links cancer chemoresistance and metastasis. *Cell* 2012, 150, 165–178.

[31]   Liu, S., Ginestier, C., Ou, S.J.S.J., Clouthier, S.G.S.G., et al., Breast cancer stem cells are regulated by mesenchymal stem cells through cytokine networks. *Cancer Res.* 2011, 71, 614–624.

[32]   Kishimoto, K., Liu, S., Tsuji, T., Olson, K.A., Hu, G.F., Endogenous angiogenin in endothelial cells is a general requirement for cell proliferation and angiogenesis. *Oncogene* 2005, 24, 445–456.

[33]   Fagiani, E., Christofori, G., Angiopoietins in angiogenesis. *Cancer Lett.* 2013, 328, 18–26.

[34]   Capoccia, B.J., Robson, D.L., Levac, K.D., Maxwell, D.J., et al., Revascularization of ischemic limbs after transplantation of human bone marrow cells with high aldehyde dehydrogenase activity. *Blood* 2009, 113, 5340–5351.

[35]   Bell, G.I., Broughton, H.C., Levac, K.D., Allan, D. a., et al., Transplanted Human Bone Marrow Progenitor Subtypes Stimulate Endogenous Islet Regeneration and Revascularization. *Stem Cells Dev.* 2012, 21, 97–109.

[36]   Kidd, S., Spaeth, E., Watson, K., Burks, J., et al., Origins of the tumor microenvironment: Quantitative assessment of adipose-derived and bone marrow-derived stroma. *PLoS One* 2012, 7.

[37]   Huang, D.W., Sherman, B.T., Lempicki, R.A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 2009, 4, 44–57.

[38]   Bertero, A., Madrigal, P., Galli, A., Hubner, N.C., et al., Activin/Nodal signaling and NANOG orchestrate human embryonic stem cell fate decisions by controlling the H3K4me3 chromatin mark. *Genes Dev.* 2015, 29, 702–717.

[39]   Massagué, J., TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* 2012, 13, 616–630.

[40]   Quail, D.F., Zhang, G., Findlay, S.D., Hess, D. a, Postovit, L.-M., Nodal promotes invasive phenotypes via a mitogen-activated protein kinase-dependent pathway. *Oncogene* 2014, 33, 461–73.

[41]   Duan, W., Li, R., Ma, J., Lei, J., et al., Overexpression of Nodal induces a metastatic phenotype in pancreatic cancer cells via the Smad2/3 pathway. *Oncotarget* 2015, 6, 1490–506.

[42]   Coda, D.M., Gaarenstroom, T., East, P., Patel, H., et al., Distinct modes of SMAD2 chromatin binding and remodeling shape the transcriptional response to NODAL/Activin signaling. *Elife* 2017, 6, 1–31.

[43]   Yun, M.R., Choi, H.M., Kang, H.N., Lee, Y., et al., ERK-dependent IL-6 autocrine signaling mediates adaptive resistance to pan-PI3K inhibitor BKM120 in head and neck squamous cell carcinoma. *Oncogene* 2018, 37, 377–388.

[44]   Balko, J.M., Schwarz, L.J., Bhola, N.E., Kurupi, R., et al., Activation of MAPK pathways due to DUSP4 loss promotes cancer stem cell-like phenotypes in basal-like breast cancer. *Cancer Res.* 2013, 73, 6346–6358.

[45]   Park, J.-I., Lee, M.-G., Cho, K., Park, B.-J., et al., Transforming growth factor-beta1 activates interleukin-6 expression in prostate cancer cells through the synergistic collaboration of the Smad2, p38-NF-kappaB, JNK, and Ras signaling pathways. *Oncogene* 2003, 22, 4314–4332.

[46]   Gupta, J., Robbins, J., Jilling, T., Seth, P., TGFβ-dependent induction of interleukin-11 and interleukin-8 involves SMAD and p38 MAPK pathways in breast tumor models with varied bone metastases potential. *Cancer Biol. Ther.* 2011, 11, 311–6.

[47]   Bierie, B., Chung, C., Abrogation of TGF-β signaling enhances chemokine production and correlates with prognosis in human breast cancer. *J Clin Invest* 2009, 119, 1571–1582.

[48] Hendrix, M.J., Kandela, I., Mazar, A.P., Seftor, E.A., et al., Targeting melanoma with front-line therapy does not abrogate Nodal-expressing tumor cells. *Lab. Investig.* 2016, 00, 1–11.

[49] Dominici, M., Le Blanc, K., Mueller, I., Slaper-Cortenbach, I., et al., Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* 2006, 8, 315–317.

[50] Ponte, A.L., Marais, E., Gallay, N., Langonne, A., et al., The in vitro migration capacity of human bone marrow mesenchymal stem cells: comparison of chemokine and growth factor chemotactic activities. *Stem Cells* 2007, 25, 1737–1745.

[51] Chamberlain, G., Wright, K., Rot, A., Ashton, B., Middleton, J., Murine mesenchymal stem cells exhibit a restricted repertoire of functional chemokine receptors: Comparison with human. *PLoS One* 2008, 3, 1–6.

[52] Ringe, J., Strassburg, S., Neumann, K., Endres, M., et al., Towards in situ tissue repair: Human mesenchymal stem cells express chemokine receptors CXCR1, CXCR2 and CCR2, and migrate upon stimulation with CXCL8 but not CCL2. *J. Cell. Biochem.* 2007, 101, 135–146.

[53] Pricola, K.L., Kuhn, N.Z., Haleem-Smith, H., Song, Y., Tuan, R.S., Interleukin-6 maintains bone marrow-derived mesenchymal stem cell stemness by an ERK1/2-dependent mechanism. *J. Cell. Biochem.* 2009, 108, 577–588.

[54] Wolfe, A.R., Trenton, N.J., Debeb, B.G., Larson, R., et al., Mesenchymal stem cells and macrophages interact through IL-6 to promote inflammatory breast cancer in pre-clinical models. *Oncotarget* 2016, 7, 82482–82492.

[55] Zhang, T., Tseng, C., Zhang, Y., Sirin, O., et al., CXCL1 mediates obesity-associated adipose stromal cell trafficking and function in the tumour microenvironment. *Nat. Commun.* 2016, 7, 11674.

[56] Sherman, S.E., Kuljanin, M., Cooper, T.T., Putman, D.M., et al., High Aldehyde Dehydrogenase Activity Identifies a Subset of Human Mesenchymal Stromal Cells with Vascular Regenerative Potential. *Stem Cells* 2017, 35, 1542–1553.

[57] Bendall, S.C., Hughes, C., Stewart, M.H., Doble, B., et al., Prevention of Amino Acid Conversion in SILAC Experiments with Embryonic Stem Cells. *Mol. Cell. Proteomics* 2008, 7, 1587–1597.

[58] Kuljanin, M., Dieters-Castator, D.Z., Hess, D.A., Postovit, L.-M., Lajoie, G.A., Comparison of sample preparation techniques for large-scale proteomics. *Proteomics* 2017, 17, 1600337.

[59] Wessel, D., Flügge, U.I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 1984, 138, 141–143.

[60] Kulak, N. a, Pichler, G., Paron, I., Nagaraj, N., Mann, M., Minimal, encapsulated proteomic-sample processing applied to copy-number estimationa in eukaryotic cells. *Nat. Methods* 2014, 11, 319–24.

[61] Cox, J., Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 2008, 26, 1367–1372.

[62] Guo, F., Carter, D.E., Leask, A., Mechanical tension increases CCN2/CTGF expression

and proliferation in gingival fibroblasts via a TGFβ-dependent mechanism. *PLoS One* 2011, 6, 1–10.

[63]   Kuk, H., Hutchenreuther, J., Murphy-Marshman, H., Carter, D., Leask, A., 5z-7-oxozeanol inhibits the effects of TGFβ1 on human gingival fibroblasts. *PLoS One* 2015, 10, 1–13.

[64]   Irizarry, R.A., Hobbs, B., Collin, F., Beazer-barclay, Y.D., et al., Exploration , Normalization , and Summaries of High Density Oligonucleotide Array Probe Level Data 2018, 249–264.

Chapter 6

Discussion

6.1   General discussion and conclusions

Our understanding of the complex and heterogeneous nature of cancer and the TME as a whole is constantly evolving. This has significant repercussions for patient outcomes which depend heavily on early detection, accurate diagnosis, and tailored treatments options. Rapid advances in the proteomics pipeline have propelled the study of complex biological systems to an unprecedented depth. When applied to models of cancer, proteomics can provide valuable information on all aspects of disease progression including subtype specific signatures, biomarkers for screening and detection, and signalling pathways involved in malignant behaviours. In this thesis, I employed global mass spectrometry-based approaches to identify biological features that may aid in the diagnosis and detection of ovarian cancer and characterize signalling factors regulated by NODAL in the breast cancer secretome. More specifically, in Chapter 2, I systematically compared newer fractionation strategies with established methods for achieving large proteomic coverage. In Chapter 3, I contrast differences between two ovarian cancer subtypes, namely HGSC and EC, for which differential diagnosis is lacking. In Chapter 4, I performed detailed proteomic analysis of ovarian cancer EVs from malignant bio-fluids to identify novel biomarkers for detecting ovarian cancer. In Chapter 5, I demonstrated a role for NODAL in regulating cytokines and chemokines that mediate stromal cell recruitment within the breast cancer microenvironment.

6.2   Sample preparation strategies for comprehensive proteome coverage

Within each proteomic workflow, a vast array of parameters can be optimized and tweaked to achieve comprehensive proteomic information of complex biological samples. Sample preparation strategies remain an active area of the proteomics pipeline and are continually being improved. Our work in Chapter 2 has further

established high-pH/low-pH reversed phase fractionation (HpH) as a superior sample preparation scheme over other methods for large-scale proteomics when no other limitations are present (i.e. time/cost/access to specialized equipment) [1]. Importantly, our findings are in agreement with recent studies by other groups which have implemented HpH fractionation into their workflows [2–5]. For example, Kulak *et al.*, whom vastly simplified SCX StageTip fractionation used in our comparison, recently developed a Loss-less Nano-fractionator (Spider Fractionator under commercial development by PreOmics, GmbH) to help streamline and automate LC-based fractionation which is commonly employed with HpH [6]. Fortunately, when UHPLC or HPLC systems are not available, HpH fractionation can be performed with StageTips [7]. One of the earliest examples of this was by Han *et al.* whom in 2013 performed StageTip-based HpH fractionation of digests from BV-2 mouse microglial cells [8]. Although we did not compare HPLC- and StageTip-based HpH fractionation, HPLC-based fractionation provides slightly better fractionation efficiency and more protein identifications (unpublished observations and personal communication with Dr. Kuljanin) [6]. Compared to HPLC-based setups, StageTips do not typically employ concatenation schemes and fractions are eluted in steps which may reduce orthogonality [9,10].

Protein digestion with trypsin remains an integral component of sample preparation protocols in nearly all bottom-up proteomic studies [11]. Paradoxically, the majority of tryptic peptides (56%) generated by *in silico* digestion of the yeast proteome are ≤6 amino acids (AA) long while 97% of all peptides observed by mass spectrometry are between 7-35 AA with an average length of 8.4 AA [12]. These differences highlight a long standing issue with bottom-up proteomics whereby short peptides (≤6 AA) comprise the majority of species within each digest but are often excluded during acquisition or searching to avoid non-unique sequences [13]. Alternative enzymes and chemical digestions are one means to mitigate these issues and increase protein identifications and approach 100% sequence coverage [13]. While such adaptations are reported for PTM analysis and antibody sequencing,

only a handful of studies have employed multiple enzyme and/or chemical digestions for large scale proteomic analysis [14–16]. For example, Wiśniewski *et al.* combined multiple enzyme digestions with filter aided sample preparation (MED-FASP) to increase the proteome depth of HeLa cells over digestion with trypsin alone [12,17]. More recently, Bekker-Jensen *et al.* utilized trypsin, LysC, chymotrypsin and GluC digestions to report the highest protein identifications and greatest sequence coverage achieved in a mammalian cell line to date [5]. While these modifications are associated with increased sample handling, they provide additional sequence information for detecting species/proteoforms that are missed with standard tryptic digests [11].

In addition to sample preparation, ongoing advances in instrumentation hardware and software have resulted in significant improvements to the amount and quality of spectral data acquired. For instance, segmented quadrupoles on newer Orbitrap systems have reduced co-isolation of multiple precursors that are commonly encountered when working with complex samples. Impressively, current generation Orbitrap Fusion Tribrid mass spectrometers can theoretically achieve sequencing speeds of ~60Hz in parallel acquisition mode [18]. Until very recently, this potential was not realized due to limitations in assigning correct charge states of overlapping and/or low abundance precursors with the standard precursor detection algorithm. However, an Advanced Precursor Detection (APD) algorithm is now available which can handle these scenarios and maximize the number of MS/MS scans to 96% capacity up from 72% [18].

The field of mass spectrometry-based proteomics is undergoing innovation at a staggering pace and it is an exciting time to witness these innovations. In the latest 2017 report from the Human Proteome Project (HPP), highly stringent protein evidence was still missing from an estimated ~18% of the human proteome [19]. There is an ongoing effort to accelerate the identification of missing proteins such as that noted with olfactory receptors. In this section, I briefly touched on a few of

the many areas where sample preparation and data acquisition are having notable successes. In effect, a combination of the technologies and their iterations are needed to enable complete proteome coverage in future endeavors.

6.3 Proteomic profiling enables classification of cancer subtypes

Ovarian cancer remains a challenging disease to detect early, diagnose and treat. Genomic and histological studies have significantly improved our understanding of the mechanisms underlying EOC progression and pathology [20,21]. Mass spectrometry-based proteomics is a complementary approach to these techniques and is steadily being incorporated into ovarian cancer research [3,22]. In chapter 3, we performed a global proteomics analysis on fresh frozen EC and HGSC tumour samples with the goal of identifying histospecific markers of EC. Using the SVM package 'geNetClassifier' in R, ~100 proteins were identified with high classification power for either EC or HGSC. Our proteomics dataset represents a significant contribution to the field of ovarian cancer research which is striving to improve patient management through better histopathological assessment. In agreement with other studies, TP53 and CDKN2A expression were upregulated in several HGSC tumours while PR was elevated in EC tumour samples [21]. To validate novel histotype specific markers for EC, we performed IHC for a subset of candidates on a large cohort of over 300 EC and HGSC tumour samples. IHC corroborated our proteomics findings however limited specificity or sensitivity hindered our ability to identify a high performing marker of EC. KIAA1324 was the best subtype specific marker of EC next to PR. IHC, although streamlined, remains a rate limiting, but necessary step for validation. Fortunately, we now have a database with matched staining information for these markers. We will use this information for multivariate analyses that may reveal a histotype specific signature, rather than a single marker, for detecting EC and correlating patient outcomes.

With regard to proteomic coverage, we achieved relatively good depth for unfractionated tumour samples analyzed on an Orbitrap Elite (~5600 entries after filtering). For a rough comparison, Hughes *et al.* quantified 8167 proteins (98.5% present in all samples) using TMT-labelled, HpH fractionated digests from FFPE tumour sections (CCC and HGSC) on an Orbitrap Fusion[3]. Although we anticipated an unfractionated approach would be sufficient to detect markers enriched in or exclusive to EC, it is possible that low abundance proteins were missed. Notably, WT1 is a transcription factor that is specific for HGSC which was absent from our proteomic dataset. We verified the presence of numerous nuclear proteins and transcription factors and therefore, insufficient extraction of this organelle was an unlikely factor. Rather, it is plausible that WT1 expression was below the limit of detection of our instrument when analyzing unfractionated samples. This observation is supported by previous work by Hughes *et al.* and Coscia *et al.* in which WT1 expression appears significantly lower than other HGSC markers including TP53 and CDKN2A. For example, in the study by Coscia *et al.*, WT1 was absent in 19 of 30 cell lines analyzed with $\log_2$ LFQ intensity values, when present, ranging from ~20-25. TP53, on the other hand, was detected in 29 out of 30 ovarian (cancer) cell lines with $\log_2$ LFQ values ranging from ~25-30. Using these values as a conservative estimate, the abundance of WT1 is likely 2^5 or ~2 orders of magnitude lower than TP53. Interrogation of our dataset revealed an average TP53 $\log_2$ LFQ value of 24.6 in the 3 out of 10 HGSC tumour samples it was present in. Accordingly, WT1 would be at or below the limit of detection (LFQ value of around 20). If low abundance markers are indeed required to improve discrimination of EOC subtypes, it is worthwhile to perform additional analysis on fractionated EC and HGSC tumours samples as the cost of instrument time. Multiplexing with tandem mass tags (TMT) can mitigate this issue but MS2 based quantification on Orbitrap Elite and Q Exactive systems is associated with interference in the low mass range, also known as ratio distortion [23]. MS3 based quantification, however, can improve quantification accuracy at the cost of sensitivity and sequencing speed [23]. Interestingly, an updated quantification

strategy based on complementary TMT (peptide) fragments (TMTc+) was recently reported [24]. Importantly, this approach is compatible with older generation Orbitrap instruments while maintaining the high sensitivity, accuracy and speed afforded by synchronous precursor selection (SPS) MS3-based quantification on newer Orbitrap Fusion Tribrid instruments [23,24].

Advances in instrumentation and multiplexing are increasing the rate of proteomic profiling of tumours and cancer biopsies. Rapid in depth proteome coverage can be obtained within 1 day as demonstrated on recent state of the art Q Exactive HF [5]. It is probable that patient biopsies will one day be routinely profiled using proteomics. Moreover, proteogenomics is a growing area of research that incorporates both genomics and proteomics [25]. While the field is still in its infancy, personalized protein sequence databases containing single amino acid variants (SAAVs) missing from publicly available repositories can be derived from whole genome, exome and RNA sequencing technologies. However, there are a number of challenges associated with generating, searching and integrating multiple 'omic' datasets. For instance, high quality spectral information may not exist in curated databases thereby necessitating the need for *de novo* peptide sequencing tools like Peaks® [26]. Moreover, iterative search strategies are required to minimize high FDRs associated with large search spaces involving six-frame translations and potential SAAVs [25]. Tumours are also heterogeneous masses which may consisting of ~20% stromal and immune that further increases proteome complexity [27]. Laser capture microdissection combined with proteomics can help elucidate cancer- and stromal-specific factors but is manually intensive [28]. Unfortunately, we were unable to implement these strategies into our workflow given their significant technological barriers and ongoing development. Moving forward, advances in the field of proteogenomics will hopefully enable its accessibility for mainstream proteomic analysis of tumour biopsies.

6.4 Strategies for profiling cancer-derived biomarkers

When detected early (stage I/II), the 5-year overall survival for patients with EOC patients is 70% or higher [29]. Unfortunately, most patients present with advanced disease (stage III/IV) for which overall survival is dismal [30]. CA-125 plus TVUS screening may reduce mortality by a small margin but is associated with a high ratio of unnecessary follow-up surgeries to confirm a positive diagnosis [31,32]. Non-invasive blood based biomarkers may one day improve early detection of ovarian cancer. However, unbiased biomarker discovery approaches are hampered by the capacity to detect low abundance tumour derived proteins in biological fluids [33]. In chapter 4, EV preparations from different bio-fluids (conditioned media, plasma and ascites) were characterized by LC-MS/MS as an alternative method for profiling potential plasma based biomarkers of malignant EOC. Our proteomics analysis revealed a substantial number of factors associated with malignant disease. Furthermore, this analysis extends beyond traditional biomarker discovery approaches which focus on soluble signalling factors and highlights the potential value in surveying all populations of EVs for detecting EOC. This dataset constitutes the most in depth proteome of ovarian cancer ascites EVs to date and is a significant contribution to the field of biomarker research.

Although we extensively profiled and documented EV proteins associated with malignant EOC, according Pepe *et al.*, this work encroaches on the first of 5 phases of biomarker development, namely the pre-clinical exploratory phase. Therefore, to move towards the second phase of biomarker development – clinical assay and validation – we utilized a targeted PRM strategy as a proof-of-concept to determine the performance of EOC candidates in an independent cohort of plasma EVs. Notably, over a dozen candidates were found to harbour ROC-AUC values with similar or better performance than CA-125 for differentiating malignant from non-malignant plasma EVs. However, this approach can be further refined using isotopically labelled peptides for more accurate absolute quantification prior

to validation in larger patient cohorts (early versus late stage disease). While PRM assays may one day be used clinically, ELISAs may be better suited for detecting blood based markers in the near term. It remains to be confirmed whether membrane bound and intracellular EV markers can be readily detected in this format. Rapid and robust EV purification strategies are also needed to overcome the limitations of ultracentrifugation (time and low/variable sample recovery). Interestingly, one group reported a technique known as PRotein Organic Solvent Precipitation (PROSPR) to selectively isolate EVs [34]. In this approach, soluble plasma proteins are precipitated in acetone while lipid containing vesicles remain in the organic phase [34]. PROSPR was shown to be superior over ultracentrifugation and therefore merits further interrogation in our hands.

Once clinical assay development has been fully optimized, either using ELISAs or PRM assays, a substantial amount or work remains prior to clinical approval. This includes retrospective longitudinal studies, prospective screening studies and finally cancer control studies. However, TMT and advanced software detection algorithms can be implemented with PRM assays to facilitate high-throughput biomarker screening and validation in large case studies [35]. For example, Gygi's group developed and applied TOMAHAQ (triggered by offset, multiplexed, accurate-mass, high- resolution, and absolute quantification) as a high-throughput screening tool for identifying molecular targets associated with drug sensitivity [35]. In this study, 69 target proteins (131 peptides) were monitored by TOMAHAQ in the entire panel of NCI-60 cell lines. Impressively, TOMAHAQ reliably quantified 54 proteins in all cell lines (≥2 biological replicates) without fractionation in just 2 days and revealed a novel role for the DNA damage response protein BAZ1B in predicting Doxorubicin sensitivity. TOMAHAQ was only recently developed and investigated using cell lines but should be applicable to monitoring biomarkers in biological fluids.

6.5 Cancer plasticity associated with NODAL signalling differentially regulates the breast cancer secretome and stromal cell recruitment

In Chapters 3 and 4, I focused on proteomic characterization of clinical specimens. However, cancers are dynamic and plastic entities comprised of multiple cell types and extracellular factors within the TME. NODAL is one such embryonic morphogen which promotes metastatic phenotypes in cancer in part through regulating activities of cells within the TME [36,37]. In chapter 5, I combined proteomics with *in vitro* assays to test whether NODAL contributes directly or indirectly to stromal cell recruitment in breast cancer. Proteomic analysis of conditioned media from MDA-MB-231 and SUM149 breast cancer cell lines revealed unpublished alterations in a subset of cytokines and chemokines following NODAL knockdown or overexpression. Surprisingly, NODAL regulated these factors in an agonistic and antagonistic manner depending on the breast cancer cell line and model. These opposing functions are not dissimilar to its family member TGF-β which also signals though SMAD2/3 and promotes malignant progression in some cancers but suppression in others [38]. Moreover, several reports have shown TGF-β signalling to positively or negatively regulate IL6, CXCL1, and CXCL8(IL8) levels in various cell lines [39–44]. Although we show that NODAL induces global shifts in several cytokines/chemokines, additional work is needed understand the underlying mechanisms of this regulation. For example, does SMAD2/3 transiently regulate gene expression or are epigenetic changes involved that may be more permanent? Co-immunoprecipitation (CHIP) sequencing and/or bisulfite sequencing of these promoters may reveal differences in histone methylation/acetylation and DNA methylation. Moreover, phosphoproteomics and affinity-purification mass spectrometry can identify downstream factors associated with canonical and non-canonical signalling events that may vary between cell types [45,46]. Although we utilized cell lines stably

expressing constitutively active NODAL shRNA and overexpression vectors, tetracycline inducible systems can aid in evaluating the dynamics NODAL induction/inhibition. NODAL signalling can also be robustly targeted with the small molecule inhibitor SB431542 however this drug also blocks TGF- β and ACTIVIN pathways due to inhibition of their cognate receptors (ALK4/5/7) [47].

NODAL has been previously shown induce endothelial tube formation indirectly by altering the secretion pro-angiogenic factors in breast cancer cells [36]. In this thesis, I investigated whether NODAL could act could act on stromal cells (directly or indirectly) to promote migration *in vitro*. MSC were sensitive to changes in the breast cancer secretome while HFFs were not. Conversely, recombinant human NODAL protein activated SMAD2/3 and ERK1/2 in HFFs, but not MSC, and was associated with increased migration, invasion and expression of CAF markers. We examined how NODAL indirectly regulates MSC migration by profiling cognate receptors and neutralizing IL6 in breast cancer CM. Our findings suggest that factors in addition to IL6 are driving MSC recruitment. While the cognate receptors for CXCL1 and 8 (CXCR1/2) were not detected on MSC, these G-protein coupled receptors have been documented on other cell types in TME. For example CXCL8 can signal to CXCR1/2 present on endothelial cells and transactivate VEGFR2 to increase vascular permeability and tube formation [48]. Alternatively, CXCL1 can signal to CXCR2 expressing myeoloid derived suppressor cells (MDSCs) and neutrophils which are important for fighting infection but are also implicated in cancer [49]. In a study by Swarnali *et al*., CXCL1 produced by mammary tumours promoted CD11b+Gr1+ myeloid cell recruitment. Reciprocal signalling to breast cancer cells via myeloid cell derived S100A8/9 was linked to increased tumour growth and metastasis. Along this line, cell type specific labeling using amino acid precursors (CTAP) may be useful to investigate reciprocal signalling events between NODAL expressing cancer cell lines and stromal cell types [50]. In addition to our *in vitro* work, we also attempted to trace MSC recruitment to breast cancer xenografts *in vivo*. However, we experienced significant issues with MSC

lodging in the lungs following tail vein injections and/or extremely low MSC frequencies in mammary fat pad tumours following intraperitoneal injections (data not shown). Thus, mice transplanted with labelled bone marrow cells may be required to sufficiently determine the effects of NODAL on MSC recruitment in breast cancer xenografts.

In summary, NODAL promotes tumour growth *in vivo* and its expression is associated disease progression in many human cancers including melanoma, breast cancer and prostate cancer [51–53]. Our proteomics and *in vitro* findings reveal an additional, unknown layer of complexity for NODAL in regulating inflammatory factors and stromal cell recruitment in breast cancer. Consequently, future NODAL targeted therapies may block its intrinsic and extrinsic tumourigenic activities while sparing healthy, non-transformed tissues. A recent study by Hendrix *et al.* detected NODAL expression in melanoma lesions of deceased patients prior to, and following, treatment with BRAF inhibitors targeting V600E or V600K mutations [54]. This observation suggests NODAL expressing cancer cells do not respond or acquire resistance to conventional therapy. However, inhibiting BRAF with Dabrafenib in combination with an anti-NODAL mAb *in vivo* significantly reduces lung metastasis of the human melanoma cell line A375SM-L1, compared to either monotherapy alone [54]. Collectively, NODAL regulates multiple phenotypes associated with cellular plasticity in embryonic and malignant settings and therefore synergistic or complementary approaches targeting NODAL may improve outcomes for cancer patients.

6.6   Summary

The realm of proteomics is providing exciting opportunities for interrogating cancer biology. In this thesis, I employed and demonstrated the capacity of MS-based proteomics for elucidating proteins important for defining cancer subtypes, detecting disease using bio-fluids, and regulating intercellular communication and cellular behaviours within TME. Advances in proteomic technologies are beginning

to bridge the gap between the laboratory and clinical settings and identify specialized treatment modalities for targeting not only cancer cells but the TME as a whole.

6.7   References:

[1]   Kuljanin, M., Dieters-Castator, D.Z., Hess, D.A., Postovit, L.-M., Lajoie, G.A., Comparison of sample preparation techniques for large-scale proteomics. *Proteomics* 2017, 17, 1600337.

[2]   Huang, F.-K., Zhang, G., Lawlor, K., Nazarian, A., et al., Deep Coverage of Global Protein Expression and Phosphorylation in Breast Tumor Cell Lines Using TMT 10-plex Isobaric Labeling. *J. Proteome Res.* 2017, 16, 1121–1132.

[3]   Hughes, C.S., McConechy, M.K., Cochrane, D.R., Nazeran, T., et al., Quantitative Profiling of Single Formalin Fixed Tumour Sections: proteomics for translational research. *Sci. Rep.* 2016, 6, 34949.

[4]   Kelstrup, C.D., Jersie-Christensen, R.R., Batth, T.S., Arrey, T.N., et al., Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* 2014, 13, 6187–95.

[5]   Bekker-Jensen, D.B., Kelstrup, C.D., Batth, T.S., Larsen, S.C., et al., An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* 2017, 4, 587–599.e4.

[6]   Kulak, N.A., Geyer, P.E., Mann, M., Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. *Mol. Cell. Proteomics* 2017, 16, 694–705.

[7]   Ruprecht, B., Zecha, J., Zolg, D.P., Kuster, B., High pH Reversed-Phase Micro-Columns for Simple, Sensitive, and Efficient Fractionation of Proteome and (TMT labeled) Phosphoproteome Digests. *Methods Mol. Biol.* 2017, 1550, 83–98.

[8]   Han, D., Moon, S., Kim, Y.Y.Y.Y., Kim, J., et al., In-depth proteomic analysis of mouse microglia using a combination of FASP and StageTip-based, high pH, reversed-phase fractionation. *Proteomics* 2013, 13, 2984–8.

[9]   Gilar, M., Olivova, P., Daly, A.E., Gebler, J.C., Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* 2005, 28, 1694–1703.

[10]   Wang, Y., Yang, F., Gritsenko, M. a., Wang, Y., et al., Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 2011, 11, 2019–2026.

[11]   Tsiatsiani, L., Heck, A.J.R., Proteomics beyond trypsin. *FEBS J.* 2015, 282, 2612–2626.

[12]   Swaney, D.L., Wenger, C.D., Coon, J.J., Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* 2010, 9, 1323–1329.

[13]   Meyer, B., Papasotiriou, D.G., Karas, M., 100% protein sequence coverage: A modern form of surrealism in proteomics. *Amino Acids* 2011, 41, 291–310.

[14]   Davis, S., Charles, P.D., He, L., Mowlds, P., et al., Expanding proteome coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) combined with broad specificity

proteolysis Expanding proteome coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) combined with broad specificity proteolysis 2017.

[15]   Cristobal, A., Marino, F., Post, H., van den Toorn, H.W.P., et al., Towards an optimized workflow for middle-down proteomics. *Anal. Chem.* 2017, acs.analchem.6b03756.

[16]   Tran, N.H., Rahman, M.Z., He, L., Xin, L., et al., Complete de Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* 2016, 6, 1–10.

[17]   Wiśniewski, J.R., Mann, M., Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal. Chem.* 2012, 84, 2631–2637.

[18]   Hebert, A.S., Thöing, C., Riley, N.M., Kwiecien, N.W., et al., Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Anal. Chem.* 2018, 90, 2333–2340.

[19]   Baker, M.S., Ahn, S.B., Mohamedali, A., Islam, M.T., et al., Accelerating the search for the missing proteins in the human proteome. *Nat. Commun.* 2017, 8, 14271.

[20]   Cancer Genome Atlas Research Network, Integrated genomic analyses of ovarian carcinoma. *Nature* 2011, 474, 609–15.

[21]   Köbel, M., Bak, J., Bertelsen, B.I., Carpen, O., et al., Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry. *Histopathology* 2014, 64, 1004–1013.

[22]   Coscia, F., Watters, K.M., Curtis, M., Eckert, M.A., et al., Integrative proteomic profiling of ovarian cancer cell lines reveals precursor cell associated proteins and functional status. *Nat. Commun.* 2016, 7, 12645.

[23]   McAlister, G.C., Nusinow, D.P., Jedrychowski, M.P., Wühr, M., et al., MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* 2014, 86, 7150–8.

[24]   Sonnett, M., Yeung, E., Wuhr, M., Accurate, Sensitive, and Precise Multiplexed Proteomics using the Complement Reporter Ion Cluster. *bioRxiv* 2017, 205054.

[25]   Ruggles, K. V., Krug, K., Wang, X., Clauser, K.R., et al., Methods, Tools and Current Perspectives in Proteogenomics. *Mol. Cell. Proteomics* 2017, 16, 959–981.

[26]   Ma, B., Zhang, K., Hendrie, C., Liang, C., et al., PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003, 17, 2337–2342.

[27]   Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., et al., Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 2013, 4.

[28]   Staunton, L., Tonry, C., Lis, R., Espina, V., et al., Pathology-Driven Comprehensive Proteomic Profiling of the Prostate Cancer Tumor Microenvironment. *Mol. Cancer Res.* 2017, 15, 281–293.

[29]   American Cancer Society, Survival Rates for Ovarian Cancer, by Stage 2018.

[30]   Jelovac, D., Armstrong, D.K., Recent progress in the diagnosis and treatment of ovarian cancer. *CA Cancer J Clin* 2011, 61, 183–203.

[31]   Jacobs, I.J., Menon, U., Ryan, A., Gentry-Maharaj, A., et al., Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): A randomised controlled trial. *Lancet* 2016, 387, 945–956.

[32]   Buys, S.S., Partridge, E., Black, A., Johnson, C.C., et al., Effect of Screening on Ovarian Cancer Mortality. *JAMA* 2011, 305, 2295–303.

[33]   Hanash, S.M., Pitteri, S.J., Faca, V.M., Mining the plasma proteome for cancer biomarkers. *Nature* 2008, 452, 571–579.

[34]   Gallart-Palau, X., Serra, A., Wong, A.S.W., Sandin, S., et al., Extracellular vesicles are rapidly purified from human plasma by PRotein Organic Solvent PRecipitation (PROSPR). *Sci. Rep.* 2015, 5, 14664.

[35]   Erickson, B.K., Rose, C.M., Braun, C.R., Erickson, A.R., et al., A Strategy to Combine Sample Multiplexing with Targeted Proteomics Assays for High-Throughput Protein Signature Characterization. *Mol. Cell* 2017, 65, 361–370.

[36]   Quail, D.F., Walsh, L. a., Zhang, G., Findlay, S.D., et al., Embryonic protein nodal promotes breast cancer vascularization. *Cancer Res.* 2012, 72, 3851–3863.

[37]   Quail, D.F., Zhang, G., Findlay, S.D., Hess, D. a, Postovit, L.-M., Nodal promotes invasive phenotypes via a mitogen-activated protein kinase-dependent pathway. *Oncogene* 2014, 33, 461–73.

[38]   Massagué, J., TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* 2012, 13, 616–630.

[39]   Park, J.-I., Lee, M.-G., Cho, K., Park, B.-J., et al., Transforming growth factor-beta1 activates interleukin-6 expression in prostate cancer cells through the synergistic collaboration of the Smad2, p38-NF-kappaB, JNK, and Ras signaling pathways. *Oncogene* 2003, 22, 4314–4332.

[40]   Qi, W., Chen, X., Holian, J., Mreich, E., et al., Transforming growth factor-beta1 differentially mediates fibronectin and inflammatory cytokine expression in kidney tubular cells. *Am. J. Physiol. Renal Physiol.* 2006, 291, F1070–F1077.

[41]   Gupta, J., Robbins, J., Jilling, T., Seth, P., TGFβ-dependent induction of interleukin-11 and interleukin-8 involves SMAD and p38 MAPK pathways in breast tumor models with varied bone metastases potential. *Cancer Biol. Ther.* 2011, 11, 311–6.

[42]   Michaeloudes, C., Sukkar, M.B., Khorasani, N.M., Bhavsar, P.K., Chung, K.F., TGF-β regulates Nox4, MnSOD and catalase expression, and IL-6 release in airway smooth muscle cells. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 2011, 300, L295–L304.

[43]   Fang, W. Bin, Mafuvadze, B., Yao, M., Zou, A., et al., TGF-β negatively regulates CXCL1 chemokine expression in mammary fibroblasts through enhancement of smad2/3 and suppression of HGF/c-met signaling mechanisms. *PLoS One* 2015, 10, 1–21.

[44]   Bierie, B., Chung, C., Abrogation of TGF-β signaling enhances chemokine production and correlates with prognosis in human breast cancer. *J Clin Invest* 2009, 119, 1571–1582.

[45]   Dunham, W.H., Mullin, M., Gingras, A.C., Affinity-purification coupled to mass spectrometry: Basic principles and strategies. *Proteomics* 2012, 12, 1576–1590.

[46]   Batth, T.S., Francavilla, C., Olsen, J. V, Off-Line High-pH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* 2014, 13, 6176–6186.

[47]   Inman, G.J., SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor-beta Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* 2002, 62, 65–74.

[48]   Ha, H., Debnath, B., Neamati, N., Role of the CXCL8-CXCR1/2 axis in cancer and inflammatory diseases. *Theranostics* 2017, 7, 1543–1588.

[49]   Coffelt, S.B., Wellenstein, M.D., De Visser, K.E., Neutrophils in cancer: Neutral no more. *Nat. Rev. Cancer* 2016, 16, 431–446.

[50]   Gauthier, N.P., Soufi, B., Walkowicz, W.E., Pedicord, V.A., et al., Cell-selective labeling using amino acid precursors for proteomic studies of multicellular environments. *Nat. Methods* 2013, 10, 768–773.

[51]   Postovit, L.-M., Margaryan, N. V, Seftor, E.A., Kirschmann, D. a, et al., Human embryonic stem cell microenvironment suppresses the tumorigenic phenotype of aggressive cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105, 4329–34.

[52]   Lawrence, M.G., Margaryan, N. V., Loessner, D., Collins, A., et al., Reactivation of embryonic nodal signaling is associated with tumor progression and promotes the growth of prostate cancer cells. *Prostate* 2011, 71, 1198–209.

[53]   Topczewska, J.M., Postovit, L.-M., Margaryan, N. V, Sam, A., et al., Embryonic and tumorigenic pathways converge via Nodal signaling: role in melanoma aggressiveness. *Nat. Med.* 2006, 12, 925–932.

[54]   Hendrix, M.J., Kandela, I., Mazar, A.P., Seftor, E.A., et al., Targeting melanoma with front-line therapy does not abrogate Nodal-expressing tumor cells. *Lab. Investig.* 2016, 00, 1–11.

Appendices

Appendix A: Copyright permission from Springer Nature

**SPRINGER NATURE LICENSE
TERMS AND CONDITIONS**

Mar 02, 2018

This Agreement between Western University -- Dylan Dieters-Castator ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4300920501488
License date	Mar 02, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Cancer
Licensed Content Title	The disparate origins of ovarian cancers: pathogenesis and prevention strategies
Licensed Content Author	Anthony N. Karnezis, Kathleen R. Cho, C. Blake Gilks, Celeste Leigh Pearce, David G. Huntsman
Licensed Content Date	Nov 25, 2016
Licensed Content Volume	17
Licensed Content Issue	1
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	PhD candidate
Instructor name	Dr. Lynne-Marie Postovit
Institution name	Western University
Expected presentation date	Apr 2018
Portions	Figures 1 and 4
Requestor Location	Western University 1400 Western Road Siebens-Drake Research Institute Room G31 London, ON n6g2v4 Canada Attn: Dylan
Billing Type	Invoice
Billing Address	Western University 1400 Western Road Siebens-Drake Research Institute

Appendix B: Copyright permission from John Wiley and Sons

**JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS**

Mar 23, 2018

This Agreement between Western University -- Dylan Dieters-Castator ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4314900347841
License date	Mar 23, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Proteomics
Licensed Content Title	Comparison of sample preparation techniques for large-scale proteomics
Licensed Content Author	Miljan Kuljanin, Dylan Z. Dieters-Castator, David A. Hess, et al
Licensed Content Date	Jan 18, 2017
Licensed Content Volume	17
Licensed Content Issue	1-2
Licensed Content Pages	9
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	PhD candidate
Expected completion date	Apr 2018
Expected size (number of pages)	1
Requestor Location	Western University 1400 Western Road Siebens-Drake Research Institute Room G31 London, ON n6g2v4 Canada Attn: Dylan
Publisher Tax ID	EU826007151
Total	0.00 CAD
Terms and Conditions	

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work

Appendix C: Health and Research Ethic Board of Alberta Cancer Committee Certification of Ethics Approval

HREBA | Health Research Ethics Board of Alberta Cancer Committee

Health Research Ethics Board of Alberta
Cancer Committee
1500, 10104 - 103 Avenue NW
Edmonton, Alberta. T5J 0H8
Telephone: (780) 423-5727
Fax: (780) 429 3509
Email: cancer@hreba.ca

Certification of Ethics Approval - Renewal

This is to acknowledge that the renewal to the research indicated below has been reviewed and on behalf of the Health Research Ethics Board of Alberta (HREBA) – Cancer Committee (CC), I am pleased to advise that approval has been granted.

Ethics ID:	HREBA.CC-17-0450_REN1
Principal Investigator:	YangXin Fu
Co-Investigator(s):	Helen Steed
Student Co-Investigator(s):	There are no items to display
Study Title:	Identifying the altered gene expression in ovarian cancer tissues and primary ovarian cancer cells using normal ovarian surface epithelium as a control.
Sponsor:	Faculty of Medicine and Dentistry of the University of Alberta Women and Children's Health Research Institute

Effective: December 27, 2017   **Expires:** December 26, 2018

The Annual Progress Report was reviewed by delegated review on 27 December 2017.

This Committee is constituted and operates in accordance with the Alberta Health Information Act (HIA), the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2), Good Clinical Practice (GCP) Guidelines of the International Conference on Harmonization (ICH), Health Canada's *Food and Drug Regulations* (FDR), Part C, Division 5 and is registered with the U.S. Department of Health and Human Services (HHS), Office for Human Research Protections (OHRP), IRB # 00009687.

Members of the HREBA-CC who are named as principal investigators or co-investigators in this research do not participate in discussions related to, nor vote on, such studies when they are presented to the Committee. The membership of this Committee is listed at www.hreba.ca.

This renewal is subject to the following conditions:

1. Approval is granted only for the research described in this application.

Appendix D: Curriculum Vitae

Dylan Dieters-Castator

EDUCATION

Western University, London ON, Canada

    PhD in Anatomy and Cell Biology   2011 – 2018
    Proteomic Characterization of Ovarian and Breast Cancer Microenvironments for
    Improved Diagnostics and Therapeutic Targeting
    Supervisors: Dr. Lynne-Marie Postovit and Dr. Gilles Lajoie
    Bachelor of Medical Sciences   2007 – 2011
    Honours Specialization in Biochemistry and Cell Biology
    Minor in Medical Sciences
    Deans Honour Role (≥80% average)

AWARDS AND SCHOLARSHIPS

    CNPN Travel Award   2017
    CIHR Strategic Training Program in Cancer Research and   2011 – 2015
    Technology Transfer Awards (MSc and PhD)
    The Role of Stem Cell Recruitment in Nodal-Associated Neovascularization
    Schulich Graduate Research Scholarship   2011 – 2015
    Ontario Graduate Scholarship   2012 – 2015
    The Role of Stem Cell Recruitment in Nodal-Associated Neovascularization
    Ontario Stem Cell Initiative Travel Award   2013
    CIHR Strategic Training Program in Cancer Research and   2010
    Technology Transfer Summer Studentship
    Mechanisms Underlying Hypoxia-Induced Nodal Expression in Cancer

PUBLICATION CONTRIBUTIONS

Accepted

    Kuljanin M, Dieters-Castator DZ*, Hess DA, Postovit L-M, Lajoie GA. Comparison of sample preparation techniques for large-scale proteomics. Proteomics. 2017. doi:10.1002/pmic.201600337

    Quail DF, Taylor MJ, Walsh LA., Dieters-Castator DZ*, Das P, Jewer M, Zhang G, Postovit LM. Low oxygen levels induce the expression of the embryonic morphogen Nodal. Mol Biol Cell. 2011. doi:10.1091/mbc.E11-03-0263.

    Quail DF, Zhang G, Walsh LA., Siegers GM, Dieters-Castator DZ*, Findlay SD, Broughton H, Putman DM, Hess DA, Postovit LM. Embryonic Morphogen Nodal Promotes Breast Cancer Growth and Progression. PLoS One. 2012. doi:10.1371/journal.pone.0048237

In preparation

    Dieters-Castator DZ*, Piaseczny M, Zhang G, Kuljanin M, Liu J, Siegers G, Jewer M, Sherman S, Hess DA, Lajoie GA, and Postovit L-M. Embryonic protein Nodal mediates stromal cell chemotaxis to breast cancer cells and broadly regulates secretome composition.

    Dieters-Castator, DZ*, Liu J, Lajoie GA, and Postovit L-M. Proteomic profiling of ovarian cancer extracellular vessicles for biomarker discovery.

    Dieters-Castator DZ*, Rambau P, Lajoie GA, Köbel M, and Postovit LM. Proteomic-based discovery of putative biomarkers for improved classification of endometrioid and high grade serous ovarian cancer subtypes.

CONFERENCE ABSTRACTS AND PRESENTATIONS

Dieters-Castator D*, Rambau R, Peter Rambau, Lajoie G, Postovit LM, Köbel M. Mass spectrometry-based proteomics for improved discrimination of endometrioid from high-grade serous carcinomas. CCOCR, Edmonton, AB. 2018 [poster]

Dieters-Castator D*, Kuljanin M, Hess D, Lajoie G, Postovit LM. Characterizing the Nodal-regulated breast cancer secretome and its role in multipotent stromal cell recruitment. CNPN 2017, Toronto, ON. [poster]

Dieters-Castator D*, Kuljanin M, Hess D, Lajoie G, Postovit LM. Characterizing the Nodal-regulated breast cancer secretome and its role in multipotent stromal cell recruitment. 14th International Tumour Microenvironment Workshop 2015, Vancouver, BC [talk]

Dieters-Castator D*, Lajoie G, Postovit LM. Characterizing the Nodal-regulated breast cancer secretome and its role in human bone marrow multipotent stromal cell mediated tumourigenesis. CNPN 2014, Montreal, QC. [poster]

Dieters-Castator D*, Lajoie G, Postovit LM. Characterizing the Nodal-regulated breast cancer secretome and its role in human bone marrow mesenchymal stem cell mediated tumourigenesis. ASMS 2014, Baltimore, MD. [poster]

Dieters-Castator D*, Lajoie G, Postovit LM. The Role of Bone Marrow Derived Cells in Nodal-Associated Tumour Neovascularization. CCRC 2013, Toronto, ON. [poster]

Dieters-Castator D*, Lajoie G, Postovit LM. The Role of Bone Marrow Derived Cells in Nodal-Associated Neovascularization. ISSCR 2013, Boston, MA. [poster]

Dieters-Castator D*, Lajoie G, Postovit LM. The Role of Bone-Marrow Derived Cells in Nodal-Associated Tumour Neovascularization. CIHR Strategic Training Program in Cancer Research and Technology Transfer, Research and Education Day 2012. London, ON. [poster]

TEACHING EXPERIENCE
Western University, London ON, Canada

Teaching Assistantship: Translational Models of Cancer	2012
4th year Undergraduate Honours Thesis Project Mentor	2011 – 2012

PROFESSIONAL DEVELOPMENT AND WORKSHOPS
Max Planck Institute of Biochemistry, Munich, Germany

Computational Mass Spectrometry-Based Proteomics 5TH Maxquant Summer School	2013

Toronto ON, Canada

Bioinformatics Solutions PEAKS Software Training Course	2013

Western University, London ON, Canada

Schulich Biotech Entrepreneur Boot Camp #1092	2014
Pathology 9520 – Public and Private Partnerships in Cancer Research	2012 – 2013
Biochem 9523B – Advanced Methods in Biochemistry	2013
Biochem 9545B – Macromolecular Informatics	2013
DEV9000 – Developmental Biology	2012 – 2013
4th year Honours Thesis Project Supervisor: Dr. Lynne-Marie Postovit, *Dept. of Anatomy and Cell Biology*	2010 – 2011
Work Study Research Position Supervisor: Dr. Dean Betts, *Dept. of Pharmacology and Physiology*	2010 – 2011
Work Study Research Position Supervisor: Dr. Lynne-Marie Postovit,	2009 – 2010

Dept. of Anatomy and Cell Biology

EXTRACURRICULAR AND VOLUNTEER

Stem Cell Western Journal Club	2009 – 2017
Let's Talk Science Teacher Partnership	2010 – 2011

MEMBERSHIPS

American Society for Mass Spectrometry	2014 – 2015
International Society for Stem Cell Research	2013 – 2014
Canadian National Proteomics Network	2011 – 2017
Society for Developmental Biology	2011 – 2012