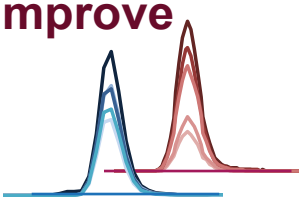
Universitat Autònoma de Barcelona

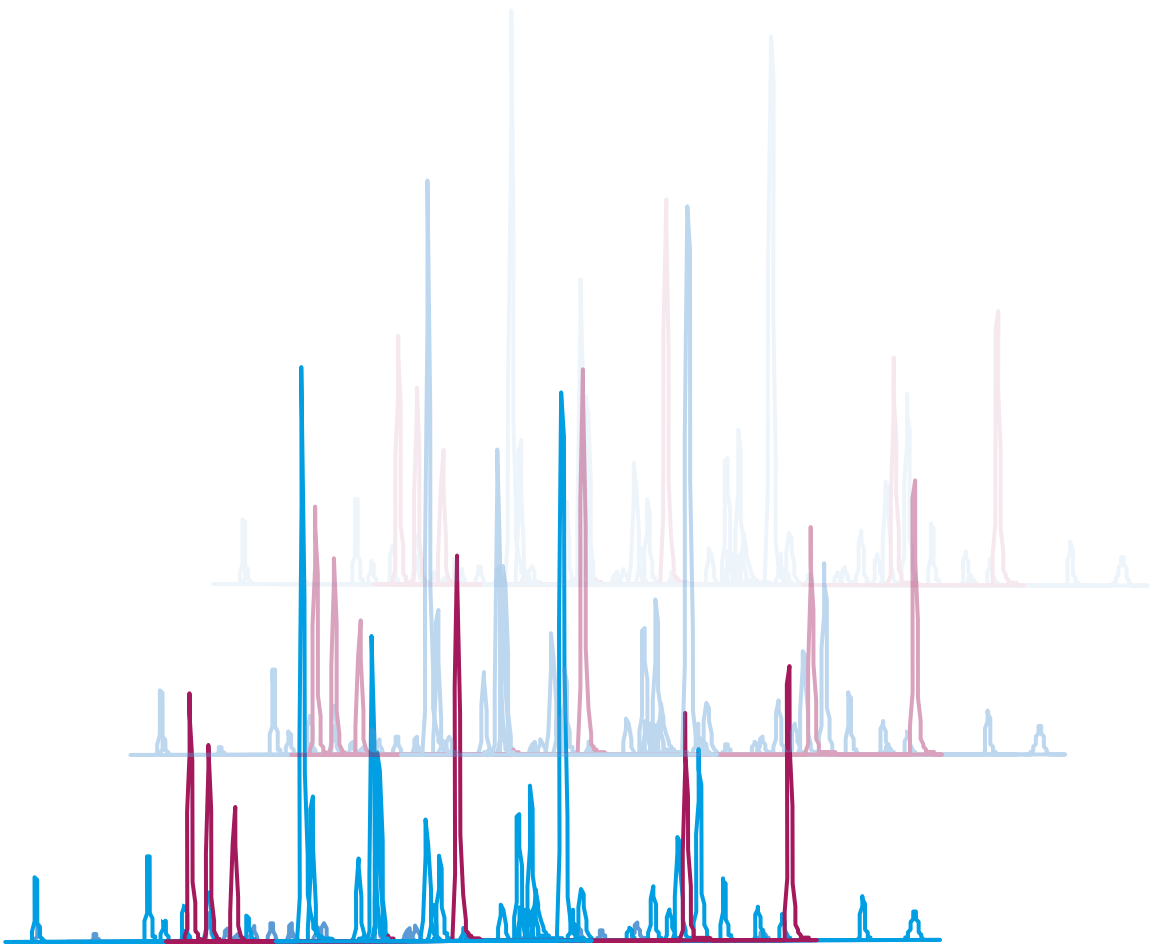
Protein signatures in uterine aspirates to improve diagnosis of endometrial cancer

Elena Martínez García

Protein signatures in uterine aspirates to improve diagnosis of endometrial cancer

The CEMARK project





Elena Martínez García

Protein signatures in uterine aspirates to improve diagnosis of endometrial cancer

"The CEMARK project"

Doctoral thesis presented by

Elena Martínez García

to obtain the degree of

Doctora por la Universitat Autònoma de Barcelona (UAB)

Doctoral thesis performed at the Vall d´Hebrón Research Institute, in the Group of Biomedical Research in Gynecology, under the supervision of **Dr. Jaume Reventós, Dr. Eva Colás and Dr. Antonio Gil**.

Doctoral study in Cellular Biology, Department of Cellular Biology, Physiology and Immunology, Faculty of Medicine at the UAB, under the supervision of **Dr. Imma Ponsa**.

Universitat Autònoma de Barcelona, 12th June 2017

Dr. Jaume Reventós (director) Dra. Eva Colás (director) Dr. Antonio Gil (director)

Dra. Imma Ponsa (tutor) Elena Martínez (student)

A mi familia, en

especial a mis padres

Index

INDEX i

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A

Ab: Antibody
AUB: Abnormal uterine bleeding
AUC: Area under the ROC curve

B

BMI: Body-mass index
BSA: Bovine serum albumin

C

c-Ab: Capture antibody
CAPG: Macrophage-capping protein
CI: Confidence interval
CID: Collision induced dissociation
cos θ: Cosine of the spectral contrast angle
C-SPE: Carbon screen-printed electrodes
CTNNB1: Beta-catenin
CV: Coefficient of variation

D

2D-DIGE: Two-dimensional differential gel electrophoresis
2DE: Two-dimensional electrophoresis
D&C: Dilatation and curettage
d-Ab: Detection antibody
DDA: Data dependent acquisition
DIA: Data independent acquisition
DMSO: Dimethyl sulfoxide
DTT: Dithiothreitol

E

EC: Endometrial cancer
EDTA: Ethylenediaminetetraacetic acid
EEC: Endometrioid endometrial cancer
EIN: Endometrial intraepithelial neoplasia
ELISA: Enzyme-linked immunosorbent assay
EMA: European Medicines Agency
ESGO: European Society of Gynaecological Oncology
ESI: Electrospray ionization
ESMO: European Society for Medical Oncology
ESTRO: European Society for Radiotherapy&Oncology

F

FBS: Fetal Bovine Serum
FDA: Food and Drug Administration
FDR: False discovery rate
FFPE: Formalin-fixed paraffin-embedded

FIGO: Federation of Gynecology and Obstetrics
FSH: Follicle stimulating hormone
FTICR: Fourier transform ion cyclotron resonance

G

GnRH: Gonadotropin releasing hormone

H

hCG: Chorionic gonadotrophin
HPLC: High pressure liquid chromatography
HRAM: High resolution accurate mass spectrometry
HRP: Horseradish peroxidase

I

IAA: Iodoacetamide
iCAT: Isotope coded affinity tagging
IgG: Immunoglobulin G
IHQ: Immunohistochemistry
IPA: Ingenuity Pathway Analysis
iTRAQ: Isobaric tags for relative and absolute quantitation
IVD: In vitro diagnostics

K

KPYM: Pyruvate kinase

L

L/H: Endogenous/synthetic peptide ratio
LC-MS: Liquid chromatography coupled to mass spectrometry
LH: Luteinizing hormone
L1CAM: L1 cell adhesion molecule

M

m/z: Mass over charge ratio
MALDI-TOF: Matrix-assisted laser desorption/ionization time-of-flight
MMP9: Metalloproteinase-9
MRI: Magnetic resonance imaging
MS: Mass spectrometry

N

NEEC: Non-endometrioid endometrial carcinoma

P

PBS: Phosphate-buffered saline
PBS-T: Phosphate-buffered saline with 0.05% Tween-20
POLE: DNA polymerase epsilon catalytic subunit
PolyHRP: Polymeric streptavidin-HRP complexes
PRM: Parallel reaction monitoring
PSA: Prostate-specific antigen

Q

Q-OT: Quadrupole orbitrap
QqQ: Triple quadrupole

QqTOF: Quadrupole time of flight

R

R coefficient: Pearson correlation coefficient
ROC curve: Receiver operating characteristic curve
RPPA: Reverse-phase protein array

S

SCNAs: Somatic copy number alterations
SD: Standard deviation
SEC: Non-endometrioid serous ECs
SEGO: Spanish Society of Gynecology and Obstetrics
SIL peptides: Stable isotopes labeled peptides
SILAC: Stable isotope labeling by amino acids in cell culture
SPE: Solid phase extraction
SRM: Selected reaction monitoring
SWATH: Sequential window acquisition of all theoretical fragment-ion spectra

T

TCGA: The Cancer Genome Atlas
TMB: 3,3′,5,5′-Tetramethylbenzidine
TMT: Tandem mass tags
TOF: Time-of-flight

W

WHO: World Health Organization

X

XIC: Extracted ion chromatograms
XPO2: Exportin-2

Introduction

1. THE ENDOMETRIUM

1.1. ANATOMY AND FUNCTION OF THE UTERUS

The uterus is a female reproductive sex organ. It is a hollow, thick-walled, muscular organ with the shape of an inverted pear. It is located in the pelvic cavity above the vagina, above and behind the bladder and in front of the rectum (Figure 1). Although the shape and position of the uterus changes dramatically during pregnancy, in its non-pregnant state, it measures on average 8 cm long, 5 cm wide and 2.5 cm in diameter [1]. The uterus has three major regions: the **fundus**, the curved upper area where the Fallopian tubes connect to the uterus – it is the wider part of the organ, of around 5 cm; the **body**, the enlarged main part which lies below the fundus and is separated from the cervix by a slight constriction, the isthmus; and the **cervix**, the inferior cylindrical narrow part of 2-3 cm length that opens into the vagina (Figure 1).

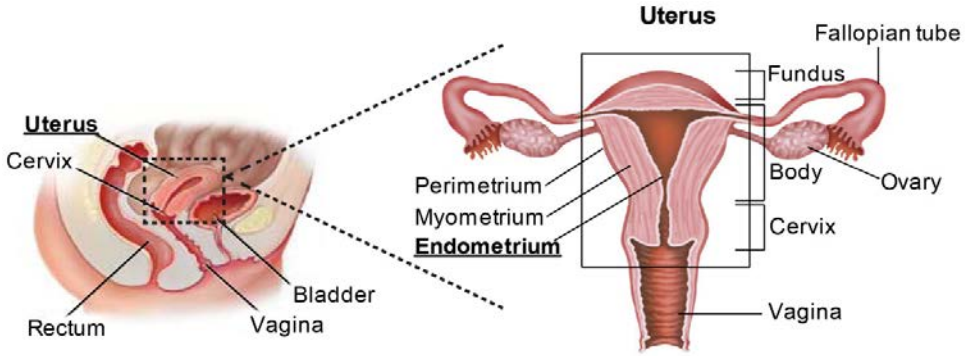


Figure 1. Location and anatomical parts of the uterus.

The wall of the uterus is thick and has 3 layers [2] (Figure 1):

- The **perimetrium**, the outer serosal layer of the uterus that consists of a thin visceral peritoneum.
- The **myometrium**, the thicker middle layer (1-2 cm thick) of the uterus composed primarily of smooth muscle cells in a framework of arterial and venous blood supply supported by an underlying connective tissue. It allows the uterus to expand during pregnancy and to contract during childbirth.
- The **endometrium**, the inner hormone-responsive mucosal lining of the uterus.

The functional uterus is an essential organ for reproduction. At ovulation, the wall of an ovarian follicle ruptures, releasing a secondary oocyte that passes into the uterine cavity through the Fallopian tubes. If fertilization occurs, the blastocyst implants on the endometrial

lining. As the embryo develops into a fetus, it triggers changes within the endometrium that lead to the formation of the placenta. The placenta acts to provide oxygen and nutrients to the fetus, whilst removing carbon dioxide and other waste products [3]. At the end of pregnancy, the uterus plays a critical role in the process of childbirth. It dilates the cervix and contracts the myometrium at term to deliver the fetus, through the vagina, out of the mother´s body [4].

1.2. HISTOLOGY OF THE ENDOMETRIUM

The endometrium consists of three main constitutive elements: i) a simple columnar epithelium of ciliated and secretory cells in the surface; ii) an underlying matrix of cellular connective tissue stroma containing a rich supply of blood vessels; and iii) simple tubular uterine glands formed by invagination of the epithelium that extend through the entire thickness of the stroma (Figure 2). Those epithelial and stromal cells coexist with other minor subtypes of cells, such as lymphocytes and other leucocytes, among others [2]. Around 90% of endometrial carcinomas originate from the epithelial glands in the endometrium.

1.3. PHASES OF THE ENDOMETRIUM

1.3.1. Functional endometrium

The functional endometrium has a lifetime equal to the reproductive life of an adult, i.e., from the menarche to the menopause. The main characteristic of the endometrium is its perfect response to the hormone stimuli from the ovaries [1]. This response produces important morphologic changes in all the elements of the endometrium during the endometrial cycle, in order to prepare itself for receiving a fertilized oocyte [5]. If the fertilization does not occur, the functional layer of the endometrium is shed and expulsed, leading to menstruation, which occurs approximately every 28 days.

Functionally, the endometrium can be divided into two layers based on their involvement in the changes of the menstrual cycle [2].

- The **functional layer** is the luminal part of the endometrium. It is the part that suffers the cyclic changes, it is shed at the time of menstruation and built up again under the stimulation of ovarian steroid hormones.

- The **basal layer** is located below the functional layer, adjacent to the myometrium. It is highly vascular, it changes little during the menstrual cycle and is not shed during the menstruation. It serves to regenerate the functional layer after each menstruation.

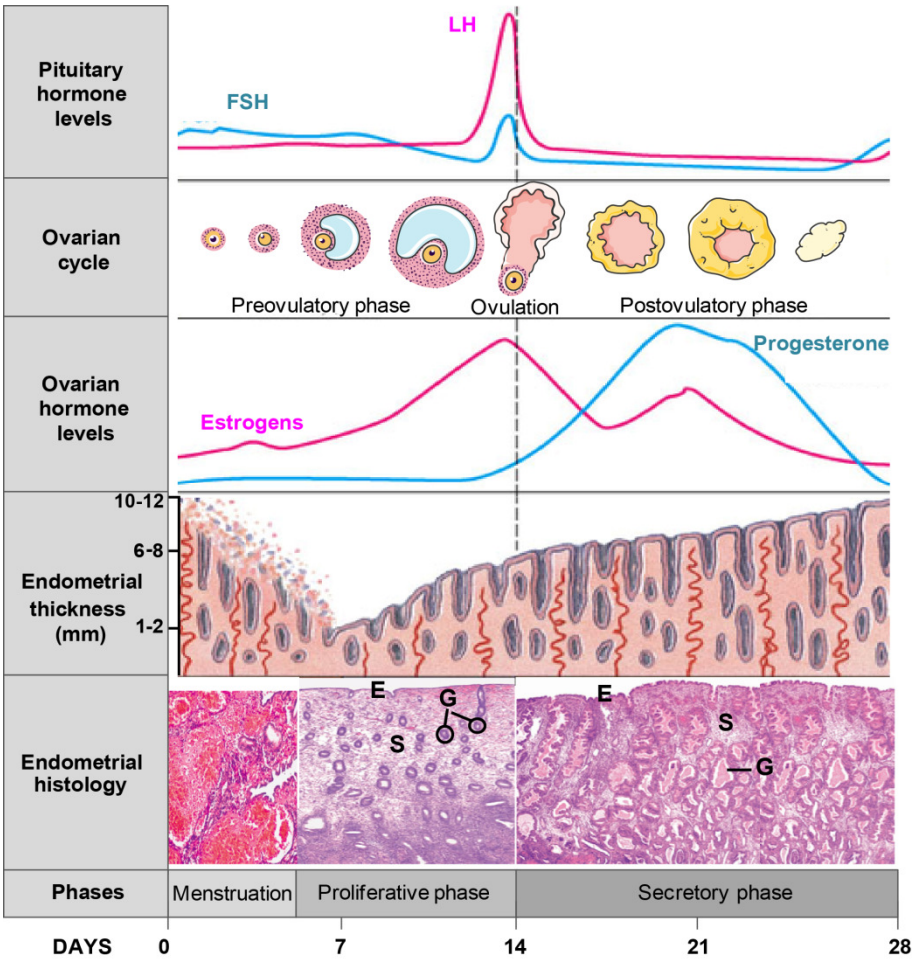


Figure 2. Endometrial cycle and endometrial histology. E, simple columnar epithelium; S, stroma; G, uterine glands.

The functional endometrium goes through four different stages, related to the three phases of the endometrial cycle (proliferative, secretory, and menstrual) and to pregnancy. All of them have different histological and molecular features (Figure 2).

1.3.1.1. Proliferative endometrium

The proliferative endometrium is present during the phase prior to the ovulation period, between days 4 and 14 [1].

Immediately after the menstruation, the thickness of the endometrium is 1-2 mm. The proliferative endometrial tissue is characterized by its great capacity to proliferate and increase its thickness as a result of estrogen stimulation.

The proliferative endometrium is composed of rectilinear endometrial glands, which are delimited by pseudo-stratified nuclear cells and mitotic components. There is a prominent mitotic activity in both the glands and the stroma. The stroma is dense with cells showing a reduced cytoplasmic fraction [6,7].

During these days, the hypothalamus releases gonadotropin-releasing hormone (GnRH), which stimulates the anterior pituitary gland to release follicle-stimulating hormone (FSH) and luteinizing hormone (LH). FSH stimulates the maturation of a follicle within an ovary. Ovulation occurs on day 14 and it is associated with a high release of estrogen and an LH surge.

1.3.1.2. Secretory endometrium

The secretory endometrium is present just after the ovulation period, between days 15 and 28. From day 20 to 24 of the cycle, the "window for implantation" occurs, which is the period of time when endometrial receptivity for blastocyst implantation is at its highest [1].

Changes in the endometrium during this phase are oriented to prepare a nutritive soil for the blastocyst implantation and are supported by the appearance of progesterone.

This endometrial tissue is characterized by tortuous endometrial glands that continue growing and by the predecidual transformation of stromal cells, which increase their cytoplasmic fraction and acquire a cubical morphology. A particular phenomenon of this phase is the appearance of subnuclear vacuoles containing glycogen and mucopolysaccharides. These vacuoles move to the luminal surface where they are secreted. The stroma becomes edematous and spiral arteries appear [6,7]. The endometrium in this phase can increase its thickness up to 8 mm.

1.3.1.3. Menstrual endometrium

This menstrual endometrium is present during the menstrual phase, between days 1 and 4 of the endometrial cycle [1], and it is formed if the oocyst is not fertilized.

In the absence of conception, chorionic gonadotrophin (hCG) is not produced. In a first phase, the levels of progesterone decrease and a structural detachment of the functional layer of the endometrium occurs along with increased leucocyte infiltration, hemorrhage and necrosis. In a second phase, and as a consequence of the hormonal stimuli, the endometrium begins its regeneration from the basal layer which was not dissociated.

1.3.1.4. Endometrium at pregnancy

This endometrial tissue is only formed if the oocyst has been fertilized and the blastocyst has implanted and begun to proliferate within the endometrium. In this case, the endometrium increases its hypertrophy and secretions. Endometrial glands are rich in glycogen and stromal cells decidualize, becoming bigger, polygonal, with a high fraction of cytoplasm [1].

1.3.2. Atrophic endometrium

The non-cycling endometrium becomes gradually atrophic after the menopause. Nowadays, almost a third of the life of a woman is postmenopausal, and over 80% of endometrial carcinomas develop during this period [8].

In this phase, the endometrium suffers important histological modifications and loses its ability to proliferate and secrete. The main reason for these modifications is the privation of estrogen and progesterone. The atrophic endometrium is thin (1 to 3 mm thick), with loss of distinction between the basal layer and the functional layer. Histologically, it maintains the three main constitutive elements but with significant alterations: i) small tubular glands may initially retain some proliferative activity, although weak; however, with further decline of estrogen secretion they become functionally inactive; ii) glands are widely spaced, lined by cuboidal epithelium showing neither secretory nor proliferative activity; iii) the stroma is dense and fibrous. With complete absence of ovarian function, the endometrium falls into cystic atrophy, ending up as a thin layer full of cystically dilated endometrial glands lined by a flattened inactive epithelium [8,9].

However, estrogen stimulation may continue to some extent, since androgens, which are secreted by the menopausal ovaries and adrenal cortices, can be converted into estrogens. In fact, the majority of non-cycling endometria are thin and atrophic, but only half of the cases are inactive [10]. The remaining show a weak proliferative activity, indicative of an endometrium that responds to continuous low levels of estrogens unopposed by progesterone. Under the influence of prolonged levels of estrogens, the postmenopausal endometrium may turn into the so-called disordered proliferative endometrium [6], or give rise to atypical endometrial hyperplasia or endometrial intraepithelial carcinoma [11,12], from which an endometrial carcinoma can develop.

2. ENDOMETRIAL LESIONS

2.1. ENDOMETRIAL HYPERPLASIA

Endometrial hyperplasia involves the abnormal proliferation of endometrial tissue, especially of the endometrial glands, that results in a greater than normal gland-to-stroma ratio. Endometrial hyperplasia develops from continuous estrogen stimulation that is unopposed by progesterone [1].

There are two main **classifications** of endometrial hyperplasias. The most common system, used by the World Health Organization (WHO) and the International Society of Gynecologic Pathologists since 1994, is based upon two features: the glandular/stromal architectural pattern, described as either simple or complex; and the presence or absence of nuclear atypia, which is the presence or absence of nuclear enlargement (greater than normal nucleus/cytoplasm ratio) [13]. Therefore, four groups are described in this classification system: simple hyperplasia without atypia, complex hyperplasia without atypia, simple atypical hyperplasia and complex atypical hyperplasia.

Progression to endometrial cancer if untreated occurs in 1% of patients with simple hyperplasia without atypia, 3% of patients with complex hyperplasia without atypia, 8% of cases with simple hyperplasia with atypia and 29% of cases with complex hyperplasia with atypia [14]. Furthermore, several studies have also found coexisting carcinomas at rates ranging from 17 to 56% of cases [15–17]. The difficulty in diagnosing hyperplasia, especially atypical hyperplasia, versus adenocarcinoma has been described in many studies [18,19]. Therefore, a simpler classification of endometrial hyperplasia (benign hyperplasia) *versus* premalignant endometrial intraepithelial neoplasia (EIN) has been proposed to better predict progression to cancer [20,21]. EIN is a premalignant clonal proliferation of endometrial glands. Correct diagnosis requires strict histological criteria, including size, architecture and cytology, and the exclusion of cancer, as well as benign conditions that mimic EIN. A diagnosis of EIN is associated with a 27% likelihood of having endometrial cancer within one year, and a 45-fold increased risk of progression to endometrial carcinoma after one year [21]. Currently, the WHO classification system remains the most widely used.

The most common **clinical presentation** of patients with endometrial hyperplasia is abnormal uterine bleeding. These women undergo the same **diagnostic process** than patients suffering from endometrial cancer (see section 3.4.3, page 19). Regarding **treatment**, the presence or absence of nuclear atypia is the most clinically relevant criterion, together with the age and wish for future childbearing. Hyperplasia without atypia is treated with progestins. More than 98% of women with hyperplasia treated with cyclic progestins

experienced regression of the disease in 3-6 months [1]. In case of atypical hyperplasia, the preferred treatment is hysterectomy. However, women younger than 40 years willing to have children receive a prolonged treatment with progestins and are only subjected to hysterectomy in case of persistent symptomatology [1].

2.2. ENDOMETRIAL POLYPS

Endometrial polyps are formed by the overgrowth of endometrial tissue. The average size of these lesions is 0.5-3 cm in diameter, although they can range from a few millimeters to larger masses that occupy the entire endometrial cavity. They can be attached to the endometrium by a thin stalk or a broad base and there may be one or several polyps present. The cause for polyps formation is unknown, but their growth appears to be linked to high levels of estrogen. Endometrial polyp prevalence rises with age and/or menopause [22].

As well as in endometrial hyperplasia and endometrial carcinoma, the most common **clinical presentation** of patients with polyps is abnormal uterine bleeding. The **diagnostic process** is the same followed in case of suspicion of endometrial cancer or hyperplasia [1] (see section 3.4.3, page 19).

The **prevalence rate of malignancy** in endometrial polyps is around 1% to 5% [22–24]. Hypertension, obesity, unopposed estrogen therapy, and postmenopausal status in women with endometrial polyps are associated with an increased risk of endometrial malignancy [23,25]. The need for **treating** polyps remains controversial; polyps may only be removed by hysteroscopy if they cause symptoms, or if they are suspected to be precancerous or cancerous [1].

3. ENDOMETRIAL CANCER

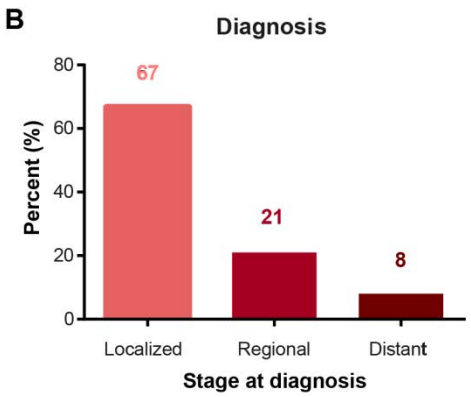
3.1. EPIDEMIOLOGY

The cancer-specific incidence and mortality rate differ between the considered "developed" (North America, Europe, Australia/New Zealand and Japan) and "developing" regions (remaining regions and countries) [26]. Endometrial cancer (EC) is the fourth most common cancer in women in developed countries, after breast, lung, and colorectal cancer and it is the most frequent of the invasive tumors of the female genital tract [1]. Recent data from the USA estimates that 61,380 new cases of EC will be diagnosed in 2017 (7% of all cancers in women) [27] (Figure 3A).

A

Incidence	Estimated new cases in 2017	
Breast	252,710	30%
Lung & bronchus	105,510	12%
Colon & rectum	64,010	8%
Uterine corpus	**61,380**	**7%**
Thyroid	42,470	5%
Melanoma of the skin	34,940	4%
Non-Hodgkin lymphoma	32,160	4%
Leukemia	25,840	3%
Pancreas	25,700	3%
Kidney & renal pelvis	23,380	3%
All Sites	852,630	100%

Mortality	Estimated Deaths in 2017	
Lung & bronchus	72,280	25%
Breast	40,610	14%
Colon & rectum	23,110	8%
Pancreas	20,790	7%
Ovary	14,080	5%
Uterine corpus	**10,920**	**4%**
Leukemia	10,200	4%
Liver &intrahepatic bile duct	9,310	3%
Non-Hodgkin lymphoma	8,690	3%
Brain & other nervous system	7,080	3%
All Sites	282,500	100%



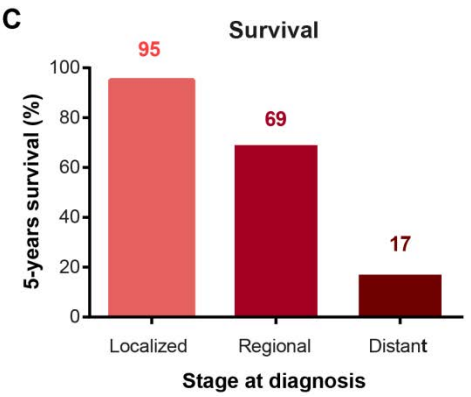


Figure 3. Epidemiology of EC. A. Estimated new cases and deaths for the ten leading cancer types in women in the United States for 2017. **B**. Stage distribution at diagnosis of endometrial cancer cases. Data from the United States for 2006-2012**. C**. Five-year survival rates by stage at diagnosis for endometrial cancer**.** Data from the United States for 2006-2012. *Adapted from Siegel et al., 2017.*

In Europe, EC represented 6.1% of the total number of cancer cases in women in 2012 [28]. In Spain, EC was the third most frequently diagnosed cancer in women (5,121 new cases in 2012), above lung cancer cases [29]. Unfortunately, its incidence is increasing as a consequence of the increased life expectancy of the population as well as a greater overall prevalence of obesity and metabolic syndromes [30–32]. EC mostly affects postmenopausal women, with a median age of 63 years [33]. However, 14% of cases are diagnosed in premenopausal women, 5% of whom are younger than 40 years [34–36].

Death rates from EC had been stable since 1992, after decreasing an average of 1.5% per year from 1975 to 1992 due to the replacement of estrogen stimulation treatments by treatments counterbalanced by progesterone. In 2012, EC caused about 23,700 deaths in Europe [28] and 1,211 deaths in Spain [29], representing the ninth leading cause of cancer

mortality in women. Remarkably, death rates for EC have been increasing in the last years. The number of expected deaths in the USA increased from 8,010 in 2012, when EC ranked eighth in cancer mortality [37], to 10,920 in 2017, EC representing now the sixth cancer type in number of deaths [27] (Figure 3A).

In spite of this, the mortality rate is not as high as it could be expected being EC the fourth most incident cancer in women (only about 18% of cases cause the death of the patient). This is mainly due to the good prognosis of the majority of cases. Nowadays, about 70% of EC patients are diagnosed at early stages of the disease, i.e., when the tumor is still localized within the endometrium, associated with an overall 5-year survival rate of 95%. However, still 30% of EC patients are diagnosed at an advanced stage of the disease associated with a 5-year survival rate of 69% when myometrial invasion is present and/or lymph nodes are affected, and of 17% when the tumor has spread beyond the pelvis [27] (Figure 3B-C).

3.2. RISK AND PROTECTIVE FACTORS

Several epidemiological studies have described different risk and protective factors for EC, which are explained in this section and summarized in Table 1.

3.2.1. Risk factors

3.2.1.1. Hormonal factors

The "unopposed oestrogen hypothesis" maintains that the exposure to estrogen that is not opposed simultaneously by a progestagen is one of the main risk factors for type I or endometrioid ECs (see full description of EC subtypes in section 3.3.1, page 14) [38]. This exposure increases the mitotic activity of the endometrial cells, causing DNA replication errors, mutations, endometrial hyperplasia and EC [1]. Different conditions associated to unopposed exposure to endogenous and exogenous estrogens are described below:

- **Endogenous estrogens**

Overweight and obesity: Excess body weight, both overweight and obesity, expressed as increased body-mass index (BMI), is associated with the risk of several adult cancers, including EC [39,40]. A study of more than one million women followed for an average of 25 years reported that overweight and obese women had an overall relative risk of EC of 1.36 and 2.51 compared with women with normal BMI, respectively [41]. Many studies have reported an approximately linear increase of EC risk with increasing body mass index in pre- and postmenopausal women, with up to 5-fold increase in EC risk for obese women [42–44]. Additionally, increased BMI is associated with increased mortality in patients with EC, with

the highest risk for those with a BMI ≥40 [45]. In premenopausal women, obesity causes a higher number of anovulatory cycles, amenorrhea and irregular menstrual cycles, all associated with a deficiency of progesterone production that can favor the development of EC. In postmenopausal women, obesity produces an increase of the peripheral production of estrogens by the conversion of adrenal precursor to estrogen and estradiol via aromatization in adipose tissue [43,46].

Table 1. Risk and protective factors associated with the development of EC.

Risk factors	Relative Risk
Estrogen replacement therapy	2-20
Tamoxifen	2-8
Overweight	1.4-1.6
Obesity	2-5
Early menarche	1.5-2.4
Late menopause	2-3
Nulliparity	2-3
Polycystic Ovary Syndrome	2-3
Advanced age	2-3
Lynch Syndrome	10
Protective factors	**Relative Risk**
Combined oral contraceptives	0.76
Grand multiparity	0.57
Cigarette smoking	0.71
Physical activity	0.79

Menstrual and reproductive factors: an elevated risk of EC has been associated with earlier menarche age (relative risk of 2.4 for ages < 12 *versus* ≥15), later age at menopause, and longer days of menstrual flow (relative risk of 1.9 for ≥ 7 *versus* < 4 days) [47,48]. Among the reproductive factors, the risk of EC is positively correlated with infertility and nulliparity, being inversely related to the number of full-term pregnancies [47,49]. Factors such as anovulation may explain the increased risk of EC found among infertile women [50]. For the most common disease associated with anovulation, the polycystic ovary syndrome, an increased risk of EC has been reported, especially in premenopausal women [51,52]. All these conditions increase the lifetime exposure to unopposed estrogen.

- **Exogenous estrogen**

Estrogen replacement therapy prescribed to control menopausal symptoms increases the risk of developing EC by 2 to 20-fold, with an increasing risk in cases of prolonged use (10 years or more). Furthermore, the relative risk remains elevated 5 or more years after discontinuation of the unopposed estrogen therapy [53]. A sequential estrogen-progestin replacement therapy with administration of progesterone at least 10 days a month, or a continuous combined therapy, significantly reduces the increased risk of EC [54,55]. In addition, the lowest dose of estrogens for each patient should be chosen.

Tamoxifen is a drug indicated for breast cancer prevention and treatment [56]. It is an antagonist of the estrogen receptor in breast tissue. However, in other tissues such as the endometrium, it behaves as an agonist and induces growth of endometrial cells. The association between the use of tamoxifen and the risk of developing EC has been widely studied. Its use has been associated with a 2 to 8-fold increase in the incidence of the disease [57–59]. A significant trend of increasing risk of EC has been reported with the duration of tamoxifen use, and with cumulative dose [60]. The risk in premenopausal women is unclear [59].

3.2.1.2. Age

Another main risk factor for EC is age, since it is a cancer of the postmenopausal period, with an average age at diagnosis of approximately 63 years. As described in the Epidemiology Section (section 3.1, page 9), only about 5% of cases are diagnosed in women younger than 40 years [34–36]. Younger women with EC generally have well-differentiated type I tumors and lower-stage disease than older women.

3.2.1.3. Hereditary factors

A hereditary component has also been described with regards to the risk of suffering EC. This component represents around 5% of all reported cases. Lynch syndrome, also known as hereditary nonpolyposis colorectal cancer syndrome, accounts for the majority of inherited cases [61,62]. It is an autosomal dominant family cancer syndrome characterized by mutations in DNA mismatch repair genes. The lifetime cumulative risk of EC for women with Lynch syndrome is 40% to 60%. These patients often develop EC at a younger age. In addition, a recent meta-analysis reported a higher risk of developing EC for women with a first-degree family history of endometrial or colorectal cancer [63].

3.2.2. Protective factors

On the contrary, the use of combined oral contraceptives [64], grand multiparity [65], cigarette smoking [66] and physical activity [67] have been described as **protective factors** that reduce the risk of EC. All these factors are associated with decreased estrogen concentrations and/or increased levels of progesterone.

3.3. EC CLASSIFICATION

3.3.1. Dualistic model

In the past decades, EC has been broadly classified into the two subtypes on the basis of clinical, pathological and molecular features (Table 2). This classification was proposed by Bokhman *et al*. in 1983 [68].

❖ **Type I or endometrioid adenocarcinoma**

Endometrioid adenocarcinoma (EEC) is the most common subtype, accounting for 75-80% of cases. It is a estrogen-dependent, diploid adenocarcinoma with endometrioid histology [69]. EECs, at least in well-differentiated forms, are composed of glands that resemble those of a normal endometrium. EECs are usually developed in perimenopausal women and can be associated with or preceded by endometrial hyperplasia [70]. These tumors are normally diagnosed at an early stage and low grade, and have a good prognosis. The overall 5-year survival rate of this subtype is around 85% [69].

Regarding molecular alterations, type I ECs are characterized by a high mutational frequency and microsatellite instability [71]. The most frequently altered pathway is the PI3K pathway. Mutations are noted in more than 90% of lesions, sometimes with multiple aberrations in this pathway, such as simultaneous loss of PTEN and PIK3CA mutations [72]. The frequency of mutations in the beta-catenin gene is around 20-25% and is almost restricted to EEC cases [73,74]. In addition, KRAS and FGFR2 mutations have been reported in about 20% and 12% of tumors, respectively [74,75].

❖ **Type II or non-endometrioid adenocarcinomas**

Non-endometrioid adenocarcinomas (NEECs) represent 10 to 15% of cases (Table 2). These tumors are high-grade, aneuploid adenocarcinomas not related to estrogen stimulation. In fact, they express neither estrogen nor progesterone receptors. They include a range of histological subtypes, being serous and clear cell EC the most common [69]. They usually develop in older women than those with type I EC and characteristically arise from

precancerous lesions (endometrial intraepithelial carcinoma) in atrophic endometrium. These tumors are associated with a higher risk of metastasis and poor prognosis. The overall 5-year survival rate of this subtype is around 55% [69]. They account for more than 50% of recurrences and deaths from EC [76].

Table 2. EC dualistic classification. *Adapted from Morice et al. 2016* [69].

Bokhman: dualistic classification			
	Type I or endometrioid endometrial carcinoma (EEC)	**Type II o non-endometrioid endometrial carcinoma (NEEC)**	
Incidence	75-80% of all cases, perimenopausal women	10-15% of all cases, older women	
Grade	Low	High	
Hormone recpetor expression	Positive	Negative	
Genomic stability	Diploid, frequent microsatellite instability	Aneploid	
Histology	**Endometrioid**	**Serous**	**Clear cell**
Molecular alterations			
TP53 mutations	Rare	>90%	35%
PI3K alterations	PTEN mutation (75-85%) PIK3CA mutation (50-60%)	PTEN mutation (11%) PIK3CA amplification (45%) PIK3CA mutation (35%)	PTEN loss (80%) PIK3CA mutation (18%)
KRAS mutation	20-30%	3%	0%
ERBB alterations	None	ERBB amplification (21-47%	ERBB mutation (12%) ERBB amplification (16%)
FGFR amplification or mutations	FGFR mutation (12%)	FGFR mutation (5%) Frequent FGFR1 and FGFR amplification	-
Wnt/b-catenin	CTNNB1 mutation (25%)	CTNNB1 mutation (3%)	-
Other	ARID1A mutation (35-40%)	PPP2R1A mutation (20%) FBXW7 mutation (20% undifferentiated EC)	ARID1A (20-40%)
Prognosis			
	Good (overall survival 85% at 5 years)	Poor (overall survival 55% at 5 years)	

NEECs are characterized by chromosomal instability and TP53 mutation [77]. However, each histological subtype shows specific molecular features. Regarding serous carcinomas, mutations in the TP53 gene occur in up to 90% of cases [78]. In addition, the reported rates of HER2/neu (ErbB2) overexpression range between 14% and 80%, with HER2 amplification ranging from 21% to 47% [79]. The PI3K pathway is also frequently dysregulated; however, these tumors are mainly characterized by amplifications in PIK3CA (45% of cases) or mutations in this gene (35%). On the other hand, the most common molecular features in

the tumors with clear cell histology are the loss of PTEN, reported in about 80% of cases, and inactivating mutations in the ARID1A gene in 20-40% of tumors [80,81].

Although this dualistic classification continues to be the most widely used in the clinical practice, its prognostic value remains limited. Around 20% of EEC cases relapse, whereas 50% of NEEC do not [68]. Moreover, 15-20% of EEC are high-grade tumors and where they fit in this model is unclear [82,83]. EC comprises a high variety of biological, pathological and molecular features impossible to simplify in a dualistic model [84].

3.3.2. Molecular classification (TCGA)

The Cancer Genome Atlas (TCGA) performed an integrated genomic, transcriptomic and proteomic characterization of endometrioid and serous adenocarcinomas to search for a more accurate classification, particularly of more aggressive tumors. A new classification of EC into four prognostically significant subgroups was suggested: POLE ultramutated, microsatellite instability hypermutated, copy-number-low (microsatellite stable), and copy-number-high (serous-like) EC (Table 3) [85].

- **POLE ultramutated**

This is a newly identified small subgroup that defines a unique subset of EC cases that represents about 10% of EECs. They are characterized by hotspot mutations in the exonuclease domain of the DNA polymerase epsilon catalytic subunit (POLE), ultrahigh somatic mutations rates, few somatic copy number alterations (SCNAs) and excellent prognosis. Among the POLE ultramutated carcinomas, 60% are high grade EEC, and 35% present TP53 mutations [86].

- **Microsatellite instability hypermutated**

This subgroup has a mutation frequency about ten-fold greater than the microsatellite stable subgroup. They are characterized by the loss of DNA mismatch repair proteins, such as MLH1, MSH2, MSH6, and PMS2, that occurs in 30-40% of EEC cases. In sporadic cases, this is due to MLH1 promoter hypermethylation, while in hereditary Lynch syndrome it can be caused by mutations in any of the DNA mismatch repair genes [87]. This subset presents an intermediate outcome.

Table 3. TCGA classification of EC. MSS, microsatellite stable; MSI, microsatellite instability. *Adapted from Morice et al. 2016* [69].

TCGA classification for endometrioid and serous histologies				
	POLE ultramutated	**MSI hypermutated**	**Copy-number low, MSS**	**Copy-number high, serous like**
Mutation load				
Somatic copy number alterations (SCNAs) load				
Histology		**Endometrioid**		**Serous and endometrioid**
Grade	G1, G2, G3	G1, G2, G3	G1, G2	G3
PI3K alterations				
KRAS mutation				
TP53 mutation	35%	5%	1%	>90%
Specific molecular alterations	Hotspot mutations in POLE	Loss of DNA mismatch repair proteins (MLH1, MSH2…)	CTNNB1 (52%)	TP53 (92%)
Prognosis				
Prognosis	Excellent	Intermediate	Intermediate	Poor
More accurate prognostic value				

❖ Copy-number-low (microsatellite stable)

This subgroup is characterized by low mutation rates, but an unusually high frequency of CTNNB1 mutations (52%), few SCNAs, and an intermediate prognosis.

❖ Copy-number-high (serous-like)

This subgroup includes most serous ECs and 25% of the tumors classified as high grade EEC which have a molecular phenotype similar to serous carcinomas. They are characterized by genomic instability, extensive SCNAs and a poor prognosis. Mutations in the TP53 gene occur in more than 90% of the cases [85].

This molecular characterization demonstrates that tumors classified as high grade EEC by pathologists are heterogeneous: 25% of them have a molecular phenotype similar to serous

carcinomas, with frequent TP53 mutations, high rate of SCNAs and poor prognosis; while other 25% are ultramutated POLE cancers with good prognosis. Therefore, it has been suggested that clinicians should consider the incorporation of these molecular features to improve management of these patients [69]. Several recent studies have developed surrogate assays that could replicate the TCGA classification in a simple, lower-cost way to enable its application in the clinical routine. These assays are based on a set of mismatch repair proteins and p53 immunohistochemistry and POLE mutational analysis performed on formalin-fixed paraffin-embedded samples [88,89].

3.4. EC DIAGNOSIS

3.4.1. Screening test

To date, there is not enough evidence to support routine screening for EC in the general population [33]. The only way of favoring the early detection of this cancer is to seek attention and report to the doctors the presence of any signs and symptoms which are indicative of EC, mainly any vaginal bleeding in postmenopausal women [90]. The Papanicolau smear is a screening test for cervical cancer and it is the only recommended periodic screening for all women. However, only 30-50% of EC patients will present malignant cells in their annual Papanicolau smear [91,92].

Patients with Lynch Syndrome have up to a 60% risk of developing EC. Therefore, in this specific subgroup of patients, an annual histological examination of an endometrial biopsy and a transvaginal ultrasonography may be recommended for women older than 35 years [62,93]. Moreover, hysterectomy may be considered as a prophylactic measure once childbearing is complete [33,94].

3.4.2. Signs and symptomatology

The most common symptom of EC is abnormal uterine bleeding, which is present in 90% of the patients [69]. Although other benign disorders generate this symptom (Table 4) [95], uterine bleeding in postmenopausal women is a reason for EC suspicion and hence, these women must enter the diagnostic process. The probability of EC in women presenting with postmenopausal bleeding is 8-15% [13,95,96], but the chances increase with age and other risk factors [76]. In premenopausal patients, especially after age 35, menorrhagia or intermenstrual bleeding might indicate the presence of an EC [97].

Other frequent symptoms are lower abdominal pain or pelvic cramping, and a thick, whitish or yellowish vaginal discharge (leucorrhoea) mixed with blood, sometimes with purulent content. Changes in bowel or bladder functions, anemia, weight loss and shortness of

breath are other signs and symptoms of EC that would suggest a more advanced disease but are totally unspecific [1,2].

Table 4. Histopathological findings in postmenopausal women with uterine bleeding.

Histology	Approximate percentage (%)
Atrophy	50-60
Proliferative	4
Secretory	1
Polyps	9-12
Hyperplasia	5-15
Endometrial cancer	**8-15**
Others	3

3.4.3. Diagnostic procedure

Any woman with suspicion of EC due to abnormal uterine bleeding and/or any other symptom related to EC, particularly if they have risk factors for this disease, should undergo a thorough diagnostic evaluation (Figure 4).

The first steps in the current diagnostic process are a gynecological examination and transvaginal ultrasound, followed by the histopathological examination of an endometrial biopsy.

3.4.3.1. Gynecological examination

The gynecological examination is performed to localize the source of bleeding and determine its physical extent [98]. The results are frequently normal, especially in the early stages of the disease. Changes in the size, shape, or consistency of the uterus and/or its surroundings may exist at advanced stages.

3.4.3.2. Transvaginal ultrasonography

Transvaginal ultrasonography is the imaging technique of choice for the assessment of the endometrium in symptomatic patients, although its routine use as a screening method in asymptomatic postmenopausal women is not justified [99]. The endometrial thickness, the presence of an endometrial mass, and/or an endometrial stripe abnormality can be evaluated by this technique.

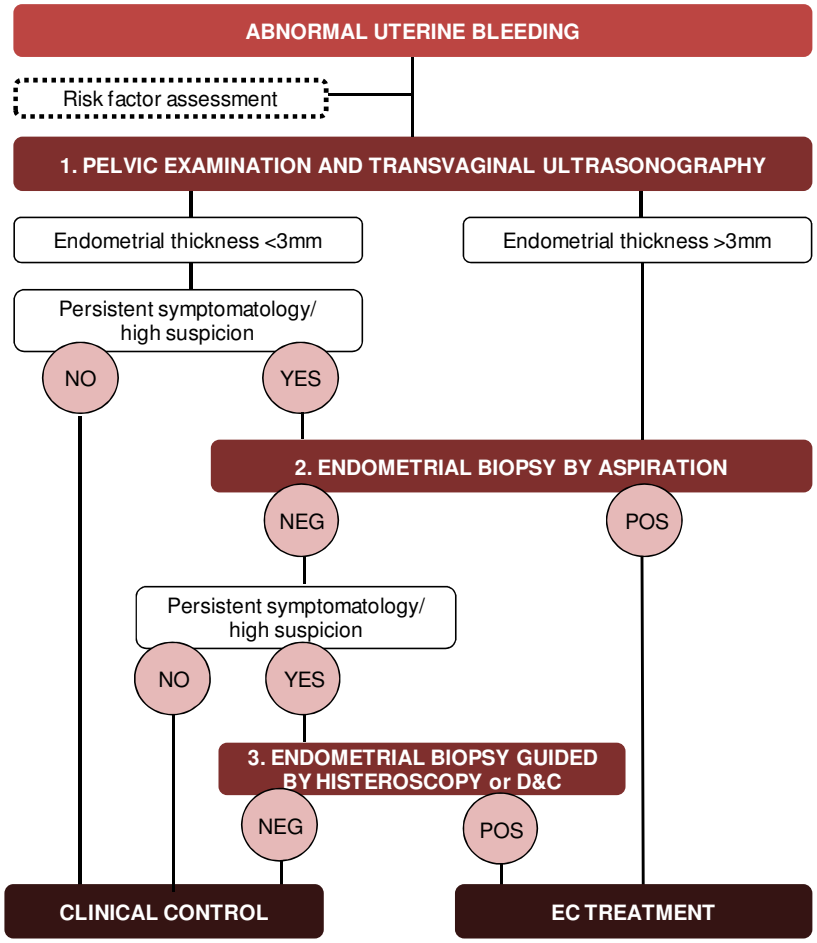


Figure 4. EC diagnostic evaluation. D&C, dilatation and curettage. *Adapted from Oncoguia SEGO 2016* [97].

The vast majority of studies have focused on the evaluation of the accuracy of transvaginal ultrasonography in detecting EC in postmenopausal women with vaginal bleeding, trying to determine the best cut-off of endometrial thickness [100–103]. This technique is very sensitive but has very limited specificity as other benign causes, such as polyps or hyperplasia, can cause the thickening of the endometrium. A review from 13 published studies including approximately 2900 patients demonstrated that an endometrial thickness cut-off of 5 mm on ultrasonography resulted in a **sensitivity** of 90% and a **specificity** of 54%, compared to a higher sensitivity of 98% but lower specificity of 35% if the cut-off was reduced to 3 mm [101]. It is up to the clinicians to decide the threshold to achieve the most optimal cost-effectiveness. Nowadays, according to the Spanish Society of Gynecology and Obstetrics

(SEGO) guidelines, an endometrial thickness exceeding 3 mm in postmenopausal women with vaginal bleeding is considered suspect and must undergo an endometrial biopsy [97].

The use of transvaginal ultrasonography, while useful in type I EC, is limited in type II EC. A thin or indistinct thickness of the endometrium has been reported in 27.5% of NEEC cases [104]. In addition, no reliable cut-off has been established in postmenopausal women receiving tamoxifen or hormone replacement therapy, as well as in pre- or perimenopausal women [98]. In premenopausal women, the thickness of the endometrium varies during the menstrual cycle, but a thickness exceeding 8 mm usually suggests the need for an endometrial biopsy [105]. Endometrial stripe abnormality is considered to be more useful than endometrial thickness in the recommendation of endometrial biopsy even in asymptomatic pre- and perimenopausal women [106].

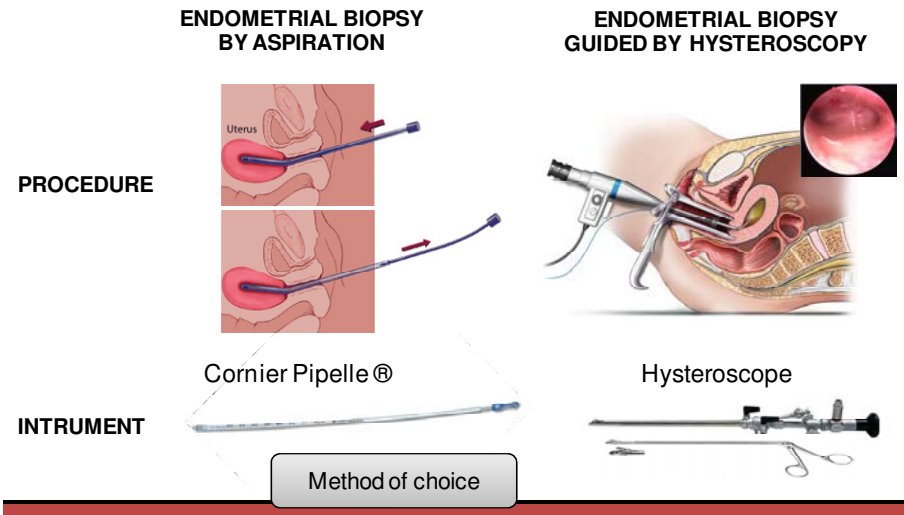
Therefore, in cases of EC suspicion due to a positive result in the ultrasonography or when it is negative but the symptomatology persists and there are risk factors, patients must undergo an endometrial biopsy to achieve a final diagnosis (Figure 4) [76,97].

3.4.3.3. Pathological examination of an endometrial biopsy

The pathological examination of an endometrial biopsy is the gold standard for EC diagnosis. This procedure allows the clinician to take a small sample of the endometrium to be microscopically examined by the pathologist for abnormal cells. In the past, these endometrial biopsies were mainly collected by a procedure called "dilatation and curettage" (D&C). Nowadays, it has been replaced in most places by endometrial biopsies obtained by aspiration or guided by hysteroscopy (Figure 5) [69,70].

- **Endometrial biopsy by aspiration**

Biopsies can be performed blindly, by aspiration, with a soft, straw-like device that suctions a small sample from inside the uterine cavity. This biopsy, named uterine aspirate or pipelle biopsy, is a liquid sample containing cells coming from the endometrium. The current diagnostic procedure relies on the cytological determination of malignant cells in this biopsy. This procedure is considered as the first method of choice to get an endometrial biopsy, as it is a fast, cost-effective and safe procedure that is well tolerated by patients [69,97,107]. It does not require any previous tests for coagulation factors, and it can be performed as an office procedure without anesthesia (Figure 5).



	Advantages	Disadvantages
Comfort	Minimally invasive	More invasive
Specific requirements	No requirements	Coagulation tests Anesthesia Hospital
Economy	Less expensive	More expensive
	Drawbacks	**Advantages**
Failure rate (% cases)	Impossibility to obtain the biopsy (8-12%) Histologically inadequate (13-22%)	Low failure rate (4%)
Conclusive results	**< 80% diagnosed patients**	**96% diagnosed patients**

Figure 5. Procedures and features of the different methodologies used for the collection of an endometrial biopsy.

A meta-analysis of 39 studies that included 7914 women who underwent endometrial sampling by aspiration using various devices revealed that the disposable Pipelle®, also known as Cornier Pipelle®, was the best technique for both pre- and postmenopausal women. Pipelle biopsy achieved **sensitivity** values ranging from 68% to 98%, and **specificity** ranging from 96% to 100% for the detection of EC when using final hysterectomy as the reference strategy [108]. Similar results were reported in a systematic quantitative review of 11 primary studies [109], where the authors concluded that, when adequate specimens are obtained, these pipelle biopsies have a high overall accuracy in diagnosing EC, performing better in symptomatic (bleeding) and postmenopausal women. However, although a positive test result is highly accurate, a negative test result is of limited accuracy

due to the moderate sensitivity reported in some studies. Therefore, if the pipelle biopsy of a patient is normal but the woman is at high risk for EC and/or her symptoms persist, further evaluation is warranted for excluding disease (Figure 4) [109].

In addition, **failure to get samples** that are adequate for histological examination is one of the main problems associated with pipelle sampling, particularly in postmenopausal women. This failure can be due to the impossibility of obtaining the sample (e.g., cervical stenosis.) or due to histologically inadequate samples (insufficient cells in the sample for adequate diagnostic assessment). In a systematic review, Clark *et al.* reported an average failure rate of 8% in obtaining samples using a Pipelle device, while 13% of the samples obtained were histologically inadequate. This was increased in postmenopausal women up to an average of 12% and 22%, respectively [109]. Several studies have reported higher ratios of up to 47-64% of inadequate pipelle specimens in postmenopausal women [110,111].

Women that do not receive a definitive diagnosis with the endometrial sampling by aspiration due to an inadequate sampling or an inconclusive result must undergo a biopsy guided by hysteroscopy.

❖ Endometrial biopsy guided by hysteroscopy

The basic hysteroscope is a thin, lighted tube with a camera usually attached to its proximal end. It is inserted into the uterus through the vagina and transmits the image of the uterus onto a screen. Carbon dioxide gas or a fluid, such as sodium chloride solution, can be pumped through the hysteroscope to distend the endometrial cavity, enabling visualization and operation. Surgical instruments such as biopsy forceps can be passed through the hysteroscope in order to perform directed endometrial biopsies that are histologically examined (Figure 5).

A systematic quantitative review of a total of 56 primary studies published between 1984 and 2001, and a more recent meta-analysis including studies published until 2011 evaluated the performance of hysteroscopy for EC diagnosis and reported an overall **specificity** higher than 99% and an overall **sensitivity** of 82.6-86.4% [112,113]. These studies underline a higher variation in sensitivity, with reported values down to 63-65% in some studies [114,115].

Importantly, the **failure rate** of this methodology is only about 4% [112,116]. Failed hysteroscopies can result from anatomic factors (e.g., cervical stenosis), patient factors (e.g., pain, intolerance) or inadequate visualization (e.g., obscured by bleeding).

This procedure is more invasive than endometrial biopsies performed by aspiration. It requires previous blood testing and might require anesthesia and a hospital setting. Some authors have also discussed that the hysteroscopy examination before surgery may increase the risk of dissemination of cancer cells into the peritoneal cavity [117]. However,

there is currently no evidence to support that these patients face worse prognosis than patients diagnosed by other procedures [118].

In conclusion, pipelle biopsies should be the first option and hysteroscopy should be performed only when diagnosis based on the pipelle sampling is not feasible or is uncertain [69,107].

3.5. PROGNOSTIC FACTORS

A prognostic factor is a situation, condition, or a characteristic of a patient, that can be used to estimate the likely outcome of an illness (i.e., chance of recovery from a disease or the chance of the disease recurring).

The prognosis of EC depends on several factors, mainly on the FIGO stage, differentiation grade, histological type, lymphovascular space invasion, tumor size and age.

3.5.1. FIGO stage

The 2009 International Federation of Gynecology and Obstetrics (FIGO) staging is the most widely used classification. It takes into account the extent of the tumor, and whether the cancer has spread to lymph nodes or distant sites. This staging system classifies EC cases into four stages [119], which are described and pictured in Figure 6. Overall 5-year survival rates significantly decrease in the higher stages [120].

3.5.2. Differentiation grade

While all NEECs are considered high grade by definition, EEC cases present different histological grades (Figure 7). The most widely used system for grading is the FIGO system, which is primarily based on the architectural grade of the tumor, and secondarily modified based on the nuclear grade.

- The **architectural grade** is based on three grades according to the proportion of solid areas of tumor cells (Table 5) [97,121].

- The **nuclear grade** is based on the presence of significant nuclear atypia. This is characterized by large, pleomorphic nuclei, irregular chromatin clumping, large irregular nucleoli, and loss of cellular polarity. The tumor is upgraded from grade 1 to 2, or from grade 2 to 3 in cases showing striking nuclear atypia [76,122]. Nuclear grading takes precedence over architectural grading in mixed types, which are considered grade 3 in cases where ≥ 10% is serous, clear cell or undifferentiated histology.

Stage	Description	5-year survival rate (%)
Stage I*	**Tumor confined to the corpus uteri**	
IA*	0-50% myometrial invasion	89.6
IB*	≥ 50% myometrial invasion	77.6
Stage II*	**Tumor invades cervical stroma, but does not extend beyond the uterus****	73.5
Stage III*	**Local and/or regional spread of the tumor**	
IIIA*	Tumor invades the serosa of the corpus uteri and/or adnexae#	56.3
IIIB*	Vaginal and/or parametrial involvement#	36.2
IIIC*	Metastases to pelvic and/or para-aortic lymph nodes#	
IIIC1*	Positive pelvic nodes	57
IIIC2*	Positive para-aortic nodes with or without positive pelvic lymph nodes	49.4
Stage IV*	**Tumor invades bladder and/or bowel mucosa, and/or distant metastasis**	
IVA*	Tumor invasion of bladder and/or bowel mucosa	22
IVB*	Distant metastasis, including intra-abbominal metastasis and/or inguinal lymph nodes	21.1

* Either grade 1, grade 2, or grade 3

** Endocervical glandular involvement only should be considered as Stage I and no longer as Stage II

Positive cytology has to be reported separately without changing the stage

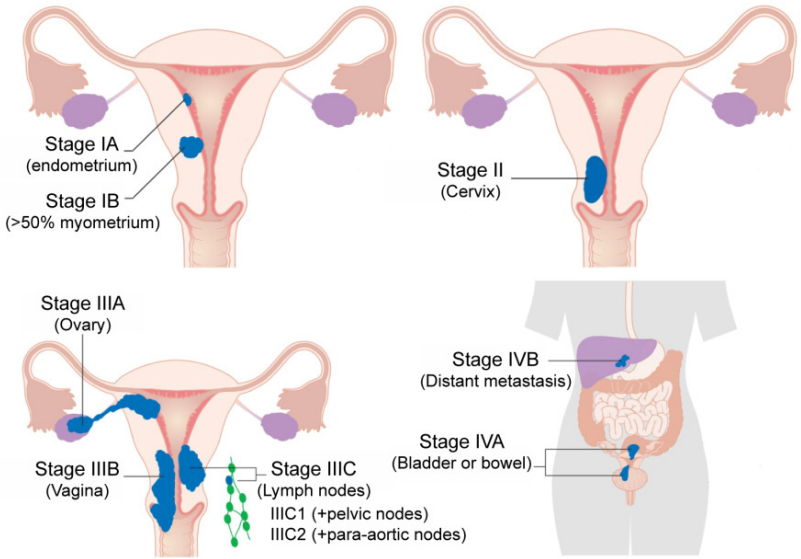


Figure 6. FIGO staging of EC. *Adapted from Pecorelli et al. 2009* [119] *and Lewin et al. 2010* [120]**.** *Image adapted from http://teachmeobgyn.com/.*

However, FIGO grading of EEC tumors requires evaluation of histological features that can be difficult to assess and reproduce. Reported interobserver agreement varies between 63% and 81% [123], being the reproducibility associated with the identification of grade 2 tumors particularly low. Moreover, the outcome of patients with grade 1 and 2 has been reported to be very similar (overall 5-year survival rates of 97% and 94% respectively), while grade 3 tumors are associated with a lower 5-year survival rate of 76% [123]. Consequently, several

studies have suggested several modifications and the conversion of this 3-tiered grading system into a binary system [124,125]. Grade 3 versus combined grades 1 and 2 still proves to be of independent prognostic significance [123,126] with higher reproducibility [126,127].

Table 5. FIGO histological grades. Adapted from Oncoguía SEGO 2016 [97].

Histologic grade	Description	% solid tumor
Grade 1	Well differentiated	<5%
Grade 2	Moderately differentiated	6-50%
Grade 3	Poorly differentiated	>50%

3.5.3. Histological type

ECs should be typed according to the 2014 WHO classification [128], since the diverse histological types present a very different overall prognosis.

The type I or **EEC adenocarcinoma** is characterized by the endometrioid histology, a gland-forming growth pattern of varying differentiation grade (Figure 7) [129]. Several subtypes or variants have been described by the WHO: adenocarcinoma with squamous differentiation (the most common variant), villoglandular carcinoma, and secretory carcinoma. **Mucinous adenocarcinomas** can be also included as type I.

The most frequent NEEC carcinomas are **serous** and **clear cell carcinomas** [130] (Figure 7). These histological types are highly invasive and aggressive, associated with a poor prognosis. Other subtypes are the **neuroendocrine tumors,** the **undifferentiated carcinomas** and the **dedifferentiated carcinomas**. **Mixed adenocarcinomas** are also common. These are a mixture of two or more different histological types. At least one of the subtypes must be a type II tumor and the minor type must account for at least 5% of the total tumor volume. Mixed adenocarcinomas with a type II tumor fraction of 25% or more have a poor prognosis. In addition, **carcinosarcomas** (also called malignant mixed Müllerian tumors) are mixed epithelial and mesenchymal tumors and they are treated in the same way as aggressive type II carcinomas [128].

3.5.4. Others

Other independent prognostic factors, related to the tumor spread and reflected in the FIGO staging, are **myometrial invasion, cervical invasion,** and **lymph node metastasis.** Deep myometrial invasion is associated with increased lymphovascular invasion, lymph node affectation and a decrease in the 5-year survival [131]. In addition, patients without lymph node

metastasis have a 5 year disease-free survival of 90%, that goes down to 60-70% in patients with pelvic lymph node metastasis and to 30-40% in those with para-aortic lymph node metastasis [69,132]. Another important prognostic factor is the **lymphovascular space invasion**. The presence of tumor cells within vascular spaces is considered an early step in the metastatic process even for patients with tumors that seem to be confined to the uterus, and a strong predictor of nodal metastasis, recurrence and cancer-specific death [133,134].

In addition, the age of the patient at the time of diagnosis is one the most widely reported prognostic factors [135]. Advanced age negatively affects survival in EC. Younger women tend to have better-differentiated lesions, while NEEC with worse prognosis typically affects older women.

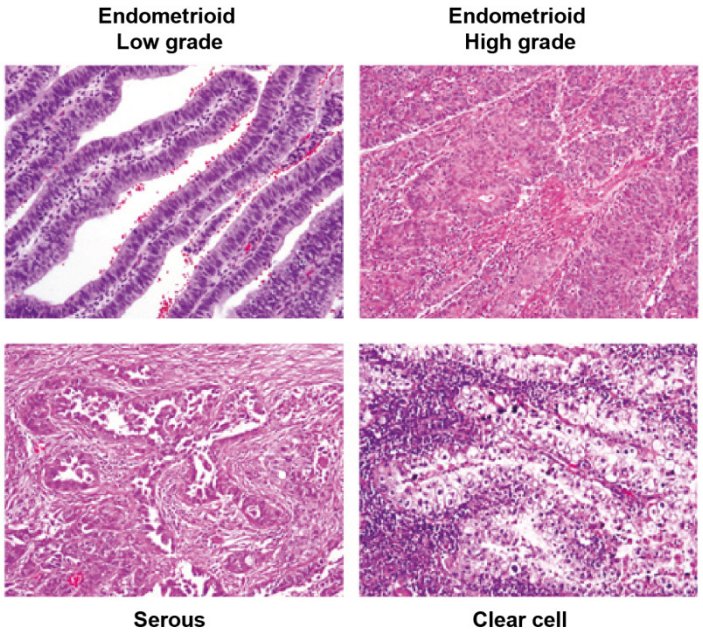


Figure 7. Histology of three common types of epithelial EC. *Adapted from Murali et al. 2014* [84].

3.5.5. Risk stratification systems

Risk stratification systems combine prognostic factors to define groups of patients with similar outcomes (low, intermediate, or high-risk of recurrence). These stratification systems are used worldwide to determine the most appropriate treatment for each group. Although the core variables of the different risk stratification systems are very similar, the combination of these variables changes. Results of a study comparing the five major risk stratification

systems suggested that the European Society for Medical Oncology (ESMO) modified system was the most accurate in the prediction of lymph node status and survival [136]. In 2016, a consensus between ESMO, the European Society for Radiotherapy & Oncology (ESTRO) and the European Society of Gynaecological Oncology (ESGO) introduced a few changes in this classification [33], which is detailed in Table 6. However, none of these systems show high accuracy in stratifying the risk of recurrence or nodal metastasis, particularly in patients with early-stage EC. It has been suggested that the incorporation of molecular and genetic characteristics would be useful to improve the accuracy of these risk stratification systems [84]. According to the TCGA classification (described in section 3.3.2, page 16), EC cases showing POLE mutations have an excellent prognosis, and patients could be spared unnecessary adjuvant treatment. Other novel candidate prognostic markers, such as stathmin or L1 cell adhesion molecule (L1CAM) have been identified. Stathmin has been described as a predictive biomarker for resistance to paclitaxel [137]. L1CAM has been reported to be a strong predictor of poor outcome in early-stage and advanced-stage EECs, but not NEECs, in several large multicenter studies [138,139]. L1CAM expression in EECs cancers indicates the need for adjuvant treatment.

3.6. PREOPERATIVE RISK ASSESSMENT AND PRIMARY TREATMENT

3.6.1. Preoperative risk assessment

The role of preoperative risk assessment is to classify patients into those groups of risk for lymphatic dissemination and disease recurrence in an early step of the diagnostic process in order to define the most optimal surgical management.

Currently, the extent of the surgical staging procedure is guided by the assessment of tumor grade and histological subtype based on the pathological examination of the preoperative endometrial biopsies, and the evaluation of myometrial and cervical invasion and lymph node metastasis based on imaging techniques (Figure 8).

3.6.1.1. Preoperative endometrial biopsies

As most information is not available preoperatively, histological subtype and grade become key factors for risk group assignment [140]. Biopsies obtained by aspiration and/or guided by hysteroscopy are not only used to discriminate between EC and non-EC cases but also to assess the tumor grade and histological subtype of the EC cases. However, high discordance rates have been reported on this matter. Regarding EEC tumors, different studies have reported that about 22% to 40% of tumors classified as grade 1 on the

preoperative biopsy were upgraded on the final surgical pathology, as were 26% to 56% of grade 2 tumors. In addition, 8% to 60% of tumors preoperatively defined as grade 3 were found to be downgraded on the final surgical evaluation and/or classified as NEEC in several studies [140–144]. Regarding high-risk NEEC histologies, some studies have reported a high concordance between preoperative and postoperative pathological interpretations (93% of the cases) [140]. However, other studies found good correlation for carcinosarcomas (90%), but worse for serous and clear cell carcinomas (67%) [145]. These discordances may be partially explained by interobserver variability [146] and the small volume of tissue available for examination in a preoperative biopsy. According to this, hysteroscopy and D&C show a higher accuracy in predicting final post-hysterectomy tumor grade than pipelle biopsy [147,148].

Table 6. Risk stratification system. FIGO 2009 staging is used. G, grade; LVSI, lymphovascular space invasion. *Adapted from Colombo et al. 2016* [33]

ESMO-ESGO-ESTRO consensus 2016	
Low risk	Stage IA G1-2 with no LVSI (Type I)
Intermediate risk	Stage IB G1-2 with no LVSI (Type I)
High-intermediate risk	Stage IA G3 with regardless of LVSI (Type I)
	Stage I G1-2, LVSI unequivocally positive, regardless of depth of invasion (Type I)
High risk	Stage IB G3 regardless of LVSI (Type I)
	Stage II
	Stage III (Type I), no residual disease
	All Type II (serous, clear cell, undifferentiated carcinoma, or carcinosarcoma)
Advanced	Stage III residual disease
	Stage IVA
Metastatic	Stage IVB

Molecular classifications such as the TCGA classification have demonstrated higher prognostic accuracy than the histomorphologic classification. Importantly, several studies have proved a high concordance between molecular alterations in preoperative endometrial

biopsies and the hysterectomy specimens [149,150] but have not yet been implemented in the routine clinical risk assessment procedure.

3.6.1.2. Magnetic resonance imaging (MRI)

Contrast-enhanced MRI has long been the preferred imaging technique for preoperative risk assessment, particularly for myometrial invasion assessment [69,151,152]. The major limitation of imaging techniques is the poor detection of lymph node metastases [153]. A recent meta-analysis including 52 studies examining MRI in the assessment of high-risk features of EC found pooled sensitivity (specificity) of 80.7% (88.5%) for ≥50% myometrial invasion, 57% (94.8%) for cervical invasion, and 43.5% (95.9%) for lymph node metastasis [154]. Considering the limited sensitivity of MRI, the authors concluded that patients with negative findings on MRI may not safely abstain from surgical staging. Furthermore, this technology is costly and requires an experienced radiologist to provide accurate interpretation. Some studies suggest that transvaginal ultrasonography has similar accuracy to that of MRI for assessment of myometrial and cervical invasion. This technology is less costly than MRI but cannot be used to determine lymph node metastases [155]. If MRI is not available, computed tomography can be also used to assess extrauterine disease.

Various studies have underlined the high accuracy of 18F-fluorodeoxyglucose Positron Emission Tomography/Computed Tomography in detection of myometrial and cervical invasion and lymph node metastasis. However, its use in preoperative risk assessment of EC remains questionable [156,157].

3.6.2. Primary treatment

The most efficient treatment for EC is surgery. Total hysterectomy (removal of the uterus) and bilateral adnexectomy (removal of both Fallopian tubes and ovaries) is the standard treatment for apparent stage I EC [33,97] (Figure 8). Alternatives to hysterectomy for women who wish future childbearing have been comprehensively reviewed [158]. Hysterectomy and adnexectomy were traditionally done with open abdominal surgery (laparotomy). However, laparoscopy is currently the preferred surgical approach, as this technique provide the same clinical outcomes than laparotomy with shorter hospital stays and fewer postoperative complications [159,160]. Robotic-assisted laparoscopic surgery appears to facilitate the surgical approach and can be used if available [161].

The necessity of performing a lymphadenectomy (surgical removal of one or more groups of lymph nodes) for the assessment of lymphatic dissemination has been a focus of debate [162], and the decision of the extent of lymphadenectomy varies tremendously between surgeons. Several randomized trials showed no evidence of benefit in terms of overall or recurrence-

free survival for pelvic lymphadenectomy when the tumor is in its early stages [163,164]. Additionally, it can be associated with long-term morbidity such as lymphedema [165]. Therefore, lymphadenectomy should only be performed in cases of advanced stages of disease or early stages in patients with bad prognostic features and/or risk factors. Following this criteria, patients with grade 1 or 2 EEC with less than 50% myometrium invasion are excluded from lymphadenectomy [33,97]. Sentinel lymph node assessment may help to determine which early-stage EC patients will benefit from lymphadenectomy and could provide important data to tailor adjuvant therapy [166,167].

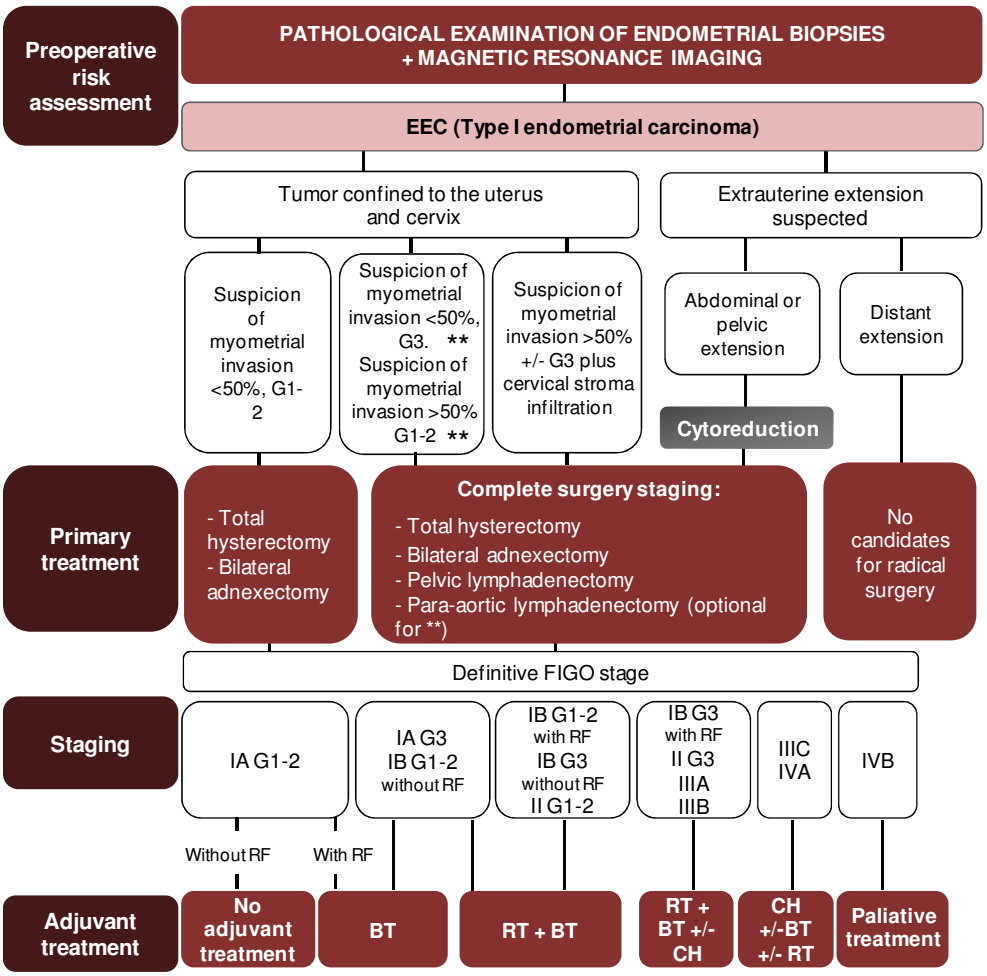


Figure 8. Treatment of endometrioid EC. G, grade; RF, risk factors; BT, brachytherapy; RT, radiotheraphy; CH, chemotherapy. *Adapted from Oncoguía SEGO 2016* [97].

A complete staging should be performed in patients diagnosed with NEECs, since these tumors present high risk of extrauterine dissemination even when they are found in their initial stages. This includes hysterectomy with bilateral adnexectomy, pelvic and para-aortic lymphadenectomy, omentectomy, and peritoneal biopsies [33,97] (Figure 9).

3.7. HISTOPATHOLOGICAL STAGING AND ADJUVANT TREATMENT

3.7.1. Histopathological staging

The final EC staging is only obtained with the pathological examination of the surgically resected tissue. The selection of the most optimal adjuvant treatment is based on the risk of recurrence established by the risk stratification systems described in section 3.5.5 (page 27) and Table 6. It is accepted that the ≈55% of EC patients that are diagnosed with FIGO grade 1 or 2 EEC, myometrial infiltration <50%, and no lymphovascular invasion (low risk group) will not benefit from any adjuvant postsurgical therapies [33], since they present an overall 5-year survival rate of 96% [69]. For all other patients, some form of adjuvant treatment has been considered based on the evidence from several large studies (Figures 8 and 9).

3.7.2. Adjuvant treatment

3.7.2.1. Radiotherapy

Radiotherapy can be delivered externally to the pelvis, internally (vaginal brachytherapy), or using a combination of both. Brachytherapy is a form of radiotherapy where a radiation source is placed inside or next to the area requiring treatment, giving a high radiation dose to the tumor while reducing the radiation exposure in the surrounding healthy tissues. Therefore, it should be the adjuvant treatment of choice over whole pelvic radiation therapy in patients with early stage EC [168].

Nowadays, radiotherapy is mainly indicated for patients with EEC carcinomas having tumors that present a high differentiation grade or that have infiltrated the myometrium [169,170]. It is also prescribed to patients who cannot undergo surgery as primary treatment.

In patients with NEEC, radiotherapy is recommended as a combined, post-chemotherapy treatment in the cases described in Figure 9 [97].

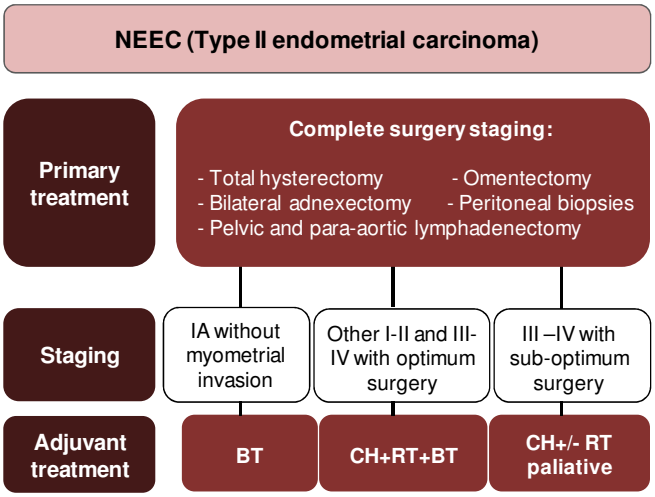


Figure 9. Treatment of non-endometrioid EC. BT, brachytherapy; RT, radiotheraphy; CH, chemotherapy. *Adapted from Oncoguía SEGO 2016* [97].

3.7.2.2. Chemotherapy

Unlike radiotherapy, brachytherapy and surgery, which are localized treatments, chemotherapy is a systemic treatment. The administration is intravenous and drugs travel throughout the entire body in order to control tumor dissemination, but also increasing the side effects. It is often delivered in combinations and in cycles.

According to the SEGO guidelines, sequential treatment with chemotherapy after radiotherapy can be considered in advanced stages of EEC cases with an increased risk of recurrence (stage IB grade 3, II grade 3 and/or any stage IIIA and IIIB) [171]. Only in stages IIIC and IV, after complete surgery, chemotherapy is recommended as primary treatment [172]. The use of subsequent radiotherapy should be evaluated. The standard pharmacological regimen for EEC tumors consists of 4 to 6 cycles of carboplatin AUC 5 and paclitaxel (175mg/m2) every 21 days [97].

In NEEC cases, chemotherapy is the adjuvant treatment of choice, since these tumors present a high risk of local and distant recurrence. Only in cases at the IA stage without myometrial infiltration, other adjuvant treatment can be considered. The standard pharmacological regimen for NEEC cases consists of 4 to 6 cycles of carboplatin AUC 5 and paclitaxel (175mg/m2) every 21 days followed by radiotherapy with or without brachytherapy [97].

3.7.2.3. Hormone therapy

The knowledge that the development of EC is associated with excess estrogen stimulation has resulted in the use of progestational agents such as medroxyprogesterone acetate, hydroxyprogesterone caproate, and megestrol acetate in the treatment of EC [173,174]. Progestational agents are an option for primary treatment only for inoperable patients or for patients who do not want to undergo surgical treatment. They can also be used as therapy for advanced or recurrent tumors that express progesterone receptors and/or are well-differentiated. The effectiveness of progestational agents has been reported to increase with the combined use of estrogenic compounds, such as tamoxifen [175].

4. BIOMARKERS IN MEDICINE

4.1. BIOMARKER DEFINITION AND TYPES

Biomarkers are a keystone of medical care. Their tremendous potential to revolutionize clinical practice has been extensively documented [176–178]. The National Institutes of Health **definition** of a biomarker is "a biological molecule that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [179].

An **ideal biomarker** should be measurable in a non-invasive clinical sample by reliable, robust and reproducible techniques. It should have excellent diagnostic and/or prognostic performance associated with high sensitivity (low rate of false negatives) and high specificity (low rate of false positives). It must be validated across a wide and representative range of populations and the final assay should be simple, cheap and thus accessible to all the populations requiring it [180].

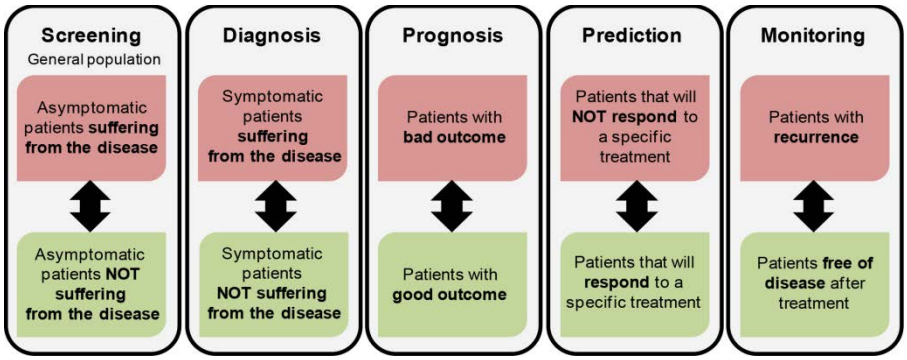


Figure 10. Clinical uses of biomarkers. Groups of patients to be differentially classified by the biomarkers.

Biomarkers have several potential applications, including screening, diagnosis, prognosis, prediction of response to treatment, and monitoring the progression of disease. Each type of biomarker must enable the discrimination of groups of patients with different status of the disease that require different management (Figure 10). Moreover, each application needs some specific biomarker characteristics [181]. In cancer, **screening biomarkers** must detect cancer cases at a curable stage before symptoms develop and result in a reduction of disease-specific morbidity and/or mortality. Importantly, a screening program must be cost-effective, minimally invasive, as it would be prescribed to a healthy population, and highly

specific to minimize false positives. As the screened population can be large, a small false positive rate could generate a large number of unnecessary diagnostic procedures associated with high costs and stress for the patients. The search of screening biomarkers is very challenging and may not even be feasible for some cancers with very low prevalence in the population. One of the most well-known and widely used cancer screening biomarkers is prostate-specific antigen (PSA) [182], although nowadays its use is controversial, as PSA screening leads to overdiagnosis and unnecessary treatment of prostate cancer [183]. A **diagnostic test** is applied to individuals already presenting symptoms and hence, it must discriminate among a narrower range of the population [181]. Nevertheless, diagnostic biomarkers must also avoid false positives and must detect cancer at the earliest stage possible while reducing the number of invasive procedures and the overall cost of the diagnostic process. **Prognostic biomarkers** provide information about the likely outcome of the cancer, independent of therapy, and hence, are used to classify patients into different risk groups. These biomarkers are useful when different therapies exist for those groups. Patients with a good prognosis could be spared from receiving more aggressive treatments, thus avoiding side effects and reducing costs [181]. **Predictive biomarkers** are those able to predict subpopulations of patients who are most likely to respond to a specific treatment. These biomarkers are the basis of the personalized medicine [184]. The evaluation of prognostic and predictive biomarkers is challenging, mainly due to the far endpoints of interest (i.e., disease-free, survival rate). One of the most widely used prognostic and predictive biomarker is the estrogen receptor in breast cancer [185–187]. Finally, **monitoring biomarkers** are used to supervise that patients remain disease free after therapy. A monitoring test must be sensitive and specific to ensure continuation of useful therapies and early replacement of ineffective treatments [181].

4.2. PROTEIN BIOMARKERS

Various classes of biological molecules can be considered as disease indicators, including DNA, RNA, proteins and metabolites. Proteins present striking advantages as biomarkers. They are the biological end products that drive both normal and disease physiology and are the targets of most current drugs [188]. Proteins are more diverse than DNA and RNA, since alternative splicing and post-translational modifications generate several proteoforms from each gene. More than 200 types of post-translational modifications have been described [188,189]. Humans have an estimated number of 20,300 genes [190], 40,000 metabolites [191] and about 100,000 mRNAs. In contrast, the human genome may potentially produce up to 1.8 million different protein species [192]. This vast diversity of proteoforms increases the probability of identifying a specific protein or a group of proteins which are associated with a disease. Another advantage is that a significant part of the proteome is detectable in

biofluids which are easily accessible and abundant. Importantly, protein biomarkers can be detected and quantified by techniques which are widely available in hospitals, such as immunohistochemistry or enzyme-linked immunosorbent assay (ELISA). The translation of a protein biomarker to a clinical test is expected to be quite rapid and efficient due to the easiness to adapt a protein-based immunoassay onto a standard clinical platform [193,194].

However, the measurement of proteins presents several challenges. A side effect of the enormous number of proteins is the vast dynamic range of concentrations that spans over six to seven orders of magnitude in cells or tissues and up to 12 orders of magnitude in plasma [195–197]. Consequently, the detection and quantification of low-abundance proteins is challenging, as they can be masked by high-abundance proteins. In addition, proteins are more dynamic than genes, and their activity can be different depending on the spatial subcellular location and the protein-protein interactions [188]. Furthermore, the difference of a protein between two groups (e.g., healthy and diseased patients) may be due to post-translational modifications, while its levels remain the same. Finally, proteomic techniques started to develop latter than genomic techniques. However, analytical approaches for the analysis of proteomes have rapidly evolved in the last years to provide better sensitivity, reproducibility, throughput and multiplexing [198], favoring the discovery of new protein biomarkers.

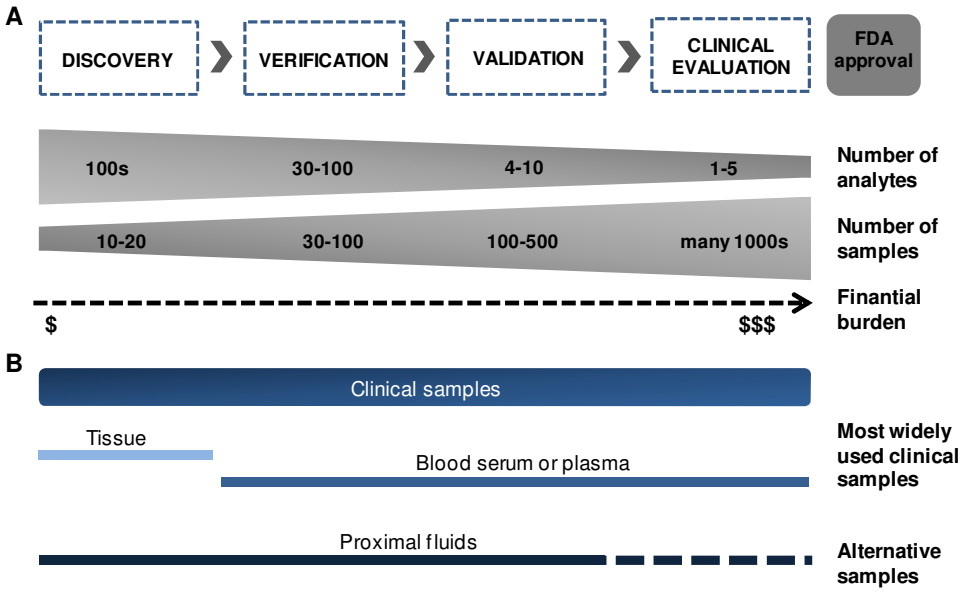
4.3. BIOMARKER PIPELINE

Typically, the phases of the biomarker development pipeline consist of discovery, qualification, verification and validation phases prior to the final clinical evaluation that leads to the implementation of the biomarker in the clinics after the approval of health regulatory agencies, such as the European Medicines Agency (EMA) or the U.S. Food and Drug Administration (FDA) [199,200] (Figure 11).

Over the last decades, proteomic studies have mainly focused on the "**discovery phase"**, using analytical methods designed for characterizing as much of the proteome as possible, though with a limited quantitative performance. This phase is an untargeted process whose goal is to look for differential proteins, usually between two simplified and homogeneous groups (e.g., healthy and diseased patients). The number of biological samples, usually tissue samples, analyzed in this first step is low (typically 5-10 samples per group) and the number of differentially expressed analytes identified ranges from 10s to 100s. In this first phase, the false discovery rate of the differentially abundant proteins is expected to be high, particularly for the low-abundance proteins, mainly due to the limited accuracy of the quantification, and the small number of samples analyzed. Therefore, they are referred as candidate biomarkers, not biomarkers [199]. The next phase, "**qualification**", is used to

assess whether the differential levels of the candidate biomarkers from the previous discovery phase can be confirmed by targeted quantitative approaches used in next steps; and whether they can be assessed in a biological sample suitable to be used in the clinic (typically, non invasive samples such as blood or urine) [199]. The "**verification phase"** is a crucial step that acts as a link between discovery and validation phases, and it has been defined as the current bottleneck of the biomarker pipeline [201–204]. The main goal of this phase is to prioritize from the large lists of candidate biomarkers generated in discovery phases, those with an increased likelihood to become a clinical assay, justifying further investment in a subsequent validation phase. In this step, the analysis of multiple biomarker candidates (30-100) is extended to a larger number of samples (30-100), now including a broader range of cases and controls, in order to confirm the sensitivity of the candidate biomarkers and begin to assess their specificity. In this phase, highly multiplexing quantitative techniques are needed. The "**validation phase"** is the final key step in the biomarker pipeline prior to the clinical evaluation. This phase requires high investment and working time and hence, it is usually set to assess only the most promising candidates (4-10). These proteins are studied in a large set of samples (100s). The composition of the sample set must represent as much as possible the diversity of the clinical conditions of the target population. In this phase, an accurate absolute quantification that allows comparison across sample cohorts, analytical platforms and laboratories is required [205] and, as the sample sets are large, high-throughput analytical methods are highly desirable. Finally, a **clinical large-scale evaluation** is needed before FDA approval and commercialization. During this phase, the final biomarker, or panel of markers, showing a great accuracy and cost-effectiveness are quantified on 500-1000s of the preferred clinical samples (typically, non-invasive body fluids) in a rigorously standardized way using the analytical approach that would be applied in the clinical practice [200].

Despite the important efforts and investments made in the search for clinically useful protein cancer biomarkers in the last decades, very few have been granted **FDA approval.** Only Pro2PSA for prostate cancer, HE4 protein and its combination with CA125 (ROMA) and OVA1 (a panel of 5 proteins) for ovarian cancer, and the fibrin/fibrinogen degradation product for monitoring progression of colorectal cancer have been approved by the FDA over the last 10 years [206]. Several comprehensive reviews have summarized the reasons for the lack of translation of research results into clinical applications [207,208]. Regarding EC, many studies have been performed in order to look for new biomarkers but yet, no protein biomarker has been applied in the clinical environment.



Tissues	Blood	Proximal fluids (Uterine aspirates)
• Higher concentration of the potential biomarkers • Most appropriate negative control (nearby healthy tissue from same patient). • Certainty that the potential biomarker comes from the affected tissue	• High availability • Minimally invasive collection • Routinely analyzed in clinical laboratories • Very constant protein concentration among people	• Minimally invasive collection • Collected in the early steps of the current diagnostic process • In contact with the tumor: enriched in proteins from the affected tissue

Figure 11. Biomarker pipeline and clinical samples usually employed in each phase. A. Characteristics of the phases of the biomarker pipeline. **B**. Clinical samples usually employed and main advantages that make them appropriate for each phase of the biomarker pipeline.

4.4. CLINICAL SAMPLES: SOURCES OF BIOMARKERS

A variety of clinical samples such as tissue specimens, blood, and proximal fluids are suitable for protein biomarker identification. Although all types of samples could be potentially used in any phase of the biomarker pipeline, their different characteristics make each of them more appropriate for some specific phases (Figure 11).

4.4.1. Tissue samples

Tissue samples collected by biopsy or surgery are the most commonly used clinical samples in the discovery phase of the biomarker pipeline. One of the main advantages of this biospecimen is the higher concentration of any potential biomarker in the affected tissue (e.g., cancerous tissue) compared to diluted in any body fluid. Additionally, biomarkers in a tissue sample can be studied within their cellular environment and importantly, their levels in the diseased tissue can be compared to nearby healthy tissue, which would represent the most appropriate control. Finally, the use of cancerous tissue samples reduces the need to demonstrate that the identified biomarker originates from the actual tumor [209]. However, tissue samples are limited and their collection requires invasive procedures for the patient. Accordingly, they are not optimal samples for advanced stages of the biomarker pipeline and clinical application when searching for screening or diagnostic biomarkers. Furthermore, the proteomic analysis of tissue samples has several methodological challenges, such as cell heterogeneity in the tissue that may mask the variation in protein levels characteristic of a specific cell type, and protein cross-linking when working with formalin-fixed, paraffin-embedded (FFPE) tissues. Some approaches to overcome these challenges are the laser capture micro-dissection [210] and novel protocols for FFPE tissues [211,212], respectively.

4.4.2. Blood serum or plasma

Blood serum or blood plasma is the preferred clinical sample for the identification of biomarkers. It is collected in a rapid, easy and minimally invasive way and it is routinely analyzed in the clinical laboratory. Its ease of collection allows to monitor the patient at several time points. Therefore, steps beyond the discovery phase of the biomarker pipeline would be preferably performed in this sample (Figure 11) [213]. As blood is in direct contact with all body organs, its content potentially includes leakage or secretion proteins from all tissues. In addition, total protein concentration in plasma is very constant among people, enabling a more straightforward comparison of specific protein levels between patients without the need of normalization methods. However, plasma is one of the most challenging samples to be analyzed by proteomic techniques. The wide dynamic range of more than eleven orders of magnitude in protein abundance [197] increases the difficulty to detect low-abundance proteins. The presence of very high-abundance proteins such as albumin (35-50 mg/ml over 70 mg/ml total protein concentration) masks potential biomarkers which are coming from the body tissues and massively diluted into circulating blood to a concentration range of ng/ml and below (cytokines and other proteins are present at the pg/ml level) [197]. The fact that blood flows through all organs also hampers the identification of biomarkers specific to a concrete process occurring in a specific part of the body.

4.4.3. Proximal fluids

In contrast to blood, a "proximal fluid" is a biofluid in direct contact or close to the site of disease. Consequently, proximal fluids are enriched in potential biomarkers coming from the diseased tissue and hence, they arise as an attractive alternative for biomarker identification [199,214] (Figure 11). Some examples of proximal fluids include cerebrospinal fluid for the study of intracranial processes, urine for renal and urological diseases, ovarian cyst fluid and ascites fluid for ovarian cancer, amniotic fluid for fetomaternal screening and nipple aspirate fluid for breast disease. The use of these proximal fluids for biomarker search is also challenging, as their proteomes are less characterized and can be more variable than the plasma proteome. In addition, each type of proximal fluid presents its own limitations, such as invasive collection procedures, small sample volume, low amounts of secreted proteins, and/or frequent blood contamination [213,215,216]. Currently, urine is one of the most widely studied proximal fluids, since it is obtained non-invasively and in large quantities and thus, is suitable for all steps of the biomarker pipeline. Urine has been employed in the search of biomarkers for diseases such as prostate and bladder cancer [217–219].

For gynecological diseases such as EC, pipelle biopsies (also called uterine aspirates) may be an interesting source of biomarkers. As explained in section 3.4.3.3. (page 21), this biofluid is in direct contact with the tumor in the endometrium and can be obtained by aspiration from inside the uterine cavity with a Cornier Pipelle. These biopsies contain a fluid fraction and a cellular fraction. Their use as a source of biomarkers has been exploited at transcriptomic and genomic level, but only through the assessment of molecular alterations in the cellular component. Remarkably our group identified, verified and validated a five gene qRT-PCR assay that improves EC diagnosis [220,221]. However, very few proteomic studies have been performed in uterine aspirates, and none of them focused on the search for EC protein biomarkers [222–224].

5. PROTEOMIC APPROACHES FOR BIOMARKER IDENTIFICATION

Ever since the field of proteomics demonstrated the ability to identify and quantify a large number of proteins and their post-translational modifications, it has been applied to various areas of biomedicine, such as in the identification of potential markers [225]. Protein biomarker identification shares characteristics with genomic and transcriptomic profiling, including the analysis of biological samples with complex matrix that generates large datasets and requires sophisticated statistical analysis. However, the study of the proteome is inherently more complex, mainly due to the extended range of analyte concentrations and the fact that proteins cannot be amplified, since an equivalent polymerase chain reaction (PCR) for proteins does not exist and therefore sensitivity is an additional challenge [198]. Nevertheless, there have been rapid advances in the field of proteomics in the last years, mainly driven by the methodological and technological advances of mass spectrometry (MS) together with the development of bioinformatics.

In this regard, **MS-based approaches** have become a driving force in the initial steps of the biomarker pipeline (i.e., discovery and verification phases), whereas **antibody-based approaches** are still the gold standard for the final validation steps of the pipeline.

5.1. MS-BASED PROTEOMICS

5.1.1. MS basics

Despite the significant achievements in the MS-based identification of intact proteins (top-down proteomics) [226], the vast majority of proteomic data has been generated by **bottom-up proteomics**. The general workflow consists of several steps (Figure 12):

- **Sample preparation**

Proteins must be extracted from the tissue samples by lysis and solubilization. This is not a requirement for biofluids such as plasma or urine because proteins are already soluble, but purification and concentration steps might be necessary in those cases. Proteins are then denatured by heat or by using denaturation reagents such as urea, the disulfide bonds are reduced and the free cysteins are alkylated in order to break the tridimensional structure of proteins and thus allow for a more efficient proteolysis by providing the proteolytic enzymes maximum access to cleavage sites within the proteins. Finally, proteins are cleaved into peptides by a protease or a mixture of proteases with high specificity [227,228]. Trypsin is the most widely used protease and it specifically cleaves peptide chains at the carboxyl side of

the amino acids lysine (K) and arginine (R), generating peptides of an average size of 14 amino acids with a predictable MS fragmentation pattern [229]. Tryptic peptides are also convenient to separate by reverse phase liquid chromatography, a separation technique fully compatible with MS analysis. Apart from trypsin, other proteases such as Lys-C or a combination of multiple enzymes can be used in order to increase protein sequence coverage [229].

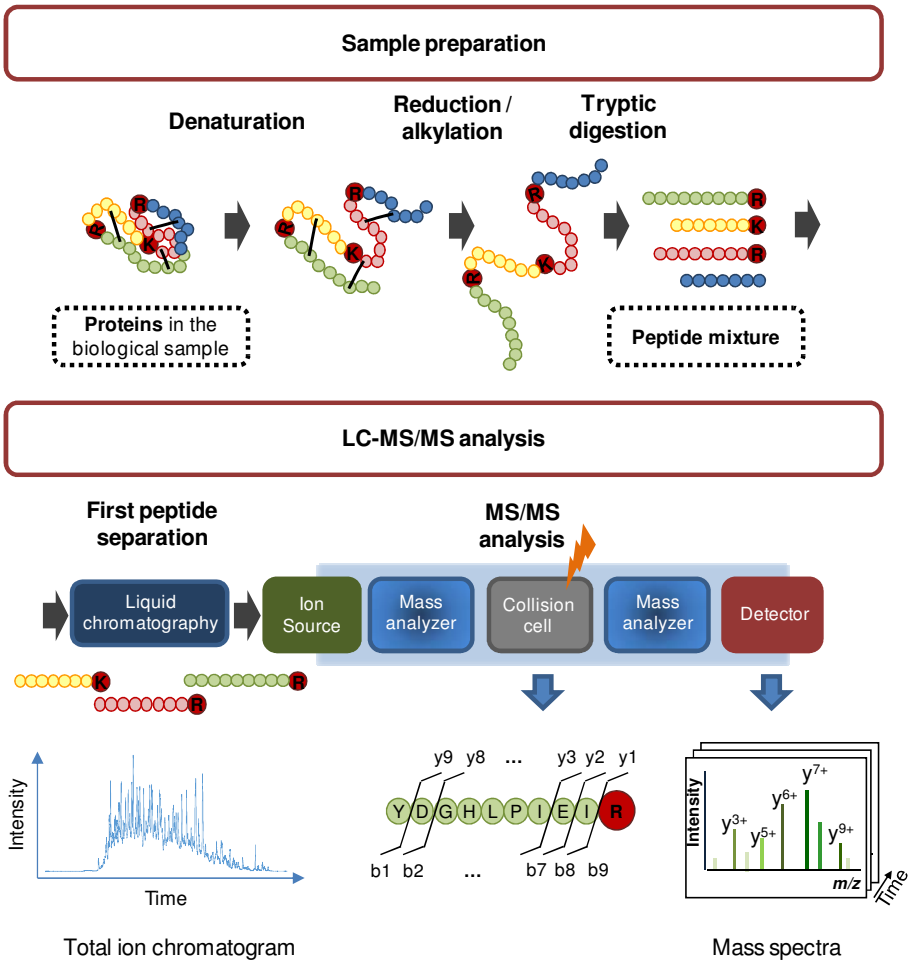


Figure 12. Bottom-up MS-based proteomics workflow. Proteins are digested with trypsin and resulting peptides are analyzed by reverse phase liquid chromatography coupled with mass spectrometry detection.

MS-based approaches achieve a good sensitivity, but the depletion of the most abundant proteins or, conversely, an enrichment of the proteins of interest can be used to improve the detection limit in clinical samples. A number of methodologies have been proposed, such as depletion of the most abundant proteins (e.g., albumin and immunoglobulin G (IgG) in plasma), glycoprotein enrichment, capture of peptides with infrequent amino-acids, or immunoaffinity purification (to be able to detect proteins at or below the ng/ml concentration level) [230].

❖ Separation prior to MS analysis

Due to the complexity of the proteome, the identification and quantification of peptides by MS is less efficient if the sample is not previously fractionated or separated [231]. The concomitance of too many compounds can generate a competition for the ionization in the source and can lead to signal suppression (i.e., matrix effect) [232]. In addition, prior sample separation reduces the risk of interferences caused by isobaric compounds. Therefore, separation techniques are employed to improve the analytical performance including the sensitivity and the coverage of the proteome. In the early stages of proteomics, two dimensional gel electrophoresis was the method of choice for the separation of proteins from complex samples prior to digestion and identification by MS (see section 5.1.3.1., page 50). Nowadays, this approach has been mostly replaced by the separation of peptides by reverse phase high pressure liquid chromatography (HPLC) that separates peptides on the basis of their hydrophobic interactions with the analytical column [227]. In this configuration, the outlet of the HPLC system is connected to a mass spectrometer (LC-MS) or several analyzers (LC-MS/MS) where the eluted peptides are directly analyzed. Off-gel fractionation is an alternative technique that fractionates proteins and/or peptides on the basis of their isoelectric point without the need of a gel-based matrix [233,234]. The fractions are further treated as regular proteomic samples and analyzed by LC-MS for an improved proteome coverage.

❖ MS analysis

Mass spectrometers measure the ratios between the masses and the charges of ions (m/z values). Therefore, peptides must be first ionized and vaporized by the ion source. The resultant ions are then separated according to their m/z by the mass analyzer and finally detected to generate a "mass spectrum", which is a plot of ions abundance against m/z values [235] (Figure 12). In the field of proteomics, the ionization process is mostly performed in positive mode, by addition of protons. The electrospray ionization (ESI) is well adapted for the analysis of peptides and proteins as it can be directly coupled with a liquid chromatography device and it generates multicharged peptides which promote efficient

fragmentation at low energy collision induced dissociation (CID). After being ionized and converted to gas phase, the peptide ions are analyzed by one or several mass analyzers, depending on the mass spectrometer configuration. These include low resolution analyzers like the quadrupole and the ion trap, and high resolution/accurate mass analyzers like the time-of-flight (TOF), the orbitrap and the Fourier transform ion cyclotron resonance (FTICR) analyzers [231,236]. Nowadays, proteomic studies employ tandem mass spectrometers which are built by a combination of several analyzers and hence, both LC-MS and LC-MS/MS terminologies usually refer to a combination of mass analyzers. Common tandem mass spectrometers configurations for proteomics studies are the quadrupole time of flight (QqTOF), the quadrupole orbitrap (Q-OT), the triple quadrupole (QqQ) and the ion trap / orbitrap, with different performances in terms of mass accuracy, resolving power, sensitivity and dynamic range [236]. In these configurations, peptides are selected in the first mass analyzer, fragmented by CID [237], and the resulting fragment ions are analyzed in the second mass analyzer. Fragmentation of peptides generates b- and y- fragment ions that allow to determine the peptide sequence [238] which, combined with the accurate mass of the precursor ion, dramatically improve the confidence of the identification (Figure 12).

5.1.2. Protein quantification by MS

MS-based proteomics has rapidly evolved over the past years from a qualitative to a more quantitative approach [239]. Although LC-MS is inherently a quantitative platform, the signal is subjected to variations. These variations are principally due to changes in the instrument performance, including variations of the injection volumes and degradation of the chromatographic column performance regarding the LC part; and the contamination or drift in the calibration regarding the MS system. In addition, the competition for the ionization in the ion source can suppress, or sometimes enhance, the signal of an ion species. Several strategies can be used for protein quantification. These can be broadly divided in label-based and label-free approaches.

5.1.2.1. Label-based approaches

An efficient approach to control for the variations in the sample preparation and/or LC-MS analysis consists in the incorporation of amino acids labeled with stable isotopes (^{13}C, ^{15}N and/or ^{18}O) into internal standards. The isotope-labeled peptides, often called heavy peptides, display the same sequences and similar physico-chemical characteristics to that of their endogenous peptides, also called light peptides, but are distinguishable by MS due to their mass increase [240]. Equal amounts of labeled internal standard are added to all samples to be analyzed and, as each heavy and endogenous peptide pair displays the same chromatographic behavior, ionization efficiency, and fragmentation patterns, the MS signal

of each endogenous peptide can be normalized by the signal of its isotopic labeled version to control for variability factors. Furthermore, the labeled internal standards should be spiked as early as possible during the sample preparation procedure in order to control the maximum number of steps of the sample processing and decrease the technical variability [239]. The addition of isotope labeled standards not only enables a more accurate relative or absolute quantification but also provides an extra step on the confidence of the peptide identification due to the co-elution of the endogenous peptide and the internal standard (used as the reference) and the possibility to perform spectral matching [241].

Several strategies are employed to introduce labeled internal standards, including the use of isotopically labeled proteins, synthetic peptides or isotopically labeled derivatization reagents.

❖ Isotopically labeled proteins

Full-length isotope-labeled proteins are the ideal standard for quantitative proteomics [242,243]. In contrast to peptide standards, adding isotope-labeled proteins in early steps of the sample preparation workflow enables to control for variations that may occur during proteolysis and pre-fractionation steps.

The chemical synthesis of proteins is almost impossible due to their size (it is difficult to synthesize over 30 amino acids) and the challenge to reproduce the specific folding and tridimensional structure. Metabolic incorporation by stable isotope labeling by amino acids in cell culture (SILAC) is a feasible method for the production of isotope-labeled proteins. In this strategy, amino acids labeled with ^{13}C and/or ^{15}N (typically, arginine and lysine) are added to an amino acid deficient culture media where cells with a specific biological condition (control, cancer or cells under a treatment) are grown. The labeled amino acids are therefore incorporated into all newly synthesized proteins [244] and after proteolysis, all tryptic peptides will be labeled on the C-terminus. The cell lines should be kept in culture until they reach an incorporation rate over 95%. The labeled cell line is then mixed with another biological condition that is not labeled and they are quantitatively compared. With this approach, the relative quantification of the complete proteome between two states is possible. However, since SILAC requires complete metabolic labeling of proteomes, it is applicable only to cultured cells or, at most, to small organisms like mice [245]. Moreover, the comparison of multiple samples is not easy.

The super-SILAC approach emerged as a variant applicable to tissue samples or biological fluids [246]. This approach consists in combining an assortment of cell lines, a super-SILAC mix, to be used as a spike-in standard. The design of the appropriate super-SILAC mix is crucial for the outcome of the experiment. The cell line mixture that better represents the

proteome of the clinical sample, the ratios in which these cells should be mixed, and the ratio of the super-SILAC to be spiked in the clinical sample must be determined [247]. It is an appropriate approach for untargeted studies such as discovery phases of the biomarker pipeline (Figure 13B), since this approach presents the advantages of conventional SILAC, enabling the relative quantification of the complete proteome. Moreover, this approach allows for the comparison of multiple samples. Its main disadvantage is that the dilution of the sample of interest by a complete exogenous proteome significantly increases the complexity of the sample, and thus may reduce the selectivity and the sensitivity of the approach.

❖ Isotope-labeled synthetic peptides

Isotope-labeled peptides can be chemically synthesized in large scale by several manufacturers on the market with different quality grades ranging from relatively inexpensive non-purified peptides used for relative quantification (crude peptides) to purified and accurately quantified peptides designed for absolute quantification (e.g., AQUA peptides) [248,249]. Furthermore, post-translational modifications such as phosphorylation, acetylation, sulfation can be incorporated [250]. It is a fast and straightforward approach that introduces less complexity to the sample compared to the super-SILAC approach, and it is widely used nowadays in targeted studies where the peptides of interest are known upfront, such as the verification and validation of biomarker candidates (Figure 13B). However, this approach also has some limitations. Synthetic peptides as internal standards do not control for variations in early steps of the sample preparation, including the tryptic digestion. Additionally, synthesis of peptides over 25 amino acids can be erratic, the stability during storage should be controlled, and the preparation of a mixture of hundreds of peptides can be a tedious task if not automated.

❖ Isotopically labeled derivatization reagents

The isobaric tags for relative and absolute quantitation (iTRAQ) technology utilizes isobaric reagents to label the primary amines of peptides and proteins [251]. The isobaric tags include a reporter group, a mass balance group and a peptide-reactive group. The function of the balance group is to make the labeled peptides from each sample isobaric (same mass). The relative abundance of the peptides is deduced from the relative intensities of the reporter group that is generated upon fragmentation in the mass spectrometer. The iTRAQ approach is often used in discovery studies since, in principle, every peptide is labeled. However, it also has some limitations, such as the underestimation of the changes in abundance reported and interferences by cross-label isotopic impurities [252]. Other approaches based on

isotopically labeled reagents include isotope coded affinity tagging (iCAT) [253] and tandem mass tags (TMT) [254].

5.1.2.2. Label-free approaches

In label-free approaches, an equal amount of each sample is analyzed by MS to estimate the relative abundance of proteins across samples, without the use of a labeling strategy [255]. The sample processing and LC-MS conditions need to be carefully controlled to minimize variations and obtain optimal results [256].

Although this method provides a less accurate quantification, its ease of execution and cost-effectiveness make it appropriate for the initial discovery phase of the biomarker pipeline (Figure 13B). The label free approach is sensitive to MS signal variation and therefore, it is more reliable when samples with similar chemical background are analyzed. Due to its limited precision and accuracy, the difference between the levels of the proteins among samples must be high to be significant (greater than two-fold) [257].

5.1.3. Major mass spectrometry acquisition strategies in proteomics

MS plays an important role in proteomics as it provides sensitive, selective and highly multiplexed analyses of the samples. The recent developments in high resolution accurate mass spectrometry and the faster acquisition rate make MS a key player not only for discovery phases but also for the subsequent verification and validation phases of the biomarker pipeline in complex clinical samples [258,259]. The selection of the most appropriate MS-based proteomic approach for each specific purpose is crucial [260].

5.1.3.1. Untargeted MS approaches for biomarker discovery

In the past, MS-based proteomics has been mainly employed in discovery phases of the biomarker pipeline to characterize as much of the proteome as possible in order to identify proteins differentially abundant between two states, such as presence or absence of disease, rather than focusing on precise quantification. Different platforms are used for this purpose.

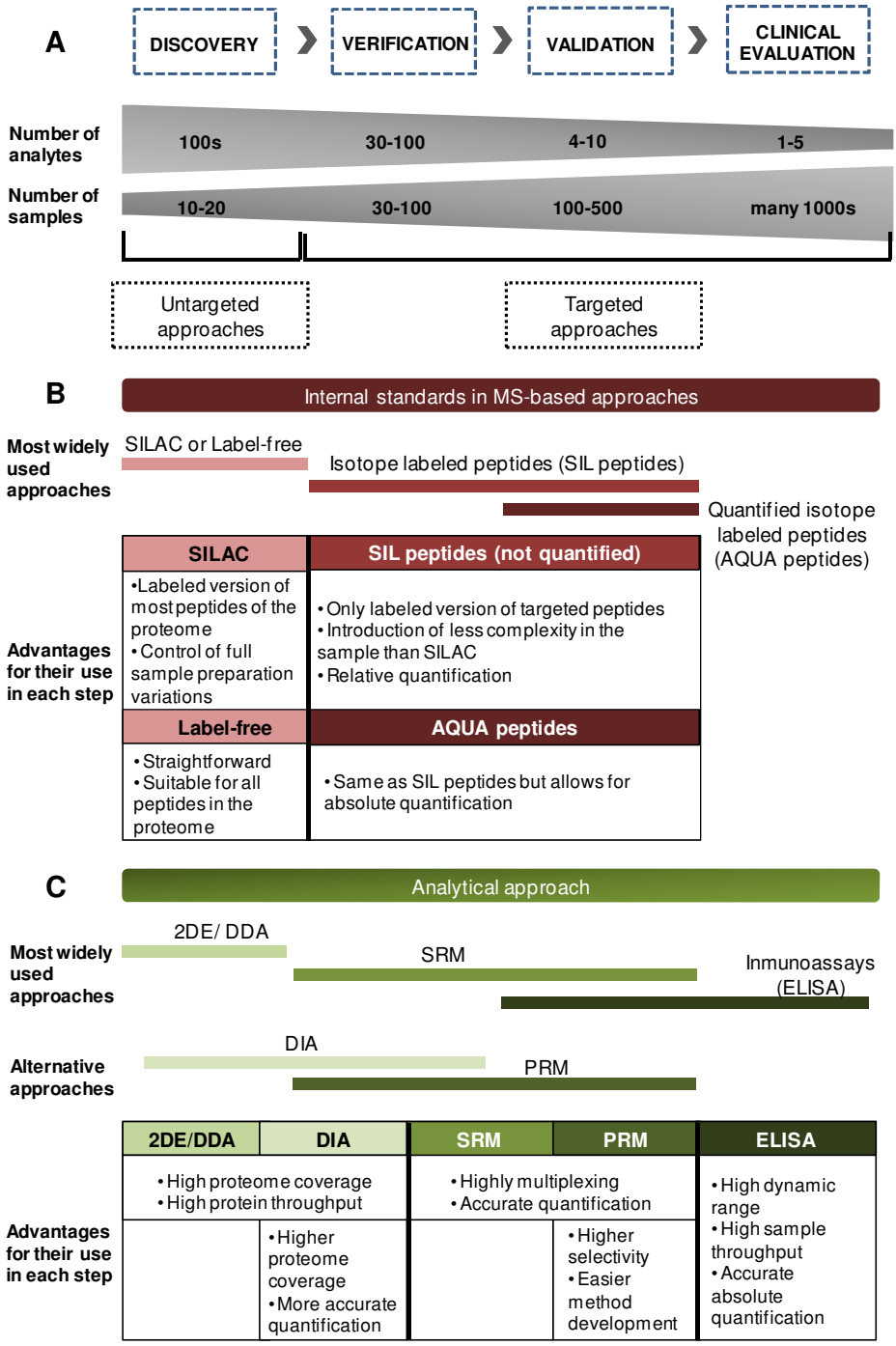


Figure 13. Internal standards and analytical approaches usually employed in each phase of the biomarker pipeline. A. Characteristics of the phases of the biomarker pipeline**. B.** Implementation of internal standards in MS-based approaches**. C.** Implementation of analytical approaches in the biomarker pipeline.

❖ Two-dimensional electrophoresis

The term "proteomics" originated in the context of **two-dimensional electrophoresis** (2DE) [261]. This gel-based strategy represented a fundamental evolution in the field of separation technologies. In 2DE experiments, proteins are separated on polyacrylamide gels based on their molecular weight and isoelectric point. After staining, each observed protein spot is quantified based on its staining intensity. This technique allows for the separation of thousands of proteins on a gel as well as the separation of intact isomeric forms of a protein for the study of post-translational modifications [262]. Each spot to be identified requires to be excised, digested in-gel and that the peptides are identified by MS, typically by the peptide mass fingerprinting method using a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer (MALDI-TOF) [263–265]. The identification confidence can be improved by tandem MS on MALDI-TOF/TOF systems [266]. The 2DE technology is easily accessible to almost any laboratory [262]. However, the lack of reproducibility of 2DE hampers the comparison of different gels (each containing the proteins from a single sample).

Two-dimensional differential gel electrophoresis (2D-DIGE) emerged as an improvement of conventional 2DE that facilitates the comparison among different gels [267]. Two samples are differentially labeled with fluorescence dyes (i.e., Cy3 and Cy5) and simultaneously resolved within the same gel, being subjected to the same procedures. Additionally, a third sample (a pool of all the samples included in the study) labeled with a third dye (Cy2) is introduced as internal standard in the same gel. This internal standard enables the comparison among all gels in the study. Although 2D-DIGE overcomes some of the problems associated with 2DE, it is still considered a very tedious, low throughput technique with a limited dynamic range. It is also dependent on the size and hydrophobicity of the proteins being studied [268]. Moreover, there is an indirect link between quantitative measurement, based on staining intensity, and protein identification, based on MS. Nowadays these techniques, although popular in the past, have been outperformed by a more powerful generation of gel-free MS proteomics, known as "shotgun proteomics", described below.

❖ Data Dependent Acquisition (DDA)

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is currently the gold analytical MS platform for the measurement of proteins in biological samples [269,270]. After a step of proteolysis, the peptide mixtures are separated by reverse phase chromatography prior to enter the mass spectrometer. The DDA acquisition is a powerful acquisition method for protein identification without the need of preliminary knowledge of the sample content prior acquisition and hence, DDA acquisition is the method of choice for biomarker discovery phase (Figure 13C). In this approach, the precursor ions are detected

by a MS1 survey scan and the most intense precursors (typically, 4 to 12) are selected for fragmentation and detection of all fragment ions, giving a full-scan MS/MS spectra. This process is repeated during the whole chromatographic separation leading to a list of several thousands of MS/MS spectra. Then, the MS data are processed by searching algorithms (e.g., Mascot, Sequest) that match the experimental MS/MS spectra with the theoretical spectra generated from protein sequence databases (e.g., Uniprot) [271–273]. The sequences of the proteins from the databases are *in silico* digested into peptides using the same cleavage rules as the actual protease used in the sample preparation and then the fragment ions masses are calculated. The search algorithm matches the best corresponding peptide sequence with the experimental MS/MS spectra and the mass of the related precursor ion. The identification of peptides that are unique for a specific protein leads to the identification of the related protein.

This method requires mass spectrometers able to select precursor ions through isolation windows of 1-3 m/z and to perform their fragmentation. The performance of DDA has been improved by the use of high resolution mass accurate spectrometers [274], particularly the time-of-flight [275] and the orbitrap mass analyzers [276,277], as the MS data acquired with these instruments reduce the ambiguity on the peptide matching.

This approach achieves a much higher sample throughput and higher proteome coverage than gel-based approaches, but can be biased toward high-abundance proteins due to the selection of precursor ions based on intensity. Another consequence of the stochastic selection of precursor ions for fragmentation is the limited reproducibility between sample replicates. Several studies showed that lists of identified peptides between replicates only overlapped by 35-60% [278,279].

Quantification in untargeted approaches can be either label-free or use SILAC or isotope labeled tag like iTRAQ and iCAT.

❖ Data Independent Acquisition (DIA)

Recently, DIA has emerged as a viable unsupervised acquisition scheme for both qualitative and quantitative analyses that overcomes several limitations of the traditional techniques previously described [280]. The principle of the DIA acquisition is to generate a comprehensive and unbiased map of a proteome by performing MS/MS fragmentation without selection of a particular precursor ion.

One popular method to improve the selectivity of DIA is the Sequential Window Acquisition of all Theoretical Fragment-ion spectra (SWATH) [281]. This DIA acquisition scheme is based on the systematic acquisition of sequential precursor ion mass windows (typically 10-50 m/z

units) that cover the entire mass range. All precursor ions in each window are fragmented and all fragments are analyzed in MS/MS mode at high resolution.

This acquisition method virtually covers all detectable peptides in a sample and thus allows for iterative data mining without the need of re-acquisition. However, the concomitant fragmentation and MS analysis of all the ions available generate extremely complex MS/MS spectra and therefore, the data processing, which is usually based on spectral matching using a reference spectra library [282,283], is very challenging. Other disadvantages of this approach are the lack of selectivity compared with methods that use a precursor selection [284], and the loss of a direct link between the precursor masses and the MS/MS spectrum, as it exists in DDA. Finally, DIA usually employs label free quantification strategies.

To date, DIA has already been applied to the identification of biomarkers in several types of cancer such as esophageal, lung and nasopharyngeal carcinomas [285–287].

5.1.3.2. Targeted MS approaches for biomarker verification

Antibody-based assays have been the most common methods of choice for targeted studies, but their low protein throughput makes them more suitable for final validation phases focused on a restricted list of biomarkers. Highly multiplexed targeted MS-based techniques have gained in popularity for verification studies, in order to evaluate a large number of potential biomarkers coming from discovery studies and prioritize the most promising candidates to enter into further validation phases [202]. In addition, the improvement in sample throughput of the MS technologies facilitates their use in larger sample sets in initial validation phases (Figure 13C).

In targeted MS approaches, a list of peptides representing the proteins to be measured must be selected before the actual MS acquisition and measured by repeated MS/MS events through acquisition windows. The data points extracted from the MS/MS events allow to recreate the chromatic elution profiles of the peptides that can be integrated for further quantification. As for DDA, targeted acquisition relies on the filtering of the precursors through a narrow mass window of 1 m/z before MS/MS fragmentation, significantly improving the selectivity compared to DIA experiments. However, targeted experiments are limited in the number of peptides that can be monitored per analysis. In order to maximize the number of measurable peptides, the liquid chromatography retention time of each peptide should be determined in a preliminary analysis to be able to monitor the fragment ions of specific peptides in the corresponding time windows. This is called scheduled acquisition [288]. Whereas DDA acquisition requires the detection of a precursor in a survey scan to trigger a MS/MS event, targeted acquisition systematically performs MS/MS events during the acquisition windows which are centered on the elution times of each targeted

peptide. Therefore, targeted acquisition does not generate missing data and replicates can be quantitatively compared as they perfectly overlap.

❖ Selected Reaction Monitoring (SRM)

The SRM acquisition performed on a triple quadrupole mass spectrometer, which was elected "Method of the year 2012" by Nature Methods, became the reference technique for MS-based quantitative methods in proteomics [289].

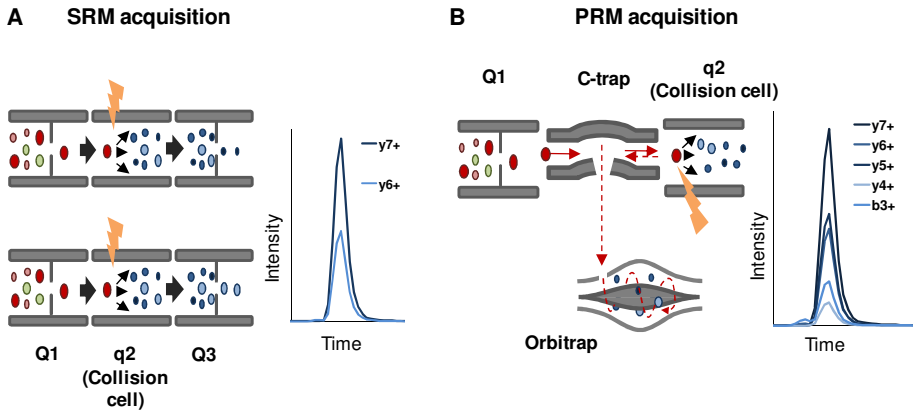


Figure 14. Comparison of the SRM and PRM acquisitions. **A.** SRM acquisition performed on a triple quadrupole mass spectrometer. **B.** PRM acquisition performed on a quadrupole-orbitrap mass spectrometer.

The SRM acquisition is performed by the selection of a precursor ion by the first quadrupole (Q1), followed by the fragmentation of the selected ion in the second quadrupole (q2) operated as a collision cell. Finally, one fragment ion is selected by the third quadrupole (Q3) and reaches the detector to generate a signal [290] (Figure 14A). The first and the third quadrupoles are usually operated to filter ions with a mass width of 0.7 to 1 m/z. A pair of a precursor associated with a fragment ion is called an SRM transition and usually three to five transitions are recorded per peptide to improve the confidence of the measurements. Regarding data processing, the quantification is usually based on the area under the peak of the elution profiles of the targeted fragment ions for improved accuracy and precision.

LC-SRM is a multiplexed acquisition method that allows for the accurate quantification of a hundred of peptides in a single LC-MS analysis. It achieves a good selectivity due to the two levels of mass filtering on the precursors and the fragment ions, and high reproducibility across many samples and different laboratories [289,291,292]. The possibility of using internal standards such as stable isotope labeled peptides or proteins brings several advantages for quantitative analysis including: i) control of the signal variation (see section 5.1.2.1., page

45); ii) a strong confidence in the peptide identification due to the co-elution of the endogenous peptide and internal standard and the similarity of the fragmentation pattern; and iii) the possibility to perform absolute quantification if the concentration of the internal standard is known [249,293].

SRM assays usually cover a linear dynamic range of 5 orders of magnitude [288], and provide a good sensitivity. However, as for most LC-MS based proteomic methods, a depletion or enrichment step can be required to quantify very low abundance proteins [230,294]. The method development in this approach is time consuming and static, as the SRM transition list must be defined prior acquisition. Since only a limited number of transitions are monitored for each peptide, the presence of interferences may jeopardize the data analysis and may require the reanalysis of the samples, decreasing the throughput of the technique [259]. Therefore, it is important to validate the specificity and sensitivity of the SRM transitions in a representative number of samples before the analysis of the full sample set [293]. The information required for the method development prior to the SRM analysis is shown in Figure 15.

❖ Parallel Reaction Monitoring (PRM)

Targeted MS-based approaches are also feasible with high-resolution accurate mass spectrometers, and the PRM acquisition performed in these instruments recently arose as a promising alternative to SRM [259,295]. The configuration of the LC-PRM shares some similarities with the LC-SRM performed on a triple quadrupole, but the third quadrupole is replaced by a high-resolution mass analyzer (TOF or orbitrap). In the case of a quadrupole-orbitrap configuration, a specific precursor ion is selected by the quadrupole and transferred via the C-trap to the collision cell where it is fragmented. The fragment ions are then transferred back to the C-trap and injected into the orbitrap to be analyzed at high resolution (Figure 14B) [296].

This acquisition method permits the detection of all fragment ions of the targeted precursor ion in one MS/MS event whereas the SRM method requires one MS/MS event per fragment ion [275,277]. The data processing is performed by extraction of the ion chromatograms (XIC) of the fragment ions of interest and, as for SRM and DIA data, quantification can be performed by integration of the areas of the elution profiles.

The PRM acquisition has several advantages over the SRM approach [259]. The high resolution dramatically reduces the risk of interferences from the background and the accurate mass improves the confidence on the fragment ions identities. In addition, because it is based on full MS/MS spectra, all potential fragment ions are recorded, instead of only the 3-5 usually targeted in SRM. A reference spectra can be used to select the fragment

ions to be extracted and to confirm the identity of the peptides by spectral matching (see Chapters 3 and 4 for details). PRM also facilitates the method development because the fragment ions to monitor are selected post acquisition and therefore, less information is required prior to the MS analysis (Figure 15). Moreover, the presence of interfering ions can be solved by using alternative fragment ions without the need of a re-acquisition, as it is the case with SRM. The PRM mode is also fully compatible with the use of isotope labeled peptides, with all the associated benefits previously described [257,297].

Although the advantages of application of these high-resolution accurate mass spectrometers in clinical studies have been evaluated [241,298], only few studies have already employed this technique for biomarker searches in clinical [299] or cell lines samples [300–302], and none of them for targeting a high number of candidate biomarkers.

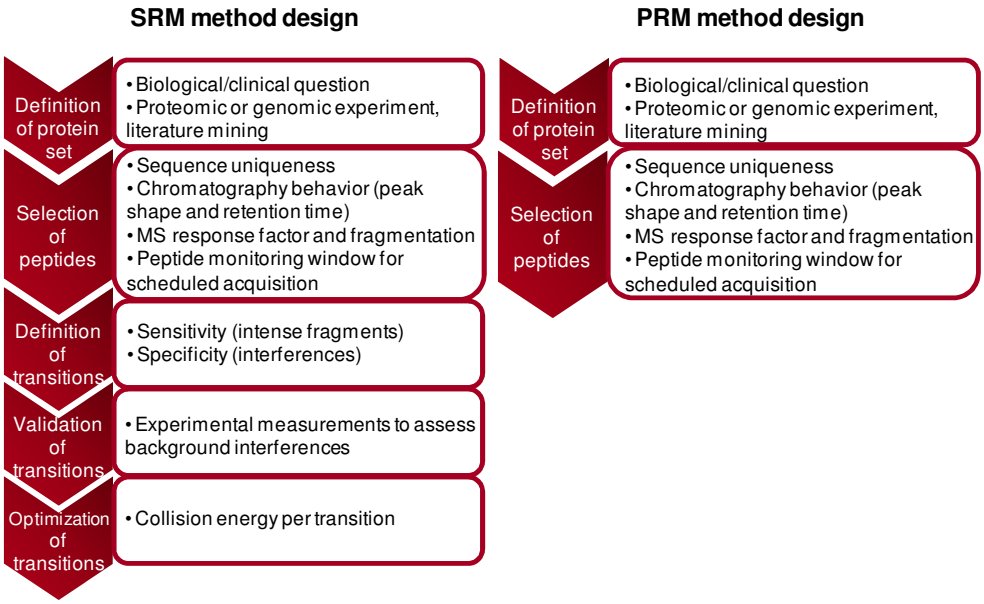


Figure 15. Steps required for the development of an SRM and a PRM method. *Adapted from Gallien et al. 2011* [292].

5.2. ANTIBODY-BASED PROTEOMICS

Although targeted MS-based approaches have significantly improved their sensitivity and sample throughput and start to be applied in validation phases of the biomarker pipeline,

antibody-based approaches continue to be the reference method for validation and clinical evaluation steps (Figure 13C).

Immunoassays use the basic immunology concept of an antigen binding to its specific antibody. Immunohistochemistry or enzyme-linked immunosorbent assays (ELISA) are widely employed in the clinical environment as they are simple and cost-effective if the antibody is already available, very sensitive and provide high sample throughput. However, they are limited by the availability of antibodies. The development of new highly specific antibodies is an expensive and time-consuming process that may take several months [303]. Other limitations are the difficulty of multiplexing, the cross-reactivity and the lack of reproducibility between platforms [202].

5.2.1. ELISA for validation and clinical evaluation

There are different types of ELISA, including direct and indirect ELISA, sandwich assays or competitive assays. Sandwich assays tend to be more sensitive and specific, being used in most commercial ELISA kits. In these assays, two different antibodies that bind to different sites on the antigen are used. The capture antibody, which is highly specific for a protein in the biofluid or sample to be analyzed, is attached to the plate surface. After nonspecific binding sites are blocked using bovine serum albumin, the antigen-containing sample is added, followed by the addition of the detection antibody, which binds the protein at a different epitope than the capture antibody. Different reporters can be used (enzyme, fluorophore, or biotin). They can be directly attached to the detection antibody or to a secondary antibody which binds the detection antibody. Finally, the substrate is added and converted by the enzyme into a colorimetric signal that can be quantified. The signal generated is proportional to the amount of antigen present in the sample [304].

To date, ELISA remains the gold standard for validation and clinical evaluation of protein biomarkers in biofluids, mainly due to several outstanding advantages. First, it is a highly sensitive technology. Body fluids such as plasma or serum present a wide dynamic range of protein abundance that spans over 10-12 orders of magnitude [197]. Therefore, a high level of sensitivity is required in order to detect and quantify putative new disease protein biomarkers that are likely to be present in biofluids at extremely low concentrations. While an SRM method reaches sensitivities within the ng/ml range in plasma, proteins such as interleukin-6 can be reliably measured by ELISA at concentrations as low as 0.15 pg/ml [199,305]. Second, when high quality antibodies are available, the ELISA immunoassay is fast and straightforward since the analysis of biofluids does not require any kind of prior sample preparation, and provide a high sample throughput. Finally, this approach allows for the absolute quantification of the protein of interest, which facilitates the comparison between

platforms and laboratories. Therefore, the ELISA immunoassay is used in clinical laboratories and consequently, an accurate biomarker detected by ELISA could be more quickly implemented in the clinics.

However, commercially available ELISA kits are often not as specific as stated by the manufacturer, and they often lack reproducibility between assays. Moreover, they are usually research tools that are not extensively validated in clinical samples, and particularly in biofluids different from plasma or serum. Therefore, *de novo* development of highly specific and reproducible ELISA assays for the proteins of interest is recommended, although it is a challenging process. In a sandwich ELISA assay, the monospecificity of the two antibodies must be validated. The development of a new ELISA assay is an expensive (>$100,000 per antibody) and time-consuming process (1-2 years) [303]. Consequently, ELISA technology is well-suited for the last steps of the biomarker pipeline where the few most promising biomarkers are quantified in a large number of samples.

Objectives

BACKGROUND

EC is the fourth most common cancer in women in developed countries and its incidence and mortality rates are increasing annually worldwide. In the United States alone about 61,380 new cases and 10,920 deaths are estimated for 2017. Patients diagnosed at early stages of the disease are associated with an overall 5-year survival rate of 95%. However, 30% of EC patients are diagnosed at an advanced stage of the disease associated with a drastic decrease in the 5-year survival rate; this rate is reduced to 69% in cases of regional metastasis, and to 17% in cases of distant metastasis [27]. Improving early diagnosis is hence a major benefit to appropriately manage EC and to decrease the mortality associated with this disease.

Abnormal uterine bleeding (AUB) is one of the most common symptoms for gynecological consultations[306]. It affects up to 11% of postmenopausal women [307] and up to 30% of women during their reproductive years [106]. Although only 5-10% of women with AUB will have EC [76], women with this symptom are alerted to consult a specialist since AUB occurs in 90% of EC cases. This means that a large number of women with benign disorders presenting AUB need to undergo a diagnostic process to rule out EC. This process consists of a pelvic examination and a transvaginal ultrasonography followed by the histopathological examination of an endometrial biopsy, which is preferably obtained by a minimally invasive aspiration from the uterine cavity using a Cornier Pipelle (i.e., uterine aspirate or pipelle biopsy) [76,107]. Diagnosis is achieved by the observation of abnormal cells in the uterine aspirate. However, high failure rates (with an average of 22%) have been reported for this procedure due to histologically inadequate specimens [109]. In those cases, a more invasive testing, i.e., dilatation and curettage or hysteroscopy, must be performed, with the added risk of anesthesia, infection and perforation, and higher healthcare costs [308,309].

In addition, these preoperative endometrial biopsies should provide information about tumor histology and tumor grade to help in the risk stratification of EC patients and guide the surgical staging procedure. Unfortunately, the limited number of cells available for examination in these biopsies and the high inter-observer variability in the pathological interpretation results in 40-50% of discordances in EC histotype and grade between biopsies and final hysterectomy specimens [140,143], which is often associated with a suboptimal surgery treatment.

GENERAL OBJECTIVE

In this context, the main goal of this thesis is the identification of specific and sensitive protein biomarkers in uterine aspirate samples in order to: i) improve early diagnosis of EC, reducing the current failure rate associated with the diagnosis based on the

histopathological examination of the cells in uterine aspirates and hence, reducing the number of more invasive sampling methods; and ii) improve the preoperative assessment of the histological type and grade of EC tumors in order to improve the risk stratification of the patients that would help to predict the most optimal surgical treatment.

In the long-term, these approaches are expected to improve the management of EC patients and decrease the mortality and morbidity associated with this disease, while reducing the healthcare costs.

SPECIFIC OBJECTIVES

In order to achieve this main objective, the thesis has been divided in five chapters that follow the sequential phases of the biomarker identification pipeline, with the following specific objectives:

CHAPTER 1. Systematic literature review

1.1. Compile a list of proteins previously associated with EC in any type of clinical sample.

1.2. Define an appropriate study design for EC biomarker identification.

CHAPTER 2. Qualification phase: selection of the most suitable clinical sample and analytical approach for the subsequent phases

2.1. Assess if the proteins previously associated with EC, mainly in tissue samples (Chapter 1), can be consistently measured by targeted MS-based approaches in uterine aspirates; and if they maintain their potential as EC diagnostic biomarkers when measured in this biofluid more suitable for use in the clinic.

2.2. Measure the degree of correlation between fluid (supernatant) and cellular (pellet) fractions of uterine aspirates and endometrial tissue, at the protein level.

2.3. Select the most appropriate fraction of uterine aspirates (supernatant or pellet) to be used in the subsequent phases of the biomarker pipeline.

CHAPTER 3. Biomarker verification

Adapted from: "Development of a Sequential Workflow based on LC-PRM for the Verification of Endometrial Cancer Protein Biomarkers in Uterine Aspirate Samples". Martinez-Garcia et al. Oncotarget 2016.

3.1. Develop a stepwise verification workflow to prioritize from a list of potential biomarkers the most promising to enter into a further validation phase.

3.2. Evaluate the performance of PRM, a targeted acquisition method employed on a high resolution accurate mass spectrometer, in clinical samples of uterine aspirates.

3.3. Assess the potential of the fluid fraction of uterine aspirates as a source of protein EC biomarkers.

CHAPTER 4. Biomarker validation

Adapted from: "Targeted proteomics identifies proteomic signatures in liquid-biopsies of the endometrium to diagnose endometrial cancer and assist in the prediction of the optimal surgical treatment". Martinez-Garcia et al. Clinical Cancer Research 2017.

4.1. Evaluate the diagnostic performance of 52 proteins on uterine aspirate samples obtained from 116 patients covering the broad clinical heterogeneity of EC cases and benign pathologies entering the EC diagnostic process.

4.2. Assess the potential of those proteins to differentiate between histological EC types.

4.3. Define panels of proteins that achieve the best performance to diagnose EC and to discriminate between the two main EC histological subtypes.

4.4. Evaluate the correlation between MS-based results and ELISA assays.

CHAPTER 5. Moving to the clinic. Development of a prototype

5.1. Assay simplification

Partially based on: *"ELISA simplification and shortening by using a polymeric signal amplifier. Application to MMP-9 detection in plasma and uterine aspirates". de la Serna et al. (under revision).*

5.1.1. Compare the performance of a shortened ELISA (1h) *versus* the standard ELISA assay (5h) to measure the concentration of MMP9 in uterine aspirates from women undergoing EC diagnosis.

5.1.2. Evaluate the performance of electrochemical biosensors to measure the concentration of MMP9 in uterine aspirates.

5.2. Simplification of the sample preparation

5.2.1. Evaluate the feasibility of using the raw fluid fraction of uterine aspirates (i.e., avoiding albumin & IgG depletion and sonication steps) for the quantification of the described diagnostic and predictive biomarker signatures.

Results

CHAPTER 1

Systematic literature review.
Proteins associated with EC

SPECIFIC BACKGROUND

Over the last 20 years, knowledge of the molecular genetics of EC has significantly increased. Many studies have been conducted to identify molecular alterations involved in tumor development and progression [77]. With the recent advances in the proteomic field, a high number of investigations have identified proteins associated with EC. The differential levels of proteins between EC and control patients can be used as biomarkers to improve early diagnosis of EC, particularly if they can be detected in non-invasive or minimally invasive clinical samples. Nevertheless, no protein biomarker has been implemented in the EC diagnostic process yet.

For this thesis, a discovery study to generate an initial list of EC candidate biomarkers was not performed *per se*, but instead an extensive literature review was performed in order to: i) compile the numerous proteins previously associated with EC in any type of clinical sample; ii) define an appropriate study design for EC biomarker identification in order to move the most promising of these candidate biomarkers into the clinical practice, as described in Chapters 2 to 5 of this thesis.

MATERIAL AND METHODS

MEDLINE database was searched for articles published from 1978 to January 2017 using the terms “endometrial cancer” and “biomarkers” *("endometrial cancer" OR "endometrial neoplasms" OR "endometrial carcinoma" AND "biomarkers" OR "marker"*). The first inclusion criteria were: 1) EC biomarker studies performed at protein level; 2) studies describing potential EC screening or diagnostic biomarkers, i.e., differentially abundant proteins between healthy or benign controls and EC patients; 3) EC biomarker studies using any biological sample from patients: tissue, plasma, serum, etc. The exclusion criteria were: 1) non-English written articles; 2) studies performed on samples from animals or cell lines; 3) studies only comparing subsets of patients with better/worse clinical outcome (prognostic biomarkers); 4) reviews; 5) articles where only the abstract was available. Importantly, from the list of candidate biomarkers derived from these filtered studies, only those proteins that have undergone at least one level of additional validation using a different technology or biospecimen type, or an independent cohort of cases and controls whether in the context of the same publication or in an independent report are shown in the tables of this chapter.

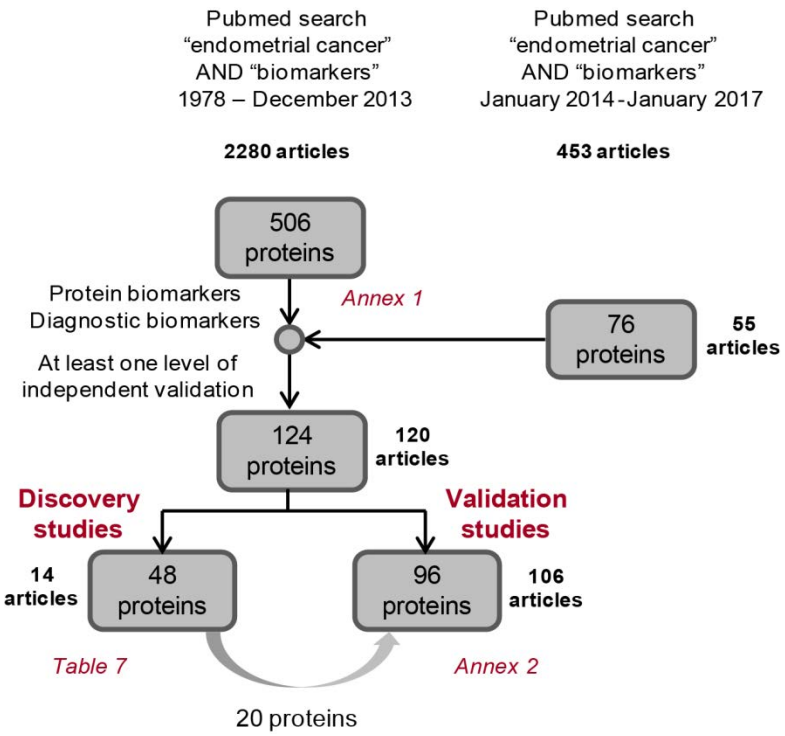


Figure 16. Flow chart summarizing the selection process of the studies and proteins associated with EC presented in this chapter.

RESULTS

Compilation of proteins associated with EC

A first search in the MEDLINE database was performed at the beginning of the thesis work and included a total of 2,280 articles published from 1978 to December 2013. As a result of this search, an initial list of 506 proteins associated with EC was obtained (Annex 1), and used as the starting point of the project described in this dissertation. Proteins in the list were ranked based on the number of articles where they were described and number of patient samples where they were evaluated.

In a second selection step, only proteins coming from studies performed at protein level and studies describing differentially abundant proteins between EC patients and non-EC controls that could be used as EC diagnostic markers were kept. In addition, the literature revision was extended to articles published from January 2014 to January 2017 (453 studies), and

55 of those articles describing potential diagnostic proteins were added to the previous selection (Figure 16).

Finally, in order to shorten the list to the most promising candidate biomarkers, and to analyze the main characteristics of the articles describing those proteins, only the candidate biomarkers validated in at least two independent cohorts of patients or by two different techniques were kept. A total of 124 proteins coming from 120 different publications matched these criteria. Fourteen out of the 120 articles were discovery studies that employed untargeted MS-based approaches for *de novo* identification of candidate biomarkers from the whole proteome of clinical samples. The other 106 articles were validation studies measuring one or few proteins in a large number of samples by targeted approaches.

❖ Discovery studies

The 14 discovery studies that followed our inclusion criteria are shown in Table 7. More than 787 differentially expressed proteins between EC and control patients have been described in these studies. According to our search, only 48 of those proteins have been further confirmed by different techniques or validated in a different set of patients and are listed in Table 7. Discovery studies are typically performed in tissue samples rather than biofluids. In our literature review, 79% of the discovery studies (11 out of the 14 studies) were described in tissue samples.

Most promising EC biomarker candidates from discovery studies

Due to the limited reproducibility of the untargeted approaches applied in discovery studies and the limited number of biological samples tested, it is important to highlight candidate biomarkers that were confirmed in several studies. In this context, 22 out of the 48 proteins (46%) were reported in more than one discovery study (in bold, in Table 7). HSPE1 (or CH10), SERPINA1, CAPG, CKB and PKM (or KPYM) were described as potential EC biomarkers in three or more articles. With the exception of HSPE1 and PKM, coming from different studies performed by the same research group, the rest of those proteins were described by independent investigators. Other two proteins, PPIA and ANXA2, may also merit further evaluation. Apart from being described as differentially expressed by two different groups using MS-based approaches, they have been validated in a large cohort of patients by tissue microarrays. The identification of biomarkers in easy-to-access biofluids is crucial for their clinical application and hence, CLU, A1BG and DJ-1 may also be interesting EC candidate biomarkers. These proteins have been reported as differentially expressed between EC and control women in both endometrial tissue and serum samples.

Table 7. Main characteristics of the discovery studies. Pre, premenopausal women; Post, postmenopausal women; Peri, perimenopausal women; EEC=endometrioid EC; EIN, endometrial intraepithelial neoplasia ; WB, western blot; IHQ, immunohistochemistry. **Panel APOA1+APOC1

Discovery phase								Validated proteins within the same publication			
Total number of differentially expressed proteins	**Study**	**Proteomic technique, quantification method**	**EC patients**	**Non-EC control patients**	**Clinical sample**	**Up-regulated proteins in EC *vs* controls**	**Down-regulated proteins in EC vs controls**	**Proteins**	**Proteomic technique (clinical sample)**	**EC patients**	**Non-EC control patients**
21 proteins (FC>2.5; P<0.05; in more than 4 of the 8 pairs)	Li et al., Mol Cell Proteomics, 2008	2DE electrophoresis	8 EEC (41-50 years old)	8 (adjacent to tumoral tissues. 41-50 years old)	Tissue	PPIA, **CAPS,** FABP5, **AGR2, DES,** TPI1, **PRDX4,** STMN1, PEBP1	**HSPB1, TF, CAPG**	PPIA	IHQ (tissue)	84 EEC (45 pre, 39 post)	29 EIN, 39 normal endometrium (all pre: 21 proliferative, 18 secretory)
19 spots increased in EC; 7 of those were identified	Byrjalsen I et al., Mol Hum Reprod., 1999	2DE electrophoresis	5 EC	11 benign controls (6 proliferative, 5 hyperplasias)	Tissue	**TF**, VIME					
	Abdul-Rahman PS et al., Electrophoresis, 2007	2DE electrophoresis	12 EEC (35-65 years old)	13 (35-65 years old)	Serum	**A1BG,** SERPINC1, **CLU**	**SERPINA1**	**A1BG,** SERPINC1, **CLU, SERPINA1**	ELISA (serum)	12 same patients	13 same patients
42 proteins (cut off: FC=1.8)	Morelli M., Int J Gynecol Cancer., 2014	2DE electrophoresis	Pool of 5 EEC (post)	Pool of 5 healthy (post, adjacent to tumor)	Tissue	**CAPG, ANXA2,** PARK7 (DJ-1)	CALR	**CAPG, ANXA2,** CALR, DJ-1	WB (tissue)	8 EEC	8 (post, adjacent to tumor)
								DJ-1	a. WB (serum) b. IHQ (tissue)	a. 15 EC. b. 23 EC	a. 20 healthy women. b. 23 healthy women
455 proteins (FC>1.3 and p-value<0.05)	Monge M, J Proteome Res., 2009	2D-DIGE	9 EEC (post from 50-80 years old)	9 (post, adjacent to tumor)	Tissue	**SERPINH1,** FSCN1	TTR, SOD1, **DES**	SOD1	WB (tissue)	5 EEC	5 (post, adjacent to tumor)
	Yang EC, J Proteome Res., 2004	MALDI-TOF (MS1 profiling)	21 EC	23: 4 proliferative, 9 secretory, 5 atrophic, 2 menstrual, 2 disordered proliferative, 1 polyp	Tissue	**HSPE1**			WB (tissue)	21 same patients	23 same patients

Discovery phase								Validated proteins within the same publication			
Total number of differentially expressed proteins	**Study**	**Proteomic technique, quantification method**	**EC patients**	**Non-EC control patients**	**Clinical sample**	**Up-regulated proteins in EC *vs* controls**	**Down-regulated proteins in EC vs controls**	**Proteins**	**Proteomic technique (clinical sample)**	**EC patients**	**Non-EC control patients**
	Guo J., Rapid Commun Mass Spectrom., 2005	MALDI-QqTOF (MS1 profiling)	8 EEC	8 (4 proliferative, 4 secretory)	Tissue	**HSPE1, S100A8**					
	Takano M., J Cancer Res Clin Oncol., 2010	SELDI TOF (MS1 profiling)- (training set)	65 EEC	40 age-matched healthy controls (not detailed)	Serum	APOC1	**APOA1**	APOC1, **APOA1****	SELDI TOF MS (serum, validation set)	40 EEC	40 age-matched healthy controls (not detailed)
11 proteins (p<0.01, welch´s test)	Negishi A., Cancer Sci., 2009	LC-ESI-QTOF (MS1 profiling)	40 EC (average age 54 +-10.6 years)	30 (15 healthy, 15 with metroptosis; average age 50 +-16.3)	Serum	**C4A**, C3	APOA4	**C4A,** C3, APOA4	WB (serum)	5 EC	5 healthy controls
9 proteins (FC>2)	DeSouza L., J Proteome Res., 2005	DDA acquisition - quantification by iTRAQ and ICAT	iTRAQ: 2 EEC; cICAT: 3 EEC	iTRAQ: 2 (1 proliferative, 1 secretory); cICAT: 1 (proliferative)	Tissue	**PKM, HSPE1, PIGR,** PEBP1, **MIF**	**SERPINA1, CKB, TAGLN**				
	Guo J., Proteomics, 2005	SELDI-QqTOF (MS1 profiling)	21 EC	23: 6 proliferative, 9 secretory, 5 atrophic, 2 menstrual, 1 polyp	Tissue	**HSPE1, S100A8**		**S100A8**	IHQ (tissue)	13 EC	11: 5 secretory, 4 proliferative, 1 atrophic, 1 menstrual, 2 atypical hyperplasia + 6 other controls
	DeSouza LV., Mol Cell Proteomics, 2007	DDA acquisition - quantification by iTRAQ	19 EC	20 (10 proliferative, 10 secretory)	Tissue	**PKM, HSPE1,** WFDC2 (HE4), **PIGR**, MUC5AC, **CLU, MIF**	**SERPINA1, CKB, TAGLN**	**PKM, HSPE1, PIGR**	IHQ (tissue)	2 EC	2 (1 proliferative and 1 secretory)
209 proteins (cut off, Q value<0,0047)	Maxwell GL, Gynecol Oncol., 2011	DDA acquisition - quantification by spectral counting	91 EC	10 (age-matched post women undergoing hysterectomy)	Tissue	COX2, **PRDX4,** PRDX3, PRDX1, **ANXA1, ANXA2**		PRDX3, **ANXA1**	WB (tissue)	12 EEC	2 (post)
								PRDX1, **ANXA2**	TMA (tissue)	139 EC	50 (26 peri + 24 post)
40 proteins (changes in expression between 40-50%)	Voisin SN., PLoS One, 2011	DDA acquisition - quantification by iTRAQ	10 EEC	10 normal proliferative endometrium	Tissue	**PKM, CAPS, CAPG, AGR2, ANXA1,** CTSB, LDHA, S100A6, HNRNPA1, CTSD	CALU, **SERPINA1, HSPB1, APOA1, TF, CKB, A1BG, CA4, SERPINH1**	CALU, CTSB, LDHA, S100A6, HNRNPA1	WB (tissue)	4 EEC	4 proliferative endometrium
								S100A6	IHQ (tissue)	5 EEC	5 proliferative endometrium

❖ Validation studies

A total of 96 proteins coming from 106 studies are shown in Annex 2 (Table 18). These studies used targeted approaches to compare the levels of one or few proteins of interest between EC and control patients. Targeted approaches can be divided in antibody-based and MS-based technologies. With the exception of 4 out of the 106 articles listed, all articles employed antibody technologies, mainly immunohistochemistry and/or ELISA. Moreover, 66 out of the 106 articles (62%) were performed on tissue samples, 24 (23%) on serum or plasma samples, 11 (10%) in both tissue and blood samples, and only 5 articles (5%) used other alternative biofluids such as uterine lavage, uterine fluid or peritoneal washings.

According to our inclusion criteria, all the 96 proteins have undergone at least one level of validation by means of a different technique, clinical sample, or different cohorts of patients. From the 96 proteins, 20 were described in the previous proteomic EC discovery studies and are here validated in larger cohorts of patients.

Most robust candidate biomarkers

Ten proteins have been described as differentially abundant between EC and control women in three or more independent validation studies (Annex 2). These proteins are HE4, CA125, SLC2A1 (GLUT1), MMP9, FOLR1, VEGFA, CD44, SAA1, TP53, and BCL2. Among them, HE4 (or WFDC2) and CA125 (or MUC16) are the two most studied biomarkers in EC. As shown in the table, their increased levels in EC have been confirmed in both tissue and serum samples and in several independent cohorts of patients. HE4 shows a considerably higher sensitivity compared to CA125 for detecting EC. The sensitivity and specificity values for these two proteins reported in the 7 articles presented in the table ranged from 31.5% to 78.8% sensitivity and 65.5% to 100% specificity for HE4; and 17.8% to 52.6% sensitivity and 33.3% to 95% specificity for CA125. All these investigations also evaluated the performance of both proteins combined. Although a slight improvement in sensitivity has been reported, the combination of CA125 and HE4 does not significantly improve the performance of HE4 alone. The limited sensitivity achieved by these proteins hamper their application in the process of EC diagnosis. Moreover, HE4 and CA125 are not specific biomarkers of EC. Serum concentrations of these two proteins are elevated in various malignancies and they have been approved by the FDA as ovarian cancer biomarkers [206].

Alternative approaches and improvements that can be applied to these studies

In spite of the high number of proteins associated with EC, none of them are currently used in the clinical practice for the diagnosis of this disease. Several reviews have highlighted some reasons for the lack of translation of biomarker research results into clinical applications [181,207,208]. In Table 8 we outline some factors that could be improved and new alternatives not yet exploited in the previously described articles that might accelerate the identification of clinically useful proteins for EC diagnosis.

In conclusion, the initial list of 506 proteins was used as a starting point of the work described in the following chapters; and the alternative approaches and recommendations indicated in Table 8 were taken into consideration for the design of the study. The benefit of these new approaches is discussed in the subsequent chapters.

Table 8. Characteristics of the described publications that could be improved and alternatives that could be exploited to improve the identification of EC diagnostic biomarkers.

	Observations	Alternatives/improvements
Preanalytical factors		
Study design	Unclear clinical question/multiple hypothesis testing	Define the clinical question to be addressed
	Small sample sizes	Sample size must be calculated to ensure adequate statistical power for each phase
	Either untargeted discovery studies or validation studies of one or few proteins. Lack of intermediate verification studies	Need to prioritize candidate biomarkers from discovery to validation with highly multiplexing technology
Patient selection	Detailed information about the patients is often absent	Inclusion/exclusion criteria for patient selection and clinicopathological characteristics of patients must be clearly defined
	The majority of studies only include premenopausal women as controls	Target population should be reflected. Postmenopausal women should represent the majority of controls (about 90% of EC cases occur after menopause).
	The majority of studies only include healthy asymptomatic women as controls	An interesting subgroup of patients to include is women with AUB (present in 90% of EC cases)
Clinical sample selection	The majority of studies are performed in tissue samples (62%), followed by the studies performed in blood samples (23%).	Use of minimally-invasive proximal fluids: uterine aspirates, vaginal secretions (only 3 out of the 106 articles reviewed)
Analytical factors		
Analytical platform	Discovery studies performed with obsolete MS-based approaches: 2DE, DIGE, SELDI-TOF..	Recent advances in MS can be applied to increase the proteome coverage (e.g., DIA acquisition)
	All but 4 out of the 106 validation studies use antibody-based technologies for the validation of one or few proteins	Recent advances in highly multiplexing targeted MS can be applied to accurately measure a high number of proteins (e.g., LC-SRM, LC-PRM)
Postanalytical factors		
Statistical analysis	Lack of robust statistics.The vast majority of publications base the statistical results only in p-values, which are not adequate to assess clinical utility and can even be misleading	Provide descriptive statistics such as fold change ratios and ROC analysis offering deeper information about the sensitivity and specificity of a candidate biomarker
	Most studies focus on single biomarkers. Only 8 of the 106 validation studies evaluate the performance of protein panels (6 of them evaluating the same combination of CA125 + HE4)	Search for biomarker signatures, evaluating their performance in independent cohorts of patients
Clinical relevance	Insufficient data showing a clear contribution to existing clinical practices	Demonstrate an improvement of the clinical scenario: improve the detection of the disease but also reduce the number of invasive biopsies and/or diagnostic costs
Implementation in the clinics		
Ease of adoption by clinical laboratories	-	Easy and straightforward assays, need of high-level assay automation
FDA approval	-	Seek FDA guidance as early as possible

CHAPTER 2

Selection of the clinical sample.
Qualification phase

SPECIFIC BACKGROUND

As shown in Chapter 1, the vast majority of EC biomarker studies used endometrial tissues and/or blood samples. Uterine aspirates (or pipelle biopsies) may be a promising alternative source of EC biomarkers. Unlike tissue samples, they are collected by minimally invasive procedures at an early stage of the current EC diagnostic process; and unlike blood, uterine aspirates are in close contact with the tumor in the endometrium and thus, may be enriched in proteins directly derived from the tumor cells. Previous studies have reported a high correlation between endometrial tissue and uterine aspirates at RNA and DNA levels [221,310] and consequently, this correlation may also be present at protein level.

From a biological point of view, proteins are key players in many cellular processes and variations of their abundance levels can be associated with pathologies such as cancer. Proteins are present in biofluids and thus may be valuable disease indicators for the development of non-invasive diagnostic tests. From a clinical perspective, proteins are more easily implemented as biomarkers in the clinics than genomic biomarkers, as they can be detected and quantified by techniques that are widely implemented in hospitals such as immunohistochemistry or ELISA. To date, only few proteomic studies have been performed on uterine aspirates and they were not focused on the search for EC biomarkers [222–224].

The work presented in this chapter aimed to i) assess if proteins previously associated with EC, mainly in tissue samples (Chapter 1), can be consistently measured by targeted MS-based approaches in uterine aspirates; and if they maintain their potential as EC biomarkers when measured in this biofluid; ii) measure the correlation at the protein level between the fluid (supernatant) and cellular (pellet) fractions of uterine aspirates and their corresponding endometrial tissue; and iii) select the most appropriate fraction of uterine aspirates (supernatant or pellet) to be used in the subsequent phases of the biomarker pipeline.

MATERIAL AND METHODS

Patient recruitment and sample collection

A total of 16 women entering the EC diagnostic process (8 women suffering from EC and 8 non-EC controls) were recruited in the Vall d'Hebron University Hospital (Barcelona, Spain) between 2012 and 2014. Informed consent forms, approved by the Vall d´Hebron Ethical Committee, were signed by all patients (approval number: PR_AMI_50-2012). Inclusion criteria were postmenopausal women with atrophic endometrium and a minimum age of 50 years. Women who had been treated previously for gynecological pelvic cancer were excluded. Patients known to be positive for the human immunodeficiency virus and/or the hepatitis virus were excluded for safety reasons.

From each woman, a uterine aspirate sample was collected by aspiration with a Cornier Pipelle (Eurogine Ref. 03040200) in the office of the clinician or in the operating room prior to surgery and transferred to 1.5 ml microtubes. Phosphate buffer saline (PBS) 1X was added in a 1/1 (v/v) ratio and centrifuged at 2,500 rcf for 20 min in order to separate the fluid fraction (supernatant) from the cellular fraction (pellet). The separated fractions were kept at -80°C until use. Apart from the uterine aspirate sample, tumor and/or normal endometrial tissue samples from all patients were provided by the pathologist after hysterectomy and kept frozen at -80ºC until use. Additionally, 7 mL of whole blood from two patients were collected in ethylenediaminetetraacetic acid (EDTA) coated tubes, and centrifuged at 1,500 rcf for 15 min to separate the plasma from the cellular fraction; both fractions were kept at -80ºC.

From the 16 women enrolled in the study, the uterine aspirate, blood and tissue samples from one EC patient and one control were used for the shotgun proteomic analysis and for the development of the LC-SRM method. Uterine aspirates and tissue samples from the remaining seven EC and seven non-EC controls were employed for the LC-SRM analysis of 54 preselected potential biomarkers. The clinical and pathological characteristics of these 14 patients are summarized in Table 9.

Table 9. Clinical characteristics of the women enrolled in the LC-SRM study. EEC, endometrioid endometrial cancer.

	EEC (n=7)	Non-EC control (n=7)
Age (years)		
Average	73	73
Minimum	56	53
Maximum	82	85
Uterine condition		
Premenopausal	-	-
Postmenopausal	7	7
Histologic grade		
Grade 1	1	
Grade 2	4	
Grade 3	2	
FIGO stage		
IA	4	
IB	3	
II-IVB	-	
Myometrial invasion		
<50%	4	
>50%	3	

Protein identification by shotgun LC-MS analysis (data dependent acquisition)

❖ Sample preparation

Supernatants from uterine aspirate and blood samples were sonicated (Labsonic M, Sartorius Stedim Biotech) at 100% amplitude during 8 cycles of 15 seconds, and 50 µl of each sample was depleted from albumin and IgG using the Albumin & IgG depletion spin trap kit (GE Healthcare) according to the manufacturer's instructions. Pellets from uterine aspirate and blood samples, as well as endometrial tissue samples, were washed 5 times with PBS1X and 300 µl of a lysis buffer containing 3.5 M urea, 1.0 M thiourea, 0.1% dithiothreitol (DTT), 5% dimethyl sulfoxide (DMSO), 10% glycerol, 2% octyl-beta-glucoside and 1% protease inhibitor cocktail (Sigma, catalog number P8340) were added after the last wash. Samples were then sonicated at 100% amplitude during 8 cycles of 15 seconds and centrifuged at 20,800 rcf for 4 min in order to separate the soluble fraction containing the extracted proteins. Total protein concentration of each sample was measured by the Bradford assay (Sigma-Aldrich), performed in triplicate. All samples were divided in two aliquots containing 7 µg of total protein to perform the proteolysis and LC-MS analysis in duplicate. Total protein was precipitated by the addition of 8 volumes of cold acetone followed by an overnight incubation at -20ºC. Afterwards, samples were centrifuged at 16,000 rcf for 15 min and the acetone was removed. Dried pellets were resuspended by sequential addition of 40 µl of 0.2% Rapigest (Waters) and 100 µl of 8 M urea, followed by protein digestion with Lys-C endoproteinase MS grade (Thermo Scientific) at a protease/total protein amount ratio of 1/150 (w/w) at 37ºC overnight. The samples were then reduced with 3 µl of 100 mM DTT for 60 min at 37°C, and alkylated with 3 µl of 200 mM iodoacetamide (IAA) at 22°C for 30 min in the dark. The concentration of urea was diluted to 1 M with 50 mM ammonium bicarbonate buffer, and samples were incubated overnight at 37°C with trypsin (Promega) at a protease/total protein amount ratio of 1/50 (w/w). The protease activity was inhibited by addition of 1 µl of neat formic acid per 100 µl of solution. Digests were desalted with solid phase extraction (SPE) C18 cartridges (Waters), vacuum dried and suspended in 0.1% formic acid to have a final concentration of 1 µg/µl.

❖ LC-MS analysis

For the proteomic profiling of both supernatant and pellet fractions of uterine aspirates and blood, as well as endometrial tissue sample, digests were analyzed using an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific) operated in data dependent acquisition (DDA) mode. The liquid chromatography system consisted of an Ultimate 3000 RSLC nano configured in binary gradient mode. The setup was operated in column switching mode and samples were loaded onto a trap column (Acclaim PepMap100 2 cm × 75 µm i.d., C18, 3

µm, 100 Å) for 3 min at 5 µl/min by an aqueous solution containing 0.05% trifluoroacetic acid and 1% (v/v) acetonitrile. Peptides were then eluted onto an analytical column (Acclaim PepMap RSLC 15 cm × 75 µm i.d., C18, 2 µm, 100 Å) by applying a 66 min linear gradient from 2 to 35% of solvent B in solvent A at 300 nl/min. The solvents A and B consisted of water with 0.1% (v/v) formic acid and acetonitrile with 0.1% (v/v) formic acid, respectively. The electrospray ionization was performed through a fused silica emitter by applying a voltage of 1.5 kV. The DDA method was based on a high resolution survey scan performed in the orbitrap (60,000 at 400 m/z) followed by the fragmentation and analysis of the 6 most intense precursor ions in the LTQ ion trap at a normalized collision energy of 35. Dynamic exclusion of precursors already selected for MS/MS experiments was set to 90 s.

❖ **Protein identification**

The identification of peptides and related proteins was performed by Mascot search engine, using the Proteome Discoverer software (v1.4) (Thermo Scientific, Waltham, MA, USA). The Swiss-Prot human database (SwissProt 201108 with 531473 sequences entries, restricted to the 20,245 entries of the human taxonomy) was used. Trypsin specificity was set to cleave after arginine and lysine residues excepted when flanked by a proline on the C-terminal side. A fragment ion mass tolerance of 0.8 Da and a precursor mass tolerance of 10 ppm were applied. Up to one tryptic missed cleavage was tolerated. Carbamidomethylation of cysteines and oxidation of methionines were specified as static and dynamic modifications, respectively. Results were filtered by Proteome Discoverer using one peptide per protein, a maximum search engine rank of 1 and a false discovery rate (FDR) below 0.01 (calculated by the node “Target decoy PSM validator”). To set FDR at 0.01, the expectation value for accepting a spectrum was below $4*10^{-3}$.

Super-SILAC mix preparation

❖ **Cell culture and protein extraction**

Three endometrial cancer cell lines (RL95, KLE, and AN3CA) were grown in SILAC DMEM/F12 medium deprived of its natural lysine and arginine and supplemented with isotopically labeled arginine ($^{13}C^{15}N$-Arg) and lysine ($^{13}C^{15}N$-Lys), proline, dialyzed fetal bovine serum (FBS) and 1% penicillin-streptomycin. Labeled aminoacids and proline were purchased from Silantes GmbH, Germany. Cells were cultured for 10 doublings to ensure complete protein labeling. The incorporation of heavy amino acids into the protein sequences was verified by LC-MS in an LTQ-Orbitrap Velos. When the labeling rate was higher than 95%, the medium was replaced by serum-free medium for 48 h to reduce the serum protein background, and cells were harvested, pelleted by centrifuging at 2,500 rcf for 5 minutes and resuspended in a lysis buffer containing 3 M urea, 1 M thiourea, 0.1% DTT,

5% DMSO, 10% glycerol, 2% octyl beta glucoside, and 1% protease inhibitor cocktail. Total protein from cells was extracted by sonication during 8 cycles of 15 seconds and centrifuged at 20,800 rcf for 4 min to separate the soluble fraction containing the extracted proteins. In addition, the conditioned medium of the three cell lines was collected and secreted proteins in the medium were precipitated by the addition of 5 volumes of cold acetone followed by an overnight incubation at -20ºC and 20 min centrifugation at 15,000 rcf. Dried pellets were re-suspended in 8 M urea buffer. Bradford assays were performed to quantify the protein content of each cell lysate and the pyrogallol red method (Sigma-Aldrich) was used for protein quantification of the secretomes (since an 8 M concentration of urea is not compatible with the Bradford reagent).

❖ **Super-SILAC mix preparation**

The super-SILAC approach consists in combining an assortment of cell lines labeled by SILAC, a super-SILAC mix, to be used as a spike-in standard in the clinical samples [311]. Several optimizations are necessary taking into account the following considerations:

a) The cell lines should differ as much as possible from one another and, at the same time, they should represent as much as possible the proteome of the clinical samples [247]. Therefore, the proteomes of the three cell lines and their medium were compared between them and also with the proteomes of the uterine aspirate and endometrial tissue samples to evaluate the overlap rate. A super-SILAC mix consisting of KLE medium, RL95 cells, RL95 medium in a 1/1/1 ratio (w/w/w total protein) was selected because it was the simplest mix that allowed for the identification of the highest number of proteins of interest.

b) The ratio of the super-SILAC mix to be spiked in the clinical samples must be established in order to favor the proper detection of both endogenous and heavy versions of the targeted peptides [247]. Three different ratios of clinical sample-SILAC mix (1/2, 1/1 and 2/1) were evaluated, and the 1/1 ratio (w/w total protein) was finally selected as the optimal ratio.

LC-SRM analysis

❖ **Sample preparation**

Seven patients suffering from EC and seven non-EC women with normal atrophic endometrium were included in this study. From each woman, a sample of uterine aspirate and endometrial tissue was collected. Total protein from these samples was extracted as performed for the shotgun LC-MS analysis (page 81), and 50 µg of protein from each clinical sample were mixed with 50ug of total protein of the super-SILAC mix. Each sample was split in two aliquots to perform the proteolysis in duplicate. First, samples were denatured by the addition of urea to a final concentration of 6 M. Proteins were then reduced in 5 mM DTT for 1 hour at room temperature, alkylated in 15mM IAA for 30 min, and digested first with LysC

protease at a protease/total protein amount ratio of 1/50 (w/w) for 4h at 37ºC, and then with trypsin overnight at 37°C at an enzyme/substrate ratio of 1/50 (w/w) after dilution of urea to a final concentration of 1 M. Proteolysis was stopped by addition of 1 µl of neat formic acid per 100 µl of solution, and the digests were desalted onto solid phase extraction cartridges. The eluates were subsequently evaporated to dryness in a vacuum centrifuge and suspended in 0.1% formic acid before LC-SRM analysis.

❖ **SRM method development**

We selected 54 proteins out of the initial 506 candidate biomarkers from Chapter 1 to be targeted in this study. The selection of these proteins was based on their detection in the shotgun proteomic analysis of uterine aspirate samples and their biological interest and relevance in literature. For those 54 proteins, we selected the two most intense proteotypic peptides per protein (peptides with an amino acid sequence uniquely associated with the protein of interest in a given proteome), with the exception of few proteins targeted with only one peptide, ending up with a total of 85 peptides. The three best transitions per peptide were then selected based on the results of the DDA analysis.

❖ **LC-SRM setup**

The LC-MS setup consisted of a Dionex Ultimate 3000 RSLC chromatography system coupled on-line to a triple quadrupole mass spectrometer (TSQ Vantage, Thermo Scientific). A total of 600 ng per sample (clinical sample/SILAC mix) were injected. Additionally, 600 ng of the SILAC mix alone were also injected as control. The LC system was configured for a high-pressure binary gradient and operated in column switching mode. The mobile phase A consisted of 0.1% formic acid in water, the phase B in 0.1% formic acid in acetonitrile and the loading phase in 0.05% trifluoroacetic acid and 1% acetonitrile in water. Each digested sample was injected and loaded onto a trap column (75 µm × 2 cm, C18 PepMap 100, 3µm) at 5 µl/min and further eluted onto the analytical column (75 µm × 15 cm, C18 PepMap 100, 2µm) at 300 nl/min by a linear gradient starting from 2 % B (98% A) to 35 % B (65% A) in 48 min. The MS analysis was performed by the triple quadrupole mass spectrometer operated in SRM mode and the list of transitions was split in two methods. The MS cycle time was set to 2.5 seconds and the first and second quadrupoles were operated at unit resolution (0.7 FWHM). The duration of the time scheduled windows for each pair of endogenous and isotopically labeled peptides was set to 2 min. The collision energy (CE) to be applied to the light precursors was calculated with the following formula: z=2, CE=0.03*m/z+2.905; and z=3, CE=0.038*m/z+2.281. The collision energy applied to the heavy precursors was the same as the one applied to their light counterparts.

❖ **SRM data processing**

Raw data from the LC-SRM analysis were imported into the Skyline program (v3.1) (McCoss Lab, University of Washington, USA) for peptide quantification. All peak boundaries were manually reviewed, checking that the retention times and the relative SRM peak intensity ratios across the transitions were the same between the endogenous peptides and their corresponding heavy peptides, and were reassigned if needed. Information including peak area and area ratio of light/heavy peptide pairs was exported for further analysis. The averaged area ratios and the coefficient of variation (CV% = (standard deviation / mean) × 100) were calculated between duplicates.

For each clinical sample, peptides were considered "detected" if the signal of each endogenous peptide had a signal-to-noise ratio higher than 3. Noise was calculated by integration of the SRM traces in the light channels within the same retention time boundaries of the respective heavy peptides when the SILAC mix alone was analyzed. Peptides detected in less than 50% of the samples in any of the groups (uterine aspirate supernatants, uterine aspirate pellets, endometrial tissues) were excluded. The linear correlation between the peptides of the same protein was evaluated using the Pearson correlation coefficient.

❖ **Statistical analysis of the LC-SRM study**

The statistical analysis of this study was performed following five different workflows in order to evaluate the robustness of the results. A straightforward analysis (herein, the M1 workflow) was performed with the SPSS software (v20.0) (IBM, Armonk, NY, USA). The comparison of the concentration levels of the monitored peptides between tumor and control samples was calculated using the nonparametric Mann-Whitney U test on the raw light/heavy ratios (without any log transformation or further normalization methods). Whenever the heavy-labeled peptide was not detected, the signal of the endogenous peptide alone was used.

Workflows M2 to M5 included a more sophisticated statistical analysis. They were performed using the R Bioconductor packages MSstats and SparseQuant in collaboration with Álex Campos from Sanford-Burnham Medical Research Institute, La Jolla, California (USA). These workflows differed among themselves in the statistical modeling approaches used to compare relative protein differences between groups (MSstats or SparseQuant) [312], in the normalization methods, and in the imputation of missing values for heavy peptides. The normalization of the SRM data set was evaluated using the R Bioconductor package Normalyzer which deploys eleven different normalization approaches. Global loess normalization (herein, Loess G), which is based on a local regression, was selected as it performed slightly better than the other ten normalization approaches evaluated here.

Overall, Loess-G reduced variation within groups and better separated the six groups of samples studied (i.e., normal and tumor groups of supernatants and pellets of uterine aspirates and tissue samples), as evaluated by Principal Component Analysis. The analysis workflows employed here are summarized as follows:

- Workflow 1 (M1): raw data > SPSS (Mann Whitney test)
- Workflow 2 (M2): No normalization (log2 data) > MSstats
- Workflow 3 (M3): Normalization (Loess-G) > MSstats
- Workflow 4 (M4): Normalization (Loess-G) > Imputation > SparseQuant
- Workflow 5 (M5): Normalization (Loess-G) > Imputation > MSstats

The correlation between every statistical method and clinical sample (fractions of uterine aspirates and endometrial tissues) was calculated pairwise using R's *cor* function. A correlation plot was generated with the pairs function from R gplots package.

For the comparison of the protein levels between EC and non-EC controls with all the methods, p-values lower than 0.05 along with fold changes greater than two were considered statistically significant.

RESULTS

Proteomic profiling of uterine aspirate samples, endometrial tissue and blood

Uterine aspirate samples present a significant blood contamination and hence, we evaluated if high-abundance blood proteins would hamper the detection of proteins coming from the endometrial tissue. The proteome of the raw blood, the endometrial tissue and the uterine aspirate of one control and one EC patient were mapped by LC-MS in DDA mode. In addition, the LC-MS detection of the 506 potential biomarkers coming from the literature review described in Chapter 1 was evaluated.

A higher number of proteins were identified in uterine aspirates compared to blood and even tissue. For the samples collected from the EC patient, a total of 1,089 proteins were detected in the uterine aspirate sample, 470 proteins were detected in the endometrial tissue, and 277 proteins were detected in the blood sample (Figure 17A). A total of 140 out of the 506 candidate biomarkers were identified in the uterine aspirate sample from the EC patient, compared to the 94 candidates detected in the tissue sample, or the 52 detected in the blood sample. Importantly, 90 out of the 94 candidates detected in endometrial tissue (96%) could be detected in any of the fractions of the uterine aspirates (Figure 17B), thus confirming the potential of this biofluid as a surrogate of the endometrial tissue for the evaluation of EC candidate biomarkers by LC-MS.

A

Protein identification by DDA				
	Total number of proteins detected		Number of candidates detected out of the 506	
SN uterine aspirate	793	1,089	121	140
Pellet uterine aspirate	573		83	
Endometrial tissue	470		94	
SN blood	191	277	37	52
Pellet blood	114		30	

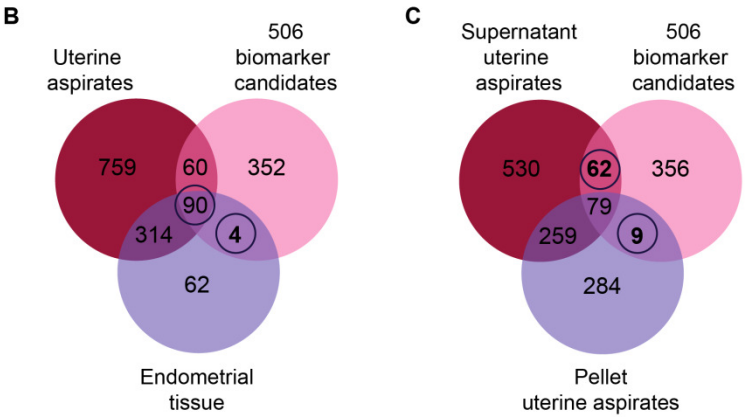


Figure 17. Proteomic profiling of uterine aspirate samples, endometrial tissue and blood.A. Protein identification in the different sample types of the same EC patient. **B.** Detection of the 506 candidate biomarkers in uterine aspirates and tissue samples**. C.** Detection of the 506 candidate biomarkers in the fractions of the uterine aspirate samples.SN, supernatant.

We then evaluated which fraction (supernatant or pellet fraction) of the uterine aspirates allowed for the identification of more proteins of interest. A total of 141 out of the 506 candidate biomarkers were detected in the supernatants of the two patients (121 in the sample derived from the EC patient and 117 in the sample from the control woman), whereas 88 candidate biomarkers were identified in the pellets (83 in the sample from the EC patient and 58 in the sample from the control woman). As shown in the Venn diagram (Figure 17C), 62 candidate biomarkers were exclusively found in the soluble fraction, while only nine were specific to the pellet fraction.

Evaluation of 54 EC candidate biomarkers in uterine aspirate and tissue samples by LC-SRM

Since further steps of the biomarker pipeline will be performed by targeted MS-based approaches, we evaluated if the candidate biomarkers previously detected by untargeted MS could be systematically quantified by LC-SRM. Moreover, we evaluated whether those candidate biomarkers described at the tissue level maintained their potential to discriminate between EC and control patients when measured in uterine aspirate samples.

A total of 54 EC candidate biomarkers (85 peptides) from the proteins detected by DDA analysis in uterine aspirates were selected based on their relevance in the literature and biological importance. The levels of these proteins in the uterine aspirate and endometrial tissue samples from seven women suffering from EC and seven non-EC controls were analyzed by LC-SRM using a super-SILAC approach as internal standard (Figure 18). We set an optimal labeled cell mixture consisting of KLE medium, RL95 cells, RL95 medium in a 1/1/1 ratio (w/w/w total protein) and added it to each clinical sample in a 1/1 ratio (w/w total protein) prior proteolysis and LC-MS analysis (data not shown).

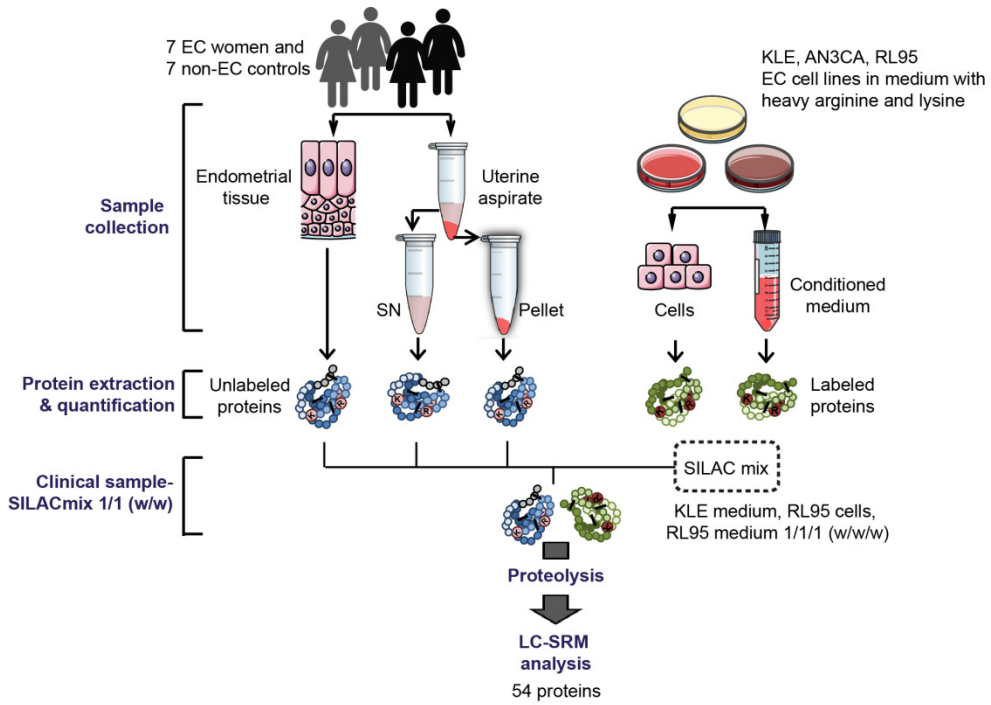


Figure 18. Workflow using the super-SILAC approach.

❖ Quality control of the SRM results

MS data was processed with the Skyline software and chromatographic peak integration was manually curated. For the 54 targeted proteins, a high incorporation rate of heavy amino acids in the SILAC mix was confirmed. The super-SILAC mixture alone was analyzed and the contribution to the light signal was below 5% for 78% of the peptides and below 10% for 95% of them, with an average contribution to the light signal of 4.9%. As illustrated in Figure 19A for the peptide LQDG[...]LAQR, the contribution of the heavy peptide to the light signal can appear artificially high when the intensity of the heavy peptide is also low, and this was the case for most of peptides showing a contribution to the light signal higher than 10%.

For final quantification, endogenous peptides below or very close to the limit of detection (signal to noise ratio lower than 3) in more than 50% of the samples in a group were excluded. In total, 74% of the targeted proteins (63 peptides corresponding to 41 proteins) were consistently detected in all the three sample groups: 47 proteins were detected in supernatants of uterine aspirates, 45 in pellets of uterine aspirates, and 42 in endometrial tissue samples. Regarding the heavy peptides, 6 proteins were not detected in the SILAC mix (MMP9, CAYP1, FABP5, PERM, PIGR and AL1A1).

The mean between duplicates in the curated dataset was calculated. The CV% between technical duplicates was below 15% for 90%, 94% and 92% of the quantified peptides in the supernatants, pellets, and tissue samples, respectively; with an average CV of 7% in the supernatant samples and 6% in pellets and tissues (Figure 19B).

Next, the correlation between the peptides derived from the same protein was evaluated by the Pearson correlation coefficient (R coefficient). In uterine aspirates and tissue samples, the correlation between peptides was high, with R coefficient over 0.9 for 23 out of 25 proteins (92%) monitored with two peptides in the aspirates and 20 out of 22 proteins (91%) in the tissues. The R coefficients below 0.9 obtained for two of the proteins were due to isoform-specific peptides. However, peptides from the same proteins in the pellet fraction of uterine aspirates showed a lower correlation, with only 14 out of the 24 proteins (58%) monitored with two peptides presenting R coefficients over 0.9 (Figure 19C). This demonstrates that more variable results were obtained in this pellet fraction of uterine aspirates.

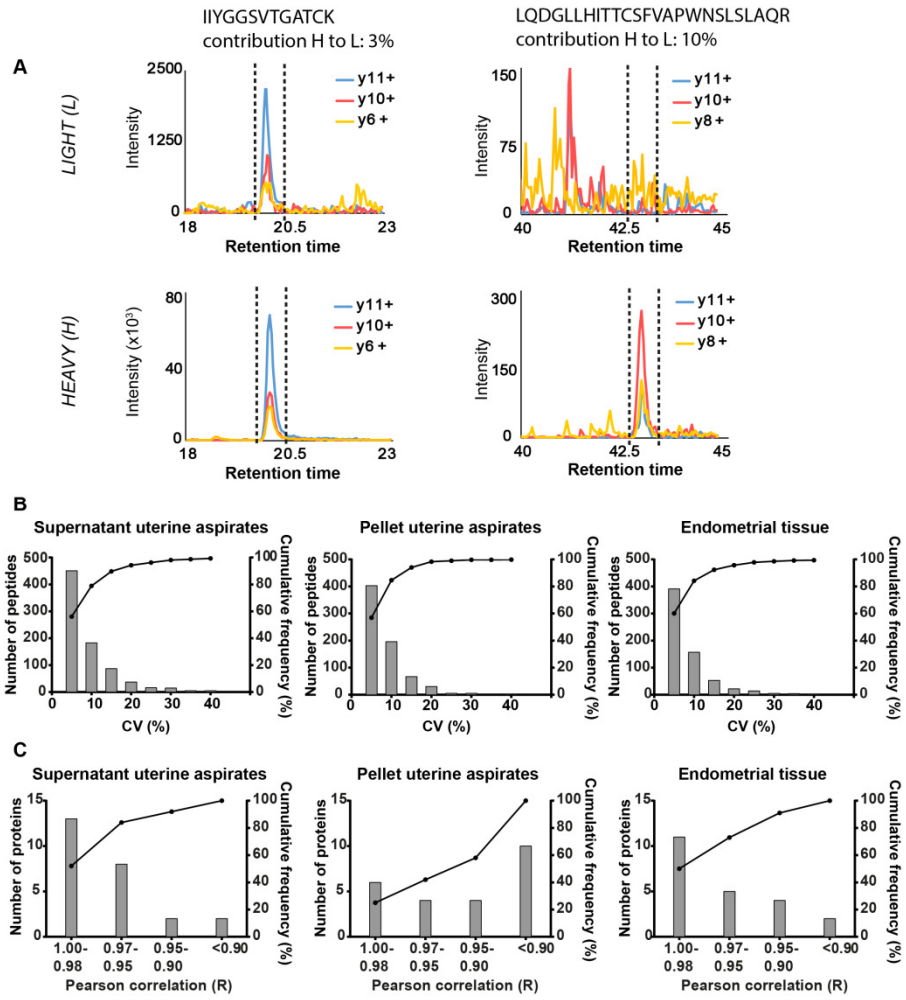


Figure 19. Data quality control of the LC-SRM analysis. A. Illustration of the LC-SRM traces of two peptides measured in a pure SILAC mixture tryptic digest. **B.** Distribution of the coefficient of variation (CV%) of the light/heavy peptide area ratios between duplicates of each supernatant of uterine aspirates, pellet of uterine aspirates and endometrial tissue samples. **C.** Pearson correlation between signature peptides coming from the same protein in supernatants and pellets from uterine aspirates and endometrial tissue samples.

❖ Correlation of the levels of EC protein candidate biomarkers between endometrial tissue and uterine aspirates

In order to evaluate the correlation between uterine aspirates and their corresponding endometrial tissue at protein level, the correlation of fold changes between EC and non-EC samples determined by different statistical methods (M2 to M5) for all 54 proteins in the

different samples, i.e., supernatants, pellets and tissues, was plotted (Figure 20). Overall, the four workflows generated very similar results for the three types of samples, showing almost perfect correlation for most pairwise comparisons (coefficients of correlation over 0.88). Nevertheless, the correlation across the three sample types was relatively poor. The coefficients of correlation calculated between supernatant/pellet, supernatant/tissue and pellet/tissue were very low for all the statistical analysis workflows, ranging from 0.24 to 0.41.

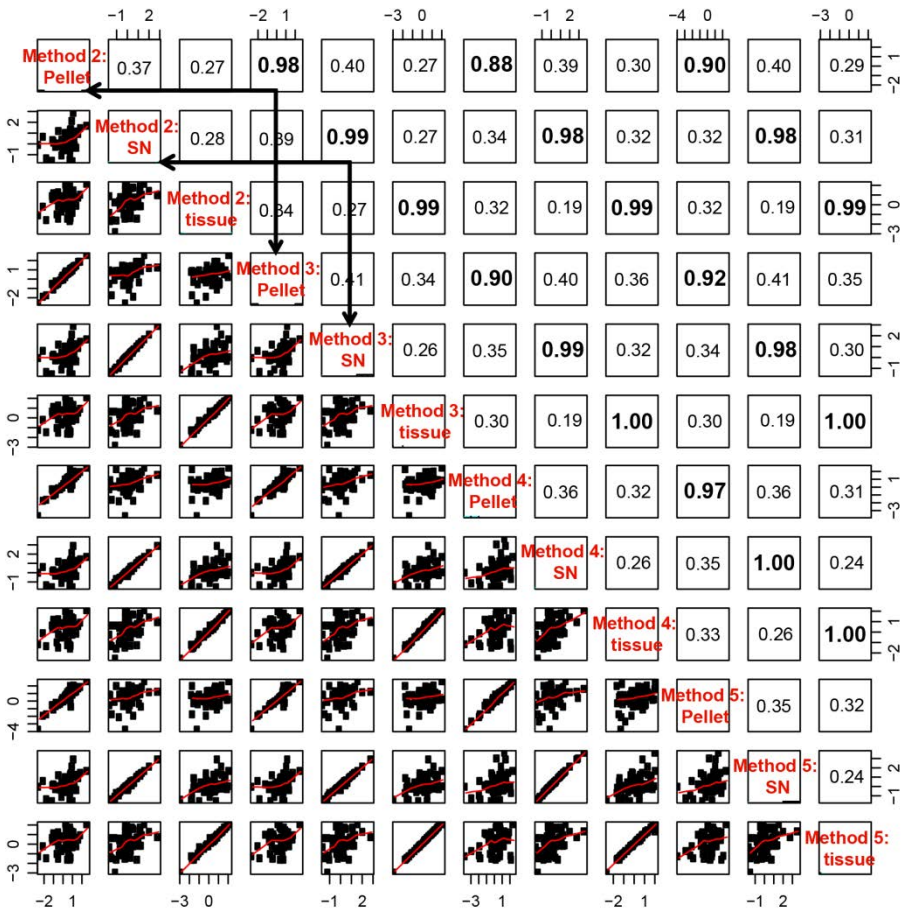


Figure 20. Correlation between uterine aspirates (supernatant and pellet fractions) and endometrial tissue samples at protein level. Correlation plots between fold changes of all 54 targeted proteins in the three sample types with the four different statistical methods using MSstats or SparseQuant. The coefficients of correlation obtained when comparing methods (in bold) and sample types are shown. SN, supernatant of uterine aspirates.

❖ Differentially abundant proteins between tumor and control patients

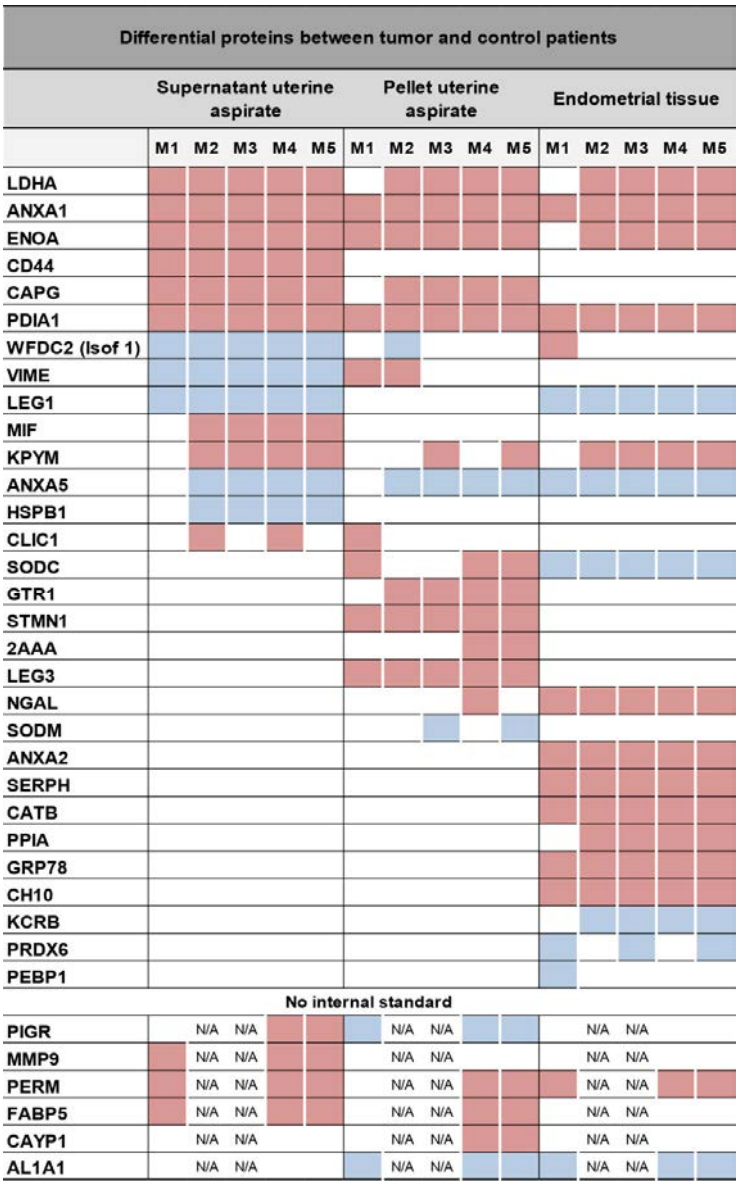
Despite the low correlation observed between uterine aspirates and endometrial tissue samples, we next evaluated if differentially abundant proteins between tumor and control patients at tissue level were also differentially expressed in the uterine aspirate samples. In order to evaluate the robustness of the results, the five different statistical analysis workflows were followed. From the 54 targeted proteins, 18, 22 and 21 proteins showed significant differences between tumor and control patients in supernatants and pellets of uterine aspirates and endometrial tissue, respectively (p-value <0.05; fold change higher than 2) (Table 10). Twelve out of the 21 differential proteins in tissue (57%) were also found differentially abundant in any of the fractions of the uterine aspirate samples. Nine of these proteins appeared differentially expressed in supernatants, pellets and tissue samples. These proteins are LDHA, ANXA1, ENOA, PDIA1, WFDC2 isoform1, KPYM, ANXA5, PERM and AL1A1.

Regarding the different statistical analysis workflows, all the four methods performed using MSstats or SparseQuant gave almost identical results regarding the differential proteins from supernatants and tissues. More variable results were obtained in the pellet fraction of uterine aspirates, probably due to the heterogeneity of the biological sample itself (Table 10). In addition, the results obtained with a more straightforward analysis using the raw data were also very similar and hence, this statistical approach was used in the next chapters. Although a lower number of proteins were found differentially abundant with this approach using the same threshold (p-value <0.05; fold change higher than 2), all but three differential proteins are in common with the results obtained using the other approaches in the three sample types. These results demonstrate the robustness of the differentially abundant proteins identified.

DISCUSSION

The "qualification phase" of the biomarker pipeline is used to assess whether the differential levels of candidate biomarkers observed in a previous discovery phase can be confirmed by MS-based targeted proteomic approaches that will be used in the next steps of the biomarker pipeline. It is also used to evaluate biological samples more suitable for a future clinical application, such as non-invasive or minimally invasive biofluids. Most of the proteins in the initial list of candidate biomarkers described in Chapter 1 have been studied in tissue samples, mainly by untargeted MS approaches or immunohistochemistry. In this chapter, we demonstrated the feasibility of measuring these proteins in uterine aspirate samples by highly multiplexing targeted proteomics.

Table 10. List of differentially abundant proteins in uterine aspirates (supernatant and pellet fractions) and endometrial tissues from 7 EC vs 7 non-EC patients (p-value<0.05, fold change >2). The differential proteins obtained with the five different statistical methods (M1-M5) are shown. For those peptides without internal standard available, the values of the endogenous peptides without normalization by the internal standard were used in method 1, whereas imputation techniques were applied to replace the missing values of the heavy peptides in methods 4 and 5. In pink, proteins with higher levels in EC samples. In blue, proteins with lower levels in EC samples. N/A, not available.



Uterine aspirate (or pipelle biopsy) is the endometrial fluid in direct contact with the endometrium that is collected by aspiration in the early steps of the process of EC diagnosis using minimally invasive methods, preferably a Cornier Pipelle [108]. The uterine aspirate is a complex biofluid composed of a cellular fraction mainly containing cells flaked off the endometrium and blood cells, and a fluid fraction or supernatant mainly containing proteins secreted by the endometrial cells and intracellular proteins coming from necrotic cells. Several investigations have demonstrated that the comprehensive study of the proteins contained in this sample can provide insights into embryo implantation and infertility and can be a promising source of biomarkers for endometrial alterations such as endometriosis [223,224]. However, no proteomic studies have focused on the identification of EC protein biomarkers in this type of clinical sample.

Regarding EC research, previous studies have demonstrated a high degree of correlation between the cellular fraction of uterine aspirates and their corresponding primary tumors in the endometrium at RNA and DNA levels [221,310]. We evaluated if this correlation was also maintained at protein level by measuring the relative abundance of 54 EC candidate biomarkers by LC-SRM. Our results showed a low correlation between the endometrial tissue and both fluid and cellular fractions of the uterine aspirates. Nevertheless, 12 out of the 21 proteins found differentially abundant in endometrial tissue samples between tumor and control patients were also differentially expressed in uterine aspirates. These results pave the way to the identification of EC biomarkers in this biofluid.

Another important objective of this study was the selection of the most appropriate fraction of uterine aspirates (i.e., supernatant or pellet) to be used in the subsequent steps of verification and validation of the candidate biomarkers by targeted proteomics. Since a clear correlation with the corresponding endometrial tissue was not observed for any of the fractions, the selection was based on the biological characteristics, LC-MS behavior, and clinical suitability of each fraction based on the results obtained in this chapter (Table 11). From a biological point of view, the supernatant fraction is easier to process since cell lysis is not required and the depletion of highly abundant plasma proteins is a well standardized process. From an analytical point of view, the shotgun LC-MS analysis of two uterine aspirate samples showed that a higher number of proteins could be detected in the supernatant fraction compared to the pellets, as well as a higher number of our initial list of EC candidate biomarkers, thus increasing the chances of finding a clinically useful protein biomarker signature in this fraction. The measurement of 54 of those proteins by LC-SRM in uterine aspirates samples from 14 women (7 EC patients and 7 controls) reported the differential expression between tumor and control samples of a higher number of proteins (22 proteins) in the pellet fraction compared to supernatants (18 proteins). However, the results obtained in the supernatant fraction were significantly more robust, since 67% of the

differential proteins in this fraction, but only 32% of the 22 proteins in pellets, were confirmed using the five different statistical analysis methods. Moreover, the correlation between peptides from the same protein was also higher in the supernatant fraction. From a clinical point of view, soluble proteins in biofluids can be directly quantified by immunoassays such as ELISA, widely implemented in clinical laboratories, with minimal sample handling. Importantly, the assessment of biomarkers in the fluid fraction of the biopsy, independent of the cellularity of the sample, could help to overcome the 22% of undiagnosed patients due to insufficient cells in the samples of the current diagnostic procedure [109]. In conclusion, the fluid fraction, or supernatant, of uterine aspirates was selected as the most convenient sample for EC protein biomarker identification.

Regarding the technical part, in this chapter we used the SRM acquisition method to quantify a set of 54 proteins. Our approach presented two limitations. On one side, selectivity and sensitivity of the SRM method can be limited when applied to complex biological samples. Indeed, we could consistently quantify 63 peptides (corresponding to 41 proteins) over the initial list of 85. On the other side, a super-SILAC approach was used for internal standardization since several EC cell lines metabolically labeled were accessible in the laboratory at that moment. However, the use of super-SILAC reduces the sensitivity of the approach, since the addition of a whole extra proteome from the cell lines increases the complexity of the clinical samples [311].

Consequently, in the next chapters we used a novel targeted MS acquisition method that takes advantage of high resolution accurate mass spectrometry, i.e., the PRM acquisition. Moreover, we used isotope-labeled synthetic peptides as internal standards, as they allow for a better control of the internal standard purity and more importantly, they introduce less complexity to the sample compared to the super-SILAC approach. LC-PRM acquisition was developed by our collaborators in this project, Prof. Domon´s team at Luxembourg Clinical Proteomics center, at this moment of the thesis [259,298]. Along several internships, I had the opportunity to apply this leading-edge technology to the next phases of the study.

Table 11. Comparison between the two fractions of uterine aspirates (the fluid fraction or supernatant, and the cellular fraction or pellet) regarding their biological, analytical and clinical characteristics.

		Supernatant uterine aspirates	Pellets uterine aspirates
Biological characteristics	**Sample preparation**	Easy processing. Do not require cell lysis	Need of a lysis buffer compatible with MS and sonication
	Sample complexity	Heterogenous among patients but more homogenous within each sample	Heterogenous among patients and within each sample
	Abundant proteins	Already optimized albumin and IgGs depletion	Unstandardized method of hemoglobin depletion
Detectability by DDA in two uterine aspirates	**Total number of proteins**	930	631
	Number of candidate biomarkers (out of the initial 506)	141 62 proteins exclusive of SN	88 Only 9 proteins exclusive of the pellet fraction
Results from LC-SRM study in 14 samples (7T vs 7N)	**Reproducibility of the process**	90% of the duplicates with a CV<15%	94% of the duplicates with a CV<15%
	Correlation between peptides from the same protein	**R>0.93** in 23/25 (92%) proteins monitored with two peptides	**R>0.9** in 14/24 (58%) proteins monitored with two peptides
	Number of differential proteins (T vs N)	18	22
	Consistency of the statistical analysis methods	67% of differential proteins between T and N obtained with the 5 statistical methods	32% of differential proteins between T and N obtained with the 5 statistical methods
Clinical suitability	**Technique for clinical evaluation and implementation in clinics**	ELISA. More straightforward approach	TMA
	Applicability in clinics	Independent of cellular content (overcome the 22% failure rate associated to insufficient cells)	Dependent on cells in the sample

CHAPTER 3

Biomarker verification

Adapted from:

"Development of a Sequential Workflow based on LC-PRM for the Verification of Endometrial Cancer Protein Biomarkers in Uterine Aspirate Samples"

Martinez-Garcia E, Lesur A, Devis L, Rosa Campos A, Cabrera S, van-Oostrum J, Matias-Guiu X, Gil-Moreno A, Reventos J, Colas E, Domon B

Oncotarget 2016

SPECIFIC BACKGROUND

The ideal biomarker pipeline consists of sequential phases of discovery, verification and validation. As shown in Chapter 1, the vast majority of biomarker studies cover either discovery phases that generate large lists of candidate biomarkers using a limited number of samples, most of which have never been further validated; or validation studies focusing on a specific protein, with an increased risk of not generating concrete application and hampering the search of biomarker panels that improve the diagnostic performance of individual proteins. The intermediate verification phase is crucial for the prioritization of candidate biomarkers to enter a validation phase in order to increase the likelihood of identifying clinically relevant biomarkers [202]. The lack of methods to guide the prioritization of candidates in the verification phase has been identified as one of the factors of poor translation of biomarkers from the discovery phase into a clinical application [201,204]. The LC-MS platform, operated in targeted acquisition mode, is ideal to achieve this task as proteins can be reliably quantified in a highly multiplexed fashion and at a fast throughput.

In this study we aimed to i) develop a stepwise verification workflow that prioritizes, from the list of 506 potential biomarkers generated in Chapter 1, the most promising ones to enter into a further validation phase; ii) evaluate the performance of the PRM, a new generation of targeted acquisition method employed on a high resolution accurate mass spectrometer, on clinical samples of uterine aspirates; and iii) assess the potential of the fluid fraction of uterine aspirates as a source of protein EC biomarkers.

MATERIAL AND METHODS

Patient recruitment and sample collection

Uterine aspirate samples were collected from a total of 42 patients (22 women suffering from EC and 20 non-EC controls, i.e., women having EC symptoms but not diagnosed with EC) recruited in the Vall d'Hebron University Hospital (Barcelona, Spain) during 2012 to 2015. Informed consent forms, approved by the Vall d´Hebron Ethical Committee, were signed by all patients (approval number: PR_AMI_50-2012). Inclusion criteria were postmenopause, a minimum age of 50 years and vaginal bleeding. Women who had been treated previously for gynecological pelvic cancer and patients positive for the human immunodeficiency virus and/or the hepatitis virus were excluded.

As described in Chapter 2, uterine aspirates were collected by aspiration with a Cornier Pipelle (Eurogine Ref. 03040200) in the office of the clinician or in the operating room prior to surgery and transferred to 1.5 ml microtubes. Phosphate buffer saline 1X was added in a

1/1 (v/v) ratio and centrifuged at 2,500 rcf for 20 min in order to separate the fluid fraction (supernatant) from the cellular fraction (pellet). The fluid fractions were kept at -80°C until use. From the 42 supernatants collected, samples coming from four patients were used for potential biomarker selection process and the development of the LC-PRM method. The list of selected biomarker candidates was then verified in the 20 EC and 18 non-EC remaining samples by LC-PRM analysis. The clinical and pathological characteristics of these 38 patients are summarized in Table 12.

Table 12. Clinical characteristics of women enrolled in the verification study.

	EEC (n=20)	Non-EC control (n=18)
Age (years)		
Average	66	61
Minimum	56	50
Maximum	81	83
Uterine condition		
Premenopausal	-	-
Postmenopausal	20	18
Histologic grade		
Grade 1	2	
Grade 2	13	
Grade 3	5	
FIGO stage		
IA	10	
IB	4	
II	3	
IIIA	2	
IIIB	-	
IIIC1	-	
IIIC2	-	
IVA	-	
IVB	1	
Myometrial invasion		
<50%	12	
>50%	8	

Evaluation of the detection of potential protein biomarkers in uterine aspirates by LC-MS analysis

Uterine aspirate supernatants from two patients diagnosed with EC and two non-EC controls were sonicated (Labsonic M, Sartorius Stedim Biotech) at 100% amplitude during 8 cycles of 15 seconds and 50 µl of each sample was depleted from albumin and IgG using the Albumin & IgG depletion spin trap kit (GE Healthcare) according to the manufacturer's

instructions. Total protein concentration was measured by the Bradford assay, performed in triplicate, and each sample was divided in two aliquots of 7 µg and processed as described in Chapter 2 (page 81). The LC-MS detection of the 506 potential biomarkers in the fluid fraction of uterine aspirate samples was then evaluated using a LTQ-Orbitrap Velos mass spectrometer operated in DDA mode, as described in pages 81-82. Peptides and related proteins identification was performed using Proteome Discoverer software (v1.4) as explained in page 82.

Effect of differential blood content on candidate biomarker detection in uterine aspirates

Uterine aspirates from one EC and one control patients were split into four equal-volume aliquots and spiked with increasing volumes of full blood (0, 10, 20 and 40% (v/v)). Samples were centrifuged at 2500 rcf for 20 min in order to separate the fluid part from the pellet. Supernatants were treated and analyzed by LC-MS as described in the previous paragraph. The elution profile areas of the peptides of 129 potential biomarkers identified in the uterine aspirates of these two patients with the different percentage of blood added were extracted from the high resolution survey scans (the identity of peptide was confirmed by MS2) using the Skyline software (v3.1) (McCoss Lab, University of Washington, Seattle, WA, USA). The levels of the surrogate peptides of each protein across the four aliquots with increasing percentage of full blood were plotted. The slope of the linear regression was calculated for each peptide and those presenting a positive slope in both patients were rejected from the study.

LC-PRM analysis

❖ Sample preparation

Fluid fractions from uterine aspirates coming from 20 EC patients and 18 non-EC controls were sonicated to disrupt potential microvesicles, protein aggregates, and/or mucus by 5 cycles at 100% amplitude during 5 seconds (Labsonic M, Sartorius Stedim Biotech). Albumin and IgG were then depleted from 50 µl of supernatant samples using the Albumin & IgG depletion spin trap kit according to the manufacturer's instructions. Total protein concentration was measured by the Bradford assay performed in triplicate. Each of the 38 samples were then separated into two aliquots of 25 µg to generate duplicates for the whole process, with exception of one sample for which the amount of material was not sufficient for duplication. The samples were diluted into a 50 mM solution of ammonium bicarbonate to a final volume of 120 µl and were denatured by addition of 185 µl of 10 M urea suspended in 50 mM ammonium bicarbonate, incubated at 22°C under agitation for 20 min, and followed by 10 min incubation in an ultrasonic bath (Branson 5510, Branson Ultrasonics). The

samples were then reduced with 7.8 µl of 200 mM DTT for 60 min at 37°C, and alkylated with 12.2 µl of 400 mM IAA at 22°C for 30 min in the dark. The samples were digested for 4 h at 37°C with Lys-C (protease/total protein amount ratio of 1/150; w/w). Afterwards, the concentration of urea was diluted to 1 M with 50 mM ammonium bicarbonate buffer, and samples were incubated overnight at 37°C with trypsin (protease/total protein amount ratio of 1/50; w/w). The trypsin activity was quenched by addition of 1 µl of neat formic acid per 100 µl of solution. A mixture of the 100 stable isotope-labeled synthetic peptides (Thermo Fisher, crude quality) was spiked in each sample (C terminal arginine $^{13}C_6$, $^{15}N_4$, Δm = 10 Da, C terminal lysine $^{13}C_6$, $^{15}N_2$, Δm = 8 Da or when it was not applicable, with a heavy leucine $^{13}C_6$, $^{15}N_1$, Δm = 7 Da or phenylalanine $^{13}C_9$, $^{15}N_1$, Δm = 10 Da). Finally, samples were purified by solid phase extraction (Sep Pak tC18, 50 mg, Waters). The eluates were subsequently evaporated to dryness in a vacuum centrifuge and suspended in 0.1% formic acid before LC-PRM analysis.

❖ **LC-PRM setup**

The LC-MS setup consisted of a Dionex Ultimate 3000 RSLC chromatography system configured for a high-pressure binary gradient and operated in column switching mode. The mobile phase A consisted of 0.1% formic acid in water, the phase B in 0.1% formic acid in acetonitrile and the loading phase in 0.05% trifluoroacetic acid and 1% acetonitrile in water. The equivalent of 250 ng of each digested sample was injected and loaded onto a trap column (75 µm × 2 cm, C18 pepmap 100, 3µm) at 5 µl/min and further eluted onto the analytical column (75 µm × 15 cm, C18 pepmap 100, 2µm) at 300 nl/min by a linear gradient starting from 2 % B to 35 % B in 48 min. The MS analysis was performed by a hybrid quadrupole orbitrap mass spectrometer (Q Exactive plus, Thermo Scientific) operated in PRM mode. The MS cycle started with a full MS1 scan performed at a resolving power of 70,000 (at 200 m/z) followed by time scheduled targeted PRM scans acquired at a resolving power of 35,000 (at 200 m/z) with a normalized collision energy of 20. The quadrupole isolation window for the PRM events was set to 1 m/z unit and the duration of the time scheduled windows for each pair of endogenous and isotopically labeled peptides were set to 2 min.

❖ **PRM data processing**

The areas of extracted ion chromatograms (XIC) of the five most intense fragment ions of each precursor (i.e., PRM transitions) were extracted using the Skyline program (v3.1) (McCoss Lab, University of Washington, USA). The selection of the best fragments was supported by a spectral library obtained from a reference LC-PRM acquisition of the synthetic peptide mix injected without biological matrix. The elution profiles were first manually reviewed and obvious interfered PRM transitions were replaced by the next most

intense available product ion. The data set was then refined using the cosine of the spectral contrast angle (cos θ) calculated between the peak areas of the five PRM transitions of the reference (PRM acquisition of the synthetic peptides mix) and the areas of the corresponding transitions for the endogenous and heavy peptides in the biological samples [313]. The formula is as follows:

$$cos\ (\theta) = \frac{\sum_{i=1}^{n}\left(A_{exp_i} \times A_{ref_i}\right)}{\sqrt{\sum_{i=1}^{n}\left(A_{exp_i}\right)^2} \times \sqrt{\sum_{i=1}^{n}\left(A_{ref_i}\right)^2}}$$

Where A_{exp} are the areas of either the endogenous or heavy PRM transitions of selected product ions for a peptide measured in a sample, and A_{ref} are the areas of the same transitions measured in a reference synthetic peptides mixture.

Peptides detection and identification were confirmed if the cos θ of the endogenous and the isotope labeled peptide were higher than 0.98 [241]. Scores below 0.98 are principally due to MS measurements below the limit of detection and in such cases the area values were replaced by an estimation of the background. Peptides with cos θ below 0.98 in more than 50% of the analyzed samples were eliminated from the study.

For the quantitative analysis, the area ratios between the endogenous and their corresponding heavy peptides were compared between samples. The area ratios were calculated as the sum of the areas of the PRM transitions of the endogenous peptide divided by the sum of the same transitions of the respective isotope labeled version.

❖ **Statistical analysis**

The statistical analysis was performed in SPSS (v20.0) (IBM, Armonk, NY, USA) and Graph Pad Prism (v.6.0) (GraphPad Software, La Jolla, CA, USA). The averaged endogenous/heavy area ratios were calculated between duplicates. The linear correlation between the signature peptides of the same protein was calculated using the Pearson correlation coefficient (R coefficient). Due to the non-normality of the data, assessed by Kolmogorov-Smirnova and Shapiro-Wilk tests, comparison of the abundance of the monitored peptides between tumor and control samples was performed using the non-parametric Mann-Whitney U test. P-values were adjusted for multiple comparisons using Benjamini-Hochberg FDR method [314]. Adjusted p-values lower than 0.05 along with fold changes greater than three were considered statistically significant. Receiver operating characteristic (ROC) curves were used to calculate the relationship between sensitivity and specificity for EC versus non-EC control group and hence, to evaluate the diagnostic performance for each biomarker candidate.

RESULTS

LC-PRM method development: selection of the candidate biomarkers

A targeted MS-based approach was selected for the verification of potential EC biomarkers in uterine aspirates as it enables the quantification of multiple peptides within a single analysis. The LC-PRM is a hypothesis driven methodology that differs from the unsupervised MS-based approaches (e.g., data dependent and data independent acquisition) as the proteins must be selected prior the actual MS acquisition. With LC-PRM, the number of targets is limited to approximately 100-150 peptides per analysis, but in return, the mass accuracy and the high resolving power of the orbitrap analyzer, in conjunction with the use of isotope labeled peptides as internal standard, allows for systematic quantitative measurements in all samples achieved with a high degree of selectivity and precision. Therefore, a pre-selection of the protein candidates to be measured is required. Starting from 506 protein candidates found in the literature from previous studies (Chapter 1), we proposed a workflow to reduce step by step this number down to 52 candidate biomarkers, leading to 104 pairs of light/heavy peptides that can be measured by a single LC-PRM method. We verified those candidates in uterine aspirates from a cohort of 20 EC patients and 18 controls by LC-PRM (Figure 21).

The starting point of this study was the extensive literature review described in Chapter 1 of this thesis, where we obtained a first list of 506 proteins associated with EC (Annex 1), which were mostly derived from studies performed in endometrial tissue samples. The second step of selection consisted in the assessment of the LC-MS detection of those 506 potential biomarkers in the fluid fraction of four samples of uterine aspirates by repeated DDA analysis. The main goal of this step was to reduce the list of protein candidates to those that can be effectively detected in the fluid fraction of uterine aspirate samples. From a total of 1,086 proteins identified in the four uterine aspirates, 158 proteins out of the initial 506 potential biomarkers list were detected (Annex 1). This first screening indicated that one third of the potential biomarkers could be easily detected by LC-MS techniques in the fluid fraction of uterine aspirates samples, thus confirming the potential of this sample as a source of protein EC biomarkers.

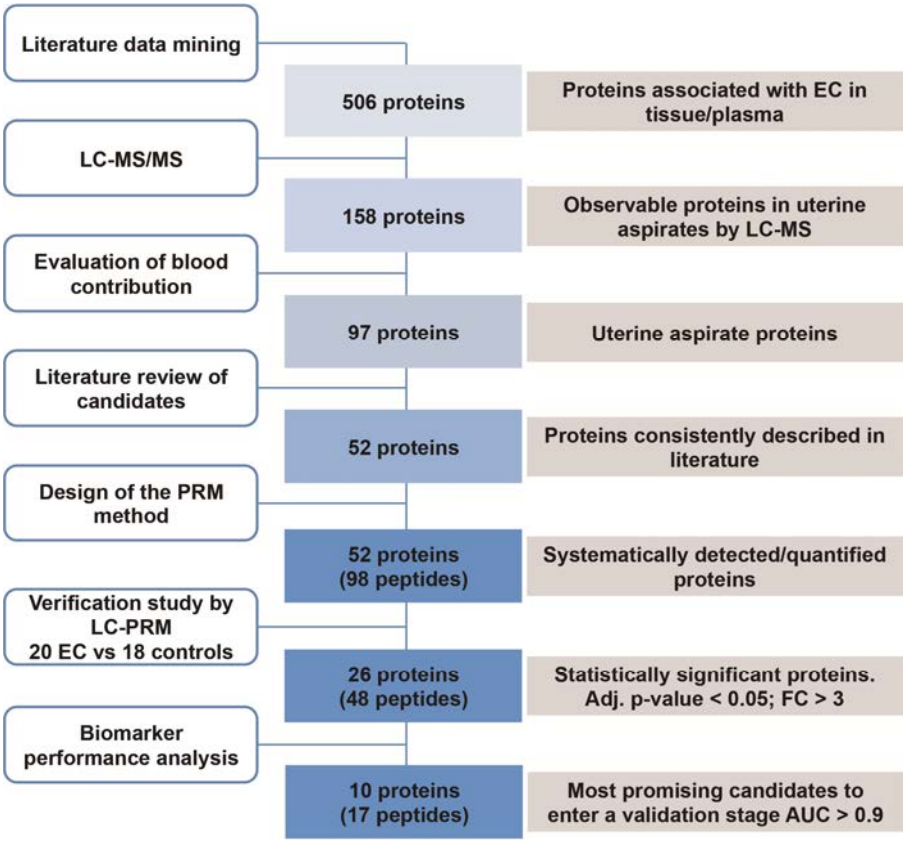


Figure 21. Stepwise workflow for the selection and prioritization of endometrial cancer candidate biomarkers, and their verification in uterine aspirates by LC-PRM. Adj p-value, adjusted p-value; FC, fold change; AUC, Area under the ROC curve.

Blood contamination of biological samples is a recurrent problem in bioanalyses, particularly in the field of biomarker research in some biofluids [215,216]. Understanding that uterine aspirates display a variable amount of blood between samples, we introduced a third step of selection to evaluate the interference of blood components during LC-MS detection of the potential biomarkers in uterine aspirates. To do that, the uterine aspirates of two patients (one control and one EC patient) were split into four equal-volume aliquots and spiked with increasing volumes of full blood: 0, 10, 20, 40% (v/v). All samples were digested and analyzed by LC-MS in duplicate. We excluded those proteins whose peptides displayed an increasing profile with an increasing concentration of spiked-in blood and maintained those proteins showing no effect or diminished levels (Figure 22). This criterion was used in order to discriminate protein biomarkers coming from the endometrial tissue rather than proteins contained in the blood proteome. Moreover, by excluding abundant proteins of blood, we

reduced analytical problems related to variable blood contamination among the samples. As a result of this analysis, 32 proteins were likely to be derived from the blood contamination of the uterine aspirates and were excluded from further analysis.

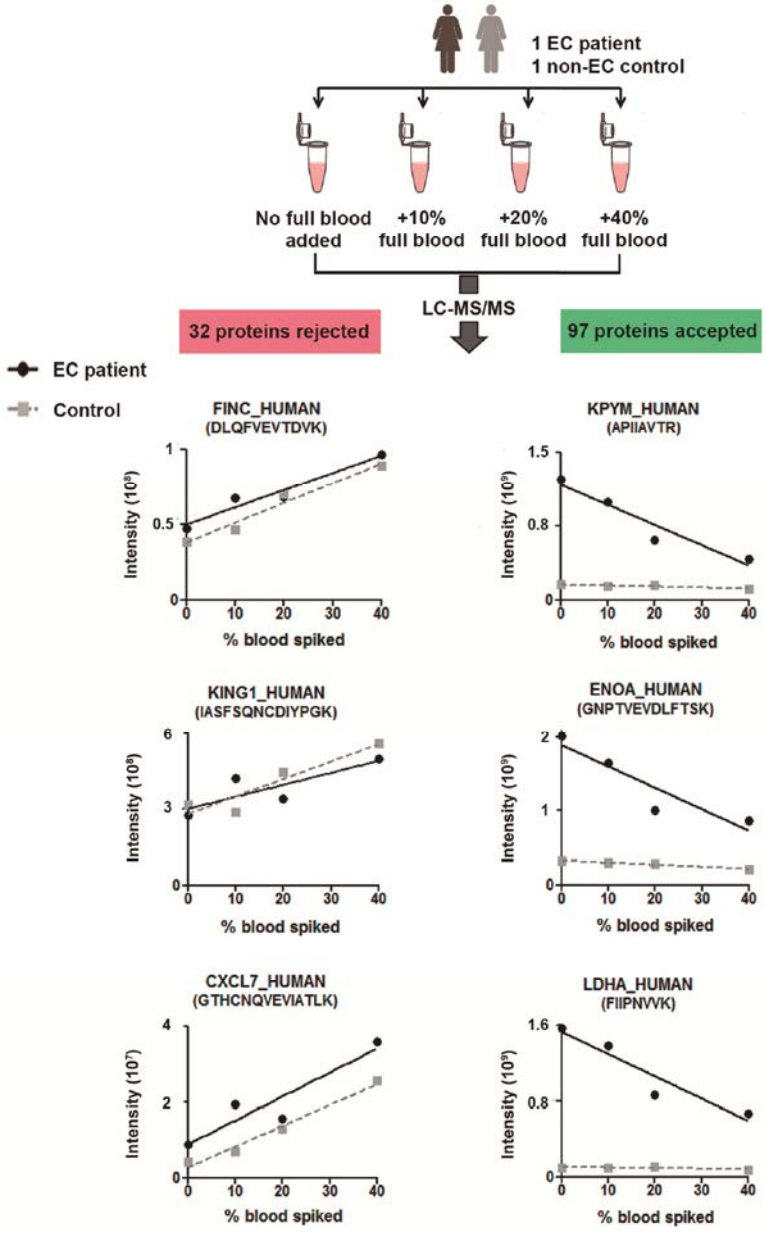


Figure 22. Effect of blood content on candidate biomarker detection. Experimental design and examples of concentration profiles of 3 potential biomarkers showing increasing and 3 decreasing profiles when uterine aspirate is diluted by increasing amount of full blood.

The remaining 97 uterine aspirate specific candidates were scaled down to 52 proteins based on their consistency in literature (Annex 1). The 52 candidates had undergone at least one level of additional validation at the protein level using a different technology, biospecimen type or different cohorts of patients in the same publication or in independent reports. A total of two peptides per each of these 52 proteins (104 peptides) were selected according to their uniqueness, detection and chromatographic behavior.

Quality control of the LC-PRM data

The 52 proteins of interest were verified in the fluid fraction of uterine aspirates by targeted MS. Uterine aspirates from 20 EC patients and 18 non-EC controls were digested in duplicate and analyzed by a quadrupole-orbitrap MS operated in PRM mode using a mix of the stable isotopes labeled (SIL) peptides of the 104 peptides (i.e., heavy peptides) as internal standards. Four of these SIL peptides could not be synthesized, leading to a final list of 100 monitored peptides in the method (available in the Panorama Public repository: https://panoramaweb.org/labkey/project/Panorama%20Public/2016/Domon%20-%20EC/begin.view?). The signals of the five most intense product ions for each precursor were extracted from the MS2 spectra to generate elution profiles. The identity of the peptides, as well as the potential interferences on the PRM traces, were evaluated by a similarity score based on the cosine of the spectral contrast angle (cos θ) calculated with the top five fragment ions of each precursor. This score was calculated against a reference LC-PRM analysis of the isotopically labeled peptides without biological matrix (Figure 23A). The signal of a peptide was accepted if the cos θ was higher than 0.98 for both the endogenous and the stable isotope labeled standard [241]. Values lower than 0.98 due to an interfered PRM transition were replaced by the next most intense available PRM transition. Six peptides were monitored with four transitions due to the absence of a clean fifth transition. Following this, a positive spectral matching was achieved for 95.1% of a total of 7,350 pairs (ratio light/heavy). The unmatched 4.9% pairs were due to two conditions: i) measurements below the limits of detection (4.7%), which were replaced with an estimation of the background value. In this account, peptides below the limit of detection in more than 50% of the samples, only two peptides -VHITSLLPTPEDNLEIVLHR and VTILELFR- fulfilled this condition, were removed from the study; and ii) measurements for which less than four clean PRM transitions are detected (0.2%). These 0.2% were due to data very close to, but below, cos θ = 0.98 and only one replicate was affected in all cases; thus the value of the accepted replicates was kept. These results illustrate the efficiency of the PRM acquisition in complex clinical samples. The use of internal standards and the availability of all transitions guarantee the correct identification of each peptide, reduce interferences and facilitate the detection and exclusion of potential interferences in large datasets. The mean between

duplicates in the cleansed dataset was calculated, as well as the correspondent coefficient of variation (CV%). The CV% of the duplicated sample preparation for each uterine aspirate sample was below 15% for 99% of the detected peptides, with an averaged CV of 3.6%. This confirmed the high reproducibility level of the full process (Figure 23B). Finally, the correlation between the peptides derived from the same protein was evaluated by a Pearson correlation coefficient (Figure 23C) and 39 out of the 46 (85%) proteins monitored with two peptides showed a very high correlation, with a R coefficient over 0.95. Only 3 proteins - ROA2, OSTP, and KPYM- presented an R coefficient below 0.9, which were due to the specificity of the monitored peptides to different isoforms of the same protein.

Differentially abundant proteins between endometrial cancer and control uterine aspirates

In order to assess the potential of the 52 selected proteins to detect EC, we compared the abundance of each candidate biomarker between 20 EC patients and 18 non-EC controls. Importantly, both patients and controls were postmenopausal women suffering from an abnormal vaginal bleeding, as these clinical features are present in 93% of patients suffering from EC. However, only 8-15% of those will be finally diagnosed with EC [7].

Based on the Bradford assays, 250 ng of the total protein concentration after albumin and IgG depletion was injected for each sample. The constant amount of injected protein among samples was further confirmed by the integration of the total ion chromatogram of the MS1 scans. After MS data curation, the relative levels (light/heavy ratios) of the final 98 monitored peptides in MS2 were subjected to Mann Whitney test for their comparison between tumor and control samples. Forty-eight peptides corresponding to 26 proteins showed significant differences between the two groups with adjusted p-value <0.05 (Benjamini corrected) and fold change greater than 3: PERM, CADH1, SPIT1, ENOA, MMP9, LDHA, CASP3, KPYM, PRDX1, OSTP, PDIA1, NAMPT, MIF, CTNB1, K2C8, ANXA2, CAPG, FABP5, MUC1, CAYP1, XPO2, NGAL, SG2A1, ANXA1, HSPB1, PIGR. All these proteins showed higher levels in tumor samples as compared to control samples (Table 13). Interestingly, from the 18 differentially abundant proteins in the fluid fraction of uterine aspirates obtained in the study described in Chapter 2, 17 were also included in this study and 14 were confirmed as diagnostic biomarkers in this independent set of patients.

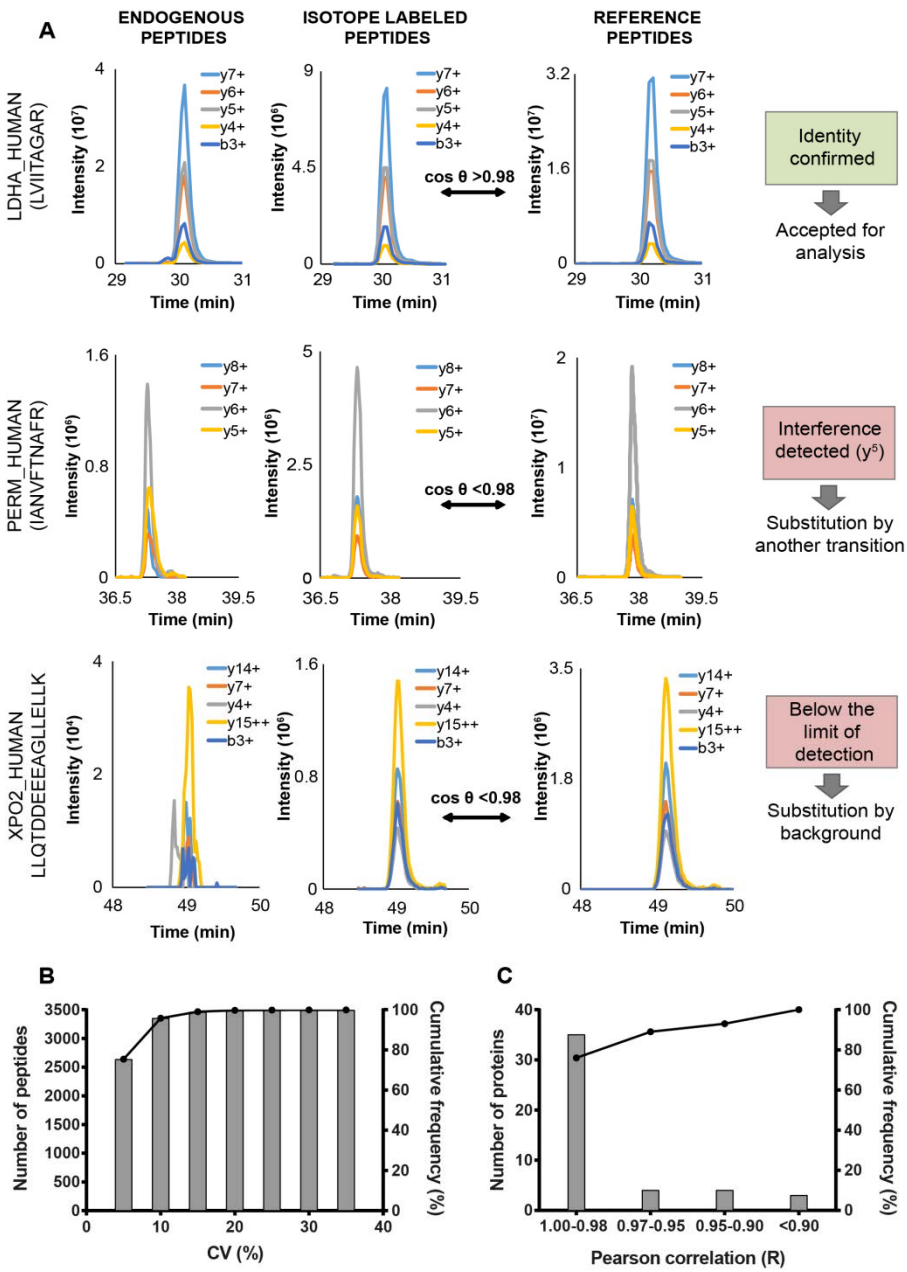


Figure 23. Principle of PRM data quality control. A. Peptide identity confirmation by comparison between PRM elution profiles of endogenous and internal standards of each candidate biomarker in the samples and a reference acquisition using the cosine of the spectral contrast angle (θ). **B.** Coefficient of variation (CV%) of the peptide signals quantified in the 38 uterine aspirates processed and analyzed by LC-PRM in duplicates. **C.** Pearson correlation between signatures peptides coming from the same protein.

To further evaluate their performance as biomarkers for EC diagnosis, we performed a ROC analysis to determine the sensitivity and specificity of each biomarker. Interestingly, these differentially abundant proteins showed Area Under the ROC Curve (AUC) values for discriminating between EC and controls patients ranging from 0.75 to 0.97. The 10 best-performing individual proteins were PERM, CADH1, SPIT1, ENOA, MMP9, LDHA, CASP3, KPYM isoform M1-M2, PRDX1 and OSTP isoform A, all of them with AUC values higher than 0.9 (Figure 24). Among those proteins, PERM, CADH1, SPIT1 and OSTP isoform A were of special interest as each of them presented sensitivities higher than 80% when specificity was fixed to 95% (Table 13). This is particularly important in EC diagnosis, as biomarkers with high specificity could complement the output of non-invasive techniques such as the transvaginal ultrasonography, which currently presents very high sensitivity but lack of specificity [103].

Furthermore, we conducted a bioinformatics analysis using Ingenuity Pathway Analysis (IPA) to better understand the association of these proteins with cancer and their origin regarding the subcellular location. As expected, integration of the data resulted in the identification of cancer, inflammatory disease, organismal injury and abnormalities, and reproductive system disease as the top diseases associated to these biomarkers. The top five molecular and cellular functions involved with these proteins included cellular movement, cellular death and survival, cellular development, cellular growth and proliferation, and cell-to-cell signaling and interaction, all of them important processes altered in cancer. These proteins are mainly found in the cytoplasm, plasma membrane and extracellular space (Table 13), indicating that they are coming either from secretion of the epithelial and inflammatory cells of the endometrium or by necrosis of cells in the proximal tissue. This is in concordance with the observation that all biomarkers in this study were found more abundant in EC patients as compared to controls, as both processes are related to the higher proliferation rate of epithelial cells in EC.

Table 13. Proteins showing statistical differences between EC (n=20) and control patients (n=18) with adjusted p-value <0.05 and fold change >3. FC, fold change; AUC, area under the ROC curve; SN, sensitivity; SP, specificity; Isof, isoform.

Uniprot Number	Protein name	Protein ID	Peptide	FC	Adjusted P-value	AUC	SN (%)	SP (%)	SN (%) at 95% SP	Protein location
P05164	Myeloperoxidase	PERM	IANVFTNAFR	14.1	6.E-05	0.97	95	89	80	Cytoplasm
			VVLEGGIDPILR	13.3	1.E-04	0.95	95	89	70	
P12830	E-cadherin	CADH1	VFYSITGQGADTPPVG-VFIIER	3.8	9.E-05	0.94	95	89	85	Plasma Membrane
			NLVQIK	3.3	2.E-04	0.93	85	94	85	
O43278	Kunitz-type protease inhibitor 1	SPIT1	SFVYGGCLGNK	3.3	1.E-04	0.93	95	94	95	Extracellular Space
			WYYDPTEQICK	3.3	1.E-04	0.93	90	94	90	
P06733	Alpha-enolase	ENOA	YISPDQLADLYK	3.8	1.E-04	0.92	75	94	75	Cytoplasm
			TIAPALVSK	4.0	2.E-04	0.89	80	83	70	
P14780	Metalloproteinase 9	MMP9	SLGPALLLLQK	5.7	1.E-04	0.91	95	83	60	Extracellular Space
			AFALWSAVTPLTFTR	5.5	1.E-04	0.91	90	83	60	
P00338	Lactate dehydrogenase A	LDHA	LVIITAGAR	6.2	1.E-04	0.91	85	89	65	Cytoplasm
			VTLTSEEEAR	5.7	1.E-04	0.91	85	89	60	
P42574	Caspase-3	CASP3	SGTDVDAANLR	4.9	2.E-04	0.91	90	89	65	Cytoplasm
P14618	Pyruvate kinase	KPYM Isof M1-M2	NTGIICTIGPASR	5.4	1.E-04	0.91	85	89	75	Cytoplasm
		KPYM Isof M1-M3	APIIAVTR	3.1	1.E-02	0.75	60	89	50	
Q06830	Peroxiredoxin-1	PRDX1	LVQAFQFTDK	4.2	2.E-04	0.90	75	94	75	Cytoplasm
			ADEGISFR	4.2	2.E-04	0.90	75	94	75	
P10451	Osteopontin	OSTP Isof A	ANDESNEHSDVIDSQE-LSK	11.4	2.E-04	0.90	80	94	80	Extracellular Space
		OSTP Isof A, B, D	AIPVAQDLNAPSDWDS-R	9.0	4.E-04	0.87	80	83	50	
P07237	Protein disulfide-isomerase	PDIA1	ILEFFGLK	3.3	3.E-04	0.89	75	89	65	Cytoplasm
			ALAPEYAK	3.0	3.E-04	0.88	75	89	65	
P43490	Visfatin	NAMPT	YLLETSGNLDGLEYK	4.2	3.E-04	0.88	90	83	40	Extracellular Space
			YDGHLPIEIK	4.0	3.E-04	0.88	90	83	40	
P14174	Macrophage migration inhibitory factor	MIF	VYINYYDMNAANVGWN-NSTFA	4.2	3.E-04	0.88	75	94	75	Extracellular Space
			LLCGLLAER	3.1	3.E-04	0.87	70	94	70	
P35222	Beta-catenin	CTNB1	LLNDEDQVVVNK	4.2	3.E-04	0.88	85	89	70	Nucleus
			LVQLLVR	4.2	3.E-04	0.87	85	89	65	
P05787	Keratin, type II cytoskeletal 8	K2C8	LSELEAALQR	3.6	3.E-04	0.88	95	67	65	Cytoplasm
			WSLLQQQK	3.1	6.E-04	0.85	60	94	60	
P07355	Annexin A2	ANXA2	GVDEVTIVNILTNR	4.8	4.E-04	0.87	75	89	45	Plasma Membrane
			QDIAFAYQR	5.1	5.E-04	0.86	95	61	50	
P40121	Macrophage-capping protein	CAPG	EGNPEEDLTADK	3.6	5.E-04	0.85	85	83	45	Nucleus
			YQEGGVESAFHK	3.5	6.E-04	0.85	80	83	45	
Q01469	Fatty acid binding protein 5	FABP5	LVVECVMNNVTCTR	3.9	6.E-04	0.85	90	78	45	Cytoplasm
			ELGVGIALR	3.6	6.E-04	0.85	90	78	45	
P15941	Mucin-1	MUC1	QGGFLGLSNIK	3.6	1.E-03	0.84	85	78	45	Plasma Membrane
Q13938	Calcyphosine	CAYP1	SGDGVVTVDDLR	3.4	1.E-03	0.83	85	78	45	Cytoplasm
P55060	Exportin-2	XPO2	ANIVHLMLSSPEQIQK	4.0	1.E-03	0.83	75	89	25	Nucleus
			LLQTDDEEEAGLLELL-K	4.4	2.E-03	0.81	70	89	40	
P80188	Lipocalin2	NGAL	VPLQQNFQDNQFQGK	5.0	1.E-03	0.83	75	89	35	Extracellular Space
			ELTSELK	4.4	4.E-03	0.79	70	83	30	
O75556	Mammaglobin-B	SG2A1	ELLQEFIDSDAAAEAM-GK	3.3	3.E-03	0.80	90	72	30	Extracellular Space
			TINSDISIPEYK	3.2	5.E-03	0.78	90	67	40	
P04083	Annexin A1	ANXA1	DITSDTSGDFR	4.8	3.E-03	0.80	60	100	60	Plasma Membrane
			GGPGSAVSPYPTFNP-SSDVAALHK	3.9	7.E-03	0.77	55	100	55	
P04792	Heat shock 27kDa protein 1	HSPB1	LFDQAFGLPR	3.6	4.E-03	0.79	85	67	40	Cytoplasm
			LATQSNEITIPVTFESR	3.1	4.E-03	0.79	85	67	40	
P01833	Polymeric immunoglobulin	PIGR	VYTVDLGR	3.4	7.E-03	0.77	80	78	30	Plasma Membrane

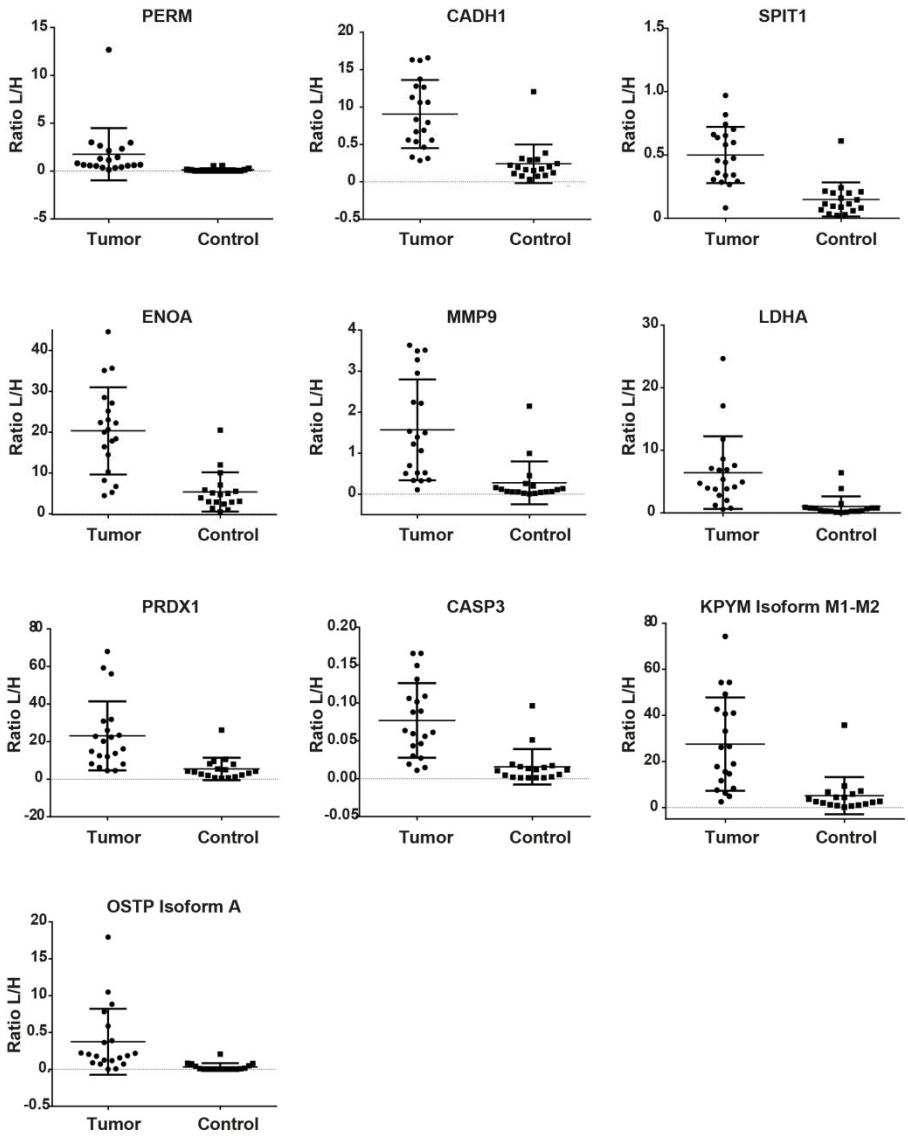


Figure 24. Scattering plots of the levels of the peptides coming from 10 biomarkers in the verification study. Scattering plots depicting the distribution of the light/heavy (L/H) ratios across the 20 EC patients and 18 controls of the best individual performing biomarkers (AUC>0.9).

DISCUSSION

About 30% of EC patients are diagnosed at advanced stages of the disease, associated with a drastic increase in the mortality and morbidity [27]. Therefore, the identification of sensitive and specific biomarkers to improve early detection of EC is an important clinical need. As discussed in Chapter 1, despite the major effort and investments made to identify EC biomarkers, no protein has yet reached the stage of clinical application. The poor translation of the results produced by those studies in the clinic can be explained by two determinant factors: on the one side, the lack of studies in biofluids to identify accessible EC biomarkers. Most of the studies were based in tissues, and/or those that used biofluids were limited to serum or plasma [315–317]. Blood presents several important advantages, as it is in direct contact with all body tissues, and its collection is rapid, easy and minimally invasive. However, the search of biomarkers in plasma or serum is extremely challenging due to the low concentration of the potential biomarkers and the wide dynamic range in protein abundance [197]. In addition, the fact that blood flows through all organs also hampers the identification of biomarkers of a disease occurring in a specific part of the body. On the other side, the lack of verification studies as a bridge between discovery and validation phases, which has been defined as the current bottleneck of the biomarker pipeline [201,202]. Discovery studies generate large lists of differentially abundant proteins. Many of those candidate biomarkers are never validated or turn to be false positives due to the small number of samples analyzed, the biological variability and/or the limited quantitative performance of the technologies employed in this phase. There is a need to verify and refine those lists to the best candidates that can enter a validation phase. This is the critical role of the verification phase. In order to overcome these limitations, we presented a stepwise workflow to select potential EC biomarkers and verified them by targeted MS-based analysis in the fluid fraction of uterine aspirate samples.

Targeted MS-based approaches have gained in popularity for biomarker verification in complex clinical samples because they combine precision, sensitivity, multiplexing and absence of missing values. Among those, SRM acquisition mode performed on a triple quadrupole mass spectrometer has been the reference method for the accurate quantification of peptides in biological matrices [292,318,319]. However, SRM is limited in selectivity and requires a substantial method development. We implemented the PRM acquisition, a new generation of targeted MS-based approach, performed on high-resolution accurate mass spectrometers (HRAM) such as the hybrid quadrupole-orbitrap. To date, the advantage of PRM for large scale analysis has been evaluated [241,298] but not yet commonly introduced as a technique for biomarker searches in clinical [299] or cell lines samples [300–302]. The high resolution and the accurate mass (i.e., 35,000 at 200 m/z and below 5 ppm) of the

orbitrap analyzer decrease the risk of inferences due to the complexity of the chemical background and we obtained clear and easily readable chromatograms profiles. This aspect, in conjunction with the use of spectral matching as a quality metric, significantly facilitates the data processing to compare with SRM data, for instance only 0.2% of the chromatographic peak needed to be manually curated in this study. A straightforward and highly automatable data processing is an important requirement of large scale studies. The PRM acquisition allowed for the quantification of 100 pairs of peptides at the reasonable throughput of one analysis per hour with an excellent precision (i.e., the CV% between full workflow duplicates was below 15% for 99% of the detected peptides). Finally, the design of an LC-PRM method is easier and faster than for LC-SRM, as the selection of the fragment ions to quantify is performed post-acquisition and the list of transitions can be refined iteratively to remove potential interferences coming from the background, without the need of a new analysis [259].

Another strength of this study is the use of the fluid fraction of uterine aspirates as a biological sample for biomarker detection. A useful diagnostic biomarker not only has to ameliorate the discrimination between patients suffering the disease and benign cases, but also should be economically profitable and advantageous in the clinical scenario. In the case of diagnostic biomarkers for EC, accelerate the diagnostic process, improving the comfort of patients, and reducing the healthcare costs are very important values. After evaluating the feasibility of the analysis of uterine aspirates by MS in Chapter 2, we here confirmed the convenience of the fluid fraction as a source of EC biomarkers.

Our final achievement was to eliminate doubtful candidate biomarkers derived from the variable amounts of blood contamination in uterine aspirates and successfully verify the differential levels between EC and non-EC patients of 26 proteins in this sample. A bioinformatics analysis confirmed their individually and collectively association with cancer, and showed that they maintain a strong association with commonly altered molecular processes in cancer such as cellular movement, cellular death and survival, etc. Among all candidates, ten provided high sensitivity and specificity, with AUC values over 0.9, and four of those, PERM, CADH1, SPIT1 and OSTP, were highlighted as they achieved sensitivity over 80% when fixing a specificity of 95%. The protein biomarkers verified in this study merit further validation in an extended study with a larger cohort of patients and controls to confirm their diagnostic power and evaluate their clinical applications. A limitation of the present study is that we did not use combinations of multiple markers to avoid overfitting due to the relatively small number of subjects included [320]. Although the diagnostic power of the individual proteins was already high, the development of panels of proteins that improve their diagnostic performance must be evaluated in the next step.

In conclusion, this study brings forward the proteomic search of EC biomarkers in the fluid fraction of uterine aspirates following an appropriate workflow. Moreover, this study proves the efficiency of high resolution MS operated in PRM mode in order to verify a large number of potential biomarkers to fill the gap between discovery and validation studies. The described workflow permitted to reduce step by step an initial list of 506 potential biomarkers down to ten proteins with an increased likelihood to reach the stage of a clinical assay after a subsequent validation phase.

CHAPTER 4

Biomarker validation

Adapted from:

"Targeted proteomics identifies proteomic signatures in liquid-biopsies of the endometrium to diagnose endometrial cancer and assist in the prediction of the optimal surgical treatment"

Martinez-Garcia E, Lesur A, Devis L, Cabrera S, Matias-Guiu X, Hirschfeld M, Asberger J, van-Oostrum J, Casares de Cal MA, Gómez-Tato A, Reventos J, Domon B, Colas E, Gil-Moreno A

Clinical Cancer Research 2017

SPECIFIC BACKGROUND

The "validation phase" is a key step in the biomarker pipeline to decide whether there is sufficient power to justify moving candidates to the final clinical evaluation. In this phase, the good performance of the biomarkers (high sensitivity and specificity) must be confirmed, but a clear contribution to existing clinical practices must be also demonstrated. This phase requires large numbers of samples to ensure statistical rigor and to represent as much as possible the diversity of the clinical conditions of the target population [181]. Consequently, high-throughput analytical methods are highly desirable.

On the other side, cancer is a multifactorial and heterogeneous disease, and it is unlikely that a single biomarker will display sufficient discriminatory power to significantly affect clinical decisions. Therefore, the search for biomarker signatures that provide more complete information is crucial [176], and highly multiplexing analytical methods would facilitate this task.

In the previous chapter, we demonstrated that the fluid fraction of uterine aspirates are minimally invasive samples with significant value for the screening of EC protein biomarkers. Following this investigation, here we employed the same analytical approach, LC-PRM acquisition, to evaluate the diagnostic performance of the 52 proteins in uterine aspirate samples obtained from 116 patients covering the broad clinical heterogeneity of EC cases and benign pathologies entering the EC diagnosis process. In addition, the potential of those proteins to differentiate between histological EC types was assessed. Finally, protein panels were developed to achieve the best performance to diagnose EC and discriminate between the two main EC histological subtypes. The correlation between MS-based results and high-throughput ELISA was also evaluated.

MATERIAL AND METHODS

Patients and sample description

A total of 116 women were recruited in the Vall Hebron University Hospital (Barcelona, Spain), the Hospital Universitari Arnau de Vilanova (Lleida, Spain) and the Freiburg University Medical Center (Freiburg, Germany) from 2012 to 2015. Informed consent forms, approved by the Ethical Committees of each Hospital, were signed by all patients. All women were patients entering the EC diagnostic process due to EC suspicion, i.e., patients showing an AUB and/or a thickness of the endometrium higher than 4mm for postmenopausal women [101] and 8mm for premenopausal women based on the results of a transvaginal ultrasonography [105]. From the 116 women, 69 were diagnosed with EC,

including 49 endometrioid EC (EEC) and 20 non-endometrioid serous ECs (SEC). The remaining 47 women were non-EC women with normal endometrium or diagnosed with benign disorders.

As proceeded with all samples employed in the previous chapters, uterine aspirates were collected by aspiration with a Cornier Pipelle (Eurogine Ref. 03040200). Phosphate buffer saline 1X was added in a 1/1 (v/v) ratio and centrifuged at 2,500 x g for 20 min. The fluid fraction of each uterine aspirate, ranging volumes from 100 µl to 1 ml, was separated from the cellular fraction and kept at -80°C until use.

Sample preparation for LC-PRM analysis

The sample preparation for the LC-PRM analysis was performed as described in Chapter 3 (pages 101-102). The only difference was that each of the 116 samples was divided in two aliquots of 12.5 µg prior to proteolysis (instead of 25 µg). Each sample was diluted into a 50 mM solution of ammonium bicarbonate to a final volume of 60 µl and the volume of each subsequent reagent added was adapted accordingly in order to maintain the same final concentrations.

LC-PRM setup and PRM data processing

The LC-PRM setup and subsequent PRM data processing were performed as described in Chapter 3 (pages 102-103).

Statistical analysis

The light/heavy area ratio of each peptide was extracted from Skyline and the average between duplicates was calculated. The correlation between the signature peptides of the same protein was calculated using the Pearson correlation coefficient. The statistical analysis was performed in SPSS (v20.0) (IBM, Armonk, NY, USA) and Graph Pad Prism (v.6.0) (GraphPad Software, La Jolla, CA, USA). Comparison of the levels of the monitored peptides between groups of patients was performed using the nonparametric Mann-Whitney U test, since the data did not follow a normal distribution. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg FDR method [314]. Adjusted p-values lower than 0.05 were considered statistically significant. ROC analysis was used to assess the specificity and sensitivity of the biomarkers and the AUC values were estimated for each individual protein.

Development of the classifiers

A logistic regression model was adjusted to the data in order to assess the power of the different combinations of proteins to classify samples in two clinical categories. ROC curves were generated for each of these regression models; the AUC, and the sensitivity and specificity at the "optimal" cutoff point for discrimination between groups were obtained. The optimal cut-off corresponded to the threshold that maximized the distance to the identity (diagonal) line. The optimality criterion was: max (sensitivities + specificities). AUCs 95% confidence intervals (CI) were computed with the Delong's method [321]. The 95% CIs of the sensitivity and specificity values were computed with bootstrap resampling and the averaging methods described by Fawcett [322]. All ROC analysis were performed using the R "pROC" package [323]. To assess the robustness of each protein panel, the “leave-one-out” cross-validation procedure was performed by applying to each sample in the dataset the logistic regression model adjusted to the remaining samples on the dataset, hence deriving a new ROC curve and afterwards performing the usual ROC analysis. In a similar way, the discrimination power of the diagnostic protein panel was further validated by applying to each sample of the independent set of 38 samples from Chapter 3 (cohort 2) the logistic regression model adjusted to the initial set (cohort 1), hence deriving a new ROC curve and afterwards performing the usual ROC analysis.

ELISA

The concentrations of MMP9 and KPYM isoform M1/ M2 were quantified in the soluble fraction of uterine aspirates with commercially available ELISA kits (R&D Systems and USCN life Science and Technology Company, respectively) according to the manufacturer´s protocol. For MMP9, 105 uterine aspirate samples were analyzed using 1/10; 1/100 or 1/1000 dilutions. For KPYM, only 39 uterine aspirate samples could be analyzed using 1/2, 1/4 or 1/10 dilutions due to a lack of sample material. All samples were assayed in duplicate and the average values were reported as ng/mL. The linear correlation between the results from LC-PRM and ELISA was calculated using the Pearson correlation coefficient.

RESULTS

Subject characteristics

A total of 116 patients entering the EC diagnostic process due to common symptoms in EC, mainly AUB (86%), and/or a thickened endometrium based on transvaginal ultrasonography (79%), were included in the study. The mean age of the patients was 61 years (range, 30-93 years) with 79% postmenopausal women and 21% premenopausal women. Among the 69 EC cases, the two most common histological types, endometrioid and serous histologies,

were represented in 49 (71%) and 20 (29%) cases, respectively. A total of 47 non-EC women included as controls suffered from benign uterine pathologies (polyps (36%), myomas (15%), and endometrial hyperplasia (19%)), or were women with a normal endometrium (30%). Clinical and demographical data from women included in this study are summarized in Table 14.

Table 14. Clinical characteristics of women enrolled in the validation study. *One case with undetermined grade.

	EEC (n=49)	SEC (n=20)	Non-EC control (n=47)
Age (years)			
Median	67	73	53
Minimum	37	51	30
Maximum	87	93	80
Collection center			
VHIR	41	12	37
Lleida	5	8	-
Freiburg	3	-	10
Uterine condition			
Premenopausal	7	1	16
Postmenopausal	42	19	31
Histologic grade			
Grade 1	5*	-	
Grade 2	33	-	
Grade 3	10	20	
FIGO stage			
IA	25	5	
IB	13	-	
II	9	3	
IIIA	-	2	
IIIB	-	1	
IIIC1	-	-	
IIIC2	1	6	
IVA	1	2	
IVB	-	1	
Myometrial invasion			
<50%	30	12	
>50%	19	8	
Lymphovascular invasion			
Yes	9	11	
No	40	9	

Final diagnosis was performed on the basis of the highest level of diagnostic evaluation. For all EC cases, this was the histopathological analysis of the surgical specimen after hysterectomy. For non-EC controls, final diagnosis was achieved with the histological examination of uterine aspirates (12%), or hysteroscopy-guided biopsies (88%).

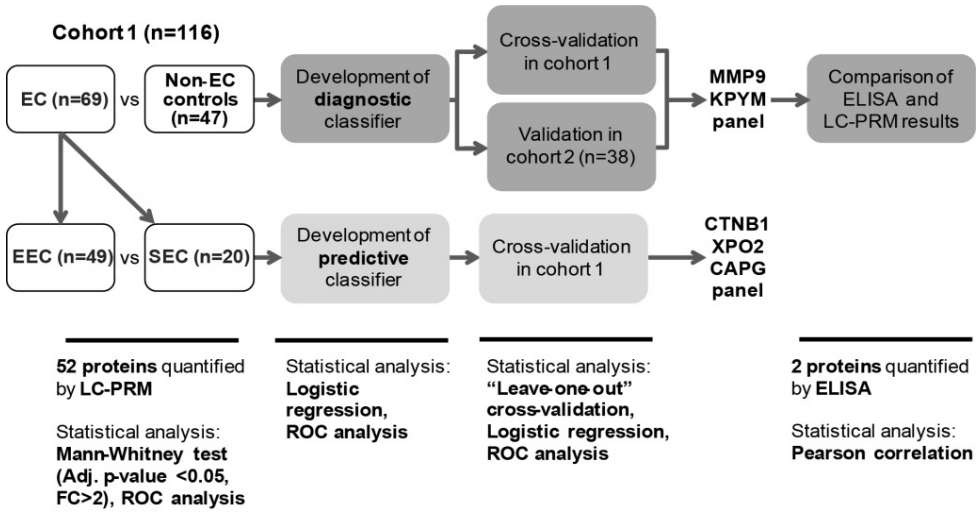


Figure 25. Overview of the validation study design. Schematic representation of the steps followed for the development of the EC diagnostic panel (upper part), and the predictive panel (bottom part). FC, fold change.

LC-PRM analysis

The general workflow followed in this study is shown in Figure 25. The fluid fraction of uterine aspirates of 116 patients were processed in duplicate and analyzed by PRM acquisition. In each sample, the relative abundance of the 52 candidate biomarkers prioritized in Chapter 3 was measured by the analysis of the same 98 peptides (~2 peptides per protein). A spectral similarity score was calculated to confirm the identity of the peptides and to detect interferences in the MS signal. Five peptides not detected in more than 50% of the samples were removed, leading to a total of 51 proteins robustly measured with 93 peptides. This data set included 21,576 measured pairs of endogenous/synthetic peptides (L/H) and 92.7% were validated by spectral matching. The rejected values were due to different causes including MS signal below the limits of detection that were replaced by an estimation of the background (6.8% cases), and high signals showing interferences that had to be manually reviewed (0.5%). The reproducibility of the analytical workflow was evaluated by duplication of the sample preparation and the coefficient of variation between duplicates (CV %) was below 15% for 98% of the detected peptides, with an average of 3.6%, confirming the high reproducibility of the full process (Figure 26A). Moreover, peptides derived from the same protein showed a high correlation, and consequently, they could be used interchangeably to evaluate the protein levels (Figure 26B). Only three proteins (KPYM, OSTP and ROA2) displayed a correlation coefficient below 0.85, which was due to

the specificity of the monitored peptides to different isoforms. In those cases, their peptides were considered as derived from different proteins.

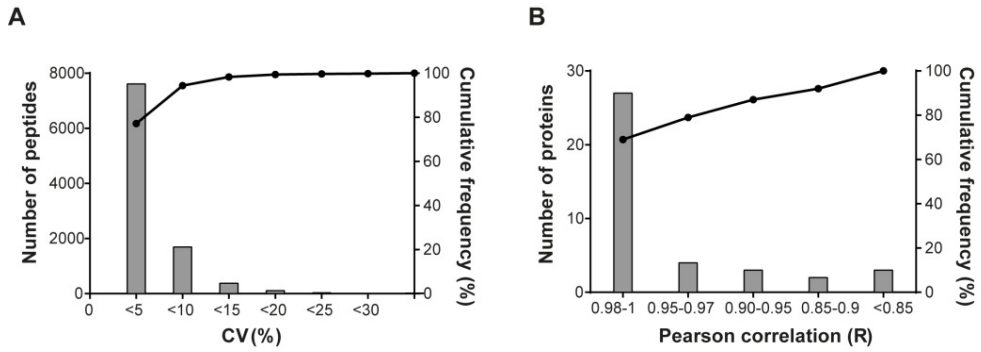


Figure 26. LC-PRM data quality control. A. Coefficient of variation (CV%) of the peptide signals quantified in the 116 uterine aspirates processed and analyzed by LC-PRM in duplicates. **B.** Plot representing the Pearson correlation coefficients between the peptides coming from the same protein.

Diagnostic biomarkers

The relative concentration levels of the 51 proteins were compared in 116 uterine aspirate samples belonging to 69 EC patients and 47 control patients in order to identify the proteins that optimally allows for the detection of EC. A total of 28 proteins showed significantly higher levels in the fluid fraction of uterine aspirates from EC patients (adjusted p-value <0.05, and fold change >2) and presented high accuracy to individually discriminate between EC and control cases (AUC values higher than 0.75) (Table 15). These results further confirmed previous results described in Chapter 3 in which the levels of these proteins were measured in a simplified set of only postmenopausal women including 20 EEC cases and 18 age-matched controls (Table 15). The best proteins in that dataset were also the best in this wider cohort of patients. Although postmenopausal women represent the vast majority of patients with EC, up to 14% of EC patients will be in the premenopause [69] and should be considered when evaluating biomarkers for EC diagnosis. Consequently, 21% patients included in the present study were premenopausal patients. Moreover, less frequent but more aggressive SEC cases were also included. Importantly, the validated biomarkers showed potential to detect EC independently of the endometrial status and the histological type of the EC cases. The five best individual biomarkers, measured by the AUC values, were LDHA with 0.91 (95% CI, 0.856-0.957), KPYM isoform M1-M2 with 0.90 (95% CI, 0.841-0.953), MMP9 with 0.89 (95% CI, 0.827-0.950), NAMPT with 0.88 (95% CI, 0.824-0.942), and SPIT1 with 0.88 (95% CI, 0.814-0.948).

Table 15. Statistical results of the 28 proteins showing significant differences between EC patients (n=69) and non-EC control women (n=47). Proteins with potential as diagnostic markers are shown. All of them have an adjusted p-value <0.05, fold change >2 and AUC>0.75. The performance of these proteins as early stage EC biomarkers was assessed by comparing EC cases in the early stage IA versus non-EC controls. The statistic results of the study performed in Chapter 3 with these proteins in a limited cohort of 38 age-matched patients is shown. Note: ** Performance of KPYM is isoform specific. FC, fold change.

Protein name	COHORT 1 (n=116)						COHORT 2 (n=38) (Chapter 3)		
	FC All EC/ All non-EC	Adjusted p-value	AUC	FC EC stage IA/ All non-EC	Adjusted p-value	AUC	FC EC/ non-EC	Adjusted p-value	AUC
LDHA	5,49	1,E-11	0,91	5,34	2,E-07	0,89	6,21	1,E-04	0,91
KPYM: Isoform M1-M2	5,67	1,E-11	0,90	5,48	2,E-07	0,90	5,37	1,E-04	0,91
**KPYM: Isoform M1-M3	3,39	9,E-05	0,72	2,82	2,E-02	0,66	3,11	1,E-02	0,75
MMP9	11,40	2,E-11	0,89	9,55	2,E-07	0,89	5,68	1,E-04	0,91
NAMPT	3,84	4,E-11	0,88	3,61	2,E-06	0,85	4,24	3,E-04	0,88
SPIT1	3,92	5,E-11	0,88	3,49	6,E-07	0,87	3,35	1,E-04	0,93
CADH1	3,33	5,E-11	0,88	3,44	5,E-07	0,87	3,75	9,E-05	0,94
ENOA	3,66	1,E-10	0,87	3,55	4,E-07	0,87	3,78	1,E-04	0,92
PERM	8,39	4,E-10	0,86	9,02	2,E-07	0,90	14,14	6,E-05	0,97
CAPG	3,74	8,E-10	0,85	3,20	8,E-07	0,86	3,56	5,E-04	0,85
CH10	3,08	1,E-09	0,85	2,98	8,E-06	0,82	2,29	6,E-03	0,77
CTNB1	3,89	2,E-09	0,84	3,85	2,E-06	0,84	4,19	3,E-04	0,88
K2C8	3,02	2,E-09	0,84	2,98	2,E-06	0,84	3,61	3,E-04	0,88
CLIC1	2,91	4,E-09	0,84	2,86	2,E-06	0,84	2,79	5,E-04	0,86
PDIA1	2,68	4,E-09	0,83	2,59	4,E-06	0,83	3,30	3,E-04	0,89
PRDX1	2,85	5,E-09	0,83	2,70	2,E-06	0,85	4,21	2,E-04	0,90
CD44	2,71	6,E-09	0,83	2,51	2,E-06	0,84	2,58	4,E-04	0,86
MIF	2,85	9,E-09	0,83	2,77	1,E-05	0,81	4,21	3,E-04	0,88
FABP5	3,19	1,E-08	0,82	2,73	2,E-06	0,84	3,89	6,E-04	0,85
XPO2	4,68	2,E-08	0,81	3,92	8,E-05	0,77	4,03	1,E-03	0,83
TPIS	2,35	7,E-08	0,80	2,37	2,E-05	0,80	2,93	3,E-04	0,87
CASP3	3,49	1,E-07	0,80	3,52	5,E-06	0,83	4,91	2,E-04	0,91
GSTP1	2,88	5,E-07	0,79	2,74	3,E-05	0,80	2,91	2,E-03	0,82
ANXA1	4,06	7,E-07	0,78	5,03	8,E-06	0,82	4,80	3,E-03	0,80
NGAL	3,63	3,E-06	0,77	4,27	3,E-05	0,80	4,41	4,E-03	0,79
ANXA2	3,22	3,E-06	0,76	3,72	6,E-05	0,79	4,83	4,E-04	0,87
GTR1	3,17	6,E-06	0,76	2,89	3,E-03	0,71	1,47	3,E-02	0,71
OSTP: Isoform A, B, D	2,28	7,E-06	0,76	1,71	1,E-02	0,68	9,02	4,E-04	0,87
MUC1	2,38	7,E-06	0,76	1,89	2,E-04	0,77	3,60	1,E-03	0,84

❖ Early-stage EC diagnostic biomarkers

Early detection of EC is directly associated to high rate of survival for the patients. Thus, the suitability of the 51 proteins as markers for early detection of EC was evaluated. A total of 30 EC cases diagnosed at IA stage according to the FIGO staging, i.e., tumors confined to the uterine corpus with less than 50% myometrial invasion [119,120], were compared with the 47 non-EC controls. All proteins previously identified for EC diagnosis were found differentially abundant with adjusted p-value<0.05 and fold-change >2, except for OSTP and MUC1

(Table 15). Remarkably, the five proteins showing the best performance in discriminating all EC cases from non-EC controls were also able to accurately discriminate the very initial stage of the disease with AUC values over 0.84 (Table 15). In addition, PERM increased its individual potential to detect early stage EC as follows: AUC_{EC} was 0.86 (95% CI, 0.786-0.929) and $AUC_{early\text{-}EC}$ was 0.90 (95% CI, 0.841-0.969).

Table 16. List of proteins showing statistical differences between patients diagnosed with endometrial hyperplasia (n=9) and the other non-EC controls (n=38) with adjusted p-value <0.05 and fold change > 2.

	FC (non-EC/ Hyperplasia)	Adjusted p-value	AUC
NAMPT	7,00	3,2E-02	0,89
ENOA	2,85	2,3E-02	0,88
CATD	3,40	1,8E-02	0,87
GSTP1	2,76	2,4E-02	0,85
KPYM	2,75	4,1E-02	0,82
PPIA	2,46	3,9E-02	0,81
CAPG	2,57	4,2E-02	0,80
TAGL	2,34	4,7E-02	0,79
MMP2	3,52	4,1E-02	0,78
MIF	2,13	4,3E-02	0,78
LEG3	2,05	4,0E-02	0,77
CASP3	3,20	4,6E-02	0,76
CLIC1	2,54	4,4E-02	0,76
LEG1	2,20	4,5E-02	0,76
FSCN1	2,89	4,6E-02	0,76
CTNB1	2,02	4,5E-02	0,75

❖ Endometrial hyperplasia

Endometrial hyperplasia is a thickening of the endometrium caused by the excess of estrogen stimuli. Although it is a benign disease, it is considered a precursor lesion of EC, and should be distinctively diagnosed [13]. As described above, our biomarkers allowed for the accurate discrimination between EC and non-EC controls, including hyperplasia cases (Figure 27A and Table 15). Interestingly, 16 proteins were found differentially abundant with adjusted p-values <0.05 and fold changes higher than 2 (Table 16) between hyperplasias (n=9) and the other non-EC controls (n=38). Four of them, NAMPT, ENOA, CATD and GSTP1 showed AUC higher than 0.85. These biomarkers, if validated, open the avenue to individually diagnose hyperplasias from EC and other benign conditions.

❖ Diagnostic biomarker panel

By using a homemade R script, individual EC biomarkers were combined into all possible panels from two to five proteins. The combination of MMP9 and KPYM significantly

improved the diagnostic power of the individual biomarkers, achieving an AUC value of 0.96 (95%CI, 0.94-0.99), with 94.2% sensitivity and 87.2% specificity (Figure 27B-D). The addition of more proteins only enhanced marginally the performance of the 2-protein panel and therefore, were not considered. The robustness of the panel was assessed using two methods: a "leave-one-out" cross-validation procedure performed over the set of 116 samples (cohort 1), and a validation of the model over the independent set of 38 samples from Chapter 3 (cohort 2). The results of the logistic regression model underlying the ROC analysis in both independent datasets are shown in Figure 27D. The MMP9-KPYM panel achieved sensitivity (specificity) of 88.86% (85.1%) and 100% (83.33%) in the first and second cross-validation, respectively.

The diagnostic biomarker panel can differentiate EC from non-EC control patients with high accuracy in the fluid of the uterine aspirate samples. Importantly, in the cohort of patients included in this study, the histopathological examination of uterine aspirates did not provide a proper diagnosis to 17% of women, either due to insufficient material or incorrect diagnosis. Moreover, the misdiagnosed patients were EC patients, which were diagnosed with no malignancy. For this 17% of women, the biomarker panel combining MMP9 and KPYM achieved a correct diagnosis. Consequently, our biomarker panel was able to perfectly complement the current diagnostic procedure to provide final diagnosis to all women by using this minimally-invasive sampling.

❖ **Transferability to ELISA**

In order to evaluate the feasibility of developing an assay that could be more easily deployed in a clinical environment, we assessed the transferability of the diagnostic panel MS-based results to ELISA assays. The levels of the two proteins of the diagnostic panel, MMP9 and KPYM, were quantified by commercially available ELISA kits and the correlation with the results obtained by LC-PRM was evaluated. The quality of the ELISA performance was assessed by the median overall CV (%), which was 1.8% for MMP9 and 6.0% for KPYM. Absolute levels in the ELISA (ng/ml) and relative levels (light/heavy ratios) in the PRM acquisition showed a linear correlation for both biomarkers, with an R^2 value of 0.93 for MMP9 and 0.61 for KPYM (Figure 27E).

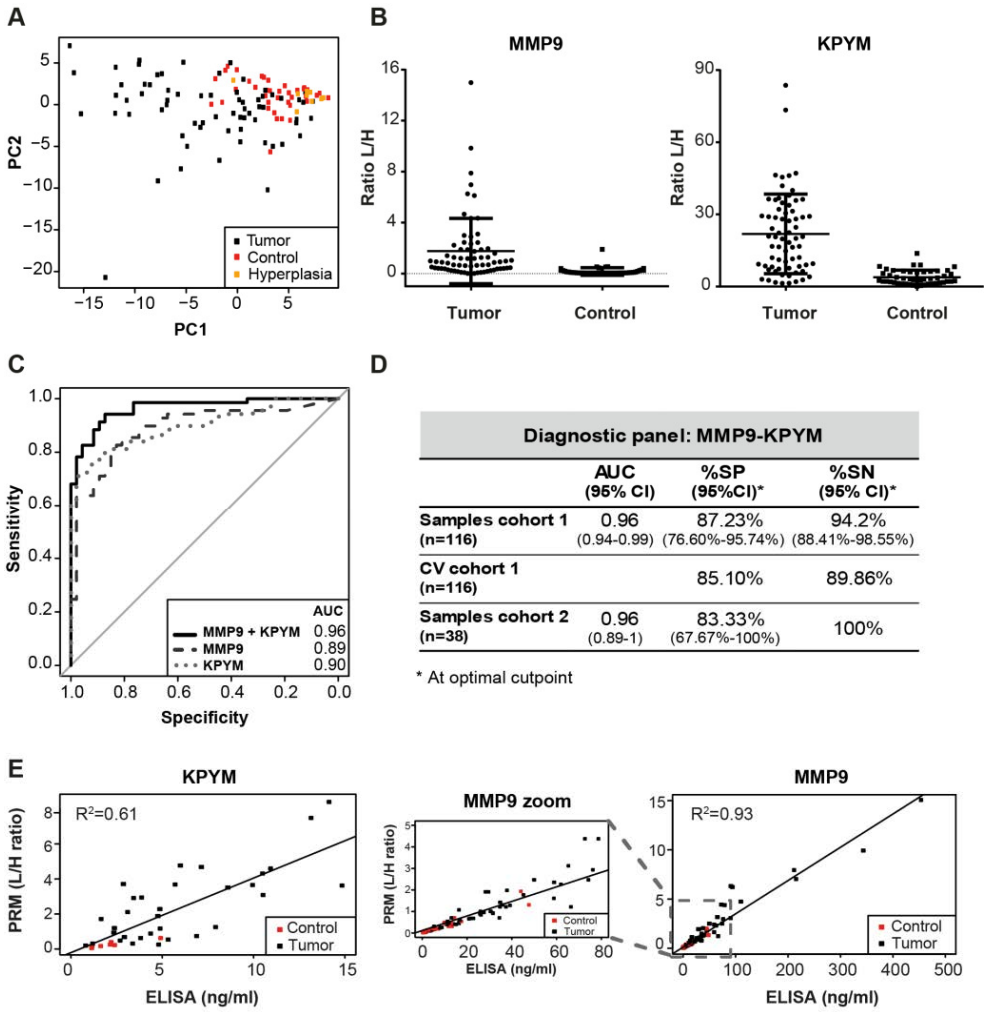


Figure 27. Diagnostic performance of biomarkers in discriminating EC patients from non-EC controls in the fluid fraction of uterine aspirates. **A.** Principal Component Analysis (PCA) plot constructed with the data of the differential proteins of the 116 patients included in the study. The plot clearly shows that most of the variance plotted in the X axis allows for the differentiation of EC patients compared to endometrial hyperplasias and other non-EC controls. **B.** Scattering plots depicting the distribution of the light/heavy (L/H) ratios obtained by LC-PRM across the 69 EC patients and 47 controls of the two proteins that compose the diagnostic panel. **C.** ROC curves of EC versus non-EC controls women for the individual proteins of the panel and the 2-protein panel. **D**. Summary table of the results of the logistic regression model adjusted to the data of cohort 1 (n=116) underlying the ROC analysis. The robustness of the panel was assessed by performing a "leave-one-out" cross-validation (CV) for each sample in the cohort 1, and by applying the model to an independent cohort 2 (n=38) from the previous study described in Chapter 3. SN (sensitivity) and SP (specificity) values obtained after "leave-one-out" cross-validation in cohort 1 and validation in cohort 2 are shown. **E.** Correlation between LC-PRM and ELISA results evaluated in 105 patients for MMP9 (68 EC patients, in black; and 37

control patients, in red), and 39 patients for KPYM (32 EC patients, in black; and 7 control patients, in red).

Predictive biomarkers

As explained in the introduction, EEC is the most common histology in EC and has a good prognosis when compared with non-endometrioid EC cases (NEEC) [69]. NEEC represents less than 20% of all EC cases but accounts for more than 50% of recurrences and deaths from EC. Among NEEC, the serous EC (SEC) is the most common subtype. We investigated the abundance of the 51 proteins in the cohort of 49 EEC and 20 SEC cases. The levels of nine proteins were significantly increased in uterine aspirate samples from EEC patients (adjusted p-value<0.05) (Figure 28A). Among those, six proteins had fold change higher than 2 and presented the highest individual AUC values: PIGR with 0.85 (95% CI, 0.734-0.958), CAYP1 with 0.83 (95% CI, 0.725-0.942), CTNB1 with 0.78 (95% CI, 0.670-0.895), SG2A1 with 0.77 (95% CI, 0.661-0.880), VIME with 0.76 (95% CI, 0.645-0.881), and WFDC2 with 0.74 (95% CI, 0.624-0.855).

❖ **Predictive biomarker panel**

Following the same procedure as described before, all possible combinations of two and three proteins were evaluated among the diagnostic and predictive biomarkers to identify panels of proteins that will improve the outcome of individual biomarkers. A combination of three proteins, consisting of CTNB1, XPO2 and CAPG was the best-performing panel to discriminate between EEC and SEC in the fluid of uterine aspirate samples with an AUC of 0.99 (95% CI, 0.90-1) (Figure 28B). This panel achieved 95% (CI 95% 85%-100%) sensitivity and 95.9% (CI 95% 89.8%-100%) specificity (Figure 28C). After completion of the "leave-one-out" cross-validation the values were 95% sensitivity and 89.8% specificity.

DISCUSSION

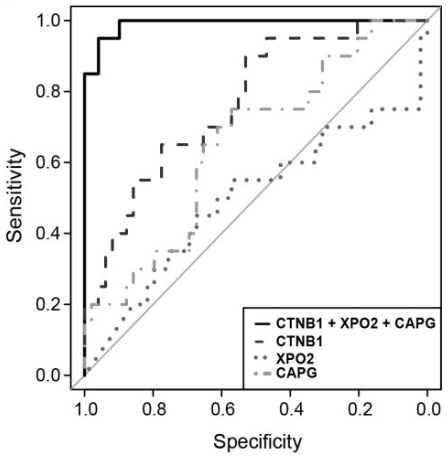
In this chapter we defined two protein biomarker signatures to detect EC and to distinguish between EEC and SEC histologies by using a liquid biopsy obtained from the female genital tract (i.e., the fluid of uterine aspirates). The current diagnostic procedure is based on the histological examination of the limited cellular content in this sample and it is associated to important drawbacks: an average of 22% of undiagnosed patients due to histologically inadequate samples [109], and up to 50% of incorrect histotype and/or grade assignment of EC cases [140]. Our approach help to overcome these limitations by the identification of biomarkers in the fluid fraction of uterine aspirates.

A

EEC vs SEC			
Protein ID	FC (EEC/SEC)	Adjusted p-value	AUC
PIGR	5.65	7.E-04	0.85
CAYP1	4.48	7.E-04	0.83
CTNB1	2.70	5.E-03	0.78
SG2A1	34.40	4.E-03	0.77
VIME	2.40	8.E-03	0.76
CADH1	1.95	2.E-02	0.74
WFDC2	4.55	2.E-02	0.74
CD44	1.76	5.E-02	0.71
LEG3	1.61	5.E-02	0.71
LEG1*	1.49	5.E-02	0.70
CAPG*	-1.67	2.E-01	0.67

*p-value < 0.05

B



C

Prognostic panel: CTNB1-XPO2-CAPG			
	AUC (95% CI)	%SP(95%CI)*	%SN (95% CI)*
Samples cohort 1 (n=69)	0.99 (0.98-1)	95.92% (89.80-100%)	95.00% (85.00%-100%)
CV cohort 1 (n=69)		89.80%	85.00%

* At optimal cutpoint

Figure 28. Predictive performance of biomarkers in classifying EC cases in the most prevalent histological subtypes, associated with different surgical treatments, in the fluid fraction of uterine aspirates. A. List of proteins showing statistical differences between patients diagnosed with EEC (n=49) and SEC tumors (n=20) with adjusted p-value <0.05. Two proteins, LEG1 and CAPG, showed an adjusted p-value > 0.05, but p-value<0.05. **B.** ROC curves of EEC versus SEC cases for the individual proteins of the panel and the 3-protein panel. **C.** Summary table of the results of the logistic regression model adjusted to the data from the 69 patients underlying the ROC analysis. The robustness of the panel was assessed by performing a "leave-one-out" cross-validation for each sample in this cohort of patients.

The diagnostic panel is composed of two proteins. Metalloproteinase 9 (MMP9) is involved in the release of tumor promoting agents and in the degradation of extracellular matrix, that may promote cell migration and invasion, favoring tumor metastasis [324,325]. Increased levels of MMP9 have been detected by immunohistochemistry in endometrial tissue from EC patients; whereas lower levels have been reported in EC patients compared to healthy women in serum samples [317,326]. Pyruvate kinase (KPYM) is a key enzyme in glycolysis and is also involved in gene transcription [327]. An increased aerobic glycolytic rate is a metabolic hallmark of malignant cells [328], and increased levels of KPYM have been described in several cancer types. In EC, higher levels of this protein have been reported in endometrial tissue from EC patients compared to controls [318,329]. When combined, MMP9 and KPYM

form a powerful panel that can detect EC with 94.2% sensitivity and 87.2% specificity in the fluid fraction of minimally-invasive uterine aspirates. This molecular panel, combined with the current diagnosis based on the histological examination of the cells in these samples, permitted to achieve diagnosis in 100% of patients in our dataset. Consequently, implementation of this panel is expected to impact on the clinical scenario by precluding the use of subsequent invasive sampling methods, (i.e., dilatation and curettage or hysteroscopy).

A clinical strength of this investigation is that women enrolled in the study covered the broad variability of women entering the EC diagnostic process. Regarding EC patients, the two most common histologies, EEC and SEC cases, were included. Non-EC patients covered all women with suspicion of EC mainly due to AUB or thickening of the endometrium [95,101]. This included women suffering from benign pathological conditions (mainly polyps, myomas, and endometrial hyperplasia), and women with normal endometrium. Moreover, this study included both premenopausal women with functional endometrium and postmenopausal women mostly presenting an atrophic endometrium. Despite the molecular differences that are known to exist between EEC and SEC tumors [74], and between endometrial tissues of women with different hormonal status [330], our diagnostic biomarkers allowed for the accurate differentiation between EC patients and non-EC women independently of these factors.

A limitation of this investigation is that we here reported several proteins that allowed for the discrimination of endometrial hyperplasias from both EC and control women. However, hyperplasias include a broad range of lesions with very different progression risk to carcinoma, from 1-3% for women with non-atypical simple endometrial hyperplasia, to almost 30% for patients with complex atypical hyperplasia [13]. Due to the limited number of cases in this study, division into risk groups was not feasible. Therefore, further exploration of the levels of these biomarkers in the different subgroups is needed.

Regarding the role of the studied proteins to improve the risk group assignment of EC patients, current major risk stratification systems, such as the European Society for Medical Oncology (ESMO) classification, focus on the histology, grade, myometrial invasion and lymphovascular invasion of the tumors. As most information is not available preoperatively, histological subtype and grade become key factors for risk group assignment and for the determination of the extent of the surgical staging procedure [140]. Our data showed that the combination of three proteins (CTNB1, XPO2 and CPG) allowed for the accurate discrimination between EEC and SEC histological EC types with a sensitivity of 95.0% and specificity of 95.9%. These three proteins have been previously related to EC. Beta-catenin (CTNB1) has an important role in epithelial cell-cell adhesion, and in the transcription of essential genes responsible for cellular proliferation and differentiation in the Wnt signaling

pathway [331]. In concordance with our observations, CTNB1 has been described in EEC tissue specimens, but not in SEC tumors [332,333]. Macrophage-capping protein (CAPG) modulates cell motility by remodeling actin filaments. It is involved in cell migration and invasiveness in several type of cancers[334]. Unlike CTNB1, higher levels of CAPG have been reported in more aggressive SEC tumors compared to EEC cases at tissue level, in agreement with our results in uterine aspirates. Finally, exportin-2 (XPO2), also known as cellular apoptosis susceptibility protein, has a role in the mitotic spindle checkpoint. Depletion of XPO2 leads to cell-cycle arrest and, consequently, it has been associated with tumor proliferation [335]; although it has also been related to tumor invasion and metastasis [336]. Higher levels of XPO2 have been observed in many cancer types, including EC, and have been positively associated with a higher cancer grade and worse outcome of the patients [337]. Although no significant differences in XPO2 levels were observed between EEC and SEC cases in our study, its inclusion in the panel formed by CTNB1 and CAPG significantly improved its performance.

This study was however limited to identify protein signatures that accurately distinguish among tumor grades, since the number of patients available for this comparison was too low to achieve statistically robust results. This should be tackled in a near future to aid in the risk group assignment of EC patients. Moreover, current histopathological risk stratification systems are limited to predict the risk of recurrence. Additional molecular information is needed to improve these stratification systems and guide a more precise surgery. Although it is not yet applied in the clinical practice, the novel molecular classification of EC developed by The Cancer Genome Atlas (TCGA) have demonstrated higher prognostic accuracy [85]. The behavior of CAPG, XPO2 and CTNB1 expression in relation to the molecular classification was assessed using the available TCGA data [338]. *CAPG* and *XPO2* presented a significant mRNA upregulation in the serous-like group in contrast to the microsatellite instability and copy-number low subgroups that were characterized by a high number of *CTNB1* missense mutations. The POLE group presented mixed features, with a high number of alterations in XPO2 and CTNB1 (Figure 29). Although these findings are at genomic level, they open up a route to assess the potential of this proteomic signature to discriminate molecular subgroups and hence, further ameliorate the stratification of ECs.

Another important strength of this study is manifested in the technical viewpoint. We here employed the last generation of targeted MS method, the LC-PRM acquisition performed on a high-resolution accurate mass spectrometer. Unlike the vast majority of EC biomarker studies that use antibody-based approaches to test one or few proteins, this multiplexing targeted MS-based approach presents two clear advantages. First, it increases the chances of finding a clinically useful biomarker; and second, it is better suited for the development of protein panels since combinations among all the studied proteins can be evaluated. Indeed,

the best individual proteins do not necessarily perform well as a combination, since they can provide redundant information, as proved in both panels described in this study. Finally, the use of PRM acquisition facilitates the application of targeted MS to large cohorts of patients in comparison to the standard SRM acquisition [259,296]. He here measured 98 peptides in 232 samples (116 samples in duplicates). Apart from the easier method development, the decrease of interferences and the use of spectral matching significantly facilitated the data processing. Moreover, we achieved a precise quantification of the 93 peptides (average CV of 3.6%).

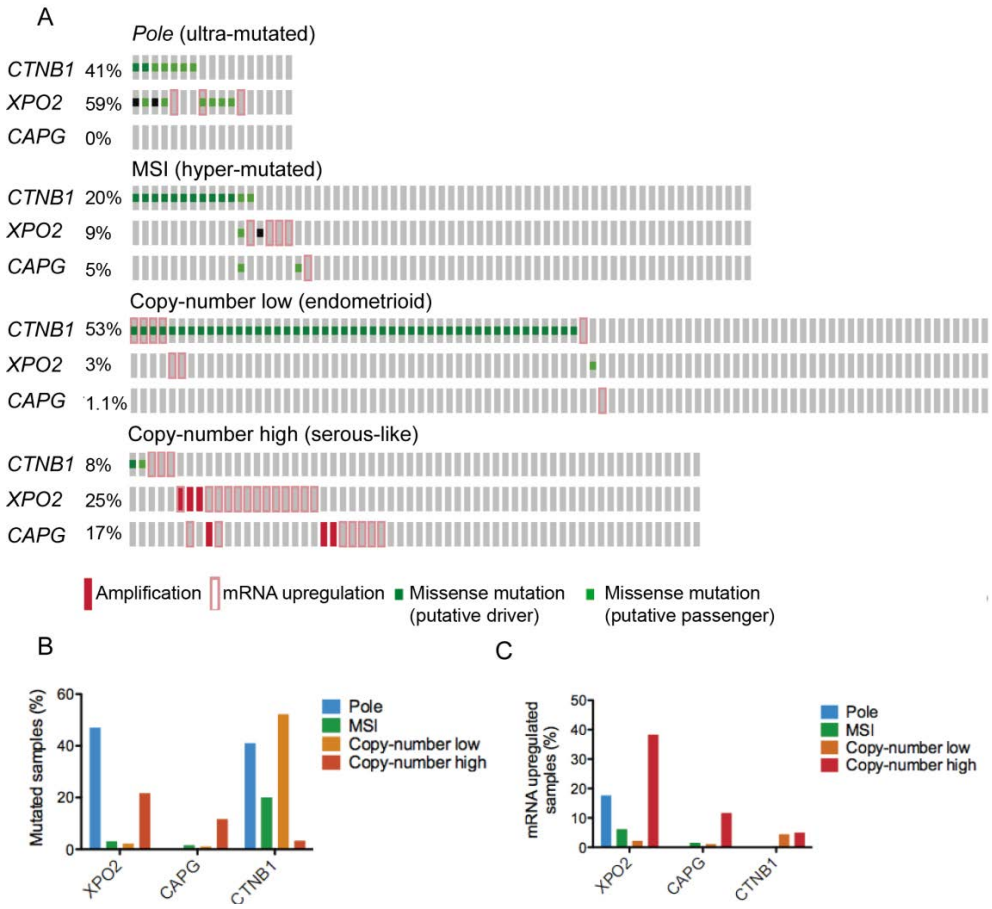


Figure 29. CTNB1, XPO2 and CAPG study according to the TCGA classification. A. Oncoprint of the genetic alterations identified in CTNB1, XPO2 and CAPG in the different TCGA subgroups. **B.** Percentage of CTNB1, XPO2 or CAPG mutated cases in each molecular subgroup. **C.** Percentage of cases with CTNB1, XPO2 or CAPG mRNA upregulation in each molecular subgroup.

As a first step towards clinical implementation, the diagnostic panel was evaluated by ELISA, since immunoassays continue to be the preferred method for clinical validation and further application in the clinical environment [199]. The commercially available ELISA kit against MMP9 yielded results perfectly matching the LC-PRM data, but the KPYM ELISA did not perform up to the same standard. As next steps, highly specific and reproducible immunoassays for the protein panels must be developed and the performance of the biomarker signatures must be validated in a prospective, multicentric study including a higher number of patients.

CHAPTER 5

Moving to the clinic.
Development of a prototype

The implementation of biomarker signatures into clinical practice requires not only a robust performance of the biomarkers, but also the development of a simple and straightforward assay. We have identified accurate diagnostic and predictive biomarker signatures by using the albumin & IgG depleted fluid fraction of uterine aspirates in combination with high-resolution mass spectrometry. We also confirmed the mass spectrometry results for MMP9 detection using a commercial ELISA, as a first step to approach the diagnostic signature to an accessible assay in the clinical practice. Following these results, we targeted two goals in this chapter: first, to simplify the ELISA assay of MMP9, which was accomplished in the context of a scientific collaboration with the Diagnostics Nanotools Group (CIBBIM-Nanomedicine, Vall d´Hebrón Institute of Research) led by Dr. Baldrich; and second, to simplify the sample preparation.

5.1. ASSAY SIMPLIFICATION

This section is partially described in the article: "*ELISA simplification and shortening by using a polymeric signal amplifier. Application to MMP9 detection in plasma and uterine aspirates".*
de la Serna E, Martinez-Garcia E, García-Berrocoso T, Penalba A, Gil-Moreno, Colas E, Montaner J, Baldrich E (currently under revision).

SPECIFIC BACKGROUND

ELISAs are widely used as diagnostic tools in clinical laboratories [339]. This technique exploits antibodies (Ab) for the specific binding of the analyte of interest, coupled to an enzyme-mediated color change that is proportional to the analyte concentration in the sample. ELISA can be performed using a detection Ab modified with an enzyme label, such as horseradish peroxidase (HRP) [340]. Alternatively, indirect detection using a biotinylated Ab and a (strept)avidin-enzyme conjugate can be carried out [341]. However, the standard assay set-up entails a time-consuming multi-step procedure that would be difficult to operate using a low-cost point-of-care detection platform.

Engineered polymeric streptavidin-HRP complexes (polyHRP) consist of hundreds of enzyme molecules chemically coupled to each other or to a polymer backbone [342]. They have been employed as signal amplifiers in immunohistochemistry [343] and in ELISA to improve assay sensitivity. The use of polyHRP has allowed to develop ELISAs which are 3.8 to 20 times more sensitive than any alternatives using Ab-HRP or streptavidin-HRP [344–346]. Interestingly, Dr. Baldrich´s team demonstrated that the signal enhancement provided by polyHRP can be successfully employed to simplify and shorten pre-existing ELISA assays.

As a proof of concept, they converted the commercial ELISA kit for MMP9 detection that we used in Chapter 4, and that requires 5 hours of working time, into a shortened ELISA that can be performed in about 1 h.

In this context, we aimed to compare here the performance of the shortened ELISA *versus* the standard ELISA assay and the previous results obtained by LC-PRM to detect MMP9 in uterine aspirates from women suspected of having EC. Moreover, we evaluated the potential applicability of the shortened assay format to other platforms such as electrochemical biosensors, providing the basis for the development of point-of-care devices for EC diagnosis [347].

MATERIAL AND METHODS

Detection of MMP9 by ELISA

ELISA was performed using the human MMP9 DuoSet ELISA development kit (R&D Systems). The kit contains a capture monoclonal mouse anti-human MMP9 Ab (c-Ab), a detection biotinylated goat anti-human MMP9 Ab (d-Ab), a recombinant human MMP9 calibrator, and streptavidin-HRP.

Unless otherwise stated, incubations were performed with 100 µL of solution per well, inside an incubator at 24ºC, and were followed by three consecutive washes with PBS-T (200 µL per well), consisting of PBS 1X (Gibco) supplemented with 0.05% of Tween-20 (Sigma-Aldrich).

❖ **MMP9 detection by the "standard" ELISA**

The standard ELISA was carried out according to the provider's instructions. Briefly, microtiter plates were modified overnight with c-Ab (1 µg/mL in PBS). After washing, the plates were blocked with PBS supplemented with bovine serum albumin (PBS-1% BSA) for 1 h, followed by a 2 h incubation with recombinant MMP9 (30-2000 ng/mL, in triplicate) for the calibration curve and the uterine aspirate samples to be analyzed, and a 2 h incubation with d-Ab (100 ng/mL in PBS-2% BSA). After washing, streptavidin-HRP (Thermo Scientific) at 1/200 in PBS-1% BSA was added and incubated for 20 min. The plates were washed four times with PBS-T, 100 µL of the enzymatic substrate (3,3′,5,5′-Tetramethylbenzidine (TMB) liquid substrate supersensitive for ELISA from Sigma-Aldrich) was added per well, and the reaction was allowed to proceed for 20 min before being stopped with 50 µL of 1 M sulphuric acid (Sigma-Aldrich) per well and measured at 450 nm using a Sunrise plate reader (Tecan) (Figure 30A).

❖ **MMP9 detection by the "shortened" ELISA**

The "shortened" ELISA consisted of the modification of the microtiter plate with c-Ab and BSA following the standard protocol, followed by a 30 min simultaneous incubation with recombinant MMP9 (30-2000 ng/mL, in triplicate) for the calibration curve and the uterine aspirate samples together with d-Ab (100 ng/mL in PBS-2% BSA). The plate was then washed, incubated for 10 min with streptavidin polyHRP (diluted 1/10000 in PBS-1% BSA) and washed four more times with PBS-T. Detection was performed as described for the standard ELISA (Figure 30B).

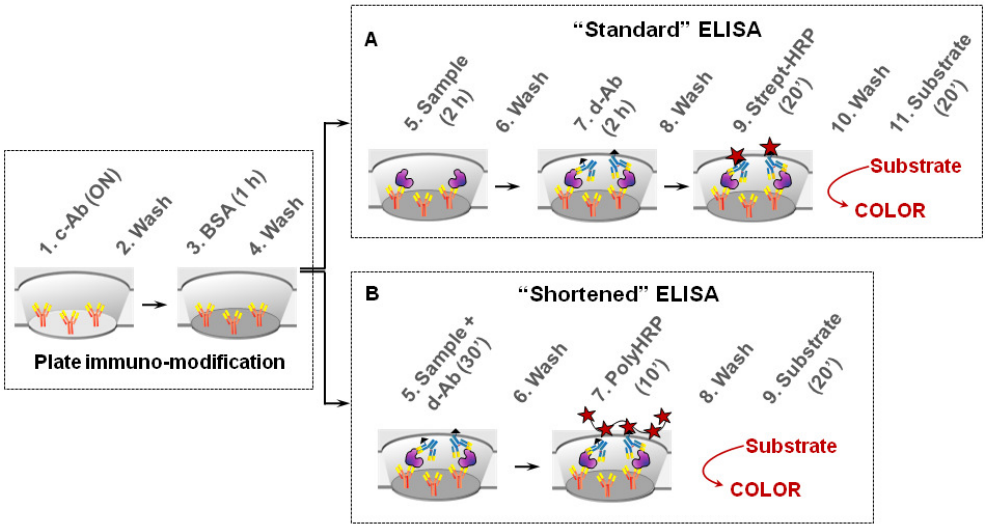


Figure 30. Schematic representation of the standard ELISA (A) and the shortened assay (B). In both cases, the microtiter plate is modified with c-Ab overnight (ON) and blocked with BSA for 1 h. However, while the standard ELISA continues with a series of incubations that last for nearly 5 h, the shortened assay can be carried out in about 1 h.

Detection of MMP9 by electrochemical immunoassays

The same bioreagents included in the human MMP9 DuoSet ELISA development kit used in the ELISA were employed to produce the electrochemical immunosensors.

Unless otherwise stated, incubations were performed with 50 µL of solution per chip, except for the capture antibody, in an incubator at 24ºC and inside a humid chamber to prevent reagent evaporation, and were followed by three consecutive washes with 100 µL of PBS-T per sensor.

For the production of electrochemical immunosensors we used commercial carbon screen-printed electrodes (C-SPE; DropSens, Spain). Each device contained a carbon working electrode (4 mm diameter), a carbon counter electrode and a silver pseudoreference electrode, all stamped onto a ceramic substrate (L33 x W10 x H0.5 mm). Before their utilization, the C-SPE were serially washed with ethanol and water, dried under an air flow, and then characterized by cyclic voltammetry in 0.1 M KCl, 0.1 mM K_4 $Fe(CN)_6$. All the amperometric measurements were performed at room temperature using a µSTAT 8000P portable multipotentiostat controlled by the DropView 8400 software (Dropsens).

For sensor immune-modification, the working electrode of each chip was modified overnight with c-Ab (6 µl of a 12.5 µg/mL dilution in PBS). After washing, the chips were blocked with 60 µl of PBS-2% BSA for 45 min, washed again, and incubated with a mixture of the uterine aspirate samples and d-Ab (100 ng/mL in PBS-2% BSA), like in the shortened ELISA. The chips were then washed, incubated for 10 min with streptavidin polyHRP (diluted 1/10000 in PBS-1% BSA) and washed three times with PBS-T and one time with PBS diluted 1/10. For the electrochemical detection, 50 µL of PBS 1/10 was added onto the electrodes and the current was allowed to stabilize at 0 V *versus* the silver (Ag) pseudo-reference for 150 s. The measurement was halted, PBS was substituted by TMB and the measurement was resumed for other 500 s.

Electrochemical detection of the levels of MMP9 in each sample was based on the amperometric monitoring of the enzymatic product. That is, polyHRP reduces H_2O_2 into H_2O coupled to the oxidation of TMB in the substrate solution. Oxidized TMB is then reduced in real time at the electrode surface, which results in an electron flow of negative charge (from the electrode towards the TMB in solution). The current intensity is proportional to MMP9 concentration.

Processing of uterine aspirates

The sonicated and depleted fluid fractions of uterine aspirates from 16 women included in the validation study described in Chapter 4 (all women suspected of EC: 10 EC patients and 6 non-EC patients) were employed in this study. The depleted samples were diluted in PBS-BSA (1% in the standard ELISA, and 2% in the shortened ELISA and the electrochemical sensors). Three different dilutions (1/10, 1/100, 1/1000) were tested for each sample and each dilution was analyzed in duplicate by each assay, blinded to clinical data. Each ELISA experiment included its own calibration curve with triplicates run in parallel. In contrast, the electrochemical immunoassays could only be performed simultaneously in the 8 sensors available on the measurement equipment. Hence, only 8 out of the 16 samples were measured (4 EC patients and 4 controls) and an averaged calibration curve was obtained with data gathered from three independent experiments.

Data analysis

The averaged signals registered for each sample dilution were then interpolated in the corresponding calibration plot using a polynomic approximation to obtain post-dilution MMP9 concentrations. These were then multiplied by the appropriate dilution factor to estimate the MMP9 concentration in pre-diluted samples.

The linear correlation between the results obtained with the different approaches (i.e., shortened and standard ELISA, electrochemical immunosensors and previous LC-PRM results) was calculated using the Pearson correlation coefficient. For the differences between EC and control samples, a cut-off was established as the average concentration of MMP9 in non-EC samples, plus 3 times its standard deviation; the comparison of the levels of MMP9 between both groups was performed using the Student t-test.

RESULTS

Correlation between the standard and shortened ELISAs

The levels of MMP9 in 16 uterine aspirate samples from EC and control patients were measured in parallel using the shortened and the standard ELISAs. The results produced by the shortened ELISA presented a strong linear correlation with the results obtained with the standard procedure (R^2 = 0.97) (Figure 31A), and with the previous LC-PRM results (R^2 = 0.93) (Figure 31B).

The concentrations of MMP9 in the samples from EC patients were in average one order of magnitude higher (3-60 ng/mL in the shortened ELISA) than in the non-EC control patients (0.33-1.44 ng/mL) (Figure 31C). Accordingly, MMP9 could be quantified in EC samples using dilutions of 1/100 or higher, whilst in samples from non-EC women MMP9 could only be quantified consistently in 1/10 diluted samples. In spite of the limited number of samples, a tentative cut-off was established as the average concentration of MMP9 in non-EC samples, plus 3 times its standard deviation (cut-off = 3.14 and 2.60 ng/mL in the standard ELISA and the shortened ELISA, respectively; insert in Figure 31C). All samples from EC patients displayed concentrations of MMP9 significantly higher than the corresponding cut-off at a 99% confidence level (p-value < 0.01), independently of the assay used for their determination.

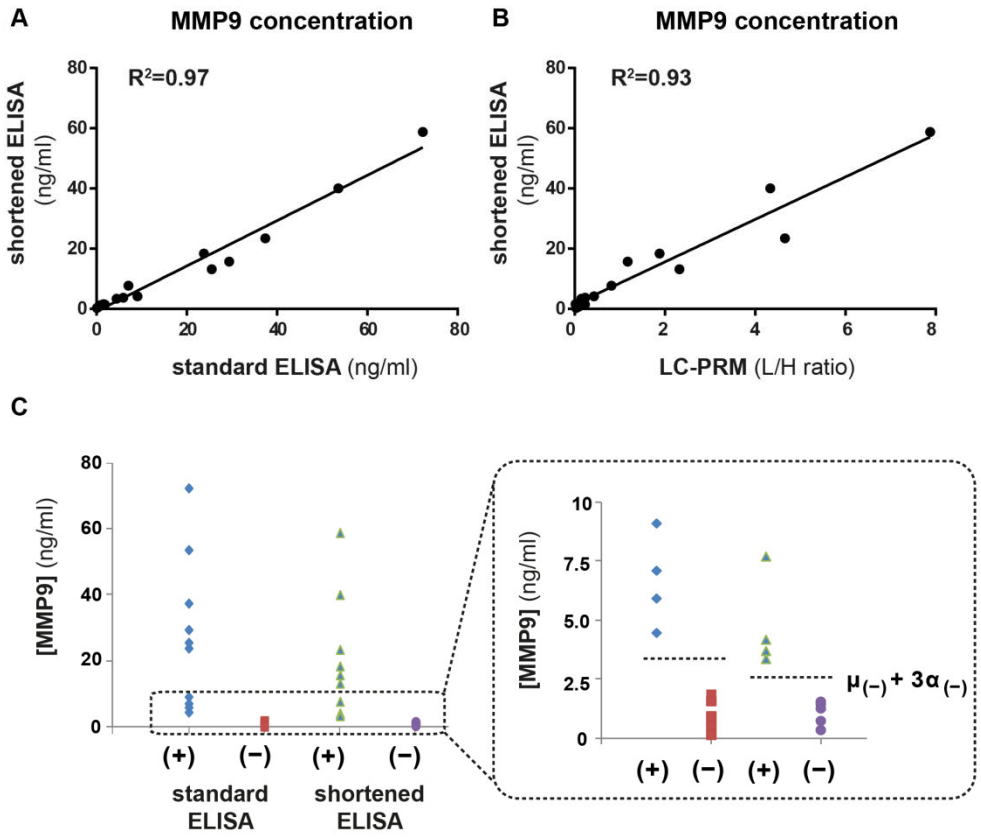


Figure 31. Correlation between the MMP9 levels measured by the standard ELISA and both the shortened ELISA and LC-PRM. A. Correlation between the MMP9 concentrations in uterine aspirate samples from 16 patients measured by the standard and the shortened ELISA assays. **B**. Correlation between the MMP9 concentrations in uterine aspirate samples from 16 patients measured by the shortened ELISA and the relative abundance of MMP9 in the same samples measured by LC-PRM. **C.** MMP9 concentrations classified according to patient status (+/– are presence/absence of EC) and the ELISA used. Right insert: amplification of C. The dotted lines illustrate the cut-off for each assay.

Correlation between the shortened ELISA and electrochemical biosensors

The levels of MMP9 in 8 of the previous uterine aspirate samples were measured using a similar shortened immunoassay protocol, applied this time to an electrochemical immunosensor format (Figure 32A), and compared to the paired results obtained with the shortened ELISA and with LC-PRM. A high correlation was achieved for both comparisons, with R^2=0.93 and R^2=0.96, respectively, as shown in Figure 32B-C.

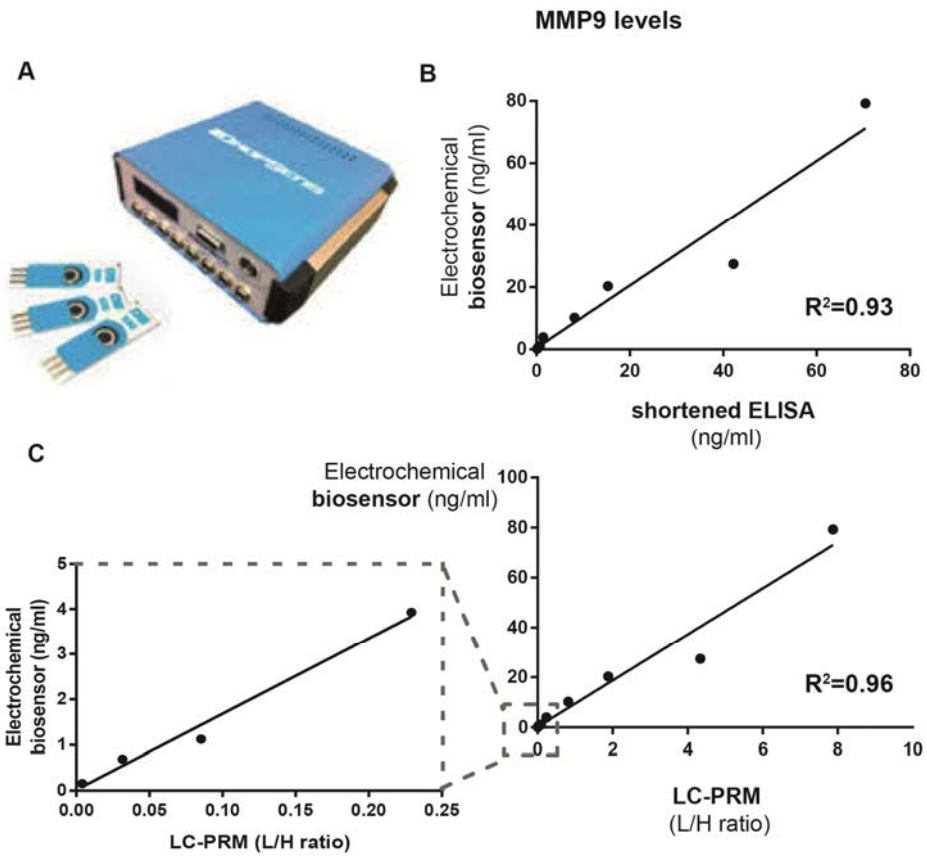


Figure 32. Correlation between the MMP9 levels measured by the electrochemical biosensor and both the shortened ELISA and LC-PRM. A. Equipment for electrochemical signal measurement. **B.** Correlation between the concentration of MMP9 in uterine aspirate samples from 8 patients measured by electrochemical biosensors and the shortened ELISA assays **C.** Correlation between the concentration of MMP9 in uterine aspirate samples from 8 patients measured by electrochemical biosensors and the relative abundance of MMP9 in the same samples measured by LC-PRM.

5.2. SIMPLIFICATION OF THE SAMPLE PREPARATION

SPECIFIC BACKGROUND

MS-based approaches achieve good sensitivity, but several methodologies including the depletion of abundant plasma proteins or the enrichment of the proteins of interest can be

required to improve the proteome coverage in complex clinical samples, particularly in the initial phases of the biomarker pipeline [230].

In this project, sonication followed by albumin & IgG depletion of the samples was selected as the best cost-effective option among other tested methods (not included in this thesis). This methodology enabled the detection of a higher number of potential biomarkers and a more precise quantification of those proteins by LC-MS.

However, once the final biomarker panels have been developed (Chapter 4), we aimed to evaluate the feasibility of using the raw fluid fraction of uterine aspirates (i.e., avoiding albumin & IgG depletion and sonication steps) for the quantification of the described diagnostic and predictive biomarker signatures.

MATERIAL AND METHODS

Patients recruitment and sample collection

A total of 39 patients entering the EC diagnostic process (20 women suffering from EC and 19 non-EC controls) were recruited in the Vall d'Hebron University Hospital (Barcelona, Spain) during 2016 to 2017. Informed consent forms, approved by the Vall d´Hebron Ethical Committee, were signed by all patients. From each patient, a uterine aspirate sample was collected as described in the previous chapters (page 99) and the fluid fraction of each sample was kept at -80ºC.

Sample preparation for LC-PRM analysis

Ten out of the 39 samples were divided in two identical aliquots: one was kept at -80ºC without any treatment before proteolysis (here referred to as "raw uterine aspirates"), and the another aliquot was processed as previously described in Chapters 3 and 4, with sonication and albumin & IgG depletion (here referred to as "processed uterine aspirates").

The 39 raw uterine aspirates and the ten processed uterine aspirates were submitted to the same proteolysis procedure described in the previous chapters (pages 101-102) with a few changes: i) a total amount of 12.5 µg of protein was digested per sample; ii) only the sample preparation of a pool of four samples was performed in triplicate in order to estimate the reproducibility of the procedure; iii) after digestion, samples were desalted onto a 25 mg tC18 96 wells plate (Waters).

LC-PRM setup and data processing

An equivalent of 250 ng of total protein per sample was analyzed by LC-PRM using a set-up similar to the previous chapters. However, for this experiment the Q Exactive Plus was

replaced by a Q Exactive HF and thus, the resolution was set to 60,000 at 200 m/z with a maximum fill time of 110 ms and an automatic gain control of 1E6 ions.

The settings for the extracted ion chromatograms (XICs) extraction in Skyline were adapted accordingly. Data processing based on spectral matching was performed as described in the previous chapter (pages 102-103).

Statistical analysis

The light/heavy area ratio of each peptide was extracted from Skyline. The statistical analysis was performed in SPSS (v20.0) (IBM, Armonk, NY, USA) and Graph Pad Prism (v.6.0) (GraphPad Software, La Jolla, CA, USA). The correlation between the levels of the proteins from the diagnostic (MMP9 and KPYM) and predictive panels (CAPG, CTNB1 and XPO2) was calculated using the Spearman correlation coefficient, which is based on the ranked values.

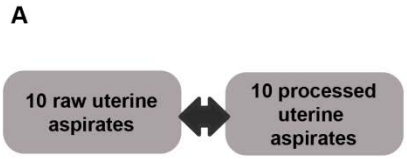
Comparison of the levels of MMP9 and KPYM between EC and control patients was performed using the nonparametric Mann-Whitney U test. P-values lower than 0.05 were considered statistically significant. The AUC values for MMP9 and KPYM were also estimated.

RESULTS

The levels of the 52 candidate biomarkers studied in Chapters 3 and 4 were measured by LC-PRM in the raw uterine aspirates and in their corresponding sonicated and depleted version, i.e., processed uterine aspirates, collected from 10 patients entering the EC diagnostic process. A high correlation (Spearman correlation coefficient over 0.82) was obtained for the proteins that form the diagnostic panel (MMP9 and KPYM) and the CAPG protein from the predictive panel (Figure 33A). However, the correlation coefficient was slightly lower for CTNB1, and it was not calculated for XPO2 due to the low signal obtained for this protein in most of the samples.

Taking into account the high correlation for the diagnostic panel, we then evaluated if the levels of those proteins were able to discriminate between 20 EC patients and 19 non-EC controls by using the raw uterine aspirates. As shown in Figure 33B-C, both proteins from the diagnostic panel, MMP9 and KPYM, showed significantly higher levels in the uterine aspirates from EC patients (p-value <0.05, and fold change >3) and presented high accuracy in individually discriminating between EC and control cases (AUC values of 0.82 for MMP9 and 0.93 for KPYM). These results further confirmed the results obtained in

Chapter 4, where MMP9 and KPYM achieved AUC values of 0.89 and 0.90, respectively, when measured in the processed uterine aspirates of an independent set of patients.



	Protein	Spearman coefficient
Diagnostic panel	MMP9	0.95
	KPYM	0.82
Predictive panel	CAPG	0.89
	CTNB1	0.76
	XPO2	N/A

B

Protein	FC (T/N)	P-value	AUC (95%CI)
MMP9	3.51	5.87E-05	0.82 (95%CI, 0.69-0.96)
KPYM	4.98	4.65E-06	0.93 (95% CI, 0.83-1)

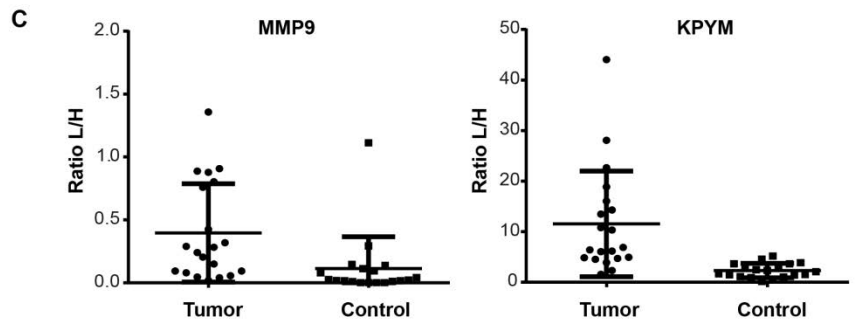


Figure 33. Feasibility of using raw uterine aspirates for the final measurement of the biomarker signatures. A. Correlation between the levels of the 5 proteins that form the diagnostic and predictive panels measured in raw and processed (sonicated+depleted) uterine aspirate samples from 10 patients. **B.** Summary table of the statistical results obtained from the comparison of the abundance of the diagnostic proteins in the raw uterine aspirates from 20 EC patients and 19 controls. **C**. Scattering plots depicting the distribution of the light/heavy (L/H) ratios of the two proteins that compose the diagnostic panel obtained by LC-PRM in the raw uterine aspirates from 20 EC patients and 19 controls.

DISCUSSION

In the first part of this chapter we demonstrated that signal amplifiers such as polyHRP can be exploited for immunoassay simplification and shortening of the assay time, which is essential for assay automation and integration in low-cost point-of-care detection platforms. Although a larger number of samples must be tested to validate these preliminary results,

we demonstrated that the shortened ELISA can be transferred to an electrochemical immunosensor format which offers important advantages compared to the ELISAs. First, upon immune-modification and blocking with BSA, immunosensors can be stabilized, dried and stored at 4ºC for weeks. Accordingly, they can be used straightforward for sample analysis. Moreover, they require less volume of sample and reagents than conventional ELISA. This is very important for those uterine aspirate samples where the material available is very limited. The reduction of the volume of reagents needed would also decrease the cost of each assay. Due to their high specificity, speed, portability, and low cost, biosensors offer important opportunities for decentralized clinical testing.

The preliminary results showed in the second part of this chapter suggest that the raw fluid fraction of uterine aspirate samples can be directly used for the measurement of the diagnostic panel, without any prior treatment. This is an important step towards the automation, simplification and reduction in the costs of the final assay. However, further investigations are needed regarding the predictive panel.

Global discussion and future perspectives

FROM THE BENCH...

Cancer is today one of the most common cause of death in the world, with over 8 million annual deaths, mainly due to the aging of the population [348]. Interest in precision diagnostics has significantly increased in the last years [176]. Early detection, i.e., finding tumors when they are still confined to the organ of origin, is crucial to favor complete resection of the tumor and to ensure a more effective treatment. The World Health Organization (WHO) has proposed that the combination of current therapies with effective early detection would significantly impact survival [349]. Since therapies for advanced stages are elusive, there is currently little evidence that survival among patients with disseminated disease will improve [350]. In this context, biomarkers have become a keystone of medical care to favour early detection. Their tremendous potential to transform the clinical practice has been extensively documented [176–178]. Apart from early diagnosis, an ideal biomarker could also provide clinicians with useful information including prognostication of disease progression or selection of the optimal therapy at an early stage of the disease.

In this post-genome era, clinical proteomics is receiving major attention, as reflected in the increasing number of published manuscripts in this field during the last years (Figure 34). The search for disease-specific biomarkers, and cancer biomarkers in particular, is one of the areas where proteomics is having a significant impact [351]. However, despite the important efforts and investments made to find protein cancer biomarkers of clinical utility, less than one protein biomarker per year has received FDA approval over the last 10 years [206].

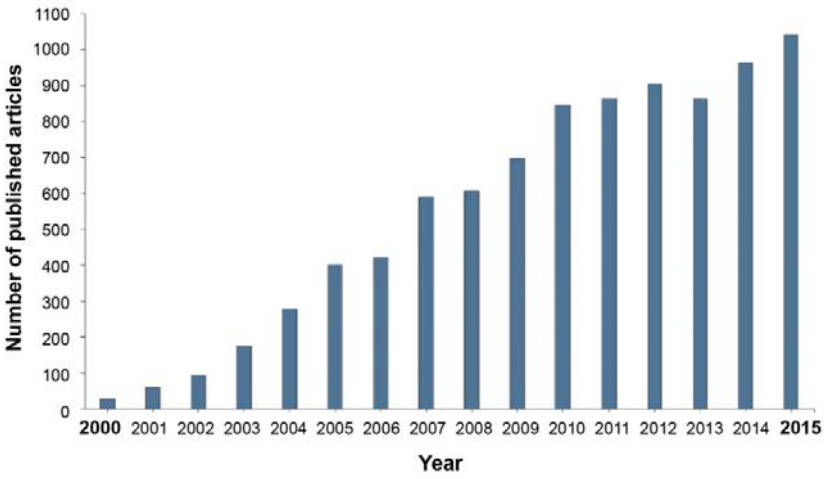


Figure 34. Evolution of clinical proteomics research. Number of published manuscripts from 2000 to 2015 based on the results obtained by searching the Web of Science database using the terms "clinical" and "proteom*". *Adapted from Nkuipou-Kenfack et al. 2017* [352].

Regarding EC, about 30% of the patients are still diagnosed at an advanced stage of the disease associated with a drastic decrease in the 5-year survival rate. Diagnosis is achieved by the histopathological examination of an endometrial biopsy, which is preferably obtained by a minimally invasive aspiration from the uterine cavity (i.e., uterine aspirate) [76,107]. Unfortunately, the limited number of cells available for examination in these biopsies is associated with two important drawbacks: i) about 22% of undiagnosed patients due to histologically inadequate specimens [109]; and ii) up to 50% of incorrectly assigned EC histotype and grade [140].

In this challenging but exciting context, the main goal of this thesis was to establish an effective workflow for the identification of specific and sensitive proteomic signatures able to: i) accurately detect all EC cases in uterine aspirate samples, and thus to reduce the number of more invasive sampling methods; and ii) provide clinicians more accurate preoperative information regarding EC histological type and tumor grade that could improve the risk stratification of the patients and further guide the selection of the optimal surgical treatment of EC patients.

An overview of the steps followed to achieve these objectives, described in the results section of this thesis, and future steps required for the implementation of the biomarker signatures in the clinical practice are shown in Figure 35.

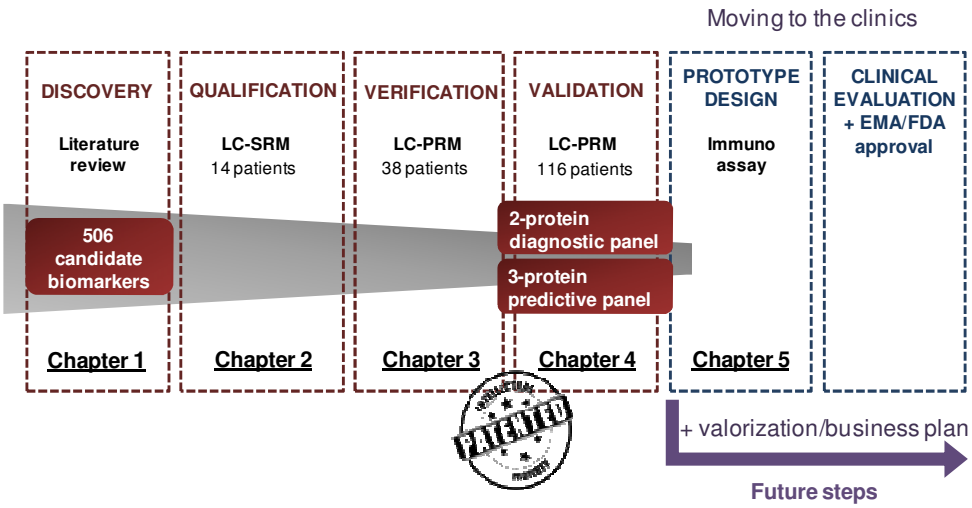


Figure 35. General overview of the biomarker identification process followed in the thesis and future steps.

In **Chapter 1** an extensive literature review was performed, and a list of 506 proteins associated with EC was generated. The analysis of the main characteristics of the studies included in this literature review allowed us to outline some factors that could be improved and new approaches not yet exploited that might accelerate the identification of clinically useful proteins for EC diagnosis. These approaches included the need for highly multiplexing verification studies to prioritize the most promising proteins among the initial list of 506 potential biomarkers, and the exploitation of easy-to-access proximal fluids such as uterine aspirates.

In **Chapter 2** we demonstrated that those proteins previously associated with EC, mainly studied by immunohistochemistry at the tissue level, can be measured in uterine aspirate samples by highly multiplexing targeted proteomics. Moreover, we selected the fluid fraction of these uterine aspirate samples, corresponding to the supernatant after centrifugation, as the most appropriate fraction for biomarker identification. The selection was based on the following arguments: a higher number of candidate biomarkers could be detected by LC-MS in this fraction compared to the cellular fraction; the results obtained by targeted LC-SRM approach were significantly more robust and reproducible than those obtained in the cellular fraction; proteins in this fraction are soluble proteins that do not require cell lysis and thus, can be directly measured by technologies widely applied in the clinical laboratories, such as ELISA; and importantly, proteins in the fluid fraction are independent of the limited cellular content in the uterine aspirate samples.

In **Chapter 3** we described a stepwise workflow to reduce the initial list of 506 candidate biomarkers generated in Chapter 1 down to 52 proteins based on their detection by LC-MS in uterine aspirates, the elimination of doubtful candidate biomarkers derived from the variable amounts of blood contamination in uterine aspirates, and their consistency in literature. Those 52 proteins were then analyzed in the fluid fraction of uterine aspirates from 20 EC patients and 18 non-EC controls by LC-PRM. The differential expression of 26 biomarkers was observed (adjusted p-value <0.05 and fold change >3), and among those, ten proteins showed a high sensitivity and specificity (AUC > 0.9). We demonstrated the importance of targeted MS approaches in biomarker verification studies to prioritize the most promising biomarkers to enter a validation study, and highlighted the benefits of LC-PRM acquisition performed on high-resolution accurate mass spectrometers for this purpose, being this one of the first studies applying this technology to clinical samples.

The same 52 proteins were then validated in the fluid fraction of uterine aspirates from 116 patients by LC-PRM, as described in **Chapter 4**. The levels of 28 proteins were significantly higher in EC patients (n=69) compared to controls (n=47), with an adjusted p-value <0.05 and fold change >2. The combination of MMP9 and KPYM exhibited 94% sensitivity and

87% specificity for detecting EC cases. This panel perfectly complemented the standard histopathological diagnosis, achieving 100% of correct diagnosis in this dataset. Nine proteins were significantly increased in endometrioid EC (n=49) compared to less common but more aggressive serous EC (n=20), with an adjusted p-value <0.05. The combination of CTNB1, XPO2 and CAPG achieved 95% sensitivity and 96% specificity for the discrimination of these subtypes.

In order to facilitate the implementation of these biomarker signatures in the clinical practice, the translation of the LC-PRM results to commercially available ELISA kits was evaluated in Chapter 4. Moreover, simplification of the analytical assay and the sample preparation was assessed in **Chapter 5.** We demonstrated the feasibility of measuring the levels of MMP9 with a shortened ELISA assay that can be performed in 1 h instead of the 5 h of the standard ELISA and that can be applied to electrochemical sensors for MMP9 analysis. The results obtained with these biosensors and the LC-PRM results showed a good correlation (R^2=0.96), providing the basis for the development of point-of-care devices for EC diagnosis, and reducing the volume of sample and reagents needed compared to the commercial ELISA kit. Regarding sample preparation, we demonstrated the feasibility of using the raw fluid fraction of uterine aspirates (i.e., avoiding Albumin/IgGs depletion and sonication steps) for the quantification of the described diagnostic panel.

Beyond these important achievements in the field of EC diagnosis, two remarks are worthy to be highlighted from this work:

On one hand, the proteomic workflow described in this thesis, based on targeted MS approaches, and LC-PRM in particular, for the verification and validation of the candidate biomarkers in independent cohorts of patients, followed by immunoassay development for the best-performing panel of proteins, represents a significant improvement over the current state-of-the-art of biomarker identification. Especially, the inclusion of targeted MS approaches in the biomarker pipeline allowed for the screening of a large number of candidate biomarkers and all possible combinations among them, which translates into a more efficient identification of clinically useful biomarker signatures. Importantly, MS-based approaches are independent of the availability of high-quality antibodies.

Many studies have reported serious problems regarding the reliability of commercially available antibodies and immunoassays such as ELISA [202]. Berglund *et al.* reported that fewer than half of around 6,000 commercial antibodies were specific for their intended targets [353], and a poor correlation between antibody lots has also been reported, highlighting the low reproducibility between assays [354]. More than 50-60% of commercially available antibodies are not suitable for biomedical research, resulting in an elevated waste of time and money (estimated annual cost of US$350 million in the United States alone) [355].

Moreover, most of the validated antibodies and ELISA kits have been validated in serum or plasma samples, but not in other biofluids such as uterine aspirates.

Consequently, *de novo* development of highly specific and reproducible immunoassays for the selected proteins and validated in the sample of interest is highly recommended. A recent study has emphasized the importance of defining antibodies by their sequence and produce them as recombinant proteins [355]. Because of the high costs and development time associated with *de novo* assay generation and validation, this is often out of scope of academic laboratories, and usually carried out by specialized companies [303] and, importantly, it must be applied only to the most promising biomarkers. Therefore, the use of MS-based approaches to define the final biomarker signatures before immunoassay development is crucial.

On the other hand, the use of liquid biopsies for the diagnosis of a disease is the preferred method over the collection of more invasive tissue samples and thus, the interest in the search for disease-specific biomarkers in liquid biopsies has significantly increased in the last few years [176,356]. Although currently the vast majority of studies in liquid biopsies employs serum or plasma samples, the use of proximal fluids for biomarker search is a promising alternative. MS instrumentation has undergone rapid advances and it presents a good sensitivity, but limited when applied to complex clinical samples such as blood, with a dynamic range that expands over more than eleven orders of magnitude in protein abundance, and where potential biomarkers are massively diluted to a concentration range of ng/ml and below [197]. In contrast, proximal fluids in contact with the diseased tissue are enriched in proteins directly secreted by the affected cells, favoring the identification and quantification of potential biomarkers. Moreover, blood flows through all organs, and thus, it can be affected by processes occurring in any part of the body, whereas proximal fluids better reflect the specific condition of the diseased tissue.

Proximal fluids as a source of biomarkers are particularly interesting when they can be obtained in a noninvasive or minimally invasive way and therefore could be directly used in the clinical practice for routine screening. In this thesis we demonstrated the usefulness of uterine aspirates as a source of biomarkers to improve early diagnosis of EC. Additionally, we believe that uterine aspirates might also be a promising source of biomarkers for other gynecological diseases such as endometriosis and ovarian cancer, two diseases that lack of an effective diagnostic procedure.

...TO THE CLINICS

Knowledge *per se* has no impact in human health; it needs to be translated into products or services. Translational medical research refers to the conversion of discoveries generated through basic scientific research into improvements in the diagnosis and therapy of human diseases. It has been reported that up to 95% of early-phase studies may not translate into tangible improvements in clinical management [357].

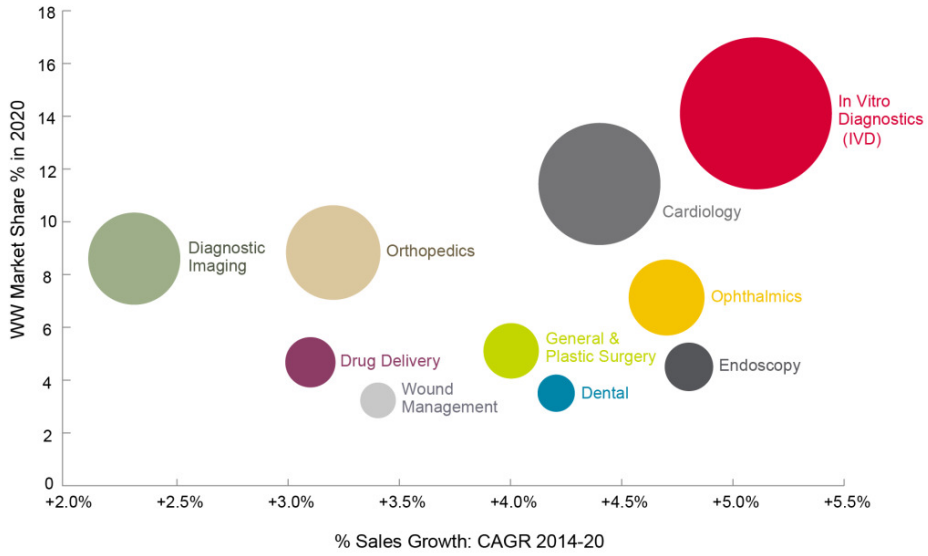


Figure 36. World medical technology market by area and sales growth, 2014-2020. WW, worldwide. *Extracted from Evaluate MedTech October 2015* [358]. Note: Size of the Bubble= Worldwide Sales in 2020**.**

Our next objective is to apply the results described in this thesis to the clinical practice as an *in vitro* diagnostics (IVD) test. According to the European Medical Technology industry, an IVD test is "any medical device which is a reagent, calibrator, control material, kit, instrument, apparatus, equipment or system, whether used alone or in combination, intended to be used *in vitro* (i.e., without direct contact with the patient) for the examination of specimens derived from the human body, including tissue samples, blood and any other biofluid, to provide information concerning a physiological or pathological state, a congenital abnormality, or to determine the safety and compatibility with potential recipients, or to monitor therapeutic measures" [359].

IVD testing has become an indispensable tool in the clinical environment. There are over 40,000 different IVD products currently available that provide information regarding

diagnosis, prognosis, prediction of a treatment response, risk of developing a disease, etc, improving patients health, reducing healthcare costs and increasing the economic growth of the society [360]. IVD is the largest sector of medical technology worldwide followed by cardiology and orthopedics, and it is expected to suffer the greatest sales growth of the medical technology sector during the period 2014-2020 [358] (Figure 36) .

In this context, our final aim would be to license the IVD test for EC diagnosis and prognosis to a specialized company in IVD products and/or oncology IVDs. However, several steps are first required regarding scientific and valorization activities:

A) SCIENTIFIC ACTIVITIES

Step 1. Prototype design

The ease of adoption of a biomarker signature by routine clinical laboratories is a key factor for its clinical application. Therefore, the biggest challenge is now the design of the most simple and cost-effective prototype for the measurement of the described proteomic signatures. Several options are available, each of them offering important advantages and challenges.

❖ **Mass spectrometry**

The most straightforward way to apply the protein signatures described in this thesis would be the use of MS platforms. MS has a promising future in clinical laboratories for numerous reasons including, but not limited to, its high multiplexing capabilities and the high analytical specificity and sensitivity achieved. MS would allow to measure several biomarker panels, such as the diagnostic and predictive signatures described in this thesis, in just one assay, with minimum waste of sample and time. Moreover, it would accelerate the direct application of the high number of biomarkers that are currently being discovered and validated with MS-based approaches, without the need to develop specific immunoassays.

MS-based approaches have already been applied for the quantification of small molecules for several decades and the use of FDA-cleared tests employing mass spectrometers in clinical laboratories is well established for drugs of abuse confirmations, identification of newborn inherited errors of metabolism from dried blood spots, and analysis of steroid hormones such as testosterone [361]. There are also some FDA-approved MS assays and several commercially available MS kits with *in vitro* diagnostic-CE (Conformité Européene) certification for the measurement of therapeutic drugs such as immunosuppressants (e.g., Recipe-ClinMass®, Waters-MassTrack®) [361,362]. In addition, MALDI-TOF for the identification of microbes, and software and databases of microorganisms from different manufacturers have also been FDA-approved [363].

The application of MS in the proteomics field is more recent and thus is still further from its adoption into routine use in clinical laboratories. However, the rapid advances in MS-based proteomics in the last years and those expected to occur in the next five years appear promising to accelerate the clinical use of this technology [364]: i) important improvements in sensitivity, specificity and throughput have been achieved over the past years; ii) automated sample processing has been already addressed with robotics such as the AssayMap Bravo [365] or the Tecan Freedom Evo [366]; iii) FDA-approved instrumentation and software that can be employed by non-experts would be required. Several open source platforms are already available, such as Skyline or Qualis-SIS [367], where the latest has been developed following the FDA requirements; iv) standardized methods that can be executed by non-experts and that satisfy the specific thresholds for precision, accuracy and reproducibility defined by the regulatory agencies are also needed; v) standardized reagents are required, including stable isotope-labeled peptides to improve detection confidence and measurement precision, and standardized quality control kits, several of which are already commercially available [368,369].

Nevertheless, until this technology becomes widely available in clinical laboratories, immunoassays continue to be the gold standard in clinical diagnostics.

❖ **Multiplex ELISA assays**

At the moment, the most feasible approach to apply our findings in the clinics would be the development of a multiplexed ELISA assay. ELISAs present high sensitivity and sample throughput, are fast and straightforward assays once the antibodies are available, and allow for an absolute quantification of the proteins of interest. In addition, the development of a multiplexed immunoassay for the protein panels developed in this thesis would enable the simultaneous measurement of all the proteins in a patient sample with considerable savings in overall assay time, reagent costs and sample volume compared to singleplex ELISAs [370].

Development of multiplexed immunoassays requires extensive validation of assay configuration and performance to ensure assay precision and accuracy. Apart from the important challenges related to the design and validation of high-quality antibodies for each protein (associated to high costs and development time, as previously described), other challenges of these multiplexed assays include the immobilization of the capture ligands, calibration, assessment of the cross-reactivity of reagents, compatibility of assay limits of quantification, and the need for high-level assay automation for clinical application [370,371].

Despite all this, multiplexed immunoassays have tremendous potential for IVD testing [176,372].

Several diagnostic protein arrays have been already approved by the FDA or CE-marked for use in the European Union, most of them focused on the diagnosis of autoimmune and infectious diseases and using the Luminex™ format, such as QuantaPlex (INOVA

Diagnostics), MyAllergyTest (Immune Tech), FIDIS System (Theradiag), Bio-Plex 2200 Autoimmune and Infectious Disease Panels (Bio-Rad Laboratories), etc. [370,373].

Other analytical approaches that can be evaluated are:

❖ **Turbidimetry or nephelometry**

As the ELISA assays, these techniques are also based on antibodies, but the readout of the method is different. In these techniques, antibodies and antigens bind to form a complex that increases the turbidity of the sample. When light is passed through the solution, some light is absorbed by the sample, some is scattered by the sample, and the rest passes through. Turbidimetry assays measure the absorbance of the light by the sample, whereas nephelometry measures the light scattered at various angles by the antigen-antibody complexes [374]. These techniques are widely used in clinical laboratories because they are fast, cost-effective and relatively easily automated. For instance, certain proteins that form the FDA-approved biomarker signature for ovarian cancer diagnosis, known as OVA1, are measured by nephelometry (e.g., Siemens Healthcare Diagnostics' BNII System)[375].

❖ **Electrochemical immunosensors**

Electrochemical biosensors have played a major role in the simplification of IVD testing. One of their major applications is the point of care testing (POCT) [347], including home-use devices, such as the glucose strips that have revolutionized the diabetes monitoring market [376]. Although assays for cancer-related protein biomarkers are more complex, electrochemical immunosensors have high potential for POCT.

In Chapter 5 we demonstrated that the levels of MMP9 in uterine aspirates can be determined in 1 h using electrochemical biosensors. Importantly, electrode arrays with multiple individually addressable electrochemical transducers can be designed to simultaneously measure protein signatures, as described in several studies [377,378].

Step 2. Clinical evaluation

Once the final prototype is optimized for the biomarker signatures, it needs to be analytically and then, clinically validated in a rigorously standardized prospective multicenter clinical trial. Clinical validation will require to comply with the regulatory requirements in order to obtain the CE mark.

B) VALORIZATION ACTIVITIES

Apart from the scientific work, other fields are equally important in order to successfully introduce a product in the market. During the last year of this thesis I have participated in other activities in parallel to the experimental work:

- Regarding the intellectual property strategy, the results generated in this thesis have been protected with two patents (Annex 3, page 211).

- Moreover, a "freedom to operate" analysis of the proteins that form the panels has been carried out to evaluate if it is possible to produce and commercialize an IVD test including those proteins without infringing third party intellectual property rights.

- I have taken part in mentoring and networking activities offered by several grants that the project has received (Annex 3, page 212), including the "CaixaImpulse program" organized by La Caixa Bank, and the "IDEA2 global program" organized by the Massachusetts Institute of Technology (MIT).

- I have also participated in activities related to market research and business plan, such as interviews with stakeholders (patients, gynecologists, hospital managers, phamaceutical companies, etc.), pharmaco-economic studies in the different countries where the product would be introduced, and evaluation of competitors.

- Apart from public funding, private funding will also be required due to the high costs associated with the valorization activities as well as the prototype development and clinical evaluation steps. Therefore, a robust financial plan is needed to present the project to investors and possible licensees.

In summary, the studies included in this thesis constitute an important contribution to the field of proteomic biomarker identification in liquid biopsies in general, and EC in particular. The two uterine aspirate-based proteomic signatures developed in this thesis are expected to significantly improve EC diagnosis. From one side, the diagnostic signature would perfectly complement the current diagnostic procedure based on the histological examination of the cellular fraction of these samples, precluding the use of subsequent invasive sampling methods. On the other hand, the uterine aspirate-based proteomic approach here described paves the way for the identification of proteomic signatures that classify EC tumors into more clinically relevant risk groups to help on surgical treatment prediction. Although further studies are needed in the future, a first step has been achieved with a signature to accurately classify the most common histologies. In the long term, the implementation of these signatures is expected to improve the management of EC patients and substantially reduce disease related healthcare costs.

Conclusions

The main conclusions derived from this thesis are as follows:

1. The high number of molecular studies of EC published between 1978 and 2017 allowed us to generate a list of 506 candidate diagnostic biomarkers, mainly identified in endometrial tissue samples.

2. Although it has been described that uterine aspirates can be used as surrogates for their endometrial tissue counterpart to assess molecular alterations at DNA and RNA levels, this correlation is not maintained at protein level.

3. Despite this low correlation, a high number of proteins that showed significant differences between EC and control patients at tissue level maintained these differential levels in uterine aspirates, enabling the use of this biofluid as a source of protein EC biomarkers.

4. The fluid fraction of uterine aspirates presented important advantages for EC biomarker identification by targeted proteomics and further implementation in the clinical practice compared to the cellular fraction.

5. Verification studies to evaluate the performance of a high number of candidate biomarkers are essential for the prioritization of the most promising biomarkers to enter a subsequent validation phase.

6. PRM acquisition performed on high-resolution accurate mass spectrometers is an ideal platform for biomarker verification and initial validation in complex clinical samples (10s-100s samples). It allowed for the quantification of 100 pairs of peptides per sample in one hour of analysis with excellent precision.

7. LC-PRM acquisition presents important advantages compared to LC-SRM acquisition, particularly for large scale studies: the method design is easier, it has an increased selectivity that reduces the risk of interferences, the list of transitions can be refined iteratively to remove potential interferences without the need for new analyses, and the use of spectral matching as a quality metric significantly facilitates the data processing.

8. A total of 26 proteins showed significantly higher levels in the fluid fraction of uterine aspirates from EC patients compared to non-EC women in two independent cohorts of 38 and 116 patients, the latter covering the broad clinical heterogeneity of women with suspicion of EC.

9. The diagnostic panel composed of MMP9 and KPYM exhibited 94% sensitivity and 87% specificity for detecting EC cases in the fluid fraction of uterine aspirate samples.

10. The combination of this diagnostic panel and the standard histological examination of uterine aspirates achieved 100% of correct diagnosis in our dataset of 116 patients.

11. The predictive panel composed of CTNB1, XPO2 and CAPG achieved 95% sensitivity and 96% specificity for the discrimination of the two main histological subtypes, EEC and more aggressive SEC cancer, in the fluid fraction of uterine aspirate samples.

12. The levels of MMP9 can be measured with a shortened ELISA that can be performed in 1 h and that can be applied to electrochemical sensors. The results obtained with these immunoassays and the LC-PRM results showed a good correlation. This approach would facilitate the use of MMP9 in the clinical practice and must be extended to the other biomarkers.

13. The fluid fraction of uterine aspirates can be used for the direct quantification of the diagnostic panel, without the need for prior treatment (i.e., sonication and albumin/IgGs depletion steps).

14. Finally, this thesis brings forward the proteomic search of biomarkers in uterine aspirates to improve EC diagnosis. In combination with the current histopathological examination, these uterine aspirate-based signatures are expected to preclude the need for more invasive diagnostic samplings and help to predict the optimal surgical treatment for EC patients.

Bibliography

1. Cabero L. *Tratado de Ginecología y Obstetricia. Ginecología y medicina de la reproducción*. (Editorial médica Panamericana, 2013).

2. Baggish MS, Valle RF & Guedj H. *Hysteroscopy: Visual perspectives of uterine anatomy, physiology and pathology*. (Lippincott Williams Wilkins, 2007).

3. Gude, N. M., Roberts, C. T., Kalionis, B. & King, R. G. Growth and function of the normal human placenta. *Thromb. Res.* **114,** 397–407 (2004).

4. Arrowsmith, S., Kendrick, A., Hanley, J.-A., Noble, K. & Wray, S. Myometrial physiology--time to translate? *Exp. Physiol.* **99,** 495–502 (2014).

5. Fox, C., Morin, S., Jeong, J.-W., Scott, R. T. & Lessey, B. A. Local and systemic factors and implantation: what is the evidence? *Fertil. Steril.* **105,** 873–884 (2016).

6. Buckley CH & Fox H. *Biopsy Pathology of the Endometrium*. (Arnold, 2002).

7. Hacker NF, Gambone JC & Hobel CJ. *Hacker & Moore's Essentials of Obstetrics and Gynecology*. (Elsevier, 2016).

8. Sivridis, E. & Giatromanolaki, A. The pathogenesis of endometrial carcinomas at menopause: facts and figures. *J. Clin. Pathol.* **64,** 553–560 (2011).

9. Deligdisch, L. Hormonal Pathology of the Endometrium. *Mod. Pathol.* **13,** 285–294 (2000).

10. Sivridis, E. & Giatromanolaki, A. Proliferative activity in postmenopausal endometrium: the lurking potential for giving rise to an endometrial adenocarcinoma. *J. Clin. Pathol.* **57,** 840–844 (2004).

11. Archer, D. F. *et al.* Endometrial morphology in asymptomatic postmenopausal women. *Am. J. Obstet. Gynecol.* **165,** 317-320; discussion 320-322 (1991).

12. Mutter, G. L., Zaino, R. J., Baak, J. P. A., Bentley, R. C. & Robboy, S. J. Benign endometrial hyperplasia sequence and endometrial intraepithelial neoplasia. *Int. J. Gynecol. Pathol. Off. J. Int. Soc. Gynecol. Pathol.* **26,** 103–114 (2007).

13. Lacey, J. V. & Chia, V. M. Endometrial hyperplasia and the risk of progression to carcinoma. *Maturitas* **63,** 39–44 (2009).

14. Kurman, R. J., Kaminski, P. F. & Norris, H. J. The behavior of endometrial hyperplasia. A long-term study of 'untreated' hyperplasia in 170 patients. *Cancer* **56,** 403–412 (1985).

15. Chen, Y.-L. *et al.* Risk factor analysis of coexisting endometrial carcinoma in patients with endometrial hyperplasia: a retrospective observational study of Taiwanese Gynecologic Oncology Group. *J. Gynecol. Oncol.* **24,** 14–20 (2013).

16. Janicek, M. F. & Rosenshein, N. B. Invasive endometrial cancer in uteri resected for atypical endometrial hyperplasia. *Gynecol. Oncol.* **52,** 373–378 (1994).

17. Rakha, E. *et al.* Clinical outcome of atypical endometrial hyperplasia diagnosed on an endometrial biopsy: institutional experience and review of literature. *Am. J. Surg. Pathol.* **36,** 1683–1690 (2012).

18. Zaino, R. J. *et al.* Reproducibility of the diagnosis of atypical endometrial hyperplasia: a Gynecologic Oncology Group study. *Cancer* **106,** 804–811 (2006).

19. Kendall, B. S. *et al.* Reproducibility of the diagnosis of endometrial hyperplasia, atypical hyperplasia, and well-differentiated carcinoma. *Am. J. Surg. Pathol.* **22,** 1012–1019 (1998).

20. Committee on Gynecologic Practice, Society of Gynecologic Oncology. The American College of Obstetricians and Gynecologists Committee Opinion no. 631. Endometrial intraepithelial neoplasia. *Obstet. Gynecol.* **125,** 1272–1278 (2015).

21. Baak, J. P. *et al.* The molecular genetics and morphometry-based endometrial intraepithelial neoplasia classification system predicts disease progression in endometrial hyperplasia more accurately than the 1994 World Health Organization classification system. *Cancer* **103,** 2304–2312 (2005).

22. Anastasiadis, P. G. *et al.* Endometrial polyps: prevalence, detection, and malignant potential in women with abnormal uterine bleeding. *Eur. J. Gynaecol. Oncol.* **21,** 180–183 (2000).

23. Giordano, G., Gnetti, L., Merisio, C. & Melpignano, M. Postmenopausal status, hypertension and obesity as risk factors for malignant transformation in endometrial polyps. *Maturitas* **56,** 190–197 (2007).

24. Baiocchi, G. *et al.* Malignancy in endometrial polyps: a 12-year experience. *Am. J. Obstet. Gynecol.* **201,** 462.e1-4 (2009).

25. Lee, S. C., Kaunitz, A. M., Sanchez-Ramos, L. & Rhatigan, R. M. The oncogenic potential of endometrial polyps: a systematic review and meta-analysis. *Obstet. Gynecol.* **116,** 1197–1205 (2010).

26. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136,** E359–E386 (2015).

27. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer Statistics, 2017. *CA. Cancer J. Clin.* **67,** 7–30 (2017).

28. Ferlay, J. *et al.* Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur. J. Cancer Oxf. Engl. 1990* **49,** 1374–1403 (2013).

29. Sociedad Española de Oncología Médica (SEOM). Las cifras del cáncer en España. (2016).

30. Lacey, J. V. *et al.* Incidence rates of endometrial hyperplasia, endometrial cancer and hysterectomy from 1980 to 2003 within a large prepaid health plan. *Int. J. Cancer* **131,** 1921–1929 (2012).

31. Sheikh, M. A. *et al.* USA endometrial cancer projections to 2030: should we be concerned? *Future Oncol. Lond. Engl.* **10,** 2561–2568 (2014).

32. Trabert, B. *et al.* Metabolic syndrome and risk of endometrial cancer in the united states: a study in the SEER-medicare linked database. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **24,** 261–267 (2015).

33. Colombo, N. *et al.* ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: diagnosis, treatment and follow-up. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.* **27,** 16–41 (2016).

34. Evans-Metcalf, E. R., Brooks, S. E., Reale, F. R. & Baker, S. P. Profile of women 45 years of age and younger with endometrial cancer. *Obstet. Gynecol.* **91,** 349–354 (1998).

35. Lee, T. S. *et al.* Feasibility of ovarian preservation in patients with early stage endometrial carcinoma. *Gynecol. Oncol.* **104,** 52–57 (2007).

36. Duska, L. R. *et al.* Endometrial cancer in women 40 years old or younger. *Gynecol. Oncol.* **83,** 388–393 (2001).

37. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2012. *CA. Cancer J. Clin.* **62,** 10–29 (2012).

38. Key, T. J. & Pike, M. C. The dose-effect relationship between 'unopposed' oestrogens and endometrial mitotic rate: its central role in explaining and predicting endometrial cancer risk. *Br. J. Cancer* **57,** 205–212 (1988).

39. Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. & Zwahlen, M. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. *Lancet Lond. Engl.* **371,** 569–578 (2008).

40. Bergström, A., Pisani, P., Tenet, V., Wolk, A. & Adami, H. O. Overweight as an avoidable cause of cancer in Europe. *Int. J. Cancer* **91,** 421–430 (2001).

41. Bjørge, T., Engeland, A., Tretli, S. & Weiderpass, E. Body size in relation to cancer of the uterine corpus in 1 million Norwegian women. *Int. J. Cancer* **120,** 378–383 (2007).

42. Wise, M. R. *et al.* Obesity and endometrial hyperplasia and cancer in premenopausal women: A systematic review. *Am. J. Obstet. Gynecol.* **214,** 689.e1-689.e17 (2016).

43. Kaaks, R., Lukanova, A. & Kurzer, M. S. Obesity, endogenous hormones, and endometrial cancer risk: a synthetic review. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **11,** 1531–1543 (2002).

44. Ward, K. K. *et al.* The risk of uterine malignancy is linearly associated with body mass index in a cohort of US women. *Am. J. Obstet. Gynecol.* **209,** 579.e1-5 (2013).

45. Secord, A. A. *et al.* Body mass index and mortality in endometrial cancer: A systematic review and meta-analysis. *Gynecol. Oncol.* **140,** 184–190 (2016).

46. Siiteri, P. K. Steroid hormones and endometrial cancer. *Cancer Res.* **38,** 4360–4366 (1978).

47. Ali, A. T. Reproductive factors and the risk of endometrial cancer. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **24,** 384–393 (2014).

48. Brinton, L. A. *et al.* Reproductive, menstrual, and medical risk factors for endometrial cancer: results from a case-control study. *Am. J. Obstet. Gynecol.* **167,** 1317–1325 (1992).

49. Xu, W.-H. *et al.* Menstrual and reproductive factors and endometrial cancer risk: Results from a population-based case-control study in urban Shanghai. *Int. J. Cancer* **108,** 613–619 (2004).

50. Escobedo, L. G., Lee, N. C., Peterson, H. B. & Wingo, P. A. Infertility-associated endometrial cancer risk may be limited to specific subgroups of infertile women. *Obstet. Gynecol.* **77,** 124–128 (1991).

51. Barry, J. A., Azizia, M. M. & Hardiman, P. J. Risk of endometrial, ovarian and breast cancer in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Hum. Reprod. Update* **20,** 748–758 (2014).

52. Palomba, S., Santagni, S., Falbo, A. & La Sala, G. B. Complications and challenges associated with polycystic ovary syndrome: current perspectives. *Int. J. Womens Health* **7,** 745–763 (2015).

53. Grady, D., Gebretsadik, T., Kerlikowske, K., Ernster, V. & Petitti, D. Hormone replacement therapy and endometrial cancer risk: a meta-analysis. *Obstet. Gynecol.* **85,** 304–313 (1995).

54. Pike, M. C. *et al.* Estrogen-progestin replacement therapy and endometrial cancer. *J. Natl. Cancer Inst.* **89,** 1110–1116 (1997).

55. Schindler, A. E. Progestogen deficiency and endometrial cancer risk. *Maturitas* **62,** 334–337 (2009).

56. Jordan, V. C. The role of tamoxifen in the treatment and prevention of breast cancer. *Curr. Probl. Cancer* **16,** 129–176 (1992).

57. Fornander, T. *et al.* Adjuvant tamoxifen in early breast cancer: occurrence of new primary cancers. *Lancet Lond. Engl.* **1,** 117–120 (1989).

58. Fisher, B. *et al.* Endometrial cancer in tamoxifen-treated breast cancer patients: findings from the National Surgical Adjuvant Breast and Bowel Project (NSABP) B-14. *J. Natl. Cancer Inst.* **86,** 527–537 (1994).

59. Mourits, M. J. *et al.* Tamoxifen treatment and gynecologic side effects: a review. *Obstet. Gynecol.* **97,** 855–866 (2001).

60. van Leeuwen, F. E. *et al.* Risk of endometrial cancer after tamoxifen treatment of breast cancer. *Lancet Lond. Engl.* **343,** 448–452 (1994).

61. Lu, K. H. & Broaddus, R. R. Gynecologic Cancers in Lynch Syndrome/HNPCC. *Fam. Cancer* **4,** 249–254 (2005).

62. Meyer, L. A., Broaddus, R. R. & Lu, K. H. Endometrial cancer and Lynch syndrome: clinical and pathologic considerations. *Cancer Control J. Moffitt Cancer Cent.* **16,** 14–22 (2009).

63. Win, A. K., Reece, J. C. & Ryan, S. Family history and risk of endometrial cancer: a systematic review and meta-analysis. *Obstet. Gynecol.* **125,** 89–98 (2015).

64. Collaborative Group on Epidemiological Studies on Endometrial Cancer. Endometrial cancer and oral contraceptives: an individual participant meta-analysis of 27 276 women with endometrial cancer from 36 epidemiological studies. *Lancet Oncol.* **16,** 1061–1070 (2015).

65. Hinkula, M., Pukkala, E., Kyyrönen, P. & Kauppila, A. Grand multiparity and incidence of endometrial cancer: a population-based study in Finland. *Int. J. Cancer* **98,** 912–915 (2002).

66. Zhou, B. *et al.* Cigarette smoking and the risk of endometrial cancer: a meta-analysis. *Am. J. Med.* **121,** 501–508.e3 (2008).

67. Voskuil, D. W. *et al.* Physical activity and endometrial cancer risk, a systematic review of current evidence. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **16,** 639–648 (2007).

68. Bokhman, J. V. Two pathogenetic types of endometrial carcinoma. *Gynecol. Oncol.* **15,** 10–17 (1983).

69. Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N. & Darai, E. Endometrial cancer. *The Lancet* **387,** 1094–1108 (2016).

70. Colombo, N. *et al.* Endometrial cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO* **22 Suppl 6,** vi35-39 (2011).

71. Catasus, L., Machin, P., Matias-Guiu, X. & Prat, J. Microsatellite instability in endometrial carcinomas: clinicopathologic correlations in a series of 42 cases. *Hum. Pathol.* **29,** 1160–1164 (1998).

72. Mutter, G. L. PTEN, a Protean Tumor Suppressor. *Am. J. Pathol.* **158,** 1895–1898 (2001).

73. Okuda, T. *et al.* Genetics of endometrial cancers. *Obstet. Gynecol. Int.* **2010,** 984013 (2010).

74. Matias-Guiu, X. *et al.* Molecular pathology of endometrial hyperplasia and carcinoma. *Hum. Pathol.* **32,** 569–577 (2001).

75. Byron, S. A. & Pollock, P. M. FGFR2 as a molecular target in endometrial cancer. *Future Oncol. Lond. Engl.* **5,** 27–32 (2009).

76. Amant, F. *et al.* Endometrial cancer. *Lancet* **366,** 491–505 (2005).

77. Yeramian, A. *et al.* Endometrial carcinoma: molecular alterations involved in tumor development and progression. *Oncogene* **32,** 403–413 (2013).

78. Lax, S. F., Kendall, B., Tashiro, H., Slebos, R. J. & Hedrick, L. The frequency of p53, K-ras mutations, and microsatellite instability differs in uterine endometrioid and serous carcinoma: evidence of distinct molecular genetic pathways. *Cancer* **88,** 814–824 (2000).

79. Buza, N., Roque, D. M. & Santin, A. D. HER2/neu in Endometrial Cancer: A Promising Therapeutic Target With Diagnostic Challenges. *Arch. Pathol. Lab. Med.* **138,** 343–350 (2014).

80. Fadare, O. *et al.* The clinicopathologic significance of p53 and BAF-250a (ARID1A) expression in clear cell carcinoma of the endometrium. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **26,** 1101–1110 (2013).

81. Zhang, Z. *et al.* The clinicopathologic significance of the loss of BAF250a (ARID1A) expression in endometrial carcinoma. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **24,** 534–540 (2014).

82. Zannoni, G. F. *et al.* Does high-grade endometrioid carcinoma (grade 3 FIGO) belong to type I or type II endometrial cancer? A clinical-pathological and immunohistochemical study. *Virchows Arch. Int. J. Pathol.* **457,** (2010).

83. Brinton, L. A. *et al.* Etiologic heterogeneity in endometrial cancer: evidence from a Gynecologic Oncology Group trial. *Gynecol. Oncol.* **129,** 277–284 (2013).

84. Murali, R., Soslow, R. A. & Weigelt, B. Classification of endometrial carcinoma: more than two types. *Lancet Oncol.* **15,** e268-278 (2014).

85. Cancer Genome Atlas Research Network *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497,** 67–73 (2013).

86. Hussein, Y. R. *et al.* Clinicopathological analysis of endometrial carcinomas harboring somatic POLE exonuclease domain mutations. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **28,** 505–514 (2015).

87. Modica, I. *et al.* Utility of immunohistochemistry in predicting microsatellite instability in endometrial carcinoma. *Am. J. Surg. Pathol.* **31,** 744–751 (2007).

88. Talhouk, A. *et al.* A clinically applicable molecular-based classification for endometrial cancers. *Br. J. Cancer* **113,** 299–310 (2015).

89. Talhouk, A. *et al.* Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer* **123,** 802–813 (2017).

90. Robertson, G. Screening for endometrial cancer. *Med. J. Aust.* **178,** 657–659 (2003).

91. Lai, C.-R., Hsu, C.-Y., Hang, J.-F. & Li, A. F.-Y. The Diagnostic Value of Routine Papanicolaou Smears for Detecting Endometrial Cancers: An Update. *Acta Cytol.* **59,** 315–318 (2015).

92. Ponce J, Torrejón R & Barahona M. *Oncoguía SEGO. Cáncer de endometrio 2010. Guías de práctica clínica en cáncer ginecológico y mamario*. (Publicaciones SEGO, 2010).

93. Renkonen-Sinisalo, L. *et al.* Surveillance for endometrial cancer in hereditary nonpolyposis colorectal cancer syndrome. *Int. J. Cancer* **120,** 821–824 (2007).

94. Schmeler, K. M. *et al.* Prophylactic surgery to reduce the risk of gynecologic cancers in the Lynch syndrome. *N. Engl. J. Med.* **354,** 261–269 (2006).

95. Gredmark, T., Kvint, S., Havel, G. & Mattsson, L. A. Histopathological findings in women with postmenopausal bleeding. *Br. J. Obstet. Gynaecol.* **102,** 133–136 (1995).

96. van Hanegem, N. *et al.* Diagnostic evaluation of the endometrium in postmenopausal bleeding: an evidence-based approach. *Maturitas* **68,** 155–164 (2011).

97. *Oncoguía SEGO: Cáncer de Endometrio 2016. Guías de práctica clínica en cáncer ginecológico y mamario*. (Publicaciones SEGO, 2016).

98. Denschlag, D., Ulrich, U. & Emons, G. The diagnosis and treatment of endometrial cancer: progress and controversies. *Dtsch. Ärztebl. Int.* **108,** 571–577 (2010).

99. Yasa, C. *et al.* Evaluation of the diagnostic role of transvaginal ultrasound measurements of endometrial thickness to detect endometrial malignancy in asymptomatic postmenopausal women. *Arch. Gynecol. Obstet.* (2016).

100. Smith-Bindman, R. *et al.* Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. *JAMA* **280,** 1510–1517 (1998).

101. Timmermans, A. *et al.* Endometrial thickness measurement for detecting endometrial cancer in women with postmenopausal bleeding: a systematic review and meta-analysis. *Obstet. Gynecol.* **116,** 160–167 (2010).

102. Gupta, J. K., Chien, P. F. W., Voit, D., Clark, T. J. & Khan, K. S. Ultrasonographic endometrial thickness for diagnosing endometrial pathology in women with postmenopausal bleeding: a meta-analysis. *Acta Obstet. Gynecol. Scand.* **81,** 799–816 (2002).

103. Dørum, A., Kristensen, G. B., Langebrekke, A., Sørnes, T. & Skaar, O. Evaluation of endometrial thickness measured by endovaginal ultrasound in women with postmenopausal bleeding. *Acta Obstet. Gynecol. Scand.* **72,** 116–119 (1993).

104. Billingsley, C. C. *et al.* The Use of Transvaginal Ultrasound in Type II Endometrial Cancer. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **25,** 858–862 (2015).

105. Ozdemir, S., Celik, C., Gezginç, K., Kıreşi, D. & Esen, H. Evaluation of endometrial thickness with transvaginal ultrasonography and histopathology in premenopausal women with abnormal vaginal bleeding. *Arch. Gynecol. Obstet.* **282,** 395–399 (2010).

106. Kim, M.-J., Kim, J.-J. & Kim, S. M. Endometrial evaluation with transvaginal ultrasonography for the screening of endometrial hyperplasia or cancer in

premenopausal and perimenopausal women. *Obstet. Gynecol. Sci.* **59,** 192–200 (2016).

107. SGO Clinical Practice Endometrial Cancer Working Group *et al.* Endometrial cancer: a review and current management strategies: part I. *Gynecol. Oncol.* **134,** 385–392 (2014).

108. Dijkhuizen, F. P., Mol, B. W., Brölmann, H. A. & Heintz, A. P. The accuracy of endometrial sampling in the diagnosis of patients with endometrial carcinoma and hyperplasia: a meta-analysis. *Cancer* **89,** 1765–1772 (2000).

109. Clark, T. J. *et al.* Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial cancer: a systematic quantitative review. *BJOG Int. J. Obstet. Gynaecol.* **109,** 313–321 (2002).

110. Elsandabesee, D. & Greenwood, P. The performance of Pipelle endometrial sampling in a dedicated postmenopausal bleeding clinic. *J. Obstet. Gynaecol. J. Inst. Obstet. Gynaecol.* **25,** 32–34 (2005).

111. Batool, T., Reginald, P. W. & Hughes, J. H. Outpatient pipelle endometrial biopsy in the investigation of postmenopausal bleeding. *Br. J. Obstet. Gynaecol.* **101,** 545–546 (1994).

112. Clark, T. J. *et al.* Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia: a systematic quantitative review. *JAMA* **288,** 1610–1621 (2002).

113. Gkrozou, F. *et al.* Hysteroscopy in women with abnormal uterine bleeding: a meta-analysis on four major endometrial pathologies. *Arch. Gynecol. Obstet.* **291,** 1347–1354 (2015).

114. Loiacono, R. M. R. *et al.* Hysteroscopy as a valid tool for endometrial pathology in patients with postmenopausal bleeding or asymptomatic patients with a thickened endometrium: hysteroscopic and histological results. *Gynecol. Obstet. Invest.* **79,** 210–216 (2015).

115. Deckardt, R. *et al.* Comparison of transvaginal ultrasound, hysteroscopy, and dilatation and curettage in the diagnosis of abnormal vaginal bleeding and intrauterine pathology in perimenopausal and postmenopausal women. *J. Am. Assoc. Gynecol. Laparosc.* **9,** 277–282 (2002).

116. van Dongen, H., de Kroon, C. D., Jacobi, C. E., Trimbos, J. B. & Jansen, F. W. Diagnostic hysteroscopy in abnormal uterine bleeding: a systematic review and meta-analysis. *BJOG Int. J. Obstet. Gynaecol.* **114,** 664–675 (2007).

117. Chang, Y.-N., Zhang, Y., Wang, Y.-J., Wang, L.-P. & Duan, H. Effect of hysteroscopy on the peritoneal dissemination of endometrial cancer cells: a meta-analysis. *Fertil. Steril.* **96,** 957–961 (2011).

118. Obermair, A. *et al.* Impact of hysteroscopy on disease-free survival in clinically stage I endometrial cancer patients. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **10,** 275–279 (2000).

119. Pecorelli, S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int. J. Gynaecol. Obstet. Off. Organ Int. Fed. Gynaecol. Obstet.* **105,** 103–104 (2009).

120. Lewin, S. N. *et al.* Comparative performance of the 2009 international Federation of gynecology and obstetrics' staging system for uterine corpus cancer. *Obstet. Gynecol.* **116,** 1141–1149 (2010).

121. Zaino, R. J., Kurman, R. J., Diana, K. L. & Morrow, C. P. The utility of the revised International Federation of Gynecology and Obstetrics histologic grading of endometrial adenocarcinoma using a defined nuclear grading system. A Gynecologic Oncology Group study. *Cancer* **75,** 81–86 (1995).

122. Ayhan, A., Taskiran, C., Yuce, K. & Kucukali, T. The prognostic value of nuclear grading and the revised FIGO grading of endometrial adenocarcinoma. *Int. J. Gynecol. Pathol. Off. J. Int. Soc. Gynecol. Pathol.* **22,** 71–74 (2003).

123. Scholten, A. N., Smit, V. T. H. B. M., Beerman, H., van Putten, W. L. J. & Creutzberg, C. L. Prognostic significance and interobserver variability of histologic grading systems for endometrial carcinoma. *Cancer* **100,** 764–772 (2004).

124. Taylor, R. R., Zeller, J., Lieberman, R. W. & O'Connor, D. M. An analysis of two versus three grades for endometrial carcinoma. *Gynecol. Oncol.* **74,** 3–6 (1999).

125. Lax, S. F., Kurman, R. J., Pizer, E. S., Wu, L. & Ronnett, B. M. A binary architectural grading system for uterine endometrial endometrioid carcinoma has superior reproducibility compared with FIGO grading and identifies subsets of advance-stage tumors with favorable and unfavorable prognosis. *Am. J. Surg. Pathol.* **24,** 1201–1208 (2000).

126. Alkushi, A. *et al.* Description of a novel system for grading of endometrial carcinoma and comparison with existing grading systems. *Am. J. Surg. Pathol.* **29,** 295–304 (2005).

127. Guan, H. *et al.* Prognosis and reproducibility of new and existing binary grading systems for endometrial carcinoma compared to FIGO grading in hysterectomy

specimens. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **21,** 654–660 (2011).

128. Kurman RJ, Carcangiu ML, Herrington CS & Young RH. *WHO Classification of Tumours of Female Reproductive Organs.* (IARC Press, 2014).

129. Clement, P. B. & Young, R. H. Endometrioid carcinoma of the uterine corpus: a review of its pathology with emphasis on recent advances and problematic aspects. *Adv. Anat. Pathol.* **9,** 145–184 (2002).

130. Clement, P. B. & Young, R. H. Non-endometrioid carcinomas of the uterine corpus: a review of their pathology with emphasis on recent advances and problematic aspects. *Adv. Anat. Pathol.* **11,** 117–142 (2004).

131. Abal, M. *et al.* Molecular determinants of invasion in endometrial cancer. *Clin. Transl. Oncol. Off. Publ. Fed. Span. Oncol. Soc. Natl. Cancer Inst. Mex.* **9,** 272–277 (2007).

132. Barrena Medel, N. I. *et al.* Comparison of the prognostic significance of uterine factors and nodal status for endometrial cancer. *Am. J. Obstet. Gynecol.* **204,** 248.e1-7 (2011).

133. Guntupalli, S. R. *et al.* Lymphovascular space invasion is an independent risk factor for nodal disease and poor outcomes in endometrioid endometrial cancer. *Gynecol. Oncol.* **124,** 31–35 (2012).

134. Mannelqvist, M., Stefansson, I., Salvesen, H. B. & Akslen, L. A. Importance of tumour cell invasion in blood and lymphatic vasculature among patients with endometrial carcinoma. *Histopathology* **54,** 174–183 (2009).

135. Abeler, V. M. & Kjørstad, K. E. Endometrial adenocarcinoma in Norway. A study of a total population. *Cancer* **67,** 3093–3103 (1991).

136. Bendifallah, S. *et al.* Just how accurate are the major risk stratification systems for early-stage endometrial cancer? *Br. J. Cancer* **112,** 793–801 (2015).

137. Werner, H. M. J. *et al.* Stathmin protein level, a potential predictive marker for taxane treatment response in endometrial cancer. *PloS One* **9,** (2014).

138. Zeimet, A. G. *et al.* L1CAM in early-stage type I endometrial cancer: results of a large multicenter evaluation. *J. Natl. Cancer Inst.* **105,** 1142–1150 (2013).

139. van der Putten, L. J. *et al.* L1CAM expression in endometrial carcinomas: an ENITEC collaboration study. *Br. J. Cancer* **115,** 716–724 (2016).

140. Helpman, L. *et al.* Assessment of endometrial sampling as a predictor of final surgical pathology in endometrial cancer. *Br. J. Cancer* **110,** 609–615 (2014).

141. Eltabbakh, G. H., Shamonki, J. & Mount, S. L. Surgical stage, final grade, and survival of women with endometrial carcinoma whose preoperative endometrial biopsy shows well-differentiated tumors. *Gynecol. Oncol.* **99,** 309–312 (2005).

142. Francis, J. A., Weir, M. M., Ettler, H. C., Qiu, F. & Kwon, J. S. Should preoperative pathology be used to select patients for surgical staging in endometrial cancer? *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **19,** 380–384 (2009).

143. Petersen, R. W., Quinlivan, J. A., Casper, G. R. & Nicklin, J. L. Endometrial adenocarcinoma--presenting pathology is a poor guide to surgical management. *Aust. N. Z. J. Obstet. Gynaecol.* **40,** 191–194 (2000).

144. Batista, T. P., Cavalcanti, C. L. C., Tejo, A. a. G. & Bezerra, A. L. R. Accuracy of preoperative endometrial sampling diagnosis for predicting the final pathology grading in uterine endometrioid carcinoma. *Eur. J. Surg. Oncol. J. Eur. Soc. Surg. Oncol. Br. Assoc. Surg. Oncol.* (2016). doi:10.1016/j.ejso.2016.03.009

145. Sany, O., Singh, K. & Jha, S. Correlation between preoperative endometrial sampling and final endometrial cancer histology. *Eur. J. Gynaecol. Oncol.* **33,** 142–144 (2012).

146. Nofech-Mozes, S. *et al.* Interobserver Agreement for Endometrial Cancer Characteristics Evaluated on Biopsy Material. *Obstet. Gynecol. Int.* **2012,** (2012).

147. Leitao, M. M. *et al.* Comparison of D&C and office endometrial biopsy accuracy in patients with FIGO grade 1 endometrial adenocarcinoma. *Gynecol. Oncol.* **113,** 105–108 (2009).

148. Bleu, G. *et al.* How to improve the preoperative staging of presumed early-stage endometrioid endometrial cancer? *Eur. J. Gynaecol. Oncol.* **36,** 698–702 (2015).

149. Talhouk, A. *et al.* Molecular classification of endometrial carcinoma on diagnostic specimens is highly concordant with final hysterectomy: Earlier prognostic information to guide treatment. *Gynecol. Oncol.* **143,** 46–53 (2016).

150. Stelloo, E. *et al.* High concordance of molecular tumor alterations between pre-operative curettage and hysterectomy specimens in patients with endometrial carcinoma. *Gynecol. Oncol.* **133,** 197–204 (2014).

151. Kinkel, K. *et al.* Radiologic staging in patients with endometrial cancer: a meta-analysis. *Radiology* **212,** 711–718 (1999).

152. Frei, K. A. & Kinkel, K. Staging endometrial cancer: role of magnetic resonance imaging. *J. Magn. Reson. Imaging JMRI* **13,** 850–855 (2001).

153. Cabrita, S. *et al.* Magnetic resonance imaging in the preoperative staging of endometrial carcinoma. *Eur. J. Gynaecol. Oncol.* **29,** 135–137 (2008).

154. Luomaranta, A., Leminen, A. & Loukovaara, M. Magnetic resonance imaging in the assessment of high-risk features of endometrial carcinoma: a meta-analysis. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **25,** 837–842 (2015).

155. Savelli, L. *et al.* Preoperative local staging of endometrial cancer: transvaginal sonography vs. magnetic resonance imaging. *Ultrasound Obstet. Gynecol. Off. J. Int. Soc. Ultrasound Obstet. Gynecol.* **31,** 560–566 (2008).

156. Horowitz, N. S. *et al.* Prospective evaluation of FDG-PET for detecting pelvic and para-aortic lymph node metastasis in uterine corpus cancer. *Gynecol. Oncol.* **95,** 546–551 (2004).

157. Chang, M.-C. *et al.* 18F-FDG PET or PET/CT for detection of metastatic lymph nodes in patients with endometrial cancer: a systematic review and meta-analysis. *Eur. J. Radiol.* **81,** 3511–3517 (2012).

158. Rodolakis, A. *et al.* European Society of Gynecological Oncology Task Force for Fertility Preservation: Clinical Recommendations for Fertility-Sparing Management in Young Endometrial Cancer Patients. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **25,** 1258–1265 (2015).

159. Janda, M. *et al.* Quality of life after total laparoscopic hysterectomy versus total abdominal hysterectomy for stage I endometrial cancer (LACE): a randomised trial. *Lancet Oncol.* **11,** 772–780 (2010).

160. Ju, W. *et al.* Comparison of laparoscopy and laparotomy for management of endometrial carcinoma: a meta-analysis. *Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc.* **19,** 400–406 (2009).

161. Zanagnolo, V., Garbi, A., Achilarre, M. T. & Minig, L. Robot-assisted Surgery in Gynecologic Cancers. *J. Minim. Invasive Gynecol.* **24,** 379–396 (2017).

162. Koskas, M., Rouzier, R. & Amant, F. Staging for endometrial cancer: The controversy around lymphadenectomy - Can this be resolved? *Best Pract. Res. Clin. Obstet. Gynaecol.* **29,** 845–857 (2015).

163. ASTEC study group *et al.* Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. *Lancet Lond. Engl.* **373,** 125–136 (2009).

164. Benedetti Panici, P. *et al.* Systematic pelvic lymphadenectomy vs. no lymphadenectomy in early-stage endometrial carcinoma: randomized clinical trial. *J. Natl. Cancer Inst.* **100,** 1707–1716 (2008).

165. Abu-Rustum, N. R. *et al.* The incidence of symptomatic lower-extremity lymphedema following treatment of uterine corpus malignancies: a 12-year experience at Memorial Sloan-Kettering Cancer Center. *Gynecol. Oncol.* **103,** 714–718 (2006).

166. Ballester, M. *et al.* Detection rate and diagnostic accuracy of sentinel-node biopsy in early stage endometrial cancer: a prospective multicentre study (SENTI-ENDO). *Lancet Oncol.* **12,** 469–476 (2011).

167. Amant, F. & Trum, H. Sentinel-lymph-node mapping in endometrial cancer: routine practice? *Lancet Oncol.* **18,** 281–282 (2017).

168. SGO Clinical Practice Endometrial Cancer Working Group *et al.* Endometrial cancer: a review and current management strategies: part II. *Gynecol. Oncol.* **134,** 393–402 (2014).

169. Creutzberg, C. L. *et al.* Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. PORTEC Study Group. Post Operative Radiation Therapy in Endometrial Carcinoma. *Lancet Lond. Engl.* **355,** 1404–1411 (2000).

170. Keys, H. M. *et al.* A phase III trial of surgery with or without adjunctive external pelvic radiation therapy in intermediate risk endometrial adenocarcinoma: a Gynecologic Oncology Group study. *Gynecol. Oncol.* **92,** 744–751 (2004).

171. Hogberg, T. *et al.* Sequential adjuvant chemotherapy and radiotherapy in endometrial cancer--results from two randomised studies. *Eur. J. Cancer Oxf. Engl. 1990* **46,** 2422–2431 (2010).

172. Randall, M. E. *et al.* Randomized phase III trial of whole-abdominal irradiation versus doxorubicin and cisplatin chemotherapy in advanced endometrial carcinoma: a Gynecologic Oncology Group Study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **24,** 36–44 (2006).

173. Thigpen, J. T. *et al.* Oral medroxyprogesterone acetate in the treatment of advanced or recurrent endometrial carcinoma: a dose-response study by the Gynecologic Oncology Group. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **17,** 1736–1744 (1999).

174. Reifenstein, E. C. The treatment of advanced endometrial cancer with hydroxyprogesterone caproate. *Gynecol. Oncol.* **2,** 377–414 (1974).

175. Whitney, C. W. *et al.* Phase II study of medroxyprogesterone acetate plus tamoxifen in advanced endometrial carcinoma: a Gynecologic Oncology Group study. *Gynecol. Oncol.* **92,** 4–9 (2004).

176. Borrebaeck, C. A. K. Precision diagnostics: moving towards protein biomarker signatures of clinical utility in cancer. *Nat. Rev. Cancer* **17,** 199–204 (2017).

177. Trusheim, M. R., Berndt, E. R. & Douglas, F. L. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nat. Rev. Drug Discov.* **6,** 287–293 (2007).

178. Li, D. & Chan, D. W. Proteomic cancer biomarkers from discovery to approval: it's worth the effort. *Expert Rev. Proteomics* **11,** 135–136 (2014).

179. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69,** 89–95 (2001).

180. Kimhofer, T., Fye, H., Taylor-Robinson, S., Thursz, M. & Holmes, E. Proteomic and metabonomic biomarkers for hepatocellular carcinoma: a comprehensive review. *Br. J. Cancer* **112,** 1141–1156 (2015).

181. Pavlou, M. P., Diamandis, E. P. & Blasutig, I. M. The long journey of cancer biomarkers from the bench to the clinic. *Clin. Chem.* **59,** 147–157 (2013).

182. Catalona, W. J. *et al.* Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. *N. Engl. J. Med.* **324,** 1156–1161 (1991).

183. Schröder, F. H. *et al.* Screening and prostate-cancer mortality in a randomized European study. *N. Engl. J. Med.* **360,** 1320–1328 (2009).

184. Nalejska, E., Mączyńska, E. & Lewandowska, M. A. Prognostic and predictive biomarkers: tools in personalized oncology. *Mol. Diagn. Ther.* **18,** 273–284 (2014).

185. Duffy, M. J. Biochemical markers as prognostic indices in breast cancer. *Clin. Chem.* **36,** 188–191 (1990).

186. Isaacs, C., Stearns, V. & Hayes, D. F. New prognostic factors for breast cancer recurrence. *Semin. Oncol.* **28,** 53–67 (2001).

187. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **14,** 2843–2877 (1996).

188. Ebhardt, H. A., Root, A., Sander, C. & Aebersold, R. Applications of targeted proteomics in systems biology and translational medicine. *Proteomics* **15,** 3193–3208 (2015).

189. Prabakaran, S., Lippens, G., Steen, H. & Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **4,** 565–583 (2012).

190. Legrain, P. *et al.* The human proteome project: current state and future direction. *Mol. Cell. Proteomics MCP* **10,** M111.009993 (2011).

191. Wishart, D. S. *et al.* HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* **41,** D801–D807 (2013).

192. Jensen, O. N. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8,** 33–41 (2004).

193. Aebersold, R. *et al.* Perspective: a program to improve protein biomarker discovery for cancer. *J. Proteome Res.* **4,** 1104–1109 (2005).

194. Tainsky, M. A. Genomic and Proteomic Biomarkers for Cancer: A Multitude of Opportunities. *Biochim. Biophys. Acta* **1796,** 176–193 (2009).

195. Ebhardt, H. A., Sabidó, E., Hüttenhain, R., Collins, B. & Aebersold, R. Range of protein detection by selected/multiple reaction monitoring mass spectrometry in an unfractionated human cell culture lysate. *Proteomics* **12,** 1185–1193 (2012).

196. Omenn, G. S. Plasma proteomics, the Human Proteome Project, and cancer-associated alternative splice variant proteins. *Biochim. Biophys. Acta* **1844,** 866–873 (2014).

197. Anderson, N. L. & Anderson, N. G. The Human Plasma Proteome History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **1,** 845–867 (2002).

198. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130,** 395–398 (2007).

199. Rifai, N., Gillette, M. A. & Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **24,** 971–983 (2006).

200. Surinova, S. *et al.* On the development of plasma protein biomarkers. *J. Proteome Res.* **10,** 5–16 (2011).

201. Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N. & Wang, P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. *Proteomics Clin. Appl.* **2,** 1386–1402 (2008).

202. Whiteaker, J. R. *et al.* A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **29,** 625–634 (2011).

203. Makawita, S. & Diamandis, E. P. The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification. *Clin. Chem.* **56,** 212–222 (2010).

204. Parker, C. E. & Borchers, C. H. Mass spectrometry based biomarker discovery, verification, and validation--quality assurance and control of protein biomarker assays. *Mol. Oncol.* **8,** 840–858 (2014).

205. Ye, X., Blonder, J. & Veenstra, T. D. Targeted proteomics for validation of biomarkers in clinical samples. *Brief. Funct. Genomic. Proteomic.* **8,** 126–135 (2009).

206. 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* **10,** 13 (2013).

207. Diamandis, E. P. Cancer Biomarkers: Can We Turn Recent Failures into Success? *JNCI J. Natl. Cancer Inst.* **102,** 1462–1467 (2010).

208. Ioannidis, J. P. A. Biomarker Failures. *Clin. Chem.* **59,** 202–204 (2013).

209. DeSouza, L. V. *et al.* Endometrial carcinoma biomarker discovery and verification using differentially tagged clinical samples with multidimensional liquid chromatography and tandem mass spectrometry. *Mol. Cell. Proteomics MCP* **6,** 1170–1182 (2007).

210. Wiśniewski, J. R. *et al.* Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **8,** 611 (2012).

211. Casadonte, R. & Caprioli, R. M. Proteomic analysis of formalin-fixed paraffin-embedded tissue by MALDI imaging mass spectrometry. *Nat. Protoc.* **6,** 1695–1709 (2011).

212. Gustafsson, O. J. R., Arentz, G. & Hoffmann, P. Proteomic developments in the analysis of formalin-fixed tissue. *Biochim. Biophys. Acta* **1854,** 559–580 (2015).

213. Drabovich, A. P., Martínez-Morillo, E. & Diamandis, E. P. Toward an integrated pipeline for protein biomarker development. *Biochim. Biophys. Acta* **1854,** 677–686 (2015).

214. Teng, P., Bateman, N. W., Hood, B. L. & Conrads, T. P. Advances in proximal fluid proteomics for disease biomarker discovery. *J. Proteome Res.* **9,** 6091–6100 (2010).

215. Schwartz, E. B. & Granger, D. A. Transferrin enzyme immunoassay for quantitative monitoring of blood contamination in saliva. *Clin. Chem.* **50,** 654–656 (2004).

216. Aasebø, E. *et al.* Effects of blood contamination and the rostro-caudal gradient on the human cerebrospinal fluid proteome. *PloS One* **9,** e90429 (2014).

217. Chou, R. *et al.* Urinary Biomarkers for Diagnosis of Bladder Cancer: A Systematic Review and Meta-analysis. *Ann. Intern. Med.* **163,** 922–931 (2015).

218. Truong, M., Yang, B. & Jarrard, D. F. Toward the detection of prostate cancer in urine: a critical analysis. *J. Urol.* **189,** 422–429 (2013).

219. Øverbye, A. *et al.* Identification of prostate cancer biomarkers in urinary exosomes. *Oncotarget* **6,** 30357–30376 (2015).

220. Perez-Sanchez, C. *et al.* Molecular diagnosis of endometrial cancer from uterine aspirates. *Int. J. Cancer* **133,** 2383–2391 (2013).

221. Colas, E. *et al.* Molecular markers of endometrial carcinoma detected in uterine aspirates. *Int. J. Cancer J. Int. Cancer* **129,** 2435–2444 (2011).

222. Casado-Vela, J. *et al.* Comprehensive proteomic analysis of human endometrial fluid aspirate. *J. Proteome Res.* **8,** 4622–4632 (2009).

223. Ametzazurra, A. *et al.* Endometrial fluid is a specific and non-invasive biological sample for protein biomarker identification in endometriosis. *Hum. Reprod. Oxf. Engl.* **24,** 954–965 (2009).

224. Salamonsen, L. A. *et al.* Proteomics of the human endometrium and uterine fluid: a pathway to biomarker discovery. *Fertil. Steril.* **99,** 1086–1092 (2013).

225. Boja, E. *et al.* Evolution of clinical proteomics and its role in medicine. *J. Proteome Res.* **10,** 66–84 (2011).

226. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480,** 254–258 (2011).

227. Feist, P. & Hummon, A. B. Proteomic Challenges: Sample Preparation Techniques for Microgram-Quantity Protein Analysis from Biological Samples. *Int. J. Mol. Sci.* **16,** 3537–3563 (2015).

228. Gundry, R. L. *et al.* Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Curr. Protoc. Mol. Biol. Ed. Frederick M Ausubel Al* **CHAPTER,** Unit10.25 (2009).

229. Trevisiol, S. *et al.* The use of proteases complementary to trypsin to probe isoforms and modifications. *Proteomics* **16,** 715–728 (2016).

230. Mesmin, C., van Oostrum, J. & Domon, B. Complexity reduction of clinical samples for routine mass spectrometric analysis. *Proteomics Clin. Appl.* **10,** 315–322 (2016).

231. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422,** 198–207 (2003).

232. Trufelli, H., Palma, P., Famiglini, G. & Cappiello, A. An overview of matrix effects in liquid chromatography-mass spectrometry. *Mass Spectrom. Rev.* **30,** 491–509 (2011).

233. Hörth, P., Miller, C. A., Preckel, T. & Wenz, C. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Mol. Cell. Proteomics MCP* **5,** 1968–1974 (2006).

234. Magdeldin, S. *et al.* Complementary Protein and Peptide OFFGEL Fractionation for High-Throughput Proteomic Analysis. *Anal. Chem.* **87,** 8481–8488 (2015).

235. de Hoffmann E & Stroobant V. *Mass Spectrometry. Principles and Applications*. (John Wiley & Sons, 2007).

236. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312,** 212–217 (2006).

237. Sleno, L. & Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom. JMS* **39,** 1091–1112 (2004).

238. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 6233–6237 (1986).

239. Ong, S.-E. & Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1,** 252–262 (2005).

240. van der Wal L & Demmer JAA. Quantitative Mass Spectrometry-based proteomics. in *Recent Advances in Proteomics Research* (InTechOpen, 2015).

241. Gallien, S. & Domon, B. Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. *Methods San Diego Calif* **81,** 15–23 (2015).

242. Heudi, O. *et al.* Towards absolute quantification of therapeutic monoclonal antibody in serum by LC-MS/MS using isotope-labeled antibody standard and protein cleavage isotope dilution mass spectrometry. *Anal. Chem.* **80,** 4200–4207 (2008).

243. Huillet, C. *et al.* Accurate quantification of cardiovascular biomarkers in serum using Protein Standard Absolute Quantification (PSAQ™) and selected reaction monitoring. *Mol. Cell. Proteomics MCP* **11,** M111.008235 (2012).

244. Ong, S.-E. *et al.* Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics MCP* **1,** 376–386 (2002).

245. Krüger, M. *et al.* SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134,** 353–364 (2008).

246. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **7,** 383–385 (2010).

247. Geiger, T. *et al.* Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* **6,** 147–157 (2011).

248. Kirkpatrick, D. S., Gerber, S. A. & Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods San Diego Calif* **35,** 265–273 (2005).

249. Kettenbach, A. N., Rush, J. & Gerber, S. A. Absolute quantification of protein and post-translational modification abundance with stable isotope–labeled synthetic peptides. *Nat. Protoc.* **6,** 175–186 (2011).

250. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci.* **100,** 6940–6945 (2003).

251. Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics MCP* **3,** 1154–1169 (2004).

252. Evans, C. *et al.* An insight into iTRAQ: where do we stand now? *Anal. Bioanal. Chem.* **404,** 1011–1027 (2012).

253. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17,** 994–999 (1999).

254. Thompson, A. *et al.* Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **75,** 1895–1904 (2003).

255. Neilson, K. A. *et al.* Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* **11,** 535–553 (2011).

256. Wang, W. *et al.* Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.* **75,** 4818–4826 (2003).

257. Rauniyar, N. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int. J. Mol. Sci.* **16,** 28566–28581 (2015).

258. Mermelekas, G., Vlahou, A. & Zoidakis, J. SRM/MRM targeted proteomics as a tool for biomarker validation and absolute quantification in human urine. *Expert Rev. Mol. Diagn.* **15,** 1441–1454 (2015).

259. Lesur, A. & Domon, B. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. *Proteomics* **15,** 880–890 (2015).

260. Domon, B. & Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **28,** 710–721 (2010).

261. Wilkins, M. R. *et al.* From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnol. Nat. Publ. Co.* **14,** 61–65 (1996).

262. Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up the mountains. *Proteomics* **2,** 3–10 (2002).

263. Henzel, W. J. *et al.* Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 5011–5015 (1993).

264. Pappin, D. J., Hojrup, P. & Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol. CB* **3,** 327–332 (1993).

265. Shevchenko, A. *et al.* Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 14440–14445 (1996).

266. Izquierdo, I. *et al.* Proteomic identification of putative biomarkers for early detection of sudden cardiac death in a family with a LMNA gene mutation causing dilated cardiomyopathy. *J. Proteomics* **148,** 75–84 (2016).

267. Viswanathan, S., Unlü, M. & Minden, J. S. Two-dimensional difference gel electrophoresis. *Nat. Protoc.* **1,** 1351–1358 (2006).

268. Shao, S., Guo, T. & Aebersold, R. Mass spectrometry-based proteomic quest for diabetes biomarkers. *Biochim. Biophys. Acta* **1854,** 519–527 (2015).

269. Ducret, A., Van Oostveen, I., Eng, J. K., Yates, J. R. & Aebersold, R. High throughput protein characterization by automated reverse-phase chromatography/electrospray tandem mass spectrometry. *Protein Sci. Publ. Protein Soc.* **7,** 706–719 (1998).

270. Hebert, A. S. *et al.* The one hour yeast proteome. *Mol. Cell. Proteomics MCP* **13,** 339–347 (2014).

271. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989 (1994).

272. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567 (1999).

273. Geer, L. Y. *et al.* Open mass spectrometry search algorithm. *J. Proteome Res.* **3,** 958–964 (2004).

274. Mann, M. & Kelleher, N. L. Precision proteomics: The case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci.* **105,** 18132–18138 (2008).

275. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M. & Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* **83,** 5442–5446 (2011).

276. Hu, Q. *et al.* The Orbitrap: a new mass spectrometer. *J. Mass Spectrom. JMS* **40,** 430–443 (2005).

277. Michalski, A. *et al.* Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics MCP* **10,** M111.011015 (2011).

278. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76,** 4193–4201 (2004).

279. Tabb, D. L. *et al.* Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **9,** 761–776 (2010).

280. Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **1,** 39–45 (2004).

281. Collins, B. C. *et al.* Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* **10,** 1246–1253 (2013).

282. Toprak, U. H. *et al.* Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol. Cell. Proteomics MCP* **13,** 2056–2071 (2014).

283. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics MCP* **11,** O111.016717 (2012).

284. Bern, M. *et al.* Deconvolution of mixture spectra from ion-trap data-independent-acquisition tandem mass spectrometry. *Anal. Chem.* **82,** 833–841 (2010).

285. Hou, G. *et al.* Biomarker Discovery and Verification of Esophageal Squamous Cell Carcinoma Using Integration of SWATH/MRM. *J. Proteome Res.* **14,** 3793–3803 (2015).

286. Ortea, I., Rodríguez-Ariza, A., Chicano-Gálvez, E., Arenas Vacas, M. S. & Jurado Gámez, B. Discovery of potential protein biomarkers of lung adenocarcinoma in bronchoalveolar lavage fluid by SWATH MS data-independent acquisition and targeted data extraction. *J. Proteomics* **138,** 106–114 (2016).

287. Luo, Y. *et al.* SWATH-based proteomics identified carbonic anhydrase 2 as a potential diagnosis biomarker for nasopharyngeal carcinoma. *Sci. Rep.* **7,** 41191 (2017).

288. Stahl-Zeng, J. *et al.* High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics MCP* **6,** 1809–1817 (2007).

289. Method of the Year 2012. *Nat. Methods* **10,** 1–1 (2013).

290. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9,** 555–566 (2012).

291. Hüttenhain, R., Malmström, J., Picotti, P. & Aebersold, R. Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol.* **13,** 518–525 (2009).

292. Gallien, S., Duriez, E. & Domon, B. Selected reaction monitoring applied to proteomics. *J. Mass Spectrom. JMS* **46,** 298–312 (2011).

293. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4,** 222 (2008).

294. Kuzyk, M. A. *et al.* Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. *Mol. Cell. Proteomics MCP* **8,** 1860–1877 (2009).

295. 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 MCP* **11,** 1475–1488 (2012).

296. Domon, B. & Gallien, S. Recent advances in targeted proteomics for clinical applications. *Proteomics Clin. Appl.* **9,** 423–431 (2015).

297. Bourmaud, A., Gallien, S. & Domon, B. Parallel reaction monitoring using quadrupole-orbitrap mass spectrometer: Principle and applications. *Proteomics* (2016). doi:10.1002/pmic.201500543

298. Gallien, S., Kim, S. Y. & Domon, B. Large-Scale Targeted Proteomics Using Internal Standard Triggered-Parallel Reaction Monitoring (IS-PRM). *Mol. Cell. Proteomics MCP* **14,** 1630–1644 (2015).

299. Kim, Y. J. *et al.* Quantification of SAA1 and SAA2 in lung cancer plasma using the isotype-specific PRM assays. *Proteomics* **15,** 3116–3125 (2015).

300. Jaffe, J. D. *et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat. Genet.* **45,** 1386–1391 (2013).

301. Kim, H.-J., Lin, D., Lee, H.-J., Li, M. & Liebler, D. C. Quantitative Profiling of Protein Tyrosine Kinases in Human Cancer Cell Lines by Multiplexed Parallel Reaction Monitoring Assays. *Mol. Cell. Proteomics MCP* **15,** 682–691 (2016).

302. Dun, M. D. *et al.* Proteotranscriptomic Profiling of 231-BR Breast Cancer Cells: Identification of Potential Biomarkers and Therapeutic Targets for Brain Metastasis. *Mol. Cell. Proteomics MCP* **14,** 2316–2330 (2015).

303. Wang, P., Whiteaker, J. R. & Paulovich, A. G. The evolving role of mass spectrometry in cancer biomarker discovery. *Cancer Biol. Ther.* **8,** 1083–1094 (2009).

304. Cox, K. L. *et al.* Immunoassay Methods. in *Assay Guidance Manual* (eds. Sittampalam, G. S. et al.) (Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2004).

305. Solier, C. & Langen, H. Antibody-based proteomics and biomarker research - current status and limitations. *Proteomics* **14,** 774–783 (2014).

306. Telner, D. E. & Jakubovicz, D. Approach to diagnosis and management of abnormal uterine bleeding. *Can. Fam. Physician Med. Fam. Can.* **53,** 58–64 (2007).

307. Astrup, K. & Olivarius, N. de F. Frequency of spontaneously occurring postmenopausal bleeding in the general population. *Acta Obstet. Gynecol. Scand.* **83,** 203–207 (2004).

308. Abdelazim, I. A., Aboelezz, A. & AbdulKareem, A. F. Pipelle endometrial sampling versus conventional dilatation & curettage in patients with abnormal uterine bleeding. *J. Turk. Ger. Gynecol. Assoc.* **14,** 1–5 (2013).

309. Bradley, L. D. Complications in hysteroscopy: prevention, treatment and legal risk. *Curr. Opin. Obstet. Gynecol.* **14,** 409–415 (2002).

310. Esteller, M., García, A., Martínez-Palones, J. M., Xercavins, J. & Reventós, J. Detection of clonality and genetic alterations in endometrial pipelle biopsy and its surgical specimen counterpart. *Lab. Investig. J. Tech. Methods Pathol.* **76,** 109–116 (1997).

311. Shenoy, A. & Geiger, T. Super-SILAC: current trends and future perspectives. *Expert Rev. Proteomics* **12,** 13–19 (2015).

312. Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinforma. Oxf. Engl.* **30,** 2524–2526 (2014).

313. Wan, K. X., Vidavsky, I. & Gross, M. L. Comparing similar spectra: from similarity index to spectral contrast angle. *J. Am. Soc. Mass Spectrom.* **13,** 85–88 (2002).

314. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57,** 289–300 (1995).

315. Tanabe, K. *et al.* Midkine and its clinical significance in endometrial carcinoma. *Cancer Sci.* **99,** 1125–1130 (2008).

316. Tian, W. *et al.* Visfatin, a potential biomarker and prognostic factor for endometrial cancer. *Gynecol. Oncol.* **129,** 505–512 (2013).

317. Yurkovetsky, Z. *et al.* Development of multimarker panel for early detection of endometrial cancer. High diagnostic power of prolactin. *Gynecol. Oncol.* **107,** 58–65 (2007).

318. DeSouza, L. V., Romaschin, A. D., Colgan, T. J. & Siu, K. W. M. Absolute quantification of potential cancer markers in clinical tissue homogenates using multiple reaction monitoring on a hybrid triple quadrupole/linear ion trap tandem mass spectrometer. *Anal. Chem.* **81,** 3462–3470 (2009).

319. DeSouza, L. V. *et al.* Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. *J. Proteome Res.* **7,** 3525–3534 (2008).

320. Baker, S. G., Kramer, B. S. & Srivastava, S. Markers for early detection of cancer: statistical guidelines for nested case-control studies. *BMC Med. Res. Methodol.* **2,** 4 (2002).

321. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44,** 837–845 (1988).

322. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recogn Lett* **27,** 861–874 (2006).

323. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12,** 77 (2011).

324. Vandooren, J., Van den Steen, P. E. & Opdenakker, G. Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase-9 (MMP-9): the next decade. *Crit. Rev. Biochem. Mol. Biol.* **48,** 222–272 (2013).

325. Deryugina, E. I. & Quigley, J. P. Matrix metalloproteinases and tumor metastasis. *Cancer Metastasis Rev.* **25,** 9–34 (2006).

326. Amalinei, C. *et al.* Immunohistochemical analysis of steroid receptors, proliferation markers, apoptosis related molecules, and gelatinases in non-neoplastic and neoplastic endometrium. *Ann. Anat. Anat. Anz. Off. Organ Anat. Ges.* **193,** (2011).

327. Iqbal, M. A., Gupta, V., Gopinath, P., Mazurek, S. & Bamezai, R. N. K. Pyruvate kinase M2 and cancer: an updated assessment. *FEBS Lett.* **588,** 2685–2692 (2014).

328. Hsu, P. P. & Sabatini, D. M. Cancer cell metabolism: Warburg and beyond. *Cell* **134,** 703–707 (2008).

329. Dube, V. *et al.* Verification of endometrial tissue biomarkers previously discovered using mass spectrometry-based proteomics by means of immunohistochemistry in a tissue microarray format. *J. Proteome Res.* **6,** 2648–2655 (2007).

330. Rai, P., Kota, V., Sundaram, C. S., Deendayal, M. & Shivaji, S. Proteome of human endometrium: Identification of differentially expressed proteins in proliferative and secretory phase endometrium. *Proteomics Clin. Appl.* **4,** 48–59 (2010).

331. Brembeck, F. H., Rosário, M. & Birchmeier, W. Balancing cell adhesion and Wnt signaling, the key role of beta-catenin. *Curr. Opin. Genet. Dev.* **16,** 51–59 (2006).

332. Schlosshauer, P. W., Ellenson, L. H. & Soslow, R. A. Beta-catenin and E-cadherin expression patterns in high-grade endometrial carcinoma are associated with

histological subtype. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* **15,** 1032–1037 (2002).

333. Scholten, A. N., Creutzberg, C. L., van den Broek, L. J. C. M., Noordijk, E. M. & Smit, V. T. H. B. M. Nuclear beta-catenin is a molecular feature of type I endometrial carcinoma. *J. Pathol.* **201,** 460–465 (2003).

334. Glaser, J. *et al.* Macrophage capping protein CapG is a putative oncogene involved in migration and invasiveness in ovarian carcinoma. *BioMed Res. Int.* **2014,** 379847 (2014).

335. Behrens, P., Brinkmann, U. & Wellmann, A. CSE1L/CAS: its role in proliferation and apoptosis. *Apoptosis Int. J. Program. Cell Death* **8,** 39–44 (2003).

336. Tai, C.-J., Hsu, C.-H., Shen, S.-C., Lee, W.-R. & Jiang, M.-C. Cellular apoptosis susceptibility (CSE1L/CAS) protein in cancer metastasis and chemotherapeutic drug-induced apoptosis. *J. Exp. Clin. Cancer Res. CR* **29,** 110 (2010).

337. Peiro, G., Diebold, J., Baretton, G. B., Kimmig, R. & Lohrs, U. Cellular apoptosis susceptibility gene expression in endometrial carcinoma: correlation with Bcl-2, Bax, and caspase-3 expression and outcome. *Int. J. Gynecol. Pathol. Off. J. Int. Soc. Gynecol. Pathol.* **20,** 359–367 (2001).

338. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci. Signal.* **6,** pl1 (2013).

339. Gan, S. D. & Patel, K. R. Enzyme immunoassay and enzyme-linked immunosorbent assay. *J. Invest. Dermatol.* **133,** e12 (2013).

340. Avrameas, S., Ternynck, T. & Guesdon, J.-L. Coupling of Enzymes to Antibodies and Antigens. *Scand. J. Immunol.* **8,** 7–23 (1978).

341. Savage, M. D. *Avidin-biotin chemistry: a handbook*. (Pierce Chemical Company, 1992).

342. Vasilov, R. G. & Tsitsikov, E. N. An ultrasensitive immunoassay for human IgE measurement in cell-culture supernatant. *Immunol. Lett.* **26,** 283–284 (1990).

343. Mayer, G. & Bendayan, M. Amplification methods for the immunolocalization of rare molecules in cells and tissues. *Prog. Histochem. Cytochem.* **36,** 3–85 (2001).

344. Marquette, C. A., Hezard, P., Degiuli, A. & Blum, L. J. Macro-molecular chemiluminescent complex for enhanced immuno-detection onto microtiter plate and protein biochip. *Sens. Actuators B Chem.* **113,** 664–670 (2006).

345. Charbgoo, F., Mirshahi, M., Sarikhani, S. & Saifi Abolhassan, M. Synthesis of a unique high-performance poly-horseradish peroxidase complex to enhance sensitivity of immunodetection systems. *Biotechnol. Appl. Biochem.* **59,** 45–49 (2012).

346. Li, D., Ying, Y., Wu, J., Niessner, R. & Knopp, D. Comparison of monomeric and polymeric horseradish peroxidase as labels in competitive ELISA for small molecule detection. *Microchim. Acta* **180,** 711–717 (2013).

347. Wang, J. Electrochemical biosensors: towards point-of-care cancer diagnostics. *Biosens. Bioelectron.* **21,** 1887–1892 (2006).

348. IARC Publications Website - World Cancer Report 2014. Available at: http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014. (Accessed: 17th April 2017)

349. *World Health Organization. Cancer control: Early detection WHO guide for effective programmes*. (WHO, 2007).

350. Etzioni, R. *et al.* The case for early detection. *Nat. Rev. Cancer* **3,** 243–252 (2003).

351. Khalilpour, A., Kilic, T., Khalilpour, S., Álvarez, M. M. & Yazdi, I. K. Proteomic-based biomarker discovery for development of next generation diagnostics. *Appl. Microbiol. Biotechnol.* **101,** 475–491 (2017).

352. Nkuipou-Kenfack, E., Zürbig, P. & Mischak, H. The long path towards implementation of clinical proteomics: Exemplified based on CKD273. *PROTEOMICS – Clin. Appl.* n/a-n/a (2017). doi:10.1002/prca.201600104

353. Berglund, L. *et al.* A genecentric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteomics MCP* **7,** 2019–2027 (2008).

354. Bordeaux, J. *et al.* Antibody validation. *BioTechniques* **48,** 197–209 (2010).

355. Bradbury, A. & Plückthun, A. Reproducibility: Standardize antibodies used in research. *Nature* **518,** 27–29 (2015).

356. Shigeyasu, K., Toden, S., Zumwalt, T. J., Okugawa, Y. & Goel, A. Emerging Role of MicroRNAs as Liquid Biopsy Biomarkers in Gastrointestinal Cancers. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* (2017). doi:10.1158/1078-0432.CCR-16-1676

357. Lippi, G., Simundic, A.-M., Rodriguez-Manas, L., Bossuyt, P. & Banfi, G. Standardizing in vitro diagnostics tasks in clinical trials: a call for action. *Ann. Transl. Med.* **4,** 181 (2016).

358. EvaluateMedTech® World Preview 2015, Outlook to 2020. http://info.evaluategroup.com/ evaluatemedtech-world-preview-2015. (2015).

359. Europe, MedTech. The European Medical Technology Industry-in figures. On line. (2015).

360. Rohr, U.-P. *et al.* The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report. *PloS One* **11,** e0149856 (2016).

361. Jannetto, P. J. & Fitzgerald, R. L. Effective Use of Mass Spectrometry in the Clinical Laboratory. *Clin. Chem.* **62,** 92–98 (2016).

362. Adaway, J. E. & Keevil, B. G. Therapeutic drug monitoring and LC-MS/MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life. Sci.* **883–884,** 33–49 (2012).

363. Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* **26,** 547–603 (2013).

364. Percy, A. J. *et al.* Clinical translation of MS-based, quantitative plasma proteomics: status, challenges, requirements, and potential. *Expert Rev. Proteomics* **13,** 673–684 (2016).

365. Hoofnagle, A. N., Becker, J. O., Wener, M. H. & Heinecke, J. W. Quantification of thyroglobulin, a low-abundance serum protein, by immunoaffinity peptide enrichment and tandem mass spectrometry. *Clin. Chem.* **54,** 1796–1804 (2008).

366. Percy, A. J., Mohammed, Y., Yang, J. & Borchers, C. H. A standardized kit for automated quantitative assessment of candidate protein biomarkers in human plasma. *Bioanalysis* **7,** 2991–3004 (2015).

367. Mohammed, Y., Percy, A. J., Chambers, A. G. & Borchers, C. H. Qualis-SIS: Automated Standard Curve Generation and Quality Assessment for Multiplexed Targeted Quantitative Proteomic Experiments with Labeled Standards. *J. Proteome Res.* **14,** 1137–1146 (2015).

368. Percy, A. J., Chambers, A. G., Smith, D. S. & Borchers, C. H. Standardized protocols for quality control of MRM-based plasma proteomic workflows. *J. Proteome Res.* **12,** 222–233 (2013).

369. Lebert, D. *et al.* DIGESTIF: a universal quality standard for the control of bottom-up proteomics experiments. *J. Proteome Res.* **14,** 787–803 (2015).

370. Tighe, P. J., Ryder, R. R., Todd, I. & Fairclough, L. C. ELISA in the multiplex era: potentials and pitfalls. *Proteomics Clin. Appl.* **9,** 406–422 (2015).

371. Ellington, A. A., Kullo, I. J., Bailey, K. R. & Klee, G. G. Antibody-Based Protein Multiplex Platforms: Technical and Operational Challenges. *Clin. Chem.* **56,** 186–193 (2010).

372. Haab, B. B. Applications of antibody array platforms. *Curr. Opin. Biotechnol.* **17,** 415–421 (2006).

373. Yu, X., Schneiderhan-Marra, N. & Joos, T. O. Protein microarrays for personalized medicine. *Clin. Chem.* **56,** 376–387 (2010).

374. Koivunen, M. E. & Krogsrud, R. L. Principles of Immunochemical Techniques Used in Clinical Laboratories. *Lab. Med.* **37,** 490–497 (2006).

375. Andrew John Li, M. D. New biomarkers for ovarian cancer: OVA1 and ROMA in diagnosis. *Contemporary OB/GYN* (2012). Available at: http://contemporaryobgyn.modernmedicine.com/contemporary-obgyn/news/modernmedicine/modern-medicine-feature-articles/new-biomarkers-ovarian-cance. (Accessed: 16th April 2017)

376. Newman, J. D. & Turner, A. P. F. Home blood glucose biosensors: a commercial perspective. *Biosens. Bioelectron.* **20,** 2435–2453 (2005).

377. Kojima, K. *et al.* Electrochemical protein chip with arrayed immunosensors with antibodies immobilized in a plasma-polymerized film. *Anal. Chem.* **75,** 1116–1122 (2003).

378. Wilson, M. S. Electrochemical immunosensors for the simultaneous detection of two tumor markers. *Anal. Chem.* **77,** 1496–1502 (2005).

Annexes

ANNEX 1

Table 17. Initial list of 506 proteins associated with endometrial cancer (based on the literature review described in Chapter 1). In bold, 158 proteins identified by DDA in uterine aspirate samples in Chapter 2. In blue, 52 proteins targeted by LC-PRM in Chapters 3-4.

Uniprot Number	Gene name	Uniprot Number	Gene name	Uniprot Number	Gene name
P04792	**HSPB1**	P10415	Bcl2	P33527	ABCC1/MRP
P07339	**CTSD**	P11802	CDK4	P35226	BMI-1
P12830	**CDH1**	P15692	VEGF	P35900	KRT20
P07355	**ANXA2**	P17813	ENG	P41159	LEP
P11166	**Glut1**	P24864	CCNE1	P48061	CXCL12
P61604	**HSPE1**	P35354	COX2	P49767	VEGF-C
Q14508	**WFDC2 / HE4**	P38936	P21	P52799	EFNB2
P09603	**M-CSF**	P42771	P16	P67870	CSNK2B
P14780	**MMP9**	P43405	SYK	P78396	CCNA1
P15941	**MUC1/CA 15-3**	P46013	MKI-67	P78504	JAG-1
Q13938	**CAPS**	P60484	PTEN	Q01196	RUNX1
P15328	**FOLR1**	Q15118	PDK1	Q02763	TEK
P08253	**MMP2**	Q92878	RAD50	Q03405	PLAUR
P04083	**ANXA1**	Q9Y251	HPSE	Q12983	BNIP3
P09211	**GSTP1**	O96017	pCHK2	Q14457	BECN1
P10451	**SPP1**	P13647	KRT5	Q15389	ANGPT1
P14618	**KPYM**	O00425	IGF2BP3	Q15506	SPA17
P30086	**PEBP1/ RKIP**	P00533	EGFR	Q15910	EZH2
P60174	**TPI1**	P01241	GH	Q6XLA1	CASC2
P07237	**P4HB**	P02771	AFP	Q8N307	MUC20
P16949	**STMN1**	P0DJI8	SAA1	Q96I34	PPP1R16A
P17931	**LGALS3**	P40763	pSTAT3	Q99814	EPAS1
P35222	**CTNNB1**	Q15465	SHH	Q9H013	ADAM19
Q01469	**FABP5**	O14493	CLDN4	Q9NR61	DLL4
Q01995	**TAGLN**	P06401	PR	Q9UMX1	SUFU
P16035	**TIMP-2**	P08254	MMP-3	Q9Y6Q5	AP1M2
P16070	**CD44**	P19544	WT1	Q9Y6Q9	AIB1
P01833	**PIGR**	P40692	MLH1	P08123	COL1A2
P00441	**SOD1**	P48023	FasL	P16422	EPCAM
P06733	**ENO1**	Q07864	POLE	P19404	NDUFV2
P08670	**VIM**	Q969H0	FBXW7	P35232	PHB
P80188	**LCN2**	P37275	ZEB1	Q10589	BST2
Q06830	**PRDX1**	Q15672	TWIST1	Q13162	PRDX4
P05787	**KRT8**	A2N4P8	IL-2R	O95967	EFEMP2
P22626	**HNRNPA2B1**	O95997	PTTG1	P09874	PARP1
P62937	**PPIA**	P01106	MYC	P11717	IGF2R
Q16658	**FSCN1**	P01236	PRL	P18827	SDC1
P05164	**MPO**	P02741	CRP	O00762	UbcH10
P42574	**CASP3**	P08476	INHBA	P0DJI9	SAA2
P43490	**NAMPT**	P14635	CCNB1	P10696	ALPP2
P00338	**LDHA**	P17948	FLT1	P52701	MSH6

Uniprot Number	Gene name	Uniprot Number	Gene name	Uniprot Number	Gene name
P09382	LGALS1	P21583	SCF	O14508	SOCS2
P30041	PRDX6	P21802	FGFR2	O43570	CA12
O00299	CLIC1	P27986	PIK3R1	O60259	KLK8
O43278	HAI-1 / SPINT1	P28906	CD34	O60911	CTSL2
P14174	MIF	P29317	EphA2	O95049	TJP3
P40121	CAPG	P31749	AKT1	O95832	CLDN1
P50454	SERPINH1	P35638	CHOP	P01112	HRAS
P55060	CSE1L	P42345	MTOR	P05093	CYP17
O75556	SCGB2A1 / MGB2	P43246	MSH2	P07585	DCN
P09466	PAEP	P46527	P27	P08581	MET
O43852	CALU	Q13635	PTCH1	P09110	ACAA1
P02787	TF	Q14865	ARID5B	P09769	FGR
Q8WXI7	MUC16 / CA125	Q15848	ADIPOQ	P10145	IL8
P11021	GRP78 / HSPA5	Q16665	HIF-1A	P11362	FGFR1
P30153	PPP2R1A	Q99835	SMO	P11678	EPX
P35268	RPL22	P02461	COL3A1	P14902	IDO1
P02647	APOA1	O00755	WNT7A	P15735	PHKG2
P05109	S100A8	O75390	CS	P17301	ITGA2
P09237	MMP7	O76070	SNCG	P18850	ATF6
P02766	TTR	P02538	KRT6A	P19838	NFKB1
P11388	TOP2A	P05362	ICAM1	P20809	IL11
P29508	SERPINB3	P12109	COL6A1	P21860	ERBB3
P07858	CTSB	P22223	CDH3	P23443	RPS6KB1
P01033	TIMP1	Q13753	LAMC2	P25054	APC
P21741	MDK	Q16629	SRSF7	P25116	PAR1
P0C0L4	C4A	Q99988	GDF-15	P27658	COL8A1
P09493	TPM1	Q93052	LPP	P31321	PRKAR1B
Q13421	MSLN	P06396	GSN	P35670	ATP7B
Q16851	UGP2	Q9H3D4	TP63	P41134	ID1
P01009	SERPINA1	P21980	TGM2	P42830	CXCL5
Q14624	ITIH4 (IHRP)	P24844	MYL9	P53350	PLK
P07951	TPM2	Q5TCI8	LMNA	P78556	CCL20
P26447	S100A4	O00559	EBAG9	Q02833	RASSF7
P04179	SOD2	O15120	AGPAT2	Q08345	DDR1
P39060	COL18A1	O15551	CLDN3	Q12756	KIF1A
P08758	PP4	O43597	SPRY2	Q12809	KCNH2
P12004	PCNA	O43683	BUB1	Q13111	CHAF1A
P12277	CKB	O54908	DKK1	Q13485	SMAD4
P30044	PRDX5	O96028	MMSET	Q13541	EIF4EBP1
Q14103	HNRNPD	P01135	TGFA	Q13761	RUNX3
P01023	A2M	P01137	TGFB1	Q14164	IKBKE
P02768	ALB	P01233	CGB	Q14315	FLNC
P10909	CLU	P04259	KRT6B	Q14626	IL11R
P01024	C3	P05019	IGF1	Q15915	ZIC1
O94788	ALDH1A2	P05111	INHA	Q16678	CYP1B1
Q8NBS9	TXNDC5	P05121	SERPINE1	Q16790	CA9

Uniprot Number	Gene name	Uniprot Number	Gene name	Uniprot Number	Gene name
P21926	CD9	P05231	IL-6	Q32Q65	FGB
P31947	SFN	P07333	CSF1R	Q53R41	FASTKD1
P00352	ALDH1A1	P08833	IGFBP-I	Q5TEJ8	THEMIS2
P19971	TYMP	P09038	FGF2	Q6NUR7	EZR
Q9NY33	DPP3	P10147	CCL3	Q86XX4	FRAS1
P04217	A1BG	P13236	CCL4	Q86YT6	MIB1
P06727	APOA4	P13501	CCL5	Q8N6T7	SIRT6
O95994	AGR2	P14373	TRIM27	Q8ND30	PPFIBP2
P05090	APOD	P16581	SELE	Q96D59	RNF183
P02654	APOC1	P19320	VCAM1	Q96GD4	AURKB
Q07654	TFF3	P19438	TNFRSF1A	Q99102	MUC4
Q9H6S3	EPS8L2	P19440	GGT1	Q99571	P2RX4
P00450	CP	P21810	BGN	Q99784	OLFM1
P02749	APOH	P22894	MMP-8	Q9BY76	ANGPTL4
P02751	FN1	P24385	CCND1	Q9NSC7	ST6GALNAC1
P02765	AHSG	P27338	MAOB	Q9NY64	GLUT8
P05198	EIF2S1	P29965	CD40LG	Q9P107	GMIP
P05783	KRT18	P33993	MCM7	Q9P2M7	CGN
P06703	S100A6	P35968	KDR	Q9UBT3	DKK4
P08107	HSPA1A	P41161	ETV5	Q9UBU3	GHRL
P09651	HNRNPA1	P42336	PIK3CA	Q9UIQ6	LNPEP
P15531	NME1	P46531	Notch1	Q9UM07	PADI4
P19338	NCL	P49789	FHIT	Q9Y263	PLAA
P19823	ITIH2	P50225	SULT1A1	Q16625	OCLN
P21333	FLNA	P51671	CCL11	P15924	DSP
P23142	FBLN1	P51911	CNN1	P09486	SPARC
P23526	AHCY	P52926	HMGA2	P24592	IGFBP6
P28838	LAP3	P54760	EphB4	P35613	BSG
P29966	MARCKS	P62736	ACTA2	Q01650	SLC7A5
P31943	HNRNPH1	P63261	ACTG1	Q13642	FHL1
P31949	S100A11	Q02962	PAX2	Q15303	ERBB4
P36222	CHI3L1	Q04609	PSMA	P21912	SDHB
P42330	AKR1C3	Q04743	EMX2	O00204	SULT2B1
P47756	CAPZB	Q13043	MST1	O00744	Wnt10b
P49327	FASN	Q15124	PGM5	O43294	TGFB1I1
P55072	VCP	Q15831	STK11	O43623	SNAI2
P62851	RPS25	Q16512	PKN1	O60894	RAMP1
Q01518	CAP1	Q29983	MICA	O75795	UGT2B17
Q12906	ILF3	Q92597	NDRG1	O95863	SNAI1
Q15121	PEA15	Q92731	ESR2	P05120	SERPINB2
Q16555	DPYSL2	Q99572	P2RX7	P07288	KLK3
Q99832	CCT7	Q9BZM4	ULBP3	P09544	WNT2
P24821	TNC	Q9BZM6	ULBP1	P10275	AR
P00738	HP	Q9UBP4	DKK3	P16112	ACAN
P02671	FGA	Q9Y6M1	IGF2BP2	P16662	UGT2B7

Uniprot Number	Gene name	Uniprot Number	Gene name	Uniprot Number	Gene name
O43490	**PROM1**	P06731	CEACAM5	P17252	PKCA
P00918	**CA2**	Q13309	SKP2	P24462	CYP3A7
P01008	**SERPINC1**	O43291	SPINT2	P26715	KLRC1
P01042	**KNG1**	P00750	PLAT	P27701	CD82
P02452	**COL1A1**	P01111	NRAS	P29536	LMOD1
P02652	**APOA2**	P01344	IGF2	P35749	MYH11
P02675	**FGG**	P09529	INHBB	P41221	WNT5A
P02679	**FGG**	P09758	TACSTD2	P49765	VEGF-B
P02774	**GC**	P17936	IGFBP3	P51692	STAT5B
P02775	**PPBP**	P21964	COMT	P54852	EMP3
P02788	**LTF**	P26006	ITGA3	P58335	ANTXR2
P02790	**HPX**	P35625	TIMP3	P78545	ELF3
P07108	**DBI**	P48509	CD151	Q01826	SATB1
P09455	**CRBP-1**	Q99623	PHB2	Q02224	CENPE
Q6FHJ7	**SFRP4**	Q9UM47	Notch3	Q02413	DSG1
Q9HC84	**MUC5B**	P35916	FLT4	Q02447	SP3
P02776	**PF4**	O14746	TERT	Q03135	CAV1
P22392	**NME2**	O14965	AURKA	Q03167	TGFBR3
P26038	**MSN**	O15123	ANGPT2	Q07820	Mcl-1
P67936	**TPM4**	O60216	RAD21	Q12964	MUC8
P22352	**GPX3**	O60271	SPAG9	Q13641	TPBG
P30043	**BLVRB**	P01116	KRAS	Q6UXB2	CXCL17
Q15436	**SEC23A**	P01375	TNF	Q8IUK5	PLXDC1
P40261	**NNMT**	P05412	JUN	Q8N2R0	OSR2
P02649	**APOE**	P06400	pRb	Q96AM1	MRGPRF
P06396	**GSN**	P08069	IGF1R	Q96D21	RASD2
P00749	PLAU	P08151	GLI1	Q99527	GPER
P32004	L1CAM	P10070	GLI2	Q9GZU2	PEG3
P35221	CTNNA1	P10071	Gli-3	Q9NYL2	ZAK
Q07812	BAX	P10914	IRF-1	Q9UHF7	TRPS1
P17661	DES	P11511	CYP19A1	Q9UMD9	COL17A1
O14497	ARID1A	P16234	PDGFRA	Q9UMS6	SYNPO2
O15392	BIRC5	P22309	UGT1A1	Q9UNQ0	ABCG2
P03372	ESR1	P23219	COX1	Q9Y243	AKT3
P04626	ERBB2	P25063	CD24		
P04637	TP53 (p53)	P25445	Fas		
P06493	CDK1	Q9UGI8	TES		
P08243	ASNS	Q9NQW6	ANLN		

ANNEX 2. Table 18. Main characteristics of the 106 validation studies reviewed in Chapter 1. IHC, Immunocytochemistry; TMA, tissue microarray; AUB, abnormal uterine bleeding; MRM, multiple reaction monitoring; N/A, not available; ROC curve, receiver operating characteristic curve; AUC, are under the ROC curve.

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Michihiro et al, Acta Cytologica, 2012	21 EC	111: 29 proliferative + 27 secretor + 24 atrophic + 31 benign disorder	Tissue	IHC	**NOTCH1**		p-value, 88% sensitivity- 90%specificity	
Guissepe et al, Gynecologic Oncology, 2012	12 EEC	12 benign disease	Tissue	IHQ, WB	**HSPA5, ATF6**		p-value	
Amalinei C et al., Annals of Anatomy, 2011	17 EC	21: 10 non-neoplastic endometria, 11 endometrial hyperplasias	Tissue	IHQ	**ESR1, PGR, FAS, FASLG, BCL2, BAX, TP53, MMP2, MMP9, TIMP2**	**TIMP1**	p-value	
Ioachim et al, International Journal of Gynecological Cancer, 2003	79 EC	67: 18 proliferative + 14 secretory + 35 hyperplasias	Tissue	IHQ		**CTSD**	p-value	
Gadducci et al, European Journal of Gynaecological Oncology, 1990	42EC	68 benign uterine pathology	Serum	ELISA	**CA125, CA72.4**		p-value	
DeSouza et al, Analytical Chemistry, 2009	10 EC	10 normal endometrium	Tissue	Targeted MS: MRM	**PKM, PIGR**		Fold-change	
Li et al, Oncotarget, 2015	87: a. 72 EC, b. 15 EC	47: a. 32 normal endometrium, b. 15 healthy	a. Tissue, b. Serum	a. IHQ, b. ELISA	**GPI**		p-value	
Saito et al, Pathology International, 2007	85 EC	62: 18 proliferative + 13 secretory + 31 benign	Tissue	IHQ	**VEGFA**		p-value	
Krakstad et al, Oncotarget, 2015	564 EC	18 hyperplasias	Tissue	IHQ, TMA	**ATAD2**		p-value	
Li et al, International Journal of Oncology, 2015	31 EC	31 normal endometrium	Tissue	IHQ	**GPI**		p-value	
Shaarawy et al, Acta Oncologica, 2001	72 EC	57: 30 healthy + 27 hyperplasias	Serum	ELISA	**VEGFA**		p-value	
Wojciechowski et al, Archives of Gynecology and Obstetrics, 2015	51 EC	20 normal endometrium	Tissue	IHQ	**CD44**		p-value	
Peiró et al, Int J Gynecol Pathol., 2001	89 EC	56: 24 atrophic + 32 hyperplasias	Tissue	IHQ		**BAX, CASP3**	p-value	

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Driák et al , Folia-Biol Prague, 2011	6 EC	19: 8 proliferative+ 5 atrophic + 6 hyperplasias	Tissue	WB	**BCL2, BAX, CASP3**		p-value	
Chen et al, Gynecologic Ongology, 2013	97 EC	16 normal endometrium	Tissue	IHQ	**MUC20**		p-value	
Bian et al, Technol Cancer Res Treat., 2016	105 EC	87 healthy	Serum	Electrochemilumin escence	**HE4, CA125, CA72.4, CA19-9**		p-value, ROC analysis (AUC)	
Peng et al, Ai Zheng., 2002	39 EC	80 healthy	Serum	ELISA	**VEGFA**		p-value	
Gun et al, Diagnostic Pathology, 2012	47 EEC	10 proliferative	Tissue	IHQ	**FSCN1**		p-value	
Reithdorf et al, Virchows Arch, 1996	120 EEC	54: 15 normal endometrium + 39 hyperplasias	Tissue	IHQ	**TP53**		% of cases with p53 accumulation	
Chen et al, Cancier Sciencie, 2008	103 EC	74: 20 proliferative + 20 secretory + 34 hyperplasias	Tissue	IHQ	**NDRG1**	**PTEN**	p-value	
Erkanli et al, Gynecologic Oncology, 2007	50 EEC	50: 20 proliferative + 30 hyperplasias	Tissue	IHQ	**COX2, BIRC5**		p-value	
Shevra et al, Journal of Postgraduate Medicine, 2015	20 EC	41: 10 normal endometrium + 31 hyperplasias	Tissue	IHQ	**MKI67**		p-value	
Li et al, Gynecologic Oncology, 2007	43: a. 40 EC, b. 3 EC	44: a. 28 normal endometrium + 13 hyperplasias, b. 3 NE	Tissue	a. IHQ, b. WB		**P2RX7**	p-value, fold change, 89-100% sensitivity-90-100% specificity	
Yurkovetsky et al, Gynecologic Oncology, 2007	115 EC	135 healthy	Serum	ELISA, multiplex bead-based immunoassay	**TNFRSF1A, IL2R, TSH, PRL, CA125, CA19-9, MMP7, SERPINB3, SAA1**	**VEGFA, ERBB2, AFP, CGA/FSH, CGA/LH, ICAM1, SERPINE1, ADIPOQ, MMP2, MMP9, TTR, FASLG**	p-value; **PRL**: 98,3% sensitivity and 98% specificity	
Cho et al, Cancer Investigation, 2009	88: a. 32 EC, b. 56 EC	166: a. 12 normal endometrium, b. 154 healthy	a. Tissue, b. Serum	a. IHQ, b. ELISA	**SPP1**		p-value, ROC analysis (AUC)	
Bignotti et al, British Journal of Cancer, 2011	291: a. 153 EC, b. 138 EC	109: a. 16 proliferative + 17 secretory, b. 76 atrophic	a. Tissue, b. Serum	a. TMA, b. ELISA	**HE4, CA125**		p-value, sensitivity-specificity	
Kemik et al, Gynecologic Oncology, 2016	50 EC	50 normal endometrium with AUB	Serum	ELISA	**CA125, CHI3L1/YKL-40, HE4**		p-value, ROC analysis	

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Simaga et al, Eur J Cancer., 1998	61: a. 21 EC, b. 40 EC	32: a. 15 normal endometrium, b. 17 NE	a. Tissue, b. Serum	a. WB, B. ELISA	**DPP3**		p-value, fold change	
Hevir et al, Mol Cell Endocrinol, 2011	49: a. 15 EC, b. 18 EC, c. 16 EC	45: a. 15 normal endometrium, b. 14 NE, c. 16 EC	Tissue	a. WB b. IHQ c. WB		**CYP1B1**	p-value, fold change	
Avall Lundgvist et al, European Journal of Gynecological Oncology, 1989	61 (cervical carcinoma + ovarian carcinoma +	104 benign	Serum	ELISA	**SERPINB3, CA125, CEA**		N/A	
Deng et al, Journal of Experimental and Clinical Cancer Research, 2015	84 EC	48: 9 proliferative + 9 secretory + 30 hyperplasia	Tissue	IHQ	**ANXA2, HE4**		p-value	
Yu et al, Int J Gynecol Cancer. 2012	97: a. 47 EC, b. 50 EC	74: a. 22 normal endometrium + 22 hyperplasia, b. 10 healthy + 20 benign	a. Tissue, b. Serum	a. IHQ, WB, b. ELISA	**SPAG9**		p-value	
Xiong et al, European Journal of Gynecologic Oncology, 2010	24 EEC	93: 10 proliferative + 83 hyperplasias	Tissue	IHQ	**CTNNB1, SLC2A1 (GLUT1)**		N/A	
Yang Liu et al,Gynecologic Oncology, 2014	110: a. 53 EC b. 57 EC	168: a. 78: 42 normal endometrium, 36 atypical hyperplasia b. 90: 52 NE, 38 atypical hyperplasia	a. Endometrial tissue b. Serum + peritoneal washings	a. IHQ b. ELISA	**BGN**		p-value	
Chen H. et al, Oncology Reports, 2013	102 EC	138: 48 proliferative, 36 secretor, 54 hyperplasias	Tissue	IHQ	**WNT10B**		p-value	
Kurai M. et al, Human Pathology, 2005	73 EC	70: 30 hyperplasias, 15 proliferative, 25 secretor	Tissue	IHQ, WB	**AURKA**		p-value	
Pan XY et al, Int J Gynecol Cancer, 2007	30 EEC	97: 25 proliferative, 25 secretory, 47 hyperplasias	Tissue	IHQ, WB	**CLDN3, CLDN4**		p-value	
Yu CG et al, Asian Pacific Journal of Cancer Prevention, 2015	67 EC	106: 53 normal endometrium, 53 atypical hyperplasias	Tissue	IHQ	**ERBB2, MKI67**	**ESR1, PGR**	p-value	
Li X et al., Biomed Research International, 2015	102 EC	50: 30 hyperplasias, 10 secretory, 10 proliferative	Tissue	IHQ	**HE4**		p-value	
Wahl H et al., Gynecologic Oncology, 2010	75 EC	20: benign endometrial samples	Tissue	TMA	**SLC2A1 (GLUT1)**		p-value	

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Feng ZZ et al., Medical Oncology, 2012	124 EEC	63: 16 proliferative, 19 secretory, 28 atypical hyperplasias	Tissue	TMA		**PTEN**	p-value	
Erkanil S et al., Int J Gynecol Cancer, 2006	29 EC	48: 38 hyperplasias, 10 proliferative	Tissue	IHQ	**BIRC5**		p-value	
Eskander RN et al., Int J Gynecol Cancer, 2016	60 EEC	27: NE	Tissue	WB		**DKK3**	p-value	
Erdem O et al., Int J Gynecol Cancer, 2007	31 EC	35: 12 proliferative, 23 hyperplasias	Tissue	IHQ	**VEGFA**		p-value	
Eissa S et al., Clinical Biochemestry, 1997	32 EC	141 with AUB: 79 hyperplasias, 28 secretory, 34 proliferative	Tissue	Enzyme imunoassay	**ERBB2, TP53**		p-value	
Senol S et al., Int J Clin Exp Pathol, 2015	95 EEC	88: 58 hyperplasias, 30 secretory and proliferative	Tissue	TMA	**FOLR1**		p-value	
Qiu M et al., BMC Cancer, 2014	76 EC	68: 57 normal endometrium, 11 atypical hyperplasias	Tissue	IHQ	**FOXA1, AR**		p-value	
Brustmann H et al., Pathol Res Pract. 2003	49 EC	52: 8 proliferative, 4 secretory, 40 hyperplasias	Tissue	IHQ	**LGALS3**		p-value	
Goldman NA et al., Modern Pathology, 2006	27 EC	38: 13 atrophic, 13 proliferative, 12 secretory	Tissue	Immunoblot, Immuno-fluorescense	**SLC2A8 (GLUT8)**		p-value	
Zagorianakou N et al., Eur J Gynaecol Oncol, 2003	100 EEC	86: 40 hyperplasia, 46 normal endometrium	Tissue	IHQ	**CD44**		p-value	
Lawicki S et al., Growth Factors, 2012	65 EC	94: 40 uterine myoma, 45 normal endometrium	Plasma	a. ELISA, b. Chemiluminescent microparticle immunoassay	**CSF1, CA125**		p-value	
Cherubini A et al., British Journal of Cancer, 2000	18 EC	11: 1 proliferative, 4 secretory, 6 hyperplasias	Tissue	IHQ	**HCNH2**		p-value	
Zanotti L et al., Clinical Chemistry and Laboratory Medicine, 2012	193 EC	125 normal endometrium: 24 premenopausal, 101 postmenopausal	Serum	Chemiluminiscent microparticle immunoassays	**HE4, CA125**		p-value, ROC analysis (sensitivity-specificity)	**HE4+CA125**. 90% specificity, 76% sensitivity

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Elbasateeny SS et al., Pathology, Research and Practice, 2016	62 EC	15: 15 proliferative	Tissue	IHQ	**CD44**		p-value	
Zagorianakou N et al., Eur J Gynaecol Oncol, 2003	65 EC	59: 28 hyperplasias, 31 proliferative and secretory	Tissue	IHQ		**HSPB1**	p-value	
Wang BY et al., Cancer, 2000	41EC	58: 9 proliferative, 8 secretory, 41 hyperplasias	Tissue	IHQ	**SLC2A1 (GLUT1)**		N/A	
Horowitz IR et al., Int J Gynecol Cancer, 2001	21 EC	60: 19 premenopausal, 41 postmenopausal	Plasma, Tissue	ELISA, IHQ	**PAEP**		p-value	
Chang X et al., BMC Cancer, 2009	32 EC	76 hyperplasias	Tissue, blood	IHQ, ELISA	**PADI4**		p-value	
Elsässer-Beile U et al., Tumour Biology, 1994	32 EC	241 normal endometrium	Plasma	ELISA	**TNFRSF1A, IL2R**		p-value	
Yap J et al., Reprod Biol Endocrinol., 2010	16 EC	14: 4 atrophic, 10 proliferative	a. Uterine lavage fluid, b. Tissue	a. ELISA, b. IHQ	**IL11**		p-value	
Ege CB et al., Arch Gynecol Obstet., 2011	64 EC	40: 20 hyperplasias, 10 proliferative, 10 secretory	Tissue	IHQ		**LGALS3**	p-value	
Martinho O et al., J Clin Pathol., 2012	209 EC	97: 49 polyps, 48 hyperplasias	Tissue	TMA		**PEBP1**	p-value	
Walmer DK et al., Cancer Res., 1995	12 EC	39: 22 cycling endometrium , 7 atrophic endometrium, 10 hyperplasias	Tissue	WB, IHQ	**LTF**		Increased staining	
Graesslin I et al., Ann Oncol., 2006	38 EC	59: 10 proliferative, 10 secretory, 39 hyperplasias	Tissue	IHQ	**MMP2, MMP9**	**MMP7, TIMP1, TIMP2**	p-value, sensitivity-specificity	
Tanabe K et al., Cancer Sci., 2008	205: a. 85 EC b. 120 EC	76: a. 21 proliferative, 12 secretory, b. 45 normal endometrium	a. Tissue, b. Serum	a. IHQ, b. ELISA	**MDK**		p-value	
Colas E et al., Int J Cancer, 2011	70: a. 70 EC b. 4EC	11: a. 4 atrophic, 3 proliferative, 1 secretory, 3 hyperplasias; b. 4 adjacent normal tissue	Tissue, uterine fluid	a. TMA, b. WB	**P4HB**		p-value, 100% sensitivity-87,5% specificity	

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Berg et al, Oncotarget, 2014	650: a. 156 EEC, b. 494 EEC	95: a. 18 atypical hyperplasia, premenopausal, b. 77 atypical hyperplasias, premenopausal	Tissue	a. TMA, b. FISH, c. RPPA	**STMN1, PIK3CA, PTEN**		p-value	
DeSouza et al, Proteomics, 2010	10 EEC	15 proliferative	Tissue	Targeted MS: MRM (mTRAQ)	**PIGR**	**PKM**		
DeSouza et al, Journal of Proteome Research, 2008	4: a. 1 EEC, b.3EEC	6: a. 3 normal proliferative, b. 3 normal proliferative	Tissue	a. MRM (mTRAQ), b. ELISA	**PKM**		Fold-change	
Panici et al, Gynecol Obstet Invest 1989	47 EEC	20 hyperplasia	Plasma	ELISA	**CEA, CA125, CA15-3**		p-value	
Faloppa et al, Am J Clin Pathol., 2014	218 EEC	107: 53 polyps, 37 hyperplasia, 17 EIN	Tissue	TMA		**COX2**	p-value	
Shang et al, Tumour Biol., 2014	207 EC	85 normal endometrium	Tissue	IHC	**ATAD2**		p-value	
Pan XY et al, Oncol Lett, 2013	62 EEC	60: 34 proliferative, 26 secretory	Tissue	IHC	**CLDN4**		p-value	
Dainty et al, Gynecol Oncol, 2007	58 EC	10 normal endometrium	Tissue	IHC	**FOLR1**		p-value	
Allard et al, Gynecol Oncol, 2007	485 EC	60: 50 normal endometrium, 10 ovarian cancer	Tissue	TMA	**FOLR1**		p-value	
Zhao et al, Tumour Biol, 2014	188 EC	84: 43 beningn lesions, 41 normal endometrium	Tissue	a. IHC, b. WB	**HABP1**		p-value	
Sivridis et al, Histopathology, 2002	111 EC	87: 42 normal endometrium, 45 hyperplasias	Tissue	IHC		**MUC1**	p-value	
Kanat-Pektas et al, Arch Gynecol Obstet, 2009	135 EC	135 normal endometrium	Blood	Electrochemilumin escence	**TSH, PRL, CA-125, CA 19-9**	**AFP, CA 15-3**	CA125: 42,5% sensitivity-87,4% specificity	
Diefenbach et al, Gynecol Oncol, 2006	34 EC	44 normal endometrium	Blood	ELISA	**YKL-40**		p-value, 76% sensitivity-89% specificity	
Zhengyu et al, Int J Cancer, 2008	84 EC	68: 39 normal endometrium, 29 EIN	Tissue	a. MS, b. IHQ	**E-FABP, CAPS**		p-value	

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Mitsuhashi et al, Histopathology., 2012	76 EC	37 normal endometrium	Tissue	IHC	**Notch1**		p-value	
Ilhan et al, Asian Pac J Cancer Prev, 2015	42 EC	42 controls with suspicion of EC (AUB)	Serum	ELISA	**NAMPT**		p-value	
Cocco et al, Cancer, 2010	112: a. 10 EEC, b. 102 EEC	102: a. 10 normal endometrium, b. 92: 50 normal endometrium, 22 benign	a. Tissue, b. Blood	a. IHC, b. bead-based immunoassay	**SAA1**		p-value	
Dobrzycka et al, Eur Cytokine Netw. , 2001	98 EC	30 normal endometrium	Blood	ELISA	**VEGF, VEGFC**		p-value	
Konno et al, Clin Cancer Res., 2000	18 EC	24 normal endometrium	Blood	ELISA	**FAS**		p-value	
Li et al, Gynecol Obstet Invest., 2016	8 EC	19 hyperplasias	Tissue	IHC	**SPAG9**		74% sensitivity-83% specificity	
Laas et al, Gynecol Oncol. , 2014	20 EC	22 hyperplasias	Tissue	IHC	**MMP9, Bcl2**	**ER, PGR , CD44v6**	p-value	
Jóźwik et al, BMC Cancer, 2015	42 EC	43: 23 normal endometrium, 20 benign disorders	Blood	Flow cytometry	**ICAM1/CD54**		p-value	
Kacinski et al, Int J Radiat Oncol Biol Phys, 1990	24 EC	11 secretory	a. Tissue, b. Blood	a. IHC, b. Radio-immunoassay	**CSF1**		Increased staining	
Omer et al, Tumour Biol., 2013	64 EC	60 normal endometrium	Blood	Immuno-nephelometry	**HE4, SAA**		sensitivity-specificity	**HE4+SAA:** 73.3% sensitivity, 64% specificity
Matsuo et al, Gynecol Oncol., 2013	179 EC	15 benign endometrium	Tissue	IHC	**GRP78/HSPA5**		p-value	
Li et al, Cancer Epidemiol Biomarkers Prev., 2006	29 EC, 16 cervical cancer	42 normal: 15 endometrial, 27 cervical tissues	Tissue	a. WB, b. IHC		**P2RX7**	sensitivity-specificity	
El Behery et al, Arch Gynecol Obstet, 2010	10 EC	110 benign endometrium (with AUB)	Uterine fluid	LDH isoenzyme activity profile	**LDH isoenzymes**		100% sensitivity, 90.1% specificity	
Angioli et al, Tumour Biol., 2013	101 EC	103 benign uterine diseases	Blood	ELISA	**HE4, CA125**		p-value	**HE4+CA-125**: 60.4% sensitivity, 100% specificity
Jiang et al, Journal of Obstetrics and Gynaecology Research, 2010	28 EC	30 normal endometrium	Serum	ELISA	**DKK3**		p-value	

Study	EC patients	Non-EC control patients	Clinical sample	Proteomic technique	Up-regulated proteins in EC *vs* controls	Down-regulated proteins in EC vs controls	Statistical criteria	Protein panels
Moore et al, Gynecol Oncol., 2008	171 EC	156 healthy postmenopausal	Blood	Immunoassays	**HE4 , CA125**		sensitivity-specificity	**HE4+CA-125**: 57.3% sensitivity, 90% specificity
Apostolou et al, Diagn Cytopathol., 2014	48 EC	59: 10 disordered proliferative, 13 hyperplasias, 36 normal endometrium	Tissue	IHC		**Bcl2**	p-value	
Farias-Eisner et al, Am J Obstet Gynecol., 2010	343 EC	90 normal endmetrium	Blood	Immunoassay	**ApoA-I, TTR, TF**		sensitivity-specificity	**ApoA-I+TTR+TF:** 71% sensitivity, 88% specificity
Nakopoulou et al, Gynecol Oncol., 1990	39 EC	43 hyperplasias	Tissue	IHC	**VIM**		p-value	
Li et al, Clin Chim Acta., 2016	74 EC	775: 118 normal endometrium, 637 benign uterine diseases	Blood	ARCHITECT Immunoassays	**HE4, CA125**		sensitivity-specificity	**HE4+CA125:** 74.3% sensitiv, 78.6%specif (pre); 32.9%sensitiv, 94.4% specif (post)
Dubé et al, J Proteome Res., 2007	53 EC	95: 10 hyperplasias, 25 proliferative, 25 secretory, 25 atrophic, 10 menstrual	Tissue	TMA	**AAT/SERPINA1, HSPE1, PKM**	**PIGR**	sensitivity-specificity	**AAT/SERPINA1 + CPN10+ PKM2**: 85% sensitivity, 93% specificity
Tian et al, Gynecol Oncol., 2013	234 EC: a. 70 EC, b. 114 EC	a. 120 normal endometrium, b. 135: 49 hyperplasias, 86 normal endometrium	a. Blood, b. Tissue	a. ELISA, b. TMA	**NAMPT**		p-value, sensitivity-specificity	
Wang et al, Tumour Biology, 2015	a. 7 EC , b. 95 EC	37: a. 7 matched normal endometrium, b. 26 hyperplasias + 10 normal endometrium	Tissue	a. WB, b. IHQ	**KHDRBS1**		p-value	
Nagymanyoki et al, Cancer Citopathology, 2015	17 EC	16 endometriosis	Peritoneal washings	IHQ		**ARID1A**	p-value, 47% sensitivity-100% specificity	
Mao et al, American Journal of Surgical Pathology, 2013	143 EEC	103: 29 proliferative, 16 secretory, 4 atrophic, 2 gestational, 38 hyperplasias, 14 polyps	Tissue	IHQ		**ARID1A**	p-value	
Soliman et al, Cancer, 2006	117 EC	238 normal endometrium	Serum	ELISA		**ADIPOQ**	p-value	

ANNEX 3

Main achievements of this thesis

❖ Publications

"ELISA simplification and shortening by using a polymeric signal amplifier. Application to MMP-9 detection in plasma and uterine aspirates*" (partially described in Chapter 5).* de la Serna E, Martinez-Garcia E, García-Berrocoso T, Penalba A, Gil-Moreno, Colas E, Montaner J, Baldrich E **(under revision).**

"Targeted proteomics identifies proteomic signatures in liquid-biopsies of the endometrium to diagnose endometrial cancer and assist in the prediction of the optimal surgical treatment" *(described in Chapter 4).* Martinez-Garcia E, Lesur A, Devis L, Cabrera S, Matias-Guiu X, Hirschfeld M, Asberger J, van-Oostrum J, María de los Ángeles Casares de Cal MA, Gómez-Tato A, Reventos J, Domon B, Colas E, Gil-Moreno A**. Clinical Cancer Research 2017 (accepted in April).**

"Development of a Sequential Workflow based on LC-PRM for the Verification of Endometrial Cancer Protein Biomarkers in Uterine Aspirate Samples" *(described in Chapter 3).* Martinez-Garcia E, Lesur A, Devis L, Rosa Campos A, Cabrera S, van-Oostrum J, Matias-Guiu X, Gil-Moreno A, Reventos J, Colas E, Domon B. **Oncotarget. 2016 Aug 16;7(33):53102-53115**

❖ Patents

"Markers of endometrial cancer"

Inventors: Martinez-Garcia, E; Colas E; Gil-Moreno A; Reventós, J; Domon B¸ Lesur A
Holding Institution: Vall d´Hebrón Research Institute (Spain), Luxembourg Institute of Health (Luxembourg) and Biomedical Research Institute of Lleida Foundation Dr. Pifarré (Spain).
Priority application number: EP16168328.9
Priority application date: 04th May 2016
PCT Application number: PCT/EP2017/057635
PCT application date: 30th March 2017

"Markers for the differential diagnosis of endometrial cancer"

Inventors: Martinez-Garcia, E; Colas E; Gil-Moreno A; Reventós, J; Domon B¸ Lesur A
Holding Institution: Vall d´Hebrón Research Institute (Spain), Luxembourg Institute of Health (Luxembourg) and Biomedical Research Institute of Lleida Foundation Dr. Pifarré (Spain).
Priority application date: to be filed in June 2017

❖ Internships

During a total of 17 months, I worked in the Laboratory of Proteome and Genome Research Unit at the Luxembourg Institute of Health (Luxembourg), under the supervision of Dr. Bruno Domon. Here I performed the sample preparation and LC-MS analysis of the studies described in Chapters 2 to 4.

❖ Grants awarded to continue with the development of the asset

- CaixaImpulse (Caixa Bank): prototype development and mentoring activities (160 h of training, networking and mentoring that we received during December 2016-March 2017).
- FIPSE (Fundación para la Innovación y la Prospectiva en Salud en España): freedom to operate analysis and intellectual property costs. Mentoring activities (February-July 2017).
- VALUNI (Generalitat de Catalunya): regulatory and market studies, attendance to partnering, lawyers (January-December 2017).
- MAP EADA ACCIÓ (EADA business school and Agencia para la Competitividad de la Empresa): development of the business plan (2017-2018).
- IDEA2 GLOBAL PROGRAM (Massachusetts Institute of Technology, MIT): mentoring and networking in Madrid and Boston (March-December 2017. The first meeting I attended took place in Boston on the 5th-6th June 2017).
- Mind the gap (Botín Foundation, Santander Bank): launching of business initiatives.

Other publications in collaboration

"ALCAM shedding at the invasive front of endometrioid endometrial cancer is a marker of cancer progression, promotes myometrial invasion and tumour dissemination". Devis L, Martinez-Garcia E, Moiola C, Quiles M.T, Arbos M.A., Vasilica Stirbat T, Brochard-Wyart F, García A, Alonso-Alconada L, Abal M, Diaz-Feijoo B, Thomas W, Dufour S, Mancebo G, Alameda F, Reventos J, Gil-Moreno A, Colas E. **Under revision**

"Activated leukocyte cell adhesion molecule (ALCAM) is a marker of recurrence and promotes cell migration, invasion and metastasis in early stage endometrioid endometrial cancer". Devis L, Moiola C, Masia N, Martinez-Garcia E, Santacana M, Stirbat Tomita V, Brochart-Wyart F, Garcia A, Alameda F, Cabrera S, Palacios J, Moreno-Bueno G, Abal M, Thomas W, Dufour S, Matias-Guiu X, Santamaria A, Reventos J, Gil-Moreno A, Colas E**. Journal of Pathology. 2017. Mar;241(4):475-487**

"Nidogen 1 and Nuclear Protein 1: novel targets of ETV5 transcription factor involved in endometrial cancer invasion". Pedrola N, Devis L, Llauradó M, Campoy I, Martinez-Garcia E, Garcia M, Muinelo-Romay L, Alonso-Alconada L, Abal M, Alameda F, Mancebo G, Carreras R, Castellví J, Cabrera S, Gil-Moreno A, Matias-Guiu X, Iovanna JL, Colas E, Reventós J, Ruiz A**. Clinical and experimental metastasis. 2015 Jun;32(5):467-78**

"The EMT signaling pathways in endometrial carcinoma". Colas E, Pedrola N, Devis L, Ertekin T, Campoy I, Martinez-Garcia E, Llauradó M, Rigau M, Olivan M, Garcia M, Cabrera S, Gil-Moreno A, Xercavins J, Castellvi J, Garcia A, Ramon y Cajal S, Moreno-Bueno G, Dolcet X, Alameda F, Palacios J, Prat J, Doll A, Matias-Guiu X, Abal M, Reventos J. **Clinical & translational oncology. 2012 Oct;14(10):715-20**

Acknowledgements

Aún no me creo que haya llegado este momento. Tengo la impresión de que fue ayer cuando entraba en el VHIR sin estar segura de si la decisión que había tomado de hacer una tesis doctoral era la más adecuada. Han pasado casi 5 años desde entonces y ahora estoy convencida de que fue una buena decisión. Cuando echo la mirada atrás veo mucho esfuerzo, frustraciones y días de estar a punto de dejar todo y buscar algo totalmente diferente. Pero sobre todo veo un trabajo distinto a todos los demás, donde tú mismo te marcas las metas y ambiciones. Veo 5 años de aprendizaje continuo, de un gran reto personal, de una altísima formación científica y humana. Y veo muchos buenos momentos que han sido posibles gracias a todas las personas que han estado conmigo en este viaje, y a las que estoy enormemente agradecida.

En primer lugar me gustaría agradecer a mis directores de tesis, el Dr. Jaume Reventós, la Dra. Eva Colás y el Dr. Antonio Gil, por acogerme en esta familia "endométrica", valorarme, creer en mí y darme la oportunidad de hacer este proyecto. A **Jaume**, por tus buenos consejos y nuestras intensas conversaciones de política y fútbol (sigo siendo del Madrid, lo siento; pero ya reconocí que me gusta el himno del Barça, por algo se empieza). A **Eva**, por el apoyo, la tranquilidad y el optimismo que me has transmitido todos estos años y por despertar mi espíritu emprendedor. A **Antonio**, por tu sonrisa constante, el apoyo y el cariño que has dado al grupo estos últimos años.

Al servicio de Anatomía Patológica y de Ginecología del Hospital, y en especial a **Berta** y **Silvia**, por toda vuestra ayuda, paciencia y vuestro interés en el proyecto. A todas las pacientes que voluntariamente han querido participar en los estudios, porque sin ellas esto no habría sido posible. A **Anna, Mire** y **Miquel**, por los buenos consejos no sólo en las evaluaciones de seguimiento, sino en el día a día, en el pasillo, comiendo o tomando un café. Porque a veces cinco minutos de conversación con vosotros fueron claves para continuar.

Por supuesto agradecer a mis compañeros del lab, los que pasan a ser tu nueva familia durante la tesis. Sin duda ellos han sido la gran motivación para ir feliz al trabajo los días en los que los resultados no salían como querías (y de estos hay unos cuantos!). Ellos son con los que pasas la mayor parte del día y aún así con los que quieres ir a comer unos pinchos después del trabajo. A **Ire**, mi Jire, mi siamesa, porque contigo he vivido muchos de los mejores momentos en el lab: nuestros días de risas en cultivos con Delfi, nuestro burbujal...Brugal! y tu EHHHHHH que me hace reír hasta cuando lo escribo!; nuestras miles de frases sin desperdicio; nuestras risas escandalosas; por tener el honor de ser las nuevas paseantes; por mil y una razones pero, sobre todo, por ser la primera en estar ahí para ayudarme con cualquier cosa. A **Lauri**, por tu ayuda constante y tus buenos consejos, por estar siempre dispuesta a escuchar y a echar una mano, porque te he mareado como a

nadie en estas últimas semanas con los papeleos de la tesis pero siempre contestas con una sonrisa. A **Lucía**, por ser tan buena amiga, por nuestras despedidas diarias de 2h en el metro de Diagonal, por nuestras tardes de playa y compras, por nuestras risas dentro y fuera del lab. Mis motivaciones para salir de la cama cada día: **Nuri**, mi compi de bus, por sacarme una sonrisa desde primera hora de la mañana y por hacer que te eche tanto de menos cada vez que me voy de Barcelona; **Eli** y **Juli**, porque nuestros cafés mañaneros terapéuticos (con Ire) me daban fuerza cada día, porque no hay nada mejor que tener a una psicóloga y a unas chicas siempre sonrientes cerca mientras haces la tesis. A **Blanca**, por tus consejos en el lab, tu fuerza y optimismo, por tener la puerta de tu casa siempre abierta para pasar un rato genial con nosotras: las tardes en tu jardín, las calçotadas en l'Ampolla, los días de playa y de buen comer en Llavaneras. A **Tati**, mi pequeña hiperactiva, por la energía que das al laboratorio y las miles de risas con el singstar en tu casa. A **Gabriel**, por tus saludos siempre con una sonrisa, tus bromas y tus ayudas en momentos críticos de algún experimento. A **Alfonso**, por estar siempre ahí para hablar y escuchar, el mejor para animar en los malos momentos. A **Berta**, por tu increíble optimismo, tu eterna sonrisa y tus exquisitas tartas. A **Cristian**, por la alegría y la tranquilidad que transmites desde el primer día. A **Eva Coll**, por venir cada día con una sonrisa, por ayudarme tanto en los últimos meses y trabajar siempre con tantas ganas. Te deseo de corazón lo mejor para los próximos años. A las que os fuisteis yendo, **Marta Llauradó, Nuria Pedrola, Marta García, Marina Rigau, Tamara**, por vuestra ayuda y vuestros sabios consejos de quien tiene reciente la escritura y defensa de una tesis doctoral; y a los que ahora empezáis, **Melissa, Manuel, Carlos**, **Leticia, Andrea, Ariadna** os deseo lo mejor en esta etapa llena de retos.

También quiero agradecer a toda la gente de otros grupos de investigación con la que he tenido el enorme placer de colaborar a lo largo de esta tesis: **Eva Baldrich, Erica, Álex Campos, Antonio y Mª Ángeles**. Ha sido un verdadero placer trabajar con todos vosotros, aprendiendo un sinfín de cosas nuevas y haciendo posible este proyecto.

Many thanks to Prof. **Bruno Domon** for giving me the precious opportunity to join your group and to have access to the most leading-edge technology and high-level knowledge in mass spectrometry. And I can´t be more thankful to **Cristina** and **Jan**, because you kindly offered me your help without asking. Cristina, the most cheerful and optimistic person in the lab, thank you so much for your thorough revision of the thesis; and Jan, many thanks for all your good advice and constructive suggestions and comments that have been of key importance for the success of this work.

These, not one nor two but five, internships in Luxembourg have been, without any doubt, the best experience of this thesis. It has been a real privilege to work in such a culturally and professionally rich environment, with people from more than 15 different countries. I want to

thank all my other colleagues: **Katriina, Stephane, Suruchi, Marie Aline, Adèle, Lizianne, Sara, François, Vinçent, Daniel, Victoria, Miriam, Sang Yoon, Yeoun Jin, Sebastien, Giselle, Nina, Cedric.** Thank you for making me feel at home, for being so nice and helpful with me since the very first minute. For all the funny moments that we have lived together inside and outside the lab. Special thanks go to "mis chicas de la física", **Lina** and **Daniela,** for all the affection and support you gave me since the beginning, for the best italian food I have ever tried and the best "macedonian" massage I have received while doing SPE.. :P For all the laughs, dinners and nice moments that we have enjoyed together. You are all very welcome to come to Barcelona whenever you want!

Pero no sólo quiero agradecer a los que han vivido conmigo el día a día de esta tesis. El apoyo de muchas otras personas ha sido igualmente indispensable para alcanzar esta meta.

Por ello quiero agradecer a mis **burgalesas** por amenizarme los días de escritura en Burgos. Al **pack biotec** y en especial a **Cris** y **Lauri**, por ser mi mejor compañía y mayor apoyo en Barcelona. A **mis sagradas**, mi consejo de sabios, porque cada éxito compartido con vosotras es doble éxito, porque nuestros skypes han llenado de luz y risas hasta los días más oscuros de esta tesis. Porque sois uno de los pilares más importantes de mi vida.

A toda la gente que he conocido durante estos años en Barcelona: mis chicas del máster, **Jessi, Esther, Patri**, porque con vosotras hice mis primeras cenas y fiestas en Barcelona, porque con vosotras descubrí lo mejor de esta ciudad... y de Tarragona! **Sonia, Carol**, por ser el mejor regalo de mi primer trabajo en Barcelona.

A mis compañeras de piso, **Alba** y **Georgi,** porque si hay algo importante es sentirte a gusto en casa, y con vosotras ha sido así desde el primer momento. Es una suerte poder volver a casa y desconectar en torno a una buena cena con nuestras sobremesas (aunque al día siguiente lamentes no haberte ido antes a la cama). Gracias por estar siempre ahí para compartir alegrías y penas, siempre dando apoyo y buenos consejos.

Y por supuesto agradecer a toda mi **familia, mis tíos, primos** y **abuelos**, la mejor familia que una puede tener. Los que queréis tener la revista donde se publica mi artículo aunque no sepáis una palabra de inglés, los que no dudáis en ayudarme con portadas y formatos de la tesis, y los que creéis en mí más que yo misma. Y en especial agradecer a **mis padres** y a **mi hermano**. Porque aun estando lejos os he sentido a mi lado cada día. Porque me habéis dado toda la fuerza y el optimismo cuando más lo necesitaba. Porque habéis creído siempre en mí y os alegráis casi más que yo de cada artículo, póster o cualquier objetivo conseguido. Por vuestra infinita paciencia frente a mis quejas y mis ensayos de presentaciones. Esas presentaciones que, aunque lo hiciera en inglés y no entendierais una palabra, os parecía un trabajo magnífico. Estoy segura de que sin vosotros

no hubiese llegado hoy aquí. Mil gracias de todo corazón!! Y a **Cruz, Carlita** y **Alonso**, porque aun sin saberlo, me han dado la energía y la felicidad que necesitaba mientras escribía esta tesis. Porque sus vídeos, fotos y sus "hola hola" me alegraron cada día.

Finally, I want to thank **Antoine**, the best gift of this thesis! Because you have helped me and supported me like nobody else. Because if I had not met you in 2013, this thesis would not have been the same. People say that working with your boyfriend is not a good idea, but we enjoy working and living together, being 24h a day together, and also do fine when we are separated by 1155 km. Thank you so much for your good advice, your master classes in mass spectrometry, your patience with me, and all your affection, amazing food and elaborate tapas while I was writing (my engine to write). I just hope to have you closer in this new stage of my life.

Muchísimas gracias a todos por haber hecho esto posible.

Thank you so much for making this possible.