SHAPING ENDOMETRIAL CANCER DIAGNOSIS VIA THE IDENTIFICATION OF PROTEIN BIOMARKERS IN GYNECOLOGICAL FLUIDS

Doctoral thesis presented by

Eva Coll de la Rubia

to obtain the degree of
Doctor for the Universitat Autònoma de Barcelona (UAB)

Doctoral thesis performed at the Vall Hebron Institute of Research, in the Group of Biomedical Research in Gynecology, under the supervision of
Dr. Antonio Gil Moreno, Dr. Eva Colás Ortega and Dr. Silvia Cabrera Díaz

Doctoral study in Biochemistry, Molecular Biology and Biomedicine, Department of Biochemistry and Molecular Biology, Faculty of Medicine at the UAB, under the supervision of
Dr. Joan Xavier Comella Carnicé.

Universitat Autònoma de Barcelona, 11th February 2022

Dr. Antonio Gil Moreno	Dr. Eva Colás Ortega	Dr. Silvia Cabrera Díaz
(director)	(director)	(director)

Dr. Joan Xavier Comella Carnicé	Eva Coll de la Rubia
(tutor)	(student)

Index

Index

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AB: Antibody

A **AUB**: Abnormal uterine bleeding

AUC: Area under the ROC curve

B **BMI**: Body-mass index

CAPG: Macrophage-capping protein

CE mark/certification: Conformité Européene mark/certification

CI: Confidence interval

CID: Collision induced dissociation

C **CLIA**: Chemiluminescence immunoassay

CPTAC: Clinical Proteomic Tumor Analysis Consortium

CT: Computed tomography

CV: Coefficient of variation

2DE: Two-dimensional electrophoresis

D&C: Dilatation and curettage

DDA: Data dependent acquisition

D **DIA:** Data independent acquisition

DOR: Diagnostic odds ratio

DTT: Dithiothreitol

EAH: Endometrial atypical hyperplasia

EC: Endometrial cancer

ECLIA: Electrochemiluminescence immunoassay

EEC: Endometrioid endometrial cancer

EIN: Endometrial intraepithelial neoplasia

ELISA: Enzyme-linked immunosorbent assay

E **EMA:** European Medicines Agency

ESGO: European Society of Gynecological Oncology

ESI: Electrospray ionization

ESMO: European Society for Medical Oncology

ESP: European Society of Pathology

ESTRO: European Society for Radiotherapy & Oncology

FDA: Food and Drug Administration

FDR: False discovery rate

FFPE: Formalin-fixed paraffin-embedded

F **FIGO:** Federation of Gynecology and Obstetrics

FN: False negative

FP: False positive

FSH: Follicle stimulating hormone

G   **GnRH:** Gonadotropin releasing hormone

H   **hCG:** Chorionic gonadotrophin
HPLC: High pressure liquid chromatography
HPV: Human papillomavirus
HRT: Hormone replacement therapy

I   **IAA:** Iodoacetamide
IFG-I: Insuline-like growth factor-I
iCAT: Isotope coded affinity tagging
IgG: Immunoglobulin G
IHQ: Immunohistochemistry
IPR: Intellectual property rights
ISGP: International Society of Gynecological Pathologists
iTRAQ: Isobaric tags for relative and absolute quantitation
IVD: In vitro diagnostics

K   **KPYM:** Pyruvate kinase

L   **L/H:** Endogenous/synthetic peptide ratio
LC: Liquid chromatography
LC-MS: Liquid chromatography coupled to mass spectrometry
LFQ: Label-free quantitative
LH: Luteinizing hormone
LR: Likelihood ratio
LVSI: Lymphovascular space invasion

M   **m/z:** Mass over charge ratio
MMR: mismatch repair proteins
MMRd: mismatch repair proteins deficient
MRI: Magnetic resonance imaging
MRM: Multiple reaction monitoring
MS: Mass spectrometry
MSI: Microsatellite instability
MSS: Microsatellite stable

N   **NEEC:** Non-endometrioid endometrial cancer
NGS: Next-generation sequencing
NLR: Negative likelihood ratio
NPV: Negative predictive value
NSMP: No specific molecular profile

O   **OS**: overall survival

P

p53mut: p53 mutated

PBS: Phosphate-buffered saline

PET: Positron emission tomography

PLR: Positive likelihood ratio

POLE: DNA polymerase epsilon catalytic subunit

POLEmut: DNA polymerase epsilon catalytic subunit mutated

PPV: Positive predictive value

PRM: Parallel reaction monitoring

ProMisE: The Proactive Molecular Risk Classifier for Endometrial Cancer

PSA: Prostate-specific antigen

Q

QMS: Quality management system

Q-OT: Quadrupole orbitrap

QqQ: Triple quadrupole

Q-TOF: Quadrupole time of flight

R

R coefficient: Pearson correlation coefficient

RFS: Recurrence-free survival

ROC curve: Receiver operating characteristic curve

RPPA: Reverse-phase protein array

S

SCNA: Somatic copy number alteration

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

SN: Sensitivity

SNV: Single nucleotide variant

SOPs: Standard operating procedures

SP: Specificity

SPE: Solid phase extraction

SRM: Selected reaction monitoring

SWATH: Sequential window acquisition of all theoretical fragment-ion spectra

T

TCGA: The Cancer Genome Atlas

TILs: Tumor-infiltrating lymphocytes

TMB: Tumor mutation burden

TN: True negative

TP: True positive

TMT: Tandem mass tags

TOF: Time-of-flight

TVUS: Transvaginal ultrasonography

U   **UICC:** Union for International Cancer Control
USA: United States of America

W   **WHO:** World Health Organization

X   **XIC:** Extracted ion chromatogram

Introduction

1 ENDOMETRIUM

1.1 ANATOMY AND FUNCTIONS OF THE UTERUS

The principal female sexual organs in humans are the ovaries, fallopian tubes, uterus, and vagina. Specifically, the **uterus** is situated in the center of the female pelvis between the urinary bladder and the rectum, and laterally adjacent to a broad of ligaments (Figure 1A). The uterus is a thick-walled, hollow, muscular organ with the size and shape of an inverted pear. The size and weight of the normal uterus depends on individual hormonal status and previous pregnancies. The uterus size of a nulliparous women is approximately 8 cm long, 5 cm wide, and 2.5 cm thick and weighs 40-50 grams. Anatomical subdivisions of the uterus include from the upper to the lowest part: the **fundus**, the dome-shaped superior area connected to the Fallopian tubes; the **body**, the largest central part; the **cervix**, separated from the body by a constructed region of about 1 cm long called the isthmus, is the element that opens to the vagina [1] (Figure 1B):
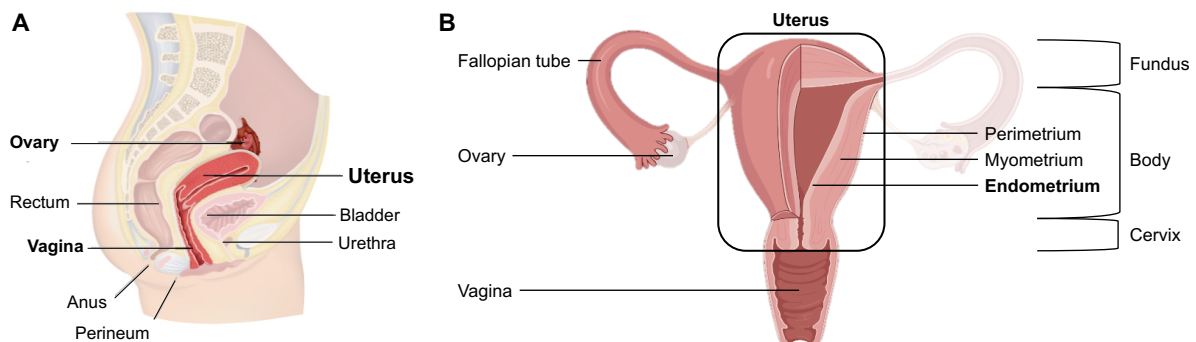
Figure 1. The uterus. (**A**) Location and (**B**) anatomical parts of the uterus.

The uterus has three layers [2] (Figure 1B):

- The **peritoneum**, the thin external serosal layer firmly attached to the uterus consisting of a visceral peritoneum.
- The **myometrium**, the thicker middle muscular layer composed of three indistinct layers of smooth muscle, is responsible for hormone response.
- The **endometrium** or mucosa layer, the inner layer of the uterus. It is a glandular and highly vascularized layer. It is characterized by being the responsive layer from the hormonal stimuli of the ovaries.

The uterus or womb is an essential organ for reproduction. It serves as part of the sperm pathway transporting them from the vagina to the Fallopian tubes through rhythmic peristaltic contractions

caused in response to the hormone oxytocin [3]. Its main function is to receive a fertilized ovum, nourish the developing fetus during pregnancy, and labor. Briefly, a fertilized ovum coming from a Fallopian tube will arrive in the uterine cavity and will divide to become a blastocyst, which will be implanted on the endometrium. As the embryo develops into a fetus, the placenta will be formed from the embryonic tissue thanks to changes in the endometrium. The placenta provides oxygen and nutrients to the fetus, while removing waste products until the birth. Also, the uterus is the source of menstrual flow during reproductive cycles when the implantation does not occur [2].

1.2 HISTOLOGY OF THE ENDOMETRIUM

The endometrium is a reddish mucous membrane varying from 1-6mm in thickness, depending on hormonal stimulation (Figure 2). It has three components: (i) the inner layer composed of simple columnar epithelium that lines the lumen formed by ciliated and secretory cells; (ii) a thick underlying endometrial stroma region of lamina propria formed by areolar connective tissue and highly supplied of blood vessels; (iii) endometrial glands developing as invaginations of the luminal epithelium and extending to the entire thickness of the stroma almost to the myometrium [4].

1.3 ENDOMETRIAL CYCLE

During the uterine cycle the endometrium is the only layer undergoing changes. The endometrium itself is divided into two layers [2]:

- The **basal layer or stratum basalis**. It is the deeper layer located adjacent to the myometrium, below the functional layer. It is a highly vascular and permanent layer during the menstruation. It gives rise to a new functional layer after each menstruation.
- The **functional layer or stratum functionalis**. It lines the uterine cavity and sloughs off during menstruation. It is shed at the time of menstruation and builds up again under the stimulation of ovarian steroid hormones.

The functional endometrium has a lifetime equal to the reproductive life of an adult, which goes from menarche to the menopause. In order to prepare itself for receiving a fertilized oocyte, during the human menstrual cycle the endometrium is exposed each month to sequential patterns of circulating ovarian sex steroids crucial to the regulation of growth and the differentiation of the endometrium. If fertilization does not occur, the functional layer of the endometrium is shed and expulsed, leading to menstruation, which occurs approximately every 28 days [5].

During the reproductive life of a woman, the functional endometrium goes though different stages: three of them are related to the endometrial cycle (proliferative, secretory, and menstrual) and one is related to pregnancy. All of them have different histological, molecular and time specific features (Figure 2). After the menopause, the non-cyclic endometrium becomes gradually atrophic [6].
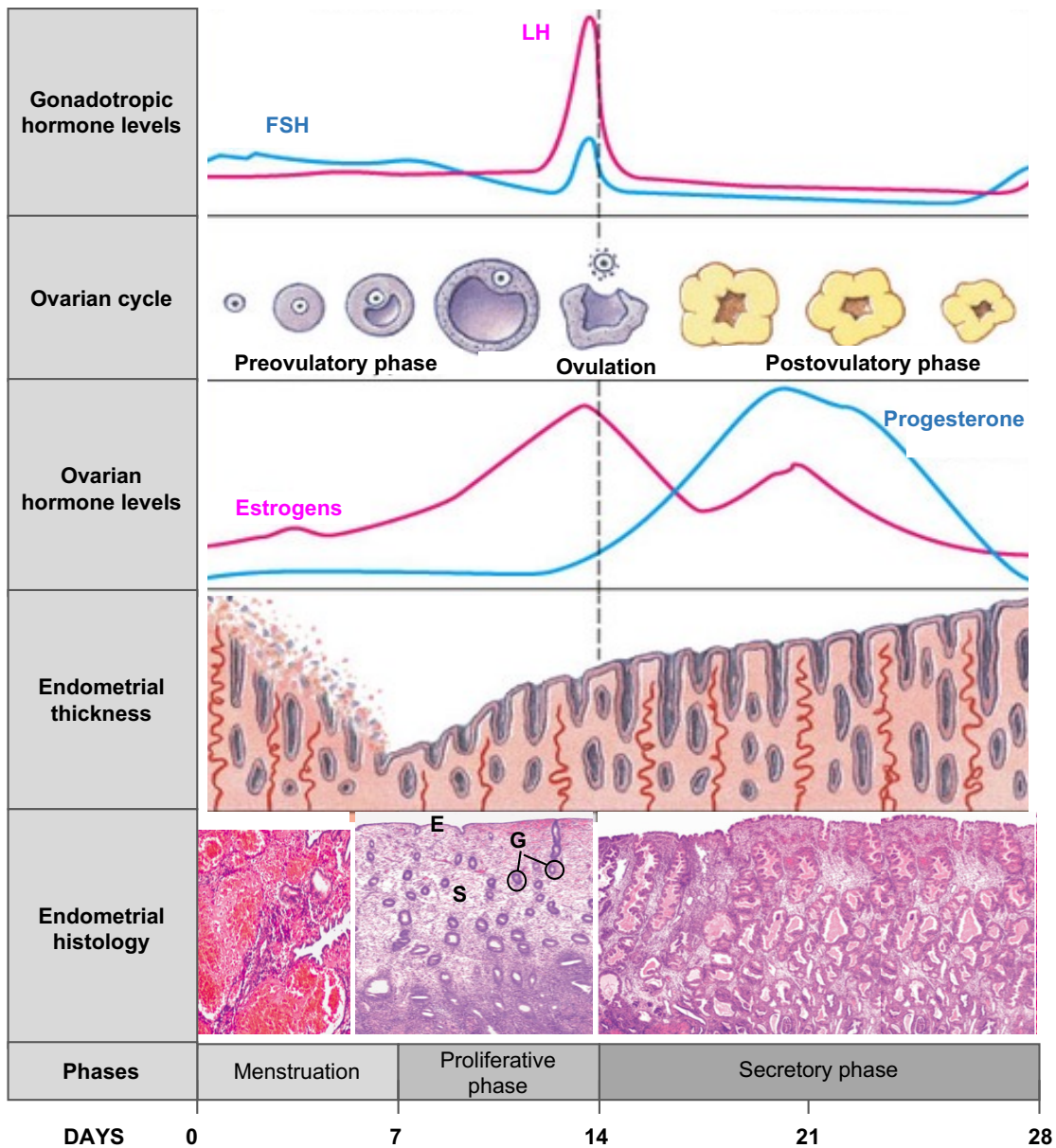
Figure 2. Endometrial cycle and endometrial histology. E, simple columnar epithelium; S, stroma; G, uterine glands.

1.3.1 Proliferative endometrium

The proliferative endometrium is present during the proliferative phase between the days 5-14 of the cycle, immediately after the menstruation and prior to the period of ovulation. At this state, the endometrium thickness is 1-2 mm. The proliferative endometrial tissue is characterized by its great capacity to proliferate and increasing its thickness due to 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 glands and stroma. The stroma is dense with cells showing a reduced cytoplasmic fraction.

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. Considering that the menstrual cycle takes 28 days, ovulation occurs on day 14 and it is associated with a high release of estrogen and an LH surge.

1.3.2 Secretory endometrium

Secretory endometrium is present during the secretory phase between days 15-28 of the menstrual cycle, just after the ovulation period. In this secretory endometrium, the "window of implantation" occurs from day 20 to 24 of the cycle, which corresponds to the period of time when endometrial receptivity for blastocyst implantation is at its highest. At this state the endometrium can increase its thickness up to 8 mm.

The tissue is characterized by increased luminal secretion of the endometrial glands and the predecidual transformation of stromal cells, which increase the cytoplasmic fraction and acquire a cubical morphology. The stroma, which was dense in the beginning, becomes adenomatous, and spiral arteries appear. All changes occurring during this phase are oriented to prepare a nutritive soil for the blastocyst implantation and are supported by the expression of progesterone.

1.3.3 Menstrual endometrium

The menstrual endometrium is present during the menstrual phase between days 1-4 of the menstrual cycle, and it is only formed when the oocyst is not fertilized.

In the absence of conception, chorionic gonadotrophin (hCG) is not produced. In the 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 the second phase, and as a consequence of the hormonal stimuli, the endometrium begins its regeneration from the basal layer which was not dissociated.

1.3.4 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. After this event, 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.

1.3.5 Atrophic endometrium

The atrophic endometrium presents at the onset of the menopause. Menopause is defined as the absence of menses for 12 months without a pathologic cause. Despite the influence of socioeconomic status and use of substances such as tobacco, the age in which menopause occurs ranges from **45-55**

years, being 51.4 the average age [7]. Nowadays, almost one third of a woman's lifetime in developed countries is postmenopausal, and it is during this period when over 80% of endometrial carcinomas develop [8].

The atrophic endometrium is thin (1 to 3 mm thick), with loss of distinction between the basal layer and the functional layer. In this phase, the endometrium suffers important histological modifications and loses its ability to proliferate and secrete. Specifically, 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.

The main reason for these modifications is the deprivation of estrogen and progesterone. Finally, 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.

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. Actually, the majority of non-cycling endometria are thin and atrophic, but only half of the cases are inactive [9]. 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** [10], or give rise to **atypical endometrial hyperplasia** or **endometrial intraepithelial carcinoma** [11], from which an endometrial carcinoma can develop.

2 ENDOMETRIAL LESIONS

2.1 ENDOMETRIAL HYPERPLASIA

Endometrial hyperplasia is a common disorder involving the abnormal proliferation of endometrial tissue, especially of the endometrial glands, that results in a greater than normal gland-to-stroma ratio [6]. It results from prolonged estrogen exposure unopposed by progesterone or pregestational agents acting on the entire endometrial field. Consequently, risk factors include perimenopause, obesity, polycystic ovarian syndrome, and diabetes. In fact, women exposed to unopposed estrogen have a 3-4-fold increased risk of endometrial carcinoma, rising to 10-fold after a decade.

Endometrial hyperplasia is a **precursor of endometrial carcinoma** and, when it is not treated, endometrial hyperplasia has the propensity to developed into the malignancy. The current system of **classification** for endometrial hyperplasia was proposed by the World Health Organization (WHO) and the International Society of Gynecological Pathologists (ISGP). It divides endometrial hyperplasia into two groups: with and without atypia [12].

2.1.1 Endometrial hyperplasia without atypia

Endometrial hyperplasia without atypia is a proliferation of endometrial glands of irregular size and shape without significant cytological atypia. **Clinically**, it is most commonly diagnosed in perimenopause (between 50-54 years), with symptoms of abnormal, non-cyclical vaginal bleeding. They will present a thickened endometrium on ultrasound. Additionally, the endometrium varies from having the uniform, tan appearance of the late proliferative phase to appearing highly thickened and sometimes polypoid or spongy with cysts. It will be diagnosed due to increased endometrial gland-to-stroma ratio; tubular, branching, and/or cystically dilated glands resembling proliferative endometrium; uniform distribution of nuclear features across submitted tissue (Figure 3A).

Progression to well-differentiated endometrial carcinoma occurs in 1-3% of women with hyperplasia without atypia. Hence, women diagnosed with hyperplasia without atypia are treated with cyclic progestins, showing regression of the disease in 3-6 months in 98% of cases [6].

Other terms also accepted for this lesion are benign endometrial hyperplasia, simple endometrial hyperplasia without atypia, and complex endometrial hyperplasia without atypia.

2.1.2 Atypical hyperplasia of the endometrium

Endometrial atypical hyperplasia / endometrioid intraepithelial neoplasia (EAH/EIN) is the simultaneous change of epithelial cytology and an increased number of endometrial glands in comparison with the

stroma (crowded gland architecture) within a morphologically defined region, distinct from the surrounding endometrium or from entrapped normal glands (Figure 3B).

EAH/EIN emerges as a clonal expansion of mutated glands that begins as a localized lesion and may expand to the entire endometrial compartment. EAH/EIN contains many of the genetic changes seen in endometrioid endometrial carcinoma, including microsatellite instability (including Lynch syndrome); *PAX2* inactivation; and *PTEN, KRAS,* and *CTNNB1* mutation.

The average patient age is 60-64 years. **Clinically**, postmenopausal or perimenopausal bleeding is the most common presenting symptom. Hyperestrogenism is a risk factor for this type of lesion. Besides, hereditary susceptibility parallels those heritable syndromes associated with increased risk for endometrioid endometrial carcinoma, including Cowden syndrome and Lynch syndrome.

Progression to well-differentiated endometrial carcinoma in women with a biopsy of EAH/EIN occurs in 25-33% at immediate hysterectomy or during the first year of follow-up. Longer-term risk elevation estimates vary 45-fold among EIN studies. The efficacy of conservative treatment of grade 1 endometrioid carcinoma or EAH/EIN with hormonal agents may be monitored by histology, but this is not yet standard clinical practice. Therefore, the recommended **treatment** for patients diagnosed with hyperplasia with atypia is the same as endometrial cancer, which is hysterectomy and bilateral salpingo-oophorectomy due to the underlying risk of malignancy or progression to endometrial cancer [13].

To diagnose this lesion, it is essential to morphologically define endometrial changes with crowded glandular architecture and altered epithelial cytology, distinct from the surrounding endometrium and/or entrapped non-neoplastic glands. However, many lesions have no distinguishing macroscopic features, others present as a focal thickening resembling a polyp, or within a diffusely thickened endometrium. The gross appearance is often obscured by hyperplasia without atypia, endometrial polyp, or carcinoma. Thus, the assessment of the loss of immunoreactivity for *PTEN, PAX2*, or mismatch repair proteins is desired to diagnose EAH/EIN.
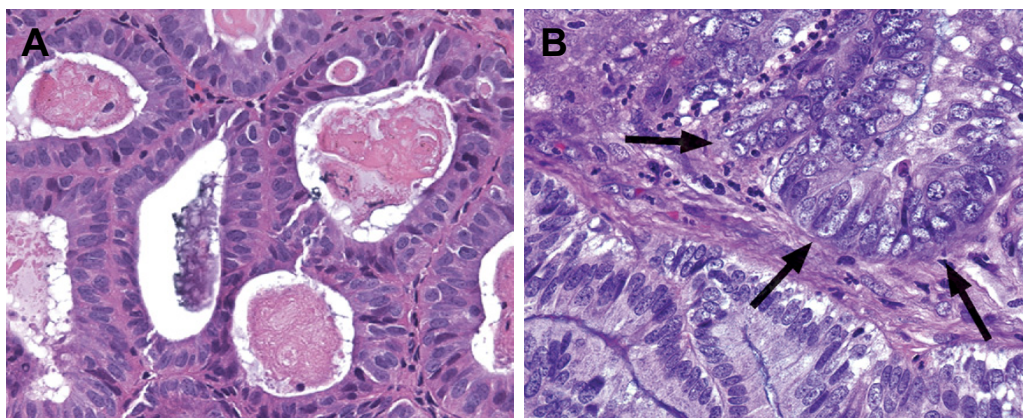
Figure 3. Hyperplasia with and without atypia. (**A**). Hyperplasia without atypia. Minimally branched, closely packed glands with a gland-to-stroma ratio of 3:1. (**B**). Atypical hyperplasia. Characterized by distinct nucleoli, nuclear pleomorphism, irregular nuclear membranes, dispersed chromatin. *Adapted from Mills AM et al. 2010* [14].

2.2 TUMOR-LIKE LESIONS

2.2.1 Endometrial polyp

Endometrial polyp is a polypoid, localized, and benign, with disorganized proliferation of endometrial glands and altered stroma, often with prominent blood vessels. It mostly arises in the fundus and the stromal component is clonal. In addition, there are rearrangements of one of the high mobility group proteins genes – *HMGA1* OR *HMGA2.*

Most affected women are perimenopausal or postmenopausal. Besides, polyps are a common endometrial lesion in patients who are on tamoxifen therapy. Clinical features related to this lesion include abnormal uterine bleeding and infertility. However, in some cases endometrial polyps may be asymptomatic. Although polyps are benign, concurrent/subsequent hyperplasia and carcinoma occur in 11-30% and 0.5-3%, respectively, of patients with endometrial polyps. Polyps in postmenopausal women with abnormal vaginal bleeding are more likely to be malignant. The frequency of carcinoma reaches 10.7% in tamoxifen-related cases, and in those cases, polyps are large and often multiple. Involvement by serous carcinoma within a polyp may be subtle and may require immunohistochemistry for confirmation. Recommended treatment for endometrial polyp varies depending on the size and regardless of the symptoms. Thus, only the polyps with >2 cm of diameter are removed, as they have higher risk to be non-diagnosed or hidden neoplasias [15]. To diagnose endometrial polyps branched and/or cystically dilated glands in altered stroma are observed. The appearance is a narrow or broad-based stalk, with a fibrotic cut surface.

2.2.2 Endometrial metaplasia

Endometrial metaplasia is a morphological alteration of the endometrial epithelium which consists of the replacement of normal epithelium at a given site by mature benign epithelium inappropriate to that site. It occurs as a consequence of altered hormonal levels, repair, endometrial breakdown, chronic inflammation, or polyp infarction. Clinically, most women with endometrial metaplasia are postmenopausal. To diagnose endometrial metaplasias, a non-neoplastic endometrial lesion or a non-neoplastic component in a neoplastic endometrial lesion will be observed. Also, identification of individual features of each histological pattern or endometrial metaplasia may be seen.

3   ENDOMETRIAL CANCER

3.1 EPIDEMIOLOGY

The cancer-specific incidence and mortality rate differ between the considered developed countries (Europe, North America, and high-income countries in Asia and Oceania), and non-developed countries (remaining regions and countries)[16]. Endometrial cancer (EC) is the most common gynecological malignancy of the female genital tract, and it represents the fourth most common cancer in developed countries, after breast, lung, and colorectal cancer. Recent data from the United States of America (USA) estimates that 65,950 new cases of EC will be diagnosed in 2022, which represents 7% of all cancers in women [17] (Figure 4A).

In Europe, EC represented 11.1% of the total number of cancer cases in women in developed countries in 2018 [16]. In Spain, EC is expected to be the fourth most frequently diagnosed cancer in women (6,923 new cases in 2021), after breast, colon, and lung cancer.
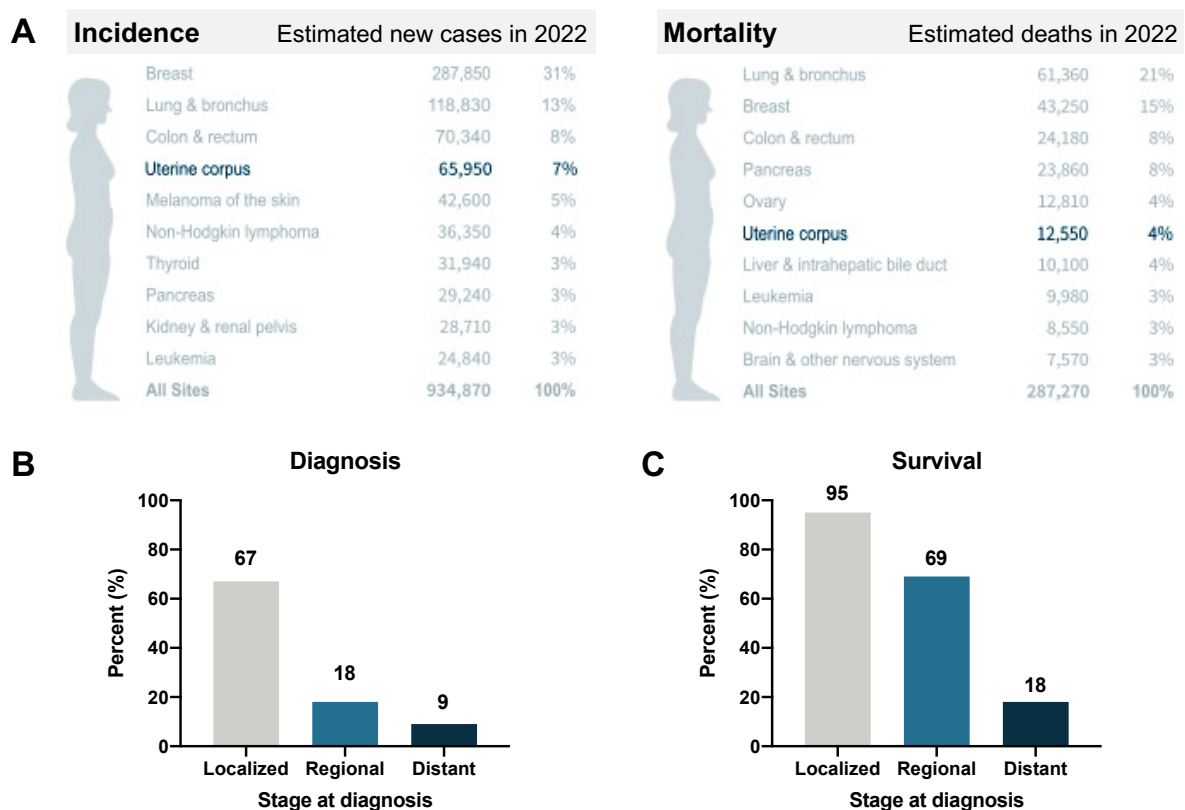
Figure 4. Epidemiology of EC. (A) Estimated new cases and deaths for the ten leading cancer types in women in the United States for 2022. **(B)** Stage distribution at diagnosis of endometrial cancer cases. **(C)** Five-year survival rates of endometrial cancer by stage at diagnosis. *Adapted from Siegel et al. 2022.*

Unfortunately, its incidence is increasing because of the increased life expectancy of the population as well as a greater overall prevalence of obesity and metabolic syndromes. In fact, it is expected that 608,000 EC cases will be newly diagnosed in 2040 worldwide compared to the 417,000 new EC cases in 2020, meaning a 46% increase in cases [18]. In particular, Europe and the USA will account for 12.5% of this increase whereby 198,453 diagnosed cases in 2020 will rise to 223,161 newly diagnosed cases in 2040. EC mostly affects postmenopausal and perimenopausal women, with a median age of 63 years. However, only 14-20% of the cases are diagnosed in premenopausal women, 6.4% of whom are younger than 45 years [8,19]. Despite its high incidence, EC has a good prognosis in the majority of cases. About 70% of EC patients are diagnosed at early stages of the disease, i.e., when the tumor is still localized within the endometrium, which is associated with an overall 5-year survival rate of 95%. However, the other 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 18% when the tumor has spread beyond the pelvis [20] (Figure 4B-C). Remarkably, while cancer survival has overall improved over the last 40 years, EC together with cervical cancer did not show any improvement related to cancer deaths [21]. This event has been associated with the rapid progress in researching new diagnostic and treatment strategies in other types of cancer in comparison to EC [22]. In 2022, 12,550 deaths are expected in the USA due to EC [20] (Figure 4A).

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.

Risk factors	Relative Risk
Estrogen Exposure	
Unopposed estrogen therapy	10-20
Tamoxifen	2-3
Obesity	2-5
Overweight	1.4-1.6
Early menarche	1.5-2
Late menopause	2-3
Nulliparity/infertility	3
Polycystic Ovary Syndrome	1.5
Others	
Age > 50	2-3
Hypertension	1.3-3
Diabetes mellitus	1.3-3
Thyroid disease	1-3
Family history: Lynch syndome	5-20
Protective factors	**Relative Risk**
Estrogen Exposure	
Combined oral contraceptives	0.76
Grand multiparity	0.57
Physical activity	0.79
Others	
Cigarette smoking	0.71

Table 1. Risk and protective factors associated with the development of EC. *Adapted from Oncoguía SEGO 2016* [23].

3.2.1 Risk factors

Patient risk factors for endometrial cancer consist of a broad number of parameters, which can be classified as estrogen related factors and other factors.

3.2.1.1 Hormone factors: estrogen exposure

Women's hormone balance plays an important role in the development of most EC. Many of the risk factors for EC affect estrogen levels either endogenously or exogenously. This exposure of estrogen increases the mitotic activity of the endometrial cells, causing DNA replication errors, mutations, etc. and is highly related to endometrial hyperplasia and endometrioid EC (see full description of EC subtypes in section 3.4.1) [6].

Exogen factors having an impact on the unopposed estrogen exposure are estrogen replacement therapy and tamoxifen. **Estrogen replacement therapy** is commonly prescribed to combat symptomatology related to menopause. It alleviates the symptoms, but stimulates the endometrium, increasing the risk of EC from 1.45 to 4.46 times [24]. Currently, estrogen replacement therapy is still the most effective treatment for vasomotor symptoms and the genitourinary syndrome of menopause. However, to reduce the risk effects of the therapy, progestins are nowadays added to the treatment [25].

Tamoxifen is a selective estrogen receptor modulator indicated for breast cancer prevention and treatment of breast cancer. It acts as an antagonist of the estrogen receptor in breast tissue. However, in the endometrium, it acts as an agonist inducing growth of endometrial cells. Tamoxifen has been associated with a significantly increased risk of EC from 2 to 8-fold. The risk increases with longer administration of the drug, but it is only evident after stopping the treatment [26].

Endogenously, exposure of estrogens can be increased due to being **overweight and obesity**. Excess body weight, expressed as increased body-mass index (BMI), is associated with the risk of several adult cancers, including EC [27]. In comparison to non-overweight women, women with a high BMI have 1.6-5 times more chance to suffer EC [28]. This condition is linked to hormone changes. Fat tissue will change androgens (hormones) into estrogens, impacting on the levels of the latter, especially after menopause. Obese women tend to have more fat tissue, which can increase women's estrogen levels and consequently, EC risk. Related to this, **menstrual and reproductive factors**, can also have an impact. Thus, **early menarche and late menopause** have been associated with a higher risk of EC. Before menopause, the ovaries are the major source of estrogen and progesterone, the main types of female hormones, and they are produced in a controlled manner according to the women's menstrual cycle. A shift in this balance towards an increase of estrogen is one of the main risk factors for EC. Additionally, after menopause, the ovaries stop synthetizing these two hormones. At this point, after menopause, the major source of estrogen is produced naturally in fat tissue [9]. Likewise, the anovulation and dysregulation of estradiol and progesterone levels caused by **infertility and nulliparity** [29,30], as well as

the most common endocrine disorder in females, **polycystic ovary syndrome** [31], have been associated with an increased risk of EC.

3.2.1.2 Other factors

Other factors, more loosely associated with EC and the excess of estrogen exposure, include **diabetes mellitus** [32], **thyroid disease** [33], and **hypertension** [34]. However, all these factors are related to an excess of estrogen exposure through their frequent association with obesity, though the importance of increased levels of insulin-like growth factor-I (IFG-I) has also been indicated [35].

The main risk factor for EC is **age**, and the average age at diagnosis is approximately 63 years. EC is a cancer of the postmenopausal and perimenopausal period, and only 14-20% of patients with endometrial cancer are premenopausal, as stated previously [8]. Thus, 90% of women suffering from EC are older than 40 years [19]. Additionally, younger women with EC generally have well differentiated endometrioid tumors and lower-stage disease than older women. Therefore, with the current and estimated increase in life expectancy, the incidence of EC will become greater.

Family history has also been described with regards to the risk of EC development. A hereditary component represents around 5% of all reported cases. **Lynch syndrome**, also known as hereditary nonpolyposis colorectal cancer syndrome, represents 3% of all EC and 9% of EC cases in women under the age of 50 years [36,37]. It is an autosomal dominant family cancer syndrome characterized by mutations in DNA mismatch repair genes (*MLH1, MSH2, MSH6* or *PMS2*). The lifetime cumulative risk of EC for women with Lynch syndrome is 40% to 60% with a median age at onset of 48 years [38]. Earlier than the average for EC patients.

3.2.2 Protective factors

On the contrary, the use of combined oral contraceptives [39], multiparity [40], cigarette smoking [41], and physical activity [42] have been described as protective factors that reduce the risk of EC. All these factors are associated with a balance role of the two main hormones in charge of menstrual cycle: estrogen and progesterone. Thus, they act by decreasing estrogen concentrations and/or increasing progesterone levels.

3.3 EC CLASSIFICATION

EC arises from the lining of the uterus. It is a heterogeneous disease composed of biologically and histologically diverse neoplasms with different pathogeneses. In order to include this heterogeneity on current classification systems, clinical, pathological and molecular features are considered. Traditionally, EC has been broadly classified into two different subtypes proposed by Bokhman *et al.,* in 1983 [43]. Recently, The Cancer Genome Atlas (TCGA) identified and described a new molecular

classification based on four molecular subclasses, that has gained relevance due to its diagnostic information, prognostic utility and potential to predict benefit from adjuvant treatment [44].

3.3.1 Pathological features: Dualistic classification

In the past decades, EC has been broadly classified into two subtypes based on clinical data and morphological features (Table 2). This model was proposed by *Bokhman et al.* in 1983 as the result of 20 years of investigations and was published in Gynecological Oncology [43]. His research consisted of a prospective study including 366 patients with EC, where he compiled clinical and morphological data based on observations, and the results from 14 years of EC patient follow-up after treatment. His observations allowed him to classify EC in two subtypes used to manage these patients during 30 years until the molecular classification arose in 2013 [44].

	Bokhman: dualistic classification		
	Type I or Endometrioid Endometrial Carcinoma (EEC)	**Type II or Non-Endometrioid Endometrial Carcinoma (NEEC)**	
Risk factors			
Incidence	> 80% of all cases, perimenopausal women	10-15% of all cases, older women	
Clinical features	Metabolic syndrome	None	
Precedent	Hyperplasia, EIN	Atrophy, SEIN	
Grade	Low (63% G1-G2, 37% G3)	High (100% G3)	
Stage	Early (84% I-II, 16% III-IV)	Late (45% I-II, 55% III-IV)	
Histology	**Endometrioid**	**Non-endometrioid: Serous**	**Non-endometrioid: Clear cell**
Molecular characteristics			
Genomic stability	Diploid (40% microsatelite instability)	Aneuploid	
Hornone receptor expression	Positive	Negative	
TP53 mutation	Rare	>90%	35%
PI3K alterations	PTEN mutation (75-85%)	PTEN mutation (11%)	-
	-	-	PTEN loss (80%)
	PIK3CA mutation (50-60%)	PIK3CA mutation (35%)	PIK3CA mutation (18%)
	-	PIK3CA amplification (45%)	-
KRAS mutation	20-30%	3%	0%
ERBB alterations	None	ERBB amplification (21-47%)	ERBB amplification (16%)
	-	-	ERBB mutation (12%)
FGFR amplification or mutations	FGFR mutation (12%)	FGFR mutation (5%)	-
		Frequent FGFR1 and FGFR3 amplification	-
Wnt/b-catenin	CTNNB1 mutation (25%)	CTNNB1 mutation (3%)	-
Other	ARID1A mutation (35-40%)	-	ARID1A mutation (20-40%)
	-	PPP2R1A mutation (20%)	-
	-	FBXW7 mutation (20% undifferentiated EC)	-
Prognosis			
5-year Overall Survival	Good (85%)	Poor (55%)	

Table 2. EC dualistic classification. *Adapted from Morice et al. 2016* [45].

3.3.1.1 Type I or Endometrioid adenocarcinomas

Endometrioid Endometrial Cancer (EEC) represents the most common subtype of EC. It constitutes more than 80% of newly diagnosed cases. The development of EEC is strongly associated with epidemiological risk factors leading to an excess of estrogen relative to progesterone. Thus, it is an estrogen-dependent, diploid adenocarcinoma with endometrioid histology [45]. EECs are usually developed in perimenopausal women and can be associated with or preceded by endometrial hyperplasia [46].

In histopathology, EECs are described as varying proportions of glandular, papillary, and solid architecture, with the neoplastic cells showing endometrioid differentiation [12]. They are diagnosed in 63% and 84% of the cases at low grade and early stage, respectively. EEC have a good 5-year overall survival rate in 85% of cases [45].

Regarding molecular alterations, type I ECs are characterized by a high mutational frequency and microsatellite instability [47]. *PTEN* mutation, in the PI3K pathway, is the most frequent somatic mutation among EECs, occurring in more than 90% of the cases. Multiple aberrations of the pathways can be noted, such as simultaneous loss of PTEN and PIK3CA. In fact, in mouse models it has been demonstrated that biallelic *Pten* loss leads to the development of Complex Atypical Hyperplasia (CAH), whereas biallelic *Pten* loss together with mutational activation of *Pik3ca* results in progression of CAH to EC [48]. Other common mutations are found in the beta-catenin gene in 20-25% of cases, and are almost restricted to EEC. *KRAS, FGFR* and *ARID1A* mutations have also been reported in about 25%, 12%, and 35% of cases, respectively [49].

3.3.1.2 Type II or Non-Endometrioid adenocarcinomas

Non-Endometrioid Endometrial Cancers (NEEC) represent 10-15% of all EC cases. Their development occurs mainly on atrophic endometrium. These tumors are high-grade, aneuploid adenocarcinomas, not related to estrogen stimulation. In fact, they express neither estrogen nor progesterone receptor. They include different minor histologies such as serous (~10%), clear cell (~3%), which are the most common types but also mixed cell adenocarcinoma or carcinosarcomas, and other rare types [43]. They usually develop in women older than those with type I EC.

In histopathology, serous carcinomas are defined with diffuse, market nuclear plomorphism, typically exhibiting papillary and/or glandular growth pattens, while clear cell tumors are defined as carcinomas demonstrating papillary, tubulocystic, and/or solid architectural patterns and variably pleomorphic polygonal, cubcoidal, flat or hobnail cells with clear or eosinophilic cytoplasm [12]. They are grade 3 carcinomas diagnosed at late stages in 55% of cases. NEEC are associated with a poor 5-year overall survival rate in 55% of cases, and they account for more than 50% of recurrences and deaths from EC [50].

NEECs are characterized by chromosomal instability and TP53 mutation [51], but each specific histological subtype shows specific molecular features. Regarding serous carcinomas, mutations in the TP53 gene occur in up to 90% of cases. Additionally, about one third of NEEC were reported to have ERBB amplificated [52]. Also, the PI3K pathway is frequently dysregulated. However, these tumors are mainly characterized by amplifications or mutations in the PIK3CA gene in 45% and 35% of the cases, respectively. Concerning the most common molecular features in the tumors with clear cell histology, they are more prompt to loss PTEN (80% of cases), and ARID1A gene was reported to be mutated in 20-40% of the tumors [53,54].

Although this dualistic classification is broadly used in the current clinical practice for pre-operative assessment and surgical planning [45], its **prognostic value remains limited**. Using this classification model, approximately 20% of EEC cases relapse, whereas 50% of NEEC do not [43]. Besides, the binary classification, 15-20% of EEC tumors are classified as a high-grade lesions that do not fit in this model, and where they fit remains unclear [55]. EC comprises a high variety of biological, pathological and molecular features impossible to simplify in a dualistic model [56].

3.3.2 Molecular classification (TCGA)

The TCGA network performed a multiplatform analysis of EC providing a new framework from which to approach EC classification [44]. The TCGA performed an integrated genomic, transcriptomic and proteomic characterization (i.*e.,* next-generation sequencing of the whole exome, methylation profiles, microRNA profiles, gene expression analysis, and reverse-phase protein lysate arrays) of 373 tumors, including 307 EEC, 66 SEC, and 13 mixed EC. Although the TCGA is continuously adding new cases to this huge molecular study, the 373 tumors analyzed were the basis to develop a more accurate classification of EC. The consortium exploited specifically single nucleotide variant (SNV), somatic copy number alterations (SCNAs) and tumor mutation burden (TMB) to break with the traditional dualistic view and proposed four prognostically significant subgroups: POLE ultramutated, microsatellite instability (MSI) hypermutated, copy-number-low (microsatellite stable, MSS), and copy-number-high (serous-like) EC (Table 3) [44].

	POLE-ultramutated	MSI-hypermutated	Copy-number low, MSS (endometrioid)	Copy-number high, high SCNA (serous-like)
CLINICAL FEATURES				
Prevalence	5-15%	25-30%	30-40%	5-15%
Associated features	Younger age at presentation; Lower BMI	Might be associated with LS; Higher BMI	Higher BMI	Advanced stage at presentation; Lower BMI
RISK FACTORS				
Histology (24)	Mainly endometrioid			Serous and endometrioid
Grade	G3 > G1, G2		G1 , G2 > G3	G3
Stage	I , II , III , IV			
Histological features	Ambiguous morphology, broad front invasion, tumor giant cells, prominent TILs, peri-tumoral lymphocytes	Mucinous differentiation, LVSI substantial, MELF-type invasion, Crohn's-like peri-tumoral reaction, low uterine segment involvement	Squamous differentiation, notable absence of TILS, ER/PR diffuse positive	LVSI, destructive invasion, Diffuse cytonuclear atypia, tumor giant cells, hobnailing, slit-like spaces
MOLECULAR CHARACTERISTICS				
Mutation load				
SCNAs load				
PI3K alterations				
KRAS mutation				
TP53 mutation	35%	5%	1%	>90%
Specific alterations	Hotspots mutation in POLE	Loss of DNA MMR proteins (MLH1, MSH2,...)	CTNNB1 (52%)	TP53; ~25% ERBB2 amplified
Top5 recurrent gene mutations	*POLE* (100%), *DMD* (100%), *CSMD1* (100%), *FAT4* (100%), *PTEN* (94%)	*PTEN* (88%), *PIK3CA* (54%), *PIK3R1* (42%), *RPL22* (37%), *ARID1A* (37%)	*PTEN* (77%), *PIK3CA* (53%), *CTNNB1* (52%), *ARID1A* (42%), *PIK3R1* (33%)	*TP53* (92%), *PIK3CA* (47%), *FBXW7* (22%), *PPP2R1A* (22%), *PTEN* (10%)
DIAGNOSIS				
DNA – Sequencing (25)	*P286R, V411L, S297F, A456P, S459, F367S, L424I, M295R, P436R, M444K, D368Y*	POLE-wt or Non-pathogenic POLE		
Protein - IHC		**MMR-loss:** MLH1, MSH2, MSH6, PMS2	MMR-proficient; TP53-wt	TP53-mut TP53-abnormal
PROGNOSIS				
Progression free survival	Excellent	Intermediate	Intermediate	Poor
TREATMENT				
Treatments to explore (26)	Immune checkpoint inhibitors		PI3K/AKT/mTOR-i; hormonal therapy	Cell cycle regulators; PI3K/AKT/mTOR-I; PARPi

Table 3. TCGA molecular classification for endometrioid and serous histologies. Clinical features, risk factors, molecular characteristics, diagnosis (clinical diagnosis: ProMiSe classification), prognosis, and treatment associated to each subgroup of the TCGA classification system. SCNA, somatic copy number alterations load; IHC, immunohistochemistry; MMR, mismatch repair proteins; mut, mutated; wt, wild-type; -i, inhibitors. *Adapted from Coll-de la Rubia E et al. 2020* [57].

3.3.2.1 POLE ultramutated

The newly identified small subgroup is constituted by **5-15% of all EC cases**, generally in younger women and with lower BMI compared to other subtypes. They are mainly EEC, predominantly from grade 3 (60%), but also lower grades.

They are characterized by **pathogenic somatic mutations in the exonuclease domain of the replicative DNA polymerase epsilon, so called, *POLE***. The five most common *POLE* mutations are P286R, V411L, S297F, A456P, and S459F [58–60], however other mutations have been described [61]. Substitutions in DNA polymerases inactivates or suppresses proofreading abilities, causing increased replicative error rates resulting in the ultramutated phenotype characterizing this subtype. Thus, this defined subgroup of ECs presents ultramutations (frequently ≥100 mutations/Mb), *TP53* mutation in 35% of cases, enhanced immune response, and **excellent clinical outcome** [44,62,63].

Consequently, despite the high-grade histologic factors, these patients are considered as low-risk patients when they have stage I-II and minimal treatment will be performed. However, their classification is not clear when patients are diagnosed with stage III-IV [64]. In any case, due to its characteristics, immune checkpoint inhibitors have been suggested as treatment for this subgroup [65].

3.3.2.2 Microsatellite instability hypermutated

The MSI subgroup accounts for **25-30% of EC** and it is associated with women with higher BMI. MSI-hypermutated tumors are mainly endometrioid, but other non-endometrioid histologies have been described such as clear cell and mixed endometrioid / clear cell carcinomas [53,66,67]. Additionally, similarly to POLE ultramutated subgroup, grade 3 histology is predominant to grades 1 and 2, and display tumor-infiltrating lymphocytes (TILs), and tumor heterogeneity [44].

These tumors are characterized by a **microsatellite unstable phenotype that arises from defects in postreplicative DNA mismatch repair system**. Thus, the loss of mismatch repair (MMR) proteins MLH1, MSH2, MSH6, and/or PMS2. Most of these tumors are sporadic, arising as a result of epigenetic silencing of MLH1 (by promoter hypermethylation). However, in a minority of ECs, it is caused by somatic or germline mutations in the MMR genes, in this case being diagnosed with Lynch Syndrome [68]. This MSI subgroup has a mutation frequency of 100-10 mutations/Mb, and shows multiple mutations affecting the PI3K/AKT/mTOR pathway [69]. Specifically, these tumors showed low PTEN and high phoshpo-AKT levels, and accumulation of mutations within the pathway including *PIK3CA, PIK3R1,* and *KRAS.*

MSI hypermutated subgroup is considered of **intermediate prognosis**. Due to the high mutation load and rich immune environment presented by this tumors, Immune checkpoint inhibitors such as anti-PD-L1 therapy seem to be the most adequate therapy, and different drugs have been approved by the European Medicines Agency (EMA) [70–72].

3.3.2.3 Copy-number low / Microsatellite stable, MSS

Copy-number low subgroup represents the biggest subgroup accounting for **30-40% of all EC cases**. This subgroup is associated with high BMI. They are constituted mainly by endometrioid tumors, predominantly with a low-grade (87%) [44].

These group of tumors are characterized by **low somatic copy number alterations**, with frequent mutations in *PTEN, PIK3CA, ARID1A,* and mutations in genes related to Wingless-related integration site (Wnt) signaling, such as mutations in *CTNNB1* in 52% of cases [44]. They present notable absence of TILs and ER/PR are presented as diffuse positive.

Copy-number low tumors have **intermediate prognosis**. According to the biology of these tumors, inhibitors of the PI3K/AKT/mTOR or hormonal therapies have been suggested for treating these tumors.

3.3.2.4 Copy-number high (serious-like)

Copy-number high is a group comprised by serous endometrial tumors and approximately 25% of high-grade endometrioid tumors. About **5-15% of all EC cases** will be classified into this subgroup associated with advanced stage at presentation and lower BMI [44].

This subgroup is characterized by presentation of **extensive SCNAs and TP53 mutations**. They also frequently present amplifications in *MYC, ERBB2,* and *CCNE1* oncogenes.

These tumors are considered to have the **poorest prognosis** within the molecular classification, with significantly worse progression-free survival. The natural biology of these tumors may suggest that the most suitable treatments for these tumors might be the ones targeting cell cycle regulators, PI3K/AKT/mTOR inhibitors, or PARP inhibitors.

3.3.2.5 Multiple classifier

Recently, many authors identified a new small molecular subgroup of EC tumors. These are characterized by harboring more than one molecular feature and were named as 'multiple-classifier' [73]. Specifically, *Castillo et al., 2020* performed a study intended to understand the behavior of tumors compiling abnormal p53 (p53mut) in addition to another classifier(s). The analyzed cohort of 3,518 EC showed 3% of multiple-classified tumors (n=107) containing p53mut. From those, 64 (1.8%) were MMRd-p53mut EC, 31 (0.9%) POLEmut-p53mut EC, and 12 (0.3%) had the triple classifier (MMRd-POLEmut-p53mut EC). The investigations used IHC and hierarchical clustering by SNV type and SCNAs, and suggested that *TP53* mutation was a secondary event and, therefore, tumors bearing p53mut mostly clustered with the other prevalent molecular alteration, either MMRd tumors and/or POLEmut tumors. In line with this, the outcome of stage I patients measured by the 5-year recurrence-free survival (RFS) was 92.2% and 94.1% for MMRd-p53mut and POLEmut-p53mut, respectively, whereas single-classifier p53mut EC had an RFS of 70.8%. Thus, supporting the idea of *TP53* mutation as a secondary event. However, it is still unclear how these patients should be managed [74].

Clinically, the characterization of the molecular subgroups by the TCGA has led to a paradigm shift from the traditional dualistic model of Type I and Type 2 EC described by Bokhman *et al.,* to a classification where the understanding of the heterogeneous nature of the disease is much reflected [43,44].

The 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 15% are ultramutated POLE cancers with good prognosis. Therefore, it has been recommended that clinicians incorporate these molecular features to improve management of EC patients (see section 3.6) [45,64]. It demonstrated to be a potential tool to improve the risk stratification system as it provides additional prognostic information. In fact, studies performed in large cohorts of patients, validated its prognostic relevance and pointed out specific subsets of patients who would benefit from this classification system [44,58,59,75–77] (Figure 5). Particularly, it has been reported that 10% of patients diagnosed with a good prognosis cancer (EEC-Grade 1 or EEC-Grade 2) but with a copy-number high (serous-like) molecular diagnosis, will now be stratified in a poorer prognosis group. In contrast, 17% of POLE ultramutated tumors, which currently have a good prognosis, were previously classified as high-risk tumors (EEC-Grade 3) [44].
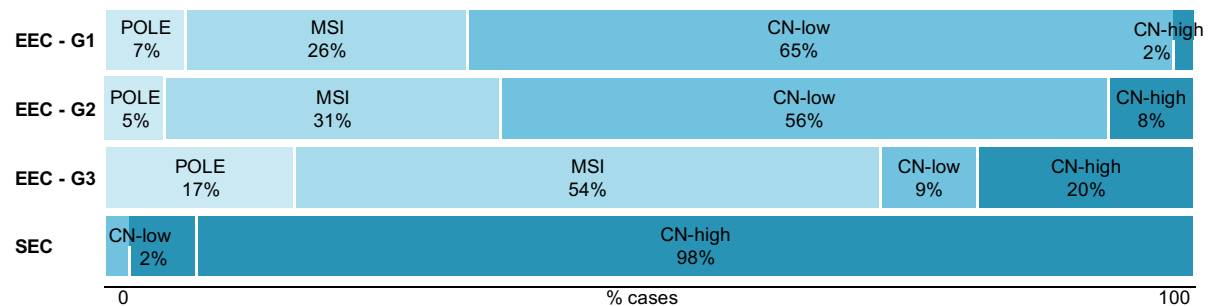
Figure 5. Assessment of TCGA novel classification. Itemization of the TCGA subgroups in the dualistic model [44]. G, grade. Adapted from *Coll-de la Rubia E et al. 2020* [57].

The characterization of EC subtypes has been achieved thank to the integration of genetic, transcriptomic and proteomic high-throughput techniques. However, these molecular techniques remain fairly inaccessible, and the WHO classification continuous relying on morphological features with the aid of immunohistochemistry (IHC). Hence, in order to molecularly classify EC, a surrogate of the molecular classification named The Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) algorithm was developed by Talhouk *et al.* [59] and was validated in several studies [58,75,77–80], allowing the molecular classification of patients as MMR proteins deficient (MMRd), POLE mutated (POLEmut), non-specific molecular profile (NSMP), and p53 mutated (p53mut or p53abn), using clinically applicable techniques. The ProMisE algorithm consists of a decision-tree based on the assessment of the three surrogate biomarker tests: (i) IHC of mismatch repair (MMR) proteins to assess their loss of expression, and therefore, whether they are deficient or proficient; (ii) *POLE* sequencing;

and (iii) IHC of the p53 protein. Regarding the assessment of p53 IHC could be positive, existing three different patters: mutated, in cases with intense and diffuse nuclear positivity; null, in cases with no nuclear staining; cytoplasmatic, in tumoral cells (exceptional). Alternatively, p53 IHC can be negative when the nuclear staining is weak and/or focal [73]. These surrogate biomarkers categorize most ECs according to their molecular classification, but the method is limited by: (i) the need of a systematic evaluation of the pathogenicity of POLE mutations [61]; (ii) the suboptimal correlation of p53 immunohistochemistry with TP53 copy-number alterations, and its use in these algorithms may misclassify some copy-number high tumors [81]; and (iii) the lack of evaluation in the significant heterogeneity seen in the copy-number-low group [82].

The application of this algorithm represents an improvement particularly relevant in EEC-G3, as it is the group where molecular classification has its greatest prognostic discriminative value [78]. This group includes tumors morphologically that is difficult to distinguish from SEC and EEC-G3. Hence, it is highly recommended to always classify these patients according to the molecular classification (*i.e., POLE* sequencing and other IHQ markers). Although in other endometrial carcinomas the IHQ study is always highly recommended (MMR proteins and p53), in high-grade endometrial carcinomas the *POLE* sequencing should be considered, while in low-grade EEC the *POLE* sequencing can be omitted. Additionally, the assessment of MMR proteins will not only allow molecular classification of patients and thus facilitate the diagnostic and identification of therapeutic options in advanced or recurrent MMRd cases, but also enables Lynch syndrome screening.

Recent research has led to the rise and validation of the molecular classification, which has also been assessed through clinically applicable techniques. In addition, the inter-observer variability in the diagnosis of endometrial tumors is well-known (see section 3.6.1). Thus, in accordance with this, up-to-date global guidelines introduced and integrated molecular classification into the clinical practice [12,64,83]. European recommendations suggest the use of the surrogate diagnostic algorithm provided by Vermij *et al.* (Figure 6), using IHC to assess p53, and MMR proteins, and somatic mutation analysis of *POLE* pathogenic mutations (exons 9, 11, 13, 14) described by Leon-Castillo et al. [61] , to classify patients in 5 different categories [64]. Moreover, updated 2020 WHO classification of Female Genital Tract tumor updates the definition of histological types and it also recommends not to use the dualistic classification as it is not fully adapting the reality. Hence, the 2020 WHO classification provides morphological, immunohistochemical and molecular diagnostic features to avoid discrepancies and develop a better understanding of the molecular bases of each tumor [12]. Furthermore, evolution within the WHO classification is expected in future editions towards personalized medicine and targeted molecular therapies in the treatment of EC in order to improve patient outcomes [84].
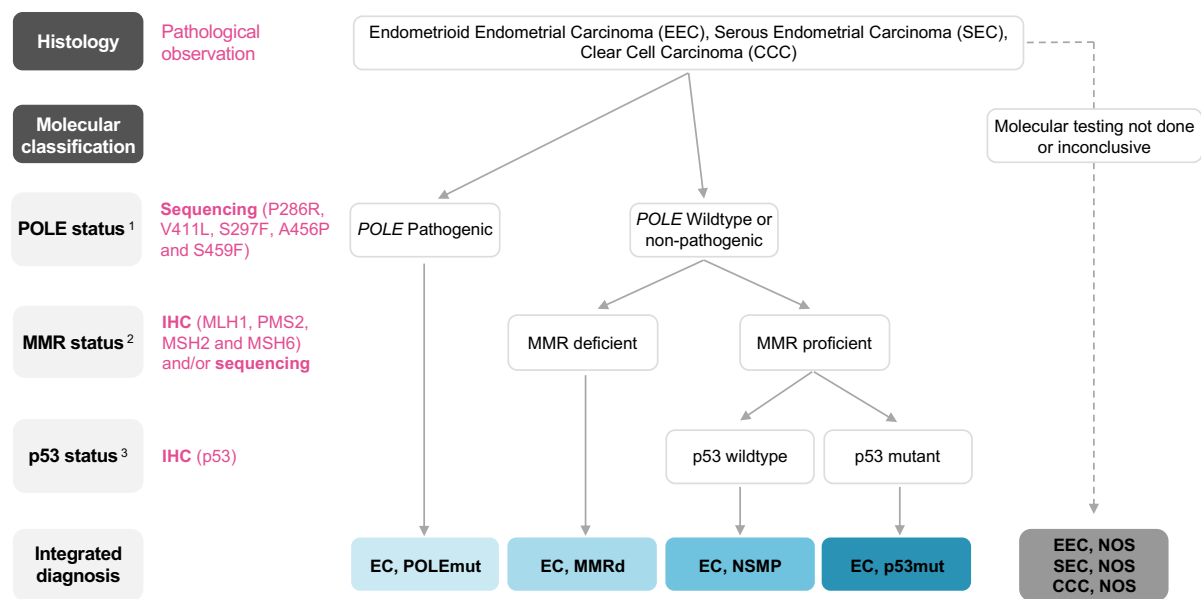
Figure 6. Diagnostic algorithm integrating histology and molecular EC classification recommended by the ESMO-ESGO-ESP 2021 Guidelines [64]. [1] Somatic pathogenic mutation in polymerase-epsilon (*POLE*) include P286R, V411L, S297F, A456P and S459F. [2] Mismatch repair protein (MMR) deficiency is assessed by immunohistochemistry (IHC) and defined by the loss of one or more MMR proteins (MLH1, PMS2, MSH2 and MSH6). Further confirmation by sequencing can be performed. [3] p53 status will be assessed by IHC, which is accepted as a surrogate marker for *TP53* mutational status in MMR proficient, and POLE wildtype EC. *Adapted from Vermij L et al., 2020* [85].

3.4 EC PROGNOSTIC FACTORS

A prognostic factor is a situation, condition, or characteristic of a patient providing information about the outcome of the disease, independent of the therapy, at the time of diagnosis. The most important prognostic factors of EC patients include histopathologic types, histopathologic grade, FIGO stage, depth of myometrial invasion, cervical involvement, tumor size, lymphovascular space invasion (LVSI) and lymph node status.

3.4.1 Histological subtype

EC tumors should be classified according to the recently updated 2020 WHO classification [12]. Histologies of the more common types of EC are represented in Figure 7.

EEC adenocarcinomas represent 80% of ECs, being the majority of them (60% of the total) low-grade tumors. EEC adenocarcinomas are characterized by invasive neoplastic cells showing endometrioid differentiation and with varying proportions of glandular, papillary, and solid architecture. The EEC histology is encountered in the Type I classification, and in the four different molecular subtypes of the molecular classification (see section 3.3). Remarkably, high-grade EEC adenocarcinomas are molecularly heterogenous and also regarding prognostic, therefore, this group has high-preference to use the molecular classification.

The other 20% of ECs are **NEEC adenocarcinomas**. The most frequent NEEC carcinomas are **serous** representing about 10% of all EC but accounting for as many as 40% of EC-related deaths. They usually have papillary architecture and abnormal expression of p53 and p16 with worse prognosis. In the TCGA molecular classification, they are classified within the copy number-high subgroup. **Clear cell carcinomas** are rare tumors (<10%) and they do not have any histotype-specific molecular profile associated. These two histological types are highly invasive and aggressive, associated with poor prognosis. Other subtypes are **undifferentiated and dedifferentiated carcinomas**, which are rare subtypes (2% of all EC) associated to aggressiveness and recurrence or death from the disease in 55-95% of the cases. **Mixed carcinomas** are also common on this subtype, accounting for about 10% of all EC. They are composed of two or more discrete histological types of EC, where at least one component is either serous or clear cell. Other ECs are **mesonephric adenocarcinoma** (1%), **primary squamous carcinoma NOS** (< 0.5%), and **primary gastric-type mucinous carcinoma** (or mucinous adenocarcinoma). Additionally, **carcinosarcomas** (or malignant mixed Müllerian tumors) are biphasic tumors composed of high-grade carcinomatous and sarcomatous components accounting for 5% of all uterine malignances and treated as an aggressive EC.
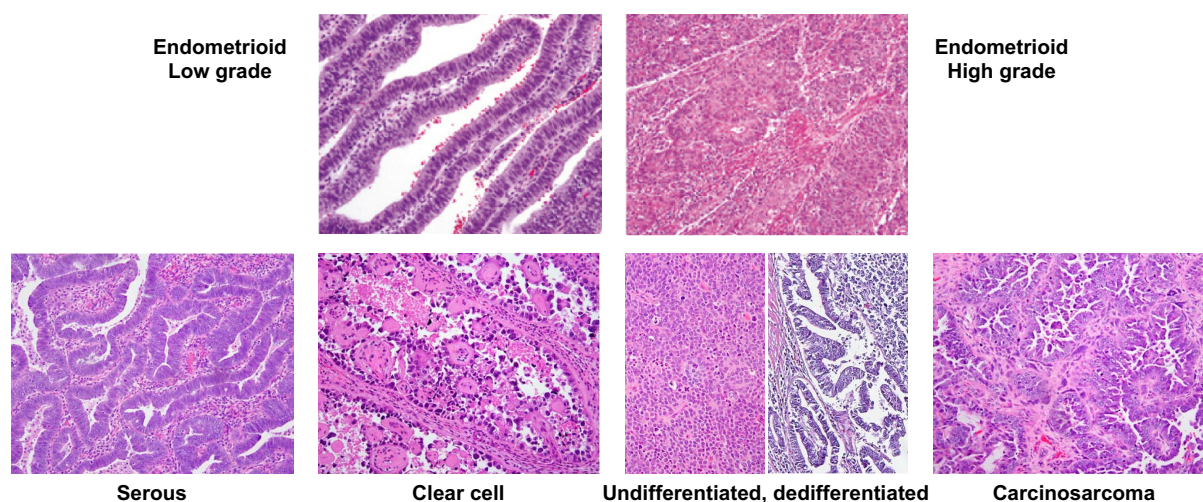
Figure 7. **Histology of the most common types of EC**. Endometrioid EC low grade; Endometrioid EC high grade; Serous carcinoma with typical papillary and micropapillary architecture; Clear cell carcinoma with papillary architecture; prototypical undifferentiated carcinoma and dedifferentiated carcinoma; homologous carcinosarcoma. *Adapted from Murali R et al., 2014* [56]*, Murali R et al., 2019* [86].

3.4.2 Histopathologic grade

All NEECs are considered high grade by definition. However, EEC are graded using the 2009 International Federation of Gynecology and Obstetrics (FIGO) grading criteria based on the assessment of the architectural grade of the tumor followed by the assessment of the nuclear grade [87].

The **architectural grade** is based on three grades according to the proportion of solid areas of tumor cells [88]:

- <u>Grade 1</u> are well differentiated tumors exhibiting ≤ 5% of solid non-glandular, non-squamous growth.
- <u>Grade 2</u> are moderately differentiated tumors exhibiting 6-50% of solid non-glandular, non-squamous growth.
- <u>Grade 3</u> are poorly differentiated tumors exhibiting >50% of solid non-glandular, non-squamous growth.

The **nuclear grade** is based on the presence of significant nuclear atypia. It is characterized by large, pleomorphic nuclei, irregular chromatin clumping, large irregular nucleoli, and loss of cellular polarity. The presence of severe cytological atypia in the majority of cells (>50%) increases the grade by one level [89].

Currently, binary grading is recommended. Thus, grade 1 and 2 tumors are classified as low-grade, while grade 3 tumors as high-grade [64,90]. Despite this, in low-grade EEC, endometrioid differentiation can be diagnosed on routine histology. Whilst in high-grade tumors, squamous differentiation strongly favors EEC over other histological endometrial carcinoma types. Moreover, patients with low-grade tumors have better 5-year overall survival rates. In one study including 1,281 patients (n=925 low-grade, n=356 high-grade) patients with low-grade EC were associated to 87% of 5-year OS rates, while high-grade showed decreased rates to 51.5% of 5-year OS [91].

3.4.3 FIGO stage

The 2009 International Federation of Gynecology and obstetrics (FIGO) staging is the most widely used classification. This staging system classifies EC cases into four stages [87], assessing the same parameters used by the Union for International Cancer Control (UICC): the extent of the tumor (T) , and whether the cancer has spread to the lymph nodes (N) or distant sites causing metastasis (M). Overall survival is highly related to FIGO stage, Stage I being the one with better outcome and Stage IV the one with worse outcome [92]. The four FIGO stages are described and pictured in Figure 8.

FIGO Stage				Union for International Cancer Control (UICC)		
Stage	Description		5-year survival rate (%)	T (tumor)	N (lymph nodes)	M (metastasis)
Stage I [a]	**Tumor confined to the corpus uteri**			T1	N0	M0
IA [a]	NO or les than half myometrial invasion		88.9	T1a	N0	M0
IB [a]	Invasion equal to or more than half of the myometrium		77.6	T1b	N0	M0
Stage II [a]	**Tumor invades cervical stroma, but does not extend beyond the uterus [b]**		73.5	T2	N0	M0
Stage III [a]	**Local and/or regional spread of the tumour**			T3	N0-N1	M0
IIIA [a]	Tumor invades the serosa of the corpus uteri and/or adnexae [c]		56.3	T3a	N0	M0
IIIB [a]	Vaginal and/or parametrial involvement [c]		36.2	T3b	N0	M0
IIIC [a]	Metastasis to pelvic and/or para-aortic lymph nodes [c]					
IIIC1 [a]	Positive pelvic nodes		57	T1-T3	N1	M0
IIIC2 [a]	Positive para-aortic lymph nodes with or wihout positive pelvic lymph nodes		49.4	T1-T3	N1	M0
Stage IV [a]	**Tumor invades bladder and/or bowel mucosa, and/or distant metastases**					
IVA [a]	Tumor invasion of bladder and/or bowel mucosa		22	T4	Any N	M0
IVB [a]	Distant metastases, including intra-abdominal metastases and/or inguinal lymph nodes		21.2	Any T	Any N	M1

[a] Either grade 1, grade 2, or grade 3
[b] Endocervical glandular involvement only should be considered as Stage I and no longer as Stage II
[c] Positive cytology has to be reported separately without changing the stage
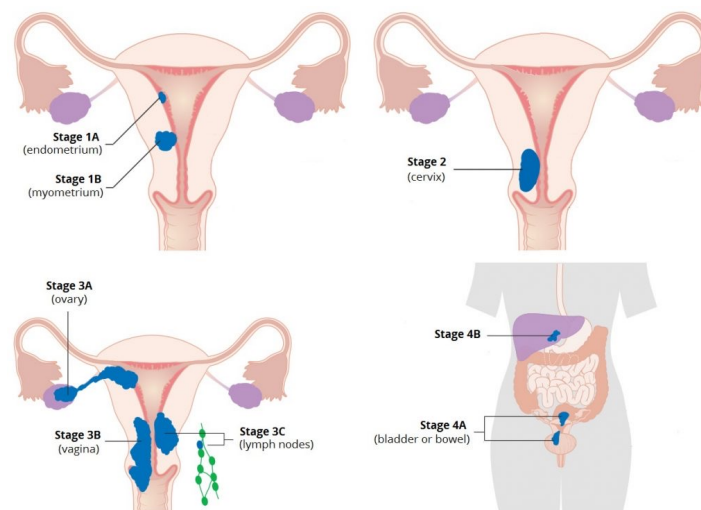
Figure 8. FIGO staging of EC and comparison with the TNM classification. *Adapted from Lewin et al. 2010* [92] *and Frédéric Amant et al. 2018* [93]. *Image adapted from http://teachmeobgyn.com/.*

3.4.4 Others

Other independent prognostic factors, related to the tumor spread and reflected in FIGO staging are **myometrial invasion** and **cervical invasion**. Deep myometrial invasion correlates with more undifferentiated tumors, LVSI, node affectation, cervical stromal invasion, and reduction of progression free survival and overall survival to 67% at 58 months and 66% at 60 months, respectively [94,95]. Additionally, indicating the presence of **LVSI** is also determinant. The presence of tumor cells within vascular spaces is considered an early step in the metastatic processes and a strong predictor of nodal metastasis, recurrence, disease-specific death, and causes an impact on the adjuvant treatment [96]. LVSI-positive tumors have a significantly worse prognosis, not only if extensive LVSI is found, but also in low-risk EC tumors decreasing 11.5% and 10.3% the 5-year recurrence-free survival and 5-year overall survival rates, respectively [97,98].

3.5 EC DIAGNOSIS

3.5.1 Screening test and prevention of EC

To date, there is **not enough evidence to support standard or routine screening test for EC** in the general population or in asymptomatic women with risk factors such as obesity, diabetes mellitus, polycystic ovary syndrome (PCOS), infertility, nulliparity or late menopause. Hence, specific examination of the endometrium should be performed only when suspicious symptoms are present. In some cases, women diagnosed with Lynch Syndrome are referred to an annual surveillance of the endometrium by transvaginal ultrasound and annual or biennial biopsy starting from the age of 35 years and until they receive prophylactic surgery preferably before the age of 40 years [64]. However, uncovering early signs or symptoms can favor the early detection of EC.

3.5.2 Signs and symptomatology

The most common symptom of EC is **abnormal uterine bleeding (AUB)**, which is present in 90% of EC patients [45,99]. Uterine bleeding in menopausal women represents 5% of office gynecology visits [100]. Although many benign disorders generate this symptom (Table 4A) [101], any woman presenting with postmenopausal bleeding must enter the diagnostic process to rule out the presence of the disease [102,103], as the risk of EC, mainly in postmenopausal women with AUB, is high (varies widely from 3% to 25% between studies). The incidence of EC in postmenopausal women presenting AUB is 8-11%, depending on the presence of coexisting risk factors (Table 4B) [99].

A Estimates of EC	%
Prevalence	90
Risk	3-25
Incidence	8-11

* Estimates in postmenopausal women with AUB

B Histology	Approximate percentage (%)
Atrophy	50-60
Endometrial cancer	**8-11**
Hyperplasia	5-15
Polyps	9-12
Proliferative	4
Secretory	1
Others	3

Table 4. Abnormal vaginal bleeding and EC. (**A**) Estimates of EC in women with postmenopausal bleeding. (**B**) Histopathological findings in postmenopausal women with uterine bleeding.

Other frequent symptoms of EC are lower abdominal pain or pelvic cramping, clear vaginal discharge, as well as changes in bowel or bladder functions, anemia, weight loss, and shortness of breath, whereby the latter suggesting a more advanced stage of the disease [104]. However, these symptoms are unspecific and could be related to many other gynecological or non-gynecological diseases.

3.5.3 Diagnostic procedure

Any women with suspicion of EC due to AUB and/or any other symptom related to the disease, should undergo the multistep diagnostic process for EC (Figure 9). Currently, this process consists of a first pelvic examination and transvaginal ultrasonography, followed by the histopathological examination of an endometrial biopsy, preferably obtained by aspiration (i.e., pipelle biopsy), but it can also be obtained by hysteroscopy and/or dilatation and curettage (D&C) in a second line. Finally, women diagnosed with EC, will be subjected to imaging techniques to determine the extent of the tumor.
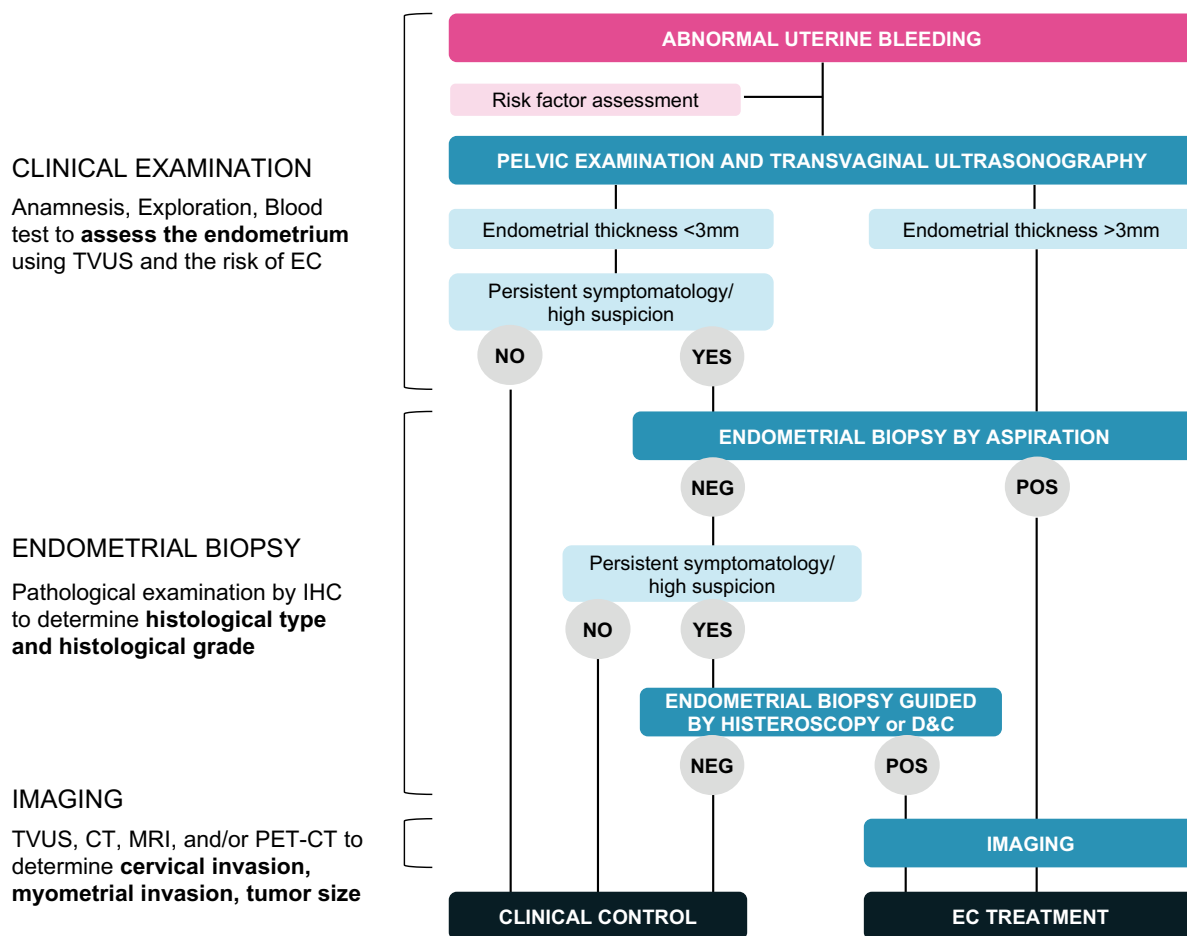
Figure 9. EC diagnostic evaluation. TVUS, transvaginal ultrasonography; D&C, dilatation and curettage; IHC, immunohistochemistry; TVUS, transvaginal ultrasonography; CT, computed tomography; MRI, magnetic resonance imaging; PET-CT, positron emission tomography – computed tomography. *Adapted from Oncoguía SEGO 2016* [23].

3.5.3.1 Clinical examination

√ *Pelvic examination*

Pelvic examination is performed to identify the location and etiology of the bleeding. In women with AUB it is recommended to do a visual inspection of the perineum, vulva, vagina, urethral meatus, and anus; speculum examination of the vulva and cervix; and bimanual examination of the cervix, uterus, and

adnexa [105]. The results are frequently normal, especially at early stages of the disease. Changes in size, shape, or consistency of the uterus and/or its surroundings may exist at advanced stages.

√ *Transvaginal ultrasonography*

Transvaginal ultrasonography (TVUS) is the imaging technique of choice for the assessment of the endometrium in symptomatic patients. It serves to evaluate the endometrial thickness, the presence of an endometrial mass, and/or an endometrial stripe abnormality of the patients. However, it is not recommended as a screening tool in asymptomatic postmenopausal women as it can mislead the physician to perform unnecessary additional biopsies due to false-positive results [106]. Also, it is not recommended in premenopausal women, since endometrium thickness varies during the menstrual cycle as a result of the dynamic hormonal changes, so it has a limited application.

TVUS is described as a **highly sensitive imaging technique, but with very limited specificity** as other benign causes can produce the thickening of the endometrium. Indeed, *Van der Bosch et al.,* recently assessed endometrial thickness in 2,216 women (1,373 premenopausal and 843 postmenopausal) presenting with different gynecological disorders. Results showed an overlap in size of the endometrial thickness between them: women diagnosed with EC had a median of 16mm endometrium, while other benign disorders such as hyperplasia without atypia or endometrial polyp showed a median endometrium thickness of 12.1 mm and 10 mm, respectively. Contrary to this, women with atrophic endometrium reported a median thickness of 5 mm (Figure 10A) [107].

Consequently, a universal consensus regarding the cut-off of endometrial thickness to diagnose EC is lacking. Hence, many studies have focused on the evaluation of the accuracy of TVUS in detecting EC in postmenopausal women with AVB, trying to determine the best cut-off of endometrial thickness. A review and metanalysis of EC diagnosis, including 62 studies with more than 17,000 women, showed how sensitivity increases, while specificity decreases as the cut-off value for endometrial thickness becomes lower. A cut-off of 3 mm demonstrated a 96.2% sensitivity and 42.1% specificity, while setting the cut-off at 5 mm would provide a 96.2% sensitivity and 51.5% specificity. In contrast to this, when the cut-off is increased at 15 mm, sensitivity decreases to 58.9%, and specificity increases up to 94.2% (Figure 10B) [108]. Another study analyzed data from more than 2,000 patients to evaluate the risk of EC when the cut-off was 11 mm, showing how different gynecological malignancies causing thickness of the endometrium would be delimited. In this scenario, 98% of atrophies and 84% of the normal endometrium would present endometrium thickness lower than the cut-off, similar to the 46% and 40% of atypical hyperplasia (precursor of EC) and EC, respectively. Thus, the risk for EC or atypical hyperplasia was 2.6 times greater (Figure 10C) [109]. To better understand the statistical measures terminology, see section 4.2.

A Cut-off	Sensitivity % (95% CI)	Specificity % (95% CI)	NPV	PPV
≥ 3 mm	96.2 (92.4–98.2)	42.1 (26.2–59.8)	99.7	7.3
≥ 4 mm	95.7 (88.1–98.5)	46.0 (36.7–55.6)	99.4	12.7
≥ 5 mm	96.2 (92.3–98.1)	51.5 (42.3–60.7)	99.3	21.1
≥ 6 mm	85.2 (70.2–93.3)	64.0 (53.0–73.6)	99.1	17.6
≥ 8 mm	88.0 (71.4–95.5)	66.2 (52.3–77.8)	98.8	22.3
≥ 10 mm	78.2 (64.7–87.5)	83.7 (74.4–90.1)	98.1	33.7
≥ 15 mm	58.9 (48.6–68.5)	94.2(76.6-98.5)	94.0	55.7

C Histology	Endometrial thickness	
	< 11 mm	≥ 11 mm
Atrophy	703 (98%)	14 (2%)
Endometrial cancer	12 (40%)	18 (60%)
Hyperplasia with atypia	6 (46%)	7 (54%)
Hyperplasia without atypia	16 (70%)	7 (30%)
Insuficient material	42 (81%)	10 (19%)
Myoma	9 (35%)	17 (65%)
Normal endometrium	83 (84%)	16 (16%)
Polyp	319 (55%)	256 (45%)
Total	1190	345

Figure 10. Endometrial thickness assessment. (**A**) Summary estimates of sensitivity and specificity for different combined cut-off values of endometrial thickness. *Adapted from Long B et al. 2020* [108]. (**B**) Median endometrial thickness in the different histological endpoints. *Adapted from Van den Bosch et al., 2021* [107]. (**C**) Distribution of histological diagnosis according to endometrial thickness. *Adapted from Alcázar et al. 2018* [109].

As there is not a global consensus on the precise value, different gynecological guidelines might recommend management of EC patients using different thresholds. Whilst the Spanish Society of Gynecology and Obstetrics (SEGO) guidelines recommend an endometrial biopsy in postmenopausal patients with AUB and endometrial thickness over 3mm in TVUS, or in patients with persistent vaginal bleeding regardless of the TVUS findings [23], others such as British Gynaecological Cancer Society (BGCS) guidelines [110] in Britain, the German Cancer Society and German Cancer Aid (AWMF) guidelines [111] in Germany, or the Associazione Italiana di Oncologia Medica (AIOM) guidelines in Italy [112], establish different cut-offs depending on different variables. Specifically, the British guidelines recommend a cut-off of 4mm or higher, however, if a woman with AUB is being treated with hormone replacement therapy (HRT) or tamoxifen, the cut-off increases to 5mm, and if a woman has no symptoms but is treated with HRT the cut-off is over 8mm. The Italian guidelines distinguish between premenopausal and postmenopausal women and recommend cut-offs over 14mm and 5mm, respectively. Similarly, the German guidelines distinguish between premenopausal and postmenopausal women, but also consider other factors such as HRT, and it establishes different cut-offs and protocols depending on the given scenario ranging from 3mm in postmenopausal women with AUB and smooth and homogeneous endometrium to 20mm in premenopausal women with AUB with inhomogeneous, not sharply delineated endometrium, polyps, BMI > 30, suspicious cytology, diabetes or Lynch syndrome. In contrast, the National Comprehensive Cancer Network (NCCN) guidelines [83] in the US it is recommended that under specific symptomatology (AUB), the physician should take action and collect an endometrial biopsy without the need of TVUS assessment (see section 3.5.3.2).

√ ***Blood tests***

Additionally, blood tests to determine serum tumor markers cancer antigen 125 (CA-125) and human epididymis protein 4 (HE-4) are often included as part of the clinical evaluation. They have demonstrated correlation with prognostic factors [113–115]. However, the appropriate cut-off is not determined and their clinical relevance is still questioned [116,117].

3.5.3.2 Endometrial biopsy

The pathological examination of an endometrial biopsy is the gold standard for EC diagnosis. This methodology is based on the collection of an endometrial sample, which is added to a fixing solution to protect and preserve it from degradation. The sample will then be processed to finally be included in a paraffin embedded block which will be microscopically examined by a pathologist. Through the observation of the cells contained in the sample, the pathologist will be able to provide a diagnosis. Additionally, information regarding **histological type and histological grade** will be provided for those cases where malignant cells are present in the sample.

There are different procedures to obtain endometrial biopsies. Traditionally, they were collected by "dilatation and curettage" (D&C). Nowadays, it has been replaced in most places by aspiration or guided by hysteroscopy (Figure 11) [45,116].

√ ***Endometrial biopsy by aspiration***

Endometrial biopsies can be performed by aspiration, blindly, with a soft, straw-like device that suctions a small sample from inside the uterine cavity. The biopsy obtained is called **uterine aspirate or pipelle biopsy**. It is a liquid and complex sample expected to contain cells coming from the endometrium. Aspiration is considered as the method of choice to get an endometrial biopsy, as it is a fast, easy to perform, cost-effective, and safe procedure that is well tolerated by patients [118]. Additionally, it is easy to perform in the same office, as there is a window of opportunity for taking the biopsy after the TVUS and it does not require any previous tests for coagulation factors or anesthesia [23,45,116,119,120].

The overall accuracy of endometrial biopsy by aspiration is high when an adequate specimen is obtained, performing better in symptomatic (bleeding) and postmenopausal women [121]. In order to determine the accuracy of pipelle biopsy to diagnose EC patients, many studies have been performed and compiled in different reviews and meta-analyses [120–122]. Specifically, the most recently performed study by Narice BF et al. in 2018, compiled 60 articles reporting data from over 7,300 patients. The results of this meta-analysis revealed a **sensitivity** ranging from 62% to 99.2%, and a **specificity** of 99.1-100% of pipelle biopsies to diagnose EC patients [120].

In the last decade, studies reporting the performance of pipelle biopsies to diagnose EC continued reproducing the results obtained to date. Reported sensitivities range from 56.8 [123] to 100% [124–130], being lower in those cases where the gold standard approach was hysterectomy [123,131,132], while the reported

specificities are near or even reaching 100% [124–129,131], with the minimum specificity reported in 89.5%[123]. Likewise, PPV rage from 92.6% [133] to 100% [123,126,130,132], whereas the NPV and the accuracy of the technique are also near 100% in most of the studies [125,126,130–133]. Remarkably, Taraboanta C et al. recently performed a large retrospective cross-sectional study including 1,677 hysterectomy specimens containing atypical hyperplasia or EC. Those were matched against a database including 24,380 endometrial biopsies reported negative in order to assess the performance of pipelle biopsies in atypical hyperplasia and/or EC. They determined an overall 89.3% sensitivity and 99.3% specificity to diagnose atypical hyperplasia or EC. Specifically, pipelle demonstrated 91% sensitivity for diagnosing EC (92% and 99% sensitivity for EEC and SEC, respectively) with 99.8% PPV and 99.4% NPV. Additionally, pipelle biopsy showed 98% sensitivity for atypical hyperplasia [134].

Likewise, pipelle biopsies reported 100% of NPV for hyperplasia, while 89.6% NPV for endometrial polyps [120]. Despite the exceptionally high NPV of pipelle biopsies, women that obtain a negative result yet are still considered to be at high risk for EC and/or her symptoms persist, should undergo further evaluation (Figure 9) [121].

However, pipelle biopsies are associated with two main drawbacks:

First, in women with postmenopausal bleeding, **endometrial sampling fails in up to 42% of cases, of which a (pre)malignancy is found in 7%**. This failure is due to (i) technical problems in 11% of cases (ranging between 1% and 53%), mainly due to cervical stenosis, or (ii) insufficient material found at histology in 31% of cases (ranging from 7% to 76%) [133,135]. Therefore, in case of a failed or inconclusive pipelle biopsy, endometrial biopsies will have to be repeated using hysteroscopy and/or D&C.

Second, the often limited amount of tissue contained in pipelle biopsies and the subjectivity of the pathological examination drives in an inaccurate diagnosis of the **tumor type and grade** in some cases. A recent meta-analysis showed **discrepancies in 11.4% of cases regarding tumor histology** (specifically, 5%, 18% and 18% discrepancies in EEC, EEC grade 3, and NEEC, respectively)**, while 27% of disagreement was found in grade staging** when using pipelle biopsies [136].

Although pipelle biopsies are the method of choice thanks to their high accuracy and low invasiveness, women whose sample could not be obtained due to technical problem, women with inadequate sampling or an inconclusive result, and women with negative result but still under suspicion of high risk of EC, won't receive a final and conclusive diagnosis using this procedure and must undergo a biopsy guided by hysteroscopy.

√ *Endometrial biopsy by hysteroscopy*

The basic hysteroscope is a thin, lighted tube (endoscope) that carries a camera usually attached to its proximal end. It is introduced into the uterus through the vagina and the cervical canal, transmitting the image of the uterus to a screen. To achieve this procedure, it is required to dilatate the cervix, which

might need administration of dilators such as misoprostol. Furthermore, for the insertion, inspection and operation of the uterine cavity, a fluid, such as sodium chloride solution, must be pumped though the hysteroscope. Finally, surgical instruments such as biopsy forceps can be passed though the hysteroscope to perform directed endometrial biopsies that are histologically examined.

The use of pipelle biopsies is less common than performing hysteroscopy, which is still the preferred method for many physicians, hospitals and/or countries. This is due to the feasibility to visualize the endometrial cavity while taking the biopsy. However, this has also been considered controversial as it can lead to subjective diagnostic orientation. In the last decade studies reporting the performance of hysteroscopies to diagnose EC have been performed. Different studies reported **sensitivities** ranging from 60.8% [123] to 100% [137,138], **specificities** from 94.2% to 99.6% [137,138], and PPV and NPV ranging from 66.7%-100% and >99%, respectively [123,137,138], to diagnose EC. Likewise, hysteroscopies reported an overall sensitivity and specificity of 72.2% and 91.5%, respectively for endometrial hyperplasia, while 95.4% and 96.4% for endometrial polyps [139] (Figure 11).

Hysteroscopies also present some drawbacks. Importantly, it has a **failure rate of 4.7%** ranging from 4.5% to 5% [140,141]. 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). Additionally, endometrial biopsies acquired by hysteroscopy reported discrepancies between pre-operative and final histological type ranging between 0-14.7% [142,143]. Also, in a recent meta-analysis, **discrepancy in the definition of histological grade of the tumors in endometrial biopsies was shown in 11%** of cases when using hysteroscopies [136].

The procedure is more invasive than endometrial biopsies performed by aspiration. It requires a hospital setting, prior blood testing and might also require anesthesia. During the procedure, it might cause complications and, although the rate of **complications** is 2%, severe cases such as infections, and uterine perforations, among others, have been recorded [144,145].

Lastly, some authors described a risk when performing hysteroscopies for disseminating cancer cells into the peritoneal cavity, however there is a lack of evidence to support these idea [146,147].
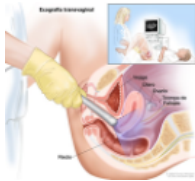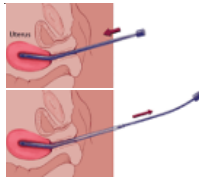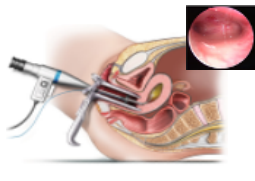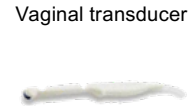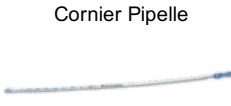
	TRANSVAGINAL ULTRASOUND (≥ 3mm)	ENDOMETRIAL BIOPSY BY ASPIRATION	ENDOMETRIAL BIOPSY BY HYSTEROSCOPY
Procedure			
	Vaginal transducer	Cornier Pipelle	Hysteroscope
Instrument			
Diagnostic features			
Comfort	Non-invasive	Minimally invasive	Invasive
Specific requirements	No requirements	No requirements	Coagulation tests Anesthesia Hospital
Economy	Less expensive	Less expensive	More expensive
Complications	No	No	Yes
Diagnostic performance			
Diagnostic performance	0%. EB required in all cases	< 80% diagnosed patients	96% diagnosed patients
Technical failure	-	11%	4%
Insufficient material		31%	
Accuracy diagnosis			
Sensitivity	96.2%	56.8 - 100%	60-100%
Specificity	41.1%	89.5 - 100%	94.2 - 99.6%
PPV	12.7%	92.6 - 100%	>99%
NPV	99.4%	>99%	66.7 - 75%
Accuracy prognostic factors			
Pre- *vs* final diagnosis		70% (CI 95% 0.60-0.79)	89% (CI 95% 0.80-0.98)
Histological type	-	88.6%	~90%
Histological grade	-	73%	89%
Molecular classification	-	89% (78-96%)	

Figure 11. Procedures and features of the different methodologies to diagnose EC. Transvaginal ultrasonography to assess the endometrium and aspiration and hysteroscopy to obtain an endometrial biopsy. EB, endometrial biopsy; PPV, positive predictive value; NPV, negative predictive value, CI, confidence interval.

3.5.3.3 Imaging techniques

Imaging techniques represent a cornerstone in the pre-operative evaluation of patients with EC. **Myometrial invasion, tumor size, and cervical invasion** are properly assessed using imaging techniques such as TVUS, or magnetic resonance imaging (MRI). However, both techniques demonstrated poor to moderate sensitivity in detecting high-risk features. Regarding myometrial invasion assessment, a study comparing TVUS and MRI reported pooled sensitivity and specificity of 75% and 82% for TVS and 83% and 82% for MRI [148], respectively. Likewise, MRI reported sensitivities and specificities of 53-57% and 95%, respectively, for cervical invasion [149,150].

Lymph node status can be studied by computed tomography (CT), MRI or positron emission tomography (PET)-CT, and those techniques seem to have similar accuracies. Specificity is high but the sensitivities are moderate to low [151], with reported sensitivity and specificity values of 44% and 96%, respectively [149]. In fact, imaging cannot replace lymph-node dissection for staging purposes since the detection rate for metastatic lesions of 4mm or less is only 12% [152].

Another important prognostic factor is the **LVSI** since the presence of tumor cells within vascular spaces is considered an early step in the metastatic process and consequently a strong predictor of nodal metastasis, recurrence and cancer-specific death [153]. However, it cannot be assessed using imaging techniques and is usually not observed in pre-operative biopsies. Thus, the decision to perform lymphadenectomy can be made based on intra-operative frozen sections or pre-operative risk assessment based on histology and imaging tests, where the most important criteria are myometrial invasion and tumor grading. In this regard, sentinel lymph node mapping has emerged as a new strategy to determine the lymph node status while sparing many lymph node dissections [114].

3.6 RISK ASSESSMENT AND TREATMENT

3.6.1 Pre-operative risk assessment

The purpose of pre-operative risk assessment is to classify patients according to the groups of risk for lymphatic dissemination and disease recurrence in an early stage of the diagnostic process to define the most optimal surgical management. Hence, the last update on the risk stratification system to manage EC patients was published by the ESGO-ESTRO-European Society of Pathology (ESP) consortium in 2021 [64]. This system integrates not only clinical and pathological information coupled to the definition of the spread of the tumor observed from imaging techniques, but it also includes molecular information described by the TCGA consortium [44]. As a result, it provides additional prognostic information that permits an improvement of the previous classification system of the disease, individualizing the treatment of each patient (Table 5).

Currently, the extent of the surgical procedure is guided by the assessment of known recurrence risk factors such histological type and tumoral grade, myometrial invasion, lymph node and/or extrauterine affectation using FIGO criteria and, when possible, molecular profiling. Pre-operatory staging will also collect a detailed anamnesis with clinical assessment, whereby age and comorbidities are important features associated with the patient that could contraindicate the surgery or limit its extension [154]. The most appropriate imaging technique to perform on each patient in order to determine the above-mentioned prognostic factors will depend on the extension of the disease (confinement to the uterus *vs* extrauterine disease), and the histological type and grade.

Molecular classification unknown

Risk group	FIGO stage	Histological type	Histological grade	LVSI	Myometrial invasion	Residual disease
Low	IA	EEC	low-grade	Negative or focal		
Intermediate	IB	EEC	low-grade	Negative or focal		
Intermediate	IA	EEC	high-grade	Negative or focal		
Intermediate	IA	NEEC	high-grade		No	
High-Intermediate	I	EEC	any	Substantial	any	
High-Intermediate	IB	EEC	high-grade	any		
High-Intermediate	II	EEC				
High risk	III-IV	NEEC				No
High risk	I-IVA	NEEC			Yes	No
Advanced metastatic	III-IVA					Yes
Advanced metastatic	IVB					

Molecular classification known *†

Risk group	FIGO stage	Molecular subgroup	Histological type	Histological grade	LVSI	Myometrial invasion	Residual disease
Low	I-II	POLEmut					No
Low	IA	MMRd or NSMP	EEC	low-grade	Negative or focal		
Intermediate	IB	MMRd or NSMP	EEC	low-grade	Negative or focal		
Intermediate	IA	MMRd or NSMP	EEC	high-grade	Negative or focal		
Intermediate	IA	p53abn	NEEC			No	
Intermediate	IA	p53abn	NEEC			No	
Intermediate	IA					No	
High-Intermediate	I	MMRd or NSMP	EEC	any	Substantial	any	
High-Intermediate	IB	MMRd or NSMP	EEC	high-grade	any		
High-Intermediate	II	MMRd or NSMP	EEC				
High risk	III-IVA	MMRd or NSMP	EEC				No
High risk	I-IVA	p53abn				Yes	No
High risk	I-IVA	MMRd or NSMP	SEC, Undiff CA, CS			Yes	No
Advanced metastatic	III-IVA						Yes
Advanced metastatic	IVB						

*For stage III–IVA POLEmut endometrial carcinoma and stage I–IVA MMRd or NSMP clear cell carcinoma with myometrial invasion, insufficient data are available to allocate these patients to a prognostic risk group in the molecular classification. Prospective registries are recommended.

†See text on how to assign double classifiers (eg, patients with both POLEmut and p53abn should be managed as POLEmut).

‡According to the binary FIGO grading, grade 1 and grade 2 carcinomas are considered as low-grade and grade 3 carcinomas are considered as high-grade.

LVSI, lymphovascular space invasion; MMRd, mismatch repair deficient; NSMP, non-specific molecular profile; p53abn, p53 abnormal; POLEmut, polymerase-mutated; SEC, serous endometrial carcinoma; Undiff CA, undifferentiated carcinoma; CS, carcinosarcoma

Table 5. 2020 ESGO-ESTRO-ESP prognostic risk group to guide adjuvant therapy use. *Adapted from Concin et al. 2021* [64].

Histological type and grade have become key factors for risk group assignment and are based on the histopathological examination of pre-operative endometrial biopsies (see section 3.5.3.2) [155]. These are used not only to discriminate between EC and non-EC patients, but also to assess these prognostic factors of the tumor. However, it is known that the concordance between pre-operative endometrial sampling and final diagnosis is moderate and variable among different studies mainly due to the limited tissue available in the pre-operative biopsies. Specifically, a recent meta-analysis including 45 studies described the agreement between histological subtypes and grades assessed in endometrial sampling (either by aspiration, hysteroscopy or D&C) versus the final diagnosis [136]. Results of the meta-analysis reported an agreement of 95% (95% CI 94-97%) in EEC grade 1 and 2, while an agreement of 81% (95% CI 74%-91%) was shown in EEC grade 3. Regarding histological grade, an overall grade correlation of 67% was found in EEC (75% for grade 1, 61% for grade 2, and 75% for grade 3). Importantly, clinically relevant downgrading was reported in 26% of cases and upgrading in 8% of patient's samples (see section 3.5.3.2 for method-specific inter-observer variability). Regarding NEEC tumors, an overall correlation of 81% (69%-92%) was reported, with greater agreement in carcinosarcomas (90%) and worse in serous and clear cell carcinomas (67%) [156]. As the determination of these factors is crucial to choose the most suitable imaging technique, these discordances in histological type and grade can lead to an underestimation of the risk of lymph node metastasis (undertreatment) or to the performance of unnecessary surgical procedures associated with complications (overtreatment), which might lead to comorbidities or even life-threatening risks.

Hence, independently of the technique used, the accuracy of endometrial biopsy diagnoses is widely variable among studies, being dependent on the physiologist acquiring the biopsy, the amount of tissue available, and the pathologist assessing the sample. The recent introduction of **molecular profiling** in the risk classification of EC is a potential tool to accurately classify EC patients at pre-operatory level. As the use of clinically applicable surrogate marker tests assessing few biomarkers can provide molecular profiling (see section 3.3.2) in endometrial biopsies, whereby high correlation between those pre-operatory biopsies and the resected tumor tissue has been shown [76,157,158].

Finally, evaluation of the myometrial invasion, cervical invasion and lymph node metastasis are assessed by imaging techniques. Extension of the disease will determine the imaging techniques that should be performed. Whereas in pre-operatory early stages (FIGO I-II) and low-grade histology, pelvic MRI and/or TVUS is recommended to evaluate the uterus locally (myometrial invasion, cervical status), in pre-operatory late stages (FIGO III-IV) and/or high-grade histologies, metastatic disease should be discarded (nodal, peritoneal, and metastatic) through a CT of the thorax, abdomen and pelvis [23].

3.6.2 Primary treatment: Surgery

3.6.2.1 EC at initial stage (stage I-II)

Surgery is the primary treatment of EC confined to the uterus (Table 6). The surgical treatment includes **total hysterectomy** (removal of the uterus) and **bilateral salpingo-oophorectomy** (removal of both Fallopian tubes and ovaries). Information regarding lymph node status is essential to reduce the risk of recurrence and to reduce cause-specific death in EC patients, as patients identified with lymph node metastases will receive a different adjuvant treatment that associates chemotherapy to radiotherapy (5-year overall survival of 78.5% when both therapies are performed versus 68.5% only with radiotherapy) [159]. However, the staging lymphadenectomy is associated with intra-operative morbidity (such as increase in bleeding or operative time) and post-operative sequelae as lymphocele or lymphedema. Moreover, its therapeutic benefit has never been shown in prospective, randomized studies[160], and for these reasons an alternative strategy has emerged as a validated tool to identify the lymph node status: the sentinel lymph node biopsy.

Sentinel lymph node biopsy (SLNB) assessment diagnoses lymph node metastases with a sensitivity of 96% (95%CI 93-98%) and a NPV of 99.7% [161]. SLNB assessment should be performed in addition to an ultrastaging pathologic analysis of the node (ultra-sectioning and cytokeratin's IHQ), allowing the maximum diagnostic benefit. [64]. The procedure has different advantages in each group of patients: (i) low-risk patients, it allows a diagnostic study of the lymph node status, which are affected in 10% of cases, and allows re-stratification [162]; (ii) intermediate-risk patients, SLNB assessment is a surrogate for the pelvic lymphadenectomy, as in this group 15% of patients will have lymph disease and systemic lymphadenectomy is recommended [162]; (iii) high-risk patients, SLNB will give additional information allowing the detection of low volume disease, as in this group 40% of the patients will have lymph node disease [163]. Additionally, this group has the highest number of FN (~5%), in particular in NEEC tumors.

Conventional oncologic surgery for EC is performed using **minimally-invasive surgery** (laparoscopy or robotic surgery) , as they offer the same oncologic results but with faster recuperation and reduced morbidity associated with the procedure [164].

3.6.2.2 EC at advance stage (stage III-IV)

However, when the tumor is diagnosed at an advanced stage the approach should be open surgery (laparotomy) aiming to remove the presence of residual disease. Those patients who are not candidates for surgery due to medical criteria, should be offered external radiotherapy and braquitherapy treatment as the primary treatment, even when the cure rates are smaller. The surgical procedures recommended according to the stage and pre-operative risk assessment are summarized in Table 6 [64].

		Early stage disease. Stage I/II			Advanced stage disease. Stage III/IV
		Low	**Intermediate**	**High**	**Advanced metastatic**
No MC		EEC, G1G2, Stage IA	EEC, G1G2, Stage IB / EEC, G3, Stage IA / NEEC, no MI	Stage II / EEC, G3, Stage IB / NEEC with MI	
MC		POLEmut, Stage I-II	p53mut, no MI / EEC, MMRd-NSMP, G1G2, Stage IB	p53mut, with MI / EEC, MMRd-NSMP, G3, Stage IB	
		MMRd-NSMP, G1G2, Sta…	EEC, MMRd-NSMP, G3, Stage IA / NEEC, MMRd-NSMP, no MI	EEC, MMRd-NSMP, Stage II / NEEC, MMRd-NSMP, with MI	
Surgical procedures	**Total hysterectomy**				
	Bilateral salpingo-oophorectomy				
	Pelvic lymph-adenectomy	**SLNB** recommended. For staging purposes only.	**SLNB** OR staging lymphadenectomy if no drain of the SLN	**SLNB** OR staging lymphadenectomy if no drain of the SLN	
	Aortic lymph-adenectomy		In case of positive pelvic node, or clinical criteria	Recommended	
	Palliative cytoreductive				Aiming absence of residual disease
	Others			Cavity exploration, biopsies of suspisious lesions, staging infracolic omenectomy (SEC, Carcinosarcoma and undifferentiated tumors).	

Table 6. Surgical treatment of EC. *Adapted from Concin et al. 2021* [64].

3.6.3 Histopathological staging: final diagnosis and final risk-stratification

The final staging of the tumor is only obtained with the anatomopathological examination of the surgically resected tissue. Thus, once all prognostic factors are identified in the resected tumor, *i.e.,* after surgical treatment, patients are reclassified according to the risk stratification system defined by three of the major consortium in EC, *i.e.,* the ESMO-ESTRO-ESP consensus [64]. It uses the clinical, molecular, and pathological information of patients from low to metastatic, measuring the risk of recurrence to classify, treat and predict the outcome of EC patients (Figure 12). The selection of the most optimal adjuvant treatment is based on the risk of recurrence established by the risk stratification systems.

To date, the proposed classification has been applied to a total number of 3,650 patients diagnosed with EC comprised in eight studies by the Vancouver and the PORTEC groups (Figure 12A) [58,59,75,77–80,165]. The mentioned studies served not only to validate the prognostic relevance of the molecular classification, but they also pointed out specific subsets of patients that would benefit from this new classification that was recently proposed by the ESMO-ESTRO-ESP consortium in terms of treatment [64]. Tumors classified as POLEmut are associated with diagnosis at an early stage, which seems to resemble the outcome of the molecular groups. Thus, the fact that POLEmut and p53mut groups are attributed to have excellent and poor prognosis, respectively, might be attributed to this fact. (Figure 12B). Controversially, according to the summarized studies, POLE ultramutated group have received the second-highest rate of adjuvant treatment (51.2%) after p53mut (58.8%), which is likely attributed to the fact that POLEmut tumors are commonly presented as high-grade EEC and, consequently,

classified as high-risk. p53wt and MMRd patients received adjuvant treatment in 44.6% and 46.5% of the cases (Figure 12C).

Figure 12. Published trials using ProMisE algorithm. (**A**) Summary of published trials using ProMisE; (**B**) analysis of the FIGO stage; (**C**) adjuvant treatment distribution in terms of the patients' molecular subtype out of the

A

Author	Patient Cohort	N patients	FIGO Stage					Subtypes			
			IA	IB	II	III	IV	POLE	MMRd	p53wt	p53mut
Stelloo 15	Portec 3 criteria	116	36.2%		18.1%	25.3%	9.5%	12.1%	16.4%	37.9%	33.6%
Talhouk 2015	"discovery"	143	71.3%			28.7%		8.4%	28.7%	44.1%	17.5%
Stelloo 2016	PORTEC 1 & 2	834	n/a			n/a		5.9%	26.3%	59.0%	8.9%
Talhouk 2017	"confirmation"	319	69.3%			29.5%		9.4%	20.1%	27.0%	43.6%
Bosse 2018	Grade 3 EEC	381	44.9%	31.5%	30.2%	13.1%	2.9%	12.9%	36.2%	30.2%	20.7%
Cosgrove 2018	NRG / GOG GOG 210	982	74.5%		9.3%	14.4%	1.8%	4.0%	38.6%	48.9%	8.6%
Kommoss 2018	"validation"	452	61.1%	19.7%	3.8%	12.2%	1.3%	9.3%	28.1%	50.4%	12.2%
León-Castillo 2020	PORTEC 3	423	13.2%	17.8%	25.6%	43.4%		12.4%	33.4%	31.5%	22.7%
Total		3650						7.9%	30.9%	46.5%	14.7%

B

Author	Patient Cohort	N patients	HR OS Multivariate				HR RFS Multivariable			
			POLE	MMRd	p53wt	p53mut	POLE	MMRd	p53wt	p53mut
Stelloo 15	Portec 3 criteria	116								
Talhouk 2015	"discovery"	143	0.28 (0.00-3.01)	0.90 (0.31-2.73)	1.00	4.28 (0.95-18.34)	0.15 (0.00-1.94)	0.32 (0.10-1.03)	1.00	1.64 (0.32-7.06)
Stelloo 2016	PORTEC 1 & 2	834	1.105 (0.394-3.101)	1.879 (1.307-2.700)	1.00	3.777 (2.364-6.037)				
Talhouk 2017	"confirmation"	319	1.01 (0.26-2.99)	1.90 (0.88-4.04)	1.00	2.61 (1.27-5.72)	0.19 (0.02-0.81)	0.64 (0.25-1.60)	1.00	1.75 (0.84-3.96)
Bosse 2018	Grade 3 EEC	381	0.56 (0.27-1.15)	0.84 (0.57-1.25)	1.00	1.37 (0.9-2.09)	0.23 (0.07-0.77)	0.61 (0.37-1.00)	1.00	1.92 (1.20-3.07)
Cosgrove 2018	IRG / GOG GOG 21	982	0.19 (0.03-1.35)	1.04 (0.70-1.56)	1.00	1.61 (0.93-2.78)	0.26 (0.06-1.05)	1.08 (0.78-1.50)	1.00	1.56 (0.99-2.48)
Kommoss 2018	"validation"	452	0.95 (0.30-2.36)	1.41 (0.82-2.41)	1.00	2.29 (1.12-4.65)	0.15 (0.00-n/a)	1.54 (0.73-3.24)	1.00	3.40 (1.30-8.81)
León-Castillo 2020	PORTEC 3	423	0.118 (0.016-0.868)	1.00	0.547 (0.302-0.993)	2.298 (1.418-3.726)	0.079 (0.011-0.576)	1.00	0.976 (0.620-1.537)	2.517 (1.621-3.907)
Total		3650								

C
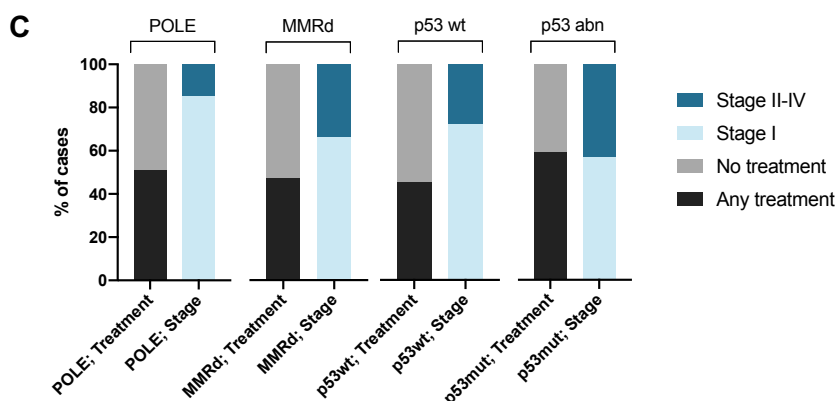
published trials (excluding *León-Castillo et al. and Bosse et al.* from this analysis) [58,59,75,77–80,165]. HR, hazard ratio; OS, overall survival; RFS, recurrence-free survival, n/a, not available.

3.6.4 Adjuvant treatment

Adjuvant treatment refers to a non-surgical treatment given in addition to primary surgery aiming to reduce the risk of recurrence. In EC, main modalities for adjuvant treatment have been chemotherapy, radiotherapy and hormonal therapy. Most of EC patients are diagnosed at early stages of the disease and, therefore, at low risk of recurrence. Consequently, it is accepted that the group of ~55% of EC patients presenting an overall 5-year survival rate of 95%, will not benefit from any adjuvant post-surgical therapies and they will be only treated with surgery. Remarkably, the last update in the risk assessment, which considers the novel molecular classification, fits POLEmut patients in the low-risk or intermediate-

risk groups, independently of the stage. However, it is not clear whether this subgroup of patients should receive adjuvant treatment. For other EC patients, the disease stage and the recurrence risk of the patient will define the recommended adjuvant treatment (see Table 7).

Postoperative staging	MC	Specific features	Recommended adjuvant treatment
Low risk	No		**Observation. No adjuvant treatment.**
	Yes	[Stage I-II, POLEmut]	Observation. No adjuvant treatment.
	Yes	[Stage III-IVA, POLEmut]	No sufficient data. Prospective registration is recommended
Intermediate risk	No		**Adjuvant brachytherapy can be recommended to decrease vaginal recurrence. Omission of adjuvant brachytherapy can be considered especially for patients aged <60 years**
	Yes	[POLEmut, p53abn, MI]	Specific recommendations
	Yes	[p53abn restricted to a polyp] OR [no MI]	Adjuvant therapy is generally not recommended
High-intermediate risk	No		**Adjuvant brachytherapy can be recommended to decrease vaginal recurrence.** **Omission of any adjuvant treatment is an option**
	No	[Substantial LVSI] AND/OR [Stage II]	EBRT can be considered
	No	[All] but especially for [high-grade] AND/OR [substantial LVSI]	Adjuvant chemotherapy can be considered
	Yes	[POLEmut AND p53abn]	Specific recommendations. See respective recommendations for low- and high-risk
High-intermediate risk (lymph node staging not performed)	No	[All], but especifically for [substantial LVSI] AND/OR [Stage II]	Adjuvant ERBT is recommended
	No	[All], but especifically for [high-grade] AND/OR [substantial LVSI]	Additional adjuvant chemotherapy can be considered
	No	[High-grade, LVSI negative] OR [Stage II, grade 1]	Adjuvant brachytherapy alone
	Yes	[POLEmut AND p53abn]	Specific recommendations. See respective recommendations for low- and high-risk
High-risk	No		**EBRT with concurrent and adjuvant chemotherapy or alternatively sequential chemotherapy and radiotherapy is recommended** **Chemotherapy alone is an alternative option** **Carcinosarcomas should be treated as high-risk carcinomas (not as sarcomas)**
	Yes	[p53abn, no MI, POLEmut]	Specific recommendations (see respective recommendations for low- and intermediate-risk)

Table 7. Recommended adjuvant treatment of EC. *Adapted from Concin et al. 2021* [64] *.* MC, molecular classification; ERBT, external bean radiation therapy; LVSI, lymphovascular space invasion.

4 BIOMARKERS IN MEDICINE

4.1 BIOMARKER DEFINITION AND TYPES

The working Group of the National Institute of Health defined a **biomarker** as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention". An ideal biomarker should be easily obtained with minimum discomfort or risk to the patient, and it should have excellent diagnostic and/or prognostic performance associated with high specificity (low rate of false positives) and high sensitivity (low rate of false negatives). It must be reproducible across a wide and representative range of populations. and the final assay should be objective, quantifiable and economical [166].

Biomarkers are a keystone of medical care, and they have several potential applications to revolutionize clinical practice to approach personalized medicine. There are different types of biomarkers, including screening, diagnosis, prognosis, prediction of response to treatment and monitoring the progression of disease. Each type of biomarker must have specific objectives and characteristics, and must enable the discrimination of groups of patients with different status of the disease that require different management (Table 8) [167].

	SCREENING	DIAGNOSIS	PROGNOSIS	PREDICTION	MONITORING
Definition	Predictor of the probability of developing a specific disease	Indicates if an individual has or not a specific disease	Usually after standard treatments to classify patients into different risk groups, independently of the therapy.	Predict subpopulations that will respond to a specific treatment.	Follow-up of patients that remain disease free after therapy.
Population	Healthy	Symptomatic	Diseased	Diseased	Recovered patients
Characteristics	**Highly sensitive** to identify all the true positives and **high PPV**, cost-effective, minimally invasive.	**Highly specific** to avoid false positives, and **high NPV**. Detect the disease at earliest stage. Cost-effective.	**Highly sensitive and high PPV**. When different therapies available, choose the most suitable for each patient avoiding side effects and costs.	**Highly sensitive and high PPV**. Predict subpopulations that will respond to a specific treatment.	**Sensitive and specific** to ensure continuation of useful therapies and early replacement of ineffective treatments.
Group 0	Assymphomatic patients **suffering from the disease**	Symphomatic patients **suffering from the disease**	Patients with **bad outcome**	Patients that will **NOT respond** to a specific treatment	Patients with **recurrence**
	↑↓	↑↓	↑↓	↑↓	↑↓
Group 1	Assymptomatic patients **NOT suffering from the disease**	Symphomatic patients **NOT suffering from the disease**	Patients with **good outcome**	Patients that will **respond** to a specific treatment	Patients **free of disease** after treatment

Table 8. Clinical uses of biomarkers. Definition, target population, characteristics, and groups of patients to be differentially classified by the biomarkers.

Hence, the goal of cancer biomarker research is the development of robust, sensitive, specific and cost-effective strategies for the above-mentioned clinical uses. Hence, the ideal diagnostic biomarker has the potential to completely discriminate subjects with and without disease. Values of a perfect test which are above a specific cut-off are always indicating the disease, while values below the cut-off exclude the disease. Unfortunately, in practical terms, optimal tests are rare, and often only a partial distinction between diseased and non-diseased patients may be performed. Therefore, the cut-off divides the population of examined subjects with and without the disease in four subgroups considering parameter values of interest: true positive, false positive, true negative, false negative [168].

4.2 STATISTICAL EVALUATION OF A BIOMARKER AND ITS IMPLICATIONS IN EC

Diagnostic tests are an application of the Bayer Theorem to Medicine. They are based on their discriminative ability between two certain conditions of interest such as health and disease. The diagnostic accuracy refers to the degree of agreement between the index test (biomarker) and reference standard (disease). The performance of any novel test needs to be evaluated in terms of its analytical performance, clinical validity, and clinical utility.

Analytical performance refers to the accuracy with which a particular characteristic of interest can be identified by a given laboratory test. The term includes both accuracy and reproducibility within and between laboratories. To achieve this, different analytical requirements are needed such as precision, reliability, limit of detection, and quantitation, linearity range, and specificity.

The **clinical validity** of a biomarker is defined as its ability to accurately identify patients with the targeted pathological state. It is described by different measures: sensitivity and specificity, positive and negative predictive values, likelihood ratio, area under the ROC curve (AUC), Youden's index, and diagnostic odds ratio [169]. The starting point to calculate sensitivity and specificity is the construction of a 2x2 table, called a decision matrix, with the index test results on one side and the reference standard outcome on the other side (Figure 13). The 2x2 matrix will be built based on the following parameters:

- **True positive (TP)** – subjects correctly diagnosed as diseased
- **False positive (FP)** – subjects incorrectly diagnosed as diseased
- **True negative (TN)** – subjects correctly diagnosed as not diseased
- **False negative (FN)** – subjects incorrectly diagnosed as not diseased

That will allow the following to be calculated:

- **Sensitivity**. It is defined as the probability of getting a positive test result in subjects with the disease. Hence, it relates to the potential of a test to diagnose diseased subjects. In EC, low sensitivity would mean that a large proportion of women with EC will be falsely re-assured leading to a delayed diagnostic and poorer survival.

- **Specificity**. It is defined as the probability of a negative test result in a subject without the disease. Hence, it describes the test's ability to recognize subjects without the disease. It is complementary to sensitivity. In EC, <u>a low specificity would mean that a large proportion of women without EC will undergo further unnecessary invasive tests/treatments.</u>

Sensitivity and specificity calculation are not influenced by the prevalence of the disease. Therefore, results can be transferred to other settings with different prevalence of the disease in the population. However, sensitivity and specificity can vary greatly depending on the spectrum of the disease in the studied group. Although sensitivity and specificity are the most provided variables in diagnostic studies, they do not directly apply to many clinical situations. Often, the physician is more interested in the probability that the disease is truly present or absent if the diagnostic test is positive or negative, rather than the probability of a positive test given the presence of the disease (sensitivity). These former, more clinically interesting probabilities are provided by the predictive values:

- **Positive predictive value (PPV)** defines the probability of having the disease in a subject with positive results. Hence, PPV represents the proportion of subjects with positive test result who are correctly diagnosed. In EC, <u>a low PPV has implications for women with a positive test, a large proportion of whom will undergo further unnecessary diagnostic tests or even treatments that are not indicated.</u>
- **Negative predictive value (NPV)** describes the probability of not having a disease in a subject with a negative test result. Hence, NPV represents the proportion of subjects with a negative test result who are correctly diagnosed. In EC, <u>a low NPV has implications for women with a negative test, a large proportion of whom will be falsely reassured and their diagnosis will be delayed</u>.

Contrary to specificity and sensitivity, predictive values are highly dependent on the prevalence of the disease in the population, which may be unknown for a particular target population. Transferring those values to another setting with different prevalence would not be appropriate.

- Likewise, **likelihood ratio (LR)** for a dichotomous test is used to assess the value of performing a diagnostic test and to determine whether a test result usefully changes the probability that a condition exists. However, these tests are not decisive. It is defined as the likelihood of a test result in patients with the disease divided by the likelihood of the test result in patients without the disease. A LR close to 1 (positive likelihood ratio, PLR) means that the test result does not appreciably change the likelihood of disease or the outcome of interest. For a negative test, the further a LR is from 1 (negative likelihood ratio, NLR), the less likely the disease or outcome.
- **Youden's index** is a global measure of a test performance, used for the overall evaluation of discriminative power of a diagnostic procedure and to compare the test to other tests. It is calculated by deducting 1 from the sum of the test's sensitivity and specificity expressed as a part of a whole number. A Youden's index of 0 means poor diagnostic accuracy, while the perfect test would have a Youden's index of 1. Its main disadvantage is that it does not differentiate between the sensitivity

and specificity of the test, leading to incorrect conclusions when it is assessed alone. Youden's index is independent of the disease prevalence but it is affected by the spectrum of the disease.

- **Diagnostic odds ratio (DOR)** can be used in dichotomous and polychotomous tests. DOR of a test is the ratio of the odds of positivity in disease relative to the odds of positivity in non-diseased. The value of DOR ranges from 0 to infinite, where higher values indicate better discriminatory test performance. DOR equal to 1 means that the test does not discriminate between non-diseased and diseased individuals. Values lower than 1 point to more negative tests among the diseased, and thus, improper test interpretation. DOR rises when both sensitivity and specificity are close to the maximum [170].

- Finally, the diagnostic **accuracy** of a test is the <u>proportion of correctly classified patients by the test under evaluation</u>. Hence, the sum of true positive and true negative tests. In addition, it depends on the prevalence of the target disorder in the study group whenever sensitivity and specificity are not equal. Furthermore, its weighs FP and FN findings equally. Although it is sometimes reported as a global assessment of the test, it can perform unsatisfactory estimates in situations such as when the prevalence of the disease substantially deviates from 50%. Thus, it is recommended that authors report more than a single estimate of diagnostic accuracy [171].

	Accuracy	Sensitivity	Specificity	PPV	NPV	PLR	NLR	OR	AUC
Strengths	Intuitive	Does not depend on prevalence	Does not depend on prevalence	Clinical relevance	Clinical relevance	Does not depend on prevalence	Does not depend on prevalence	Does not depend on prevalence; combines SN and SP	Does not depend on prevalence; combines SN and SP
Weaknesses	Depends on prevalence	Applies only to diseased individuals	Applies only to non-diseased individuals	Depends on prevalence	Depends on prevalence	Applies only to positive tests	Applies only to negative tests	Values FP and FN errors equally; not intuitive	Lack of clinical interpretation

Figure 13. The Diagnostic Matrix and the Derivation of Main Diagnostic Parameters. Definition, strengths, and weaknesses. PPV, positive predictive value; NPV, negative predictive value; PLR, positive likelihood ratio; NLR, negative likelihood ratio; OR, odds ratio; AUC, area under the curve; SN, sensitivity; SP, specificity

To graphically plot these measures, the scientific and medical community often use a **receiver operating characteristic (ROC) curve**. It illustrates the diagnostic ability of a binary classifier system

as its discrimination threshold is varied. It is created by plotting the true positive rate (or sensitivity) against the false positive rate (or 1-specificity) at various threshold settings. Often the **AUC value** is calculated as any point on the line segment between two prediction results, often given using 95% interval of confidence. It represents the probability that a classifier (biomarker) will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' ranks higher than 'negative) [172]. Thus, it is a measure of discrimination. Usually, it is considered that a biomarker has satisfactory discriminative power when AUC > 0.75, and excellent power with AUC > 0.9. ROC curves can be used to determine a specific cut-off or as an alternative for clinical discrimination, to propose two cut-offs separated by the "grey zone". The first of these cut-offs is chosen to exclude diagnosis (*i.e.,* privilege specificity), while the second is chosen to include the diagnosis (*i.e.,* privilege sensitivity), both with near certainty. The "grey zone" represents the area where uncertainty exists [173].

Clinical utility of a biomarker is defined as the risks and benefits resulting from test use. It is measured with rates of acceptably, complications and side effects. In EC, a highly invasive test is less acceptable to patients and may lead to complications. Safe, minimally invasive, sensitive, specific, and acceptable by the patients, are the preferred characteristics of the ideal EC diagnostic test criteria.

Other consideration for the development of an optimal EC diagnostic tool is the **cost-effectiveness**. In pharmacoeconomics, it is defined as the ratio of the cost of the intervention of the therapeutic or preventive intervention in comparison to a relevant measure of its effect. In diagnostic biomarkers, this means that measuring the described analyte, obtaining a diagnostic and giving the consequent treatment is less costly than current diagnostic and consequent treatment procedures. Cost-effectiveness is calculated using cost effectiveness ratio, cost of intervention and effect of intervention. In the cost-effective analysis (CEA), effects are measured in terms of years of full health lived, using measures such as quality-adjusted life years or disability-adjusted years.

4.3 PROTEIN BIOMARKERS

The term "molecular biomarker" is a broad concept that encompasses a variety of biological components that can be considered as disease indicators, such as specific cells, proteins, metabolites, hormones, enzymes, molecules, RNAs, miRNAs, genes, and specific mutations, among others. Nevertheless, proteins present striking advantages as biomarkers.

Proteins constantly change because of disease and therefore provide real-time information associated to biological processes for both disease and normal physiology. Additionally, they are the biological endpoint responsible for cellular function, the product of an active gene and, thus, the targets of most current drugs [174]. Many changes occur post-transcriptionally and post-translationally, which cannot be revealed at the DNA and/or RNA level. Consequently, proteins are more diverse due to alternative splicing and post-translational modifications generating several proteoforms from each gene. More than 200 types of post-translational modifications have been described [174]. Humans have an estimated

number of 20,300 genes [175], 114,000 metabolites [176], and about 100,000 mRNAs. In contrast, the human genome may potentially produce up to 1.8 million different protein species [177]. The vast diversity of proteoforms increases the probability of identifying a specific protein or a group of proteins which are associated with a disease. In contrast, the impact of genetic mutations cannot be directly translated from sequencing analysis, especially when genomic variables are localized in non-coding regions.

Protein assays remain a cornerstone of diagnostics, as they allow the measurement of protein biomarkers not only in tissues, but also in easy-accessible biofluids, with high precision. Protein assays are widely available in hospitals, and they can be economically used in a clinical laboratory setting, contrary to Next Generation Sequencing (NGS). Examples of protein assays include immunohistochemistry as well as enzyme-linked immunosorbent assay (ELISA) [178].

Still, the measurement of proteins and protein biomarkers presents several challenges and limitations. A side effect of the enormous number of proteins is the vast dynamic range of concentrations that expands over six to seven orders of magnitude in cells or tissues, and up to 12 orders of magnitude in plasma [179]. Consequently, the detection and quantification of low-abundance proteins is challenging, as they can be masked by high-abundance proteins. Actually, within the 300 most abundant proteins, 23% have been described as biomarkers, whereas only 4% were recorded among the following most abundant 1,200 proteins [180]. Also, as proteins are more dynamic than genes, their activity can change between different special subcellular locations and due to specific protein-protein interactions [174]. Antibody-based assays can suffer from low specificity due to cross-reactivity with interfering substances such as auto-antibodies, and may only measure few proteins at once. In contrast, highly sensitive and specific proteomic techniques such as modern selected reaction monitoring (also termed multiple reaction monitoring) or parallel reaction monitoring assays are not broadly implemented.

4.4 BIOMARKER PIPELINE

The typical biomarker pipeline for a research study consists of three different phases: discovery, verification, and validation. The latter may be extended and/or divided itself into two stages: analytical validation and clinical validation. The implementation of a biomarker in clinics requires a prospective validation to assess its clinical utility and the approval of health regulatory agencies, such as the European Medicines Agency (EMA) in the European Union or the U.S. Food and Drug Administration (FDA) in the US (Figure 14A) [181].

Over the last few decades, biomarker proteomic studies have mainly focused on the **discovery phase**, using analytical methods designed to characterize the proteome. This phase in an untargeted process aimed to identify and quantify as many proteins as possible to pinpoint many differential proteins (from 10s to 100s), usually between two different homogeneous groups (e.g., healthy, and diseased patients). This phase often includes the analysis of a low number of samples per group (5-10 samples) because of the cost, logistics and relatively low throughput of the techniques [182]. Due to the limited number of

samples, large numbers of biomarkers analyzed in parallel, and the limited accuracy of the quantification, discovery studies suffer from high risk of false positive results, particularly in low-abundant proteins [183].

Following discovery, the **verification phase** aims to test the validity of the candidate biomarkers in additional populations and reduce the list of potential biomarkers. In this phase multiple biomarker candidates (30-100), resulting from the analysis of the discovery phase, are analyzed in a larger number of samples (30-100) using an independent cohort of patients. Thus, identities and expression previously seen in the discovery phase can be confirmed, additionally to the assessment of sensitivity and specificity. In this phase, highly multiplexing quantitative techniques are required, such as SRM and/or PRM. Finally, those biomarkers with an increased likelihood to become a clinical assay will be prioritized and assessed in the next phase.

The **validation phase** is the final key step in the biomarker pipeline prior to the clinical evaluation. The main goal of this (analytical) validation phase is to confirm the utility of the biomarkers. It requires high investment and working time and hence, it is usually set to assess only the most promising candidates (3-10) to be tested in a large set of samples (10s-1000s). The set of samples must represent the diversity of the clinical conditions that can be found in the target population, as well as allowing comparison across sample cohorts, analytical platforms, and laboratories.

Finally, a **clinical large-scale evaluation** is needed before FDA and EMA approval and commercialization. In this phase, the final biomarker or panel of biomarkers that showed high accuracy in previous phases, will be quantified in a high number of clinical samples (1000s), preferably in non-invasive body fluids, using a cost-effective and highly multiplexed assay already implemented in clinical practice [167].

Although thousands of single candidate cancer biomarkers have been reported in literature and have been known for several years, very few have been granted **FDA approval.** Some examples are prostate-specific antigen (PSA) for prostate cancer, HE4 protein and its combination with CA125 (ROMA), the protein panel of OVA1, or Overa for ovarian cancer. [184–187]. The reasons associated to the unsuccessful translation of protein biomarkers to the clinic have been broadly discussed [188,189]. Regarding EC, many studies have been performed seeking new biomarkers. In fact, from 1978 to 2017, over 700 proteins have been identified as candidate biomarkers for EC, but none of them have been applied in the clinical setting [190].
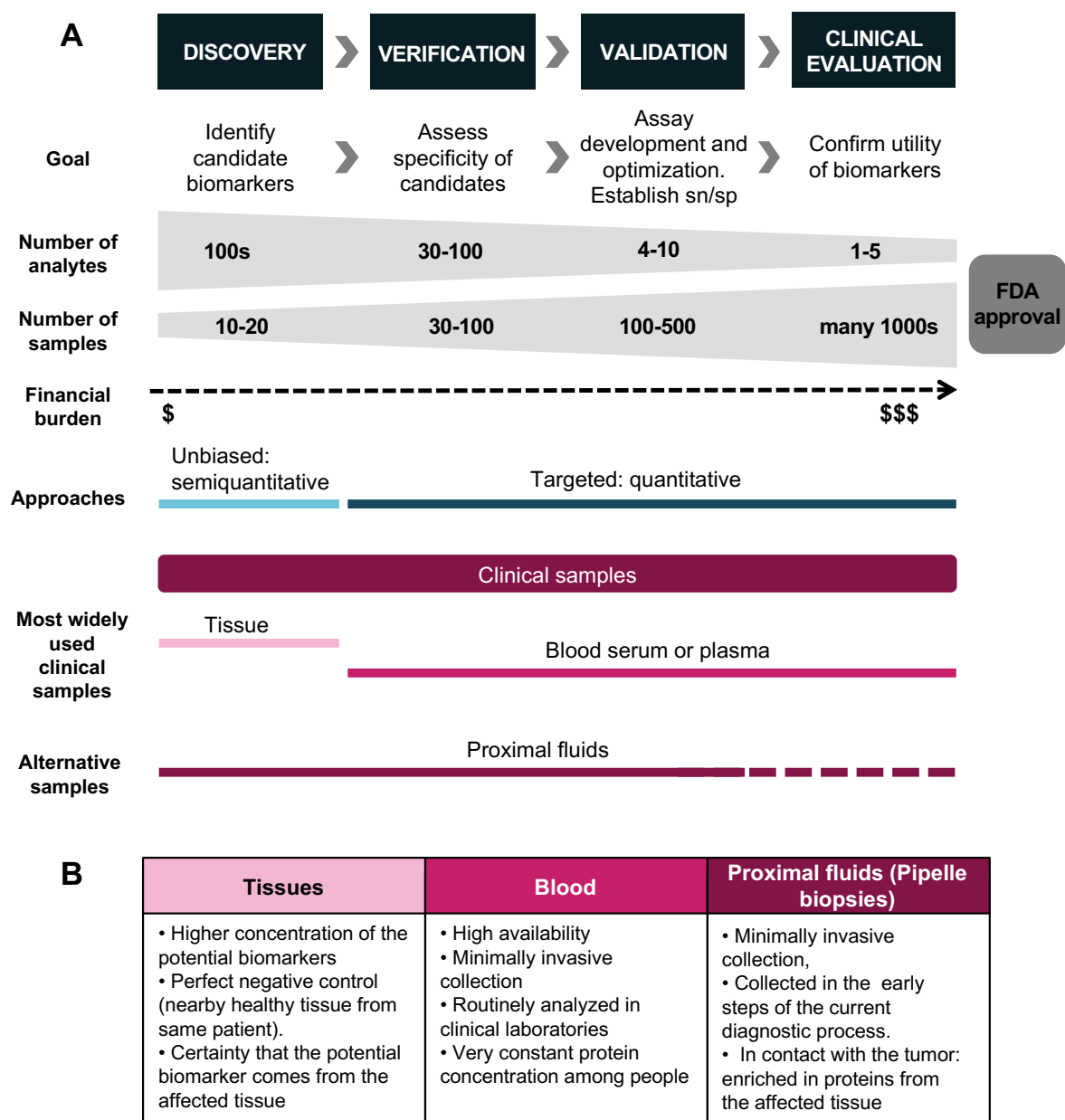
Figure 14. Biomarker pipeline and clinical samples usually employed in each phase. (**A**) Characteristics of the phases of the biomarker pipeline. (**B**) Clinical samples usually employed and advantages that make them appropriate for each phase of the biomarker pipeline.

4.5 CLINICAL SAMPLES: SOURCES OF BIOMARKERS

A variety of clinical samples such as tissue specimens, blood or proximal fluids can be used as source of biomarkers. Although all of them can be used in all phases of the biomarker pipeline and in the clinical practice, their features make them more appropriate for some specific phases (Figure 14B).

4.5.1 Tissue samples

Tissue samples collected in biopsies or surgical specimens are the most used clinical sample to diagnose diseases, as well as the most common sample used in the discovery phase of the biomarker pipeline. Tissue samples are the preferred source of diagnostic and prognostic of biomarkers, as well as patient stratification in terms of response to therapy. Tissue biospecimens are highly concentrated in potential specific biomarkers of the affected tissue compared to other diluted samples such as blood or biofluids. It is clear that the identified proteins are derived from the tumor tissue and therefore, they will represent the diseased tissue microenvironment that can be further studied not only at molecular level but also it can be evaluated by physical interaction (IHC, immunofluorescence, etc.). Despite previous macro-dissection or micro-dissection of the nearby healthy tissue might be required in order to enrich the biomarkers coming specifically from the diseased tissue, the removed healthy tissue can serve as control compared to the diseased tissue. Disadvantages of tissue samples include the need of invasive procedures to collect the sample from the patient, and often a limited quantity of sample [191]. Also, to preserve their nature, tissues require to be properly preserved. They are normally snap-frozen to be stored at -80ºC or they are fixed in formalin and paraffin-embedded for room temperature storage. The first is a more tedious storage but provides a high-quality material for biomarker research studies. This procedure is mainly performed by researchers and clinicians owning private collections and biobanks. The second is the most standardized method of preservation in clinical practice, but requires optimization of protein extraction methods and analysis in biomarker research [192].

4.5.2 Blood: serum or plasma

Blood serum or blood plasma is the preferred sample for clinical use. It is collected rapidly, easily, and is a minimally invasive and cost-effective procedure, and it is broadly accepted by the society. Additionally, it is routinely used and analyzed in clinical laboratories using high-throughput platforms of flow cytometry, ELISA, mass spectrometry, nucleic acid sequencing, etc. It serves as a source of diagnostic and prognostic biomarkers, as well as monitoring biomarkers for different diseases (serial sampling is much easier). Importantly, the molecular landscape of blood has been broadly investigated and to date, we know that 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. Besides, access to normal or control samples for comparative analysis is feasible. However, plasma is one of the most challenging samples to be analyzed, especially by proteomic techniques [180,193,194]. Blood is in direct contact with all body organs; therefore, its content potentially includes leakage or secretion proteins from all tissues, making it more difficult to identify specific disease-related biomarkers. Consequently, it also suffers from intra and inter-patient variability in its composition. Another issue is the wide dynamic concentration range of more than eleven orders of magnitude in protein abundance, which increases the difficulty to detect low-abundant proteins. In detail, very-high abundant proteins such as albumin (35 to 70 mg/ml in blood) masks other potential proteins acting as biomarkers such as cytokines diluted in blood at the pg/ml level [179]. Hence, blood is

a preferred sample for clinical use and it is mainly used for analyzing biomarkers in late phases of the biomarker pipeline, beyond the discovery phase.

4.5.3 Proximal fluids

Proximal fluids are a powerful source of potential biomarkers, presented as an attractive alternative for conducting biomarker research studies, from discovery phases to clinical use. Proximal fluids are in direct contact or close to the side of the disease, and they can be commonly collected by rapid, cost-effective, and non-invasive or minimally invasive procedures. They represent the tissue microenvironment and thus, provide an enriched source of biomarkers compared to blood. However, the use of proximal fluids for biomarker research is also challenging, as their proteomes are less characterized. Each type of proximal fluid presents its own limitations such as small or variable sample volume, low amounts of secreted proteins, variability in sample collection, intra and inter-patient variability in composition, and/or frequent blood contamination [195].

Some examples of these proximal fluids are 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, nipple aspirate fluid for breast disease, among others [196]. In gynecological diseases, different patient-friendly tools have been used to collect proximal fluids, such as the uterine brushing, cervical brush, vaginal tampon, and vaginal swab, among others. Specifically, in EC, fluids from the genital tract are considered proximal fluids, such as vaginal, cervical and uterine fluids (Table 9) [197].

Specifically, **uterine fluids** demonstrated to be an interesting source of biomarkers. This fluid is in direct contact with the tumor in the endometrium and can be obtained from pipelle biopsies, which are biopsies routinely obtained by aspiration from inside the uterine cavity using a Corner Pipelle ®. Pipelle biopsies contain a fluid fraction (uterine fluid) and a cellular fraction, and they are highly representative of the tumor at genomic, transcriptomic and proteomic level [198–200]. The cellular fraction is routinely used for histopathological examination to observe cells contained in the sample and diagnose EC patients, but it has also been exploited at transcriptomic and genomic level though the assessment of molecular alterations. Remarkably, our group identified, verified and validated a five gene qRT-PCR assay that currently improves EC diagnosis through the kit GynEC-Dx, which is commercialized by Reig Jofre from 2014 in Spain [199,201,202]. Strikingly, our group moved from the cellular part to uterine fluids, successfully identifying, verifying and validating protein biomarkers to diagnose EC by mass spectrometry and ELISA assays [200,203].

Other groups have assessed other sampling methods to identify biomarkers to diagnose EC such as **uterine brushings**[204–206], **cervical brush (or pap-smears)** [205,207], **vaginal tampons** [206], or **vaginal swabs** [208–210]. However, none of them have been used in the quest of EC protein biomarkers.

SAMPLING METHOD	BIOMARKER		N patients	SN	SP	ADVANTAGES	DISADVANTAGES
Uterine lavage or pipelle biopsies	• Genomic - mutations, methylated DNA	1	-	-	-	• Proximal to tumour, therefore rich in cancer-relevant biomarkers	• Invasive • Risk of failure to perform the collection of the sample
	• Proteomic - MMP9 and KPYM	2	69 EC; 47 non-EC	94%	87%		
Uterine brushings	• Cytology	3	-	-	-	• Established screening programme in Japan • Fewer inadequate samples than endometrial biopsy • Genomic biomarkers have high sensitivities	• Invasive • Risk of failure to perform the collection of the sample
	• Genomic - mutations, methylated DNA	4	123 EC; 125 non-EC	93%	100%		
Cervical brush	• Cytology	5	6599 EC	45%	-	• Non-invasive	• Insufficiently sensitive for early (pre)cancer detection
	• Genomic - mutations and copy number alterations (PapSEEK)	6	382 EC; 714 non-EC	81%	~99%		
Vaginal tampon	• Genomic - methylated DNA	7	38 EC; 28 non-EC	82% AUC		• Non-invasive • Suitable for self-collection at home	• Uncomfortable for elderly women • Proof-of-principle pilot data only
Vaginal swab	• Metabolomic	8	21 EC; 33 non-EC	75%	80%	• Non-invasive • Suitable for self-collection at home	• Proof-of-principle pilot data only
	• Genomic - methylated DNA	9	30 EC; 73 non-EC	83% AUC			
	• Proteins - CA125	10	37 EC; 106 non-EC	78%	57%		
Urine	• Genomic - microRNA	11	-	-	-	• Non-invasive • Suitable for self-collection at home	• Proof-of-principle pilot data only
	• Metabolomic	12	25 EC; 25 non-EC	89% accuracy			
	• Proteomic - MCM5	13	41 EC; 58 non-EC	88%	76%		
	• Spectroscopic	14	10 EC; 10 non-EC	95%	100%		
Blood	• Genomic - circulating tumor cells, ctDNA	15	92 EC	detection in 75% EC		• Routinely available	• Low concentrations of cancer-specific biomarkers in early cancer
	• Proteins - CA125, HE4	16	2281 EC; 1901 non-EC	65%	91%		
	• Metabolomic	17	88 EC; 80 non-EC	62%-99% accuracy			
	• Spectroscopic	18	342 EC; 242 non-EC	87% (91% EEC only)	78% (81% EEC only)		

Table 9. Novel EC detection tools under research development. *Adapted from Jones et al. 2021* [197]. N patients, number of patients; SN, sensitivity; SP, specificity. Data from [1] Maritschnegg E et al. 2015, [2] Martinez-Garcia et al. 2017, [3] Fambrini M et al. 2014 (review), [4] Wang Y et al. 2018, [5] Frias-Gomez et al. 2020 (review and meta-analysis), [6] Wang Y et al. 2018, [7] Bakkum JN et al. 2015, [8] Cheng SC et al. 2019, [9] Doufekas K et al. 2016, [10] Calid P et al. 2016, [11] Závesky et al. 2015, [12] Shao X et al. 2016, [13] Stockley J et al. 2020 , [14] Paraskevaidi M et al. 2018, [15] Kiss I et al. 2018, [16] Li L et al. 2018 (review and meta-analysis for HE4), [17] Troisi J et al. 2018, [18] Paraskevaidi M et al 2020.

5 PROTEOMICS IN BIOMARKER RESEARCH

Mass spectrometry (MS) is an analytical chemistry technique that helps identify the amount and type of chemicals present in a sample by measuring the mass-to-charge ration and abundance of gas-phase ions. Since MS demonstrated to be able to identify and quantify a large number of proteins and their post-translational modifications, it has been applied to different areas of biomedicine, such as the identification of potential biomarkers [211]. Like genomic and transcriptomic profiling, protein biomarker identification generates large datasets and requires sophisticated statistical analysis, especially for biological samples with complex matrixes. 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 (unlike polymerase chain reaction, PCR) and therefore, sensitivity is an additional challenge. Nonetheless, advances in the methodological, technological, and bioinformatics fields in MS, make proteomics a powerful tool to provide new insights for biomarker research.

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, as well as their application in the clinical setting. In this section, only MS and antibody techniques related to this thesis are described.

5.1 MS-BASED PROTEOMICS

5.1.1 Workflows in mass spectrometry

Proteins can be studied as intact entities by mass spectrometry, an approach called **top-down proteomics** [212]. This approach allows the measurement of all the modifications occurring on the protein, enabling the identification of the precise proteoform. Contrary to this approach, the so-called **bottom-up proteomics** is the broadest approach for protein identification and characterization, and it relies on the identification of peptides generated by proteolytic digestion of the protein mixture. Proteins are enzymatically digested into peptides (mostly by trypsin). The resulting mixture of peptides will then be analyzed in the mass spectrometer, obtaining mass spectra. By computational methods, masses will be compared with a peptide spectral library to identify those peptides assembled into a protein identification. This approach allows highly sensitive identification of the peptides, however tracking post-translational modifications is not always possible.

Different workflows are used in bottom-up proteomics [213]:

- Discovery (or shotgun) proteomics: Data-dependent acquisition (DDA) aimed to achieve unbiased and complete coverage of the proteome.

- Hybrid data-independent acquisition (DIA) is an emerging tool for studying the proteome targeting wide, evenly spaced precursor isolation windows tiled across an *m/z* range of interest. It is aimed to detect and sensitively quantify all peptides covered in a specific spectrum library, generally acquired by DDA.

- Targeted proteomics: This technology aims at the reproducible, sensitive, and streamlined acquisition and quantification of a subset of known peptides of interest. The two most common approaches used in targeted proteomics are Multiple reaction monitoring (MRM) and Parallel reaction monitoring (PRM).

The general workflow of a bottom-up proteomic experiment consists of several steps (Figure 15): from the sample collection and sample preparation to the MS analysis.

5.1.1.1 Sample obtention and preparation

Sample preparation is a crucial step in the identification of protein biomarkers. To capture the most biological significance and obtain meaningful data, the implementation of standard operating procedures (SOPs) by those responsible for sample collection, and clinical data and preanalytical variables registration is essential. Additionally, to avoid sample degradation, the time between sample collection, sample processing and number of freeze-thaw cycles should be minimized, and long-term storage should be done at -80ºC. In order to achieve that, specific experimental design targeting each of those parameters is required. For instance, stability of the samples can be tested by spiking internal standards and monitoring their abundances across freeze-thaw cycles and storage time [182].

MS-based approaches achieve good sensitivity, but highly abundant proteins represent a major challenge for proteomic analysis since the MS data collection is biased towards high-abundant peptides. To solve this issue, especially in blood or specimens containing blood such as uterine fluids, two main methodologies have been proposed to perform before specific MS sample preparation: immunodepletion (e.g., albumin and immunoglobulin G (IgG)) and fractionation by chromatography [214].

To prepare samples for MS analysis, proteins must be soluble. To achieve in-depth protein characterization by MS, protein extraction, solubilization, purification and concentration steps might be necessary except for soluble proteins. Proteins are denatured by heat or by using denaturation reagents such as urea. This will improve the solubility of proteins by disrupting hydrogen bonds and hydrophobic interactions between and within proteins. Then, to reduce the number of protein disulfide bonds, a reducing agent such as dithiothreitol (DTT) is added. Following this, the cysteines are alkylated using an alkylation reagent such as iodoacetamide (IAA) in order to prevent the reformation of disulfide bonds. At this time point, a more efficient proteolysis is possible because proteolytic enzymes have maximum access to cleavage sites within the proteins [178]. Then, proteins are cleaved into peptides by adding a protease or a mixture of proteases with high specificity. Trypsin is the most used enzyme and digestion is performed at a neutral pH and 37ºC. 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. Tryptic peptides are also convenient to separate by reversed-phase liquid chromatography, a separation technique fully compatible and widely use in MS analysis. In addition to trypsin, other proteases such as Lys-C or combination of multiple enzymes can be used to increase protein sequence coverage, as this increases protein cleavage [215]. The digestion process can be stopped by the addition of acid to lower the pH of the sample. The commonly used acid is formic acid, often added at low levels (0.1%), as it ensures the analyte is more basic than the solvent, therefore, facilitating ionization. Finally, sample desalting using solid-phase extraction is performed to remove salts and buffers not compatible with the following steps. After this step, it is recommendable to check buffer pH and to do peptide quantification (generally, only around 40-60% of the peptides will be recovered in comparison to the number of peptides before the digestion, but this might differ from different techniques).
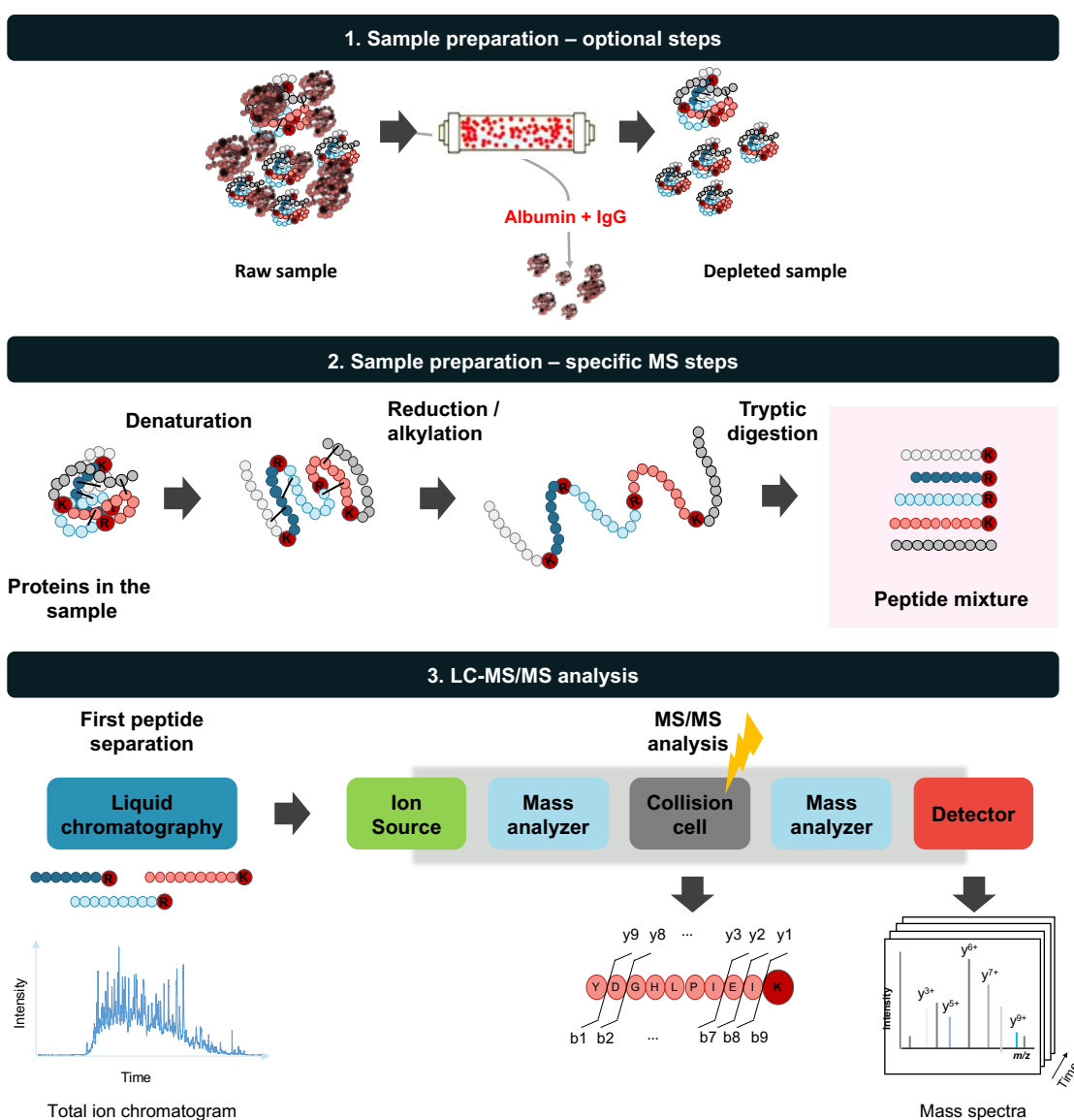
Figure 15. Bottom-up MS-based proteomics workflow. Complexity of the samples can be reduced in a previous optional step of depletion, removing albumin and IgG. Then, proteins are digested with trypsin and resulting peptides are analyzed by reversed-phase liquid chromatography coupled to a mass spectrometry detection.

5.1.1.2 Separation prior to MS analysis

Despite the effort on improving the performance in proteomic analysis for the last decades, no mass spectrometer can capture the whole proteomic landscape in a complex human proteome sample, simply because the peptide heterogeneity exceeds the resolving power of the instruments. The complexity of the sample and, therefore, the simultaneous presence of more than one component in the ion source can result in competition in the ionization process and a subsequent reduction in MS signals, which is known as ion suppression. Ion suppression can be a limitation on the ultimate sensitivity of any MS assay and it can also be caused by salts or other components that might be present in the sample [216].

Separation techniques are routinely employed to improve analytical performance, including accuracy and sensitivity as well as the coverage of the proteome [217]. Nowadays, **liquid chromatography (LC)** is the most widely used technique to physically separate peptides prior to MS analysis, being the reverse phase chromatography the most successful method [218]. LC, or more recently nano high-performance LC (HPLC), separation is based on the interactions of the compounds with the mobile and stationary phases, and the degree of compound separation is related to each compound's affinity for the mobile phase. Generally, reversed-phase HPLC, which separates peptides based on their polarity, has become increasingly popular as the first dimension for tryptic peptide fractionation in a biomarker discovery workflow [219]. The sample is loaded in low organic solvent (mobile phase) and peptides will create hydrophobic interactions with the analytical column (stationary phase) [220].

5.1.1.3 MS analysis

Mass spectrometers are comprised of three main components: ion source, mass analyzer and detector. The mass analyzer measure ions based on the mass-to-charge ratio (m/z). Therefore, peptides must be first ionized and vaporized at the ion source. The resultant ions will be then separated according to their m/z by the mass analyzer and finally detected to generate the mass spectrum, which is a plot of relative ions abundance against m/z values.

Usually, reversed-phase chromatography is coupled to an **electrospray ionization (ESI)** source that will evaporate the solvent and it will generate multicharged peptides which promote efficient fragmentation later in the collision cell. After being ionized and converted to the gas phase, the peptide ions are ready to be transferred to the vacuum of the mass analyzers. However, the number of ions inside the MS is not a direct readout of the number of proteins that were originally in the sample. This is a consequence of the efficiency of the ionization process, which depends on the orders of magnitude for different peptides, as well as the type of ionization used and the fragmentation technique. Another factor is the varying protein conversion efficacy when converting proteins to analyzable peptides due to different digestion efficiencies and/or peptide solubility [221]. Hence, MS is an inherently non-quantitative method.

Mass analyzers include low resolution analyzers such as the quadrupole and the ion trap, and high resolution/accurate mass analyzers such as the time-of-flight (TOF), and the orbitrap analyzers [222]. Current mass spectrometers combine different mass analyzers in tandem and hence, both LC-MS and LC-MS/MS terminologies usually refer to a combination of mass analyzers. Common **tandem mass analyzers** configurations for proteomic studies are the quadrupole time of flight (Q-TOF), 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 [223]. Depending on the mass spectrometer configuration, mass analyzers can be operated at scanning (scan across a range of *m/z* values, resulting in a mass spectrum) or filtering mode (setting to monitor specific *m/z* values).

In the mass spectrometer two levels of MS measurement take place in tandem. Firstly, a mass analyzer measures the *m/z* of peptide molecular ions (MS1). Hence, the MS1 spectrum represents all the intact peptides eluting at a given time, where the height of each peak represents the number of detected ions. In MS1, the high-resolution precursor ion current peaks are extracted as quantitative features. These ions are then fragmented, and secondly, the resulting fragment ions of the *m/z* values are detected in the second mass analyzer (MS2). MS2 or MS/MS spectrum is used to identify the amino acid sequence after the fragmentation of the intact peptide in the collision cell, aiming to assign peptide ID to quantitative features but not involved in determining quantitative values. Finally, the specific fragment ion pattern of each peptide ion, together with its *m/z* value, enables the identification of the peptides present in the sample by its search against a database of peptide sequences generated by *in silico* digestion. The database search step is one of the most critical steps in data analysis, as it identifies the peptide that generated each experimental MS2 mass spectrum. For each spectrum, a list of candidate peptides is extracted from a database as a result of the evaluation of the observed precursor mass and scoring peptide-spectrum matches, and eventually, the highest scored protein is assigned to that precursor [224].

For ~20k human proteins, ~10^6 tryptic peptides can be generated. These complex peptides mixtures contain thousands of peptides with large dynamic range that may be analyzed in several chromatographic runs, however, if they are analyzed in MS instruments of lower resolving power, coeluting peptides with similar *m/z* values frequently overlap. Thus, mass analyzers with high resolving power are required to accurately determine the charge and the mass of each peptide. Despite some related peptide sequences with different amino acid sequence but same chemical composition are not prevented from misidentification [225], high resolution mass analyzers permit the accurate determination of the mass of a peptide. This acts as filter to reduce the number of potential false positive assignments, as it increases certainty of identification. Ideally, all peptides detected by the mass spectrometer should be fragmented. However, due to the large dynamic ranges within the peptide mixture, more abundant peptides are fragmented multiple times, while low-level signals may never be targeted. In this regard, fragmentation cycles of modern ion traps and TOF instruments worked in increasing their speed of

sequencing and have a smarter distribution of the available sequencing capacity, but still some signals might not be targeted. This limits the comprehensiveness of analysis and causes part of the irreproducibility associated with shotgun proteomics, as different subgroups of peptides are picked for sequencing in different analysis of the same sample. Furthermore, the acquisition software controlling the MS does not exclusively target every eluting peptide for fragmentation [226]. Hence, in order to achieve the maximum number of true identifications a compromise between high resolution and low duty cycles is required.

5.1.2 Protein quantification by MS

MS-based proteomics has rapidly evolved over the past decades from a qualitative to a quantitative approach [227]. Although LC-MS is inherently a quantitative platform, the signal is subjected to variations, mainly due to changes in the instrument performance. These can be a consequence of variations in the injection volumes and degradation of the chromatographic column performance regarding the LC component, as well as the contamination or drift in the calibration regarding the MS system. In addition, ionization competition in the ion source can suppress, or sometimes enhance, the signal of an ion species. Currently, there are two main quantitative proteomics methods: 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 labelled with stable isotopes (^{13}C, ^{15}N, deuterium and/or ^{18}O) into internal standards. The isotope-labeled peptides, often called **heavy peptides**, display the same sequences and similar physicochemical properties and chemical reactivity, including chromatographic co-elution, ionization efficiency, and fragmentation pattern to that of the endogenous peptides, also called **light peptides**, but are distinguishable by MS due to their increased mass [228]. Equal amounts of labeled internal standard are added to all samples to be analyzed and, each heavy and endogenous peptides 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 variability factors. The labeled internal standards should be spiked as early as possible during the sample preparation procedure to control the maximum number of steps of the sample processing and decrease the technical variability [227]. The addition of isotope labeled standards enables a more accurate relative or absolute quantification, and provides further confidence of the peptide identification due to the co-elution of the endogenous peptide and the internal standard (used as the reference) as well as the possibility to perform spectral matching [229]. Hence, the main advantage of this quantification approach is the ability of former methods to derive differential protein ratios within a single MS analysis, as well as higher quantitative accuracy and precision.

Isotope-labeled synthetic peptides are peptides labeled with stable, non-radioactive isotopes, and they only differ from the analyte in their isotopic compositions. Therefore, they can be unambiguously

distinguished and easily detected with MS. They are normally added to the digested samples at a known concentration, and therefore the relative ratios of the resulting peptides can be a direct measure of absolute concentration of the peptides in the sample. Accuracy of the measurement will depend on the ability to discriminate from possible interfering signals in the MS spectra, the resolution of the mass analyzer, and the sample complexity.

Stable isotope labeled (SIL) peptides can be chemically synthesized in large scale by manufacturers 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). Ideally, SIL peptides should be (i) prototypic, to avoid peptides with the same sequence from different proteins; (ii) efficiently ionizable, to provide good detection by MS; (iii) without PTMs, to avoid polymorphism; (iv) with a maximum length of 25 amino acids to ensure stability [230]. This approach introduces less complexity to the sample, and it is widely used in targeted studies where the peptides of interest are known upfront, such as the verification and validation of biomarker candidates (Figure 16A).

5.1.2.2 Label-free approaches

Label-free quantitative (LFQ) proteomics provides a straightforward option of large-scale analysis of biological samples. This method has several advantages compared to label-based approaches: it is cost-effective, applicable to any sample and with a simplified and rapid workflow, as it does not need the tedious labeling steps. LFQ approaches allow the comparison of large numbers of experiments with no multiplexing limits. Also, it is a very powerful technique, highly sensitive to MS analysis thus enabling identification of several thousand proteins from complex samples such as bodily fluids and is less susceptible to technical error [231]. However, it provides a less accurate quantification than label-based approaches and difference between the levels of proteins among samples must be high to be significant (greater than two-fold) [232]. LFQ has become more popular in biomedical sciences, especially for the initial discovery phase of the biomarker pipeline (Figure16A).

Two major methods have emerged for LFQ. The first is based on **spectral counting**, which is largely outdated. Spectral counting provides relative quantitation by comparing the number of MS/MS spectra produced by the same protein across multiple LC-MS/MS runs. It counts the number of times that all peptides corresponding to a specific protein are sequenced, therefore, the more abundant the protein, the higher number of tryptic peptides of that protein are available for sequencing, resulting in more MS/MS events, referred to as spectral counts. However, spectral counting lacks precision, accuracy, and reproducibility, especially for low abundant proteins. The second method is based on the **precursor peak/ion intensity** as determined by the extracted ion chromatogram (XIC), which is the plot of intensity versus retention time of a particular *m/z* value. It relies on measuring the three-dimensional space of peptide ion intensity, *m/z,* and chromatographic elution time. In this case, MS fragmentation will determine the identity of each peak, but it will not quantify it. It is more accurate than spectral counting

and more suitable to measure relative abundances of average abundance proteins, because every MS spectra will have a corresponding MS1 chromatographic peak that can be integrated [233].
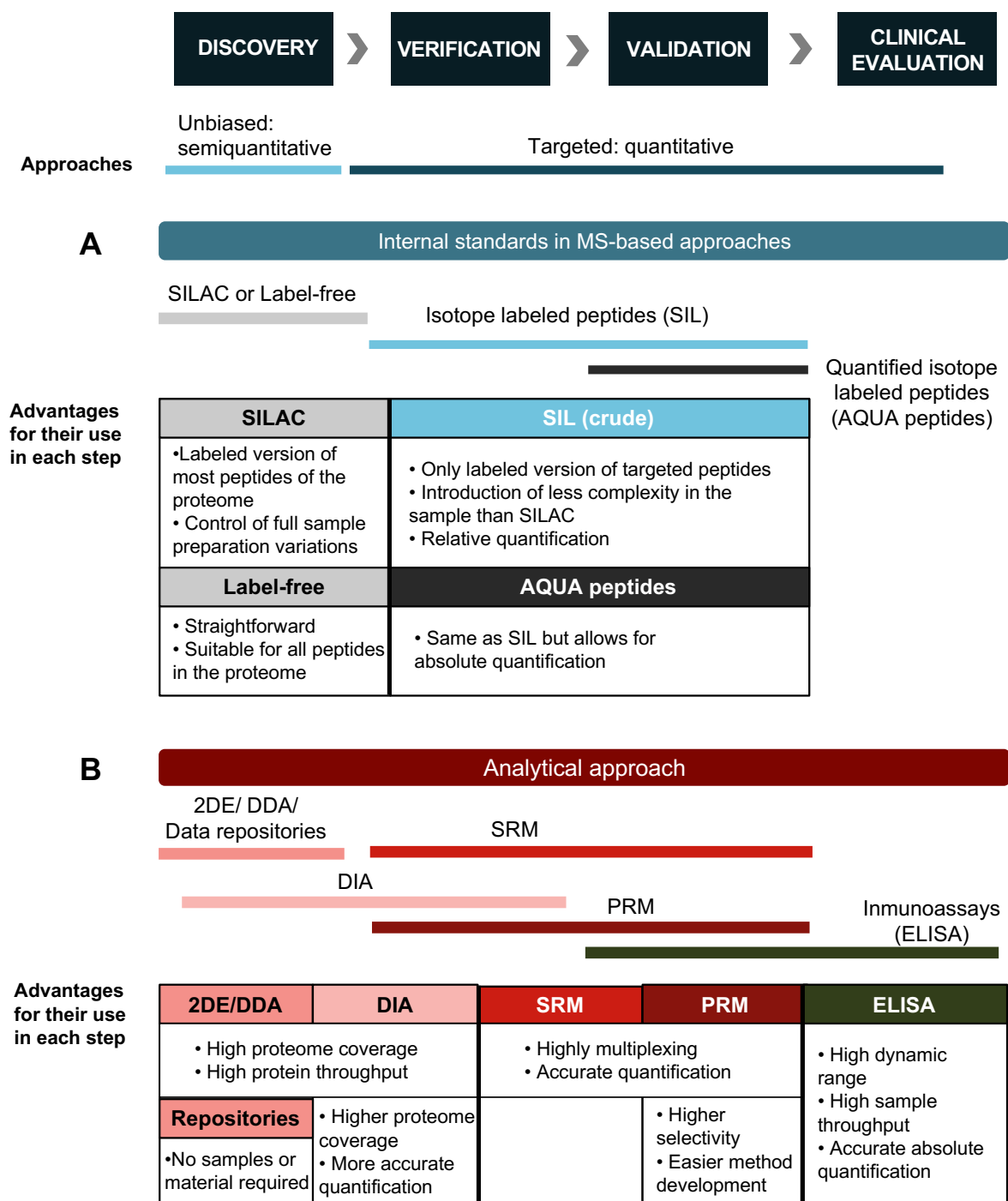
Figure 16. Internal standards and analytical approaches generally employed in each phase of the biomarker pipeline. (**A**) implementation of internal standards in MS-based approaches. (**B**) implementation of analytical approaches in the biomarker pipeline.

5.1.3 Major mass spectrometry acquisition strategies in proteomics

MS plays a key role in understanding and analyzing the proteome of different species and samples through a sensitive, selective, and highly multiplexed analysis. Recent advances in high resolution accurate MS and fast acquisition rates make MS a main player for the discovery but also subsequent verification and validation phases of the biomarker pipeline in complex clinical samples [234]. The selection of the most appropriate MS-based proteomic approach for each specific phase and purpose is essential.

5.1.3.1 Untargeted MS approaches for biomarker discovery

Untargeted MS-based proteomics have been traditionally used in discovery phases of the biomarker pipeline to characterize as much of the proteome as possible. They are high-throughput techniques that aim to identify proteins differentially abundant between two states, rather than focusing on precise quantification.

Different platforms have been developed for this purpose. The first separation technology was the two-dimensional electrophoresis (2DE), based on polyacrylamide gels that separated proteins according to their molecular weight and isoelectric point. They were stained, compared with other gels, collected, digested, and identified by MS. Despite that this approach is still used in order to separate proteins from complex samples such as blood [235], it is outperformed by more powerful, gel-free MS proteomics, known as "shotgun proteomics" [236]. These are **Data-Dependent Acquisition (DDA)** and **Data Independent Acquisition (DIA)**, described as below.

5.1.3.1.1 Data-Dependent Acquisition (DDA)

DDA is the classic liquid chromatography-tandem mass spectrometry (LC-MS/MS) method for protein identification [237]. In this acquisition mode, tryptic peptides are separated by reversed-phase chromatography and subsequently introduced to a mass spectrometer after being protonated by an ESI source. The mass spectrometer can scan and record all the ion species that co-elute at a retention time (precursor-ion spectra) at the MS1 level (MS1 spectra). The instrument alternates between the acquisition of full-scan data and acquisition of fragment-ions spectra, in which the instrument identifies the *top N* (*N* is a preset value in the instrumental method, typically from 4 to 12) ions with the highest intensities as targets for gas-phase fragmentation. Fragmented ions are scanned and recorded as MS2 spectra. Finally, the MS data is processed by searching algorithms (e.g., Mascot, Sequest, MaxQuant Andromeda) that match the experimental MS/MS spectra with the theoretical spectra generated from protein sequence databases (e.g., Uniprot), which are *in silico* digestions imitating cleavages of the protease used during sample preparation [238,239]. 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 though isolation windows of 1-10 m/z and to perform their fragmentation. The performance of DDA has been improved using high-resolution and speed mass spectrometers [226], particularly the time-of-flight [240], and the quadrupole-orbitrap mass analyzers [241].

Data analysis of a DDA dataset is quite straightforward to perform, and a multitude of pipelines and software are available such as MaxQuant [242] or Perseus (for post-data analysis) [243]. In addition, most of MS also provide DDA workflows nowadays. It achieves a high sample throughput with high sensitivity (4 orders of magnitude per MS injection) and high proteome coverage, allowing the quantification of thousands of peptides per MS injection (depending on the sample and method). However, it can be biased toward high-abundant proteins, because the selection of precursor ions is based on intensity. Another consequence of the stochastic selection of precursor ions for fragmentation is the limited reproducibility between sample replicates. Several studies focusing on standard DDA approaches showed that lists of identified peptides between replicates only overlapped by 35-60% [244]. Retrospective targeting is possible only in MS1 level. These features make DDA acquisition a powerful method for protein identification without the need of preliminary knowledge of the sample content prior to acquisition. Hence, DDA is the most used workflow for the biomarker discovery phase (Figure 16B-17).

Quantification in untargeted approaches can be either label-free or with the use of SILAC or isotope labeled tags such as, iTRAQ, TMT and dimethyl labelling.

5.1.3.1.2 Data-Independent Acquisition (DIA)

Thanks to faster computers and sophisticated data-mining algorithms, first Sequential Window Acquisition of all Theoretical Fragment-ion spectra (SWATH) [245], or later data-independent acquisition (DIA), gained popularity. DIA acquisition aims to generate a comprehensive and unbiased map of a proteome by performing MS/MS fragmentation without selecting a particular precursor ion [246–249]. In DIA mode the entire mass spectrum is divided in sequential precursor ion mass windows (typically ranging from 10-50 *m/z* units). Each pre-defined window is first measured by MS1, and then the precursors fragmented simultaneously. Hence, in this mode the mass spectrometer periodically alternates either the collision energy (AIF [250], MSE [251]) or precursor ion isolation window (MSX-DIA [252,253], SWATH [249]) so that each cycle covers a broad range of spectral information. Finally, all fragments are analyzed in MS/MS mode at high resolution. This approach allows to scan all peptides avoiding that some peptides are lost during the acquisition (Figure 17).

DIA data acquisition requires definition of the mass range to cover, precursor isolation window width and, calculated according to the previously two, the number of MS2 scans per cycle. It is a high throughput technique able to quantify thousands of peptides per MS injection, while reaching a high sensitivity (4 orders of magnitude). However, the concomitant fragmentation and MS analysis of all the ions available generate extremely complex MS/MS spectra, which requires peptide query parameters, sophisticated software and large informatic resources. Whilst data is subjected to a classic database

search based on spectral matching, the analysis of DIA data requires *a priori* construction of a fragment-ion spectra for the query (reference spectra library) to recover the original signal from these peptides acquired in DIA data, which is challenging [254]. In any case, new *in silico* approaches are emerging to solve this problem [255], as well as data analysis software tools such as OpenSWATH, MaxDIA or Spectronaut, among others [256]. The search of DIA data with a spectral library is much easier than using a theoretical library and, therefore, DDA is still needed previously to DIA acquisition to construct the spectral library. Another issue is the loss of a direct link between the precursor masses and the MS/MS spectrum, as exists in DDA, and obtained data is usually not available before in-depth analysis and processing of resulting DIA datasets containing complex spectra of fragment ions (retrospective data extraction). However, DIA permits high reproducibility in its experiments and allows a retrospective targeting in both MS1 and MS2 level. 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 kidney [257], lung [258], and prostate [259] among other carcinomas, as well as female-specific cancers such as breast cancer [260] and endometrial cancer [261].
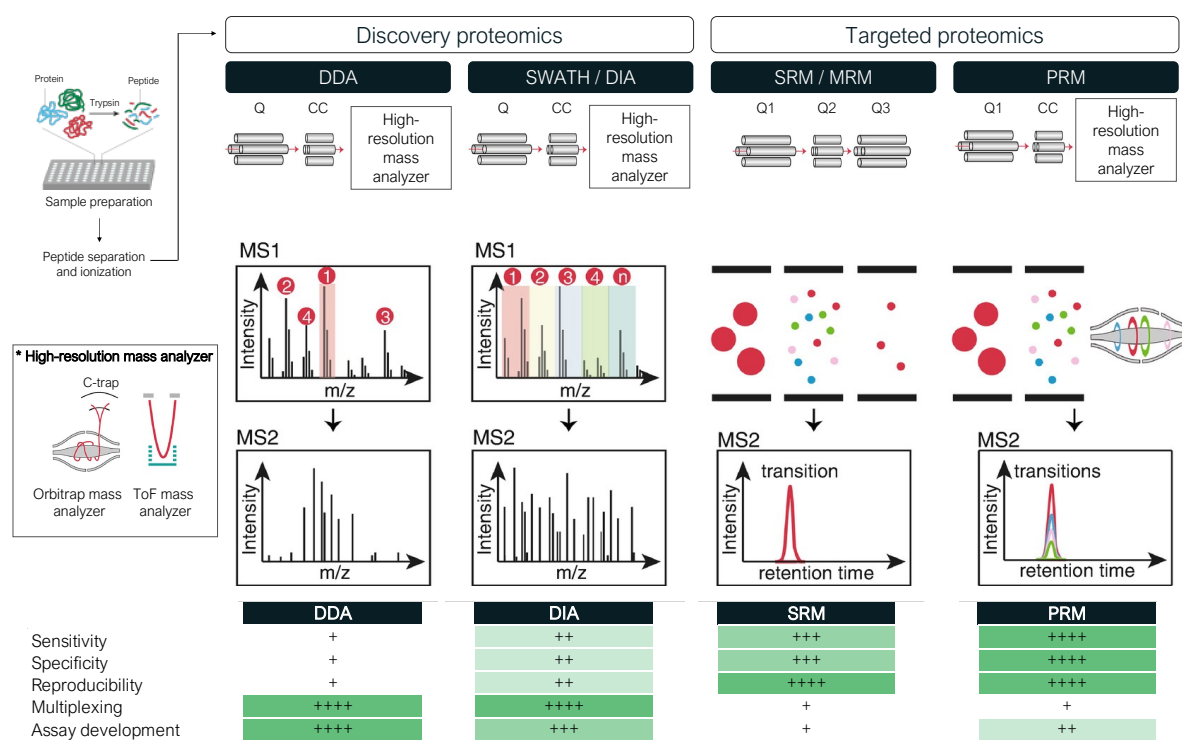
Figure 17. Acquisition methods for discovery and targeted proteomics and comparison of their performances. Q, quadrupole; CC, collision cell. *Adapted from Niu L et al. 2018* [262].

5.1.3.2 Targeted MS approaches for biomarker verification

Antibody-based assays have been the most common methods for protein studies for decades. However, their low throughput limits its application in detection of large number of protein candidates, mostly due

to antibody availability, and it makes them more suitable for final validation phases focused on a restricted list of biomarkers. Highly multiplexed targeted MS-based techniques have gained popularity for verification studies to evaluate numerous biomarkers identified in previous discovery studies and prioritize the most promising candidates to enter in further validation phases. Additionally, the improvement in sample throughput of the MS technologies facilitates their use in larger cohorts of samples in initial validation phases.

In targeted proteomics, proteins of interest are predetermined and known before the actual MS acquisition, and selected peptides are measured by repeated MS/MS events though acquisition methods. To achieve higher multiplexity and more accurate detection, targeted MS methods require optimization for data acquisition. With pre-acquired information, prototypic peptides can be selectively and recursively isolated and fragmented over their chromatographic elution time. The multiple data points acquired from the MS/MS events allow recreation of the chromatographic elution profiles of the targeted peptides, which can be integrated for further quantification at MS2 level. Targeted acquisition relies on the filtering of the precursors through a narrow mass window, typically 1 *m/z,* before MS/MS fragmentation, and this targeted selection significantly improves the selectivity and sensitivity compared to non-targeted experiments. This is done by setting the first quadrupole of a QqQ instrument to the expected precursor ion *m/z* ratio and the third quadrupole to the *m/z* ratio off an abundant fragment ion that is specific for the targeted peptide. To achieve selectivity, the process is multiplexed to several fragments per peptide, known as **multiple reaction monitoring (MRM) or Selected Reaction Monitoring (SRM)** and throughput is increased by multiplexing it to many peptides [263]. The implementation of high-resolution and accurate instruments such as quadrupole orbitraps allowed the implementation of the **parallel reaction monitoring (PRM)**, which uses the entire MS spectrum [264]. Data analysis of these experiments is different compared to DDA or DIA data, since no spectra library is needed, neither theoretical nor experimental. Along with the technique, several software tools have been developed for manual and automated analysis of target spectra extraction and further quantification (*e.g.* Skyline) [265]. While targeted proteomics can be highly multiplexed, they have a limited number of peptides (10-100) that can be quantified per MS experiment. Nonetheless, those will be quantified with high selectivity (4-5 orders of magnitude), since targeted acquisition systematically performs MS/MS events during the acquisition windows which are centered on the elution times of each targeted peptides. Therefore, targeted acquisition does not generate missing data and replicates can be quantitatively compared as they perfectly overlap, which is an advantage compared to DDA acquisition that requires the detection of a precursor in a survey scan to trigger a MS/MS event. Conversely, targeted data acquisition does not permit retrospective targeting.

5.1.3.2.1   Selected Reaction Monitoring (SRM)

The SRM acquisition performed on a triple quadrupole MS, was elected "Method of the year 2012" by Nature Methods, and became the reference technique for MS-based quantitative methods in proteomics to quantify a protein of interest [266].

In a SRM acquisition, the first quadrupole (or mass analyzer) (Q1) is set to filtering mode with a narrow isolation window (e.g., ± 1 *m/z*) to isolate a specific peptide precursor ion of interest. Then, the isolated peptide is fragmented in the second quadrupole (Q2) operating as a collision cell. Finally, one of the resulting fragment ions is selected by the third quadrupole (Q3) set to filter a concrete mass-to-charge ratio. Finally, the fragment will reach the detector to generate a signal (Figure 16B-17) [263]. The double selection of a peptide precursor ion and peptide fragment ion is called a *transition*. The number of transitions that can be monitored within a method depends on the chromatographic peak width and the sampling rate, which should provide a minimum of 8-10 fragment data points per chromatographic peak for accurate and precise quantification. Usually, a number of three to five transitions are recorded per peptide and one to three unique peptides with good quantitative properties are used for protein quantification to improve the confidence of the measurements [267,268]. Related to 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 accurate quantification in a single LC-MS analysis. It is highly specific due to the low probability of finding the same mass, deriving from two different isobaric peptides, and exhibiting the same retention time. Thus, it benefits from low chemical noise resulting in high sensitivity measurements detecting low concentrated proteins. Also, it provides high reproducibility across samples and laboratories [269–271]. The use of internal standards such as SIL peptides or proteins bring several advantages for quantitative analysis including control of the signal variation, strong confidence in peptide identification due to co-elution of the endogenous peptide and internal standard and the similarity of the fragmentation pattern, as well as the possibility to perform absolute quantification if the concentration of the internal standard is known [272]. However, SRM also has some drawbacks as it might require additional depletion steps to detect low abundant proteins [273]. Also, SRM method development is time consuming and static, and is required prior to acquisition. Additionally, due to a limited number of monitored transitions, the presence of interferences may jeopardize the data analysis and require re-analysis of the samples, decreasing the throughput of the technique [234].

5.1.3.2.2   Parallel Reaction Monitoring (PRM)

High-resolution accurate mass spectrometers are being increasingly used in targeted proteomics with the so-called PRM acquisition method, an alternative to SRM. Similarly to SRM, the configuration of the LC-PRM is performed on a triple quadrupole, where the first analyzer (Q1) is set to filtering mode with a narrow isolation window to isolate peptides of interest that will be fragmented in the collision cell (Q2) and the resulting ions are analyzed using a mass analyzer (Q3) set to screening mode. However, 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 though the C-trap to the collision cell to be fragmented. The resulting fragment ions are

transferred back to the C-trap and injected into the orbitrap to be analyzed at high resolution, as compared to QqQ instruments (Figure 16B-17) [274,275].

The acquisition method of PRM allows simultaneously analysis of all fragment ions of a pre-selected peptide of interest, in contrast to SRM that performs sequential acquisition of each fragment from a given peptide. Thus, although both methods require peptide selection prior to MS acquisition, PRM does not need fragment ions to be selected in advance, allowing a more simple and rapid approach. PRM provides both identification and quantification information by extraction of the ion chromatograms (XIC) of the fragment ions of interest and quantification can be performed by integration of the areas of the elution peptides, as for SRM, DDA and DIA.

In PRM the use of chromatographic peaks to retrieve fragment ions increases the resolution, reducing the risk of interferences from the background and provides more sequence information, since the accurate mass improves the confidence on the fragment ions identities [234,276]. Therefore, PRM has more flexibility in selecting the fragment ions *a posteriori,* since it is based on full MS/MS spectra and all potential fragment ions are recorded, instead of only 3-5 that are regularly targeted in SRM. This can also be beneficial to remove ions showing interferences, or re-extract others from data without the need to re-acquire the sample [277]. Also, PRM exhibits a high dynamic range and it is fully compatible with the use of SIL peptides, with all the associated benefits previously described [278].

5.1.4 Public data repositories for assay development

Nowadays, an increasing volume of proteomics data is shared though multiple centralized public repositories [279]. As a high throughput technique, proteomics also generates large datasets of results that are often not fully explored. The availability and accessibility to proteomics data can facilitate the validation and development of protein biomarkers by members of the scientific community for the benefit of patients. Equivalent to genomic and transcriptomic databases such as cBioportal [280,281], there are a variety of available MS-based proteomics databases including knowledge bases (sequences, abundance, isoforms, localization, functionality, PTMs, etc.), peptide knowledge bases (uniqueness, LC retention time, gas-phase fragmentation, ionization, etc.), and LC-MS/MS instrumentation knowledge bases (instrument specifications, parameters, and QC approaches).

Among these repositories, **ProteomeXchange** [282] is recognized as the largest public consortium of proteomics containing raw LC-MS/MS data files, organized spectral datasets, and lists of identified proteins and quantified surrogate peptides with annotated supporting spectral evidence. It can serve as a peptide knowledge base to identify potential biomarkers that could be assessed in a targeted MS assay. Another public consortium is **Proteomics DB** [283] , which was designed to enable cross-dataset comparison of protein abundance, as well as their underlying spectral evidence for peptide identification. From the same research group, **ProteomeTools** [284] offers public peptide knowledge of tryptic peptides covering the entire canonical human proteome.

Complex systems such as cells in the human body require highly accurate tools to improve our understanding of them. However, this generates large amounts of data that is not yet fully analyzed or understood. Therefore, it is essential not only to share the acquired data, but also to boost bioinformatics and biostatistics tools to further analyze this data to enhance our comprehension of the human body and its diseases such as cancer.

5.2 ANTIBODY-BASED PROTEOMICS

Targeted MS-approaches have undoubtedly improved the sensitivity and sample throughput for biomarker discovery. In addition, these platforms have started to be applied during the validation phases of the biomarker pipeline. Nevertheless, antibody-based approaches continue to be the reference method for validation and clinical evaluation steps (Figure 16B).

Antibody-based techniques that are commonly used include Western Blotting, enzyme-linked immunosorbent assays (ELISA) and IHC. They use the basic immunology concept of an antigen binding to its specific antibody. *De novo* generation of highly specific antibodies is expensive and time-consuming (~6 months), which limits the scale of antibody-based approaches. Also, they are highly dependent on the availability and quality of the antibody reagents and appropriate protein standards. Other limitations are the difficulty of multiplexing, the cross-reactivity, and the lack of reproducibility between platforms. Nonetheless, they are still the method of choice in the clinical environment as they are simple and cost effective, when the antibody is available, very sensitive and provide high sample throughput [180].

5.2.1 ELISA for validation and clinical evaluation

ELISA is a sensitive method used to detect the presence and/or quantify the abundance of a determined antigen. There are three main constructs of ELISA: indirect, sandwich and competitive. Sandwich assays tend to be more sensitive and specific, being used in most commercial ELISA kits. The sandwich ELISA is based on two different antibodies: primary antibody (or capture antibody) and the secondary antibody (or detection antibody), both bind to a different specific site on the selected antigen. First, a primary antibody (or capture antibody) is attached on the solid surface of a microtiter well. After blocking the binding sites using bovine serum albumin, the sample is added, and the target antigen can bind to the capture antibody in a specific incubation time. After washing, the detection antibody is added and allowed to bind with the bound antigen. Different reporters can be used (enzyme, fluorophore, or biotin) to detect and quantify the abundance. They can be directly attached to the detection antibody or to a secondary antibody which binds the detection antibody. Finally, a substrate is added and converted by the enzyme into a colorimetric signal that can be quantified. Typically for ELISA, recombinant proteins are used as a positive control in order to generate standard curves, by which samples can be quantified [285].

ELISA is considered as the gold standard for protein detection and quantification, thus it is widely accepted and used for both clinical diagnostics (validation and clinical evaluation) and basic research, as it has several incomparable advantages. Firstly, ELISA is able to measure proteins quantitatively with high sensitivity down to the pg/mL range, and with a wide dynamic range (4 orders of magnitude). Hence, proteins of very low or unknown concentration in the sample can be detected due to the high sensitivity of the technique, because the capture antibody essentially concentrate the low-abundant analyte. This capture antibody enrichment is essential, especially when using body fluids such as plasma or serum presenting a dynamic range of protein abundance over 12 orders of magnitude [179]. Secondly, 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 provides high sample throughput, as the antibodies are mostly conducted in 96- or 384-well microtiter plates. Thirdly, despite its relatively long incubation times and expensive washing, it does not suffer from possible instrument calibration and detector drifts over the time, as happens in MS. Instead, the entire ELISA plate can be read in seconds to a few minutes [286]. Finally, this approach allows the absolute quantification of the protein of interest. Hence, for the reasons mentioned above, ELISA is a cheap (only plate and plate reader required), straightforward and easy to use technique, and therefore, no specific training is required. This not only facilitates the performance, but also the comparison between platforms and laboratories. Consequently, the ELISA immunoassay is broadly used in clinical laboratories, and 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 using clinical samples, especially in biofluids different from plasma and serum [287]. Additionally, ELISAs are not always available for specific antibodies or no robustly correlated results with MS are found due to confounding factors of clinical samples such as proteoform complexity. Thus, *de novo* development of highly specific and reproducible ELISA assays for the proteins of interest is recommended, even though it is a challenging process. The development of a new ELISA assay is an expensive (10,000-100,000$) and time-consuming process (6-24 months), and it can be subjected to large batch variation in antibody specificity, particularly for polyclonals. In sandwich ELISA assays, the mono-specificity of both capture and detection antibodies must be validated using spiked-in proteins that are structurally similar to the target, but no specificity can be difficult to predict, and this approach is not exhaustive. It is therefore of critical importance to keep comprehensive documentation of lot numbers and batch numbers when performing antibody-based proteomic experiments to monitor potential unexpected cause of variation [288]. Another drawback is that immunoassays require 100-200 µl of sample, which is not always available. Also, although the equipment costs are low when using immunoassays, the reagent costs are high and therefore, the overall cost per sample and daily running costs may be quite high depending on the number of proteins that need to be analyzed, and antibody royalties. However, the average cost per sample for a single analyte ELISA is 4-5$, 80% of which comes from the antibody [289]. Although it can be adapted to different platforms, the ELISAs can vary among

them. Finally, multiplexing antibodies for ELISA is still not easy, though Luminex Multi-Analyte Profiling (xMAP) technology is emerging and being optimized which allows up to 500 targets in a single assay [290].
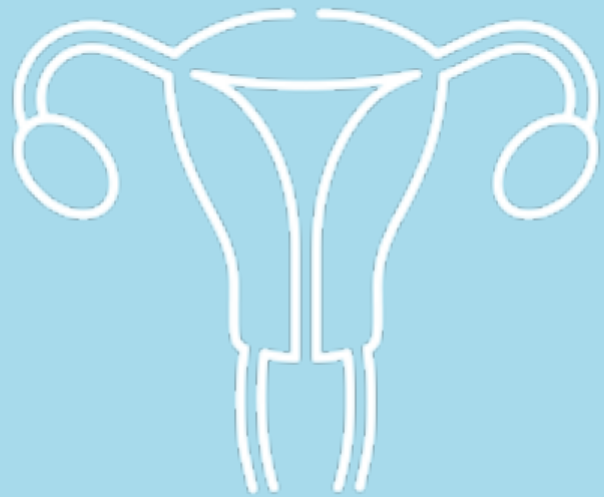
Objectives

BACKGROUND

Endometrial cancer (EC) is the fourth most common cancer in developed countries, and its incidence and mortality rates are increasing annually worldwide. In the United States, it is estimated that about 65,950 new cases of EC will be diagnosed in 2022 and 12,550 deaths are expected. Patients diagnosed at early stages of the disease are associated with a 5-year overall survival rate of 95%. However, nearly 30% of EC patients are diagnosed at an advanced stage of the disease which is associated with a 5-year survival rate of 69% when myometrial invasion is present and/or lymph nodes are affected. In addition, this survival rate decreases to 18% when the tumor has spread beyond the pelvis [17]. Improving early diagnosis is therefore a major benefit to appropriately manage EC and to decrease the mortality associated with the disease.

Current EC diagnosis relies on the follow-up of women at high risk of suffering EC and the presence of AUB. AUB is one of the most common symptoms for gynecologic consultations, affecting up to 11% of postmenopausal women [291] and up to 30% of women during reproductive years [292]. Despite only 8-11% of the postmenopausal women with AUB will ultimately have EC, all of them will be referred to a specialist on the suspicion of cancer, as AUB occurs in over 90% of EC cases [99]. Consequently, a large number of women with benign diseases presenting AUB will be included in the multistep process to rule out EC, which includes invasive, unpleasant and painful tests. This process starts with a pelvic examination and TVUS. However, TVUS lacks in specificity [108]. The first clinical examination is followed by the histopathological examination of an endometrial biopsy. This biopsy can be obtained through minimally invasive methods (*i.e.,* aspiration from the uterine cavity using a Cornier Pipelle, so called uterine aspirates or pipelle biopsies) or invasive methods (i.e., biopsy obtained by hysteroscopy). Although it is recommended to use minimally invasive aspiration [116], the selection is practitioner and center-dependent. In addition, diagnosis though uterine aspiration fails in up to 42% of cases due to technical problems (11%), mainly cervical stenosis, or due to insufficient material found at histology (31%) [133,135]. Thus, these women will have to undergo surgical diagnosis by hysteroscopy, which is associated with the need of pre-operatory tests, the use of anesthesia (or even sedation), nearly 2% of patients complications, such as infections, uterine perforations, among others, and higher healthcare costs [144,145].

In addition to diagnosis, endometrial biopsies should provide information regarding tumor histology and tumor grade to help in the assessment of risk of recurrence in EC patients, and consequently, use this information to guide their surgical treatment. Unfortunately, the limited number of cells available for examination in these biopsies and the high inter-observer variability in the pathological interpretation results in 11.4% and 27% of discordances between pre-operative biopsies and the final hysterectomy specimens in the determination of EC histological subtype and grade, respectively. Remarkably,

clinically relevant downgrading was reported in 26% and upgrading in 8% of patient samples [136], which can lead to either undertreatment or overtreatment.

Recently, the identification of molecular subgroups in EC has led to a shift of paradigm in the management of EC patients. Current EC guidelines have introduced molecular classification as a variable to consider when assessing the pre-operative risk of recurrence in these patients [12,64,83], particularly after it was demonstrated that it could be assessed in pre-operative biopsies [157,158]. However, the pre-operative risk assessment is still not perfect. Thus, the objective measurement of prognostic factors and/or the identification of novel prognostic biomarkers is expected to improve EC diagnosis.

The current process for EC diagnosis and risk classification is still inaccurate. As a result, an important aspect of EC research is directed to the hunt for biomarkers, particularly to provide accurate information at the pre-operative time of diagnosis. In this regard, previous research in the group focused on the identification of highly sensitive and specific biomarkers in pipelle biopsies and uterine fluids to preclude unnecessary hysteroscopies, aiming to early-diagnose EC patients. In the first study performed using pipelle biopsies, Colas E *et al.* proved that pipelle biopsies are reliable surrogates of the primary tumor reporting a gene expression correlation of $R^2 = 0.98$ in a set of 20 genes [199]. Additionally, in a prospective study including 514 women, they defined a diagnostic algorithm composed of 5 genes achieving a sensitivity of 81% and specificity of 96% [201]. Although the study led to GynEC-Dx, a commercialized molecular tool by Reig Jofre SL to diagnose EC, it has important limitations since it is a laboratory developed test, does not provide information on the tumor features, and it is based on the cellularity of the sample so the study only provided results in 72% of cases. In order to utilize the potential of pipelle biopsies, the group moved to proteomics demonstrating that described protein biomarkers in literature could be measured in the fluid of pipelle biopsies (*i.e.,* uterine fluid), and that the levels of 26 proteins from the 52 measured were able to significantly distinguish between EC and non-EC patients [293]. The 52 proteins were further measured in a retrospective study including uterine fluids of 116 women, which allowed the development of a 2-protein panel exhibiting 94% sensitivity and 87% specificity to detect EC cases. In the same study, the discrimination power of these 52 proteins, previously described as diagnostic biomarkers, to discriminate histological subtypes was assessed, resulting in a 3-protein panel with 95% sensitivity and 96% specificity to determine histological subtypes [200]. Therefore, the study suggested that uterine fluids might be a potential gynecological fluid to identify protein biomarkers that could accurately determine tumor features such as histological subtype, histological grade, molecular classification, spread of the tumor or even predict recurrence.

The implementation of a tool based on the assessment of protein biomarkers in uterine fluids to diagnose EC is especially useful in those 31% of cases where insufficient material is found in pipelle biopsies, and consequently, pathologists cannot report any diagnosis. However, the 11% of women presenting cervical stenosis or suffering from intolerable pain will not be diagnosed using pipelle biopsies, as this approach is technically not possible in this setting. Instead, diagnosis in these cases is

achieved by hysteroscopy. In this sense, EC diagnostic research recently shifted to the identification of biomarkers in non-invasive samples to detect EC, and this is expected to not only improve diagnosis, but also screening of EC. The proximity of the endometrial cavity to the low gynecological tract creates an opportunity to collect samples from the cervix or vagina whilst retaining the ability to identify potential EC biomarkers. In recent years, many studies have been conducted in order to test different non-invasive gynecologic sampling methods, such as the use of cervical brushes [205,207], vaginal tampons [206], or vaginal swabs [294–296], to evaluate the feasibility of using those samples as source of EC biomarkers, and to further identify potential biomarkers by the use of different technologies such as genomics, epigenomics, proteomics, and metabolomics. Unfortunately, most of the studies are proof-of principle investigations. In this regard, it is likely that the use of cervical cytologies to detect EC has been the most investigated by different authors. In a recent review and meta-analysis of 45 studies including 6,599 EC patients, it was reported that cervical cytologies have around 45% sensitivity to diagnose EC based on morphological evaluation [207]. An increased sensitivity of vaginal cytologies was observed by O'Flynn H *et al*., when combining those results with uterine cytologies, reporting a combined sensitivity of 92% and specificity of 89% in a cohort of 216 patients (103 EC and 113 non-EC) [297]. Furthermore, Wang *et al*. demonstrated the use of Pap brush (cervical brush) to diagnose EC, achieving 81% sensitivity and 98.6% specificity by amplifying 139 exonic regions within 18 genes, as well as assessing aneuploidy in a cohort of 1,096 patients (382 EC and 714 non-EC) [205]. However, a non-invasive and highly sensitive tool that avoids false negatives and that could be used in screening settings is still missing. Hence, considering that the uterine fluid drains from the uterine cavity to the vagina, we hypothesize that protein EC diagnostic biomarkers could be identified in the fluid of cervical samples and will allow the diagnosis EC patients with high sensitivity.

GENERAL OBJECTIVE

The main goal of this thesis is the identification of highly sensitive, specific, and highly reproducible biomarkers that improve the diagnosis and pre-operative risk assessment of EC tumors.

To achieve that, this thesis followed two different approaches. The first was to investigate the uterine fluid collected from pipelle biopsies as a potential source of prognostic biomarkers in EC, specifically biomarkers related to histological subtype and grade, and recurrence prediction. This approach was conducted as a clinical retrospective study including uterine fluids from 149 patients quantified by data-independent acquisition mass spectrometry. In relation to this, a previously developed spectral library from uterine fluids of 42 EC patients was used as a reference. Those biomarkers are expected to improve the pre-operative risk assessment of EC, specifically by providing an accurate and objective diagnosis.

As a step toward advancing EC diagnosis, the second approach was directed to investigate protein biomarkers in cervical fluids to accurately and non-invasively diagnose EC, as well as to provide prognostic information, *i.e.,* histological subtype and grade. In this thesis, we conducted two clinical retrospective studies, including a discovery study on 60 patients by label-free mass spectrometry, and a verification study on 241 patients by targeted LC-MS/MS PRM analysis. Resulting biomarkers foresee a non-invasive, accurate and objective EC diagnosis, reducing invasive sampling methods and failure rates associated with those procedures.

In the long-term, these approaches are expected to improve the overall management of EC patients, diagnosing more patients at early stages of the disease and, therefore, decreasing the mortality and morbidity associated with EC, while decreasing the burden on the healthcare system.

SPECIFIC OBJECTIVES

To achieve the main objective, the thesis has been divided in the 4 chapters that follow the sequential steps of the patient pathway. Following this, a novel and improved alternative to the current patient pathway is provided at the end. Thus, this thesis consists of the following specific objectives:

Chapter 1: In-silico approach to identify, validate and unveil new applications for prognostic biomarkers in endometrial cancer.

1.1 Compile a list of proteins previously described in the literature and associated with prognostic factors and/or prognosis in EC.

1.2 *In-silico* validation and/or identification of new prognostic applications of the protein biomarkers compiled in objective 1.1 by using the accessible datasets of TCGA and CPTAC studies [44,298].

Chapter 2: Protein biomarkers in uterine liquid biopsy for an objective and accurate pre-operative risk assessment in endometrial cancer

2.1 Create a spectral library for data-independent acquisition (DIA) mass spectrometry quantification of proteins expressed in uterine fluids of women diagnosed with EC.

2.2 Identify protein biomarkers associated to histological subtype and grade, and recurrence prediction of EC by analyzing the fluid from pipelle biopsies (*i.e.,* uterine fluids) from a cohort of 149 women (69 low-grade EEC, 35 high-grade EEC, 16 SEC, 29 other NEEC excluding SEC) using a DIA approach.

2.3 Develop protein panels to achieve the highest diagnostic accuracy for histological subtype, histological grade, and recurrence prediction.

Chapter 3: Potential of five different cervical fluids as a source of protein biomarkers for endometrial cancer diagnosis

3.1 Study of five cervical fluids that were collected through different methods as a source of potential EC diagnostic biomarkers. The proteome of the five cervical fluids collected from 4 patients (2 EC, 2 non-EC) were analyzed and compared among them and against the proteome of the uterine fluids to select the most suitable sampling methods for proteomic analysis.

3.2 Evaluate and select the most promising cervical fluid as potential source of EC protein biomarkers for proteomic studies. A total number of 52 known EC protein biomarkers in uterine fluids (described in Martinez-Garcia E *et al*. 2017 [200]) were analyzed in two different cervical fluids and uterine fluids from 41 patients (22 EC, 19 non-EC) by a targeted MS-approach.

Chapter 4: Protein signatures in cervical fluids permit a highly accurate and non-invasive diagnosis of endometrial cancer

4.1 Identify potential protein biomarkers to diagnose EC patients in cervical fluids during a discovery phase investigation. The study is a clinical retrospective study on cervical fluids from 60 patients (20 EC, 20 non-EC, and 20 presenting some cervical pathology) performed by DDA.

4.2 Perform a verification phase study to develop the method in LC-MS/MS PRM, a targeted acquisition method employed on a high-resolution mass spectrometer, in order to evaluate the performance of the identified and selected potential protein biomarkers in objective

4.3 Determine the diagnostic performance of the previously identified proteins in cervical fluids during the verification phase. The study is a clinical retrospective study on cervical fluids from 241 patients entering the EC diagnostic process (128 EC, 106 non-EC, and 7 hyperplasia with atypia), covering the broad clinical heterogeneity of EC tumors and benign pathologies, and analyzed by LC-MS/MS PRM.

4.4 Assess the potential of the newly discovered and verified protein biomarkers to differentiate between histological EC subtypes and grades.

4.5 Develop protein panels to achieve the highest accuracy to diagnose EC and to discriminate between EC histological subtypes and grades.

4.6 Evaluate the feasibility to use an antibody-based assay, such as ELISA, to measure the best performing diagnostic biomarker and evaluate the correlation of ELISA results with previous mass spectrometry results.

Results
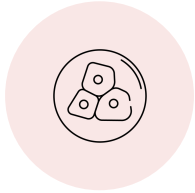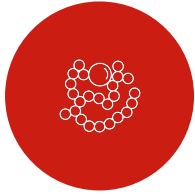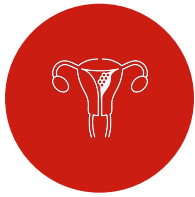
x

CHAPTER 1

In-silico approach to identify, validate and unveil new applications for prognostic biomarkers in endometrial cancer

Adapted from,

"Prognostic biomarkers in endometrial cancer, a systematic review and meta-analysis"

Coll-de la Rubia E, Martinez-Garcia E, Dittmar G, Gil-Moreno A, Cabrera S, Colas E

Journal of Clinical Medicine 2020

AND

"In silico Approach for Validating and Unveiling New Applications for Prognostic Biomarkers of Endometrial Cancer"

Coll-de la Rubia E*, Martinez-Garcia E, Dittmar G, Nazarov PV, Bebia V, Cabrera S, Gil-Moreno A, Colas E*

Cancers 2021

SPECIFIC BACKGROUND

In the age of personalized medicine, the detailed classification of patient subgroups is imperative before and after surgical treatment. In EC this translates into the improvement of stratification tools, including pathologic parameters, imaging techniques and molecular markers. The TCGA molecular classification offers a basis for such an integrative approach. Considering the inaccuracy of current techniques in determining the risk classification of EC, an important part of EC research is directed to the hunt for biomarkers, particularly to provide accurate information at the time of diagnosis.

In Chapter 1, we systematically reviewed the existing literature, compiling an overview of numerous proteins which are associated with EC prognostic factors or that are directly related to recurrence and survival. Among those, we highlight the proteins with an increased potential to become prognostic biomarkers in the clinical setting after prospective validation. Unfortunately, the vast majority of these biomarkers have not been introduced in clinical practice, probably due to a lack of validation in independent studies, reliability or existing evidence. Therefore, we aimed to validate and identify new prognostic applications for the prognostic biomarkers identify by performing a statistical analysis using the accessible datasets of TCGA and CPTAC studies [44,298]. Finally, we discuss possible improvements and new approaches not yet applied to EC biomarker research that could accelerate the identification of clinically relevant biomarkers.

For this thesis, we used this approach to compile proteins associated with prognostic or prognostic factors in EC in any type of clinical sample and define the most promising prognostic biomarker candidates by validating them *in-silico*. The proteins identified here will form the basis for further investigation in Chapter 2 of this thesis, aiming to move the most promising of these candidate biomarkers in the clinical practice.

MATERIALS AND METHODS

Literature Revision

Search strategy, screening and inclusion criteria

Literature searches were performed in MEDLINE from 1991 to February 2020 using the terms "endometrial cancer" or "endometrial neoplasms" or "endometrial carcinoma", "biomarkers" or "markers", and "prognosis or prognostic" or "recurrence" or "survival".

To screen the articles found, we discarded duplicate hits were. Unrelated studies were excluded through careful browsing of the title and/or abstract of each publication. Articles where only the abstract was available were also rejected.

The inclusion criteria were (1) studies including endometrial cancer with an epithelial origin; (2) biomarker studies performed at protein level; (3) prognostic biomarker studies, *i.e.,* studies that identify or validate biomarkers that are associated to EC risk factors, recurrence or survival; (4) studies performed on any biological human sample, but not on cultured cells or animal models; and (5) studies based on the expression of biomarkers. Exclusion criteria were articles (1) not written in English; (2) based on the characterization of one specific EC type; (3) based on response-to-treatment biomarkers; (4) articles performed using less than 10 samples in total; and (5) reviews, meta-analysis, opinion articles or case report studies.

Data extraction and quality assessment

All selected articles were reviewed and data was compiled in a comprehensive database which contained, general information (name of the first author, country, journal, year of publication); number of patients and analytical technique used; association of the described biomarkers with different prognostic factors (histological type, histological grade, FIGO stage, myometrial invasion, lymph node status, LVSI, cervical invasion, TCGA molecular classification, recurrence, risk, overall survival (OS), disease-free survival (DFS), disease-specific survival (DSS), progression-free survival (PFS), and recurrence-free survival (RFS)); and statistical information of the identified biomarkers (e.g. p-value, adjusted p-value, fold-change, area under the receiver operating characteristic curve (AUC), etc.). The guidelines from Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) [299,300] were used to evaluate the quality of studies that were eligible.

Functional enrichment analysis

To investigate the potential functions of the most studied proteins regarding EC prognosis, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID) (https,//david.ncifcrf. gov/home.jsp). The GO terms refer to biological processes (BP) [301]. KEGG was used to identify the most deregulated EC pathways [302].

In-silico validation

Data Source

Expression data profiles of EC patients were collected from the TCGA database though cBioPortal (https,//www.cbioportal.org/, accessed on 27th June 2020). The RNA-Seq expression data of 333 EC patients from the Uterine Corpus endometrial Carcinoma (TCGA, Nature 2013) study (TCGA-RNAseq) was used. CPTAC—Uterine Corpus Endometrial Carcinoma data was obtained from LinkedOmics

database (http,//www.linkedomics.org/login.php, accessed on 10th June 2020). RNA-Seq (CPTAC-RNAseq) and proteomic (CPTAC-proteome) data corresponding to 95 EC patients was used for the analysis. Table 10 details the clinical information of the EC patients used for the analysis of this study.

Features	TCGA RNA-Seq (n=333)	CPTAC RNA-Seq + Proteome (n=95)
Age		
Mean	63.23 ± 10.91	63.19 ± 9.78
Maximum	90	86
Minimum	33	38
Histological type		
Endometrioid	271	83
Serous	52	12
Mixed	10	0
Grade		
Grade 1	79	37
Grade 2	90	38
Grade 3	164	8
FIGO stage		
I	226	71
II	19	8
III	70	13
IV	16	3
NA	2	0
Molecular Classification		
POLEmut	31	7
MMRd	92	25
NSMP	110	43
p53mut	78	20
Overall Survival		
0: Living	282	36
1: Deceased	51	7
Censored		52
Recurrence Free survival		
0: Censored	249	36
1: Progression	70	13
Censored		43

Table 10. Clinical, pathological, and molecular information of the patients. Detailed clinical, pathological, and molecular information of the patients included in this study. Age, means and standard deviations are shown. FIGO stage, Federation of Gynecologists and Obstetricians for staging. POLEmut, POLE ultramutated; MMRd, mismatch repair protein deficient; NSMP, non-specific molecular profile; p53mut, p53 mutated.

Data Processing and Identification of Differentially Expressed Genes (DEGs) and Proteins (DEPs)

The 255 proteins identified were subtracted from the three datasets of expression data, which were separately analyzed using the *limma* and *reportROC* packages of R software. The criteria of false discovery rate (FDR) adjusted p-value < 0.25, | logFC | > 1, and Area Under the ROC Curve (AUC) > 0.75 were applied to screen the DEGs and DEPs (DEG/Ps). The DEG/Ps that were overlapping in the TCGA-RNAseq and [CPTAC-RNAseq OR CPTAC-proteome] EC datasets were named as validated biomarkers.

Survival Analysis

We used the TCGA dataset to identify the potential genes with an impact on OS and RFS. DEGs with FDR < 0.05 and AUC > 0.6 at time points of 12, 24, 36 or 48 months were subsequently used to construct the Cox proportional hazards regression model to predict OS and RFS.

Statistical Analysis

Comparisons between histological types (endometrioid EC vs. non-endometrioid EC) and histological grade (low-grade -G1 and G2- EC vs. high-grade -G3- EC) were performed using T Test. Tukey's honest significance test was used to perform multiple comparisons for FIGO stage (I *vs* II *vs* III *vs* IV) and for molecular classification (POLE ultramutated (POLE) *vs* microsatellite instability (MSI) *vs* Copy number low (CN-LOW) *vs* copy number high (CN-HIGH)). An AUC value for each comparison was also calculated. The univariate Cox proportional hazards regression analyses were completed using the *survival* package of R software. AUC values were calculated using *survivalROC* package of R for OS and RFS at time points 12, 24, 36 and 48 months. Risk scores for each patient were calculated as follows,

$$Risc\ score = \sum_{i=1}^{n} exp_i * coef_i$$

where "*n*" is the number of related prognostic genes, "*exp_i*" is the expression value of the gene *i*, and "*$coef_i$*" is the log hazard ratio (LHR) in univariate Cox regression analysis [303]. Then, the median risk value was used to divide the patients into high and low-risk groups, while the Kaplan–Meier curve was applied to assess the survival difference between the two groups using the log-rank test. Receiver operating characteristic (ROC) curves for assessing the sensitivity and specificity of the prognostic signatures was generated using the *survivalROC* package implemented in R.

Functional Analysis of DEG/Ps, Interactions, and Tractability Information.

We used the Panther database to identify the biological processes and pathways associated with the DEGs. A FDR of < 0.05 was considered as statistically significant [304]. We reported the subcellular location of each protein using UNIPROT [305]. The potential relationship between DEGs encoding proteins was analyzed using the STRING database [306]. Finally, to assess the current available drugs against our DEG/Ps we used the Open Targets Platform [307].

RESULTS

Study Workflow

Our search retrieved 2,507 hits in the initial PubMed Search, that were reduced to 1,557 after the first screening step. Of those, 398 met our criteria and were included in the revision (Figure 18A). From the 398 reviewed studies, a total of 255 protein biomarkers were identified as potential prognostic biomarkers, defined as proteins that are associated to one or more of the known clinical prognostic factors in EC, recurrence, or survival. The complete list of the 255 proteins is shown in Annex 1 (Table

A1.1). Remarkably, only 6% of articles have categorized the recruited patients and/or analyzed their results based on the TCGA classification from 2013 to date.
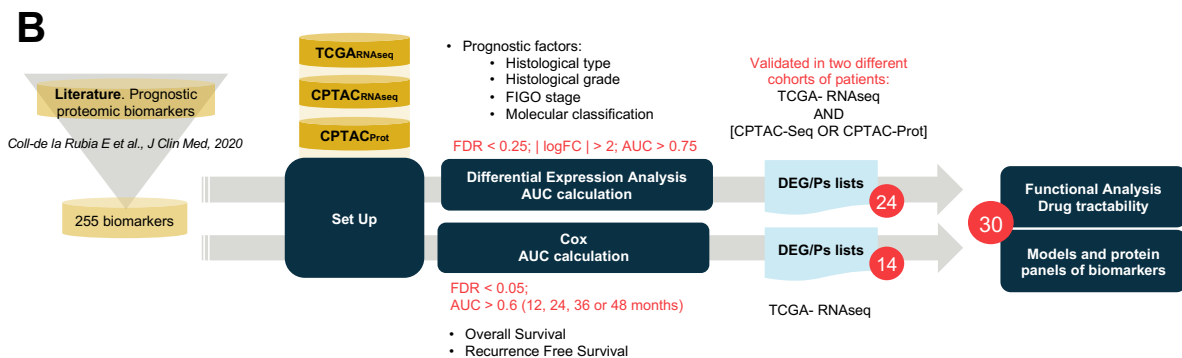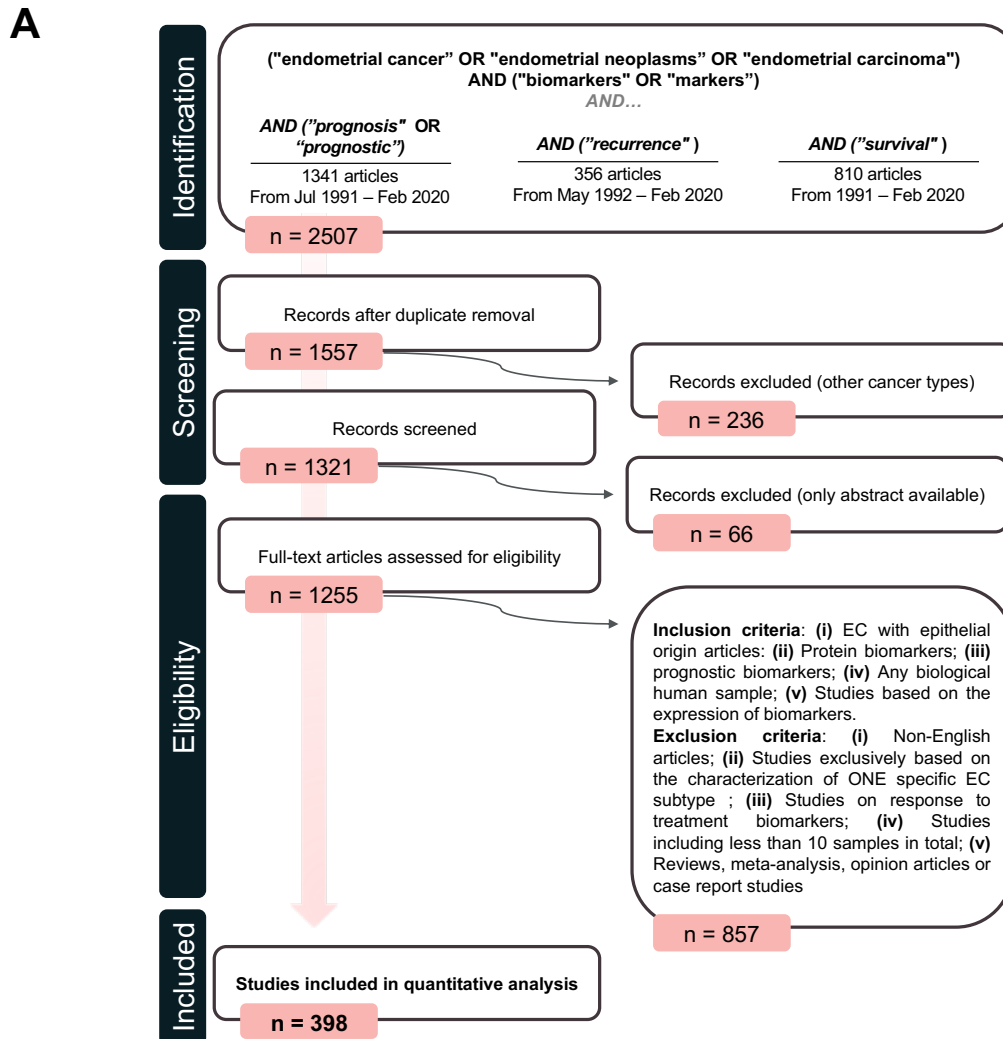
Figure 18. **Workflow of the analysis**. **(A)** Flow diagram depicting the steps followed for the selection of the studies included in this review. **(B)** The 255 biomarkers compiled in our literature revision were assessed in two independent cohorts of patients from the TCGA and CPTAC studies. RNA-Seq data from both cohorts and proteomic data from the CPTAC cohort was used. Different prognostic factors were analyzed (histological type, histological grade, FIGO stage, Molecular classification), as well as overall survival and recurrence free survival. Finally, 30 biomarkers were identified as promising for the stratification of EC tumors.

From the 255 protein biomarkers compiled in this review, only 21% were validated by using either an independent technique, an independent cohort, or in an independent study. Curiously, 60% of the studies were based on the study of a single protein. Regarding the clinical sample used, 79% of the studies were performed in tissue specimens, followed by 16% of studies that used serum samples. Other sources were plasma, imprint smears, peritoneal cytology or uterine aspirates. Additionally, 6 studies were performed in tissue and validated in serum samples and 5 articles did it *viceversa*.

The exhaustive literature revision compiled 255 protein prognostic biomarkers associated with one or more known clinical prognostic EC factors, including histological type, tumor grade, depth of myometrial invasion, cervical involvement, tumor size, LVSI, LNS, tumor spread; as well as the molecular classification, recurrence and/or survival. Most of these biomarkers have not been validated in independent studies, jeopardizing their implementation in the clinical practice. To validate the potential of those proteins as EC prognostic biomarkers and unveil novel potential prognostic associations, we performed an *in-silico* analysis of those proteins in 428 EC patients belonging to the CPTAC and TCGA studies. The workflow of this study is depictured in Figure 18B. Briefly, the 255 biomarkers were assessed in the RNA-Seq data of the TCGA and CPTAC datasets, in addition to the proteomic data of the CPTAC dataset. The most relevant prognostic factors, which are histological type, histological grade, FIGO stage, and molecular classification, were analyzed using a differential expression analysis and the calculation of the area under the ROC curve (AUC) values. Additionally, OS and RFS were assessed using a Cox analysis. Statistically significant biomarkers were identified for each parameter and dataset, and those that appeared significant in at least two datasets were considered as validated biomarkers. Among the 255 potential prognostic biomarkers, only 30 biomarkers were validated, and those were further studied using functional analysis and drug tractability studies.

Literature Revision

Prognostic protein biomarkers in EC

As shown in Table 11, most biomarkers identified in this systematic review were associated with histological grade, FIGO stage and OS, with more than 100 biomarkers described for each of these parameters. Other biomarkers were associated with lymph node status, histological type, myometrial invasion, LVSI, DFS, recurrence, DSS, PFS, risk, RFS, cervical invasion and the TCGA subgroups. The vast majority of biomarkers are related to more than one of the above-mentioned parameters, indicating that they provide relevant prognostic information but are not specifically linked to one feature in particular. In fact, those that were associated with a specific parameter (in bold in Table 11) generally corresponded to those biomarkers that have been scarcely studied. Thus, further research needs to be performed to understand whether they are truly significant as prognostic factors and specific of that parameter in particular or might be also related to other parameters.

Prognostic factor	Number validated proteins	Gene names of the validated proteins
Histological grade	158	**ADM** AKT1 ALB ALCAM ANXA2 AR ARID1A ASRGL1 ATAD2 ATP7B AURKB BECN1 BGN BIRC5 BUB1 C1QBP CCNA2 CCNB1 CCND1 CCNE1 CD44 CD68 **CD82** CDC20 CDCP1 CDH1 CDH3 **CDK4** CDKN2A CFLAR CHI3LI CIP2A CLU CRIP1 CRP **CTNNA1** CTNNB1 CTNND1 CYP19A1 DPP10 **EDIL3** EGFR EIF3A EIF4EBP1 ELF1 ENG EPHA2 **ERAP1** ERBB2 ESR1 **ESR2 ETV5 F2R** FASN FGA FGFR2 FHIT FOXA1 FSCN1 GDF15 GGT1 GLP1R H3C1 **HDAC1** HDAC2 HIF1A HIF1AN HLA-A HLA-E HSF1 HSPA5 INHA INHBA INHBB JUN KDM1A KHDRBS1 L1CAM LGALS3 LNPEP LRP1 MACC1 MAD2L1 MAPK1 MCM6 MCM7 MIB1 MIF MKI67 MME MMP12 **MMP7** MT1A MT2A MUC1 MUC16 MYCBP NAMPT **NFKB1** NOTCH2 NOTCH3 NSD2 NTSR1 OVGP1 PARK7 PAX8 PBXIP1 PCNA PDPN **PECAM1** PGR PKM PLAU **PLK1** PODXL PRL PROM1 PTCH1 PTEN PTGS2 PTPN6 REPP86 RICTOR ROBO1 RRBP1 S100A4 SAA1 SATB1 SCYL3-p SERPINE1 **SHH** SKP2 SLC2A1 **SLC7A5** SLIT1 SNAI1 SNAI2 SNCG **SOX2** SPAG9 SPP1 ST6GALNAC1 STMN1 TFGA TIMP2 TMEFF2 **TNC** TNFAIP8 TOP2A TP53 TRA2B TRIM44 TWIST1 TYMP UCHL1 VCAN WFDC2 WNT7A
FIGO stage	140	ALB ALCAM ANXA2 ASRGL1 ATAD2 BCL2 BGN BIRC5 BSG C1QBP CCNA2 CCNB1 CCND1 CD274 CD34 CD68 CD82 CDC20 CDCP1 CDH1 CDH3 CDK20 CDKN1A CDKN2A CFLAR CHI3LI CLU CRIP1 CRP CSF1 CTNNB1 CTSB **CXCL10** CYP19A1 DPP10 EGFR ELF1 ENG **EPCAM** EPHA2 ERBB2 ESR1 **FAS FASLG** FGA FGF1 FGFR2 FHIT FOXA1 GDF15 GH1 GINS3 GPER1 GSTO1 HDAC2 HDGF HIF1A HLA-E **HMGA1** HSPA5 IDO1 **IL31** IL33 INHA IRAK1 KDM1A KHDRBS1 L1CAM LGALS3 LNPEP LRP1 MACC1 MAPK1 MCM6 MIF MKI67 MME MMP7 MMP9 MMR MSI1 MTOR MUC1 MUC16 MYCBP NAMPT NOTCH1 NOTCH2 NOTCH3 NSD2 OVGP1 PARK7 PBXIP1 PDPN PGR **PIGF** PKM PLAU PODXL PRL PROM1 PTEN PTGS2 **PTK2** PTP4A3 PTTG1 RBMX REPP86 RICTOR ROBO1 RRBP1 S100A4 S100A8 SAA1 SATB1 SCYL3 SERPINE1 SLIT1 SNAI1 SNAI2 SNCG SPAG9 SPP1 ST6GALNAC6 STC2 STMN1 TMEFF2 TNFAIP8 TOP2A TP53 TRIM44 UCHL1 VCAN VEGFA VIM WFDC2
Overall survival	138	ALCAM ALDH1A1 ANXA2 AR ARID1A ATAD2 ATP7B AURKB BCL2 BECN1 BGN BIRC5 **BNIP3** BSG C1QBP CBR1 CCND1 CCNE1 CD274 CD3E CD44 CD68 CD82 CD8A CDCP1 CDH1 CDH3 CDKN1A CDKN2A CHI3L1 CHI3LI **CRHR1** CRHR2 CRIP1 CRP CTNNB1 CTNND1 CTSB CTSD CXCL12 CXCL8 DFFB EGFR ENG ERBB2 ESR1 FGA FGF1 FHIT **FLT1 FLT4 FOSL1** GDF15 GH1 GINS3 GLP1R GPER1 GSTO1 H3C1 HDAC2 HDGF HTATIP2 IDO1 INHA IRAK1 JAG1 JUN KDM1A **KDR** KHDRBS1 KLRG1 L1CAM LCN2 **LDHA** LGALS3 MAD2L1 MCM6 MCM7 MIB1 MKI67 MMP2 MMR(*) **MS4A1** MSI1 MTOR MUC1 MUC16 NAMPT NME1 NOTCH1 NOTCH2 NOTCH3 NSD2 NTSR1 PAEP PBXIP1 PDCD1 PDPN PGR PRL PROM1 PTEN PTPN6 PTTG1 REPP86 RICTOR RRBP1 S100A1 S100A4 S100A8 SATB1 SERPINE1 SKP2 **SLC2A4** SNAI1 SNAI2 SPINT1 SPINT2 SPP1 TFGA TIMP2 TMEFF2 TNFAIP8 **TNP** TOP2A TP53 TRIM27 TRIM44 TWIST1 TYMP UCHL1 VCAN VEGFA WFDC2 WNT7A **YTHDC1** YWHAQ
Lymph node status	85	ALCAM ALDH1A1 ANXA2 AR ASRGL1 ATAD2 BCL2 BGN BSG BUB1 C1QBP CD34 CD44 CD68 CDH1 CFLAR CHI3LI CSF1 CTNNB1 DFFB ERBB2 ESR1 **ESR3** FGF1 FSCN1 GDF15 GINS3 GSTO1 HIF1AN IDO1 IL33 IRAK1 JUN KDM1A L1CAM LGALS3 LNPEP MACC1 MAD2L1 MDK MIF MKI67 MTOR MUC1 MUC16 MYCBP NAMPT NME1 NSD2 PBXIP1 PDPN PGR PLAU PRL PTEN PTGS2 PTP4A3 RBMX **RETN** ROBO1 RRBP1 S100A4 SATB1 SERPINE1 SKP2 SLIT1 SPAG9 SPINT1 SPINT2 ST6GALNAC1 ST6GALNAC6 STMN1 TCGA TMEFF2 TNFAIP8 TP53 TRA2B TRIM44 TWIST1 TYMP UCHL1 VCAN VEGFA WFDC2 WNT7A
Histological type	82	**ABCB1** AKT1 ALCAM ARID1A ASRGL1 ATAD2 BIRC5 BSG **CAPG CAPS** CCNA2 CCND1 CCNE1 CD151 CD274 CD34 CD44 CD68 CD8A CDH1 CDH3 CDKN2A CHI3L1 CIP2A CRIP1 CRP CTNNB1 DPP10 EGFR EIF4EBP1 ERBB2 ESR1 FGF1 FLR1 FOLR1 FOXA1 GDF15 H3C1 HIF1A HSF1 INHBA INHBB L1CAM LCN2 **LGALS1** LGALS3 LRP1 **LTF** MAPK1 MKI67 MMR MTOR MUC1 MUC16 NOTCH2 NOTCH3 PARK7 PAX8 PCNA PGR **PIGR** PLAU PTEN PTPN6 ROBO1 S100A4 **SCGB2A1** SLIT1 SNAI1 SNAI2 SNCG SPAG9 SPP1 STMN1 TIMP2 TP53 TYMP **URI1** VEGFA VIM WFDC2
Myometrial invasion	82	ANXA2 ARID1A ASRGL1 ATAD2 BCL2 BECN1 BGN BMI1 BSG C1QBP CD34 CD44 CD68 CDH1 CDH3 CFLAR CRIP1 CSF1 CTNNB1 CTSB DFFB EGFR **EHMT2** EIF4EBP1 ELF1 EPHA2 ERBB2 ESR1 FBXW7 FGFR2 FHIT GDF15 GPER1 GSTO1 HIF1AN HLA-E IDO1 IL33 INHA INHBA IRAK1 JUN KDM1A KHDRBS1 L1CAM LGALS3 LNPEP MIF MKI67 MSI1 MUC1 MUC16 NAMPT NOTCH1 NSD2 OVGP1 PARK7 PBXIP1 PGR PTEN PTGS2 REPP86 ROBO1 RRBP1 SATB1 SLC2A1 SLIT1 SNAI1 SNAI2 SNCG SPP1 STMN1 TNFAIP8 TOP2A TP53 TRIM44 TWIST1 TYMP UCHL1 VEGFA WFDC2 WNT7A
Lymphovascular space invasion	63	ALCAM AR ASRGL1 ATAD2 BCL2 BIRC5 BMI1 BSG C1QBP CCNA2 CCNE1 CD34 CD44 CD68 CDH1 CDH3 CDK20 CTNNB1 DFFB ERBB2 ESR1 FBXW7 GSTO1 H3C1 HLA-E IDO1 KDM1A KRT1 L1CAM LNPEP MACC1 MKI67 MME MSI1 MUC1 MUC16 MYCBP NAMPT NOTCH1 NSD2 PGR PROM1 PTGS2 RICTOR ROBO1 SATB1 SKP2 SLC2A1 SNAI1 SNCG SPINT1 SPINT2 ST6GALNAC STMN1 TNFAIP8 TP53 TYMP UCHL1 VCAN VEGFA WFDC2 WNT7A
Disease-free survival	60	ALB ALDH1A1 ATAD2 ATP7B BCL2 BIRC5 BSG C1QBP CDCP1 CDKN2A CRHR1 CRP CTSB CTSD CXCL8 DFFB ERBB2 ESR1 FASN **FOLH1** HLA-E HTATIP2 JAG1 KDM1A KRT1 L1CAM **LNM** LNPEP MKI67 MMR MUC1 MUC16 NOTCH2 NOTCH3 NSD2 NUCB2 PBXIP1 PCNA PGR PROM1 PTCH1 PTEN PTGS2 RRBP1 S100A4 SATB1 SERPINE1 SPINT1 SPINT2 SPP1 TNFAIP8 TP53 TRIM44 UCHL1 VCAN VEGFA WFDC2 WNT7A **WT1** YWHAQ
Recurrence	42	ALCAM ALDH1A1 ATAD2 BCL2 BIRC5 C1QBP CD3E CD68 CDCP1 CHI3L1 CRP DFFB ERBB2 ESR1 FGA FGF1 GDF15 GINS3 H3C1 KDM1A KLRG1 L1CAM MDK MIB1 MMR(*) MUC1 MUC16 NSD2 PGR PLAU SATB1 SERPINE1 SLIT1 SNAI2 SNCG ST6GALNAC6 STC2 STMN1 TNFAIP8 TP53 WFDC2 WNT7A
Disease-specific survival	34	ALCAM ASRGL1 BCL2 CCNB1 CD151 CD274 CD8A CDH1 CXCL8 EIF3A EIF4EBP1 EPHA2 ERBB2 ESR1 FOXA1 GDF15 HLA-A HLA-E HSF1 INHA INHBA MIB1 MKI67 MUC16 NUCB2 PAEP PCNA PDCD1 PGR RBMX STMN1 TIMP2 TP53 VEGFA
Progression-free survival	31	ALB CBR1 CD274 CD68 CD8A CDH1 CHI3L1 CHI3LI ERBB2 ESR1 FOLR1 GGT1 GLP1R IDO1 INHA L1CAM MCM6 MKI67 MMR MUC1 MUC16 NTSR1 PGR RBMX S100A4 SNAI1 SNAI2 TP53 TRIM27 TYMP WFDC2
Risk	29	BCL2 BIRC5 CCNA2 CCND1 CD44 CDKN1A **CDKN1B** CDKN2A CSF1 ELF1 ERBB2 FHIT H3C1 HSPA5 L1CAM MUC1 MUC16 PARK7 PKM S100A1 SLC2A1 ST6GALNAC1 STMN1 TFGA TMEFF2 TP53 **TP63** VIM WFDC2
Recurrence-free survival	28	ALCAM BCL2 CD151 CDH1 CDKN1A CXCL12 ERBB2 FGF1 GH1 GINS3 L1CAM MIB1 MMP2 MMR(*) MUC16 NAMPT PGR PLAU PRL PTGS2 REPP86 SERPINE1 SNAI2 ST6GALNAC1 STC2 STMN1 TP53 WFDC2
Metastasis	13	CCND1 CRP ESR1 FGA FGF1 IL33 L1CAM MUC1 MUC16 PGR RBMX ST6GALNAC6 UCHL1
Cervical invasion	10	BSG CD44 CIP2A CTSB ESR1 MUC16 PGR SPP1 UCHL1 WFDC2
TCGA	3	L1CAM MCM6 PTEN

Table 11. List of proteins associated to each prognostic factor. Proteins linked only to one specific parameter are highlighted in bold.

Most of the candidate prognostic biomarkers are involved in common biological processes, such as cellular processes, biological regulation, metabolic processes, response to stimulus, cellular component organization or biogenesis and signaling; and, proteins that are part of the basic structural and functional units of the cell. In order to become promising biomarkers, the potential of any specific protein should be validated in different cohorts, and if possible, by independent groups, and in prospective studies. According to our findings, 14 proteins (ERBB2, CDH1, PTEN, TP53, MMR proteins, ESR1, PGR, MKI67, L1CAM, MUC16, and WFDC2) have been extensively studied, i.e., in more than 5 independent studies (Figure 19A). All the 14 proteins have been studied by IHC in primary tissue specimens. Notably, MKI67 and PTEN were also validated in tissue samples from imprint smears [308–310], as well as CDH1, which was analyzed in uterine aspirates using mass spectrometry-based approaches [200]. MUC16 and WFDC2 have been extensively studied in serum samples by antibody-based techniques such as ELISA or chemiluminescence techniques [115,311–313] and L1CAM was also validated in serum but just in one study

[314]. Importantly, all these proteins have also been described as diagnostic biomarkers and are main drivers of the oncogenic pathways related to EC (Figure 19B-C) [117,190].
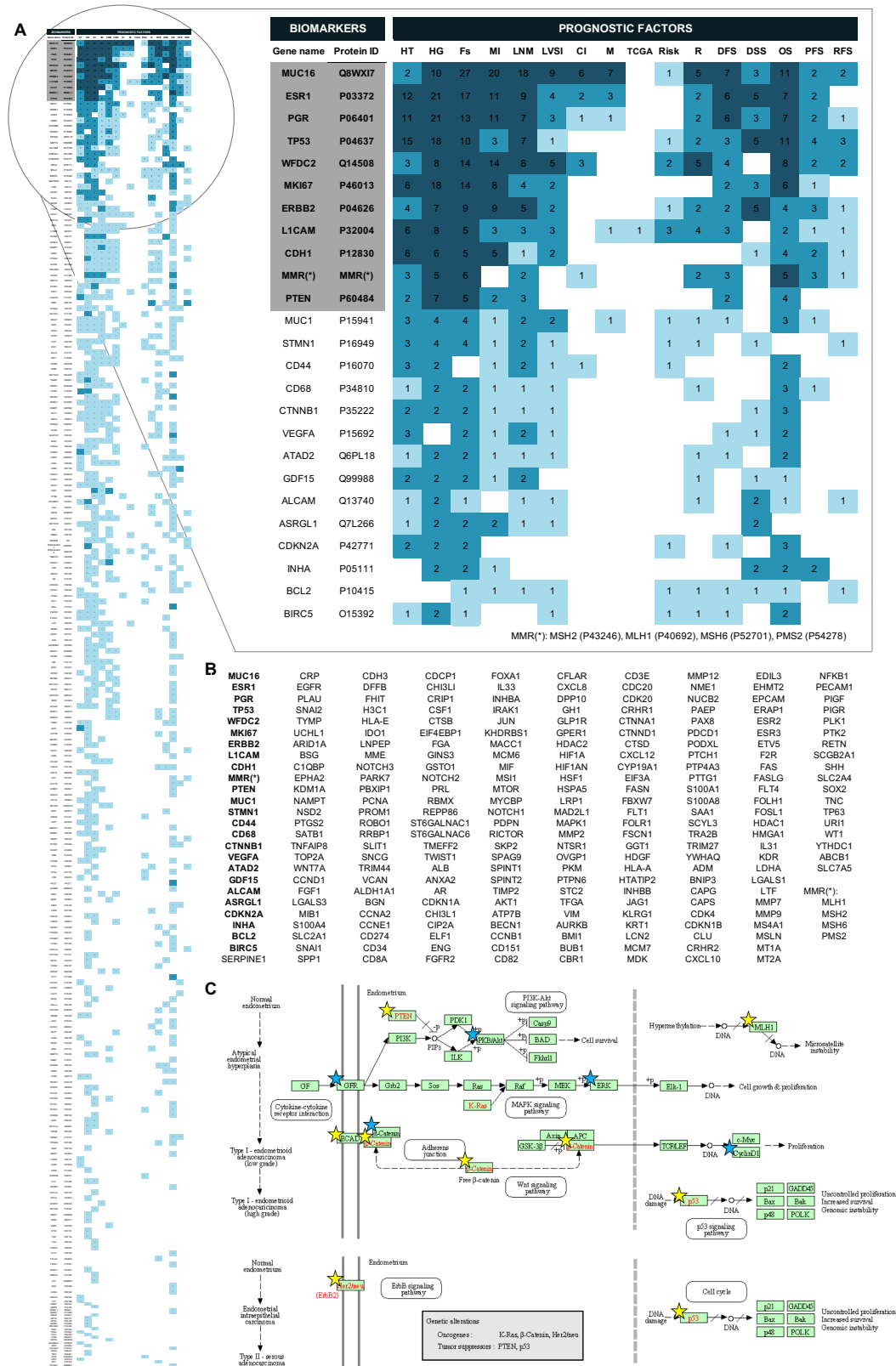
Figure 19. Overview of the validated biomarkers. (A) Full perspective of all validated biomarkers, i) dark red when protein was validated 5 or more times for that parameter; ii) red, when protein was validated in more than

one study; iii) light red, protein validated in one study. List of the top-25 most studied proteins as prognostic factor biomarkers is zoomed in. For each protein, the number of studies in which it was validated appears. (**B**) List of proteins validated, at least in one study, for one of the considered parameters. Ordered regarding the number of independent studies where they were validated. In bold, the top-25 most studied proteins. (**C**) EC disease Pathway Map obtained from the KEGG DISEASE Database [302,315,316]. The proteins from the 14 most studied proteins list are highlighted by yellow stars, while the top-25 are highlighted by blue stars. HT, histological type; HG, histological grade; Fs, FIGO stage; MI, myometrial invasion; LNS, lymph node status; LVSI, lymphovascular space invasion; CI, cervical invasion; M, metastasis; TCGA, TCGA classification; R, recurrence; DFS, disease free survival; DSS, disease specific survival; OS, overall survival; PFS, progression free survival; RFS, recurrence free survival.

In-silico validation

Statistical analysis from two big public repository datasets of EC patients resulted in 30 validated biomarkers, in which we encountered significant proteins for one or multiple prognostic parameters (Table 12). TPX2 is protein associated with a major number of prognostic parameters, including histological type and grade, molecular classification, and OS and RFS. This protein was previously studied in *Jiang T et al., 2018* where it was associated with bad prognosis [317], and recently its prognostic value in EC was further demonstrated [318,319].

Gene Name	Uniprot ID	Protein Name	HT	HG	FIGO	MC	OS	RFS
ASRGL1	Q7L266	Isoaspartyl peptidase/L-asparaginase		HG			OS	
ATAD2	Q6PL18	ATPase family AAA domain-containing protein 2		HG		MC		RFS
BUB1	O43683	Mitotic checkpoint serine/threonine-protein kinase BUB1	HT	HG				RFS
CAPG	P40121	Macrophage-capping protein				MC		
CCNE1	P24864	G1/S-specific cyclin-E1	HT	HG		MC		
CDC20	Q12834	Cell division cycle protein 20 homolog	HT	HG				
CDKN1A	P38936	Cyclin-dependent kinase inhibitor 1	HT					
CDKN2A	P42771	Cyclin-dependent kinase inhibitor 2A	HT			MC		
ERBB2	P04626	Receptor tyrosine-protein kinase erbB-2	HT					
ESR1	P03372	Estrogen receptor	HT			MC	OS	
FASN	P49327	Fatty acid synthase					OS	
HDGF	P51858	Hepatoma-derived growth factor					OS	
HMGA1	P17096	High mobility group protein	HT			MC		
L1CAM	P32004	Neural cell adhesion molecule L1	HT			MC		
MACC1	Q6ZN28	Metastasis-associated in colon cancer protein 1					OS	
MCM6	Q14566	DNA replication licensing factor MCM6					OS	
MCM7	P33993	DNA replication licensing factor MCM7					OS	
MSH2	P43246	DNA mismatch repair protein Msh2				MC	OS	
MSH6	P52701	DNA mismatch repair protein Msh6				MC	OS	RFS
PAX8	Q06710	Paired box protein Pax-8	HT			MC		
PGR	P06401	Progesterone receptor			FIGO			
PIGR	P01833	Polymeric immunoglobulin receptor	HT					
PTK2	Q05397	Focal adhesion kinase 1					OS	
S100A1	P23297	Protein S100-A1	HT			MC		
SCGB2A1	O75556	Mammaglobin-B				MC		
TMEFF2	Q9UIK5	Tomoregulin-2 (TR-2)				MC		
TPX2	Q9ULW0	Targeting protein for Xklp2	HT	HG		MC	OS	RFS
TRA2B	P62995	Transformer-2 protein homolog beta				MC		RFS
UCHL1	P09936	Ubiquitin carboxyl-terminal hydrolase isozyme L1	HT			MC		
VIM	P08670	Vimentin	HT					

Table 12. Summary for the 30 validated biomarkers in the analysis. Prognostic factors or prognostic value described in *Our literature revision* for each of these biomarkers [57] are highlighted in grey. Additionally, the specific prognostic factor or prognostic value validated in our analysis are indicated for each protein. The color green represents prognostic factors found in both, literature revision and statistical analysis, whereas the color blue represents new prognostic features not described before in literature. HT, histological type; HG, histological grade;

FIGO, FIGO stage; MC, molecular classification defined by The Cancer Genome Atlas Network; OS, overall survival; RFS, recurrence-free survival.

Validated Prognostic Biomarkers in EC

Regarding histological type, a total of 20, 36 and 18 genes/proteins were differentially expressed between endometrioid (EEC) and non-endometrioid (NEEC) histologies in the TCGA-RNAseq, CPTAC-RNAseq, and CPTAC-prot datasets, respectively (Figure 20A). However, only 15 were validated in at least two datasets (Figure 20B). From those, eight biomarkers were previously described—CCNE1, CDKN2A, ERBB2, ESR1, L1CAM, PAX8, PIGR, VIM—and seven are newly associated to the histological type, BUB1, CDC20, CDKN1A, HMGA1, S100A1, TPX2, UCHL1. The most confident biomarker proteins (present in the three datasets) were ERBB2, L1CAM, PIGR and TPX2. Among the non-endometrioid types analyzed in the TCGA dataset, 62 NEEC cases were included, 10 cases of which were mixed type and 52 serous carcinomas. As seen in Annex 1, Figure A1.1, while some of the biomarkers behaved similarly between the NEEC, others such as CDC20, CDKN1A, ERBB2, HMGA1, L1CAM and PAX8 were expressed in mixed tumors as a mixture between SEC and EEC, similarly to the nature of these tumors.

Regarding histological grade, six proteins in the TCGA-RNAseq cohort, 21 proteins in the CPTAC-RNAseq cohort, and seven proteins in the CPTAC-prot cohort showed different protein abundances between low-grade (grade 1 and grade 2) and high-grade (grade 3) EC (Figure 20C). From those, six proteins were validated in two datasets, and were also previously described in other studies, ASRGL1, ATAD2, CDC20, and TPX2 (Figure 20D). Thus, highlighting their importance and the need of further validation of those in further prospective clinical studies.

Regarding FIGO stage, three and 17 proteins showed differential abundances in the TCGA and CPTAC cohorts of patients, respectively, 10 of which were previously described in the literature. However, only PGR was validated in two cohorts of patients and its performance was limited to the comparison between stage I and II (Annex 1, Figure A1.2).

Finally, this study permitted identifying a significant number of biomarkers that allow separating between different groups of the molecular classification. A total of 16 proteins (ATAD2, CAPG, CCNE1, CDKN2A, ESR1, HMGA1, L1CAM, MSH2, MSH6, PAX8, S100A1, SCGB2A1, TMEFF2, TPX2, TRA2B, UCHL1) were confirmed, i.e., were statistically significant in at least two datasets (Annex 1, Figure A1.2). Specifically, L1CAM, ATAD2, CAPG, CNNE1, CDKN2A, ESR1, HMGA1, MSH2, MSH6, PAX8, S100A, TPX2, TRA2B, UCHL1, showed capacity to distinguish between CN-LOW vs. CN-HIGH; L1CAM, CDKN2A, HMGA1, MSH6, TMEFF2, UCHL1 between MSI vs. CN-HIGH; and L1CAM and CDKN2A to differentiate between POLE vs. CN-HIGH subgroups. Remarkably, L1CAM seems to be the most informative biomarker to differentiate the CN-HIGH from the other molecular groups.
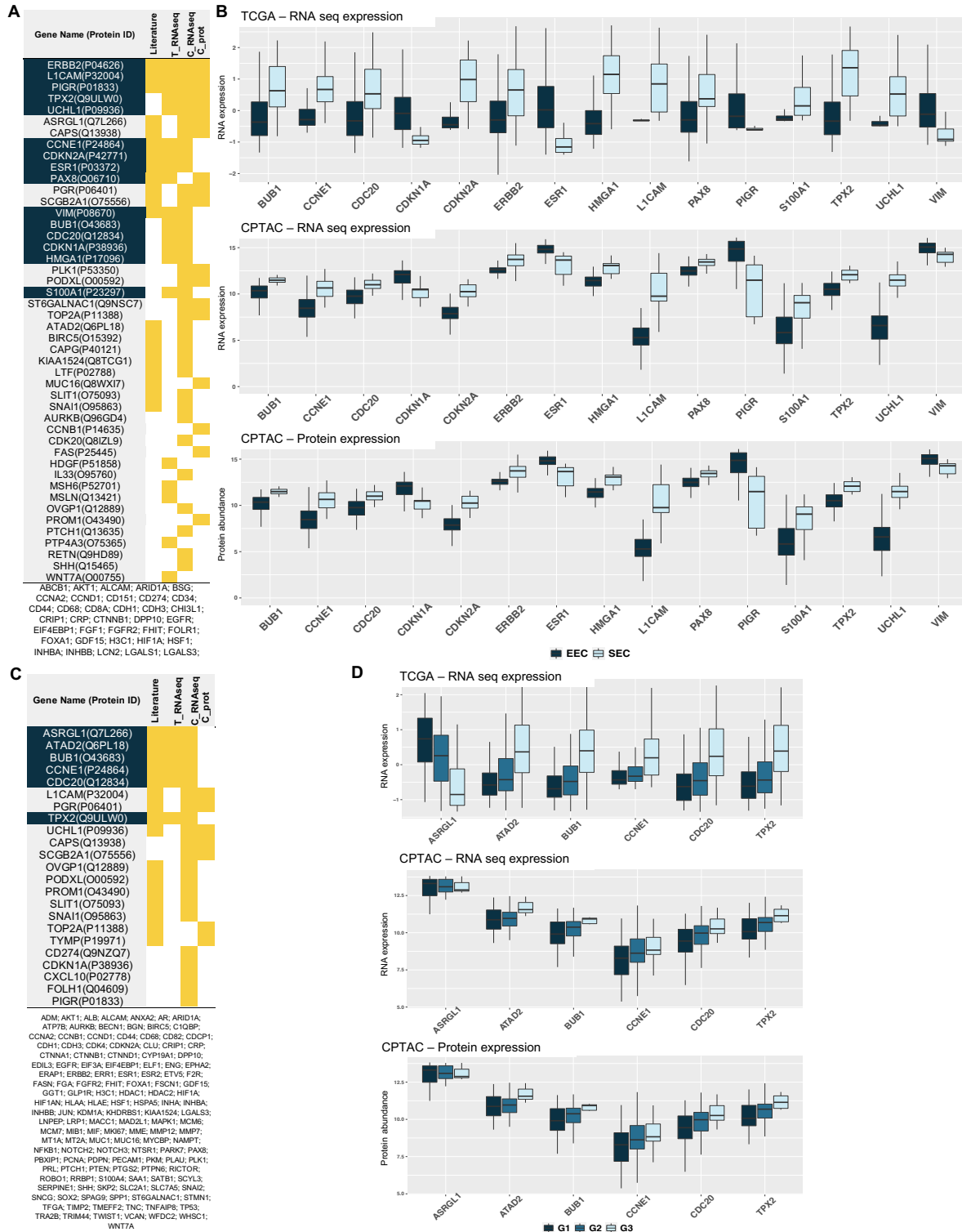
Figure 20. Biomarkers related to histological type and histological grade. (**A, C**) Table of the proteins that were found differentially expressed between, (**A**) endometrioid (EEC) and non-endometrioid (NEEC), and (**C**) G1G2 and G3, respectively, in any of the tested cohorts. Highlighted in yellow, the specific cohort in which that protein was found to be differentially expressed between histologies and/or grades. Proteins highlighted in blue are those validated in more than one cohort, and therefore, the ones that we considered as validated biomarkers. (**B**) Boxplots showing the expression of the 15 validated biomarkers for histological type in each cohort of patients, TCGA RNA-Seq data, CPTAC RNA-Seq data, and CPTAC proteomic data, respectively. (**D**) Boxplots showing the expression of the six validated biomarkers for histological grade in each cohort of patients, TCGA RNA-Seq data, CPTAC RNA-

Seq data, and CPTAC proteomic data, respectively. Literature, literature revision from our literature search [57]; T_RNAseq, RNA-Seq data of the TCGA's cohort; C_RNAseq, RNA-Seq data of the CPTAC's cohort; C_prot, proteomic data of the CPTAC's cohort.

Validated Biomarkers Associated to OS and RFS in EC

Among the 255 biomarkers studied, a total of 11 and 5 genes presented significant correlation with survival rates and recurrence, respectively (FDR < 0.05 and AUC > 0.6 at time points of 12, 24, 36, and/or 48 months) in two datasets (Figure 21A-B). In particular, the genes with significant association to OS were ASRGL1, ESR1, FASN, HDGF, MACC1, MCM6, MCM7, MSH2, MSH6, PTK2, and TPX2, while the ones associated with RFS were ATAD2, BUB1, MSH6, TPX2, and TRA2B. Among them, ASRGL1 and ESR1 were characterized as low risk, while the remaining 14 were categorized as high-risk genes. An increased prediction for OS and RFS was achieved by the development of biomarker panels. This was performed by using all the significant genes associated to OS and RFS in a Cox regression analysis. We used the prognostic signature to calculate a risk score (see equation in materials and methods) for each patient, while the median value was used to divide the patients into a high-risk (n=166), and low-risk groups (n = 167) (Figure 21A-B). An 11-protein model reached an AUC of 0.827 to predict OS, while a 5-protein model of RFS reached an AUC of 0.712, both at 48 months' time.

A Model OS using 11 genes
ASRGL1,ESR1,FASN,HDGF,MACC1,MCM6,MCM7,MSH2,MSH6,PTK2,TPX2

Overall survival		12 months	24 months	36 monts	48 months
Best performing individual biomarkers	PTK2	0.627	0.579	0.594	0.578
	MCM7	0.586	0.640	0.600	0.573
	TPX2	0.586	0.612	0.633	0.627
	TPX2	0.586	0.612	0.633	0.627
	Model 11 proteins	0.731	0.744	0.773	0.827

B Model RFS using 5 genes
ATAD2,BUB1,MSH6,TPX2,TRA2B

Recurrence-Free survival		12 months	24 months	36 monts	48 months
Best performing individual biomarkers	SCYL3	0.622	0.643	0.561	0.553
	SCYL3	0.622	0.643	0.561	0.553
	ATAD2	0.561	0.567	0.631	0.644
	MSH6	0.477	0.550	0.630	0.652
	Model 5 proteins	0.562	0.613	0.688	0.712

Figure 21. Biomarkers related to overall survival and recurrence free survival and model performance. Best performing individual biomarkers are shown for (**A**) overall survival (OS) and (**B**) recurrence free survival (RFS) for a period of 12, 24, 36, and 48 months, respectively. Additionally, models for both were performed. Regarding prediction of OS, a model of 11 proteins was used, (**A**) while a model of 5 proteins was used to predict RFS (**B**).

Biological significance of the validated biomarkers

Gene ontology and KEGG enrichment analysis were used to explore the biological functions of the initial 255 proteins, as well as the subgroup of the 30 proteins that were validated in this study. All of them were significantly associated to the following biological processes, cellular processes, biological regulation, response to stimulus, signaling, and metabolic processes. The set of validated genes had an increased association with reproductive processes (Figure 22A). Additionally, we studied the function of those 30 validated biomarkers. While BUB1, CCNE1, CDC20, CDKN1A, CDKN2A, MCM6, MCM7 and TPX2, played a role in cell-cycle, proteins such as ERBB2, ESR1, L1CAM, PGR, PIGR, PTK2 are molecules involved in the activation cascade that enhance the tumor growth. Interestingly, 14 out of the 30 proteins are described as proteins related to the epithelial-mesenchymal transition (EMT), crucial for the malignant progression [320] (Annex 1, Figure A1.3).

Regarding the pathways analysis, the 255 proteins had a balanced association with multiple pathways, in contrast to the 30 validated biomarkers having a strong relation with the gonadotropin-releasing hormone receptor pathway, p53 pathway, p53 pathway feedback loops 2, interleukin signaling pathway, cell cycle, and integrin signaling pathway, as shown in Figure 22B. Additionally, while downregulated validated biomarkers were associated to interleukin signaling pathway among others, upregulated biomarkers were associated to other pathways such as cell cycle, cadherin signaling pathway, angiogenesis, integrin signaling pathway, VEGF signaling pathway, Parkinson disease, CCLR signaling map, EGF receptor signaling pathway and FAS signaling pathway. The sub-cellular location of the validated biomarkers was mainly in the nuclear and cytoplasm component (Figure 22C). In Figure 22D, the STRING analysis of all validated biomarkers pointed to MCM6, MCM7, CDC20, CCNE1, MSH2, CDKN1A, CDKN2A, ESR1 and ERBB2 at the core of the interaction network. These are fundamental proteins involved in the most altered described pathways triggering EC, ERK, PI3K, WNT, and transcription signaling pathways. Moreover, these are key pathways widely described in cancer. Thus, to further support our findings, we explored each of our 30 validated biomarkers and their prognostic association between other types of cancer in The Human Protein Atlas. All but ESR1, PAX8, PGR, TMEFF2, were associated with prognosis of breast, cervical, colorectal, head and neck, liver, lung, melanoma, ovarian, pancreatic, renal, or urothelial cancers. Interestingly, proteins such as CCNE1, FASN, HDGF, MCM7, PIGR, PTK2, or TRA2B have been described as having a prognostic relation with some other type of gynecological cancer (breast, cervical or ovarian cancer) (see Figure S3).

Finally, as part of our literature search, we identified chemical probes that have been developed and are currently in different phases of clinical trials targeting some of our validated biomarkers. Specifically, small molecules for different applications have been developed against AURKB, ERBB2, ESR1, PGR, PLK1, and PTK2, as well as therapeutic antibodies against ERBB2 and VIM. In addition, current treatment for EC includes medroxyprogesterone acetate, hydroxyprogesterone caproate, and megestrol acetate targeting PGR, for inoperable patients or for advanced or recurrent tumors [321,322].

Their effectiveness has been reported to increase with the combined use of estrogenic compounds such as tamoxifen targeting ESR1 [323].
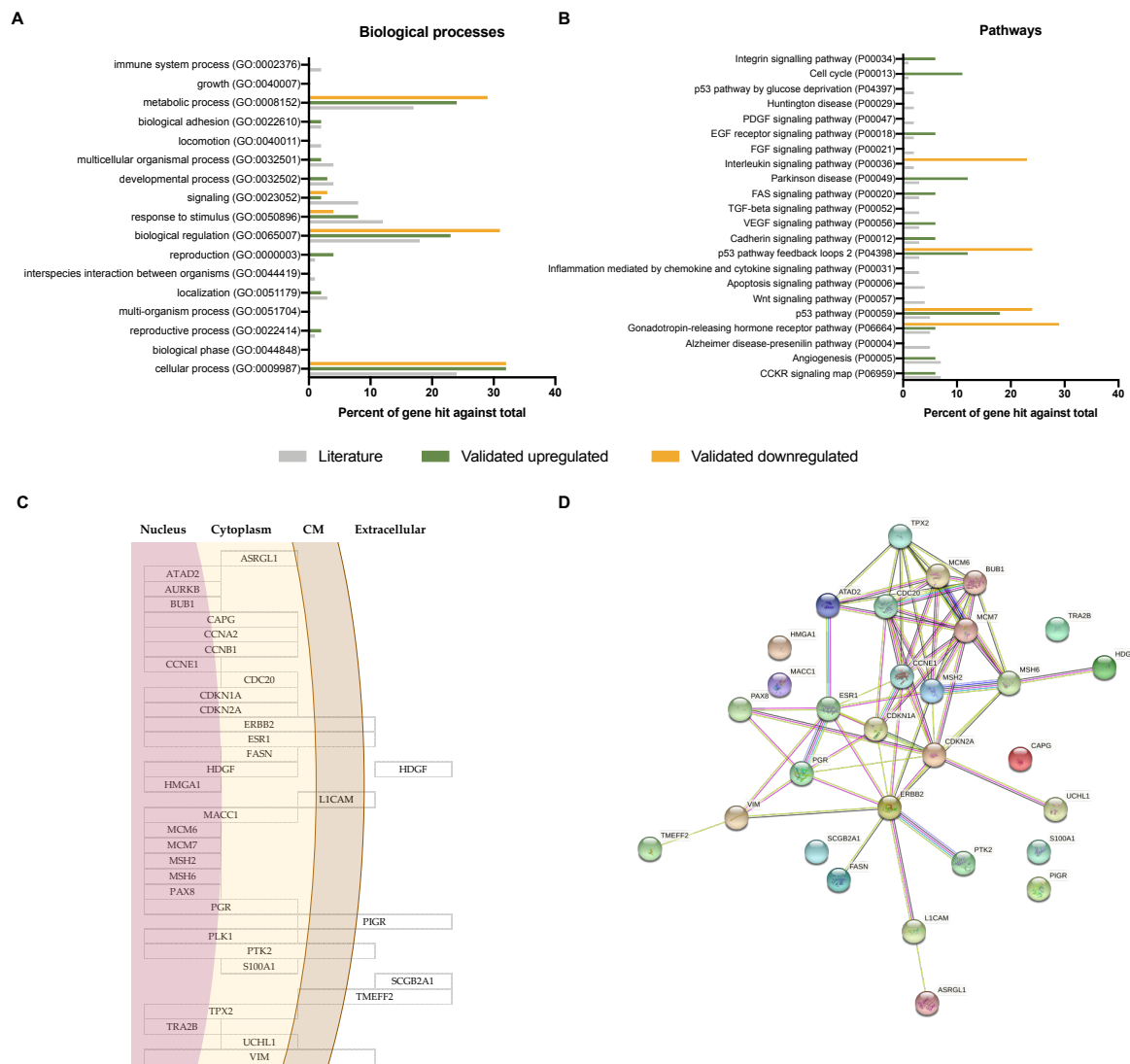
Figure 22. Functional analysis. Top 20 represented biological processes (**A**) and pathways (**B**) from the 255 proteins reviewed in the literature (out of 56 pathways) and all the pathways related to the set of validated proteins. Highlighted in green and yellow are the upregulated and downregulated pathways in relation to the literature, respectively, and in grey, the pathways represented by the proteins compiled. (**C**) Subcellular location of the 30 validated prognostic biomarkers. (**D**) String analysis of the 30 biomarkers.

DISCUSSION

To increase the accuracy and objectivity on the diagnostic process, several studies have discovered and described prognostic biomarker candidates, but none of them have reached the clinical practice. This is because there exist critical issues that hamper their clinical application and that are discussed in

this chapter, from a conceptual, methodological, and analytical point of view, together with the exposure of new strategies to follow in biomarker research (Table 13).

The systematic review described in this Chapter 1, underlines the lack of potential prognostic biomarkers of EC. Among the 2,507 articles identified in this review, 398 were deeply analyzed. As a result, this review compiles information of 255 potential biomarkers, which are related to one or more of the clinical prognostic factors in EC, recurrence, or survival. Moreover, we performed an *in-silico* validation of those biomarkers in two of the currently available molecular EC studies, specifically the RNA sequencing data generated by the TCGA and CPTAC and the proteomic data generated by the CPTAC, which compiled data from a total of 428 EC patients.

Our results revealed 30 biomarkers that show strong evidence for being prognostic EC biomarkers. Remarkably, only ESR1, PGR, ERBB2, L1CAM, MSH2, and MSH6 have been broadly studied as EC biomarkers in the literature [57]. This study adds 30 proteins that may carry important prognostic information in EC and, therefore, should be prioritized for external validation studies. Most of the validated biomarkers discriminate histological type and grade, molecular classification, OS and RFS. Regarding FIGO stage, we could only validate PGR as a good prognostic factor, as described by others [324,325]. Among the most outstanding biomarkers, the spindle assembly factor TPX2 merits further attention since its differential expression allowed for the discrimination of all the prognostic factors studied, except for the FIGO stage. TPX2 is a spindle assembly factor required for normal assembly of mitotic spindles, which mediates AURKA localization to spindle microtubules. It has been studied in a broad range of cancers as prognostic marker, including renal, liver, pancreatic and lung cancers [326]. Importantly, TPX2 was also described as prognostic biomarker in gynecological cancer such as breast cancer [327] or ovarian cancer [328]. Regarding EC, it was identified in two bioinformatics studies [317,329], and it has been further investigated in vitro to demonstrate its prognostic ability [318,319,330,331].

Another highlight in this study is the identification of biomarkers to classify EC according to the molecular classification. This classification has acquired an increasing relevance in this last year since it has been incorporated in the most recent clinical guideline of EC to improve risk assessment [64]. Despite the existence of a simplified classification, the Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE), this molecular classification is not implemented in all centers due to the technical complexity of analyzing the POLE mutations. Our study identified 16 previously described biomarkers with high discrimination potential regarding the molecular groups. We could accurately separate the CN-HIGH group from the others by using the L1CAM biomarker. L1CAM was found frequently expressed in the CN-HIGH group by *Kommoss FK et al., 2018* [332] and they also showed that the L1CAM status was predictive of worse outcome in tumors with no specific molecular profile. More studies are needed to facilitate the classification of EC patients within this molecular classification, particularly to identify the POLE group and the groups classified as multiple classifiers.

	Observation	Alternative / Improvement
Preanalytical factors		
Study design	Unclear clinical question/multiple hypothesis testing	Clearly define the clinical question to be addressed
	Almost no discovery studies. Mostly validation studies	Evident need of discovery studies and need to go through the different steps of the biomarker pipeline: discovery, verification and subsequent validations
	Small sample size	Sample size must be adequated to ensure adequate statistical power at every step of the process
	Patients not classified according to TCGA classification	Need to classify patients in the new TCGA molecular classification
Patient selection	No consensus on which parameters need to be reported and detailed information about patients is often absent	Define inclusion/exclusion criteria for patient selection and clinicopathological characteristics of patients
Clinical sample selection	Most studies were performed using tissue samples	Use of liquid biopsies is necessary to develop prognostic biomarkers for preoperative implementation. It is highly recommended the use of minimally invasive proximal fluids
Analytical factors		
Analytical platform	The largest proteomic study was performed using RPPA	Apply recent advances in MS to increase proteome coverage
	Most studies use one or few proteins for validation	Apply recent advances in multiplexed antibody-based or MS-based techniques to accurately measure higher number of proteins (e.g. Multiplexed ELISA, RPPA, LC-SRM, LC-PRM)
Post analytical factors		
Statistical anaysis	Lack of robust statistics	Provide descriptive statistics such as fold change ratios and ROC analysis
	Most studies focus on single biomarkers	Search for biomarker signatures in order to increase prognostic value
	Regression analysis are not always calculated in an optimal manner	Multivariate analysis adjusted for prognostic factors and relevant patient characteristics would be the desirable testing. Reporting hazard ratios and confidence intervals
	Lack of external and/or independent validation studies	Validation is always needed. The TCGA (RPPA) or CPTAC (MS) cohorts are publicly available.
Implementation in clinics		
Clinical relevance	Insufficient data to provide clinical relevance	Studies should incorporate data to measure the benefit of the results over clinical tools that are currently being used
Feasibility to implement the assay	Most of the studies are based in IHQ technique	Easy and straightforward assays, need of assays with high-level of automotion and reproducibility
Comercialization an intellectual property	-	The path to clinical implementation should consider protection of the results (if needed), to understand the freedom-to-operate, etc
FDA approval	-	Seek FDA guidance as early as possible in the development phases

Table 13. Outline of the preanalytical, analytical, and postanalytical factors detected in the articles reviewed and recommendations of alternatives to consider for future studies.

Related to OS and RFS, we developed an 11-biomarker signature, including ASRGL1, ESR1, FASN, HDGF, MACC1, MCM6, MCM7, MSH2, MSH6, PTK2, and TPX2, to predict OS at 48 months with an AUC=0.827. Additionally, we also developed a model of 5 proteins, including ATAD2, BUB1, MSH6, TPX2, and TRA2B, that predicts RFS only reached AUC=0.712 at time point 48 months. These findings highlight the importance of designing new studies to assess the prognosis of EC patients. Particularly, the need of studies specifically designed to identify biomarkers that can help in the prediction of EC recurrence, as currently it is not clear which are the proteins highly influencing the recurrence of the patients.

Moreover, the proteins validated in this study have been described to have an important function in the biology of the tumors, having a role in the key pathways triggering cancer such as ERK, PI3K, WNT, as well as transcription signaling pathways. To support our findings, we also explored the prognostic role of these 30 biomarkers in other types of cancer (such as of breast, cervical, colorectal, head and neck, liver, lung, melanoma, ovarian, pancreatic, renal, or urothelial cancers), finding association of those with favorable or unfavorable outcomes of the patients depending on the protein and the cancer type.

Considering all the above mentioned, our results in Chapter 1 confirm that literature revision and further *in-silico* validation of the previously described biomarkers in currently available, broadly documented, and considerable in size cohorts of patients such as the TCGA and CPTAC datasets is a valid approach for prioritizing robust biomarkers in future studies. Furthermore, we aimed to validate the 30 biomarkers described in this study in the following Chapter 2 to achieve the main goal of improving the risk assessment in EC by measure protein biomarkers in uterine fluids.
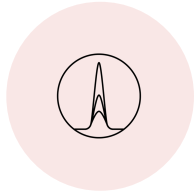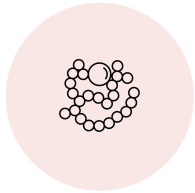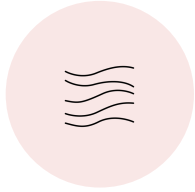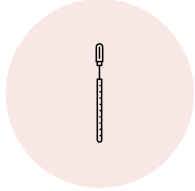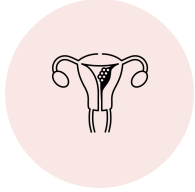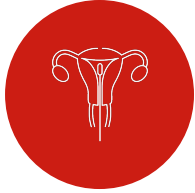
CHAPTER 2

Protein biomarkers in uterine liquid biopsy for an objective and accurate pre-operative risk assessment in endometrial cancer

SPECIFIC BACKGROUND

As shown in Chapter 1, most EC biomarker studies used endometrial tissues. Unlike tissue samples, pipelle biopsies are collected by minimally invasive procedures at an early stage of the current EC diagnostic process. They are in close contact with the tumor in the endometrium and this may be enriched in proteins directly derived from the tumor cells. Previous studies have reported a high correlation between endometrial tissue and pipelle biopsies at RNA and DNA levels [198,199], and also described the fluid of the pipelle biopsies as source of EC protein biomarkers [200]. Hence, uterine fluids may be a promising alternative source of EC prognostic protein biomarkers.

From a biological point of view, proteins are key players in many cellular processes and variations in their abundance levels can be associated with aggressiveness of different tumors. Proteins are present in biofluids and thus may be valuable prognostic indicators for the development of non-invasive prognostic tests. From a clinical perspective, proteins are widely implemented as biomarkers in comparison with genomic biomarkers, as they can be detected and quantified by accessible techniques such as IHQ or ELISA.

The work presented in Chapter 2 aims to explore the proteomic landscape of uterine fluids, which are isolated from pipelle biopsies, by generating a comprehensive EC-specific spectral library built on those uterine fluids. Moreover, the spectral library serves as reference to unveil type- and grade-specific biomarkers as well as tumor recurrence biomarkers using the novel and advantageous data-independent acquisition approach. This study aims to shed light on improving the pre-operative risk assessment of EC patients, and to facilitate treatment decision-making for each patient.

MATERIALS AND METHODS

Patients and sample description

A total of 178 patients were recruited in the Vall Hebron University Hospital (Barcelona, Spain), and the University Medical Center Freiburg (Freiburg, Germany) from 2011 to 2019. This study was approved by the Ethical Committees of each hospital. The inclusion criteria of the study were any women diagnosed with EC and received surgery treatment that agreed to sign the informed consent forms to participate in the study. Patients recruited in this study were spitted in two, corresponding to a first set of 42 patients used for the generation of the spectral library and a second set of 149 women used for the discovery study. Among those, 16 samples overlapped in both sets. From the second set of women, 104 were diagnosed with endometrioid EC (EEC), and 45 with non-endometrioid histologies including 16 serous (SEC), 14 carcinosarcomas, 6 mixed EC, 4 clear cells, 1 mucinous and 4 undifferentiated EC.

Regarding EEC, 69 were grade 1 or 2 (low-grade), while 35 were grade 3 (high-grade). Final diagnosis of EC cases was achieved by the histopathological analysis of an endometrial biopsy and a surgical specimen after hysterectomy. Clinicopathologic features of the patients are summarized in Table 14.

In all patients, pipelle biopsies were collected by aspiration using a Cornier Pipelle (Eurogine Ref. 03040200), prior to the surgical treatment of the patients. These pipelle biopsies were transferred to 1.5 mL eppendorfs and diluted with phosphate buffer saline in a ratio of 1:1 (v/v). Samples were then mixed and centrifuged at 2,500 x g for 20 minutes at 4ºC. Samples were then separated into supernatant and pellet, which corresponded to the fluid and the cellular fraction, respectively. Samples were kept at -80ºC until its use.

	Library generation (n=42)				Clinical study (n=149)				
	Endometrioid low-grade (n=20)	Endometrioid high-grade (n=10)	Type II: Serous (n=12)		Endometrioid low-grade (n=69)	Endometrioid high-grade (n=35)	Type II: Serous (n=16)	Type II: Other histologies (n=29)	
Age (years)									
Mean	63	62	66		63	71	67	66	
Minimum	31	41	46		36	38	45	52	
Maximum	84	72	81		90	92	78	79	
Body Mass Index									
Mean	30.84	29.28	33.79		33.38	31.02	29.54	30.64	
Minimum	19.72	20.85	23.05		17.42	20.85	23.05	20.00	
Maximum	43.69	48.44	46.66		53.92	48.44	41.41	44.08	
Uterine condition									
Premenopausal	3	-	-	7%	18	2	2	-	15%
Postmenopausal	17	10	12	93%	51	33	14	29	85%
FIGO stage									
IA	12	6	3	50%	39	9	4	9	41%
IB	-	-	1	2%	5	10	1	2	12%
II	-	-	3	7%	19	7	3	6	23%
IIIA	3	1	-	10%	1	-	2	2	3%
IIIB	-	-	-	0%	-	-	-	-	0%
IIIC1	1	1	-	5%	2	1	-	1	3%
IIIC2	2	1	1	10%	3	2	4	7	11%
IVA	-	-	-	0%	-	3	-	1	3%
IVB	-	-	1	2%	-	3	2	1	4%
Miometrial invasion									
<50%	13	6	7	62%	46	11	7	13	52%
>50%	7	3	2	29%	23	24	9	16	48%
Lymphobascular invasion									
Yes	3	4	4	26%	13	13	8	8	28%
No	14	5	5	57%	49	19	6	20	63%
Unknown	3	1	3	17%	7	3	2	1	9%
Molecular classification									
POLEmut	1	-	-	4%	2	2	-	-	4%
MMRd	4	4	2	36%	20	14	2	4	33%
NSMP	6	4	1	39%	27	10	-	8	38%
p53mut	-	-	6	21%	2	2	12	14	25%
NA	9	2	3		18	7	2	3	
Risk assessment - Molecular classification unknown									
Low	11	-	-	26%	39	-	-	-	27%
Intermediate	-	-	-	0%	4	8	4	9	17%
High-intermediate	1	6	3	24%	19	17	-	-	25%
High	6	3	6	36%	6	8	10	20	31%
Risk assessment - Molecular classification known									
Low	6	-	-	21%	24 = +	2 ++	-	-	22%
Intermediate	-	-	-	0%	4	6	4	8	18%
High-intermediate	-	5	3	29%	17 -	13 - -	-	-	25%
High	5	3	6	50%	6	7	10	18	34%
Recurrence									
Yes (<12 months)	-	-	-	0%	-	6	1	5	8%
Yes (12-24 months)	1	2	1	10%	5	4	1	5	10%
Yes (>24 months)	-	-	3	7%	5	3	4	3	10%
No	17	8	8	79%	59	22	10	16	72%

Table 14 Clinical characteristics of women enrolled in both datasets: (1) library generation (n=42); (2) clinical study (n=149). In the risk assessment with molecular classification known, the changes in risk classification due to molecular classification (patients classified as POLEmut) are shown. = means that being POLEmut did not change its group. + means the group where POLEmut patient was added, while – the group where POLEmut patient was removed from. POLEmut, POLE ultramutated; MMRd, mismatch repair deficient; NSMP, nonspecific molecular profile, and p53mut, TP53 mutated.

Sample preparation for LC-MS analysis

The liquid fraction of each pipelle biopsy sample i.e., uterine fluids, was sonicated (Branson sonifier) during 5 cycles of 5 seconds on ice and diluted four-fold in buffer A (Agilent, 5185-5987). Possible particles were removed with a 0.22 μm spin filter, spinning for 1 min at 16,000 g. All samples were depleted of albumin and IgGs with the multi affinity removal column HSA/IGG (Agilent, 5188-8826) following the method described by the manufacturer.

Total protein concentration was measured by the BCA assay (Thermo Scientific) for the first set of samples and by the Bradford assay (Sigma) for the second set. Equal amounts of total protein per sample were processed. The first set of samples was denatured in sodium deoxycholate (SDC) buffer to have a final concentration of 3% SDC. The second set of samples was denatured by addition of urea to a final concentration of 6 M, incubated at 22°C under agitation for 20 min. Proteins from all samples were then reduced in 5 mM dithiothreitol (DTT) for 1 h at 37°C and alkylated in 15mM iodoacetamide (IAM) for 30 min at 22°C in the dark. Samples were then digested with LysC protease (Wako) at a protease/total protein ratio of 1/50 (w/w) for 3 h at 37ºC, and then with trypsin (Promega) overnight at 37°C at a 1/25 ratio (w/w) after dilution of urea to a final concentration of 1 M. Proteolysis was stopped by addition of 1 μl of formic acid per 100 μl of solution. The tryptic peptides were desalted onto solid phase extraction cartridges (Sep Pak tC18, 25mg, Waters) and evaporated to dryness. Peptide quantification was performed by spectrophotometry (205 nm) with a Nanodrop One (Thermo Scientific) to analyze the same amount of peptides by LC-MS.

LC setup

The chromatography setup consisted of a Dionex Ultimate 3000 RSLC chromatography system 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. Samples were injected onto a trap column (75 μm × 2 cm, C18 pepmap 100, 3 μm) with the loading phase and further eluted onto an analytical column at 300 nl/min by a linear gradient. Fractionated samples for the spectral library generation were separated by a 33 min gradient from 2 to 35% B in 33min (75 μm × 15 cm, C18 pepmap 100, 2 μm). Samples analyzed by data independent acquisition (DIA) were separated by a 120min gradient from 2 to 35% B (75 μm × 25 cm, C18 pepmap 100, 2 μm).

LC-DDA analyses for spectral library generation

In order to generate a comprehensive uterine fluid spectral library for the processing of the DIA data, a pool of 300 ug of the tryptic peptides from the first set of uterine fluids was used. The peptides of this pool were fractionated by high pH reversed phase (XBridge, BEH C18, 130Å, 3.5μm, 4.6mm X 150mm, Waters) liquid chromatography (Agilent LC1260) using a 45 min gradient from 2% to 45% B. Phase A consisted of ammonium acetate 15mM, pH= 10 and phase B of 90% acetonitrile complemented 10% A

phase. Fractions were collected every 1 min ending up with 59 different fractions. Peptide concentration of each fraction was determined with Nanodrop. Similar amount of peptides of each fraction was individually analyzed with a Q Exactive HF (Thermo Scientific) using a top12 data dependent acquisition (DDA). The MS data was searched by the Spectronaut (Biognosys) Pulsar search engine against the TrEMBL UniProt Homo sapiens (February 2019, 73,928 entries) with the default settings to generate the spectral library.

LC-DIA analysis

The DIA acquisition scheme was designed with 51 isolation windows of variable width ranging from 6.5 to 200.5 m/z unit and with an overlap of 1m/z. Acquisition windows width and distribution across the m/z range were designed to be inversely proportional to the average ion density of typical sample. The orbitrap resolution was set to 30,000 (at 200 m/z) with an automatic gain control set to 1e6 charges and maximum fill time optimized for maximizing the MS duty cycle. DIA data were processed with Spectronaut software (version 13.10.191212.43655) using the experimental spectral library and with default settings. We excluded proteins detected with a single peptide and applied a data global normalization across samples based on the median.

Data analysis

Data exported from the Spectronaut software was used for the data analysis. First data processing was performed with the Perseus software (version 1.6.12.0). Raw data was log2-transformed, and proteins identified in less than 70% of all the samples were filtered out. The missing values were imputed from a normal distribution (downshifted mean by 1.8 standard deviation (SD) and scaled SD (0.3) relative to that of proteome abundance distribution). This final dataset was used for the subsequent statistical analysis performed with R software. Limma package was used to determine the significantly changed proteins between the different clinical groups of interest. Results were filtered to have a significant *adjusted P-value* lower than 0.05, a minimum fold change of ± 2, and an AUC value higher than 0.7.

Development of protein panels

A logistic regression model was fitted to the data to assess the power of different protein combinations in classifying samples into two clinical categories. To avoid overfitting, only combinations of a maximum of three proteins were tested. ROC curves were generated for each of these regression models; AUC, sensitivity and specificity were obtained at the "optimal" cutoff point for discrimination between groups. This optimal cutoff point was chosen as the threshold that maximized the distance to the line of identity (diagonal) or equivalently to the threshold where the sum of sensitivity and specificity was maximum. To assess the robustness of each protein panel for the different parameters, a leave-one-out cross-validation method was used. The 95% confidence intervals (CI) of the AUCs and the 95% CIs of the sensitivity and specificity values (at the best threshold) were calculated with a stratified bootstrap

resampling of size 2000. All statistical calculations were performed with Rstudio 1.2.15033 running R version 4.0.3 [333]. Analysis of ROC curves was performed with the R pROC package (1.18.0) [334].

In-silico validation

The 255 proteins identified as prognostic EC biomarkers in literature [57] were subjected to further *in-silico* validation using two available and broad datasets (*i.e.* RNA-Seq data of the n=333 patients included in the analysis of EC tumors in The Cancer Genome Atlas (TCGA) and the n=95 EC patients included in the Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset) as described in Coll-de la Rubia E *et al.* [335] . Resulting proteins from the review and the *in-silico* validation were compared to the identified proteins in our spectral library from a pool of the fluid of uterine fluids in EC patients (n=42), to the proteins identified in the clinical study (n=149) and to the proteins identified as potential prognostic biomarkers (n=175) for any of the prognostic factors studied.

RESULTS

Workflow

The study here conducted follows a novel strategy to identify proteomic biomarkers. It includes the generation of a spectral library obtained by the analysis of a fractionated pool of uterine fluids performed by DDA acquisition, followed by a clinical study performed in DIA acquisition using the generated spectral library, and final *in-silico* validation of identified biomarkers using literature and large available datasets of patients analyzed by RNA-sequencing and proteomics. From the clinical study, development of 2-protein signatures was also performed to improve the accuracy in the determination of histological type and grade as well as recurrence prediction. The workflow used in this study is shown in Figure 23.
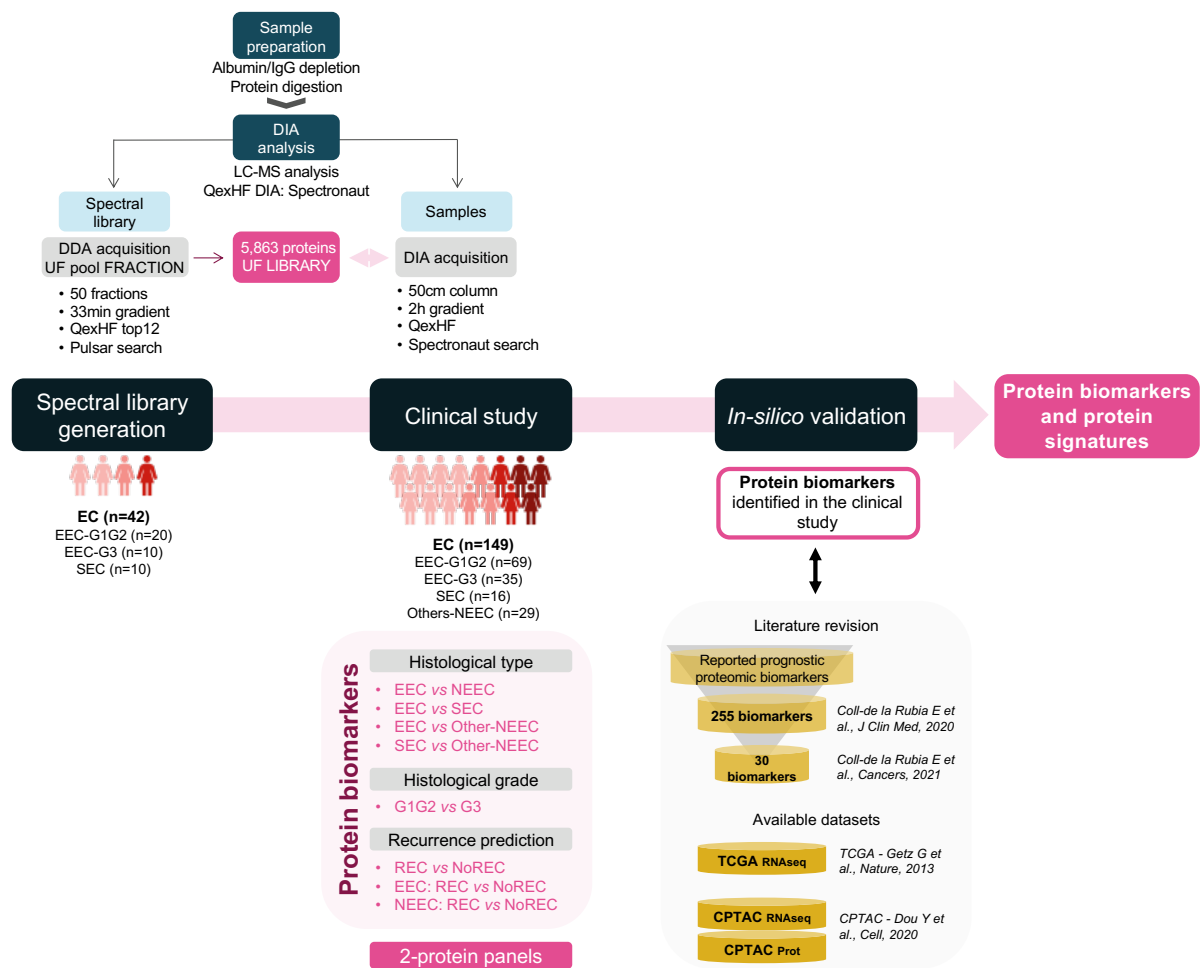
Figure 23 General workflow of the study. EC, endometrial cancer patients; UF, uterine fluids; MS: mass spectometer; DDA, data-dependent acquisition; DIA, data-independent acquisition, EEC, endometrioid EC; NEEC, non-endometrioid EC; SEC, serous EC; G, grade; REC, recurrent patients; NoREC, non-recurrent patients.

Generation of a uterine fluid spectral library

We generated a comprehensive uterine fluid spectral library from a pool of uterine fluid samples fractionated in 59 fractions and each of them analyzed by DDA. From this library we obtained reference spectra for 54,448 peptides belonging to 5,863 proteins. This library covers most of the key proteins involved in cancer-related molecular functions such as cell cycle, apoptosis, extracellular matrix-receptor interactions, adherent junctions or angiogenesis; as well as the main cancer-related pathways such as Wnt/beta-catenin, PI3K/Akt or P53 pathways, all of them important deregulated pathways in EC [45,336]. Moreover, among the 5,863 proteins identified in the liquid fraction of pipelle biopsies, 96% overlapped with proteins identified in EC tissues from the Clinical Proteomics Tumor Analysis Consortium (CPTAC), as shown in Figure 24A. From the almost 11,000 proteins identified in those EC tissue samples, we identified 51% in uterine fluid samples. These proteins are not only extracellular or

secreted proteins, but we also identified a high number of nuclear and cytoplasmic proteins proving that uterine fluids are enriched in proteins coming from necrotic or lysed cells from the tumor.

We previously reported a list of 255 proteins described in literature as potential EC prognostic biomarkers, mostly derived from studies performed in endometrial tissue samples [57]. Importantly, 62% of the 255 proteins were identified in uterine fluid samples. Altogether, these results highlight the potential of uterine fluids as a minimally invasive source of EC prognostic biomarkers.

This unique spectral library of 5,863 proteins detected in the fluid fraction of pipelle biopsies can be of great utility for other proteomic studies of gynecological diseases such as EC, endometriosis or ovarian cancer.
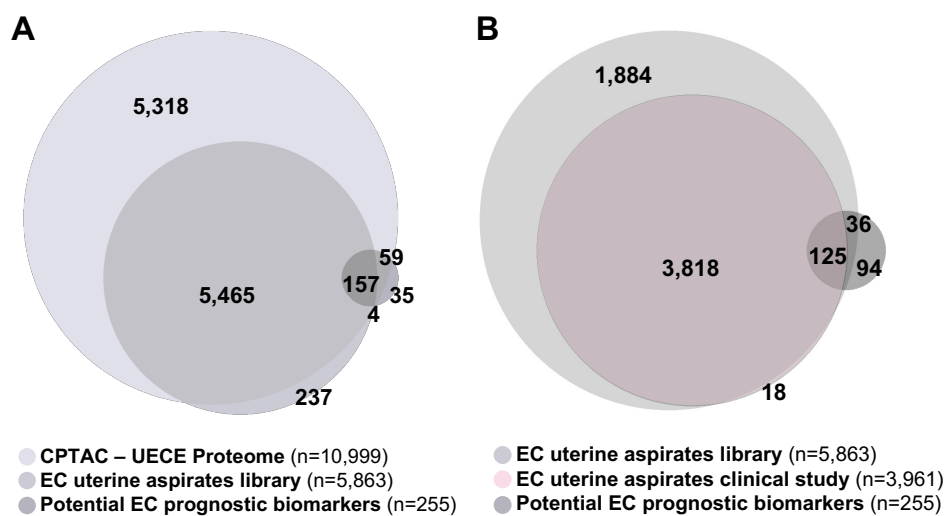
Figure 24 Venn diagrams representing the number of proteins identified in our sample sets in comparison to available tissue proteomes of EC (*i.e.,* CPTAC dataset) and to described prognostic protein biomarkers in EC. (**A**) Overlap between the 5,863 proteins composing our spectral library from a pool of EC patients' uterine fluids, the EC proteome defined by the CPTAC in formalin-fixed paraffin-embedded (FFPE) EC tissues [298], and the previously described 255 protein prognostic biomarkers in EC [57]; (**B**) Overlap between our spectral library of uterine fluids, the 3,961 proteins identified in our clinical study (n=149), and the detection of EC prognostic biomarkers in this set of samples.

Clinical study

We then analyzed the uterine fluid samples from 149 patients by DIA. We used the experimental spectral library to extract quantitative protein information from the DIA data in a targeted manner. A total of 3,961 of the 5,863 proteins that composed our library were detected in these 149 samples (68%), with an average of 2,743 proteins identified per sample. We also detected 125 (49%) of the 255 proteins previously described as potential EC biomarkers (Figure 24B).

Pre-operative risk assessment: biomarkers to determine histological type and grade

Preoperative risk-assessment is crucial to classify EC patients into groups of risk for disease recurrence in an early step of the diagnostic process in order to define the most optimal surgical treatment. Currently, this preoperative risk assessment is achieved by the determination of different prognostic factors: i) the assessment of tumor grade and histological subtype based on the pathological examination of the tissue contained in an endometrial biopsy, and ii) the evaluation of the myometrial invasion, cervical involvement, tumor size and LNS by imaging techniques.

Regarding the acquisition methodology of the endometrial biopsies, they are preferably obtained by aspiration (*i.e.,* pipelle biopsies), as it is a less invasive technique compared to hysteroscopy that often requires of anesthesia. However, in some cases, pipelle biopsies can't be obtained due to technical issues (*i.e.,* cervical stenosis), or their cellular material is insufficient to provide a final diagnosis. In our cohort of 149 patients, 52% were diagnosed using pipelle biopsies, while in 42% of cases hysteroscopy was the method used to obtain the final endometrial biopsy. Importantly, 14% of these women had to undergo a hysteroscopy due to problems associated with pipelle biopsies (3% cervical stenosis, 4% insufficient material, 7% other issues) (Table 15).

	Final diagnosis (n=149)				
	Endometrioid low-grade (n=69)	Endometrioid high-grade (n=35)	Type II: Serous (n=16)	Type II: Other histologies (n=29)	
Preoperative diagnosis methodology					
Pipelle biopsy	32	24	9	12	52%
Pipelle biopsy* + hysteroscopy	2	1	1	-	3%
Pipelle biopsy** + hysteroscopy	3	-	-	3	4%
Pipelle biopsy + hysteroscopy	7	2	-	2	7%
Hysteroscopy	23	7	4	7	28%
NA	2	1	2	5	7%
Preoperative diagnosis: Histological type					
Hyperplasia	1	-	1	-	
Endometrioid	**66**	**31**	-	9	
Serous	2	1	**13**	2	
Other non-endometrioid	3	3	1	**16**	
Agreement	31P; 35H	20P; 3H	11P; 2H	9P; 6H	85%
Disagreement	5P; 1H	4P	2H	4P; 5H; 2NA	15%
Preoperative diagnosis: Histological grade					
Low-grade	**62**	14	-	6	
High-grade	3	**18**	**13**	**20**	
Agreement	17P; 21H	13P; 5H	9P; 4H	9P; 8H	65%
Disagreement	14P; 13H	9P; 5H	-	2P; 2H; 2NA	35%

* Stenosis; ** Insufficient material; P: Pipelle biopsy; H: hysteroscopy

Table 15 Clinical methodology used to acquire endometrial biopsies in the 149 patients enrolled in the study, and comparison between the tumor type and grade determined by the preoperative examination of the endometrial biopsy and the final diagnosis obtained after surgery by the examination of the resected tumor. P, pipelle biopsy; H, hysteroscopy; NA, not available.

Regarding the information of the histological type and tumor grade obtained from the endometrial biopsies, a high inter-observer variation has been reported with a clear impact on the surgical. Among the 149 patients included in our study, 15% and 35% of the patients were inaccurately diagnosed regarding histological type and grade, respectively, independently of the method of acquisition of the

endometrial biopsy. Importantly, 87% (20/23) and 49% (23/47) of the misdiagnosed patients regarding histological subtype and histological grade had a relevant clinical impact (Table 15).

Histological type

In the 149 uterine fluids analyzed in this study, we quantified an average of 2,743 proteins per sample. In order to identify proteins that allowed for the discrimination between the main EC histological types, we compared the levels of all the identified proteins in the different histologies and we found a total of 61 proteins differentially expressed between different histologies (see Annex 2, Table A2.1 and Table 2.4 for specifications of each biomarker). Among them, we identified 34 differentially expressed proteins (*Adjusted P* value <0.05, *Fold-Change* > |2|, AUC > 0.7) between endometrioid EC (EEC) and more aggressive non-endometrioid EC (NEEC) tumors. Most of these proteins were less abundant in NEEC, with the exception of PROT, PROT, PROT, PROT, PROT, and PROT. Interestingly, 8 of those proteins (PROT, PROT, PROT, PROT, PROT, PROT, PROT, and PROT) demonstrated great performance (*Adjusted P* value <0.05, *Fold-Change* >|2|, and AUC > 0.8) to distinguish between EEC and SEC tumors, being SEC the most common histology for non-endometrioid tumors. When comparing EEC and other non-endometrioid histologies excluding SEC, 3 proteins (PROT, PROT, and PROT) were able to accurately discriminate between the two groups (*Adjusted P* value <0.05, *Fold-Change* >|2|, AUC > 0.8). Finally, levels of PROT and PROT were significantly different between SEC and other non-endometrioid histologies (*Adjusted P* value <0.05, *Fold-Change* >|2|) showing AUC values of 0.86 and 0.79 for PROT and PROT, respectively. Figure 25A shows the proteins with AUC values > 0.75 for each comparison. Violin plots of the best performing proteins for each comparison are shown in Figure 25B. Remarkably, the most promising proteins identified here in minimally-invasive pipelle biopsies showed same differential levels between histologies in EC tissues, as demonstrated in the Clinical Proteomic Tumor Analysis Consortium (CPTAC) [298], dataset at protein level, and in the TCGA dataset at RNA level, validating our results and demonstrating the tumoral origin of our highlighted biomarkers.

A

Gene Name	FC (EEC vs NEEC)	Adjusted P value (EEC vs NEEC)	AUC (EEC vs NEEC)	FC (EEC vs SEC)	Adjusted P value (EEC vs SEC)	AUC (EEC vs SEC)	FC (EEC vs Others)	Adjusted P value (EEC vs Others)	AUC (EEC vs Others)	FC (SEC vs Others)	Adjusted P value (SEC vs Others)	AUC (SEC vs Others)
PROT	9.58	4E-11	**0.85**	3.82	0.00	**0.88**	6.48	0.00	**0.83**	-2.18	0.38	0.65
PROT	-6.00	2E-06	**0.80**	-2.92	0.00	**0.87**	-4.75	0.00	0.77	1.60	0.61	0.59
PROT	-4.28	2E-05	0.78	-2.21	0.01	**0.80**	-3.95	0.00	0.77	1.18	0.88	0.51
PROT	-2.40	3E-04	0.76	-0.84	0.24	0.68	-3.21	0.00	**0.80**	-1.79	0.34	0.65
PROT	-2.24	2E-05	0.76	-1.20	0.01	0.77	-2.19	0.00	0.75	1.05	0.94	0.56
PROT	-3.06	3E-04	0.76	-1.41	0.09	0.76	-3.52	0.00	0.76	-1.33	0.73	0.58
PROT	-2.52	2E-03	0.76	-0.69	0.46	0.63	-3.93	0.00	**0.82**	-2.43	0.28	0.73
PROT	-6.00	3E-04	0.76	-2.95	0.02	**0.81**	-4.65	0.01	0.73	1.66	0.69	0.56
PROT	-3.12	6E-05	0.75	-1.65	0.03	0.78	-3.10	0.00	0.74	1.01	0.99	0.52
PROT	-3.71	5E-05	0.75	-1.56	0.08	0.72	-4.65	0.00	0.76	-1.57	0.56	0.58
PROT	-3.31	2E-03	0.75	-1.17	0.32	0.66	-4.87	0.00	0.79	-2.17	0.39	0.69
PROT	2.48	3E-04	0.68	2.18	0.00	**0.88**	1.36	0.49	0.58	-3.35	0.08	**0.84**
PROT	2.03	2E-05	0.64	1.97	0.00	**0.87**	1.06	0.87	0.53	-3.70	0.00	**0.86**
PROT	-2.45	5E-04	0.74	-1.54	0.02	**0.84**	-2.07	0.02	0.69	1.40	0.58	0.57
PROT	-5.13	4E-05	0.74	-2.82	0.00	**0.82**	-3.72	0.00	0.70	1.90	0.50	0.61
PROT	-2.39	2E-03	0.71	-1.60	0.02	0.80	-1.88	0.06	0.67	1.61	0.45	0.59
PROT	3.07	8E-04	0.69	2.41	0.00	0.80	1.77	0.23	0.63	-3.00	0.22	0.75
PROT	-3.52	1E-03	0.72	-2.23	0.02	**0.80**	-2.64	0.03	0.68	1.78	0.52	0.59
PROT	-2.03	5E-03	0.55	-2.00	0.00	0.77	-1.03	0.96	0.56	3.89	0.02	0.79
PROT	-2.09	3E-03	0.74	-0.74	0.33	0.64	-2.62	0.00	0.78	-1.57	0.44	0.67
PROT	-1.44	3E-01	0.70	0.15	0.89	0.53	-2.31	0.01	0.78	-2.57	0.21	**0.82**
PROT	-2.06	1E-02	0.70	-0.53	0.56	0.57	-2.94	0.00	0.77	-2.04	0.30	0.69
PROT	-3.64	7E-04	0.74	-1.31	0.23	0.69	-5.31	0.00	0.76	-2.14	0.39	0.63
PROT	-3.66	7E-04	0.74	-1.67	0.11	0.71	-4.22	0.00	0.76	-1.33	0.78	0.58
PROT	-1.80	5E-03	0.71	-0.31	0.66	0.61	-2.60	0.00	0.75	-2.10	0.21	0.69
PROT	-1.67	5E-03	0.73	-0.40	0.49	0.68	-2.11	0.00	0.75	-1.59	0.30	0.58
PROT	-1.99	8E-02	0.70	-0.41	0.74	0.60	-2.97	0.00	0.75	-2.23	0.34	0.67
PROT	-1.68	4E-02	0.72	-0.46	0.56	0.65	-2.07	0.01	0.75	-1.51	0.45	0.59
PROT	-1.84	3E-03	0.73	-0.69	0.24	0.70	-2.11	0.00	0.75	-1.31	0.56	0.59
PROT	-1.40	4E-01	0.67	0.35	0.72	0.51	-2.50	0.00	0.75	-3.19	0.09	0.78
PROT	-2.51	2E-02	0.71	-0.72	0.56	0.63	-3.83	0.00	0.75	-2.33	0.36	0.67
PROT	-1.48	3E-01	0.63	0.31	0.75	0.57	-2.72	0.00	0.72	-3.36	0.08	0.77
PROT	-1.53	2E-01	0.64	-0.01	0.99	0.53	-2.33	0.01	0.72	-2.31	0.26	0.75
PROT	1.94	8E-03	0.67	1.58	0.01	**0.88**	1.26	0.59	0.56	-2.37	0.21	**0.86**
PROT	-1.32	2E-01	0.63	-0.27	0.63	0.59	-1.45	0.07	0.65	-1.20	0.65	0.55

B

C **D** **E** **F**

G

Comparison	Protein 1	Protein 2	TP	TN	FP	FN	Specificity (%)	Sensitivity (%)	AUC (%)
103 EEC vs 44 NEEC			40	85	18	4	**82.5** (95%CI 74.8-89.3)	**90.9** (95%CI 81.8-97.7)	**89.4** (95%CI 83.2-94.5)
103 EEC vs 15 SEC			15	81	22	0	**78.6** (95%CI 70.9-86.4)	**100** (95%CI 100-100)	**95.6** (95%CI 90.4-98.9)
193 EEC vs 29 NEEC-Others			28	72	31	1	**69.9** (95%CI 61.2-78.6)	**96.6** (95%CI 89.7-100)	**89.3** (95%CI 83.5-94.4)
15 SEC vs 29 NEEC-Others			27	14	1	2	**93.3** (95%CI 80-100)	**93.1** (95%CI 82.8-100)	**97.7** (95%CI 93.1-100)

● EEC ● SEC ● Others

Figure 25 Performance of proteins to discriminate between different EC histological subtypes: endometrioid adenocarcinomas (EEC) and NEEC, including serous adenocarcinoma (SEC) and other NEEC histologies (Others). (**A**) Statistical results of the 32 proteins showing an AUC value greater than 0.75 to

discriminate between EEC (n=104) *vs* NEEC (n=45) and/or EEC (n=104) *vs* SEC (n=16) and/or EEC (n=104) *vs* Others (n=29) and/or SEC (n=16) *vs* Others (n=29). Highlighted in grey significant proteins showing an *Adjusted P* value <0.05, *Fold-Change* (FC) >|2| and AUC value > 0.7 for the specific comparison, and AUC values > 0.8 are indicated in bold. Highlighted in pink proteins contained in the combinations. (**B**) Violin plots picturing the distribution of the intensities obtained by DIA across patients (EEC, SEC and Others) of the best performing proteins for each comparison. (**C-F**) Scattering plots representing the probabilities of patients being correctly classified with each 2-protein panel (in pink the cut-off value), ROC curves with the individual proteins of the panel and the 2-protein panels. (**G**) Tables with true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), % specificity, % sensitivity, AUC values, and their respective 95% confidence intervals (CI) for the comparisons EEC *vs* NEEC, EEC *vs* SEC, EEC *vs* Others, and SEC *vs* Others, respectively.

To improve the performance of the individual biomarkers, we developed protein panels of two proteins combining the top 300 performing proteins for each comparison identified in the study. Then, cross-validation of the most promising combinations using the leave-one-out method was performed. As shown in Figure 25C, the combination of PROT and PROT achieved an AUC value of 89.4% (95% confidence interval (CI) 83.2-94.5) with a sensitivity of 91% and a specificity 83% to differentiate between EEC and all NEEC. To differentiate between EEC and specific NEEC histologies, we identified two 2-protein panels. A 2-protein panel composed by PROT and PROT was able to accurately discriminate between EEC and SEC with an AUC value of 95.6% (95% CI 90.4-98.9) (Figure 25D), and 100% sensitivity and 79% specificity. The combination of PROT and PROT could correctly distinguish between EEC and NEEC histologies (excluding SEC) with an AUC value of 89.4% (95% CI 83.2-94.5) (Figure 25E), achieving 97% sensitivity and 70% specificity. Finally, the combination of PROT and PROT achieved an AUC value of 97.7% (95% CI 93.1-100) (Figure 25F) to distinguish between NEEC histologies (SEC *vs* Others), with a sensitivity and specificity of 93%. CI for the sensitivities ad specificities for each comparison, as well as true positives (TP), true negatives (TN), false positives (FP), and false negative (FN) are shown in Figure 25G. Considering that 15% of the 149 EC patients included in this study were incorrectly classified regarding the histological type following the current diagnostic approach, the introduction of the biomarker panels here described would improve patient classification and thus, help to guide the most appropriate surgical treatment.

Histological grade

Histological grade is another crucial EC prognostic factor that pathologists determine from the examination of the endometrial biopsy. NEEC tumors are considered grade 3 by definition, whereas EEC should be graded based on their proportion of solid areas of tumor cells following the International Federation of Gynecology and Obstetrics (FIGO) [87]. Currently, a binary grading is recommended classifying grade 1 and 2 (G1, G2) tumors as low-grade (well and moderately differentiated tumors exhibiting <50% of solid non-glandular growth), and grade 3 (G3) as high-grade tumors (poorly differentiated tumors exhibiting >50% of solid non-glandular growth). Grade definition is not trivial, especially in high-grade tumors, where squamous differentiation strongly favors EEC over other histological endometrial carcinoma types. Additionally, low-grade tumors demonstrated better 5-year survival rates. Our analysis allowed us to identify 37 differentially abundant proteins in uterine fluids from patients with low grade endometrioid EC (grades 1 and 2) and high grade endometrioid EC (grade 3) (*Adjusted P* value <0.05, *Fold-Change* >|2|, AUC > 0.7) (see Annex 2, Table A2.2 and Table 2.4 for

specifications of each biomarker). Significant proteins with AUC values > 0.75 are shown in Figure 26A, and the expression profiles of the top3 proteins (PROT, PROT, and PROT) are represented with violin plots in Figure 26B. Similarly to histological type, most aggressive tumors (non-endometrioid histology or high-grade tumors) tend to have lower levels of the identified biomarkers.

In order to improve the power of the individual proteins in discriminating low-grade and high-grade EEC, we developed dual combinations of the top 300 best performing proteins. The most promising combinations were validated by LOOCV, and we selected the one that better identified high-grade-EEC. The combination of PROT and PROT proteins was able to discriminate low-grade-EEC from high-grade-EEC with 85% sensitivity, and 81% specificity achieving an AUC value of 85.3% (95% CI 75.9-93.1) (Figure 26C-D). Determination of histological grade of patients included in the study failed in 35% of cases, 17% of which in distinguishing between low-grade *vs* high-grade. As shown in Table 15, conventional histopathological determination of the histological grade of the tumors included in this study failed in 35% of cases, 17% of which in distinguishing between low-grade and high-grade. Therefore, our molecular signature measured in minimally invasive uterine fluids would improve the current determination of the histological grade of EC tumors.
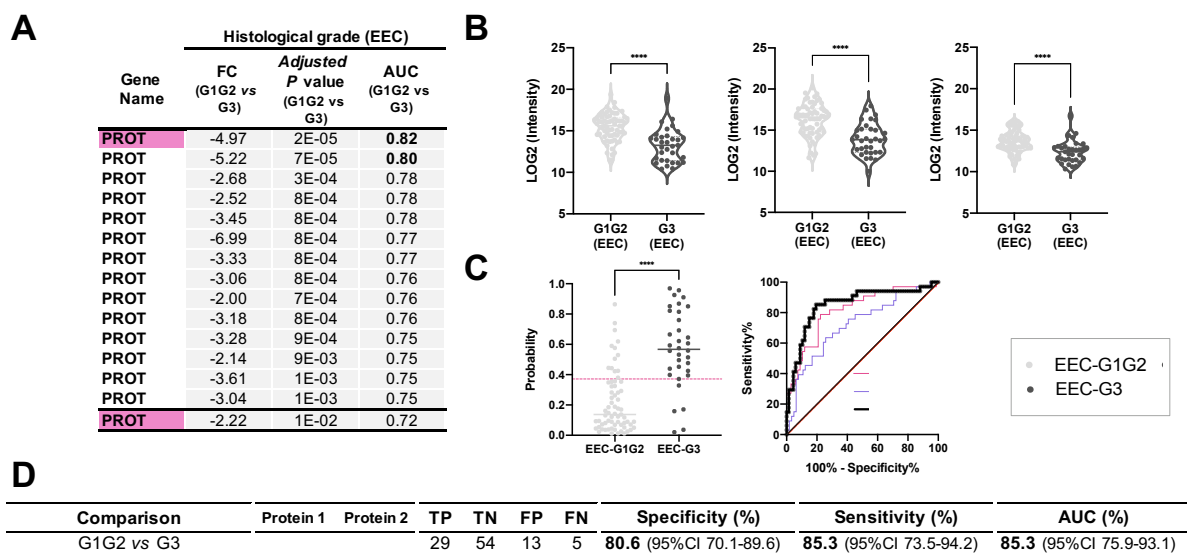
Figure 26 Performance of proteins to discriminate between different EC histological grades in endometrioid tumors: grade 1 (G1) and grade 2 (G2) defined as low-grade EEC (n=69) *vs* grade 3 defined as high-grade EEC (n=35). (**A**) Statistical results of the 14 proteins showing an AUC value greater than 0.75 to discriminate between low-grade EEC *vs* high-grade EEC. Highlighted in grey significant proteins showing an *Adjusted P* value <0.05, *Fold-Change* (FC) >|2| and AUC value > 0.7, and AUC values > 0.8 are indicated in bold. Proteins contained in the protein panel are highlighted in pink. (**B**) Violin plots picturing the distribution of the intensities obtained by DIA across patients (low-grade EEC *vs* high-grade EEC) of the top3 proteins. (**C**) Scattering plot representing the probability of patients being correctly classified with the 2-protein panel (in pink the cut-off value), ROC curves of the individual proteins of the panel and the 2-protein panel. (**D**) Table with true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), % specificity, % sensitivity, AUC values, and their respective 95% confidence intervals (CI) for the comparison low-grade EEC *vs* high-grade EEC.

Histological type and grade using pathological and molecular features

As mentioned above, pre-operative assessment of histological type and grade failed in 42% of patients using the pathological examination of the endometrial biopsy. Specifically, 23 patients were incorrectly diagnosed regarding the histological subtype, 16 of which with clinical impact on the surgery plan, whereas 47 patients had an incorrect determination of the histological grade, 22 of which with significant impact on the surgery plan. The molecular tools purposed in this article aim to diagnose with high accuracy all EC patients before surgery, particularly, those that were incorrectly diagnosed by pathological observation. Thus, combination of both techniques would reach the maximum number of correct diagnostics as possible. When we evaluate those patients with incorrect pre-operative diagnosis regarding histological type, only 6 out of 23 were incorrectly classified using our first protein panel (PROT + PROT) able to differentiate EEC *vs* NEEC. Nonetheless, from those 6 only 1 was not able to be diagnosed in any of our signatures (Figure 27A, misclassified patient indicated with a red arrow). Specifically, this patient was classified as SEC at pre-operatory level by the endometrial biopsy, and according to our protein panels is a NEEC tumor. The resected tumor revealed that this tumor was an EEC-G3. Histological grade was properly determined in both cases (endometrial biopsies and molecular signature), and therefore, the patient would have the proper surgery. Additionally, the tumor of this patient was classified as p53mut (serous-like EEC) regarding the molecular classification of TCGA. There is another patient (indicated with an arrow in orange in Figure 27A) that is classified as EEC at pre-operatory level, however, it turns to be a carcinosarcoma classified as NSMP. Using our molecular signatures, we would be able to classify the patient as NEEC-Others, but it is not identified as NEEC in first place.

Regarding the determination of histological grade, in Figure 27B are shown the 22 patients which were incorrectly classified with clinical impact (*i.e.,* from low-grade to high-grade or vice versa). From those patients using our signature that allows differentiation between EEC low-grade *vs* high-grade (PROT + PROT), we can correctly determine the grade of all those 22 patients but 2 (indicated with a red arrow in Figure 27B). Specifically, the first patient is classified as high-grade (G3) at the preoperatory level using both endometrial biopsy and our molecular tool but staged at low-grade (G2) and NSMP after surgery. However, this patient recurred, which might indicate the classification of this tumor as G3 when using our protein signature. Contrary to this, the second misclassified patient, it is classified as low-grade (G2) by both pathological assessment and using the protein signature here purposed but re-classified as high-grade with the pathological examination of the resected tumor and classified as NSMP. Hence, our molecular panel correctly classified all EEC-high-grade (but the one mentioned above). Importantly, the 2-protein panel classified all EEC-high-grade with FIGO stage IA cases that were incorrectly diagnosed as low grade preoperatively by the histopathological examination and should have had a more extensive surgery.
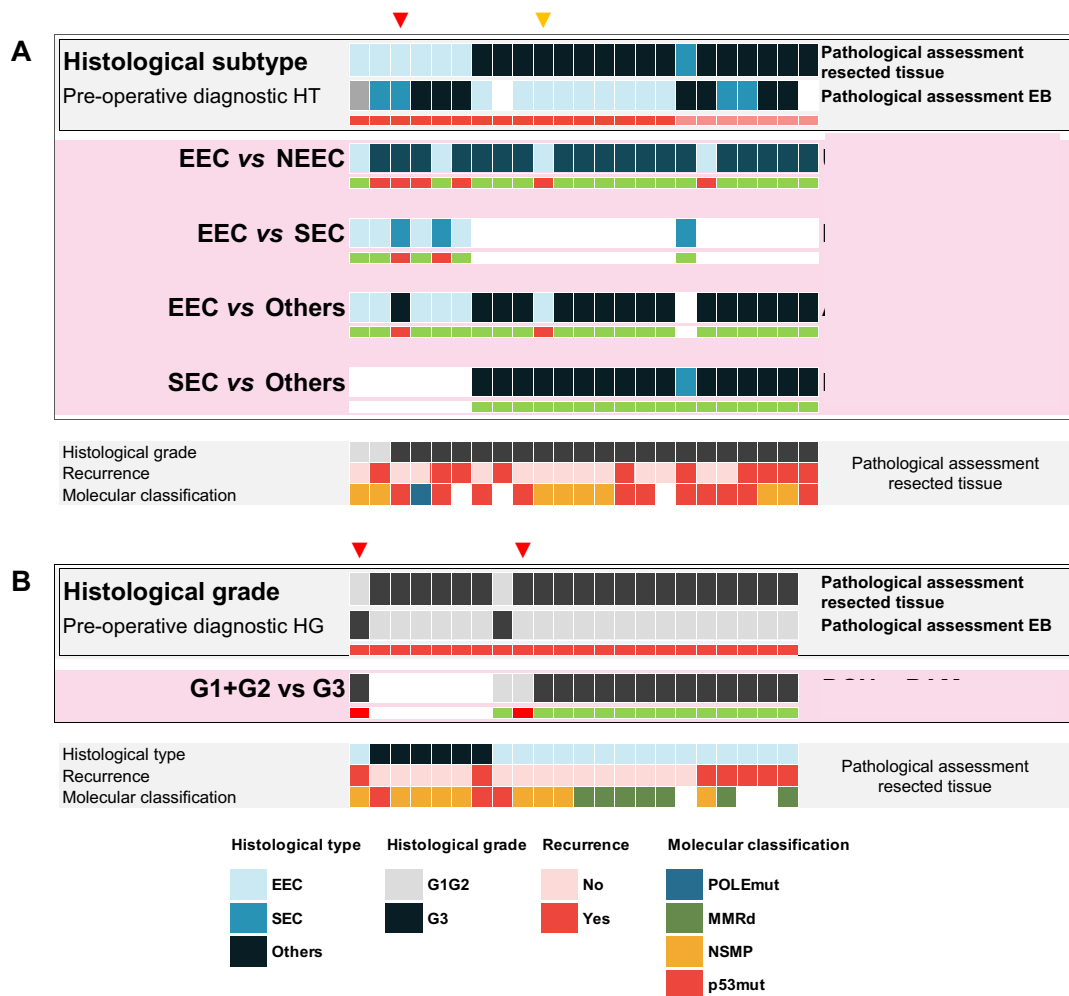
Figure 27 Pre-operative misclassified EC patients regarding histological type and grade through pathological examination of the endometrial biopsy. A comparison between pre-operative pathological diagnostic, pre-operative molecular diagnostic, and final diagnosis of tumors. The 23 patients incorrectly classified regarding histological type are shown in **A**, while the 22 patients that had an incorrect determination of the histological grade that was clinically relevant are represented in **B**, independently of A. Final diagnostic achieved through the pathological examination of the resected tissue is shown in grey together with the pre-operative diagnostic using the endometrial biopsy. In pink are shown the diagnostic results from the different 2-protein signatures identified to classify histological type and grade. In green and red are highlighted the correct or incorrect diagnostic in each case. Indicated with red arrows are the 3 patients that were misclassified at pre-operative level using the endometrial biopsy and both pathological and molecular analysis, and in orange a patient that was misclassified with pathological assessment and only correctly diagnosed with a specific protein signature. HT: histological type, HG: histological grade, EB: endometrial biopsy.

Therefore, using our molecular tools, only 4 patients could not be correctly diagnosed in comparison to the 62 patients when using only the pathological examination of the endometrial biopsy. These observations suggest that using a combination of pathological examination of the pre-operative endometrial biopsy and our protein signatures all EC patients would benefit. First, 97% of patients would have a correct diagnosis before surgery, which would optimize current management of the patient, and secondly, regarding the 3% that were not accurately diagnosed, the results obtained by both approaches combined could indicate the outcome of those patients.

117

Biomarkers to preoperatively predict the risk of recurrence in endometrial biopsies

Risk stratification of the patients aims to predict the outcome of the patients (low, intermediate, or high-risk of recurrence) to determine the most appropriate treatment for each group. As there is no system that directly predicts the risk of recurrence, current strategy is to compile as much information regarding the defined prognostic factors. Apart from the histological type and grade, myometrial invasion, cervical involvement, tumor size and LNS will be also evaluated before surgery by imaging techniques to finally classify EC patients in a specific risk group. However, these techniques lack of specificity and there is no alternative to accurately determine these parameters. Especially, LVSI, an essential parameter in risk stratification schemes, cannot be determined by imaging techniques. In this context, the definition of biomarkers that can directly predict those patients that will have a recurrence, independently of the other factors, would be a game-changer. In the cohort of 149 patients included in this study, 42 patients recurred and 107 did not have a recurrence. We compared the levels of the proteins in the UA samples from these two groups of patients and we identified a total of 9 proteins were differentially expressed (*Adjusted P* value <0.05, *Fold-Change* >|2|, AUC > 0.7) between recurrent and non-recurrent EC cases. As previously observed in other comparisons, all but PROT protein showed lower levels in more aggressive recurrent tumors compared to the less aggressive tumors that did not recur. Remarkably, from these 9 proteins, PROT, PROT, PROT, PROT, and PROT, have also shown to have an impact on patient survival when analyzed at the RNA level in the EC tumor tissues from The Cancer Genome Atlas (TCGA) [44]. More aggressive NEEC have significantly higher recurrence rates than EEC. Thus, we then evaluated proteins that could predict tumors that would recur independently of their histological type. Results revealed the potential of 79 proteins as predictors of recurrence in endometrioid tumors (*Adjusted P* value <0.05, *Fold-Change* >|2|, AUC > 0.7), being PROT and PROT the ones showing higher accuracy (AUC values of 0.81 and 0.80, respectively). Regarding non-endometrioid tumors, no protein showed statistical significance between recurrent and non-recurrent tumors, probably due to the low number of samples and the high dispersion of the data. However, we could point out some proteins as promising proteins to be further analyzed in subsequent studies. Specifically, PROT, PROT, PROT, all of them showing *Fold-Change* > |2| and AUC values of 0.77, 0.79 and 0.79, respectively. In Figure 28A the top10 proteins, ordered by AUC values, for each comparison are shown. Additionally, abundance levels of the two most promising proteins of each comparison are represented in violin plots in Figure 28B. Finally, results of all the proteins showing potential in predicting recurrence (AUC values > 0.7) are shown in Annex 2, Table A2.3 and Table 2.4 for specifications of each biomarker.

A

	Recurrence								
Gene Name	FC (Rec vs No-Rec)	Adjusted P value (Rec vs No-Rec)	AUC (Rec vs No-Rec)	FC (EEC: Rec vs No-Rec)	Adjusted P value (EEC: Rec vs No-Rec)	AUC (EEC: Rec vs No-Rec)	FC (NEEC: Rec vs No-Rec)	Adjusted P value (NEEC: Rec vs No-Rec)	AUC (NEEC: Rec vs No-Rec)
PROT	-2.05	0.00	0.78	-2.37	0.00	**0.81**	-1.37	0.93	0.67
PROT	-1.85	0.00	0.77	-1.54	0.02	0.70	-2.02	0.31	0.77
PROT	-1.76	0.00	0.74	-1.61	0.01	0.71	-1.54	0.93	0.69
PROT	-2.69	0.00	0.73	-1.84	0.03	0.70	-2.74	0.93	0.72
PROT	-3.61	0.00	0.73	-3.76	0.00	0.75	-1.60	0.93	0.59
PROT	-1.52	0.00	0.73	-1.52	0.01	0.74	-1.27	0.93	0.65
PROT	-2.33	0.00	0.73	-2.75	0.00	0.79	-1.73	0.93	0.64
PROT	-1.83	0.01	0.73	-2.01	0.01	0.74	-1.38	0.93	0.66
PROT	-3.05	0.00	0.72	-2.13	0.02	0.67	-3.00	0.93	0.72
PROT	1.84	0.01	0.72	2.10	0.00	0.76	1.33	0.93	0.60
PROT	-1.45	0.10	0.65	-2.33	0.00	**0.80**	1.29	0.93	0.55
PROT	-1.47	0.05	0.64	-2.09	0.00	0.79	1.11	0.94	0.55
PROT	-1.58	0.10	0.66	-2.58	0.00	0.78	1.23	0.93	0.50
PROT	1.50	0.07	0.67	2.06	0.00	0.77	-1.25	0.93	0.56
PROT	2.11	0.01	0.69	2.70	0.00	0.77	1.22	0.93	0.53
PROT	-1.58	0.07	0.64	-2.18	0.00	0.77	-1.05	0.98	0.55
PROT	-1.59	0.00	0.72	-1.71	0.00	0.77	-1.16	0.93	0.59
PROT	-2.07	0.02	0.70	-3.16	0.00	0.77	-1.02	0.99	0.53
PROT	1.66	0.06	0.66	1.07	0.80	0.51	2.68	0.58	0.79
PROT	1.10	0.77	0.54	-1.42	0.16	0.59	2.32	0.58	0.79
PROT	-1.11	0.82	0.51	-2.32	0.02	0.68	3.07	0.65	0.76
PROT	1.26	0.41	0.63	-1.19	0.51	0.55	2.66	0.58	0.76
PROT	1.58	0.07	0.63	1.18	0.42	0.58	3.16	0.58	0.75
PROT	-2.45	0.01	0.71	-1.98	0.06	0.65	-2.20	0.93	0.74
PROT	-1.92	0.00	0.70	-1.49	0.07	0.63	-2.11	0.84	0.74
PROT	1.20	0.12	0.65	1.09	0.44	0.58	1.24	0.93	0.66
PROT	-1.41	0.01	0.71	-1.30	0.06	0.66	-1.40	0.93	0.73
PROT	-1.17	0.05	0.67	-1.28	0.01	0.76	-1.03	0.98	0.56

B

Figure 28 Performance of proteins to discriminate between recurrent (n=42) and non-recurrent (n=107) endometrial adenocarcinomas, between recurrent endometrioid adenocarcinomas (EEC) (n=23) and non-recurrent EEC (n=81), and/or recurrent NEEC (n=19) and non-recurrent NEEC (n=26). (A) Statistical results of the 25 proteins showing an AUC value greater than 0.75 to discriminate between any of the comparisons. Highlighted in grey significant proteins showing an *Adjusted P* value <0.05, *Fold-Change* (FC) >|2| and AUC value > 0.7, and AUC values > 0.8 are indicated in bold. Highlighted in pink proteins contained in the combinations. (**B**) Violin plots picturing the distribution of the intensities obtained by DIA across patients of the best performing proteins.

Protein signatures to predict recurrence. To improve the potential of the individual proteins to predict recurrence, we evaluated the discriminatory power of all possible 2-protein combinations between the top 300 best performing individual proteins. As previously described, we chose the most promising combinations, validated them by LOOCV, and selected the one that better predicted the tumors that would recur. Overall, the most powerful combination to discriminate between recurrent and non-recurrent tumors was PROT combined with PROT, which achieved an AUC value of 82.3% (95% CI 74.1-89.4) with sensitivity and specificity of 86% and 75%, respectively (Figure 29A-D). When

considering the different histological types independently, we identified a 2-protein panel composed by PROT and PROT that was able to predict recurrence in EEC tumors with an AUC value of 87.8% (95% CI 80.2-93.6) and 91% sensitivity and 80% specificity (Figure 29B-D). Regarding NEEC tumors, we determined PROT combined with PROT as the most accurate protein signature to predict recurrence for these tumors, yielding an AUC value of 91.3% (95% CI 78.9-99.8) and 100% sensitivity and 83% specificity (Figure 29C-D).

Focusing on the relation between the use of these 2-protein predictive panels and their relation to EC molecular classification defined by the TCGA, we can observe that the information they both give is quite complementary. Regarding EC molecular classification defined by the TCGA (here referred as TCGA's molecular classification), we can observe that despite p53mut are more prompt to recur, patients classified as NSMP or MMRd also do (Figure 29A). Thus, although this classification is extremely helpful in the management of EC patients and outcome prediction, still lacks of predictive power. Remarkably, assessing our first 2-protein panel (PROT + PROT) in this dataset of patients, we can predict recurrence in all patients but 6 independently from the histological type, grade, or TCGA's molecular classification. The signature panel can identify some patients that will recur classified not only as p53mut, but also MMRd and NSMP according to the TCGA's molecular classification (Figure 29A). Interestingly, when we apply the 2-protein predictive panel obtained specifically for EEC tumors (PROT + PROT), we can identify all NSMP and p53mut tumors that will recur, only missing 1 recurrent MMRd tumors, and another tumor that could not be classified in regard to the molecular classification (Figure 29B). However, both non identified patients by our combination would be stratified as high-risk patients with current system, therefore, their treatment and follow-up would not be affected. Instead, we are identifying all low-risk patients that will recur. Similarly, in our protein signature specifically developed for NEEC tumors (PROT + PROT), is able to identify 100% of patients that will recur, including all patients classified as MMRd, NSMP, and p53mut. Therefore, we could confirm the discriminative power of our novel protein signatures to predict recurrence and highlight the added and crucial value that they give to the current TCGA's molecular classification.

When we focused on current ESMO-ESGO-ESTRO stratification system [64], it would classify 39 tumors of the dataset as patients with low risk to recur, 25 as intermediate, 36 as high-intermediate, and 44 as high. However, from those, 4 patients classified as low risk will recur, (3 of which are MMRd molecularly classified, and 1 unknown), 7 patients classified as intermediate risk (3 p53mut, 2 NSMP, 2 unknown), 8 patients classified as high-intermediate risk (1 p53mut, 3 NSMP, 3 MMRd and 1 unknown), and 23 patients with high risk (12 p53mut, 6 NSMP, 3 MMRd. 2 unknown) (Figure 29D). Here two scenarios are presented, as current recommendations state two different approaches regarding the availability to classify patients according to the TCGA's molecular classification of each hospital. First approach would be the case that any of those patients would have been classified regarding TCGA's molecular classification. still lacks of predictive power. Remarkably, assessing our first 2-protein panel (PROT + PROT) in this dataset of patients, we can predict recurrence in all patients but 6 independently from the histological type, grade, or TCGA's molecular classification. The signature panel can identify some patients that will recur classified not only as p53mut, but also MMRd and NSMP according to the

TCGA's molecular classification (Figure 29A). Interestingly, when we apply the 2-protein predictive panel obtained specifically for EEC tumors (PROT + PROT), we can identify all NSMP and p53mut tumors that will recur, only missing 1 recurrent MMRd tumors, and another tumor that could not be classified in regard to the molecular classification (Figure 29B). However, both non identified patients by our combination would be stratified as high-risk patients with current system, therefore, their treatment and follow-up would not be affected. Instead, we are identifying all low-risk patients that will recur. Similarly, in our protein signature specifically developed for NEEC tumors (PROT + PROT), is able to identify 100% of patients that will recur, including all patients classified as MMRd, NSMP, and p53mut. Therefore, we could confirm the discriminative power of our novel protein signatures to predict recurrence and highlight the added and crucial value that they give to the current TCGA's molecular classification.

When we focused on current ESMO-ESGO-ESTRO stratification system [64], it would classify 39 tumors of the dataset as patients with low risk to recur, 25 as intermediate, 36 as high-intermediate, and 44 as high. However, from those, 4 patients classified as low risk will recur, (3 of which are MMRd molecularly classified, and 1 unknown), 7 patients classified as intermediate risk (3 p53mut, 2 NSMP, 2 unknown), 8 patients classified as high-intermediate risk (1 p53mut, 3 NSMP, 3 MMRd and 1 unknown), and 23 patients with high risk (12 p53mut, 6 NSMP, 3 MMRd. 2 unknown) (Figure 29D). Here two scenarios are presented, as current recommendations state two different approaches regarding the availability to classify patients according to the TCGA's molecular classification of each hospital. First approach would be the case that any of those patients would have been classified regarding TCGA's molecular classification. In this scenario, the 4 patients classified as low risk, would not have the most optimal surgery and adjuvant treatment; the 7 and 8 patients classified as intermediate or high-intermediate risk patients respectively, might not have had the most suitable adjuvant treatment for them, as omission of any adjuvant treatment is still an option. Thus, only the 23 patients classified as high-risk would have been properly treated. The second scenario, would be that where patients had access to TCGA's molecular classification (in our dataset, 36 patients that recurred had this information). In this case, all patients would have had the same risk assessment regardless of the TCGA's molecular classification. However, there would be differences in the adjuvant treatment of the 3 patients and 1 patient classified as intermediate risk and high-intermediate, respectively, but with p53 mutated, as in those cases specific recommendations are given. The reality in patients from our dataset was that only 1 patient (out of 4) classified as low-risk that recurred was treated with adjuvant treatment; only 1 patient (out of 7) classified as intermediate risk was not treated with adjuvant treatment; all patients classified as high-intermediate risk and high-risk patients were treated with adjuvant treatment.

Using our protein panels, we would identify 100% of the patients that will recur at pre-operatory level only using the fluid of the current endometrial biopsies. This would provide a highly valuable information related to those patients that would benefit either an extended surgery or adjuvant treatment (Figure 29D). The false positives from our protein signatures are in 68%, 50% and 75% of cases high-risk or high-intermediate risk patients. Thus, adding this information to the one already being in consideration under the current guidelines, we could provide a more personalized treatment for all EC patients.

A **B** **C**

D

Comparison	49 Rec vs 101 NoREC	EEC (23 Rec vs 78 NoREC)	NEEC (19 REC vs 23 NoREC)
Protein 1			
Protein 2			
TP	36	21	19
TN	76	62	19
FP	25	16	4
FN	6	2	0
Specificity (%)	75.2 (95%CI 66.3-84.2)	79.5 (95%CI 70.5-88.5)	82.6 (95%CI 65.2-95.7)
Sensitivity (%)	85.7 (95%CI 76.2-95.2)	91.3 (95%CI 78.3-100)	100 (95%CI 100-100)
AUC (%)	82.3 (95%CI 74.1-89.4)	87.8 (95%CI 80.2-93.6)	91.3 (95%CI 78.9-99.8)

E
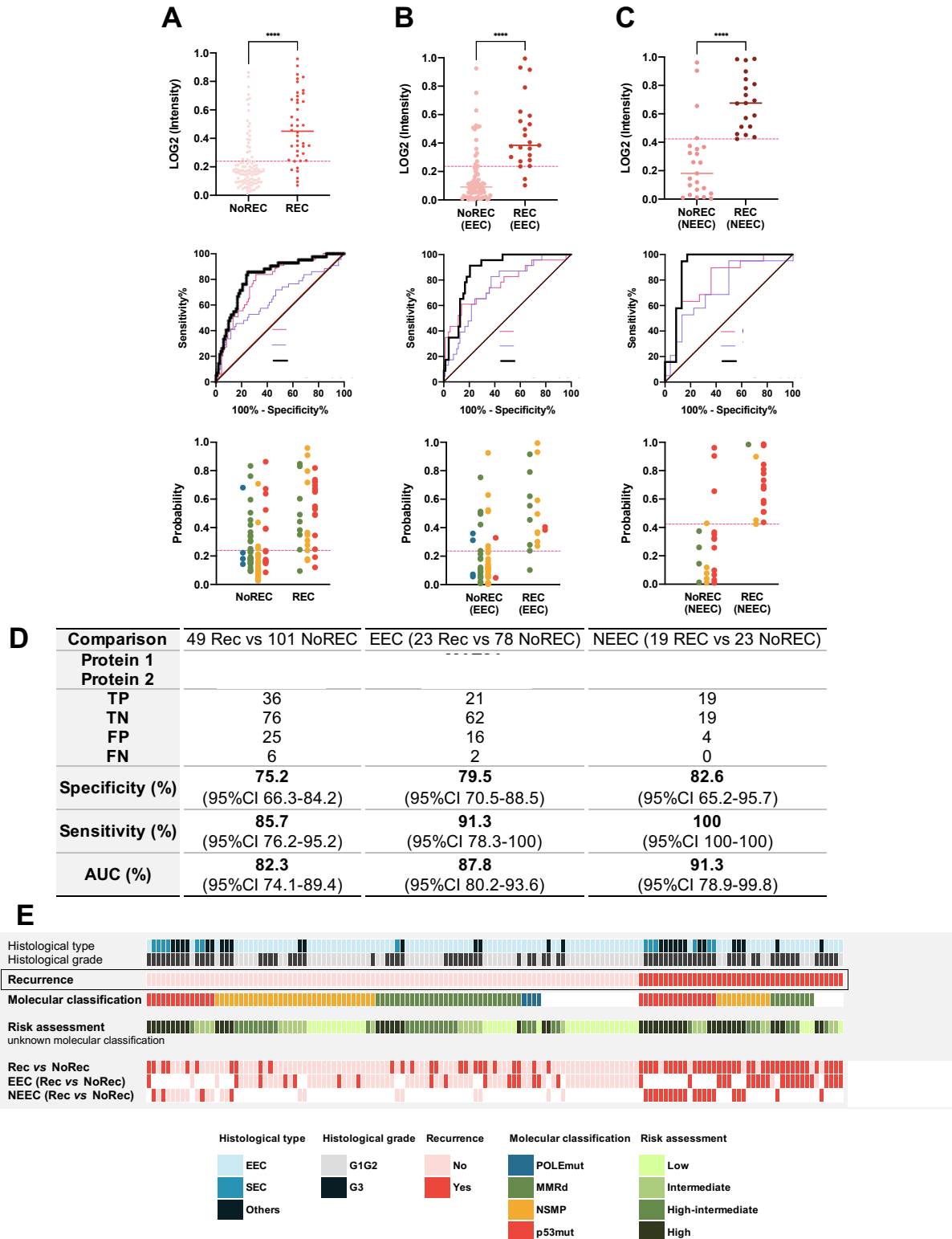
Figure 29 Performance of the protein signatures to discriminate between recurrent (n=42) and non-recurrent (n=107) endometrial adenocarcinomas, between recurrent endometrioid adenocarcinomas (EEC) (n=23) and non-recurrent EEC (n=81), and/or recurrent NEEC (n=19) and non-recurrent NEEC (n=26). Comparison between the accuracy of those panels to the accuracy of the EC molecular classification and current ESMO-ESGO-ESTRO risk stratification system. For each comparison and protein signature different elements are shown: (i) Scattering plot representing the probabilities of patients to be correctly classified regarding recurrence with the specific 2-protein panel (in pink the cut-off value); (ii) ROC curves with the individual proteins

of the panel and the 2-protein panels; (ii) and tables with sensitivity (Sn), Specificity (Sp), PPV (positive predictive value), and NPV (negative predictive value). The protein panel composed by PROT + PROT classifying recurrent (n=42) and non-recurrent (n=107) patients is shown in (A), the protein panel composed by PROT + PROT predicting recurrence in EEC (n=23 recurrent *vs* n=81 non-recurrent) are shown in panel (B), and the protein panel formed by PROT + PROT able to distinguish NEEC that will recur (n=19) *vs* NEEC that won't recur (n=26) is represented in panel (C). (D) Tables with true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), % specificity, % sensitivity, AUC values, and their respective 95% confidence intervals (CI) for each comparison. (E) Overview of the patients included in the clinical study (n=149) with data regarding recurrence, ordered by this variable, as well as TCGA's molecular classification, and ESMO-ESGO-ESTRO risk stratification system. Information regarding histological type, histological grade, and predictive power of our protein signatures is also given.

In-silico validation

Our research led to point out a total number of 175 proteins that are potential prognostic biomarkers when measured in the fluid of uterine aspirates, and they might play a role in EC. Specifically, we can highlight 22 proteins that were highlighted in more than one parameter here measured and studied (histological type, histological grade, and recurrence). Nine proteins presented significant differences in abundance between some comparison regarding histological type and histological grade (PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT), the levels of abundance of seven proteins were significantly different between some comparison of histological type and recurrence (PROT, PROT, PROT, PROT, PROT, PROT, PROT), and three proteins presented different abundant levels between some comparison regarding histological grade and recurrence (PROT, PROT, PROT). Finally, 3 proteins appeared to be significant for all the measured parameters: PROT, PROT, and PROT.

In order to further validate our biomarkers, we first compared the significant proteins identified to the proteins described as potential prognostic in EC biomarkers [57]. As previously mentioned, we could identify 125 of those proteins in our clinical study (Figure 24B), and from those, finally 15 were validated as potential biomarkers also in the fluid of pipelle biopsies (PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, and PROT). Secondly, we specifically compared our significant proteins to the 30 *in-silico* validated biomarkers using CPTAC and TCGA datasets out from the 255 identified in the review of protein prognostic biomarkers [335]. From those proteins, 22 were identified in our spectral library from the pool of the fluid of UAs of EC patients, and we were able to measure 19 of those in our clinical study (n=149). Finally, only 3 of them (PROT, PROT, and PROT) were in common with our significant proteins. Specifically, PROT was validated for histological grade and overall survival, while in our clinical study we found PROT to be differentially expressed between histological types (EEC and NEEC) and recurrent and non-recurrent patients. In accordance with our study, PROT demonstrated to play a role in histological type. Finally, PROT was mentioned to be able to distinguish between tumor histological types and molecular classification, while in our study was validated only for histological type.

Therefore, we could confirm that uterine aspirates are great source of EC biomarkers not only for ruling out EC from non-EC patients [200], but also to diagnose the tumor features and predict recurrence.

DISCUSSION

Gynecological fluids demonstrated to be a potential source of EC biomarkers [200,201,205,337–340]. They are of great value in biomarker research, as they are in direct contact or close to the tumor and therefore, they have the potential to facilitate the discovery of relevant biomarkers. Specifically, pipelle biopsies are a complex mixture of cells, tissue, blood, and other biological molecules from the different parts of the uterus. Importantly, they demonstrated to be reliable surrogates of the primary tumor reporting a gene expression correlation of $R^2 = 0.98$ in a set of 20 genes [199], and further studies of the same group defined a diagnostic algorithm composed by 5 genes achieving a sensitivity of 81% and specificity of 96% [201]. Moreover, Martinez-Garcia E et al. measured the levels of 52 described EC protein biomarkers in the fluid of the pipelle biopsies resulting in protein signatures able to accurately diagnose EC patients (94% sensitivity, 87% specificity) and to accurately determinate histological subtypes (95% sensitivity, 96% specificity).

To our knowledge, our uterine fluid spectral library is the first and larger and most comprehensive catalogue of proteins in these samples of 42 women diagnosed with EC, covering the full spectrum of EC (from different subtypes, grades, and molecular classification subgroups, to early-stage and late-stage EC). We generated a unique spectral library of 54,448 peptides belonging to 5,863 proteins detected in the pooled liquid fraction of uterine fluid samples fractionated in 59 fractions, each of them analyzed by DDA. Our library was validated as it covered most of the key proteins and pathways involved in cancer such as Wnt/beta-catenin, PI3K/Akt or P53 pathways, all of them important deregulated pathways in EC [45,336]. Additionally, 96% of proteins identified in our library overlapped with the proteins identified in EC FFPE-tissues from the CPTAC cohort [298], representing the 51% of all the nearly 11,000 identifications they had. Importantly, 62% of the 255 prognostic proteins already described in EC were also contained in our library [57]. Hence, our spectral library covered a broad range of proteins, not only secreted or extracellular, but also from inside the cells, demonstrating the potential of uterine fluids as minimally-invasive source of EC biomarkers, which could also be extrapolated to other gynecological diseases. The availability of this spectral library will also benefit the scientific community in the field of EC biomarker research.

Current risk stratification system of EC enables proper management of EC patients improving prognostic and reducing unnecessary adjuvant treatment. Surgery treatment is the first line treatment of EC, which is determined by the preoperative risk-assessment. However, the risk stratification at preoperative level is one of the biggest challenges in EC treatment, as it lacks of accuracy. In this study we present an endometrial biopsy by pipelle-based approach complementing the current stratification methods. Though the measurement of different prognostic protein signatures, we are able to provide additional and more accurate information regarding histological type and grade, as well as to provide valuable information related to the probability of a tumor to recur. Thus, providing a personalized and accurate diagnosis to all EC patients.

Preoperative risk staging is given by the determination of different prognostic factors measured by the histopathological examination of an endometrial biopsy and imaging techniques. Histological type and grade of the tumor are determined by the histopathological observation of the tumoral cells contained in the endometrial biopsies. However, it is known that considerable discrepancies exist between preoperative and postoperative determination of these parameters, mainly due to histologically inadequate specimens. Specifically, subtyping showed 7% of discrepancy, while grading 33% [136]. In our dataset of patients (n=149), 15% and 35% of the patients were inaccurately diagnosed regarding histological type and grade, respectively, independently of the acquisition method used (pipelle biopsy or hysteroscopy). Stinkingly, the 13% and the 17% of the patients included in the study not only had an incorrect determination of the histological type and grade, respectively, but also the inaccurate diagnosis at pre-operatory level was of clinical significance to plan the surgery treatment. To solve this issue, we aimed to identify protein biomarkers that could determine the histological type of the tumor, as well as its histological grade in the fluid of minimally-invasive pipelle biopsies. Thus, allowing the integration of the results obtained by the pathologists (using the tissue contained in the biopsy) and our molecular tool.

Regarding histological type, results of the study revealed 34 proteins that could significantly distinguish between EEC and NEEC tumors, being the PROT, PROT, PROT the best performing proteins for the purpose. Specifically comparing EEC and the most common NEEC histology (SEC) tumors, we identified 9 proteins with significantly different levels of abundance, where again PROT, PROT but also PROT were the best performing proteins for the determination of the subtypes. When determining proteins that can significantly distinguish between EEC and Other subtypes (including NEEC histologies but SEC: clear-cell, mixed, mucinous, undifferentiated and carcinosarcomas) 25 proteins were identified, being PROT, PROT and PROT highlighted by their potential. Lastly, to compare SEC histology to Other NEEC histologies, PROT and PROT showed discrimination power. Overall, seems that PROT, PROT, and PROT are the main players in distinguishing EEC, SEC, and other NEEC histologies (excluding SEC). Biologically, PROT, PROT, and PROT

To improve the accuracy of the individual proteins, we generated 2-protein panels for each comparison. The combination of PROT and PROT showed an AUC value of 89.4% (95%CI 83.2-94.5) to discriminate EEC from NEEC tumors. Combining PROT and PROT we achieved an AUC value of 95.6% (95%CI 90.4-98.9) and 100% sensitivity to discriminate EEC from SEC patients. To compare EEC *vs* Other NEEC histologies (excluding SEC), we used PROT and PROT achieving 89.3% (95%CI 83.5-94.4) of AUC. Finally, when combining PROT and PROT, 97.7% (95%CI 91.1-100) of AUC value. Remarkably, from the 23 tumors incorrectly diagnosed for the histological type, using our purposed protein signatures, we could correctly classify 21 of them.

Regarding histological grade, we were able to identify 14 proteins, which levels of expression are different between low-grade and high-grade EEC. Specifically, PROT, PROT and PROT showed the best performance. Biologically, PROT and PROT.

Improved and more accurate diagnosis of histological grades was achieved when combining PROT with PROT, obtaining an AUC value of 85.3% (95%CI 75.9-93.1). Remarkably, 47 tumors were incorrectly diagnosed at preoperatory level by the histopathological examination in regard to the histological grade. The misclassification of 22 of them had a clinical impact on the pre-operative risk assessment. Using our protein panel, we could classify 19 of those 22 misclassified tumors. Therefore, combining current assessment with our approach we could accurately diagnose 96% of patients regarding histological grade.

Histopathological study of the endometrial biopsies to determine histological type and grade is of huge importance to assess the risk of each patient [341]. However, inaccuracy in the determination of these prognostic factors is evidenced by different studies, especially in endometrial biopsies where there is few material to assess or in high-grade EC [86,136]. Our signatures addressed to determine those two prognostic factors are highly accurate, in particular, in defining between EEC and SEC, EEC and NEEC (excluding SEC), and SEC vs other NEEC histologies. Consecutively, the use of the signature grading EEC would be applied significantly improving current grading system only using the histopathological observation.

Another fundamental cornerstone in the risk assessment of EC patients is FIGO stage. At preoperatory level it is determined by the assessment of different factors such as depth of myometrial invasion, cervical involvement, and tumor size determined by transvaginal ultrasonography or magnetic resonance imaging, while LNS is evaluated by computed tomography, magnetic resonance imaging (MRI) or positron emission tomography. However, these techniques lack of sensitivity and fails to identify small lesions [342,343].

In the era of molecular classification of EC, the prognostic value of this classification has been stablished in the recent ESGO-ESTRO-ESP guidelines [64], as great efforts have been made to develop pragmatic substitutes to facilitate the clinical use of the molecular classification, such as the ProMisE algorithm, which allows to molecularly classify EC patients using FFPE tissues to assess four mismatch repair (MMR) proteins (MLH1, MSH2, MSH6 and PMS2) and p53 by immunohistochemistry (IHC) and the sequencing of POLE. Moreover, this approach demonstrated to have great performance also using endometrial biopsies [157,158], thus, being able to dispose of this information at preoperatory stage. However, different concerns are presented towards the molecular classification. Regarding POLE assessment, the prognostic significance of POLEmut subgroup still lacks of information, as studies performed had small number of patients falling in this group [61,344] . Additionally, 17% of POLEmut patients still recur after a median follow-up of 30 months [345], and the prognostic value of POLEmut tumors when the disease is found in an advanced stage is not clear. Besides, despite new technologies are being developed to easily assess hotspots mutations in POLE [346], these algorithms still requires sequencing not available in all hospitals. Regarding MMRd patients, they show better outcome than expected compared to MMR proficient tumors, as MMRd are often associated to poor prognostic indicators as high-grade and/or LVSI [347]. The largest number of EC patient dive into NSMP subgroup,

which mainly includes patients classified as low- to intermediate-risk. However, it also includes high-risk tumors, which prognostic is still unclear [78,80,348].

Under this scenario, it is clear that novel prognostic biomarkers to improve the current information are needed. In this sense, widely described biomarkers in EC have been studied. Vrede *et al.* recently studied the IHC abnormal expression of p53, L1 cell-adhesion molecule (L1CAM), estrogen receptor (ER) and progesterone receptor (PR) in relation to the ESMO-ESGO-ESTRO risk groups. They showed that their abnormal expression was highly related to higher risk classification groups, demonstrating their potential at preoperatory level to guide primary treatment, as well as adjuvant treatment [349]. Also, Vermij *et al.* studied HER2 and its prognostic value, suggesting that molecular subclassification testing of HER2 is superior to histotype-directed testing [350]. However, all these markers require of tissue, as they are measured through IHC. Only evidences of elevated levels of cancer antigen 125 (CA125) detected in blood could serve as moderate indicators in risk of LNM [57,351]. Hence, prognostic biomarkers in liquid biopsies are demanded.

In this study, we seek to use our approach to identify protein biomarkers in the liquid of preoperative pipelle biopsies that could predict recurrence in EC patients. The clinical study here presented included 42 recurrent patients and 107 non-recurrent patients. By analyzing those patients, we could define a total number of 6 proteins that accurately assess whether an EC will recur or not. Those proteins are PROT, PROT, PROT, PROT, PROT and PROT. Furthermore, when we assessed recurrence depending on the histological type, we identified 12 potential proteins that could indicate the recurrence of EEC patients, being PROT and PROT the two best performing, while any protein shown statistically significant differences between recurrent and non-recurrent tumors classified as NEEC. Nonetheless, PRDX4 presented a fold change higher than 2 and AUC value of 0.79. Biologically, PROT and PROT.

To better determine the prognosis of EC patients, we developed 2-protein panels that were able to predict which patients would recur and we compared their performance with the molecular classification. The first signature achieved an AUC value of 82.3% (95%CI 74.1-89.4) and it is composed by PROT and PROT. It allowed us to identify all but 6 recurrent patients, independently of the histological type, grade, or molecular classification. Thus, completely complementing current given information from the endometrial biopsy only by using the fluid that is contained. When we first classified our patients regarding their histological type, we identified a second signature related to EEC composed by PROT and PROT achieving an AUC value of 87.8% (95%CI 80.2-93.6) and sensitivity of 91.2%. It identifies all but 2 recurrent patients. Importantly, these 2 patients were classified as high-risk EC patients, thus, not identifying those patients would not affect their treatment. A final signature was developed to identify recurrent patients in NEEC. It was composed by PROT and PROT and it yielded an AUC value of 91.3% (95%CI 78.9-99.8) with a 100% sensitivity. Therefore, it was able to identify 100% of tumors that recur, independently of their molecular classification. The three signatures here presented able to identify recurrent patients at preoperatory level gives a highly valuable information that it is currently not available in the clinical practice.

Remarkably, despite the information given by the molecular classification includes all histological types and grades of EC, it presents certain overlap with the different histological subtypes. The POLEmut and MMRd subgroups are mostly composed by EECs, while the NSMP subgroup is mainly formed by low-grade EECs, and the p53mut subgroup is mainly composed by SECs. Other non-endoemtrioid histologies such as carcinosarcomas are usually found together with SECs in the p53mut subgroup, whereas clear-cell EC fall into the p53mut in 60% of cases, and into the NSMP in 40% of cases [352,353]. The potential of our signatures is that, unlike molecular classification, the knowledge provided by these signatures is independent and not overlapping to any other prognostic factor, which makes them advantageous in regard to other potential assets. Additionally, molecular classification defines treatment strategies in block, and don't focus in personalized medicine, while our protein signatures are measured independently and offer a broad range of usage opportunities.

Importantly, most of EC patients are diagnosed at early stages of the disease when the tumor is still confined in the uterus, and they can be classified as low-risk. In these patients, a total hysterectomy and bilateral salpingo-oophorectomy is performed [64]. However, in our clinical study and in accordance to the results previously reported by other researchers [162,354], 10% of low-risk patients recurred, and 45% of the recurrent patients did not fall into the high-risk group. Meaning, that there was occult metastatic spread not detected by imaging techniques [77,355]. In fact, a recent study by *Daix et al.,* they showed that the concordance between preoperative risk classification and final histology is weak in early-stage EC. Specifically, they reported underestimation and overestimation in 37% and 10% of cases, respectively, attributed exclusively to the endometrial biopsy results in 18% of cases, while 7% to the results of the biopsy in combination with the ones obtained through MRI [356]. In the study, they recorded 30.4% of an overall disagreement regarding the results of the endometrial biopsy. Hence, the potential of the tools here presented to accurately determine histological type and grade in these biopsies is also evinced.

Nonetheless, there is still a need to identify those undiagnosed small metastatic spreads to better assess FIGO stage. To solve this issue, current strategies rely on the selective lymphadenectomy. Some recent studies reported an average sensitivity of 96.5% to detect nodal metastasis through the assessment of sentinel lymph node (SLN) biopsy, and trends seem to be shedding towards this approach [357–359]. In this regard, Aboulouard *et al.* bet for the in-depth proteomic study of SLN for patients with EC, using the same novel proteomic approach as we did, to identify five proteins (PRSS3, PTX3, ASS1, ALDH2, and ANXA1) that were further validated by IHC to improve the stratification and diagnosis of EC patients [261]. However, underevaluation of the para-aortic nodes when assessing the SLN remains controversially, particularly in high-risk cancers, and the interpretation of the results it is not clear. But if the results obtained in our clinical study were validated, our protein panels could serve as a selective manner to choose lymphadenectomy before surgery in low-risk preoperative patients, which could be confirmed intra-operatory by the SLN.

An important strength of this study is the technology used. We employed a newly stablished massively parallel targeted proteomics technique to analyze our uterine fluids. The DIA-based technique permits

a highly comprehensive and reproducible analysis in complex biological samples, such as uterine fluids, by enabling a highly precise and accurate label-free quantification of the peptides contained in the samples [245]. Here we demonstrate the first protein quantification in uterine fluids from EC patients by using our own spectral library. In comparison to other high-throughput proteomic studies in EC [203], the study here described provides of not only sample high-throughput but also quantitative precision and proteome coverage in a large sample set [360], allowing the identification of novel potential biomarkers previously not described that could lead to a better management of EC patients.

Despite of the clinical strength of the investigation, where women enrolled covered the broad variability of EC tumors (including a wide variety of tumor and grade histologies, molecular subtypes, risk subgroups and recurrent patients), the study was, however, limited by the number of patients included in each group. Most of the times, we had to face and be aware of statistical problems related to unbalance and/or small groups, such as POLEmut subgroup. Nonetheless, we overcome these limitations by, for instance, limiting the number of biomarkers included in our protein panels, which is beneficial at the same time for clinical application. Moreover, we are aware that further validation of these findings is required in a larger, multicentric and independent cohort of patients and tackle further applications with our approach in other clinical gaps such as differentiation of POLEmut group with protein panels. Aiming to implement the identified biomarkers in the clinical practice, another limitation of this research is the need of testing the identified biomarkers in a more accessible technique for clinical laboratories (*i.e.* ELISA assay), as they are the preferred method not only for clinical validation but also for application in clinics [361]. Last generation of MS is advantageous in discovery studies, and despite the efforts of the mass-spectrometry community to implement clinical proteomics, the timings are still unclear.

A relevant strength of this study is that all the prognostic protein signatures developed in our investigations are measured in the fluid of endometrial biopsies. Recently, Bratulic *et al.* discussed how the assessment of molecular biomarkers in liquid biopies could change the current management in oncology and highlights the increasing interest of researches towards these type of investigations [362]. Hence, we are convinced that our approach would revolutionize the current diagnosis of EC, as it is completely independent of the pathological examination and, therefore, from the amount of tissue present in the sample. Protein signatures would not exclude pathological examination, on the contrary, they would help in the understanding of tumors and managing of patients, while solving current inaccuracy in the preoperative diagnosis. Furthermore, it provides of valuable information regarding the outcome of the patients independently of other strategies. All of it only with the usage of the fluidic part of the minimally-invasive pipelle biopsies, which is nowadays discarded, avoiding the invasiveness of tissue biopsies. In this regard, Eriksson *et al.* purposed to combine ProMisE classification with sonographic and demographic characteristics to preoperatively predict recurrence achieving similar AUC values than in our study (89%) [363], however, the measurement of variables such as waist circumference and sonographic tumor extension and size could be biased by the physician performing the measurements. While our measurements rely on quantitative and standardized techniques that precisely measure the specific biomarkers. However, for our biomarkers to be in the clinical practice,

clinical validity and utility will have to be demonstrated, which is one of the main challenges for biomarkers measured in liquid biopsies to move into the clinical practice.

Certainly, the determination of molecular profile of EC patients increases the diagnostic and prognostic precision. However, there are some clinical gaps remaining in the management of EC patients in regard to the molecular subgroups, and despite it has impact on the adjuvant treatment, yet its determination does not influence in surgery. Thus, it is reasonable to affirm that to give precise a personalized diagnosis and treatment to each EC patient, an integration of molecular classification, histopathological examination, prognostic biomarkers, and FIGO stage is required to offer a solution to current clinical gaps. This study was designed to approach as much of those gaps by the development of protein signatures. The outcome provided helps to improve the incorrect diagnosis of histological type and grade currently occurring preoperatory with the histopathological examination of the endometrial biopsy, and to provide with highly sensitive protein biomarker signatures to predict recurrence of EC patients before surgery and independently of the molecular classification. Hence, adding valuable information at preoperatory level. Despite the wide acceptance of the molecular classification and its impact on the adjuvant treatment, its determination does not influence in surgery.
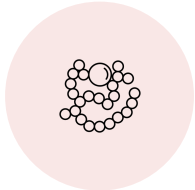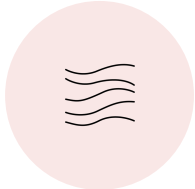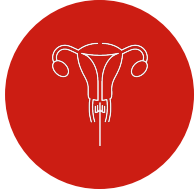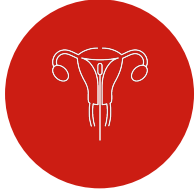
CHAPTER 3

Potential of five different cervical fluids as a source of protein biomarkers for endometrial cancer diagnosis

SPECIFIC BACKGROUND

As seen in Chapters 1 and 2, determination of prognostic factors and recurrence prediction is crucial to manage EC patients. However, identifying those patients can be extremely challenging. In the current clinical practice when a woman presents the classic symptoms of EC, as AUB is the most common symptom present in 90% of EC cases, an invasive biopsy is performed. However, more than 90% of postmenopausal and more than 98% of premenopausal and perimenopausal women with AUB have a benign underlying cause, and only a small proportion of these women will be diagnosed with EC [364]. Therefore, the development and implementation of a non-invasive test to diagnose EC is urgently needed. Such a tool could also be of extreme utility for asymptomatic women at increased risk of developing EC, including women treated with tamoxifen, women with Lynch syndrome, obesity, or diabetes.

In Chapter 3, we aim to evaluate different methods of cervical sample collection for the identification of protein biomarkers to diagnose EC. To do so, we first evaluated the protein quantity, and the proteome of cervical samples collected using five different devices and methodologies. The cervical proteome was compared with the proteome of pipelle biopsies. Finally, we assessed the expression of 52 known EC diagnostic biomarkers in the two best performing cervical sampling techniques and pipelle biopsies of 22 EC and 19 non-EC patients to decipher their potential use for EC diagnosis using protein biomarkers.

MATERIAL AND METHODS

Patient recruitment and sample collection

A total of 45 patients were recruited in this study. Informed consent forms, approved by the Ethical Committee of the Vall Hebron Hospital, were signed by the patients (PRAMI276-2018). All patients were women attending to the Gynecology Department in Vall Hebron Hospital due to the presence of an AUB, and they underwent the EC diagnostic process. Pipelle biopsies and five different exo- and endocervical samples were collected from 4 women (2 EC and 2 non-EC) for the proteomic characterization and comparison of uterine and cervical fluids. Once the two best performing cervical samplings were selected, pipelle biopsies and these two cervical samples were obtained from 41 women (22 EC and 19 non-EC) for the verification of EC biomarkers. The clinicopathological characteristics of all the patients included in this study are shown in Table 16.

		Method optimization		Verification phase	
		EC (n=2)	non-EC (n=2)	EC (n=22)	non-EC (n=19)
Age (years)					
	Mean	69	50	70	60
	Minimum	63	48	49	23
	Maximum	75	51	93	88
Uterine condition					
	Premenopausal	-	1	1	3
	Postmenopausal	2	1	21	16
Histological type					
	Endometrioid	2		18	
	Serous	-		3	
	Others (carcinosarcoma)	-		1	
Histological grade					
	Grade 1	-		7	
	Grade 2	1		7	
	Grade 3	1		8	
FIGO stage					
	IA	1		9	
	IB	1		5	
	II	-		4	
	IIIC2	-		3	
	IVB	-		1	
Miometrial invasion					
	<50%	1		13	
	>50%	1		9	
Lymphobascular invasion					
	Yes	-		7	
	No	2		15	
Samples collected					
	Pipelle biopsy (PB)	PB, PBdepl		PB	
	Cytological samples	M1, M2, M3, M4, M5		M1, M3	

Table 16. Clinicopathological features of the patients included in the study. EC: endometrial cancer; depl: depleted sample from Albumin and IgGs.

For each patient, cervical samplings were obtained first (M1 > M2 > M3 > M4 > M5) followed by pipelle biopsy. The five different cervical samples (M1-M5) were collected with five different brushes (Figure 30A-B):

- M1 was obtained with the Rovers Cervex Brush ® (Rovers Medical Devices, The Netherlands). This cervical brush is used to obtain the samples for cervical liquid cytology. It has a shape designed to obtain a good representation of endocervical and exocervical material.

- M2 was obtained with the Wooden cervical scrape or Ayres spatula (Goodwood medical care, China). It is generally used to obtain the exocervical representation of the pap-smears. It can also be used to obtain a vaginal sample, but it was not used for this purpose in the present study.

- M3 was taken with the endocervical swab HC2 DNA collection device Digene (QIAGEN, Germany) used to get an endocervical mucus sample. It is generally used to perform hybrid capture test to rule out human papillomavirus infections of the low genital tract. It has little bristles that are introduced in the endocervix to collect the mucus.

- M4 was obtained with a cotton swab (Deltalab, Spain), usually used to take superficial samples to perform bacterial cultures of any location. It is blunt and only impregnates with the secretions of the endocervix without scratching the tissue.

- M5 was taken using an endocervical brush (Bexen medical, Spain). It is the tool used to obtain an endocervical representation in pap-smears.
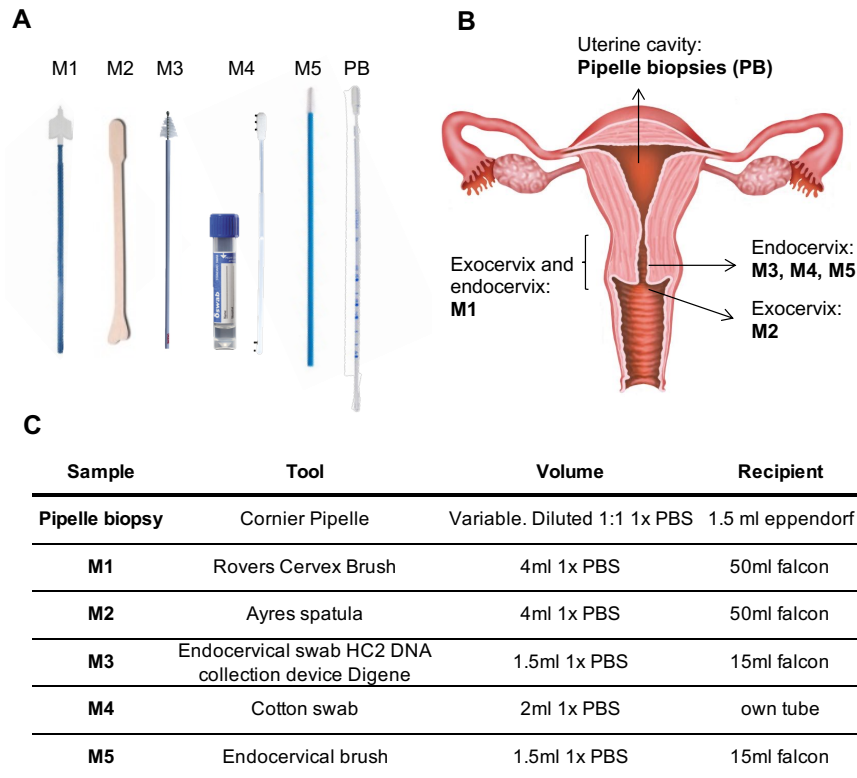
Figure 30. Clinical samples evaluated in the study. (**A**) Medical tools used to collect the samples: brushes (for M1-M5) and Cornier Pipelle (for Pipelle Biopsies - PB); (**B**) Location site for the collection of the different samples; (**C**) Description of the methodology (tools, volume, and recipient) followed to collect the samples.

After sample collection, all brushes were dipped into falcon tubes containing different volumes of PBS, which are specified in Figure 30C. Pipelle biopsies were obtained with the Cornier Pipelle (Eurogine Ref. 03040200, Spain). The device was introduced through the cervical canal into the uterine cavity, and the pipelle biopsy was obtained after applying negative pressure. This device requires cervical permeability to access the uterine cavity. Once collected, PBS was added in a 1:1 volume. All cervical and uterine samples were centrifuged at 2500g for 20 minutes to separate the fluid from the cellular fraction. Uterine fluid obtained from pipelle biopsies and cervical fluids were stored at -80ºC until used.

Sample preparation

Uterine fluids were analyzed without and with depletion of albumin and IgGs (raw and depleted respectively). For depletion, supernatants were sonicated (Labsonic M, Sartorius Stedim Biotech) during 5 cycles of 5 seconds, and 50 µl of each sample were processed with the Albumin & IgG depletion spin trap kit (GE Healthcare) according to the manufacturer's instructions.

Total protein concentration of all samples was measured by the Bradford assay. Equal amounts of protein per sample were used for the proteomic analysis. First, samples were denatured by addition of

urea to a final concentration of 6 M, incubated at 22°C under agitation for 20 min, and followed by 10 min incubation in an ultrasonic bath (Branson 5510, Branson Ultrasonics). Proteins were then reduced in 5 mM dithiothreitol (DTT) for 1 hour at 37°C and alkylated in 15mM iodoacetamide (IAA) for 30 min at 22°C in the dark. Each sample was then digested first with LysC protease (Wako) at a protease/total protein amount ratio of 1/50 (w/w) for 4h at 37°C, and then with trypsin (Promega) overnight at 37°C at an enzyme/substrate ratio of 1/25 (w/w) after dilution of urea to a final concentration of 1 M. Proteolysis was stopped by addition of 1 µl of formic acid per 100 µl of solution. For the LC-PRM studies, samples were digested with 1/150 (w/w) ratio of LysC and 1/50 (w/w) ratio of trypsin, proteolysis was stopped with formic acid, and a mixture of the stable isotope labeled synthetic peptides (Thermo Fisher Scientific, crude quality) was spiked in each sample (C terminal arginine 13C6, 15N4, C terminal lysine 13C6, 15N2, or when it was not applicable with a heavy leucine 13C6, 15N1, or phenylalanine 13C9, 15N1). All samples were desalted onto solid phase extraction cartridges (Sep Pak tC18, 25mg, Waters) and dried using a vacuum centrifuge.

DDA analysis on a tims-TOF pro mass spectrometer

Seven different matrices corresponding to 5 cervical fluids, and raw and depleted uterine fluids from four different patients (n=28 samples) were analyzed by nano-UHPLC (nanoElute, Bruker Daltonics) coupled to tims-TOF pro mass spectrometer. Samples were directly injected onto a reverse phase column (250 mm × 75µm, 1.6 µm, C18; IonOptiks) heated at 50°C. Mobile phases consisted of 0.1% (v/v) formic acid in water (phase A) and in acetonitrile (phase B). Samples were separated by a 100 min stepped gradient ranging from 2-30% B at a flow rate of 400 nl/min. The nano-UHPLC was coupled with a tims-TOF pro instrument (Bruker Daltonics) operated in dda-PASEF mode. Survey scan were acquired from 100 to 1700 m/z within an ion moblity range of 0.6 to 1.6.s/cm2. Ion mobility accumulation and separation time were both set to 100ms. Mass spectrometry (MS)/MS acquisition scheme was set to PASEF mode with 10 PASEF event per MS cycle (total cycle time 1.15s). Dynamic exclusion of fragmented ion precursors was set to 0.4min.

DDA data processing

MS files were analyzed in the MaxQuant software version 2.0.1.0 [365]. MS/MS spectra were searched by the Andromeda search engine against the TrEMBL UniProt Homo sapiens (August 2021, 78,139 entries). Trypsin/P was specified as the protease and a maximum of three missed cleavages was allowed. Carbamidomethyl (C) was set as a fixed modification and acetyl (protein N-terminus) and oxidation (M) were set as variable modifications. The "match between runs" option was enabled. The false discovery rate (FDR) for peptide and protein identifications was set to 1%. For protein quantification, label-free quantitation (LFQ) was performed with a minimum ratio count of 2 [366].

Data analysis was done in the Perseus platform version 1.6.12.0 (Tyanova et al, 2016). Raw data was filtered for "contaminants" and "reverse" identifications, log2-transformed and further filtered for 70% data completeness across all samples. The missing values were imputed from a normal distribution

(downshifted mean by 1.8 standard deviation (SD) and scaled SD (0.3) relative to that of proteome abundance distribution). This dataset was used for principal component analysis (PCA), hierarchical clustering after z-score normalization and to build multi scatter plots to evaluate the correlation between uterine and cervical samples.

PRM analysis

The separation of the peptides was performed on a Dionex Ultimate 3000 RSLC chromatography system operated in column-switching mode. The equivalent of 250 ng of digested sample was injected and loaded onto a trap column (75 µm x 2 cm, C18 pepmap 100, 3 µm) using a mobile phase of 0.05% trifluoroacetic acid and 1% acetonitrile in water at a flow rate of 5 µL/minute. Peptides were further eluted onto the analytical column (75 µm x 15 cm, C18 pepmap 100, 2 µm) at 300 nL/minute by a linear gradient starting from 2 % solvent A to 35% solvent B in 48 minutes. The solvent A was 0.1% formic acid in water and the solvent B was 0.1% formic acid in acetonitrile.

The PRM analysis was performed on a Q Exactive HF mass spectrometer (Thermo Fisher Scientific). The MS cycle consisted of a full MS1 scan performed at a resolving power of 60,000 (at 200 m/z) followed by time scheduled targeted MS2 scans, with a normalized collision energy of 20, acquired at a resolving power of 60,000 (at 200 m/z), maximum accumulation time of 110 ms and an AGC target of 1e6 charges. The quadrupole isolation window of precursor ions was set to 1 m/z unit for the MS2 events and the duration of the time scheduled windows for each pair of endogenous and isotopically labeled peptides was set to 2 minutes.

PRM data processing

The PRM data were processed as previously described [203]. Briefly, the areas of extracted ion chromatograms (XIC) of the five most intense fragment ions (i.e., PRM transitions) of each precursor were extracted using the Skyline software (McCoss Lab) [367]. The identity of the peptides was confirmed using a a spectral matching approach based on the cosine of the spectral contrast angle (cos θ) calculated between the peak areas of the five transitions of the reference (mix of synthetic peptides without biological matrix) and the areas of the corresponding transitions for the endogenous and isotope labeled peptides in the clinical samples. Peptide detection was accepted if both cos θ were higher than 0.98. Values below that threshold were replaced by an estimation of the background.

Data 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). Correlation of the expression levels of the peptides belonging to one protein were analyzed. Statistical analysis to determine the diagnostic performance of each protein was analyzed with one peptide if the protein's peptides correlation was higher than 0.80 or with two peptides treated as independent entities if correlation was lower than 0.80. Comparison of the levels of the targeted peptides between EC and non-EC patients was performed using the

nonparametric Mann-Whitney U test. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg FDR method [368]. *Adjusted P* value 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. Pearson correlation was used to compare the levels of expression of the different proteins and patients between matrixes.

RESULTS

Proteomic characterization of uterine and cervical fluids

This study included the assessment of five different cervical fluids that differed in the sampling device that was used for its collection (Figure 30A-B). M2 were used for the collection of exocervical samples; M3, M4, M5 for the collection of endocervical samples; and M1 allowed the collection of both endo- and exocervical samples. We aimed to select the most suitable sampling method as a source of protein biomarkers of EC in cervical fluid and to identify non-invasive diagnostic biomarkers for EC in them. The workflow followed in this study is depictured in Figure 31.

Firstly, we evaluated whether the protein concentration of the different cervical fluids was sufficient for proteomic analysis by MS, as all cervical samples were highly diluted in 1x PBS to obtain the cervical fluids. We measured the protein concentration of the five cervical fluids and the uterine fluid before (raw) and after depletion of albumin and IgGs (depleted) from four different women (2 EC and 2 non-EC). As expected, raw uterine fluids had a much higher protein concentration (average of 32ug/ul) than any other sample since they contained highly abundant blood proteins (Figure 32A). This protein concentration dropped to an average of 0.95ug/ul when albumin and IgGs were depleted from the uterine fluids, and a similar protein concentration was quantified in the cervical fluids. M1, M3 and M5 samples showed the highest protein concentration (average of 1.95ug/ul, 1.86ug/ul, 1.45ug/ul, respectively), which is sufficient for further proteomic analysis. Among them, M1 presented the lowest variability. On the contrary, M2 and M4 yielded a low protein concentration, which could hamper their use for MS analysis. Moreover, we did not observe any difference in protein concentration between EC and non-EC patients (Figure 3A).
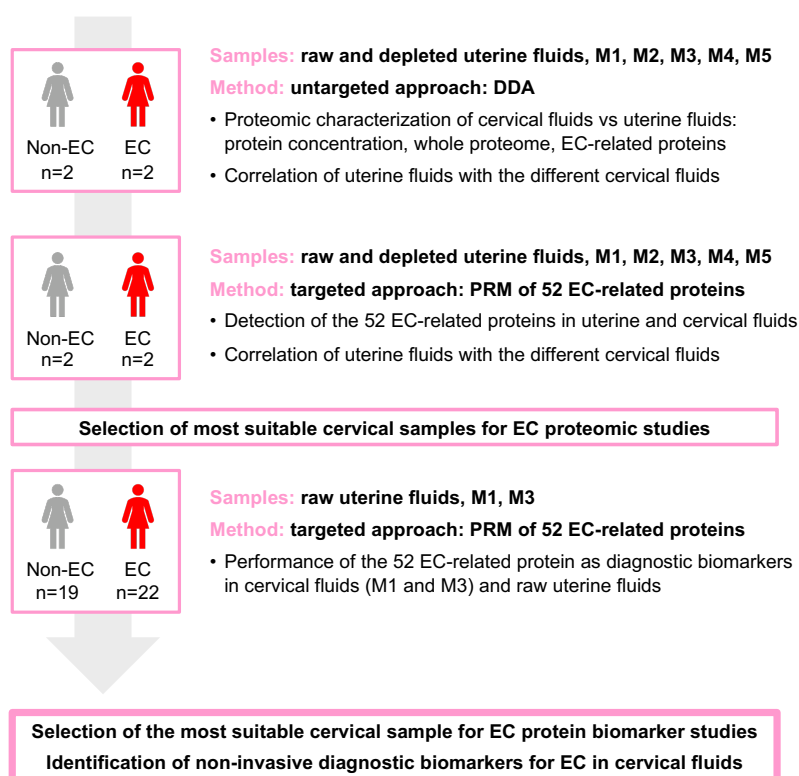
Figure 31. General workflow of the study. EC: endometrial cancer, DDA: Data Dependent Acquisition; PRM: Parallel Reaction Monitoring.

Secondly, we analyzed all uterine and cervical samples of the four patients using a shotgun MS-based approach in a tims-TOF pro mass spectrometer. This allowed us to get a global overview of the proteome of the different matrices and to evaluate the detectability of EC-related proteins in these samples. The number of protein identifications was quite similar among all samples, ranging from an average of 1,600 proteins detected in M1 and M2 samples, to an average of about 1,800 proteins detected in raw uterine fluids, M3, M4 and M5 samples. Depleted uterine fluids presented the highest number of proteins identified (average of 1,936) (Figure 32B). Interestingly, 97.6% of the proteins identified in cervical fluids were also identified in raw or depleted uterine fluids, indicating that cervical fluids share in great extense the proteome of uterine fluids (Figure 32C). In terms of protein identification, M1 was the cervical sample presenting fewer proteins in common with uterine fluids (79% of similar proteins) and with the other cervical samples (84 - 88%), whilst M2, M3, M4 and M5 shared from 83 to 86% the proteins contained in uterine fluids and from 88 to 91% the same proteins among them (Figure 32D).
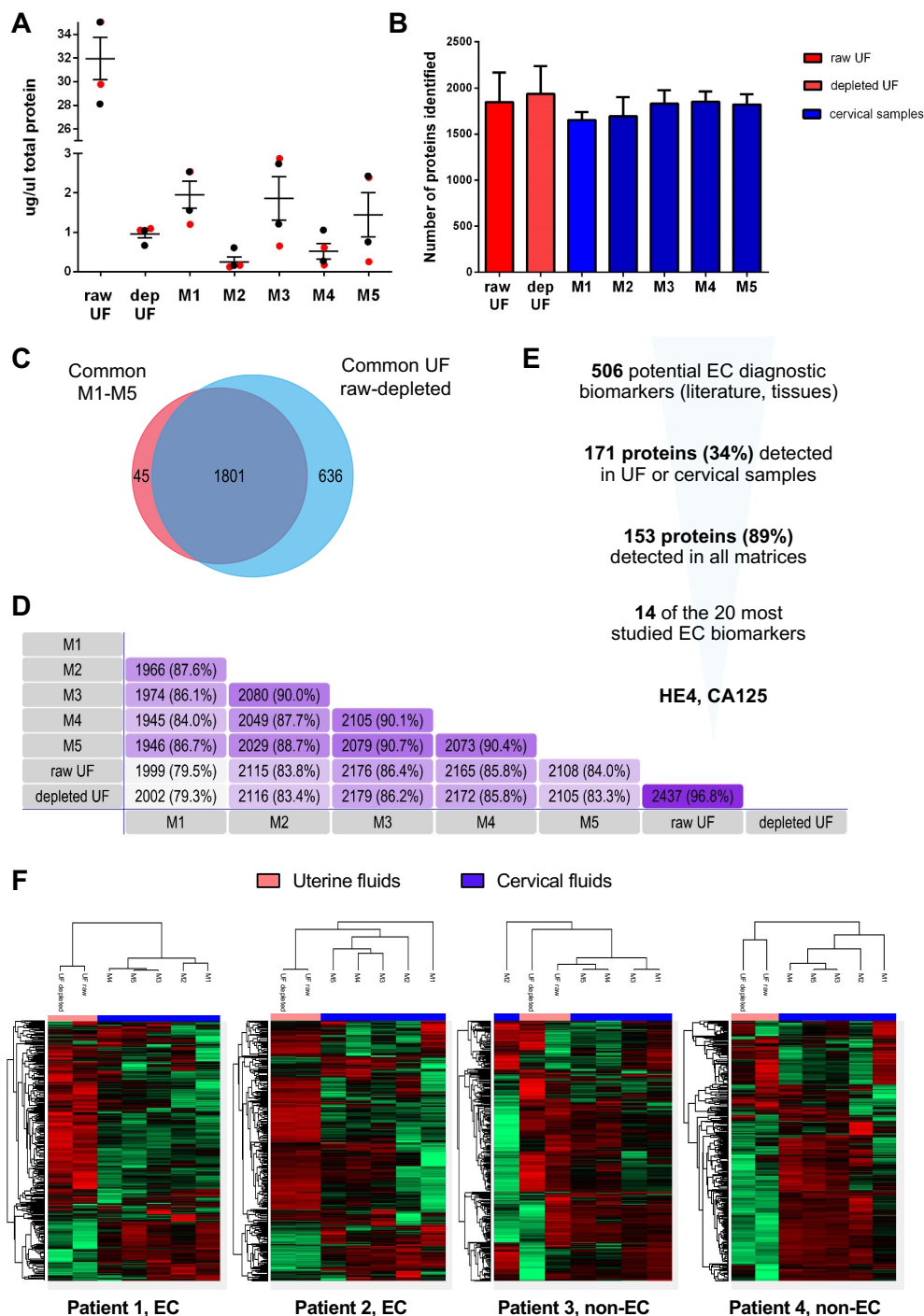
Figure 32. Protein concentration and protein identification and protein levels in uterine and cervical fluids from four patients (2 EC, 2 non-EC). (A) Total protein concentration (ug/ul) of raw and depleted uterine fluids (UF) and cervical fluids are plotted; (**B**) Bar plot representing the total number of proteins identified by MS in the four patients for each matrix analyzed (in red uterine fluids, and in blue cervical fluids); (**C**) Venn diagram with the common and differential proteins identified in raw and depleted uterine fluids versus cervical fluids; (**D**) Venn diagram showing the number of proteins in common between the different matrices; (**E**) Detectability of potential EC diagnostic biomarkers in the different matrices; (**F**) Heatmaps showing how the different matrices of the four patients analyzed are clustered. UF: uterine fluid; dep: depleted.

Next, we assessed the detectability of EC-related proteins in the uterine and cervical fluids. A list of 506 proteins described in literature as potential EC diagnostic biomarkers, mostly derived from studies performed in endometrial tissue samples, were considered as EC-related proteins [293]. From those, 169 proteins were detected in depleted uterine fluids, and 158-164 of them were also identified in cervical fluids (Figure 33E). Interestingly, 14 of the 20 most validated EC biomarkers [190] were detected in all the different matrices, including the two most studied diagnostic EC biomarkers (HE4 and CA125). Altogether, these results suggest that the proteome of uterine and cervical fluids is quite similar in terms of proteins present in each sample.

Ultimately, we compared the protein abundance in the uterine and cervical fluid proteome. As shown in Figure 33F, the protein levels in uterine fluids were different to those in the cervical samples for the four analyzed patients. The samples form two clusters except for one non-EC patient (Patient 3). The correlation of protein abundance between raw and depleted uterine fluids was very high (r=0.97-1). This was true for cervical fluids as well, but less pronounced (r= 0.81-0.90). This indicates that the depletion adds no significant improvement, and, in principle, it can be avoided. However, the correlation between uterine and the cervical fluids is significantly lower (r=0.6). Among the cervical fluids, the three samples collected at the endocervix, M3, M4 and M5, clustered together in the heatmap and showed the highest similarity; compared to M1 and M2, which were samples also representing the exocervix.

Measurement of EC-related protein biomarkers in cervical fluids

In Martinez-Garcia et al. [200], 52 proteins were studied as potential EC diagnostic biomarkers in depleted uterine fluids of 116 patients. The levels of 28 of those proteins were significantly higher expressed in EC compared with non-EC patients, and a combination of two proteins, MMP9 and KPYM, predicted EC cases with 94% sensitivity and 87% specificity [200]. The high degree of similarity in terms of protein identification between uterine and cervical fluids led us to evaluate the use of cervical fluids as a potential source of EC biomarkers. Thus, we assessed the detection of those 52 EC-related proteins in raw and depleted uterine fluids and in the five different cervical fluids, M1-M5, from 4 patients (2 EC; 2 non-EC) using a targeted LC-MS/MS PRM approach. Therefore 98 peptides belonging to those 52 proteins were measured in all samples (Annex 3, Table A3.1).

As expected, all proteins were detected in depleted uterine fluids. Importantly, detection of EC-related biomarkers was also very high in raw uterine fluids and in cervical fluids, where 46 out of the 52 proteins (88%) were detected in M2 and M4, 48 proteins (92%) in M5 and raw uterine fluids, 50 (96%) in M3 and 51 (98%) in M1 (Figure 33A). We then evaluated the correlation between the uterine fluids and cervical samples regarding the levels of the 52 targeted proteins. As observed at the whole proteome level (Figure 32F), the correlation of expression level of the 52 targeted proteins was not high between cervical and uterine fluids, on contrary to the very high correlation observed between raw and depleted uterine fluids for the four patients (r= 0.97-0.99).
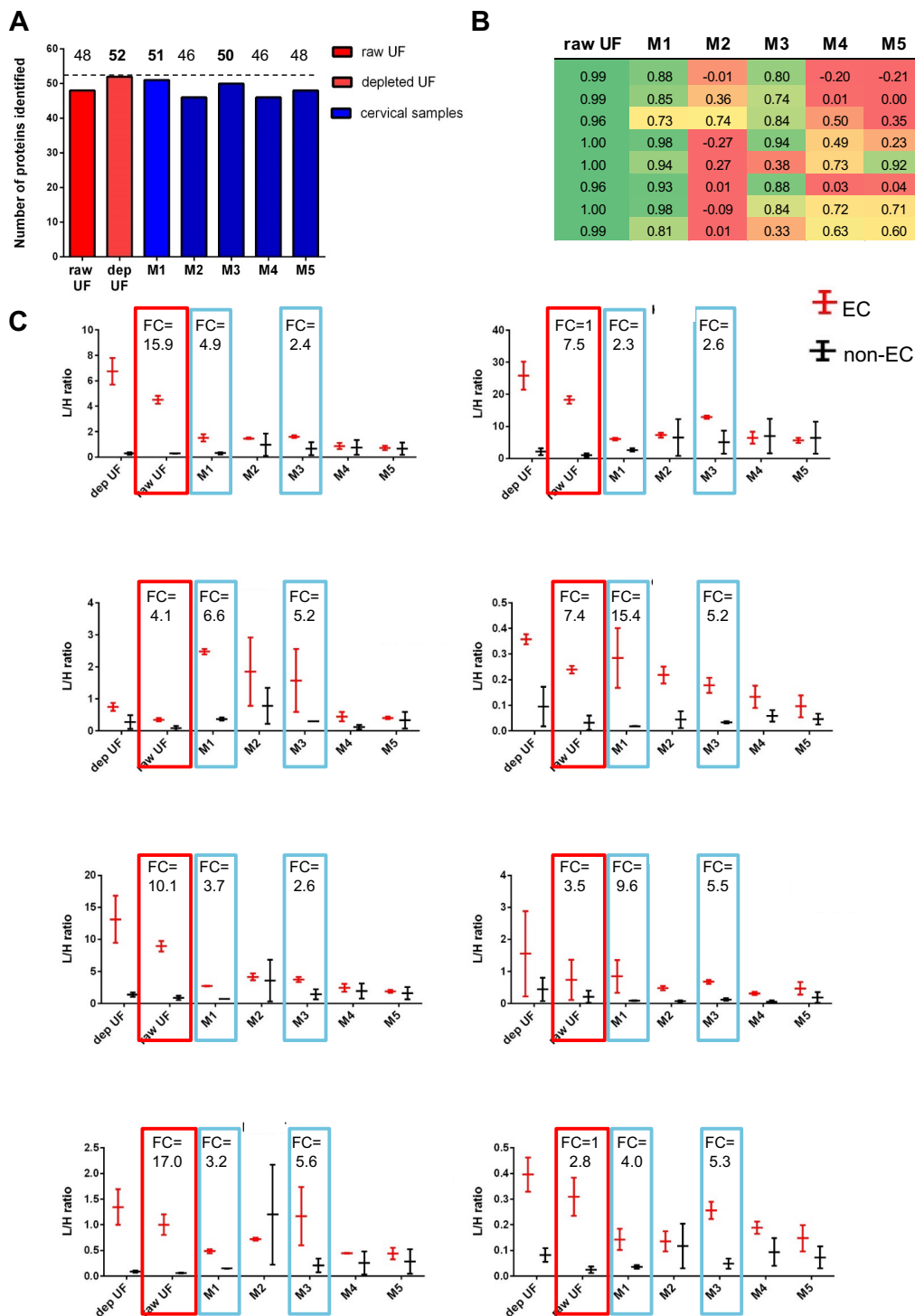
Figure 33. Quantification of 52 potential EC diagnostic biomarkers in the 7 different matrices (uterine and cervical fluids) from 4 patients (2 EC, 2 non-EC). (A) Bar plots representing the number of proteins out of the 52 EC biomarkers identified in each matrix (in red uterine fluids, and in blue cervical fluids); (**B**) Pearson correlation coefficients showing the degree of correlation between the levels of the 8 best performing EC biomarkers across the 4 patients when measured in the depleted uterine fluid and each of the other matrices analyzed; (**C**) Dot plots representing the levels (L/H ratios) of the 8 EC protein biomarkers in the different matrices of the EC (in red) and non-EC (in black) patients. The highest fold changes between EC and non-EC patients are observed in the uterine fluids (red square), and M1 and M3 cervical samples (blue squares) for all the 8 proteins. L/H ratio: light/heavy ratios; FC: fold change.

Among the 52 measured proteins, 8 proteins previously achieved the highest accuracy (AUC= 0.85-0.91) to diagnose EC in uterine fluids in two independent cohorts of 38 and 116 patients: PROT, PROT, PROT, PROT, PROT, PROT, PROT and PROT [200]. The expression levels of these proteins were analyzed in uterine and cervical fluids of 2 EC and 2 non-EC patients. In line with our previous observations, raw and depleted uterine fluids showed a high correlation for all the 8 proteins (r= 0.92-0.99). Among the cervical samples, M1 presented the highest correlation with the depleted uterine fluids (r= 0.73-0.98), followed by M3, that showed a high correlation for all proteins (r= 0.74-0.94) except for PROT and PROT. The other cervical samples showed no correlation with any of the uterine fluids analyzed (Figure 33B). Despite the limited number of patients, it was promising to observe that the 8 protein biomarkers were significantly higher in EC compared to non-EC patients not only in the depleted uterine fluids, but also when measured in the raw uterine fluids and in the M1 and M3 cervical fluids, with fold changes higher than 2 for the 8 proteins (Figure 33C).

Selection of M1 and M3 cervical samples

Among the five different cervical fluids analyzed, the one collected with the Rovers Cervex Brush® (M1) and the one collected with the endocervical swab HC2 DNA collection device Digene® (M3) were selected as the most suitable sampling methods to subsequently perform proteomic studies to identify EC protein biomarkers. While M1, M3, and M5 showed the highest protein concentrations, a similar level of total protein identifications was found across all investigated methods. Despite this, the highest number of identified EC biomarkers were found in M1 and M3. This suggests that the quality of the protein extract is more important for the biomarker discovery than the protein abundance. This might be true due to the large dynamic range of the mass spectrometric analysis. In summary, we obtained the highest correlation between these two samples and uterine fluids for the 8 best described EC biomarkers, as well as the largest differences of those 8 proteins between EC and non-EC patients.

Verification study of EC-related protein biomarkers in raw uterine fluids and cervical fluids

Expanding on our results for the M1 and M3 sampling to evaluate the potential of these cervical fluids to provide with EC diagnostic biomarkers, we measured the previously selected 52 EC-related proteins in raw uterine fluids, M1 and M3 cervical fluids from 41 patients (21 EC and 20 non-EC women). We used an LC-MS/MS PRM approach where we targeted 98 peptides that belong to the 52 EC-related protein biomarkers. Four peptides were excluded as they were detected in less than 50% of the samples, leading to a total of 51 proteins robustly measured with 94 peptides. Triplicates of four pooled uterine fluids, four pooled M1 samples and four pooled M3 samples were included to determine the reproducibility of the measurement. We obtained an average coefficient of variation (CV) of 4%, 4% and 3%, respectively. Only 3 out of the 282 measurements (94 peptides quantified in the 3 pools) showed a CV higher than 10%, and none above 20%, highlighting the robustness of the quantification in these three complex matrices (Annex 3, Figure A3.1).

A

			UTERINE FLUIDS				M1	M3
Entry	Gene Name	Peptide	Fold Change	Adjusted P value	AUC	Sn / Sp (%)	AUC	AUC
P00338	LDHA	LVIITAGAR	7.13	2E-04	0.958	100 / 80	0.758	0.748
P06733	ENO1	YISPDQLADLYK	5.09	2E-04	0.948	100 / 80	0.690	0.814
P14618	PKM	NTGIICTIGPASR	5.41	2E-04	0.935	89 / 90	0.680	0.781
P42574	CASP3	SGTDVDAANLR	6.43	5E-04	0.902	89 / 85	0.663	0.712
P16949	STMN1	SHEAEVLK	4.65	5E-04	0.899	89 / 75	0.711	0.770
O00299	CLIC1	LAALNPESNTAGLDIFAK	3.37	5E-04	0.892	89 / 80	0.627	0.801
P55060	CSE1L	ANIVHLMLSSPEQIQK	5.64	5E-04	0.892	100 / 65	0.765	0.827
Q06830	PRDX1	LVQAFQFTDK	5.11	5E-04	0.889	100 / 65	0.667	0.807
P14174	MIF	LLCGLLAER	3.67	5E-04	0.889	84 / 80	0.647	0.797
P43490	NAMPT	YLLETSGNLDGLEYK	10.48	5E-04	0.889	95 / 75	0.711	0.627
Q01469	FABP5	LVVECVMNNVTCTR	2.73	6E-04	0.886	95 / 80	0.546	0.539
O43278	SPINT1	SFVYGGCLGNK	4.42	6E-04	0.882	95 / 65	0.559	0.729
P12830	CDH1	VFYSITGQGADTPPVGVFIIER	4.50	1E-03	0.866	79 / 75	0.572	0.791
P61604	HSPE1	FLPLFDR	5.02	2E-03	0.853	95 / 60	0.725	0.781
P35222	CTNNB1	LVQLLVR	4.88	2E-03	0.848	100 / 60	0.748	0.819
P62937	PPIA	FEDENFILK	2.83	2E-03	0.837	100 / 50	0.624	0.784
P60174	TPI1	VVLAYEPVWAIGTGK	3.32	3E-03	0.833	100 / 60	0.637	0.781
P40121	CAPG	EGNPEEDLTADK	4.37	3E-03	0.827	95 / 65	0.624	0.708
P14780	MMP9	SLGPALLLQK	3.55	3E-03	0.824	89 / 65	0.595	0.624
P07237	P4HB	ILEFFGLK	3.26	4E-03	0.819	100 / 55	0.627	0.605
P10451	SPP1	ANDESNEHSDVIDSQELSK	8.45	4E-03	0.810	79 / 65	0.745	0.557
P16070	CD44	TEAADLCK	3.20	5E-03	0.807	95 / 65	0.729	0.745
P05164	MPO	VVLEGGIDPILR	9.23	6E-03	0.801	89 / 60	0.552	0.601
P07355	ANXA2	GVDEVTIVNILTNR	8.10	6E-03	0.797	100 / 45	0.562	0.745
P04792	HSPB1	LATQSNEITIPVTFESR	5.09	7E-03	0.791	84 / 60	0.650	0.824
P30041	PRDX6	DFTPVCTTELGR	2.83	9E-03	0.784	100 / 55	0.627	0.748
P09211	GSTP1	ASCLYGQLPK	2.95	9E-03	0.784	100 / 55	0.650	0.712
P04083	ANXA1	GGPGSAVSPYPTFNPSSDVAALHK	6.15	1E-02	0.775	100 / 40	0.592	0.585
O43852	CALU	TFDQLTPEESK	4.04	2E-02	0.758	100 / 45	0.725	0.691
Q13938	CAPS	SGDGVVTVDDLR	3.67	2E-02	0.755	100 / 45	0.681	0.583
P22626	HNRNPA2B1	TLETVPLER	2.21	2E-02	0.750	53 / 85	0.657	0.712
P50454	SERPINH1	GVVEVTHDLQK	2.41	3E-02	0.739	74 / 65	0.828	0.735
Q01995	TAGLN	TLMALGSLAVTK	2.05	4E-02	0.722	68 / 70	0.652	0.840
P80188	LCN2	VPLQQNFQDNQFQGK	2.54	4E-02	0.716	68 / 70	0.644	0.647
P17931	LGALS3	GNDVAFHFNPR	2.54	5E-02	0.712	84 / 50	0.627	0.797
P05787	KRT8	LSELEAALQR	12.50	5E-02	0.712	63 / 75	0.761	0.742

B
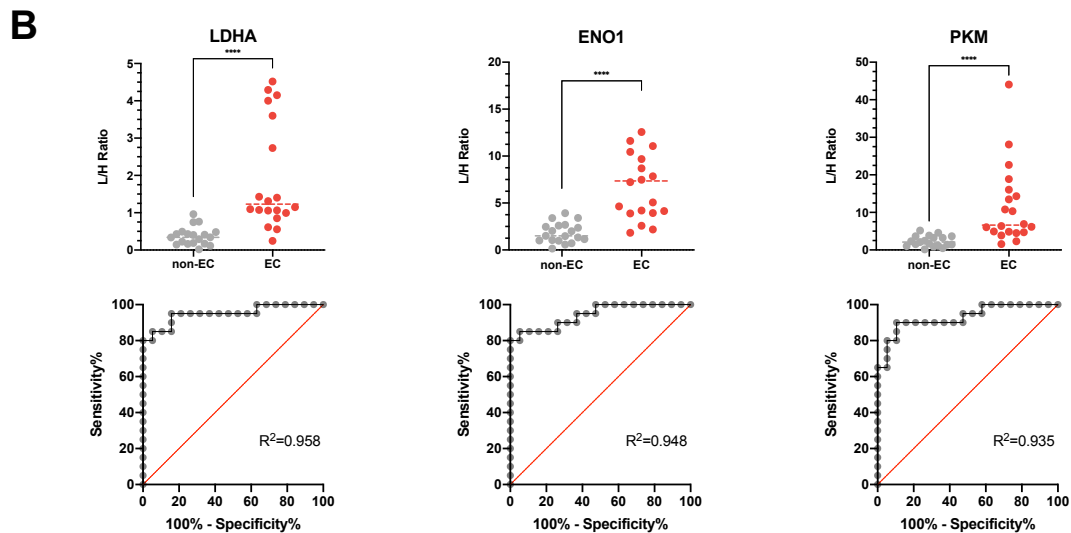
Figure 34. **Proteins showing significantly different levels in raw uterine fluids from EC (n=20) and non-EC women (n=19).** (**A**) Statistical results of the 36 proteins presenting an Adjusted P value < 0.05, Fold Change > |2| and AUC > 0.7 in the raw uterine fluids from the verification study (n=39) when comparing EC vs non-EC. Sensitivity (Sn) and specificity (Sp) are also shown. AUC values for the same proteins measured in M1 and M3 cervical fluids are also shown. AUC values are graded in grey according to their discriminative power. Proteins that demonstrated their diagnostic potential also in cervical fluids are highlighted in pink (AUC>0.7), and in dark pink the ones with AUC values higher than 0.8 in cervical fluids; (**B**) Dot plots of the top3 performing proteins representing the light/heavy ratios (L/H) obtained by LC-MS/MS PRM when analyzed in raw uterine fluids, and their corresponding ROC curve.

The statistical results for each of the 94 measured peptides (corresponding to 51 proteins) in the 3 different matrices are shown in Annex 3, Table A3.1. This study permitted to confirm the potential of 36 proteins (63 peptides) allowed the diagnosis of EC in raw uterine fluids, with all proteins showing significantly higher levels in EC patients compared to non-EC women with an *Adjusted P* value <0.05, a fold change (FC) value > 2, and an AUC higher than 0.71 (Figure 34A). Among those, LDHA, ENO1, PKM, and CASP3 showed the highest accuracy to discriminate between EC and non-EC patients with AUC values higher than 0.9. LDHA, ENO1 and PKM showed already as single protein markers significant potential all of them with sensitivities and specificities above 80% (Figure 34A-B). Interestingly, 28 of these proteins were also able to differentiate between EC and non-EC patients with AUC values higher than 0.7 in cervical fluids M1 and/or M3. However, the best performing proteins in uterine fluids were not the best performing in cervical fluids (Annex 3, Table A3.1).

The levels of all the proteins measured in the M1 cervical fluid, showed a high variability among the different patients and thus none of these proteins had an *Adjusted P* value below 0.05. Despite of this, 15 proteins discriminated EC and control patients with AUC values higher than 0.7 (Figure 35A). Among them, PROT had an AUC value of 0.83 (FC=5.89, *Adjusted P* value =0.062, and 83% sensitivity and 81% specificity), and PROT showed an AUC value of 0.80 (FC=2.65, *Adjusted P* value =0.095, and 78% sensitivity and 81% specificity). The expression of best performing 3 proteins across 41 patients and their ROC curves are shown in Figure 35C.

EC-related biomarkers measured in the M3 cervical fluid, resulted in the identification of 28 proteins differentially expressed (*Adjusted P* value <0.05) between EC and non-EC patients, all of them with AUC values higher than 0.7. Figure 35B shows the top 15 proteins, highlighting in pink the ones with AUC values > 0.8. Nine of these proteins presented fold changes higher than 2. The top 3 performing biomarkers were PROT (FC=6.38, AUC=0.84, 79% sensitivity and 72% specificity), PROT (FC=1.84, AUC=0.83, 79% sensitivity and 78% specificity), and PROT (FC=2.25, AUC=0.83, 89% sensitivity and 72% specificity) (Figure 35D).

Strikingly, in addition to the significant proteins shared in both uterine and cervical fluids, PROT, PROT, and PROT did not show diagnostic potential in uterine fluids but arise as significant diagnostic biomarkers in the M1 and M3 cervical fluids. While in uterine fluids these proteins had AUC values of 0.54, 0.64, and 0.68, respectively, they presented AUC values of 0.80, 0.75, and 0.71, respectively in M1 cervical fluids, and AUC values of 0.78, 0.73, and 0.72, respectively, in M3 cervical fluids.

A

Gene Name	Peptide	M1 CERVICAL SAMPLES			M3	UF
		Fold Change	FDR	AUC	AUC	AUC
PROT	GVVEVTHDLQK	**5.89**	6E-02	**0.828**	**0.735**	**0.739**
PROT	ISLPLPNFSSLNLR	**2.65**	9E-02	**0.804**	**0.784**	0.536
PROT	LLQTDDEEEAGLLELLK	**2.17**	9E-02	**0.788**	**0.819**	**0.892**
PROT	LSELEAALQR	**5.55**	1E-01	**0.761**	**0.742**	**0.712**
PROT	LVIITAGAR	1.49	1E-01	**0.758**	**0.748**	**0.958**
PROT	YDGHLPIEIK	1.74	1E-01	**0.748**	0.608	**0.889**
PROT	LVQLLVR	**2.03**	1E-01	**0.748**	**0.819**	**0.848**
PROT	LPDGYEFK	**2.07**	1E-01	**0.745**	**0.732**	0.641
PROT	ANDESNEHSDVIDSQELSK	1.96	1E-01	**0.745**	0.557	**0.810**
PROT	TEAADLCK	1.71	1E-01	**0.729**	**0.745**	**0.807**
PROT	TFDQLTPEESK	**2.08**	1E-01	**0.725**	0.691	**0.758**
PROT	FLPLFDR	1.99	1E-01	**0.725**	**0.781**	**0.853**
PROT	TIAPALVSK	1.32	2E-01	**0.716**	**0.791**	**0.948**
PROT	AIEENNNFSK	1.58	2E-01	**0.712**	**0.771**	**0.899**
PROT	FLIVAHDDGR	1.87	2E-01	**0.709**	**0.724**	0.683

B

Gene Name	Peptide	M1	M3 CERVICAL SAMPLES			UF
		AUC	FC	FDR	AUC	AUC
PROT	TLMALGSLAVTK	0.652	**6.35**	2E-02	**0.840**	0.722
PROT	VSFELFADK	0.634	1.84	2E-02	**0.833**	0.837
PROT	ANIVHLMLSSPEQIQK	**0.765**	**2.25**	2E-02	**0.827**	**0.892**
PROT	LATQSNEITIPVTFESR	0.650	1.33	2E-02	**0.824**	0.791
PROT	LVQLLVR	**0.748**	**2.13**	2E-02	**0.819**	**0.848**
PROT	ADEGISFR	0.693	1.59	2E-02	**0.817**	**0.889**
PROT	NLVQIK	0.516	1.82	2E-02	**0.814**	**0.866**
PROT	YISPDQLADLYK	0.690	1.69	2E-02	**0.814**	**0.948**
PROT	LAALNPESNTAGLDIFAK	0.627	1.69	2E-02	**0.801**	**0.892**
PROT	LLCGLLAER	0.647	1.63	2E-02	**0.797**	**0.889**
PROT	GNDVAFHFNPR	0.627	1.32	2E-02	**0.797**	**0.712**
PROT	ISLPLPNFSSLNLR	**0.804**	**2.72**	2E-02	**0.784**	0.536
PROT	FLPLFDR	**0.725**	1.99	2E-02	**0.781**	**0.853**
PROT	NTGIICTIGPASR	0.680	1.37	2E-02	**0.781**	**0.935**
PROT	VVLAYEPVWAIGTGK	0.637	1.66	2E-02	**0.781**	**0.833**

C **D**

Figure 35. Proteins showing significantly different levels in M1 and M3 cervical fluids from 39 EC and non-EC women. (**A**) Statistical results of the 15 proteins presenting an AUC > 0.7 in the M1 cervical fluids from the verification study when comparing EC (n=21) vs non-EC (n=18). Sensitivity (Sn) and specificity (Sp) are also shown.

147

Proteins with AUC values higher than 0.8 are highlighted in pink. AUC values for the same proteins measured in the M3 cervical fluids and the raw uterine fluids are also shown. These values are graded in grey according to their discriminative power; (**B**) Dot plots of the top3 performing proteins representing the light/heavy ratios (L/H) obtained by LC-MS/MS PRM when analyzed in M1 cervical fluids, and their corresponding ROC curve; (**C**) Top15 proteins presenting an Adjusted P value < 0.05, Fold Change > |1.3| and AUC > 0.7 in the M3 cervical fluid of 37 patients when comparing EC (n=18) vs non-EC women (n=19). Proteins with AUC values higher than 0.8 are highlighted in pink. AUC values for the same proteins measured in the M1 cervical fluids and the raw uterine fluids are also shown. These values are graded in grey according to their discriminative power; (**D**) Dot plots of the top3 performing proteins representing the light/heavy ratios (L/H) obtained by LC-MS/MS PRM when analyzed in M3 cervical fluids, and their corresponding ROC curve.

DISCUSSION

Currently, diagnosis of endometrial cancer relies on the observation of tumor cells in endometrial biopsies obtained by aspiration (*i.e.,* pipelle biopsies) but this procedure fails in up to 42% of cases to give a final diagnosis mainly due to the scarce cellularity of the pipelle biopsy samples [135].

In a previous study performed by our group, we measured the levels of 52 protein biomarkers in the liquid fraction of pipelle biopsies (*i.e.,* uterine fluids) from two independent cohorts of 38 and 116 EC and non-EC patients. We developed a 2-protein panel able to diagnose EC with 94% sensitivity and 87% specificity. Importantly, this panel achieved a correct diagnosis for all women that could not be diagnosed with the histopathologic examination of their pipelle biopsies due to insufficient cellularity in the sample or absence of tumor cells representation [200]. Uterine fluids are challenging samples to be analyzed by proteomic techniques as they contained variable amounts of blood, with albumin representing more than 80% of the whole protein content [369]. This is a problem in biomarker research since the wide dynamic range of more than eleven orders of magnitude in protein abundance in plasma and/or samples containing high amounts of plasma, such as uterine fluids, increases the difficulty to detect low-abundance proteins [179]. Thus, depletion of the top 2 (albumin and IgGs) or top 14 most abundant proteins in plasma is performed to facilitate the identification of biomarkers. Following the same approach, uterine fluids used in our previous studies were depleted of albumin and IgGs to facilitate the detection of lower abundant proteins. In this study, we demonstrated that most of the proteins previously evaluated in depleted uterine fluids can be detected in the non-depleted sample, which we named raw uterine fluid. Indeed, our results show a high correlation of the proteins in raw and depleted uterine fluids, both for the whole proteome (r=0.75-0.99) and a selection of EC-biomarker proteins (r= 0.97-0.99). Additionally, we confirmed 36 out of our previously evaluated 52 EC-biomarker proteins as biomarkers for EC diagnosis when measured in raw uterine fluids. Among the best performing biomarkers in raw uterine fluids are LDHA, ENO1, KPYM, and CASP3 with an AUC higher than 0.9, with LDHA alone reaching an AUC of 0.96. These results are promising for improving EC diagnosis, since further validation studies of larger prospective patient cohorts can be performed without the need of prior depletion steps of the uterine aspirates. This would significantly simplify the sample preparation process and increases the chances of clinical implementation.

Although uterine fluids obtained from pipelle biopsies are less invasive samples compared to endometrial biopsies obtained by hysteroscopy or D&C, they still cause substantial discomfort in women since the collection cannula (such as Cornier Pipelle®) have to go through the cervical canal. In addition, technical failure to introduce the collection cannula occurs in 11% of women, especially in postmenopausal women [370]. Therefore, the development of a simple, non-invasive test that accurately identifies EC and reassures healthy women would transform patient care. In this context, the use of samples collected in the lower genital tract (i.e., cervix or vagina) for EC diagnosis emerge as a promising alternative, since they are easier to collect, less invasive and better tolerated by women than collection of uterine samples [371]. The detection of EC-related protein biomarkers in cervical fluids can also impact on the development of screening tools, which might be applied to asymptomatic, high risk populations, such as patients undergoing tamoxifen treatment, and/or Lynch Syndrome patients. Also, cervical samples are routinely used in cervical cancer screening, being an already established screening tool in most of health systems and well accepted by women.

Many studies have evaluated the sensitivity of cytology in cervicovaginal samples for the detection of EC. Due to the anatomical continuity of the uterine cavity with the cervix, it was proven that EC shed malignant cells that can be collected in a less invasive way in the cervix. However, a systematic review and meta-analysis published in 2020 concluded that the morphological evaluation of cervicovaginal cytologies have a poor performance for EC diagnosis, with only around 45% of patients with EC showing abnormal Pap test results [207]. In contrast, a recent study demostrated a higher accuracy of urine and vaginal cytologies for gynaecological cancer diagnosis, mainly EC, reaching 91.7% sensitivity and 88.8% specificity for the combination of both cytologies, but also a 89.6% sensitivity and 88.7% specificity only for vaginal citologies [297]. Yet, the morphological assessment of cytologies can be quite subjective and require highly trained specialist pathologists, making cytologies examination less reproducible between different healthcare settings. Moreover, these approaches depend on the reliable shedding of endometrial cells into the lower genital tract. An interesting alternative to the morphological examination of cytologies is the assessment of tumor biomarkers. To date, most studies have focused on the evaluation of the methylation levels and/or mutational profiles of tumor DNA collected from the cervix and the vagina with different devices: endometrial brushes, cervical scrapes, vaginal swabs, vaginal tampons, etc. [206,372–374]. Among them, Wang et al described a PCR-based test (PapSEEK) to detect genetic alterations in samples collected from the endocervix with a Pap brush. This test allowed for the detection of 81% of EC cases with very high specificity [375]. Also, Huang et al developed a panel comprising two of the three hypermethylated genes BHLHE22, CDO1, and CELF4 that reached a sensitivity of 91.8% and a specificity of 95.5% for EC detection in cervical scrapings [376].

Very few studies have searched for protein biomarkers for EC diagnosis in cervical samples and those few have only evaluated the diagnostic power of the CA125 tumor marker. However, this biomarker did not achieve a high accuracy for EC diagnosis. The landscape of proteomic analysis to decipher new biomarkers is broad, and includes the use of multiplexing proteomic techniques that allow for the

measurement of high number of potential biomarkers which can be combined to increase the discrimination power. To apply those technologies, a high quality and quantity protein material needs to be collected. Here, we compared the proteome of uterine fluids, already proved to be relevant sources of EC protein biomarkers, and cervical fluids collected with five different sampling devices. As shown in this study, all five cervical fluids contained sufficient amount of protein to conduct MS-based proteomic studies, although M1, M3 and M5 yielded the highest concentrations. The proteomic analysis in uterine and cervical fluids from four different patients unveiled that those cervical fluids shared 97% of their proteome with uterine fluids. Moreover, all cervical fluids represented valuable specimens for the identification of EC protein biomarkers. From the 506 EC-related proteins previously described in Martinez-Garcia E et al [203], we detected 169 proteins in depleted uterine fluids, and 158-164 (93-97%) of them were also identified in the cervical fluids. In a subsequent step, we evaluated the potential of the five cervical fluids to detect the expression of the 52 EC biomarkers that were validated as diagnostic biomarkers in the Martinez-Garcia E et al study by using targeted mass spectrometry. This study unveiled that the most promising cervical fluids were collected with the Rovers Cervex Brush® (M1) and the endocervical swab (HC2 DNA collection device Digene) (M3). Firstly, the highest number of protein biomarkers was detected in these samples (M1: 51 proteins; M3: 50 proteins), and secondly, M1 and M3 cervical fluids showed the highest correlation with uterine fluids for the 8 most promising EC biomarkers and the highest differences between EC and non-EC patients.

Importantly, this study evaluated the performance of M1 and M3 cervical fluids to detect 52 EC biomarkers by analyzing 41 patients using a targeted LC-MS/MS PRM approach. A total number of 15 and 28 proteins were significantly different in M1 and M3 cervical fluids from EC and non-EC patients, respectively. Interestingly, PROT, PROT, and PROT were differential proteins in cervical fluids, but not in uterine fluids. Among the best performing biomarkers in cervical fluids, PROT and PROT achieved AUC values of 0.83 and 0.80, respectively, when measured in the M1 samples. Regarding M3 cervical samples, three proteins were able to discriminate between EC and non-EC patients with AUC values higher than 0.8: PROT (AUC=0.84), PROT (AUC=0.83), and PROT (AUC=0.82). These results validate the use of M1 and M3 cervical fluids as an untapped source of non-invasive biomarkers for EC diagnosis. However, due to the limited sample size of our study, the diagnostic performance of the biomarkers described in this study need to be further validated in a larger cohort of patients.

Finally, although we demonstrated that some EC biomarkers are shared in uterine and cervical fluids, we also showed the strong differences in the protein abundance of those samples, with low correlations at the whole proteome level and when focusing on the 52 EC biomarker candidates. This highlights the importance of designing discovery studies in the cervical fluids analyzed since we expect novel EC biomarkers to be identified.

As proved in this study, both M1 and M3 cervical samples are promising clinical samples for proteomic studies directed to identify and validate EC biomarkers. However, sample selection for subsequent

studies should also take into consideration that M1 is already introduced in the gynecologic office worldwide, as it is usually used to collect the sample of cervical liquid cytology. Its use is well known by gynecologists and well accepted by women, what could mean a higher and faster acceptance of a future developed EC diagnostic test by doctors and patients.

Moreover, this study demonstrates the potential of protein biomarkers measured in minimally invasive cervical samples for the detection of EC and lays the foundation for the validation of the here highlighted promising protein biomarkers in larger cohorts of patients.
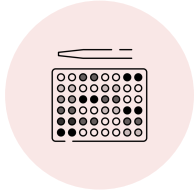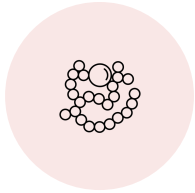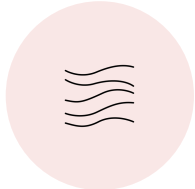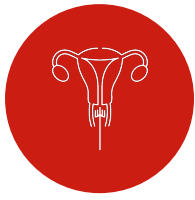
CHAPTER 4

Protein signatures in cervical fluids permit a highly accurate and non-invasive diagnosis of endometrial cancer

SPECIFIC BACKGROUND

In Chapter 3 we identified the most suitable cervical fluid to use as a source of EC diagnostic biomarkers and to be adopted in the current clinical practices. However, we also demonstrated the proteome differences between uterine and cervical fluids. Hence, to achieve the ideal diagnostic test, which yields high sensitivity and avoids false negatives, in non-invasive cervical fluids, in Chapter 4 we compiled the preclinical steps of a biomarker pipeline to identify novel EC diagnostic biomarkers.

The Discovery phase is an untargeted process focused on the identification and quantification of the largest number of proteins possible to pinpoint promising proteins between different homogeneous groups. However, it suffers from high risk of false positive results, particularly in low abundant proteins [183]. Thus, following discovery, the verification phase aims to confirm that the abundances of target peptides identified in the discovery phase are significantly different between groups. LC-MS operated in a targeted proteomics approach is ideal to achieve this task as proteins can be accurately quantified in a highly multiplexed fashion at a relatively high analytical throughput [182]. This phase is crucial to prioritize and move clinically relevant candidate biomarkers to enter in a clinical validation phase. In the verification phase, acceptable performance of the biomarkers (high sensitivity and specificity) must be demonstrated, as well as a clear contribution to existing clinical practices. Eventually, the highest performing biomarkers are selected and measured in a larger cohort of patients in a validation phase, normally using either antibody-based assays such as ELISA or targeted MS assays when specific antibodies are not available.

Additionally, cancer is a multifactorial and heterogeneous disease, and it is unlikely that a single biomarker displays sufficient discriminatory power to significantly affect clinical decisions. Therefore, the search for biomarkers signatures providing more complete information is crucial, and highly multiplexing analytical methods would facilitate this task.

Chapter 4 includes: (i) the results obtained from the discovery phase acquired by MS operated in DDA mode, where we measured the cervical fluids of 59 patients (n=20 EC, n=20 non-EC, n=19 cervical pathology); (ii) the results obtained from the verification phase in which we measured the diagnostic performance of the 75 proteins identified in the previous phase in the cervical fluids of 241 patients (n=128 EC, n=106 non-EC, n=7 atypical hyperplasia) by LC-PRM acquisition; (iii) the assessment of those proteins to determine prognostic factors in the verification cohort; (iv) the development of protein panels to increase the accuracy of individual biomarkers; (v) the verification of the top performing diagnostic biomarker by ELISA assay, as a move towards the clinical use of these biomarkers. The stepwise work conducted in this study permitted the development of protein panels that achieved an

excellent performance to non-invasively diagnose EC and discriminate between EC histological types and grades in cervical fluids.

MATERIALS AND METHODS

Study participants

A total of 300 women were recruited at the Vall Hebron University Hospital (Barcelona, Spain), and the Hospital General de Catalunya (Sant Cugat del Vallès, Spain) from 2017 to 2019. Informed consent forms, approved by the Ethical Committees of each Hospital, were signed by all patients. Women included in the study, except for those diagnosed with cervical pathology, entered the EC diagnostic process due to common symptoms in EC, mainly AUB and/or the observation of a thickened endometrium in the transvaginal ultrasonography (higher than 4mm for postmenopausal women and 8mm for premenopausal women) [377]. From these women, 148 were diagnosed with EC, 7 with hyperplasia, and 126 with benign diseases (polyps, or myomas) or benign endometrium (atrophic, desquamative, proliferative, or normal). The 19 patients with cervical pathologies, but with no suspicion of EC, were included as additional controls in the study. The clinicopathological characteristics of all the patients are shown in Table 17.

Cervical sampling

Cervical samples were collected before endometrial biopsies were obtained as part of the standard of care. Samples were collected with a Rovers Cervex Brush ® (Rovers Medical Devices), scrubbing the vagina and endo- and exo- cervix, similarly as it is performed in a conventional cytology. Then, the brush was immediately immersed and stirred in a 10 mL Screw Cap Tube PP Skited ® (Nirco P.N. 439910) containing 2 ml of 1x PBS. Each sample was transferred into 2 ml 156ppendorf tubes, and centrifuged at 2.500 rcf for 20 minutes at 4°C. The soluble fraction (supernatant) is the cervical fluid used in this study and it was separated and stored at −80°C until use.

156

	Discovery phase (n=59)				Verification phase (n=241)			
	EC (n=20)	non-EC (n=20)	non-EC (PC) (n=19)		EC (n=128)	Hyp (n=7)	non-EC (n=106)	
Age (years)								
Mean	64	63	40		64	49	64	
Minimum	90	87	61		92	55	95	
Maximum	32	51	23		32	44	31	
Uterine condition								
Premenopausal	3	0	17	33%	30	2	15	20%
Postmenopausal	17	20	2	65%	98	5	91	80%
Endometrial thickness (mm)								
Mean	24	5	-		18	11	9	
< 4 mm	-	8	-	20%	8	-	19	11%
4 mm	15	12	-	68%	114	7	57	74%
Benign gynecological condition								
Normal endometrium		1		5%			15	14%
Atrophic endometrium		12		60%			39	37%
Endometrial polyp		7		35%			36	34%
Myoma		-		0%			1	1%
Others (proliferative, descamative)		-		0%			15	14%
Histological type								
Endometrioid	13			65%	103			80%
Serous	5			25%	16			13%
Others non-endometrioid	1			5%	9			7%
Histological grade								
Grade 1	1			5%	36			35%
Grade 2	10			50%	56			55%
Grade 3 - EEC	8			40%	11			11%
Grade 3 - NEEC					24			
FIGO stage								
IA	4			20%	62			48%
IB	5			25%	18			14%
II	4			20%	26			20%
IIIA	1			5%	3			2%
IIIB	-			0%	3			2%
IIIC1	2			10%	2			2%
IIIC2	2			10%	7			5%
IVA	-			0%	2			2%
IVB	1			5%	5			4%
Miometrial invasion								
<50%	6			30%	74			58%
>50%	13			65%	53			41%
Lymphovascular invasion								
Yes	9			45%	34			27%
No	10			50%	85			66%
Unknown	-			0%	9			7%
Molecular classification								
POLEmut	2			10%	3			2%
MMRd	3			15%	25			20%
NSMP	10			50%	46			36%
p53abn	4			20%	18			14%
NA	1			5%	36			28%

Table 17 Clinicopathological characteristics of the women enrolled in the study.

Molecular classification

To classify EC patients under the molecular classification described by the TCGA [44], we followed the approach purposed by Vermij L et al. that ESGO/ESTRO/ESP 2021 guidelines recommend [64,85]. The expression of p53 and the microsatellite instability were evaluated by immunohistochemistry using antibodies directed against P53, MLH1, PMS2, MSH2, and MSH6 on histological material, and following the methods previously described [59,165]. POLE gene was tested on six of its most common hotspot mutation sites (P286R, S297F, V411C, V411T, A456P, S459F) [61] by extracting the DNA of 3 slides of 5um of FFPE tissues using the QIAamp DNA FFPE Tissue kit (Qiagen, reference 56404) and genotyping them using KASP technology V4.0 2x Mastermix (GC Genomics / Kbioscience, reference KBS-1016-

021) following manufacturer's instructions. Immunohistochemistry stainings were evaluated by a pathologist dedicated to gynecological oncology. MLH1, PMS2, MSH2 and MSH6 proteins were designed as abnormal when no nuclear staining was found, while aberrant p53 staining was recognized either by overexpression (nuclei of the tumor cell stained with an intensity higher than 75%) or no expression of the protein in the tumor cell nuclei. According to the results, patients were classified in 4 subgroups following this order: POLE mutation (POLEmut) if the patient presented a POLE mutation in any of the 6-hotspot analyzed; MMR proteins deficient (MMRd) if the patient presented an abnormal expression in one or more MMR proteins; p53 abnormal (p53mut) if the patient exhibited aberrant p53; and non-specific molecular profile (NSMP) if any molecular alteration was encountered in the patient.

Discovery study

Sample preparation

Cervical fluids from 59 patients corresponding to patients diagnosed with EC (n=20), non-EC (n=20) and cervical pathology (n=19) were included in the discovery study. To prepare the cervical fluids for MS analysis, protein concentration was measured by the Bradford assay performed in triplicate. A total of 10µg of protein of each sample was precipitated by incubation with 6 volumes of cold acetone (-20ºC) overnight; the protein pellet was resuspended in 6M urea with 200mM ammonium bicarbonate (ABC), reduced with dithiothreitol (3µL of 10mM dithioteitrol, 37ºC, 60 minutes), and alkylated with iodoacetamide in the dark (3µL of 20mM iodoacetamide, 22ºC, 30 minutes). Then, the urea concentration was diluted to 2M in 200mM ABC. Digestion was conducted with endoproteinase LysC (protease/total protein amount ratio of 1:10 weight per weight (w:w), 37ºC, overnight (o/n), Wako, P.N. 129-02541)), followed by the dilution of urea to 1M urea with 200mM ABC for a final incubation with trypsin (protease/total protein amount ratio of 1:10 w:w, 37ºC, o/n, Promega, P.N. V5111). After digestion, trypsin activity was quenched by the addition of formic acid 100% (1% of the final volume), and the samples, containing the peptide mixtures of the cervical fluids, were desalted with Ultra MicroSpin C18 columns, 300A silica (3-30 µg) (The Nest Group, Inc, P.N. SUM SS18V) before LC-MS analysis.

LC-MS Setup for shotgun proteomic analysis

Peptide mixtures of each sample were analyzed using a LTQ-Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) coupled to an EASY-nLC 1000 (Thermo Fisher Scientific (Proxeon), Odense, Denmark). Peptides were loaded directly onto the analytical column and were separated by reversed-phase chromatography using a 50-cm column with an inner diameter of 75 µm, packed with 2 µm C18 particles spectrometer (Thermo Scientific, San Jose, CA, USA). Chromatographic gradients started at 95% buffer A and 5% buffer B with a flow rate of 300 nl/min for 5 minutes and gradually increased to 22% buffer B and 78% A in 79 min and then to 35% buffer B and

65% A in 11 min. After each analysis, the column was washed for 10 min with 10% buffer A and 90% buffer B. Buffer A: 0.1% formic acid in water. Buffer B: 0.1% formic acid in acetonitrile.

The mass spectrometer was operated in positive ionization mode with nanospray voltage set at 2.4 kV and source temperature at 275°C. Ultramark 1621 for the was used for external calibration of the FT mass analyzer prior the analyses, and an internal calibration was performed using the background polysiloxane ion signal at m/z 445.1200. The acquisition was performed in data-dependent acquisition (DDA) mode and full MS scans with 1 micro scans at resolution of 120,000 were used over a mass range of m/z 350-1500 with detection in the Orbitrap. Auto gain control (AGC) was set to 1E5 and charge state filtering disqualifying singly charged peptides was activated. In each cycle of data-dependent acquisition analysis, following each survey scan, the most intense ions above a threshold ion count of 10.000 were selected for fragmentation at normalized collision energy of 28%. The number of selected precursor ions for fragmentation was determined by the "Top Speed" acquisition algorithm and a dynamic exclusion of 60 seconds. Fragment ion spectra produced via high-energy collision dissociation (HCD) were acquired in the Ion Trap, AGC was set to 1e4, isolation window of 1.6 m/z and maximum injection time of 200 ms was used. All data were acquired with Xcalibur software v4.1.31.9.

Data interrogation and statistical analysis

All MS data was processed in the MaxQuant environment [365] (Version 1.6.7.0) using the Andromeda search engine [238] against the Uniprot *Homo sapiens* database downloaded in February 2019 (72,405 entries).The default settings in the MaxQuant software were used, enabling the "match between runs" option. Analysis of the identified proteins was performed using Perseus software (version 1.6.15.0) [243], and Rstudio (v.1.4.1717, "Juliet Rose"; PBC). The analysis consisted of (i) setup of the data and generation of the *raw matrix* and quality control of the data; (ii) differential abundance analysis; (iii) absence/presence analysis. For the statistical analysis, we used the "*proteinGroup.txt*" file generated by MaxQuant. The LFQ intensity of each protein in each sample was uploaded to Perseus and the "potential contaminants", "reverse" and "only identified by site" identifications were removed and the data was log2 transformed, obtaining the *raw matrix*. On this *raw matrix*, we performed the quality control analysis based on the study of densities of intensities of the proteins identified, sum of intensities of the total identifications for each individual, principal component analysis (PCA), and heatmaps. Next, for the differential expression analysis, proteins identified in less than 75% of the patients were removed, followed by the removal of 3 patients with less than 85% of the remaining proteins identified. The missing values were imputed using N distant from the total matrix. We used an *in-house* script using the *limma* package of the R software to perform the differential abundance analysis. For the absence/presence analysis, the number of identifications of each protein contained in the *raw matrix* per each group was counted and a Fisher exact test (*stats* package) was applied. Significant proteins were considered when *Adjusted P* value <0.15, and AUC higher than 0.7 and/or FC > 1.8 for the differential abundance analysis, while *Adjusted P* value < 0.1 was considered significant in the Fisher exact test for the

absence/presence analysis. Additionally, the *reportROC* package of R software was used to calculate the Area Under the ROC Curve (AUC) for each protein.

In addition, Venn Diagrams were used to compare the proteins identified in the cervical fluids of each group of patients (EC, non-EC, CP). We also compared the proteins identified in our study with those identified by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) in formalin fixed paraffin-embedded (FFPE) tissue samples from 100 EC patients [298], and with a list of 506 potential EC diagnostic biomarkers previously described in tissue and blood samples, which were compiled in Martinez-Garcia E et al after an extensive revision of the literature searching for proteomic studies between 1978 and 2017 [190]. For the biological enrichment analysis, 262 proteins were selected based on the *Adjusted P* value. Functional annotation and characterization of the significant proteins were performed using Gene Set Enrichment Analysis (GSEA) using the *clusterProfiler* package in R was used for the biological process analysis, where the proteome of the cervical fluids was used as reference proteome for the analysis.

Verification study

Sample preparation

Cervical fluids from 241 patients corresponding to patients diagnosed with EC (n=128), non-EC (n=113) were included in the discovery study. Protein concentration of all the cervical fluids was measured by the BCA assay. For each sample, a total of 12.5µg of protein was separated and diluted with 50mM Tris-Hydrochloride (HCl) buffer to reach the same final volume of 60µL. Proteins were denatured by the addition of urea 6M, and incubated for 10 minutes under agitation at 22ºC, followed by 10 minutes incubation in the ultrasonic bath (Branson 5510, Branson Ultrasonics). Samples were then reduced with dithiothreitol (82µL of 70mM dithioteitrol, 37ºC, 60 minutes), and alkylated with iodoacetamide in the dark (261µL of 225nM iodoacetamide, 22ºC, 30 minutes). The samples were digested with endoproteinase LysC (protease/total protein amount ratio of 1:50 w:w, 37ºC, 3 hours, Thermo Scientific (cat.no. 90051)), followed by the dilution of urea to 1M urea using 50mM Tris-HCl buffer for a final incubation with trypsin (protease/total protein amount ratio of 1:25 w:w, 37ºC, o/n, Promega, P.N. V5111). After digestion, trypsin activity was quenched by the addition of formic acid 100% (1% of the final volume). Samples were spiked with 110 stable isotope labeled peptides, so called heavy peptides (C-terminal arginine, $^{13}C_6$, $^{15}N_4$, C-terminal lysine, $^{13}C_6$, $^{15}N_2$; PEPotec crude Thermo Scientific). The mixture of synthetic peptides was previously prepared from the stock solutions (50% acetonitrile, 0.1% TFA), aliquoted in Eppendorf Low Bind tubes and stored at -80°C and each aliquot was intended for a single use. Finally, samples were desalted by solid phase extraction cartridges (Sep Pak tC18, 50mg, Waters (cat.no.WAT054960)), vacuum dried, and resuspended in 0.1% formic acid before LC-PRM analysis.

LC-MS Setup for targeted proteomic analysis

The LC-MS setup consisted of a Dionex Ultimate 3000 RSLC chromatography system 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. Samples were injected onto a trap column (75 µm × 2 cm, C18 pepmap 100, 3 µm) with the loading phase at 5 µl/min for 8 min. Peptides were further eluted onto an analytical column (75 µm × 15 cm, C18 pepmap 100, 2µm) at 300 nl/min by a linear gradient starting from 2 % A to 35 % B in 66 min. The PRM acquisition was performed by a hybrid quadrupole orbitrap mass spectrometer (Q Exactive HF, Thermo Scientific). The time scheduled PRM scans were acquired with acquisition windows of 3 min centered on the expected retention times of the targeted peptides. The quadrupole isolation windows were set to 1 m/z, the maximum trapping time of the peptide ions to 110ms and the normalized fragmentation energy to 25. The resulting fragment-ions were analyzed by the orbitrap at a resolution of 60,000 (at 200 m/z).

The ion chromatograms of the most intense fragment-ions were extracted and their areas calculated by the Skyline software. Data were filtered by a similarity score (dot product>0.98) calculated for each light and heavy peptide pairs. Missing values due to MS measurements below the limit of detection were replaced by an estimation of the background (minimum accepted value /3) [229]. Three replicates were added in each plate to assess the reproducibility of the analytical workflow. Three replicates were added in each plate to assess the reproducibility of the analytical workflow.

Data analysis

The light/heavy area ratio of each peptide was extracted from Skyline. For the 36 proteins measured with more than one peptide, we evaluated the correlation between the two peptides of the same protein. The analysis of the proteins with highly correlating peptides (R^2>0.9) was based on the quantification of any of the surrogate peptides (or the one with higher number of identifications); while for the others both peptides of each protein were analyzed separately as isoform-specific. The statistical analysis was performed with Rstudio (v. 1.4.1717, "Juliet Rose"; PBC), SPSS (v.23.0; IBM), and Graph Pad Prism (v. 9.0.1; GraphPad Software). Comparison of the levels of the monitored peptides between groups of patients was performed using the nonparametric Mann-Whitney U test, as the data did not follow a normal distribution. *P* values were adjusted for multiple comparisons using the Benjamini-Hochberg FDR method [368]. Adjusted *P* values lower than 0.05 were considered statistically significant. Receiver-operating characteristic (ROC) analysis was used to assess the specificity and sensitivity of the biomarkers and the area under the ROC curve (AUC) was estimated for each individual protein.

Development of classifiers

A logistic regression model was adjusted to the data to assess the power of the different combinations of proteins when classifying samples in two clinical categories. ROC curves were generated for each of these regression models; the AUC, the sensitivity, and the specificity at the "optimal" cutoff point for

discrimination between groups were obtained. The optimal cutoff corresponded to the threshold that maximized the distance to the identity (diagonal) line. The optimality criterion was max (sensitivity + specificity). To assess the robustness of the developed classifiers, a "k-fold" cross-validation (k=10) procedure for the diagnosis model and a "leave-one-out" cross-validation for the prognostic models were performed by applying to a subset of samples (10 in case of f k-fold or 1 in case of leave-one-out) in the dataset the logistic regression model adjusted to the remaining samples on the dataset, hence deriving a new ROC curve and afterward performing the usual ROC analysis.

ELISA

The concentration of PROT in cervical fluids was quantified using a commercially available ELISA kit (R&D Systems, reference:XX) according to the manufacturer's protocol. All samples but one (due to a lack of sample material) from the 241 patients included in the verification study were analyzed by ELISA using a non-diluted or 1:5 diluted sample. All samples were assayed in duplicates and the average values were reported as ng/mL. The linear correlation between the results obtained by LC-PRM and the ones obtained with the ELISA assay was calculated using the Pearson correlation coefficient.

RESULTS

The study here conducted follows the typical protein biomarker pipeline and includes a discovery phase and a verification phase performed by untargeted and targeted mass spectrometry, respectively; that ended up on the development of protein panels to increase the accuracy of individual biomarkers for diagnosis and determination of histological subtype and grade, and an additional verification phase performed on the same cohort of patients to assess the performance of the best performing diagnostic biomarker in ELISA technology, which aimed to move forward this biomarker to a standardized technique in clinical laboratories. The workflow used in this study is shown in Figure 36.
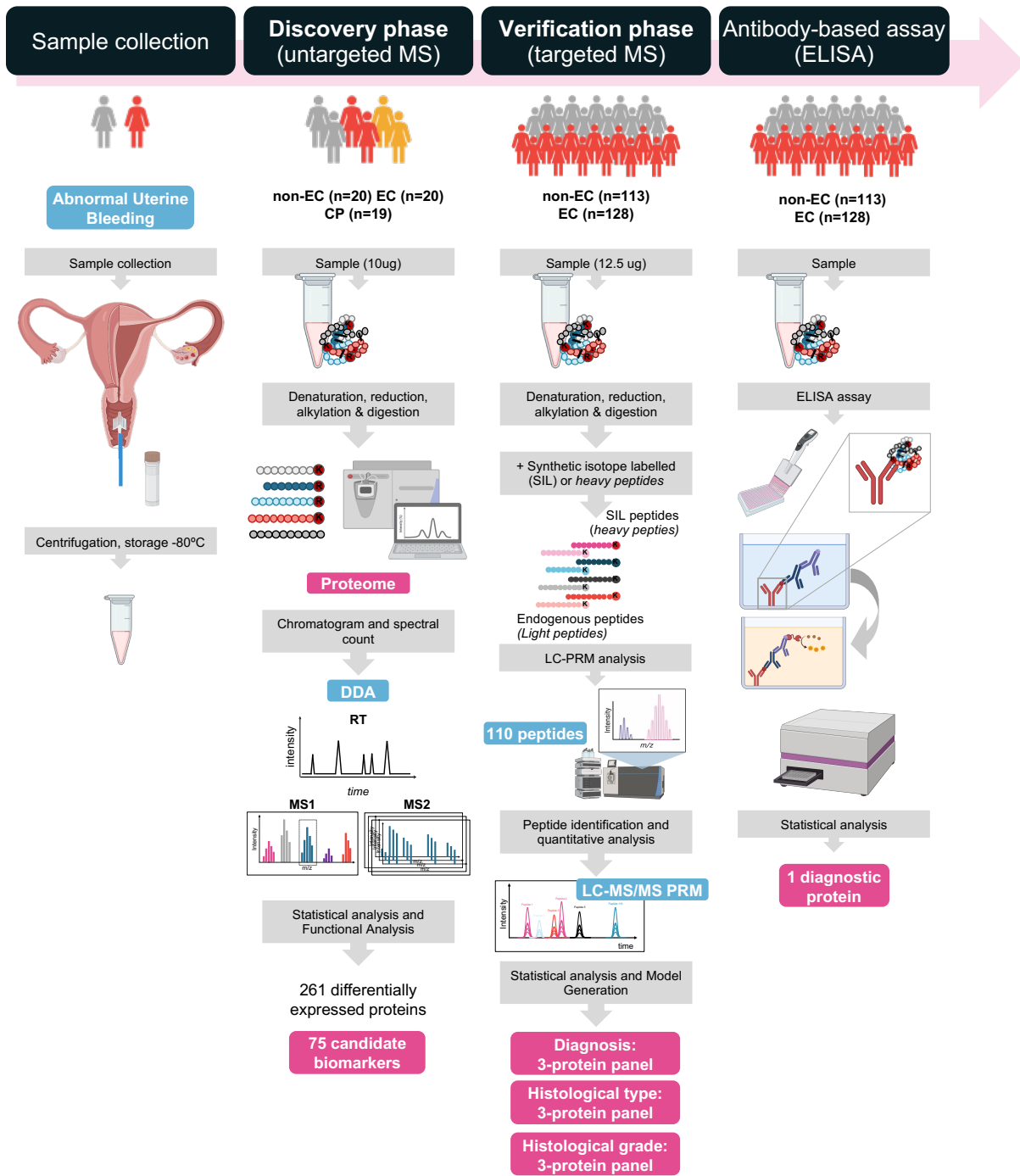
Figure 36 General workflow of the study. EC, endometrial cancer patients; non-EC, non-EC patients; CP, patients with cervical pathology; MS, mass spectrometry; DDA, data-dependent acquisition mode.

Subject characteristics

The study included 300 patients, 281 of which entered the EC diagnostic process due to common symptoms in EC, mainly AUB and/or the presence of a thickened endometrium. Final diagnosis of non-EC and EC cases was achieved by the histopathological analysis of an endometrial biopsy and a surgical specimen after hysterectomy, respectively. The remaining 19 patients were diagnosed with cervical pathology and were included in the discovery study as non-EC controls. The detailed clinicopathological features of all patients are shown in Table 18. Both discovery and verification phases included mostly postmenopausal women, representing the 77% of the women suffering EC, 88% of non-EC patients, and 71% of women with premalignant lesions.

In the discovery phase, all women diagnosed with EC presented an endometrial thickness ≥ 4 mm with an average of 24 mm, while the mean thickness of the non-EC group decreased to 5 mm. Final diagnosis of non-EC patients included the presence of atrophic endometrium in 60% of cases, followed by endometrial polyps (35%) and normal endometrium (5%). Among the 20 EC patients, 65% were diagnosed with endometrioid endometrial cancer (EEC), specifically, 81% were low-grade EEC and 19% were high-grade EEC. The other 30% were diagnosed as non-endometrioid endometrial cancer (NEEC), mostly serous EC. Most EC patients were diagnosed at early stage of the disease (considering early FIGO stages I and II), and only 25% were diagnosed at advanced stages. In addition, a subset of patients attending the gynecology department to diagnose a cervical pathology was included in this study, including 40% of women with normal cytology and negative HR-HPV (high risk HPV), and 60% with positive HR-HPV diagnosed with either normal cytology, or low-grade dysplasia (including ASC-US, LSIL and CIN1 cytologies), or high-grade dysplasia (including CIN2, CIN3, HSIL, ACG, ASC-H), or cervical adenocarcinoma.

In the verification phase, 241 women under EC suspicion were included. Among those, 74% presented a thickened endometrium (≥ 4 mm) with a mean average of 18, 11, and 9 mm for EC, hyperplasia, and non-EC cases, respectively. Specifically, a thickened endometrium was presented in 93% of EC cases, 100% of the hyperplasia, and 67% of non-EC patients. Among the 128 EC women included in the verification phase, 80% were EEC, mostly low-grade (90%), and 20% were NEEC. Additionally, 83% of the EC cases were early-stage cancers, whereas 17% were diagnosed at an advanced-stage. Regarding molecular classification, 28% of cases were not subject to classification due to insufficient amount of tissue available. Among the women that we were able to be classified, 2% were classified as POLEmut, 20% MMRd, 36% NSMP, and 14% p53mut.

In addition to the patient's features, the diagnostic procedure that the patients underwent in this study was analyzed. Data related to the diagnostic process of the patients was available in 94% of the cases (excluding the CP group). The endometrial thickness was assessable in 85% of the women, with an endometrial thickness higher than 4mm, which is a standardized threshold for EC suspicion, encountered in 94% of EC patients and 72% of non-EC patients. Final diagnosis was obtained by the

histopathological examination of an endometrial biopsy in 91% of cases either by pipelle or hysteroscopy or a combination of both techniques. The diagnostic performance of pipelle biopsies was limited in comparison to hysteroscopy, as only 77% of the patients undergoing pipelle were able to be diagnosed, in contrast to 98% of patients undergoing hysteroscopy. Diagnostic failure of pipelle biopsies was related to cervical stenosis in 7% of the patients (1% and 11% of the EC and non-EC patients, respectively) or insufficient material in 16% of the samples (6% from the EC patients and 21% of the non-EC patients). Remarkably, 31% of the patients (15% EC and 49% non-EC patients) had to undergo more than one diagnostic procedure to acquire an adequate endometrial biopsy to be diagnosed, and still 9% of the patients remained undiagnosed with close follow-up indications (pipelle biopsy was the method of choice to biopse these women). Moreover, for those women diagnosed with EC, assessment of preoperative histological type and grade in EC patients did not match with the final diagnosis (resected tissue) in 14% and 41% of cases, respectively, independently of the methodology used. Specifically, from patients diagnosed with EC, 19% of histological types were misdiagnosed by pipelle, while 7% when using hysteroscopy, and the determination of histological grade failed when using pipelle in 38% of cases, while it failed in 36% of EC patients when using hysteroscopy.

	Discovery phase (n=60)			Verification phase (n=241)			Total (n=301)	
	EC (n=20)	non-EC (n=20)	non-EC (PC) (n=20)	EC (n=128)	Hyp (n=7)	non-EC (n=106)	% EC (n=144)	% non-EC (n=123)
Biopsy acquisition issues								
Stenosis (Pipelle)	-	2		2	-	11	1%	11%
Stenosis (Hysteroscopy)	-	1		-	-	-	-	1%
Insuficient material (Pipelle)	2	4		6	-	22	6%	23%
Insuficient material (Hysteroscopy)	-	-		-	-	1	-	1%
Diagnostic procedure								
Pipelle biopsy	11	10		66	2	44	53%	47%
Hysteroscopy	7	1		39	3	1	32%	2%
Pipelle and hysteroscopy	2	9		19	2	47	15%	49%
Pipelle x2	-	-		-	-	1	-	1%
Pipelle and hysteroscopy x2	-	-		-	-	1	-	1%
Diagnostic results - Biopsy								
Histological type failure	1H	-		15P; 4H	-	-	14%	-
Histological grade failure	5P; 3H	-		24P; 21H	-	-	41%	-
Undiagnosed (Follow-up)	-	2P		-	-	9P	-	9%

Table 18 Performance of the standard of care diagnosis. P, pipelle biopsy; H, hysteroscopy

Discovery phase: Shotgun proteomics of cervical fluids

The proteome of Cervical Fluids

The proteome of 59 cervical fluids (20 EC, 20 non-EC, and 19 with CP) was analyzed by untargeted LC-MS/MS, leading to the identification of 2,816 proteins. Interestingly, 93% of the proteome of cervical fluids matched with the proteome of EC tissues analyzed in the CPTAC study, although they only represented the 22% of the EC proteome. Additionally, 46% (229) of the 506 proteins described as EC diagnostic proteins by Martinez-Garcia *et al.* [378] were identified in the proteome of cervical fluids (Annex 4, Figure A4.1-A). The proteins identified in the cervical fluids had mainly immune-related and metabolic properties, being humoral immune response, mRNA catabolic processes, negative regulation of

proteolysis, and translation initiation the main enriched biological functions, which were related to a network of proteins as seen in Annex 4, Figure A4.1-B.

The identified proteins were subjected to statistical analysis to identify potential EC protein biomarkers in cervical fluids. Briefly, the analysis consisted of three different approaches: (1) a differential abundance analysis including proteins present in more than 75% of the patients per group (non-EC vs EC); (2) statistical test to compare the significance between absence/presence of the proteins between groups (non-EC vs EC); and (3) proteins described in the literature as potential EC diagnostic biomarkers that were differentially expressed in a pilot study using cervical fluids performed in our group (Chapter 3). Additionally, to ensure the specificity of the significant proteins for EC, the patterns of expression in non-EC and EC patients for those proteins were compared with their expression in the CP patients.

A total number of 262 proteins were found significantly different between EC and non-EC samples. These proteins were subject to a biological process enrichment analysis, and we mainly observed a regulation of immune responses (cell activation involved in immune response, leukocyte activation involved in immune response, myeloid leukocyte activation, myeloid cell activation involved in immune response, inflammatory response). Importantly, the identified differential proteins were also related to cell secretion, export from cell, and exocytosis (Annex 4, Figure A4.1-B-A). Most of these proteins had binding, structural and regulatory properties mediating cellular and metabolic functions of the cell. According to what was described in the enrichment analysis, the subcellular localization of most of the proteins is the extracellular space (79%), vesicles (77%) and/or membrane (57%), giving evidence that we can identify them in the fluid of the cervical samples.

Specifically, from the 262 proteins significantly different between groups, 115 proteins were differentially expressed in the relative quantification analysis, 42 of which were upregulated and 73 downregulated in EC patients (Figure 37A); 131 proteins were identified in the absence/presence analysis, among which 67 were present and 64 absent in EC patients (Figure 37B-C); and 16 proteins were identified in our pilot study (unpublished data), all being upregulated in EC patients. Remarkably, the 262 proteins were evaluated in the CP cohort, showing that their abundance levels in CP patients were overall like non-EC patients, except for 8 proteins: MYL12A, DBNL, PRDX2, HPGD, NRBP1, PLAT, CLIC3, and PPIB. These proteins were discarded from our biomarker pipeline as they were considered unspecific EC biomarkers. The other 254 proteins were confirmed as specific potential biomarkers to detect EC and no other pathology present in the site of obtention of the sample.

In order to prioritize a set of biomarkers for the verification phase, we mined our statistical data with a threshold of *Adjusted P* value 0.01 for the absence presence study and *Adjusted P* value 0.15 and AUC higher than 0.7 and/or FC higher than 1.8 for the relative quantification analysis; we then discarded members belonging to the same family group, and we finally supported on literature to understand the biological relevance of our biomarkers. We finally chose 75 proteins to be further assessed in the

verification phase in a larger cohort of patients. From those, 40 were selected from the absence/presence study, 20 from the relative quantification test, and 15 from our previous pilot study (Chapter 3). Among the selected proteins, 35 were upregulated or present in EC whilst 40 were downregulated or absent in EC patients (Figure 37).
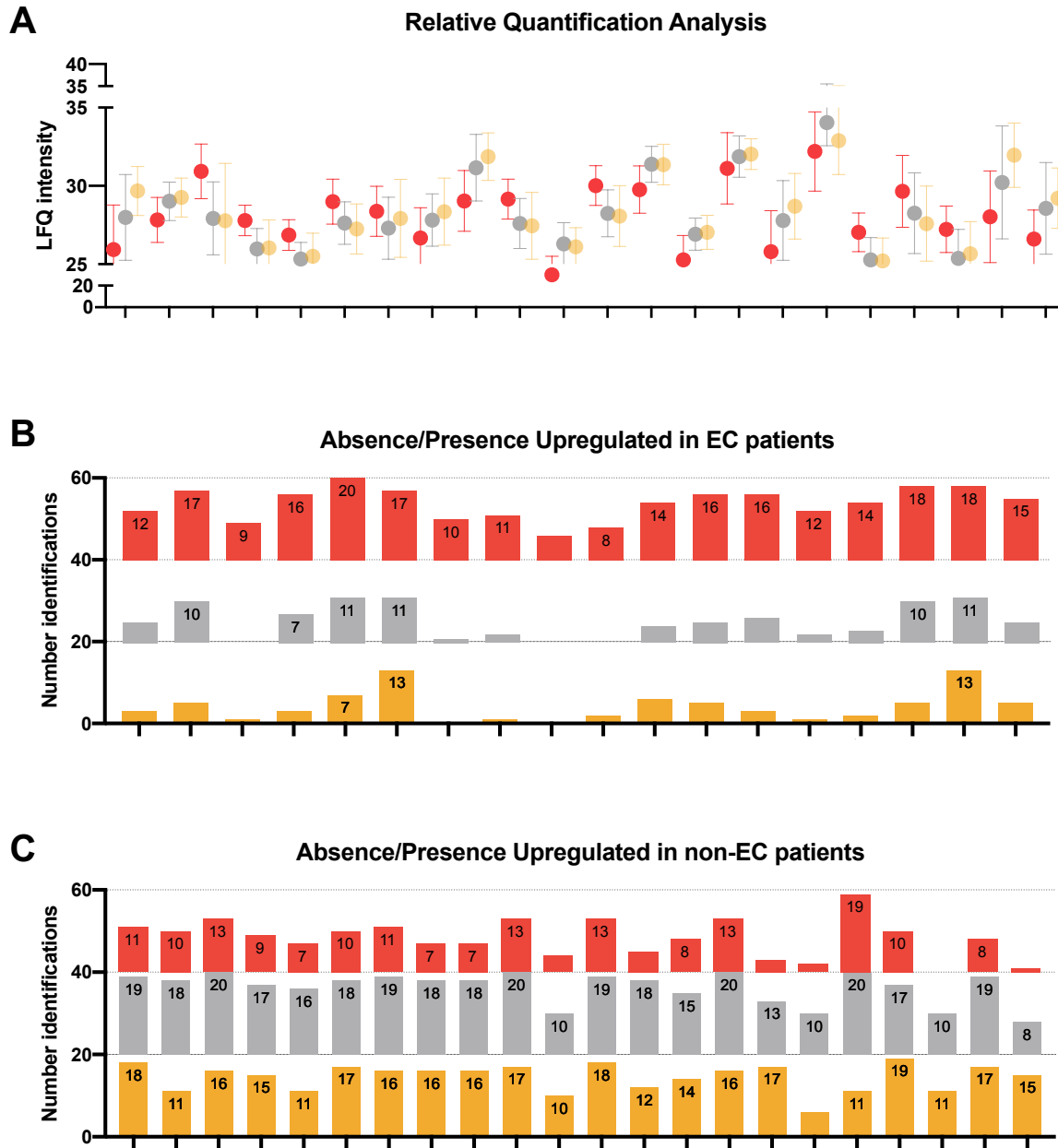
Figure 37 Levels of the selected proteins from the discovery study. (**A**) Proteins selected from the relative quantification analysis. (**B**) Proteins selected from the absence/presence analysis, upregulated in EC patients. (**C**) Proteins selected from the absence/presence analysis, upregulated in non-EC patients. EC patients are represented in red, non-EC patients in grey, and CP patients in yellow.

Verification phase: targeted proteomics of cervical fluids

Diagnostic biomarkers

The verification phase was performed by PRM acquisition to assess the performance of 75 proteins in cervical fluids of 241 patients (106 non-EC, 128 EC, and 7 patients with hyperplasia). In each sample, the relative abundance of the 75 biomarker candidates was measured by the analysis of 110 peptides (40 proteins with 1 peptide and 35 proteins with two peptides). To confirm the identity of the peptides and to detect interferences in the MS signal, we calculated a spectral similarity score. Six peptides detected in less than 50% of the samples were removed and 74 proteins were finally measured corresponding to 104 peptides. The dataset included 27,500 measured pairs of endogenous/synthetic peptides (L/H), 89.9% validated by spectral matching, the other values were rejected due to due to MS signal below the limits of detection and they were replaced by an estimation of the background. The reproducibility of the analytical workflow was evaluated adding three replicates in each batch of samples. The coefficient of variation (CV%) was on average 8%, 10% and 12% for each replicate, and the correlation between the different measurements of each replicate was $R^2>0.99$. Moreover, peptides derived from the same protein showed high correlation across samples, except for PKM that showed correlation below 0.8 because it was monitored by isoform-specific peptides or protein variants. These peptides were considered as derived from different proteins.

The statistical analysis of the verification study unveiled a total of 28 significantly expressed proteins in EC patients compared to non-EC patients (*Adjusted P* value < 0.05 and *Fold Change* > |2|). From those, 26 proteins showed high accuracy to discriminate EC with AUC values higher than 0.70, specifically 17 of those were upregulated in EC, while 9 were downregulated in EC compared to non-EC patients. Remarkably, the best performing biomarkers were PROT with 0.92, PROT with 0.83, PROT with 0.81, and PROT, PROT, PROT, and PROT with 0.8 (Table 19). Importantly, the cohort of patients evaluated in this verification study covers the wide spectra of patients that enter the EC diagnostic process, including not only postmenopausal patients, but also premenopausal women, as well as all types of EC histologies, including the less frequent but more aggressive tumor histologies such as serous carcinomas or other NEEC histologies, *i.e.,* clear cell, carcinosarcomas, mixed, undifferentiated, and mucinous.

EC is associated to high survival rates when it is diagnosed at early stages of the disease. Thus, we assessed the accuracy of the 75 proteins as markers for early detection of EC. Sixty EC patients presented an initial stage IA according to the International Federation of Gynecology and Obstetrics (FIGO) staging, which means that the tumor was still confined to the uterine corpus with less than 50% myometrial invasion [87,92]. Those 60 tumors were compared to the 106 non-EC cases. A total of 20 proteins presented significantly different abundances (*Adjusted P* value < 0.05 and *Fold Change* > |2|) between non-EC cases and early-stage IA tumors. Moreover, 11 of those proteins showed good performance in discriminating early-stage EC from control cases with AUC values higher than 0.7. These

proteins were PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, and PROT (Table 19).

| | Discovery phase (n=60) | | | | | | | | Verification phase (n=241) | | | | | |
| | RQ analysis | | | | AP analysis | | | | | | | | | |
Gene name	FC (EC vs CNT)	Adjusted P value (EC vs CNT)	AUC (EC vs CNT)	CP vs EC / CP vs CNT	PC / CNT / EC	P value	AUC (EC vs CNT)	Method of Choice	FC (EC vs CNT)	Adjusted P value (EC vs CNT)	AUC (EC vs CNT)	FC (EEC-Stage IA vs CNT)	Adjusted P value (EEC-Stage IA vs CNT)	AUC (EEC-Stage IA vs CNT)
PROT						***	0.79	AP	7.01	2E-26	0.92	7.81	8E-16	0.90
PROT						*	0.71	AP	4.26	3E-16	0.83	3.79	1E-08	0.79
PROT						***	0.81	AP	3.31	2E-15	0.81	2.35	7E-07	0.74
PROT	2.52	0.05	0.84	grey				RQ	2.97	2E-14	0.80	2.36	5E-07	0.76
PROT						***	0.79	AP	2.12	2E-14	0.80	1.75	1E-06	0.75
PROT						**	0.76	AP; PS	4.28	1E-14	0.80	4.16	9E-09	0.79
PROT						***	0.71	AP	2.91	6E-14	0.80	2.46	2E-06	0.73
PROT	1.48	0.09	0.77					RQ	2.47	9E-14	0.79	1.96	2E-06	0.74
PROT	1.52	0.11	0.78	grey				RQ	2.27	2E-13	0.79	1.73	1E-05	0.72
PROT						*	0.80	AP	2.24	9E-13	0.78	2.26	3E-07	0.76
PROT						***	0.69	AP	2.49	1E-13	0.78	3.00	4E-10	0.79
PROT						*	0.79	AP	2.16	8E-11	0.76	1.67	2E-04	0.69
PROT						**	0.79	AP	2.54	8E-11	0.76	1.76	4E-05	0.71
PROT						*	0.80	AP	2.11	2E-10	0.75	1.57	4E-04	0.68
PROT	-1.48	0.03	0.80	grey				RQ; PS	-2.04	4E-10	0.75	-1.63	2E-03	0.66
PROT						***	0.87	AP	-2.52	4E-10	0.74	-2.19	4E-05	0.71
PROT						***	0.71	AP	3.90	8E-11	0.74	2.94	1E-04	0.68
PROT	-1.93	0.12	0.68	grey				RQ	-3.26	1E-09	0.74	-2.49	2E-04	0.69
PROT	1.09	0.11	0.73	grey				RQ	2.14	1E-09	0.74	1.46	2E-03	0.65
PROT	-1.95	0.15	0.67	grey				RQ	-2.99	1E-08	0.72	-2.48	7E-04	0.67
PROT	-1.87	0.07	0.76	grey				RQ	-2.97	1E-08	0.72	-2.40	7E-04	0.67
PROT						*	0.66	AP	-2.50	1E-08	0.72	-2.09	1E-04	0.69
PROT						*	0.67	AP; PS	2.30	2E-08	0.72	2.16	3E-05	0.71
PROT						*	0.63	AP	2.33	1E-08	0.71	1.63	3E-03	0.64
PROT						*	0.74	AP	-2.83	1E-07	0.71	-2.40	3E-03	0.65
PROT	-1.64	0.12	0.70	grey				RQ	-2.26	1E-07	0.71	-2.08	3E-04	0.68
PROT						*	0.68	AP	-1.93	1E-08	0.72	-2.46	5E-08	0.78

Table 19 Diagnostic biomarker proteins for EC. Proteins presenting an *Adjusted P* value < 0.05, *Fold Change* > |2| and AUC > 0.7 in the verification phase (n=241) for any of the comparisons are shown (EC *vs* CNT or early-stage IA EC *vs* CNT). Results obtained in the discovery phase (n=59) for those proteins are also given for the relative quantification (RQ) analysis and the absence/presence (AP) analysis, and the ones also belonging in the pilot study (PS) are indicated. The proteins that were significant for the comparison of the cervical pathology (CP) with EC and CNT in the RQ analysis (*Adjusted P* value < 0.05) are highlighted in grey in the corresponding columns. Bars on the presence/absence analysis represent the number of patients in which each protein was identified. FC: *Fold Change;* EC: endometrial cancer; CNT: control or non-EC cases; CP: cervical pathology cases.

Diagnostic biomarker panel

EC is a multifactorial disease and thus, a single biomarker is unlikely to be able to detect all EC cases. The use of a panel of biomarkers is probably necessary to be both highly sensitive and specific. Using an *in-house* R script, individual EC biomarkers were combined into all possible panels from two to five proteins. The combination of PROT, PROT and PROT significantly improved the diagnostic power of the individual markers, achieving an AUC of 0.954 (95% Confidence Interval (CI) 0.927-0.976) with sensitivity of 95.4% (95%CI 91.4-98.4%) and specificity of 85.7% (95%CI 79.2-92.5%) (Figure 38A-B). The addition of more proteins only enhanced marginally the performance of the panel and, therefore, were not considered. To assess the robustness of the panel, a "k-fold" cross-validation (k=10) procedure was used. Our panel achieved 95.4% sensitivity and 85.7% specificity, including early-stage EC, in the cross-validation, providing correct diagnosis to 122/128 EC patients, and 91/106 non-EC patients. The panel was able to detect 71.4% (5/7) of hyperplasia as EC (Figure 38C). Remarkably, the panel is able to discriminate control patients from patients with EC with high sensitivity independently

of the histology, grade, or molecular subtype of the tumor, highlighting the strength of the signature (Figure 38C).
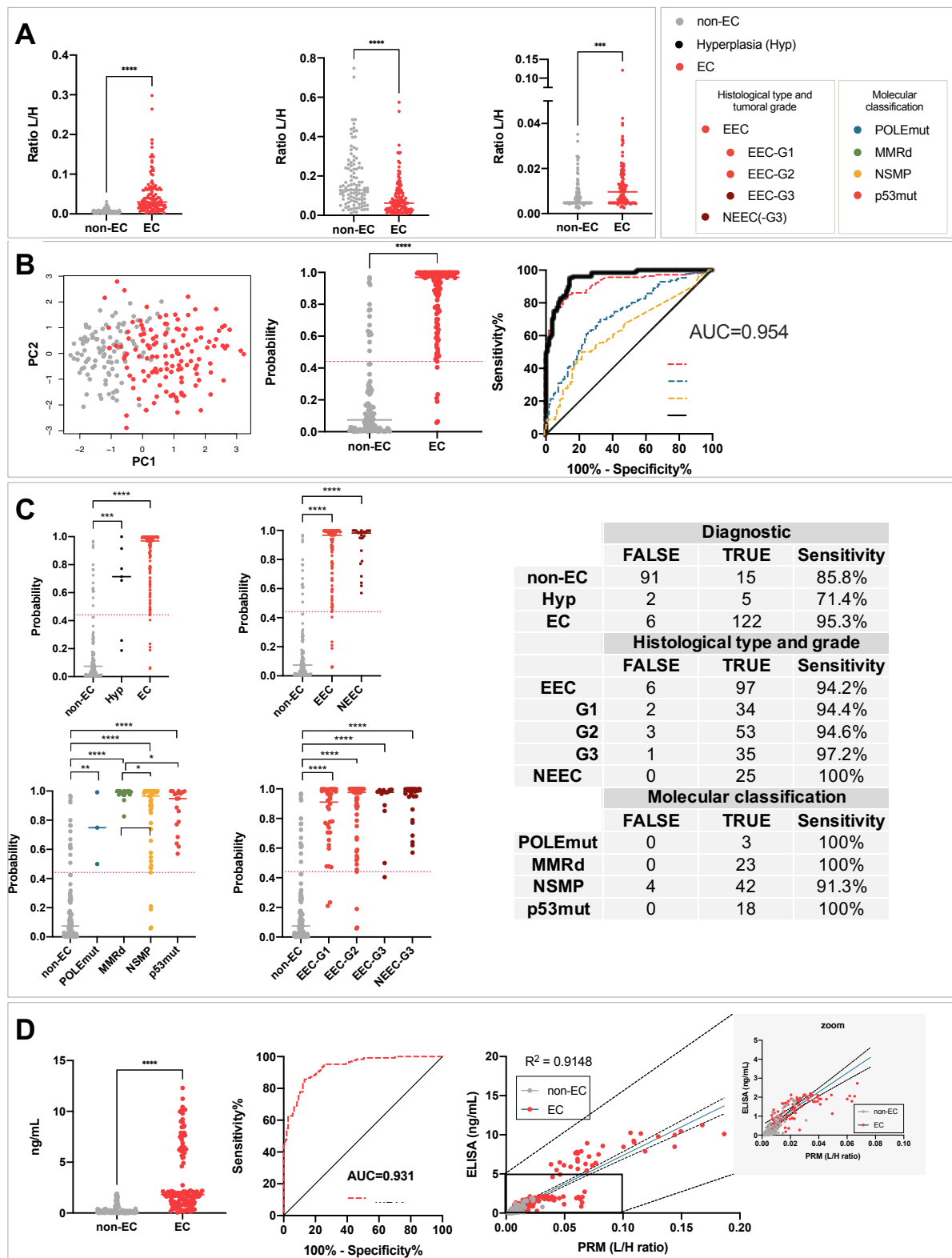
Figure 38 Performance of individual biomarkers and the diagnostic panel to discriminate between non-EC and EC patients in the cervical fluids. (**A**) Dot plots of the individual proteins composing the diagnostic panel

representing the light/heavy ratios (L/H) obtained by LC-PRM across 106 non-EC and 128 EC patients. (**B**) Principal component analysis (PCA), dot plot and ROC curve of the 3-protein panel formed by MDK+ANXA7+CSE1L constructed by the logistic regression model across the patients included in the verification phase. The ROC curve includes the individual ROC curves of the proteins included in the 3-protein panel. Additionally, intervals of confidence of the sensitivity, specificity and AUC values are shown. (**C**) Detailed description of the discrimination accuracy of the diagnostic panel across pre-malignant disease (hyperplasia) and different histopathological (histological type and tumoral grade) and molecular classifications of the endometrial tumors. Dot plots of each comparison are shown, as well as a comprehensive table showing the patients well classified, the incorrectly classified, and the consequent sensitivity of our protein-panel. (**D**) Dot plot and ROC curve depicting the distribution of MDK concentration across the 241 patients from the verification phase using ELISA assay. Additionally, representation of the correlation between LC-PRM and ELISA assay results evaluated in those 241 patients. TP, true positive; TN, true negative, FP, false positive; FN, false negative.

Importantly, we evaluated the diagnostic performance of the current standard of care in comparison to our results. In this study, the histopathological examination of the biopsies obtained by aspiration, which is currently the preferred EC diagnostic method, failed in 23% of cases (41 patients) to provide a final diagnosis, either due to technical issues (*i.e.,* stenosis) or insufficient material. In these patients, the 3-protein panel here developed provided an accurate and correct diagnosis in 91% of cases (90% when also considering hyperplasia). Remarkably, the 3-protein panel diagnosed as diseased the 8 EC patients that presented an endometrial thickness <4mm in the TVUS in this study. This is also an important finding as these patients are not usually referred for endometrial biopsy if they do not present persistent symptoms and/or risk factors (Figure 39). Additionally, those misdiagnosed patients have an endometrial thickness greater than 4, meaning that they would undergo additional tests, even if the protein tool fails.

Transferability to ELISA

Mass spectrometry allowed for the accurate measurement of a high number of proteins in one single run and therefore, it was a great technique for the development of protein panels. However, ELISA methodology is currently the gold-standard and it is implemented worldwide in clinical laboratories for measuring protein concentrations presenting several advantages [184,379]. To move our results to a clinical use, we evaluated the levels of PROT, our best-performing biomarker, using a commercially available ELISA kit in the same 241 patients included in the verification study, and the correlation with the results obtained by LC-PRM was calculated. The quality of the ELISA performance was assessed by the median overall CV (%), which was lower than 5%. By using the antibody-based technology, we confirmed the accuracy of PROT to discriminate between EC and non-EC patients, showing an AUC value of 0.93 (Figure 38D). The median concentration of PROT in non-EC patients was 0.12 (95% confidence interval (CI), 0.08-0.18), whereas in EC patients was 1.82 (95% CI, 1.63-1.98). Absolute levels of PROT in the ELISA (ng/mL) and relative levels (light/heavy ratios) in the PRM acquisition showed a linear correlation, with a R^2 value of 0.91 (Figure 38D).
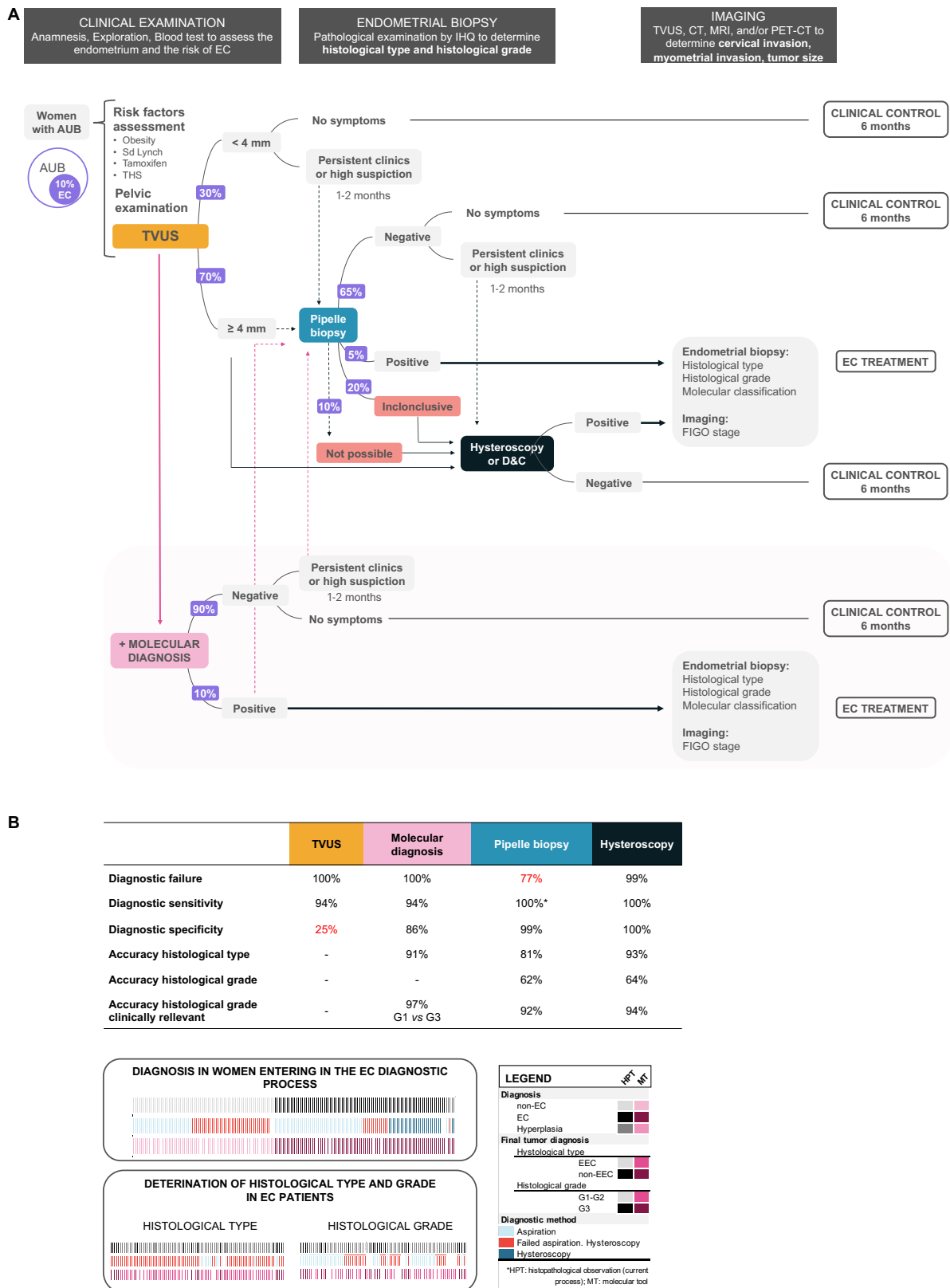
Figure 39 Current patient pathway for diagnosis of EC. (**A**)Integration of our molecular tool in the results and how it would impact on the current clinical pathway. (**B**) Performance of the different methods to diagnose EC. * 3-protein panel regarding histological grade is between grade 1 and grade 3, while results from endometrial biopsies include grade 2. Sd Lynch, Lynch Syndrome; TVUS, transvaginal ultrasonography.

172

Prognostic biomarkers

In addition to diagnosis, histopathological examination of the cells contained in endometrial biopsies should provide information on different risk factors, mainly histological type (EEC or NEEC) and grade (G1, G2 or G3), to classify EC patients in different risk groups and accordingly, guide their surgical treatment. Thus, we investigated whether the proteins assessed as diagnostic biomarkers in the verification phase had the potential to distinguish between histological subtypes and/or grades of the EC cases.

Regarding tumor histologies, NEEC cases are less frequent (~15% of all EC cases) but much more aggressive than EEC tumors (accounting for 40% of EC-related deaths) [12], and thus, they require an extensive surgery. Comparing the cohort was split in 103 EEC and 25 NEEC cases, this latter formed by 36% serous, 28% carcinosarcoma, 16% mixed, 8% clear cell, 8% others, 4% undifferentiated, cases. We found that the levels of 29 proteins were significantly different between these groups (*Adjusted P* value < 0.05) (Figure 40A). Remarkably, four of them presented a *fold change* greater than 2 and AUC values higher than 0.7. Specifically, PROT was more abundant in EEC cases distinguishing them with an accuracy of 0.85, while PROT, PROT, and PROT presented higher levels in NEEC cases with AUC values of 0.80, 0.74, and 0.72, respectively.

Regarding tumor grades, while all NEEC cases are high grade (G3), EEC cases can be graded from 1 to 3 (G1, G2 and G3). Lower grades (G1+G2) are related to good prognosis in contrast to G3. Here, we compared the levels of the 75 proteins between the cohorts of 101 EEC cases divided in 36 G1, 55 G2, and 12 G3. First, we compared low-grade (G1+G2) *vs* high-grade (G3) tumors, followed by G1 *vs* G3. A number of 15 proteins showed significantly different levels between low-grade and high-grade EEC (*Adjusted P* value < 0.05). From those, five proteins also presented a *fold change* greater than 2 and AUC values higher than 0.7. Those proteins were PROT (AUC of 0.77), PROT (AUC of 0.73), PROT (AUC of 0.73), PROT (AUC of 0.79), and PROT (AUC of 0.77), being all of them more abundant in high-grade EEC tumors. Regarding the comparison between G1 *vs* G3, eleven proteins were significantly different between groups (*Adjusted P* value < 0.05), and 7 presented a *fold change* higher than 2 and AUC values greater than 0.7. Among these proteins, PROT (AUC of 0.85), PROT (AUC of 0.82), and PROT (AUC of 0.79), were more abundant in high-grade endometrioid tumors, whereas PROT (AUC of 0.77), PROT (AUC of 0.75), PROT (AUC of 0.73), and PROT (AUC of 0.75) were more abundant in low-grade (G1) tumors (Figure 40A).

Prognostic biomarker panel

To improve the performance of the individual markers, we developed protein panels following the same procedure as previously described for the development of the diagnostic panel. Combinations from two to five proteins were assessed from all the proteins measured in the verification study. Two 3-protein panels allowed for the accurate discrimination between histological subtypes of EC (EEC *vs* NEEC) and

grades of EEC (G1 *vs* G3). The robustness of these protein panels was further assessed using the "leave-one-out" cross-validation method.

The three-protein panel able to discriminate between EEC from NEEC consisted of the combination of PROT, PROT and PROT (individual dot plots in Figure 40B). It achieved an AUC of 0.911 (95%CI 0.845-0.961) with a sensitivity of 80% (95%CI 64-96%), and specificity of 91.3% (95%CI 85.4-96.1%). The combination was able to classify 94 out of 103 EEC cases and 19 out of 25 NEEC cases (Figure 40C-F). The three-protein panel able to discriminate between grade 1 EEC and grade 3 EEC consists of PROT, PROT, and PROT proteins (individual dot plots in Figure 40D). This combination achieved an AUC value of 0.975 (95%CI 0.91-1) with a sensitivity of 91.3% (95%CI 85.4-96.1%) and a specificity of 100% (95%CI 100-100%). The combination classified 36 out of 36 grade 1 EEC tumors and 11 out of 12 grade 3 EEC tumors (Figure 40E-F).

Predictive biomarkers

In 2013, the TCGA Research Network defined a novel EC classification based on molecular features of the tumors, differentiating four different subgroups with different clinical outcome, from better to worse prognosis: POLEmut, MMRd, NSMP, and p53mut [44]. However, the translation of this classification into the clinical practice is not possible before surgery due to the lack of material, and it is still difficult to implement after surgery due to the requirement of high-throughput techniques not always available in clinical laboratories, or the lack of material, especially in early-stage carcinomas [57]. . Therefore, we also evaluated the performance of the targeted proteins to discriminate between the different molecular subtypes in our cohort of patients: 3 POLEmut, 25 MMRd, 46 NSMP, and 18 p53mut tumors. The POLEmut subgroup was discarded to perform statistical analysis, as its size was not big enough. A total of 28 proteins showed significantly different levels between at least two groups, excluding POLEmut (*Adjusted P* value < 0.05, Annex 4, Figure A4.2-A): 22 proteins for the comparison MMRd *vs* NSMP, 3 for the comparison NSMP *vs* p53mut and 3 for both comparisons (LBP, VWF, and GPLD1, Annex 4, Figure A4.2-B). Remarkably, 9 proteins (PROT, PROT, PROT, PROT, PROT, PROT, PROT, PROT, and PROT) also showed a *fold change* higher than 2. Finally, even though we could not evaluate proteins that allowed for the discrimination of the POLEmut cases *vs* other subtypes due to the low number of patients in this group (n=3), we studied the behavior of the proteins within the different subtypes, and pointed out PROT, PROT, and PROT as possible candidates to tackle in further studies designed for this purpose (Annex 4, Figure A4.2-C).

	Verification phase (n=128 EC)								
	Histological type			Histological grade (only EEC)					
Gene name	FC (EEC vs NEEC)	Adjusted P value (EEC vs NEEC)	AUC (EEC vs NEEC)	FC (G1G2 vs G3)	Adjusted P value (G1G2 vs G3)	AUC (G1G2 vs G3)	FC (G1 vs G3)	Adjusted P value (G1 vs G3)	AUC (G1 vs G3)
PROT	-5.02	5E-06	0.85	-1.20	3E-02	0.77	-4.36	5E-07	0.84
PROT	2.64	9E-05	0.80	1.24	2E-01	0.54	1.04	1E-03	0.52
PROT	2.25	3E-03	0.74	1.30	2E-01	0.56	1.02	8E-03	0.52
PROT	2.74	1E-02	0.72	1.47	2E-01	0.60	1.48	9E-03	0.63
PROT	1.65	2E-02	0.70	1.60	3E-02	0.68	2.23	2E-04	0.77
PROT	1.33	2E-02	0.69	-1.58	1E-01	0.63	2.06	1E-04	0.75
PROT	1.80	2E-02	0.69	1.62	1E-01	0.61	2.38	3E-04	0.73
PROT	-2.11	2E-02	0.68	-1.96	3E-02	0.68	-1.92	6E-04	0.76
PROT	1.71	2E-02	0.68	1.14	3E-02	0.67	2.39	6E-04	0.75
PROT	-1.84	2E-02	0.68	-3.65	3E-02	0.77	-2.16	5E-04	0.82
PROT	-2.00	5E-02	0.65	1.29	3E-02	0.75	-3.93	4E-03	0.79
PROT	-1.19	5E-01	0.55	-2.74	3E-02	0.73	-2.22	5E-02	0.77
PROT	1.11	7E-01	0.53	-2.60	3E-02	0.73	-2.06	4E-01	0.74
PROT	1.27	8E-01	0.52	-2.74	3E-02	0.79	-2.96	2E-01	0.81
PROT	1.00	9E-01	0.51	-2.60	3E-02	0.77	-3.14	9E-02	0.79

Figure 40 Prognostic biomarkers for EC. (A) Table showing proteins able to discriminate (*Adjusted P* value < 0.05, *Fold Change* > |2| and AUC > 0.7) between histological types and/or tumor grades (EEC *vs* NEEC and/or EEC-G1G2 *vs* EEC-G3 and/or EEC-G1 *vs* EEC-G3). **(B)** Dot plots of the individual proteins composing the protein panel to determine histological type representing the light/heavy ratios (L/H) obtained by LC-PRM across 102 EEC and 25 NEEC patients. **(C)** Dot plot and ROC curve constructed by the logistic regression model formed by PIGR+LBP+PKM to discriminate between the histologies (EEC or NEEC) of the 128 EC patients included in the verification phase. The ROC curve includes the individual ROC curves of the proteins composing the 3-protein panel. **(D)** Dot plots of the individual proteins composing the protein panel to determine histological grade representing the light/heavy ratios (L/H) obtained by LC-PRM across 92 EEC-G1G2 and 11 EEC-G3 patients. **(E)** Dot plot and ROC curve constructed by the logistic regression model formed by PIGR+HSPE1+SPP1 to discriminate between the tumoral grades (EEC-G1 *vs* EEC-G3) of EC patients included in the verification phase. The ROC curve includes the individual ROC curves of the proteins composing the 3-protein panel. **(F)** Table with the CI of the resulting sensitivity, specificity and AUC value of each model, the table also reports TP, true positives, TN, true negatives, FP, false positive, FN, false negatives.

DISCUSSION

Human body fluids demonstrated to be a potential source of clinically relevant proteomic biomarkers [196,380]. Cervical fluid is a dynamic fluid composed by secretions originated in the vagina, cervix, uterus, fallopian tubes and ovaries, but also other components such as exfoliated cells, bacterial products, blood, among others [381]. As its ratio volume/organ is small, the possibility of dilution of biomarkers is lower compared to plasma samples. Therefore, cervical fluid is pointed as a promising source of biomarkers related to gynecological disorders [371,382]. Regarding EC, many studies aimed to detect EC cells within cytologies, however, a recent review highlight the poor sensitivity of this approach, as the pooled sensitivity obtained from 17 different studies was 45% [207].

In this study we aimed to (i) characterize the cervical fluid of non-EC and EC patients; (ii) to identify novel protein biomarkers contained in the cervical fluids to rule out EC from other benign diseases of the female genital tract; (iii) define a protein biomarker signature to sensitively detect EC using a non-invasive sample (*i.e.,* cervical fluid obtained with a cervical brush); and (iv) define protein biomarkers related to histological type and grade. To our knowledge, this is the first study that has identified protein biomarkers in cervical samples, specifically in cervical fluids, to detect EC. The investigation here presented had followed strict practices and considerations that must be taken for mass-spectrometry-based protein biomarker discovery and verification compiled along years of experience from similar studies [182].

First, the proteome of the cervical fluid was characterized using samples from 60 patients (20 non-EC patients suffering AUB, 20 EC patients, and 20 patients with CP). The identified proteome was composed by 2,816 proteins. Interestingly, the 93% of the proteins contained in the cervical fluids were contained in the proteome of EC tissues identified by the CPTAC [298], and it covered 46% of the proteins that were already described as EC diagnostic proteins from the literature [190]. This confirmed not only that the proteome concealed a broad range of proteins but also that it covered a broad spectrum of proteins related to EC, thus enabling the use of these samples for the quest of EC biomarkers.

Second, the identified proteins were subjected to statistical analysis to identify potential protein biomarkers able to differentiate non-EC from EC patients. Three different approaches were performed: a differential abundance statistical analysis for those proteins present in >75% of patients from each group, and statistical test to compare the significance between absence/presence of proteins between groups (non-EC *vs* EC). Additionally, we consider proteins coming from a previous pilot study. A total number of 262 proteins were recognized as significant between EC and non-EC. The nature of those proteins was mainly related to immune responses, cellular and metabolic functions of the cell, and cell secretion, composed mainly with proteins that appeared in the extracellular matrix, membrane and secretions, as expected as our proteins are contained in a fluid sample.

176

Among the 262 proteins, 75 were selected to be further verified in cervical fluids of a cohort of 241 patients (128 EC, 7 atypical hyperplasias, and 106 non-EC). The results allowed us to identify X significant proteins, X upregulated and X downregulated in EC, and to define a 3-protein panel able to detect EC from other benign diseases. Importantly, a 3-protein panel was able to diagnose the percentage patients that could not be diagnosed by the histopathological examination of a pipelle biopsy due to stenosis or insufficient cellular material.

The diagnostic panel is composed of three proteins: PROT, PROT and PROT. PROT PROT PROT

Despite the main goal of the study was to find EC diagnostic biomarkers, we also tested the feasibility to use the 75 selected proteins to diagnose the histological type and grade of EC, and consequently to improve the preoperative risk assessment of EC patients [64]. Our data showed that 29 and 15 proteins are differentially expressed between histological types and grades, respectively (*Adjusted P value* < 0.05); and that two combinations of three proteins accurately discriminate between histological types (EEC *vs* NEEC) and grades (G1 *vs* G3). The combinations found were composed by PIGR, PKM and LBP for discriminating histological type, reaching a sensitivity of 80% and specificity of 91%. Regarding histological grade, the three-protein signature was composed by PROT, PROT, and PROT achieved a sensitivity of 91% and 100% specificity.

Interestingly, both combinations incorporated the PROT, PROT, PROT

A clinical strength of this investigation is that women enrolled in both discovery and verification studies covered the broad variability of women entering the EC diagnostic process. Patients without EC covered all women with suspicion of EC due to AUB or thickening of the endometrium [383], including women diagnosed with benign pathologic conditions (polyps, myomas, and endometrial hyperplasia), and women with normal endometrium. Regarding EC patients, they also covered the broad clinical spectrum of the disease, which includes the most common EEC histological type, as well as SEC, as the predominant NEEC subtype, but also other subtypes such as clear cell or mixed adenocarcinomas. EEC patients also covered all the range of grades. As in the clinical reality, the majority of women also presented an early-stage IA EC, but women diagnosed at late stages were also included. Finally, despite 25% of the women could not be classified according to the molecular classification , the groups included were in accordance to the percentages previously described [44,77]. Finally, the study included women at different stages of their reproductive age and therefore functional and atrophic endometrium for premenopausal and postmenopausal women, respectively. Despite the molecular differences and existing heterogenicity in EC, our diagnostic biomarkers permitted the accurate differentiation between EC and non-EC women independently of all these factors.

Another important strength of this study is the technical side. Discovery and verification phases were performed on high-resolution accurate mass spectrometers that enabled us to test multiple proteins at once. First, we used a non-targeted approach operated in DDA mode to characterize the proteome of

the cervical fluids and to identify potential differential proteins that might distinguish between EC and non-EC patients. Next, we used the last generation of targeted MS method, the LC-PRM acquisition. Clear advantages of the workflow here presented are the increased chances of finding a clinically relevant biomarker, and combine those ones among them to develop protein panels, as their abundances can be measured at once in large cohorts of patients [234]. Lastly, as a step toward clinical implementation, we used ELISA assay to evaluate the leading diagnostic protein of our panel, as immunoassays are the gold-standard method for clinical validation and application in clinical setting. The commercially available ELISA kit against PROT perfectly matched the results yielded in LC-PRM. Unfortunately, suitable commercial ELISA kits for the other two proteins contained in the diagnostic panel (PROT and PROT) were not available and therefore, the development of *in-house* antibodies for those proteins would be needed. Next steps in this direction would be the development of a highly specific and reproducible assay that could be automatized for the clinical application. Alternatively, our protein panels could be implemented in the clinical routine using an MS method allowing faster tests. Last year, the commercialization of new generation of fully automatized MS for clinical application have been developed and boosted, and they are currently available in some clinical laboratories. Despite they are still only available for small molecules, the trend is to promptly move into large protein analysis.

A limitation of the study to transfer the results in the clinics is that cervical fluids are currently not collected in the standard routine for the purpose reported in this study. Despite the acquisition of a cytology to inspect the presence of malignant endometrial cells or to screen for the Human papilloma virus (HPV)-test is a common practice, those samples are collected in a fixative solution which hampers its use for protein measurements. Research in the preanalytical preparation of the sample should be performed in order to adequate the sample collection process. Nonetheless, the brush used for the acquisition of the cervical fluids in this study is nowadays available in all the physician's office, as it is the one used to routinely test women for HPV. Another important limitation is to advance on the biomarker pipeline, and consequently, the results presented here need to be further validated in an independent, prospective and multicentric clinical trial.

The study covers an important clinical need in EC. From one side, the diagnostic molecular signature identified in cervical fluid would perfectly complement the current diagnostic procedure reducing the number of invasive biopsies needed and providing diagnostic to all patients. Additionally, other two signatures also provide information regarding prognostic features of the tumors, which can give support to the pathologists' observations at the time of assessing these parameters, cooperating in the decision of the optimal surgery treatment. Altogether, the development of cervical fluid protein biomarker signatures is expected to improve the comfort of the patients, the management of EC, and save great sanitary costs.

Discussion

From the identification of the clinical need..

Cancer is nowadays one of the leading causes of death globally, accounting for nearly 10M deaths and 19.3M new cases in 2020. Additionally, cancer is expected to increase its incidence in the following years reaching an estimated number of 30.2M new cases in 2040 [384]. Hence, creating a big burden for the healthcare systems. According to The World Health Organization (WHO), despite 30-50% of cancer cases can be currently prevented by avoiding risk factors and following healthy life-styles, early diagnosis of the disease and pre-cancer screening are the two main strategies to promote early detection and to ensure a more effective treatment [385]. **Early diagnosis** involves awareness of the population of potential symptoms and seek for medical advice, and their accessibility to clinical evaluation and diagnostic services. Differently, **screening** programs aim to identify individuals suffering the disease before developing symptoms.

In the era of precision medicine, interest in precision diagnostics has significantly increased during the last years. In this context, the seeking of biomarkers received major attention, as they are the cornerstone for early diagnosis of cancer [386,387]. The **ideal biomarker** for cancer diagnosis would early diagnose cancer but also provide clinicians with additional useful information regarding tumor features, predict the progression of the disease and/or allow the selection of an optional treatment at an early stage of the disease. The ideal biomarker should also be present in the clinical specimen for a sufficient long-time window and ideally in a non-invasive matrix, where it could be measurable by practical assays providing robust and highly accurate results [190,388]. Particularly proteins seem to be the perfect molecules to accomplish those requirements. They are the basic units of the cell and they define the status of an organism at a specific moment. Proteins can be measured in biofluids and using cost-effective techniques already implemented in clinics. Consequently, the proteome analysis to solve a clinical need, also known as clinical proteomics, represents a major field in the area of proteome research. Despite the increasing number of manuscript published on this topic during the last years [389], and all the efforts and investments made to find protein cancer biomarkers of clinical utility, only few biomarkers have been approved by the FDA over the last 20 years [184,390].

Regarding EC, 70% of women are diagnosed at early stages of the disease when the tumor is still localized within the endometrium, which is associated with an overall 5-year survival rate of 95% [20]. However, while most of the common cancer types benefit from advances in the early diagnosis and improvement of therapies improving their average five-year survival rates from 50.3% (over the period of 1970-77) to 67% (over the period 2007-2013), only cervix uterine cancer and specially uterine cancer, shown a decrease in the survival (from 86.9% to 82.3% over the same periods of time) [21]. Hence, indicating that current EC diagnostic procedure still needs to be improved. Especially, considering that EC incidence is expected to raise dramatically from 417,000 new cases in 2020 to 608,000 expected cases in 2040 mainly due to increased life expectancy and overall prevalence of obesity and metabolic

syndromes associated with the disease. Hence, the development of tools that allow a progress in the current diagnosis of EC would benefit to a large number of women.

Currently, there is no screening program for EC diagnosis. Only the presence of symptoms, an endometrial thickness observed during a normal exploration, and/or the diagnosis of Lynch Syndrome would made women enter the multistep process of EC diagnosis. Abnormal uterine bleeding (AUB) is the most common symptom of EC, being present in 90% of EC cases [99], however, it is highly unspecific since it is also associated to other benign diseases and only 5-10% of patients with AUB will ultimately be diagnosed with EC. Still, all women presenting AUB will start the diagnosis procedure to rule out EC. The process begins with a pelvic examination and transvaginal ultrasonography, followed by the histopathological examination of an endometrial biopsy, preferably obtained by a minimally-invasive aspiration (*i.e.* uterine aspirates or pipelle biopsies) [116]. Unfortunately, it is associated with important drawbacks, mainly due to the limited number of cells contained in the biopsy. First, up to 42% of women do not get a final diagnosis with this biopsy due to technical problems (*i.e.* cervical stenosis) or insufficient material found at histology [133,391]. Hence, non-diagnosed women will have to undergo an invasive procedure such as hysteroscopy. Once diagnosed, EC patients will need undergo a preoperative risk assessment to guide the surgical treatment, which is the cornerstone treatment. This assessment is based on different prognostic factors. Among those, the main ones are histological subtype and grade, which are determined thanks to the histopathological examination of an endometrial biopsy; and the degree of tumor extension and dissemination, which is determined by imaging tecniques. Since pathological examination is a subjective determination, there exists a great interobserver variation in diagnosing the histological type and grade, which leads to up to 55% of incorrect tumor features determination [136,155]. In addition to these parameters, there are no tools to predict recurrence in an independent manner [64].

In this challenging but exciting context, the main goal of this thesis was to identify the current clinical gaps in the diagnostic procedure of EC and fulfill that gap with highly accurate, robust, and effective protein biomarkers and biomarker signatures. To achieve that, we followed different sampling methods, approaches, and strategies. First, using the fluid of the currently used endometrial biopsies obtained by aspiration, we aimed to develop the first omics molecular study to provide clinicians more accurate preoperative information regarding EC histological type and tumor grade to improve the preoperative risk assessment of patients and further guide the optimal treatment of EC patients. Additionally, we aimed to identify an accurate biomarker and biomarker signature able to predict the risk of recurrence of EC patients, which was independent of the currently available prognostic factors. Second, we aimed to approach a non-invasive diagnosis of EC. For this, we explored the use of cervical fluids as a source of protein diagnostic, histological type and grade biomarkers of EC. The use of cervical fluids to detect EC will be a change in the paradigm on how we manage EC patients, moving a diagnosis that today requires from invasive, unpleasant, and painful techniques to a simple, quick, and non-invasive test.

This thesis originated as a result of previous work from two thesis, where diagnostic protein biomarkers and signatures were identified in the depleted fluid of pipelle biopsies and in the exosomes contained in the raw fluid of the pipelle biopsies. Both studies went from discovery to verification and validation phases where a total of 291 patients were analyzed by mass-spectrometry but also ELISA assay [190,200,203,392] (Figure 41, "Previous work"). Results from those theses were transferred to MiMARK Diagnostics S.L., a spin-off founded in January 2021, in which I actively participated while performing this thesis. MiMARK aims to transfer those results into the market through developing a prototype based on immunoassays with proprietary antibodies and achieving an analytical and clinical validation. Hopefully, the results of this thesis are also advanced to develop an IVD test that could be launched into the clinical practice in the near future. An overview of the structure of this thesis divided in different chapters describing the results to achieve the objectives purposed, as well as future steps required for the implementation of the biomarkers and biomarker signatures in the clinical practice are shown in Figure 41.
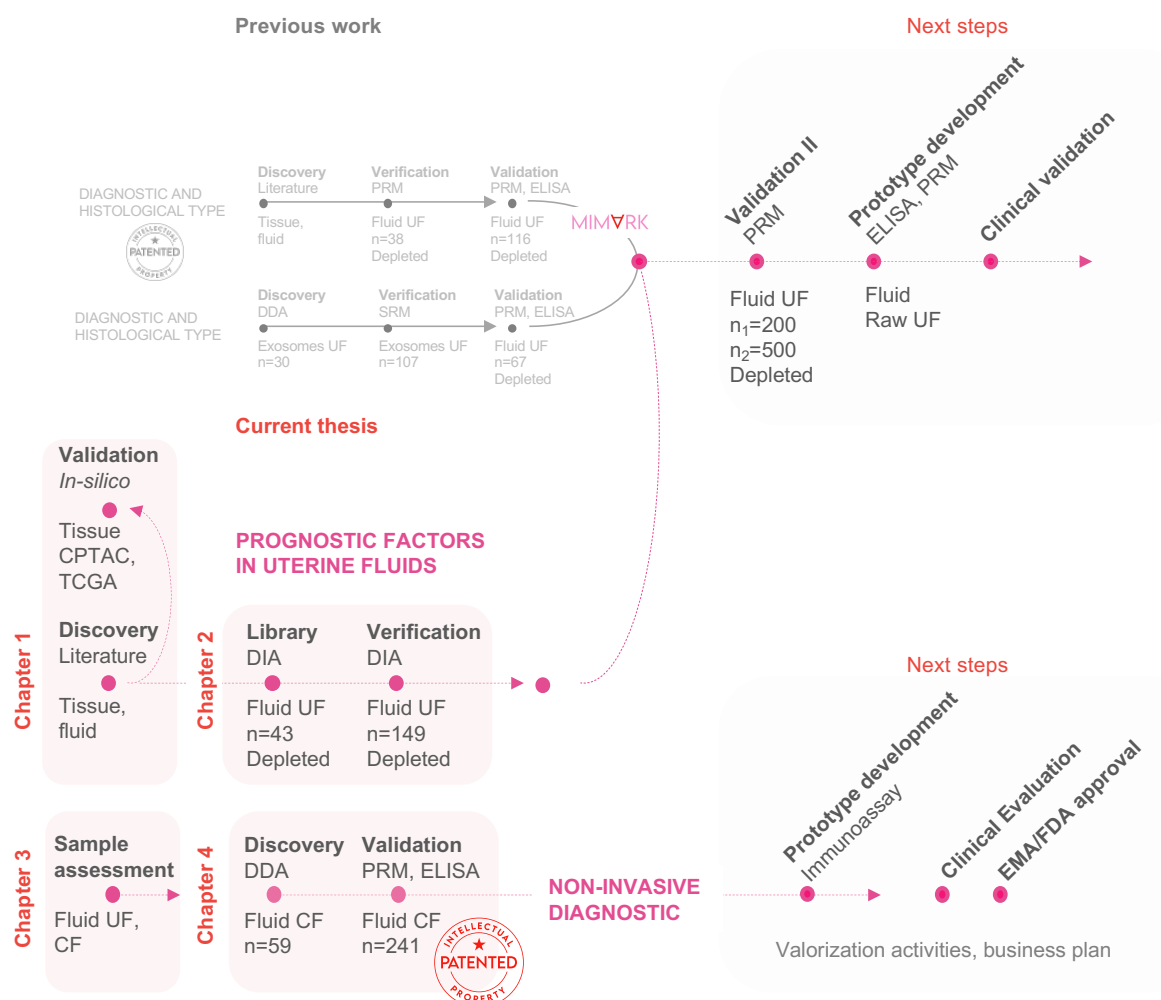
Figure 41 General overview of the biomarker identification process followed in this thesis, previous work in the group and spin-off of current company and next steps. UF, uterine fluid; CF, cervical fluid; DDA, data-dependent acquisition; DIA, data-independent acquisition; PRM: Parallel reaction monitoring, more specifically LC-MS/MS PRM, EMA, European Medicines Agency; FDA, Food and Drug Administration

Prognostic factors and recurrence prediction in uterine fluids

In the first part of this thesis, including Chapter 1 and Chapter 2, we performed an extensive literature review and an *in-silico* validation of known prognostic factors, as well as we screened the proteome of uterine fluids of EC patients for the search of biomarkers of histological grade and type, as well as of recurrence prediction. The final aim of this work was to improve the preoperative risk assessment of EC patients, by providing an objective and accurate set of biomarkers that could be measured in the uterine fluid, which is obtained thanks to a simple centrifugation of pipelle biopsies. The workflow followed is described in Chapter 1 and Chapter 2 on the results section.

In **Chapter 1** an extensive literature review was performed including articles studying protein biomarkers for prognosis (recurrence and/or overall survival) and/or related factor (histological type, tumor grade, FIGO stage, LVSI, LNS, myometrial invasion, cervical invasion, molecular classification) in EC. A resulting list of 255 proteins associated with prognosis and/or related factors in EC was generated. Additionally, we outline some limitations that were common in multiple studies and we proposed new approaches that might accelerate the identification of clinically useful proteins for EC diagnosis, including the exploitation of the identified proteins in easy-to-access proximal fluids such as uterine fluids, and their validation using data repositories before further validating them in big cohorts of patients, which is normally related to big expenses [57]. Following our rationale, Chapter 1 also include the consequent *in-silico* validation of the 255 protein biomarkers identified through our literature review in two big data repositories of known consortiums (TCGA, and CPTAC). The analysis allowed us to highlight 30 proteins with proven prognostic abilities, as well as describing new potential applications than the ones already described in literature [335]. The reasoning here described, corroborate, and stablish a judicious strategy to follow as an example for others in order to spotlight the most promising biomarkers for further validation phases.

In **Chapter 2** we used an emerging MS-based technology of high precision and accuracy to use the fluid of pipelle biopsies, i.e., the uterine fluid as a source to identify biomarkers and develop biomarker signatures that accurately determine the histological type and grade and predict the risk of recurrence for EC patients. From a technical perspective, the data-independent acquisition (DIA) method used in this chapter is a highly multiplexed and reproducible mode of targeted proteomics that allows the user to accurately quantify the proteome of the sample analyzed by matching the peptides with a spectral library of reference previously generated. Hence, in Chapter 2 we first use the fluid of uterine fluids of 43 EC patients to generate a spectral library of peptides accounting for 54,448 peptides corresponding to 5,863 proteins. Our "uterine fluid" library overlaps 96% with proteins identified in EC tissues from the CPTAC study and overlaps 62% with the 255 proteins identified in Chapter 1. Additionally, the library covered most of the key proteins involved in cancer-related pathways, highlighting the potential of uterine fluids as a minimally invasive source of EC prognostic and predictive biomarkers. From a biological and clinical perspective, we used a cohort of 149 fluids of pipelle biopsies of 149 EC patients, including 69 EEC-G1G2, 35 EEC-G3, 16 NEEC-SEC, and 29 NEEC-Other histologies to analyze their

proteome. We identified 34 proteins differentially expressed between EEC and NEEC tumors; 35 proteins which abundance levels was different between low grade EEC and high-grade EEC; and 9 proteins showing differences in abundance regarding recurrence; all with a*djusted P* value <0.05, *Fold-Change* >|2|, and AUC > 0.7). Moreover, different 2-protein signatures were developed to increase the accuracy of the individual proteins. Four different signatures were identified to specifically determine different histologies in the uterine fluids. Specifically, the combination of PROT and PROT was able to distinguish between EEC and NEEC with a sensitivity of 91% and specificity of 83%. Combining PROT and PROT we achieved 100% sensitivity and 79% specificity to determine EEC from SEC, whereas combining PROT and PROT we were able to distinguish between EEC and other NEEC histologies with a sensitivity of 97% and specificity of 70%. Last combination composed by PROT and PROT was able to determine SEC and other NEEC histologies with 93% sensitivity and specificity. A panel composed by PROT and PROT exhibited a sensitivity of 85% and specificity of 81% to discriminate between low-grade EEC and high-grade EEC. Finally, three protein signatures were defined in order to predict patients that will recur. The combination formed by PROT and PROT achieved a sensitivity of 75% and specificity of 86% to identify recurrent patients. Additionally, the combination of PROT and PROT was able to predict recurrence with a 91% sensitivity and 80% specificity in EEC patients, while the combination composed by PROT and PROT was able to predict all NEEC patient that recurred with a 100% sensitivity and 83% specificity. Interestingly, from the proteins identified in the literature revision Chapter 1, 15 were validated as potential biomarkers also in the uterine fluids, and from those 3 (PROT, PROT, PROT) were also validated in the *in-silico* validation in Chapter 1, evidencing that the stepwise workflow is a model to follow, but also highlighting the differences in detecting proteins from different matrixes (tissue *vs* uterine fluids).

In this Chapter 2 we achieved a more refined classification of EC prognostic factors and prognosis by analyzing proteomic data, acquired by DIA methodology, and comparing over 2,300 proteins. We described the application and demonstrated the suitability of the technique to generate large-scale quantitative proteomic profiles of the uterine fluids from EC patients for tumor classification. We demonstrated the importance of these novel MS approaches in biomarker discovery and verification studies to prioritize the most promising biomarkers to enter a validation study while having the whole overview of the proteome of the samples. This allows to identify and reproducibly quantify potential biomarkers at every step across samples. Thus, the workflow here presented will become reference for other studies of the field and data here generated will provide a valuable resource for future protein biomarker discoveries. Specifically, we provided the scientific community with a library of proteins contained in uterine fluids to be used in other gynecology-related studies. Biologically, the molecular landscape of pre-operatory endometrial biopsies to identify biomarkers that determine prognostic and predictive factors was studied for the first time. We demonstrated the feasibility to use uterine fluids as a source of prognostic and predictive biomarkers for EC. Moreover, we studied the disagreement between preoperative diagnostic in pipelle biopsies in comparison to the postoperative diagnosis and compared to the agreement obtained with our protein signatures measured in uterine fluids. In our

cohort of 149 EC patients, 15% and 35% disagreement were observed in the determination of histological type and grade, while the different protein signatures we described determined histological type with AUC values ranging from 0.894-0.977, and tumor grade with an AUC value of 0.853, complementing and improving current procedures. Finally, we determined that our protein signatures predicting the risk of recurrence (AUC values ranging from 0.823-0.913), were independent of the risk assessment and molecular profiling, and able to predict recurrence also in low-risk patients giving additional and valuable information to help clinicians with the management of EC patients. Hence, our protein signatures allow to give an objective and quantitative diagnosis that would complement current clinical practices and to achieve high precision diagnosis and recurrence prediction for all EC patients.

Non-invasive diagnostic

The second part of this thesis consisted in the identification of protein biomarkers to diagnose EC using the fluid of cervical samples obtained though non-invasive procedures such as cytologies (*i.e.* pap-smears). Currently, there is no screening tool for EC, and women start the EC diagnostic process only when they present AUB or there is high suspicion of EC by the physician during a routine examination. The standard diagnostic process is, in fact, a multistep process, which unfortunately includes minimally invasive to invasive tests, i.e., endometrial biopsies obtained by aspiration or hysteroscopy, respectively. The election of the method is a center- and clinician-based decision. Whilst pipelle biopsies offer a fast and less invasive diagnosis than hysteroscopies, they are associated to a high failure (up to 42%) to obtain appropriate samples for diagnosis, due to impossibility to acquire the sample (cervical stenosis) or insufficient material contained which is not evaluable by the pathologist [133,391]. Consequently, those women will have to undergo an additional invasive procedure such as hysteroscopy. Additionally, due to the few number of cells that the sample might contain, in patients diagnosed with EC, determination of histological type and tumor grade suffer from interobserver variation, which leads to up to 50% of incorrect tumor features assignment [136,155]. In this thesis, we aimed to improve the current diagnosis with the development of a non-invasive tool that could diagnose all patients and accurately determine histological type and grade of EC. To achieve our goal, we conducted the first phases of preclinical research: first, we optimized the methods to collect cervical fluids as the source of non-invasive EC protein biomarkers. second, we performed a discovery study by shotgun proteomics in 59 patients (20 EC, 20 non-EC, and 19 patients suffering with some cervical pathology) where we analyzed the whole proteome of the samples and identified 75 different expressed proteins that could act as EC diagnostic biomarkers. Using targeted proteomics, we measured the levels of those proteins in a cohort of 241 patients (128 EC, 106 non-EC, and 7 hyperplasia). Finally, we developed protein biomarker signatures to diagnose EC. The steps followed are described in Chapter 3 and Chapter 4 of the results of this thesis.

In **Chapter 3**, we characterized and compared the proteome of albumin and IgGs depleted and non-depleted (or raw) uterine fluids, as well as the proteome derived from five cervical fluids of 4 patients (2 non-EC, 2 EC). The cervical fluids were collected using different sampling devices and diluted in

different volumes and/or solution. The sampling devices were all brushes that are used in the clinical routine and that permit to obtain samples from the endocervix to the exocervix. Correlation of protein abundance between raw and depleted uterine fluids showed a high correlation (r=0.97-1), while between cervical fluids the correlation was less pronounced (r=0.81-0.9). Protein abundance between uterine and cervical fluids was lower (r=0.6). Based on total protein concentration and the number of EC diagnostic biomarkers identified, we selected two cervical samples, Rovers Cervex Brush ® and the endocervical swab HC2 DNA collection device Digene, and assessed the detection of 52 EC biomarkers previously identified in albumin and IgGs depleted-uterine fluids[200,203] . Raw uterine fluids and the two selected cervical fluids from a total of 39 patients (18 non-EC, 21 EC) were analyzed for the 52 EC-related biomarkers by LC-PRM. Results highlighted the potential of 36 proteins to be used as biomarkers in the raw uterine fluids. Those proteins reproduced the results obtained in studies previous to this thesis in depleted uterine fluids, and importantly, four of them (PROT, PROT, PROT, and PROT) achieved a great accuracy (AUC > 0.9) to distinguish between EC and control patients. Interestingly, 30 of these proteins also had diagnostic potential when measured in cervical fluids. However, the best performing proteins in uterine fluids were different from the best in cervical fluids, in line with the observed differences in the proteomes of these matrices. Using the Rovers Cervex Brush ®, 15 proteins discriminated EC and non-EC patients with AUC values higher than 0.7, where PROT and PROT were the best performing proteins, whereas when using the endocervical swab HC2 DNA collection device Digene, 28 proteins were found differentially expressed (*Adjusted P* value <0.05, AUC > 0.7), being PROT, PROT, and PROT the most accurate diagnostic proteins. Finally, we purposed Rovers Cervex Brush ® as the best brush and methodology to move to the next discovery phase to identify protein biomarkers for diagnosing EC, as it is already introduced in the gynecologic office worldwide and it could mean a higher and faster acceptance of a future developed EC diagnostic test by doctors and patients.

In **Chapter 4** we went through the different phases and methodologies of the biomarker pipeline to identify and verify protein signatures to rule out EC patients and diagnose the histological type and grade of EC patients using non-invasive cervical fluids. Starting from a discovery phase, we analyzed the fluid of cervical samples of 59 patients by MS operated in data-dependent mode (DDA). Selected patients included 20 control patients and 20 EC patients, both presenting AUB or high risk of EC due to thickness of the endometrium, and 19 patients not presenting risk for EC but suffering some type of cervical pathology. The inclusion of this last group aimed to control the specificity of our potential biomarkers for EC. The proteome of the cervical fluids consisted of 2,816 proteins. To identify potential biomarkers, we followed three different strategies: (i) absence/presence study; (ii) statistical analysis of proteins detected in >75% of patients per each group; (iii) proteins described in literature as potential EC diagnostic biomarkers showing promising results in the results obtained in Chapter 3. Pursuing these strategies, we identified 262 proteins as significant to distinguish EC from non-EC patients (115, 131, and 16, respectively). Based on statistical results, familial grouping, and literature, we finally chose 75 proteins to be further studied in the subsequent verification phase. The verification phase consisted of

analyzing those selected proteins by LC-MS/MS PRM in the cervical fluids of 241 patients (106 non-EC, 128 EC, 7 hyperplasia). Results highlighted 26 proteins presenting higher levels of abundance in EC patients compared to non-EC patients (*Adjusted P* value < 0.05 and *Fold Change* > |2|, AUC > 0.70), most of them also showing accuracy to diagnose EC at an early stage of the disease. Remarkably, PROT, PROT, PROT, PROT, PROT, PROT, and PROT, showed great discrimination power (AUC > 0.80). The combination of PROT, PROT and PROT, exhibited a sensitivity and specificity of 95.4% and 85.7%, respectively, to rule out EC patients and AUC value of 0.954. Importantly, in this dataset the histopathologic examination of the biopsies by aspiration, which are the first line and preferred method, failed in 29% of cases, either due to technical issues (*i.e.,* stenosis) or insufficient material. In this dataset, our molecular tool was able to correctly diagnose 91% of cases in non-invasive cervical fluids. Next, we also verified the best performing protein biomarker, PROT, in ELISA and demonstrated how we are able to achieve same accuracy using an antibody-based assay in comparison to MS. Moreover, the levels of 4 proteins (PROT, PROT, PROT, PROT) presented differences between EEC and NEEC histological types and 5 proteins (PROT, PROT, PROT, PROT, PROT) showed significant differences when comparing low *vs* high grade tumors for the endometrioid histology. Finally, we described two different protein panels that are able to discriminate histological types and grade. The combination of PROT, PROT and PROT was able to distinguish EEC from NEEC with AUC value of 0.911, with sensitivity of 80% and specificity of 91.3%. The second combination formed by PROT, PROT, and PROT yielded an AUC of 0.975 with sensitivity of 91.7% and specificity of 100% to determine EEC grade 1 *vs* EEC grade 3.

Remarks of this thesis

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

The **use of proteins rather than DNA or RNA** has several benefits. The Central Dogma of Molecular Biology clearly states the relationship between the genetic information within the biological system, however, still nowadays research in the field is being conducted as many questions remain unsolved. Proof of that is that despite all genes could be sequenced resulting in 19,670 human protein-coding genes described in the latest version of The Human Protein Atlas, *Uhlén et al.,* defined approximately 11% of the human proteins as missing proteins, as they lack of experimental data on the protein level [393]. Hence, <u>proteins are more diverse than DNA and RNA, more dynamic, and they are the functional players driving both normal and disease physiology</u>. In order to solve the significant gap in the proteome knowledge and to characterize different tumor proteomes, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) of the National Cancer Institute provided with public and available proteogenomic data for several cancer types, including EC [298]. Proteins seem the ideal and logical molecule to analyze for their biological significance and multiple technical and economic advantages. Proteins are more stable than DNA and RNA, thus providing <u>easy sample preparation and storage conditions</u>. Importantly, protein biomarkers <u>can be assessed by easy and low-cost methods, such as antibody immunoassay</u>

(ELISA-like), which are widely available in hospitals. Hence, the use of proteins is of key importance to transfer the results described in this thesis to the clinical practice.

The **two proteomic workflows** described in this thesis, both based on different MS approaches, represent a significant improvement over the current state-of-the-art of biomarker research. We followed two novel translational pipeline strategies to discover and validate protein biomarkers in two different proximal fluids.

On the first approach, we performed DIA technology for the identification of biomarkers, and this is an emerging highly-reproducible MS technology developed by Gilled *et al.* [245], that by July 2021 only accounted for 1% of the data annotated in public datasets compared to the 78% of the widely used label free (DDA) [394]. First, using the fluid of pipelle biopsies, we generated a spectral library representing their proteome. This is an important achievement because it impacts on the study outcome, as a peptide requires to be present in the spectral library of reference to be identified and quantified by DIA [395]. Despite many studies use public libraries, *de novo* generation was essential to success in the study because uterine fluids demonstrated to be complex samples [396], as they are a combination of tissue and blood, together with secreted molecules from cells of the uterus. Secondly, we used this library to perform a proteotyping approach of the fluid of pipelle biopsies to classify tumors according to histological type and grade and to predict recurrence, being the first study of its class in EC.

On the second approach, we faced the challenge to study cervical fluids from a proteomic perspective. We first assessed 5 different sampling methods, to select the most suitable one to obtain high quantity and quality protein material from cervical fluids. Two methods, the Rovers Cervex Brush ® and the endocervical swab HC2 DNA collection device Digene, were selected and used to assess the EC diagnostic biomarkers in a verification study in an independent cohort of patients using targeted MS approach, particularly LC-PRM. In addition to identify the most suitable sampling method, the inclusion of the untargeted approach allowed us to describe the cervical fluid proteome and discover novel potential candidates not previously described in literature as EC diagnostic biomarkers and that could be useful as biomarkers in Cervical fluids. Subsequently, targeted MS in the biomarker pipeline allowed the measurement of a large number of biomarkers at once and assess all possible combinations among them, allowing a more efficient identification of clinically useful biomarker signatures.

In fact, protein signatures have shown to provide higher discrimination power than individual biomarkers, which is comprehensive as cancer, and particularly EC, is a heterogeneous disease. In our study, we showed how combining two or three proteins we could improve the diagnostic and/or prognostic accuracy. Up to date, only three multiparametric proteomic tests have received the FDA approval: the Risk of Ovarian Malignancy (ROMA) algorithm for the prediction of ovarian cancer risk in women with pelvic mass, the Overa test (update of OVA1™) for the prediction of ovarian cancer risk in women with adnexal mass, and the DCP and AFP-L3 for the risk assessment for development of hepatocellular carcinoma [397]. Similar to our findings, ROMA is composed by HE4 and CA125 together

with the assessment of menopausal status, and its AUC value to predict ovarian cancer risk is 0.91 (0.94 for post-menopausal women), whilst the AUC values for CA125 and HE4 are 0.86 and 0.91, respectively [398]. In ROMA algorithm, they combine clinical features such as menopausal status with the measurement of HE4 and CA125, which are protein biomarkers routinary measured in blood by chemiluminescence immunoassay (CLIA) or electrochemiluminescence immunoassay (ECLIA).

In the path towards the clinical use of our findings, we finally transferred our best performing diagnostic protein to immunoassay (*i.e.,* ELISA assay). Unfortunately, we could not transfer other proteins of the diagnostic signature due to unavailability of antibodies. Despite ELISA assays remains being the gold-standard for clinical laboratories, there is a gap between high-throughput techniques used in research and the antibody-based techniques applied in clinics [183,399], mainly due to the lack of antibodies such as in our case, but also to the lack of reliable and reproducible antibodies and immunoassays [400]. In 2008, Berglund *et al.* tested over 6,100 commercial antibodies obtaining a success rate of 49% [401], demonstrating antibody quality issues, including poor correlation between antibody lots and low reproducibility between assays [402]. Irreproducible antibodies and reagents are a major challenge for the research and industry. To solve this, Uhlen *et al.* purposed five strategies (genetic, orthogonal, independent antibody, tagged protein expression and immunocapture followed by mass spectrometry) as guidelines to ensure antibody reproducibility [403]. These guidelines are used by different projects such as The Human Protein Atlas or antibodypedia to validate their antibodies. This approach has permitted to provide useful antibodies for research purposes, however, only a minority were validated to be used in ELISA assays [404]. Moreover, the validated ELISAS were assayed with serum or plasma samples, but not in other biofluids such as uterine fluids, or Cervical fluids. Under this scenario, the production of *de novo* highly specific and reproducible antibodies is highly recommended.

The **use of liquid biopsies** for the diagnosis of a disease is the preferred method over invasive tissue samples and therefore, the interest in the hunt of disease-specific biomarkers in liquid biopsies has significantly increased in the last few years [167,405]. Up to date, most of the studies have been executed using blood samples, and great efforts have been made to understand the plasma proteome [180,193,194], and the secretome and associated proteins identified in the blood [406]. However, the research of biomarkers in plasma by sensitive MS-based approaches is associated to limitations due to the complexity of the plasma, as it has dynamic range over 12 orders of magnitude in protein abundance and potential biomarkers might be diluted and masked by more abundant proteins [179]. Additionally, identified potential biomarkers might be unspecific as blood comes into contact with every organ and can be affected by other biological processes in the body. In contrast, **proximal fluids for biomarker research** are promising alternatives as they are in direct contact with the diseased tissue and enriched in proteins directly secreted by the affected cells, but also intracellular cells as diseased cells shed and naturally lysate, favoring the identification of potential biomarkers. In this thesis we worked with two proximal fluids, the uterine and cervical fluids. First, we demonstrated the uterine fluid as a source of prognostic biomarkers for EC. This finding introduces a novel and promising alternative to the

histopathological observation at pre-operatory level, which might permit to combine a subjective diagnostic test that leads to up to 55% of inaccurate determination of the histopathological type and grade, with an objective and quantitative diagnostic test that is highly accurate. The attraction of uterine fluids is not only that they are currently used in the clinical practice. Remarkably, they are in direct contact with the tumor and therefore, the proteins inside are highly representative of the carcinoma, as it was already demonstrated by Perez-Sanchez *et al.* at transcriptomic level [201]. Second, we demonstrated the cervical fluid as a source of both diagnostic and prognostic biomarkers for EC. Our findings open the avenue to develop a non-invasive diagnosis of EC, which is a change in the paradigm on how clinicians manage women with AUB and/or presenting symptoms related to EC. However, further clinical studies should be performed. The study described in this thesis, evaluated the biomarkers in balanced groups of EC and non-EC patients, however, the reality is that only 10% of women presenting AUB will be diagnosed with EC. Thus, a clinical study representing the real clinical populations must be performed in a multicentric and prospective manner including a broad number of patients covering the different pathologies causing AUB. Additionally, further evaluation of patients suffering hyperplasia should be performed. Moreover, in further studies, the molecular tool based on non-invasive cervical fluids could be assessed as a potential screening tool for EC in patients with high-risk of EC and/or over 50 years. The use of a screening tool in EC would decrease mortality rates as demonstrated with other gynecological screening tools such as mammography to screen breast cancer or cytology and/or human papillomavirus (HPV) test to screen cervical cancer [385].

Finally, we believe that both gynecological fluidic samples, i.e., uterine and cervical fluids, might be a promising source to identify biomarkers for other gynecological diseases that lack of effective diagnostic procedure such as endometriosis or ovarian cancer, among others.

… to the impact in clinics. Where research and market collide.

According to the United Nations Educational, Scientific and Cultural Organization (UNESCO), in 2017, 1.7% of the world gross domestic product was devoted to research and development activities [407], being North America, Western Europe and East Asia and the Pacific the leading regional areas of investment. However, knowledge only impacts in human health when it is translated into products or services. Translational research refers to the conversion of discoveries generated though basic research in the laboratory to clinical practices for the diagnosis and treatment of human diseases [408]. Despite all the efforts, it has been reported that only 5% of the early-phase studies will be translated into tangible improvements for the clinical management [409]. In order to overcome this gap, *Ilan* recently described two main strategies to overcome the innovation paradox. The first, looking at end-users' first and translating the innovation into the right platform, meaning that the essential is to identify the clinical needs and the market needs, and introduces the concepts adoption and penetration. As a second strategy, understanding where your competitors are and creating a differentiator accordingly [410].

The results derived from this thesis are expected to be transferred into clinical practice as an *in vitro* diagnostic (IVD) test. According to The European Parliament and of the Council, an IVD medical device is referred to "any medical device which is a reagent, reagent product, calibrator, control material, kit, instrument, apparatus, piece of equipment, software or system, whether used alone or in combination, intended by the manufacturer to be used *in vitro* for the examination of specimens, including blood and tissue donations, derived from the human body, solely or principally for the purpose of providing information concerning a physiological or pathological state, a congenital physical or mental impairments, predisposition to a medical condition or disease, or to determine the safety and compatibility with potential recipients, or to predict treatment response or reactions or to define or monitoring therapeutic measures" [411].

IVD testing is an indispensable tool in the clinical environment. IVD is the largest sector of the medical device industry worldwide followed by cardiology and diagnostic imaging. It was expected to be one of the fastest growing with annual sales of USD 79.6 billion and 13.4% share during the period 2018-2024 [412]. However, the importance of the IVD tests during the COVID-19 pandemic, made the market grew exponentially and its market size is projected to reach USD 96 billion by 2025 from USD 84.5 billion in 2020 [413]. Particularly in EU, the WHO states that are over 40,000 products available today for IVD testing, covering a wide range of conditions, including laboratory tests and point-of-care tests, giving information regarding diagnosis, prognosis, prediction of a treatment response, risk of developing a disease, etc. The importance of IVD is undoubtable since it demonstrated to improve patients health, reduce costs to the healthcare systems and promote the economic growth of the society [414]

In line with our aim to transfer results to the clinical setting, and specifically, to develop IVD test to provide solutions for unmet clinical issues in gynecology, the principal researchers of the Biomedical Research Group in Gynecology, Dr. Eva Colas and Dr. Antonio Gil, together with a Dr. Marina Rigau as a CEO, founded MIMARK Diagnostics, S.L. (MiMARK) in January 2021. The company licensed the know-how and two patents of the research group to develop their first product, WomEC. WomEC is an IVD indicated to rule out EC in women who presented AUB and to determine the histological type in women with EC by analyzing a protein panel of 5 biomarkers in the uterine fluid.

The results derived in this thesis improve our past results, which are the ones licensed to MiMARK, as:

i)   we identified accurate protein signatures to improve the diagnosis of histological type, histological grade, and predict recurrence at the time of diagnosis. Those biomarkers will also be measured in uterine fluids, and consequently, can be added to WomEC in order to complement and strength the current IVD in development.

ii)   we identified and verified accurate protein signatures in cervical fluids to diagnose EC and to determine the histological type and grade of EC patients. Since those biomarkers can be detected in a non-invasive sample, the development of an IVD test might improve and/or be combined with WomEC. Development to strength the EC diagnosis arena.

Consequently, those results are expected to be of interest of MiMARK and may be licensed to the spin-off in order to foster the IVD development. In the process to develop the IVD test, several steps are required regarding scientific and valorization activities to understand the requirements of multiple stakeholders, including industry, clinicians, patients, regulatory agencies, etc. Hence, to penetrate in the market and successful acceptance and commercialization, **a novel biomarker IVD test** must be able to solve an unmet clinical need, be widely accepted by (end) users and being implemented in a widely available platform following all the regulations indicated by that country. Success of a biomarker IVD test vary between cases and circumstances and there are different optimal ways to proceed, however, an overview and deeply understand of the different pillars is required. The four main pillars to work on are [415]:

CLINICAL PILLAR: Clinical need and patient pathway

Validation of your idea by different stakeholders is essential, and among those, gynecologists, and clinical laboratories are of special relevance. In parallel to the scientific work, during this thesis we interviewed a total number of 166 gynecologists and pathologists from different countries to:

  i)   validate the idea and the clinical need;
  ii) understand the patient journey of different countries

This permitted us to understand where the purposed IVD tests would have an optimal and impacting use, and to work accordingly to the main stakeholder's needs and requirements.

According to our results and literature review, we understood that in EC, there is currently no screening test to determine the presence of the disease in the women's population, including women at high risk of having EC, except for women with Lynch Syndrome. Lynch Syndrome are referred to an annual surveillance of the endometrium by TVUS and annual or biennial biopsy starting from the age of 35 years and until they receive prophylactic surgery preferably before the age of 40 years [64]. EC will develop in 80% of cases in the peri- and post-menospausal period of a women, being 63 years the average age [8], and diagnosis initiates mainly due to the presence of specific symptoms. Diagnosis of EC is generally achieved after a considerable broad range of clinical consultations. The range of clinical consultations can vary from 1 to 5 depending on the country, center of attention, resources, physician, biological features and physiological state of the patient, the condition of the sample, and experience of the pathologist. In depth study of the clinical pathway shows that there are too many variables causing gaps where diagnosis can be improved in both the ambulatory and hospital settings, and there is a clear need for a standardized test to rule out EC, which will reduce the burden of overdiagnosis for women with AUB, and consequently, improve the efficacy and sustainability of our healthcare systems. Moreover, gynecologists claimed a tool that can accurately and objectively diagnose tumor features, including histological type and grade, tumor stage, and molecular classification.

The results derived from this thesis aimed to cover the significantly clinical need that was raised by gynecologists and pathologists. From one side, Chapter 1 and Chapter 2 provided protein signatures able to accurately determine histological type and tumoral grade, as well as providing prognostic information at the time of diagnosis. Importantly, the application of these results in the clinics are expected to be easily incorporated since proteins are analyzed in the same pipelle biopsy that is currently obtained in the clinical procedure. On the other side, we demonstrated the use of the fluid of non-invasive cervical samples as a tool to diagnose EC and provide information regarding the histological subtype and grade. This is expected to change the paradigm of the EC diagnostic process and dramatically reduce the number of consultations for women with AUB. In addition, this is expected to increase the percentage of early detection of EC, and impact on patient's quality of life and survival. Its adoption by the physicians should be successful, as we are working with a standard cytological sampling method, which is similar to the standardized HPV testing. Additionally, the non-invasive biomarkers might serve as a screening tool for EC, which is a field to explore beyond this thesis by including not only women under suspicion of EC but patients at high risk of EC and/or normal population that do not present any symptomatology.

In Figure 42, a schematic summary and representation of the current clinical practice and patient pathway, clinical needs, and the solution purposed in this thesis is shown.
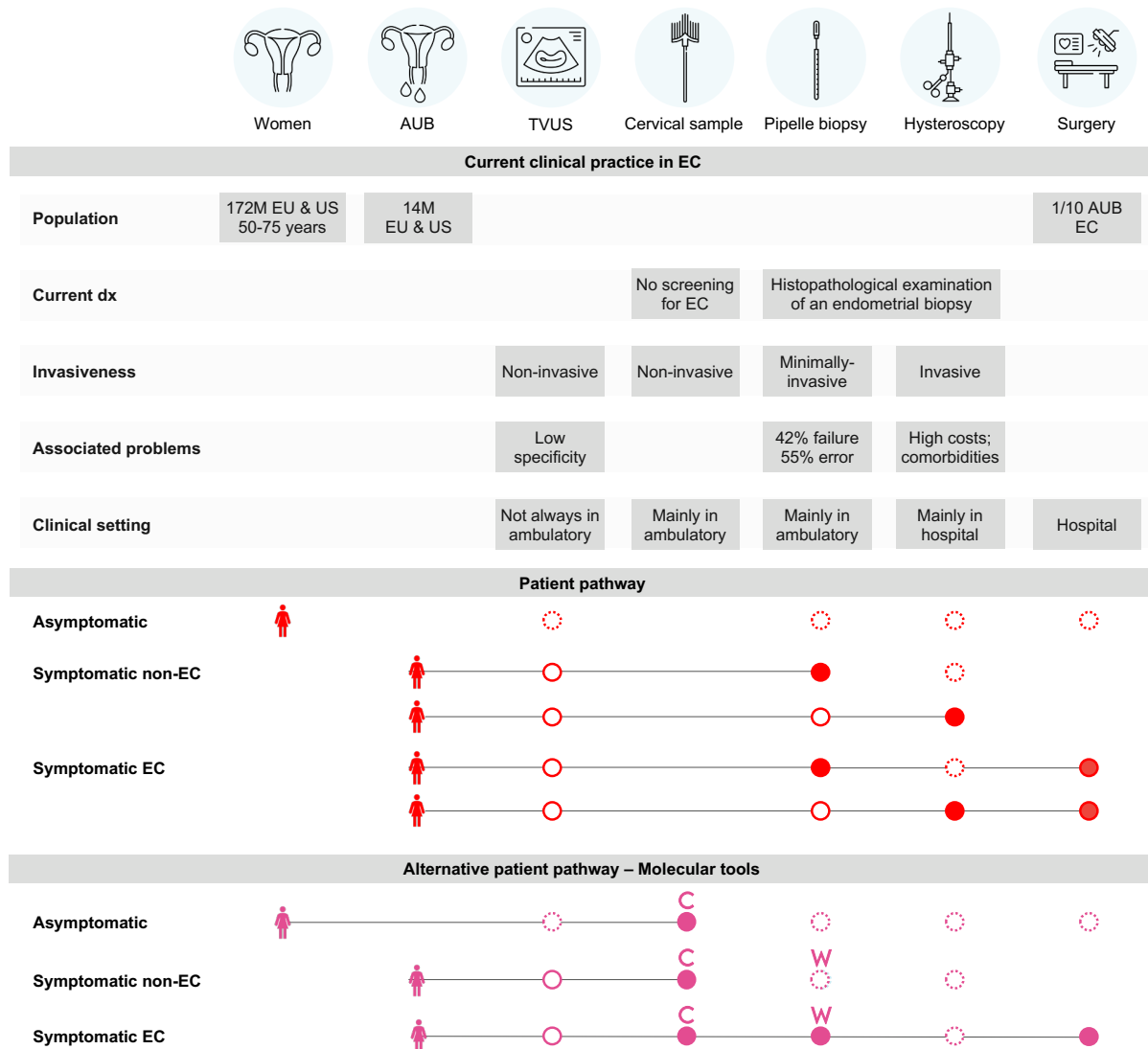
Figure 42 Clinical overview of the current EC diagnostic process. In red the current clinical scenario is represented. In pink a hypothetical clinical scenario is displayed whereby the molecular tools described in this thesis implemented. Colored dots represent that the test is performed giving co... with continuous shape means that the test is performed but do not give conclu... with discontinuous shape means that the test could be performed in specific cas... but it is not necessary for diagnosis. AUB: abnormal uterine bleeding; TVUS,... endometrial cancer; HT, histological type; HG, histological grade; C, cervical san... WomEC product in development by MiMARK diagnostics S.L., and the appli... recurrence; EEC, endometrioid EC; NEEC, non-endometrioid EC; SEC, serous E...

Additional considerations related to the clinical impact of our results are remarked:

- The two sampling methodologies used in this thesis do not require any specialized machinery or anesthesia to be acquired. Therefore, they can both be obtained easily at ambulatory level, which optimizes the speed, time, and resources.

- The studies have been performed retrospectively and trying to balance the groups included in each study. However, this does not represent the current reality in clinics, and further

prospective validation with the real prevalence of the disease and women attending to the clinical consultation is required, as well as estimation of the consequent PPV and NPV.

SCIENTIFIC PILLAR: From prototype to clinical validation

The ease of adoption of a biomarker signature by clinical laboratories is determinant for its clinical application. Therefore, one of the biggest challenges is to design a simple, fast, cheap and automatable **assay** to measure the described protein signatures. To achieve it, different phases and definition of different aspects are necessary, starting from the prototype development in the desired platform and the further analytical and clinical validation steps.

A prototype is a first, typical or preliminary model from which other forms are developed or copied. Hence, to transfer our findings into the clinical practice, the **development of a prototype** is crucial. To do that, available options should be considered in order to find the most suitable approach. Regarding protein measurement the different platforms that can be used are mainly immunoassays such as ELISA assay, multiplex ELISA, turbidimetry or nephelometry or electrochemical immunosensors, but also MS-based methodologies are becoming widespread.

Mass spectrometry (MS) is the most straightforward way to apply the protein signatures described in this thesis. MS has a promising future in clinical laboratories. Recently, Vogeser *et al.* studied the strategic landscape surrounding the implementation of MS in clinical laboratory within a SWOT analysis [416]. Among the numerous strengths of the technology, he highlighted the multiplexing capability and the sensitivity and specificity achieved. MS would allow the measurement of several panels, such as the ones described in this thesis, in just one assay with the minimum waste of sample and time. Moreover, they allow to perform matrix-independent analysis, suffering from less interferences than the ones reported in immunoassays, and they can also detect different isoforms. Importantly, MS would accelerate the direct application of the increasing number of biomarkers currently discovered and validated in research [389], without the need to develop specific and expensive immunoassays.

However, the main disadvantage of MS is that still, MS is not a broadly stablished technology in clinical laboratories for protein measurements. However, MS-based technology has stablished itself in clinical laboratories for many decades for different applications related to measurement of small molecules such as immunosuppressants [417], toxicological assessments and measurements of medication adherence and therapeutic response [418]; for the quantification of steroid and hormone analysis [419,420], vitamins [421], and measurements of dried blood spots for newborn screenings [422], among others.

In order to advance the application of clinical MS for protein measurements, there have been rapid advances in MS-based proteomics over the last years, especially with the IVD industry shaping the market by providing assay kits, certified instruments, and the first laboratory automated LC-MS/MS instruments as an analytical core [423]. Among the few protein assays that are currently on the market, we found the test of Insulin (Quest Diagnostics) and of PTHrP (Mayo, Quest, ARUP Labs) [424] for individual

proteins, and two multiplexed tests: Xpresys® Lung, which is a blood test for assessing the cancer risk of lung nodules assessing the relative expression of eleven proteins (5 diagnostic, 6 used for signal normalization) [425,426]; and PreTRM ®, which is also a blood test reporting individualized risk of spontaneous preterm birth in asymptomatic women during pregnancy assessing the levels of two proteins [427,428]. Kearney *et al.* studied both tests and described the shared strategies that contributed to their successful translation into clinical use and that could serve in our case as an example: i) Systems biology approach to selecting the initial candidates; ii) adherence to National Academy of Medicine guidelines for best practices; iii) assessing analytical performance early in development; iv) avoidance of 'loss in translation' across technology platforms; v) efficient assay design; vi) normalization techniques to address analytic and pre-analytic variation.

Whilst the path of clinical MS is being paved, immunoassays continue to be the preferred and most spread assay in clinical laboratories.

ELISA assays are the gold standard technique in clinical laboratories. Therefore, it is the most feasible approach to apply our findings in the clinics. ELISAs present high sensitivity and sample throughput, they are fast and straightforward assays once the antibodies are available, as they don't require any sample preparation, and they allow absolute quantification of the proteins of interest. Additionally, immunoassays are easy-to-use and cheap. Moreover, they are also widely accepted in regulated environments, and several diagnostic protein assays have been already approved by the FDA or CE-marketed for use in the EU.

However, the development of multiplexed immunoassays requires of antigen and antibody development, assay configuration and performance, and extensive validation to ensure assay precision and accuracy, which is time consuming and expensive [429]. Moreover, immunoassays also face different limitations that must be considered: i) sensitivity, as any change could affect the protein-antibody binding. Selectivity can be affected by autoantibodies, human anti-reagent antibodies, matrix effects or cross-reactivity between antibodies; ii) the detection of the analyte is limited, as many antibodies cannot detect differences in antigens such as protein isoforms or altered post-translational modifications; iii) immunoassay requires 100-200 μl per sample, which could be a problem in scarce samples such as uterine fluids; and iv) cost-per-sample can be quite high. Despite the equipment costs are low with immunoassays, the reagent costs are quite high, accounting for the biggest market. Additionally, the cost is dependent of the platform used, the number of analytes measured, or the antibody royalties. For a single analyte ELISA, the average range cost is $4-5, with up to 80% of the cost coming from the antibody; v) immunoassay is a multi-step process that might compromise reproducibility. Intra and inter assay variability between laboratories is dependent on the platform typically ranging from 5-15% for intra assay and up to 15-30% for inter assay; and vi) the time to run immunoassays is long taking 2-3 hours only considering the antibody-antigen binding and washing steps [289].

Finally, once the prototype is optimized for the biomarker signatures, it will require of **analytical validation** defined as the ability of a test to conform to predefined technical specifications mainly dictated by the clinical use. Analytical validation will involve <u>performance</u> characteristics for practicability (such as specimen type and its availability and invasiveness on their obtention should be studied together with the stability of the biomarker in the selected specimen type), and <u>reliability</u> characteristics aiming to demonstrate the suitability of the assay for its intended purpose [430,431] (e.g., the definition of limits of detection of the biomarkers and analytical sensitivity (<10-20%), the assessment of potential hook-effect with high analyte concentrations, recovery in clinical specimens, precision of measurement and repeatability, variation within- and between-subjects, analytical specificity and cross-reactivity, matrix effects, among others). Analytical validation will be followed by a **clinical validation** in a rigorously standardized prospective multicenter clinical trial. The clinical validation will require to comply with the regulatory requirements to obtain the CE mark.

REGULATORY PILLAR

Understanding regulatory requirements of each country is a cornerstone in the IVD test transference, as it is currently one of the major challenges. The decision to introduce a new IVD test is influenced by the reimbursement policy and regulated by the regulation agencies and each country's healthcare systems. Regulators and payers are indicating the marginal improvements and less likely to be reimbursed in the future. Also in the US, inadequate reimbursement is a major factor restraining the growth of the IVD market.

Regulatory and legal requirements applied to IVD in the EU and US are regulated by the EMA and FDA, respectively, and are becoming more stringent. Those agencies are responsible for the evaluation of the quality, safety and efficacy of marketing authorization purposes.

IVD products are defined and regulated separately from other medical devices. <u>In EU</u>, the new regulation Regulation (EU) 2017/746 [411] officially entered into force on May 26, 2017, with a date of full implementation on May 26, 2022, and adds stringent requirements for the CE-marking of new and already existing IVD medical devices. Under the new Regulation, IVD medical devices must comply with the Annex I "General Safety and Performance Requirements (GSPRs)", and those being Class A sterile, Class B, Class C and Class D (being Class A lowest risk and class D highest risk) shall undergo a conformity assessment by a notified body to demonstrate that they meet the applicable essential technical requirements, thus making sure they are safe and perform as intended. By the fact of complying with Regulation (EU) 2017/746 requirements, any IVD medical devices company must have implemented a quality management system (QMS). Particularly, ISO 13485:2016 stands as the key harmonized standard for the establishment of the QMS applicable specifically to medical devices and IVD tests.

Once the IVD device has passed the conformity assessment, products are placed a CE (Conformité Européenne) - mark to be legally marketed in the EU. Moreover, additional requirements to CE- mark

may be accomplished to market an IVD medical device. In Spain, a previous license must be obtained after CE- marking to put in the market a new IVD.

<u>In the US</u>, IVD products are regulated under title 21 CFR, which is a critical regulation for medical devices. Thus, it sets the requirements for FDA approval of IVDs. Regarding the title 21 CFR, a QMS must be in place pursuant to 21 CFR Part 820 (Quality System Regulation), though the FDA does not certify as such the QMS of the companies.

IVD products are also classified in US in Class I, II, or III, according to their risk, being Class III a high-risk product. According to this classification, the FDA have two primary premarket pathways for IVD medical devices authorization:

- The Pre-Market Approval Application (PMA) under 21 CFR Part 814 is the most stringed and applied to Class III IVD tests. In this, the FDA will conduct audits of manufacturer and critical suppliers prior to approval.
- The pathway for Class I and Class II IVD tests requires to submit a Pre-market Submission Application (510(k)) pursuant to 21 CFR Part 807 Subpart E. In this case, the FDA will not conduct compliance audits for Class I or II IVD medical device manufacturers prior to clearance, but after that, they may conduct random audits.
- When neither the class nor the pathway are clear, a *De novo* request process may be followed as to clarify the regulatory pathway (510(k) or PMA) that better fits the IVD product.

The approval or clearance by one of these pathways will be granted after demonstrating safety and effectiveness though analytical and clinical validation determining the test's suitability [432].

A particular trait of the US regulation systems are the LDTs, designated and used in single laboratories and often referred as *in-house* tests. These tests are not reviewed by the FDA to ensure their accuracy and reliability, and their exact number is unknown. LDTs are developed in research entities, private laboratories, or big companies, always in a single entity, and usually they are similar to the ones approved or cleared by FDA and sold as prepackaged kits. LDT's reporting to FDA is voluntary. However, the FDA had to intervene in several cases to ensure patient safety. Importantly, LDTs are not necessarily less accurate or reliable than their FDA-reviewed counterpart. Nevertheless, if an LDT is not accurate or safe enough, it may take substantial time before problems are identified and corrected, which can impact on the patients' health. LDTs are principally conducted though lab certification processes overseen by the Centers for Medicare & Medicaid Services (CMS), which focuses on assessing a test's clinical utility. Differently, all laboratories performing testing on human specimens are subject to regulation under the Clinical Laboratory Improvement Amendments of 1988 (CLIA), which governs the accreditation, inspection, and certification of all clinical laboratories establishing an additional set of quality standards and analytical validation. CLIA audit cannot be extrapolated to other sites or patient populations [433].

BUSINESS PILLAR

In order to create value out of the acquired knowledge, the development of a business model is essential. A business model is an outline of how a business or organization delivers value to its customers. It describes how an organization creates, delivers, and captures value in economic, social, cultural, or other contexts, including information such as the purpose, business process, target customers, offerings, strategies, infrastructure, organizational structures, sourcing, trading practices, and operational processes and polices. Thus, a business model shapes the way a company generates cash and plan its future expansions. In this thesis, we worked on two aspects related to the business model, particularly, related to commercialization and intellectual property.

Commercialization Aspects

Market analysis and market potential: target population. Understanding the market need is another key milestone to implement the IVD test in the clinical practice. Considering EC diagnostic process, our target population are women attending at the gynecological consultancy due to AUB and/or attending and being considered under suspicion of EC. In order to estimate the number of this women, we searched the population of women in the US and EU in 2020, which is estimated in 554 million women; and assume that 8% of the women between 50-75 years will suffer AUB yearly according to literature review [100,434]. As a result, we estimated that every year around 14 million women will suffer AUB in the EU and US and will initiate the diagnostic process for ruling out EC from other benign diseases [46] (Figure 42).

Market analysis and market potential: technological platform. The global proteomics market is projected to grow for the next years, reaching $55.9 billion by 2026 compared to the $25.9 billion in 2021 (16.6% CAGR during the forecast period). Owning pharmaceutical companies the largest market share in the market. [435]. Immunoassays accounted for the largest market share of the market in 2020 at $21.4 billion with a CAGR of 6.7% though 2028 [436], while clinical MS is expected to progress at CAGR of over 6% during the period 2021-2025 reaching $1.83 billions. Therefore, both approaches have advantages and disadvantages that must be evaluated. However, an option could be developing the protein panel assays in parallel for both technologies to have a higher impact in the market and clinics.

Competitors. In order to identify potential opportunities and our product's unique value proposition, a competitor analysis is crucial. A competitor analysis is the process of identifying competitors (or potential competitors) in the industry and researching their different marketing strategies to assess their strengths and weaknesses, understand our market, spot industry trends and set benchmarks for future growth.

To study the competitive landscape, we analyzed different companies and research groups focused on the search for biomarkers to diagnose EC. Using Global Data, PubMed and Google search we identified only one marketed product: GynEC®-DX by Reig Jofre S.A; and two products under development:

ADXGYNAE™ from Arquer Diagnostics, and MPap®test from GUZIP. Additionally, other products indicated for the early detection of EC, and therefore, aiming to screen EC were also identified. Similarly, there is only one marketed product in Italy which is PAP-Next™ and commercialized using a central laboratory, and two products under development, which are DoovEEgene from Mc Gill University, and a blood-based test from Manchester University. A detailed and summarized analysis of these tests is shown in Table 20. Finally, other promising studies are being conducted in the same line but they are still at a very early research stage [375,437,438].

Deeply analysis of the mentioned competitors, permit us identify the main parameters in which we should focus on: Firstly, only WomEC [200,293] and ADXGYNAE™ [439,440] are being developed as an antibody-based assay, while all other competitors are based on histopathological examination, or DNA sequencing. Secondly, unlike current pipelle biopsies, WomEC, and GynEC®-DX [199,201,202,441], other competitors use a non-invasive samples such as blood (University of Manchester [339]), pap-smears (PapNext™ [375], MPap®test from GUZIP [376], DovEEgene [442]), or urine (ADXGYNAE™ [439,440]) to diagnose EC. Thirdly, the greatest reported sensitivity in a non-invasive sample was 93% from the PapNext™ test. Fourthly, only WomEC provides with information regarding histological type, but none of them provides with any other tumor features and/or prognostic information.

Test	Description	Type of Sample	Type of test	Specimen Workflow	Time to Dx	Success Rate	Sens. Spec.	Tumor Feature info	Recurrence prediction	Regulation status
Endometrial Cancer Diagnosis										
Pipelle Biopsy (Standard of care)	Histopathological observation and IHQ markers	Pipelle biopsy	Histo-pathology analysis	Standard PB	High > 1 week	68%	81% 98%	No in 42% of cases	No	Standard of care
WomEC (Spain) MIMⱱRK	Test use an algorithm based on 5 protein-panel analyzed by Immunoassay	Uterine Fluid	IVD kit AB-based	Standard PB	Low 24h	88%	99% 79%	Yes*	No	Develop.
GynEC-DX® (Spain)	Test use an algorithm based on 5 gene-panel analyzed by RT-qPCR	Pipelle biopsy	LDT test PCR-based; Company Laboratory	Specific collection Kit	Moderate (3 days)	70%	81% 96%	No	No	Market
Assay 1 of this thesis	Test using an algorithm based on 2-protein panel analyzed by MS	Uterine Fluid	IVD kit AB-based	Standard PB	Low 24h	88%		Yes***	Yes	Develop.
							91-100% 70-93%	Histological type		
							85.3% 80.6%	Histological grade		
							82-91%% 75-83%		Recurrence	
ADXGYNAE (UK)	ELISA Test based on MCM5 protein detection	Urine	IVD kit AB-based	Urine collection cup	Low < 24h	83%	87.8% 75.9%	No	No	Develop.
MPap®test GUZIP (Taiwan)	PCR assay for methylation status of the 3 genes using the genomic DNA	Pap Smear	LDT test Methylation PCR-based; Company Laboratory	Specific collection Kit	High 1 week	ND	91.8% 95.5%	No	No	Develop.
Assay 2 of this thesis	Test using an algorithm based on 3-protein panel analyzed by MS	Cervical fluid	IVD kit AB-based	Standard Cervical sample	Low 24h	100%	95.4% 85.7%	Yes***	No	Develop.
							80.0% 91.3%	Histological type		
							91.7% 100%	Histological grade		
Endometrial Cancer Screening										
Pap-Next (Italy)	PapNext TM test is based on NGS for screening tumor DNA mutations in 30 genes	Pap Smear	LDT test NGS-based; Company Laboratory	Specific collection Kit	High 10 days	ND	93% 99.9%	No	No	Market
DOvEEgene (Canada)	Test based on single cell DNA analysis combined through AI algorithm	Pap Smear	LDT test NGS-based; Company Laboratory	Specific collection Kit	Yes 1 week	ND	70% 100%	No	No	Develop.
Manchester University (UK)	Test based on Attenuated total reflection-Fourier transform infrared (ATR-FTIR) spectroscopy and machine learning algorithms.	Blood	LDT test ATR-FTIR; Company Laboratory	Specific collection Kit	Yes 1 week	ND	87% 78%	No	No	Develop.

Table 20 Competitor landscape. Comparison of the main features of the currently marketed or under development (Develop.) molecular tools to diagnose EC in comparison to the ones described in this thesis. *Assay 1* refers to the results obtained using uterine fluids (Chapter 2), while *Assay 2* refers to the findings from Chapter 4 using non-invasive cervical samples. Tumor information regarding histological type (*), histological grade (**) or both (***). LDT, laboratory developed test; PB, pipelle biopsy; AB, antibody; Sens, sensitivity; Spec, specificity.

We envisioned the use of the results derived from this thesis are two IVD tests that will be intended to rule out EC from other benign diseases in women with AUB and/or under suspicion of having EC and to provide information regarding the tumor features to those women diagnosed with EC. Both will use antibody-based techniques to determine the protein concentrations of a set of proteins, which will be combined through the algorithm behind the protein panel. Particularly, both IVD tests differ in just two features. While the first is based on the measurement of protein biomarkers in the fluid of minimally invasive pipelle biopsies, the second is measured in the fluid of non-invasive cervical samples. And secondly, only the one measured in uterine fluids is able to predict recurrence.

Both solutions described in this thesis have a unique disruptive value proposition and provide competitive advantages to existing marketed products and products in development:

- We use a <u>protein measurement</u>, and thus, it is easily quantified by easy and low-cost methods such as ELISA assay, which are widely available in hospitals and clinical laboratories. Therefore, not requiring of centralized laboratories which is time and cost consuming.

- The use of a <u>quantitative and objective methodology</u> provides robustness and reproducibility in the results obtained in comparison to the reported interobserver variability to determine tumoral features and give information regarding prognosis, which is currently not assessable *per se* in EC.

- The use of <u>non-invasive cervical fluids provides diagnosis to 100% of the patients</u> solving the current diagnostic problems occurring in up to 42% of the patients due to technical problems (*i.e.,* cervical stenosis and/or pain presented by the patient), or insufficient material found in the pipelle biopsies.

- The measurement of the diagnostic signature in cervical samples provides a <u>95.4% sensitivity to diagnose EC</u>, the greatest value among other non-invasive tools (MPap®test, Pap-Next or DOvEEgene using pap-smears, ADXGYNAE using urine, or the blood test from Manchester University).

- The measurement of signatures related to histological type and grade permit the <u>accurate and objective determination of tumor features in cervical fluids</u>, resolving the current interobserver variabilities from the histopathological examination of endometrial biopsies.

- The measurement of different protein signatures in <u>uterine fluids allows to accurately and objectively determine histological type and grade</u> even in those cases when not enough material is found, solving the current diagnostic issues and complementing the histopathological information.

- The measurement of different protein signatures in <u>uterine fluids allows to predict recurrence</u>, independently of the current information available, which is a unique feature among other available or under development tools.

- The use of the fluid of the pipelle biopsies to measure the levels of the proteins contained in our signatures, would <u>not change the current practice </u>of physicians. This together with the fact that the hypothetical test would complement the histopathological observation of the cells would be translated into a better acceptance and absorption of the IVD test.

However, we also identified some weaknesses. The main issue with tool based on the cervical fluid in comparison to the use of the uterine fluids is that its use is currently not recommended in EC guidelines to diagnose the disease. Besides, the test would not be complemented by any histopathological observation, which could impact on less adoption by the physicians. In contrast, the main issue with the tool based in uterine fluids is that it lacks specificity when predicting recurrence. Nonetheless, the tools presented in this thesis represent innovative solutions, providing key additional features not included in the products of our competitors.

Intellectual property rights

Intellectual property rights (IPR), such as patents, copyrights, and trademarks, provide legal rights to a person or organization for their intangible creations. They are important means used by companies to

help in the protection of investments in innovations. To protect research findings, such as the ones from this thesis, two different approaches can be used: keep the results as trade secret or patenting them.

Trade secrets are a type of IPR including all types of information that have inherent economic value. They are generally not known or readily ascertainable by others, and the owner takes measures to keep secret and restrict others from disclosing it. In contrast, a patent is a license conferring a right or title for a set period. In the case of identifying a new biomarker, patent protection should be started when clear clinical indication and sufficient scientific evidence has been generated. Regarding the knowledge acquired from this thesis, we keep as trade secret the results obtained in Chapters 1 and 2, with the intention of patenting them in the near future, while we applied for a patent (EP21382680.3, see Annex 5 for detailed information) to protect the results obtained in Chapters 3 and 4.

Additionally, a "freedom to operate" (FTO) analysis of the main protein forming the diagnostic panel of the non-invasive tool (Chapter 3) was carried to evaluate the possibility of developing and commercializing an IVD test including this protein without infringing third party IPR. Furthermore, in a near future, newly results obtained from Chapter 2 will be evaluated and FTO of decisive proteins will be carried.

Final summary

The studies included in this thesis constitute an important contribution to the field of proteomic biomarker identification in liquid biopsies and, specifically, in EC. From one side, three uterine-based proteomic signatures developed in this thesis (5 prognostic, and 3 predictive) are expected to improve EC management. The signatures here described are completely compatible with the current diagnostic procedure based on the histological examination of the malignant cells contained in the pipelle biopsies at pre-operatory level. This is the first study that uses the molecular landscape of pipelle biopsies to provide information regarding the prognostic features of the tumor pre-operatively. Remarkably, one signature also gives valuable information regarding the outcome of the patient, with is currently out of scope in EC management. On the other side, this thesis developed a leading-edge molecular tool using the fluids of non-invasive cervical samples to diagnose EC and give information regarding tumor features. Up to date, any other study has described such a valuable tool that could serve as a screening, but also to replace invasive tests currently used for the diagnosis of EC. Although both studies require further validation, a first step has been achieved. In long term, the implementation of these signatures in different proximal fluids is expected to impact substantially on the improvement of EC management, as well as reducing disease-related healthcare costs.

Apart from the scientific work, other fields briefly described in this discussion are equally important to successfully introduce a product in the market. Consequently, during this thesis, and particularly during the last year after the spin-off of MiMARK diagnostics SL, I have participated in other activities in parallel to the experimental work. Those activities included mentoring and networking activities, activities related

to market research and business plan (interviews with stakeholders, health-economic and regulatory studies of different countries, and evaluation of competitors) (Annex 5). However, tools are still in an early stage. Hence, to transfer the project to the market, public and private funding will be required, as valorization activities together with the prototype development and clinical evaluation steps are expensive. Therefore, a robust financial plan is needed to present the project to investor and potential licensees.

Conclusions

The main conclusions derived from this thesis are as follows:

1. The high number of molecular studies of EC published between 1991 and 2020 allowed us to generate a list of 255 candidate prognostic protein biomarkers, mainly identified in tissue samples, which are related to histological type and grade, FIGO stage, molecular classification, and OS and RFS.

2. Among the 255 candidate biomarkers, 30 proteins were validated *in silico* using the RNA and proteomic data from the TCGA and CPTAC studies, whereby TPX2 was the protein associated with histological type and grade, molecular classification, and OS and RFS.

3. Altogether, the *in-silico* study was highly useful to validate the potential use of biomarkers, as well as to discover their new potential applications.

4. The proteome discovery of uterine fluids allowed us to create a unique spectral library which contained 5,863 proteins associated with the endometrial environment and EC. This library is now on open access for the scientific community to be used for fostering research in gynecology.

5. As a result of the clinical retrospective study conducted in a DIA-MS study on uterine fluids from 149 EC patients, we identified a total number of 59 proteins able to discriminate between histological types and histological grades. Specifically, 34 proteins were significantly different between EEC and NEEC histologies, 37 proteins had significant different levels between low-grade and high-grade EEC, and 9 proteins were significantly different between recurrent and non-recurrent EC patients.

6. We identified four different signatures to accurately diagnose histological type in uterine fluids. PROT and PROT was able to distinguish between EEC and NEEC with an AUC of 0.894. If the NEEC cases are split into SEC and other NECC, we identified PROT and PROT to differentiate EEC from SEC with an AUC of 0.956, and PROT and PROT to distinguish between EEC and other NEEC histologies with an AUC of 0.893. The last combination composed by PROT and PROT was able to determine SEC and other NEEC histologies with an AUC of 0.977.

7. Regarding histological grade, a panel composed by PROT and PROT exhibited an AUC of 0.853to discriminate between low- and high-grade EEC in uterine fluids.

8. The highest accuracy to predict tumor recurrence at the time of diagnosis was achieved with three protein signatures measured in uterine fluids. PROT and PROT predict EC recurrence with an AUC of 0.823 independently of the histological subtype. Specifically in EEC, the combination of PROT and PROT was able to predict recurrence with an AUC of 0.878, reaching a sensitivity of 91%; and in NEEC, the combination of PROT and PROT predicted recurrence with an AUC of 0.913 reaching 100% sensitivity.

9. In order to improve EC diagnosis, we unveiled that cervical fluids can be used as non-invasive sample for the detection of EC protein biomarkers. To collect cervical fluids, the most suitable tools are the Rovers Cervex Brush® or the endocervical swab (HC2 DNA collection device Digene), which was placed in a saline solution after cervical sampling.

10. Following the standard biomarker pipeline, we identified 2,816 protein biomarkers in a discovery study of cervical fluids from 59 patients, including EC, non-EC patients and patients suffering cervical pathologies, using a label-free MS approach. Then, we prioritized 75 protein biomarkers in a verification phase, in which those biomarkers were analyzed in cervical fluids from 241 patients by LC-MS/MS PRM. A total of 30 proteins were verified as EC diagnostic biomarkers. Additionally, 29 and 15 proteins had the potential to diagnose histological type and grade, respectively.

11. The highest accuracy for EC diagnosis was achieved by the combination of PROT, PROT, and PROT, achieving an AUC of 0.954, which relates to a 95% sensitivity and 86% specificity to detect EC cases in cervical fluids.

12. In addition, protein panels were developed for histological type and grade in cervical fluids to improve the pre-operative risk assessment of EC. The panel composed by PROT, PROT and PROT achieved an AUC of 0.911 to discriminate between EEC and NEEC histologies; and the panel composed by PROT, PROT and PROT reached an AUC of 0.975 to discriminate between grade 1 and grade 3 EEC.

13. In an attempt to approach the use of the identified biomarkers in clinical practice, the levels of the highest performing diagnostic biomarker, PROT, was verified by ELISA assay, obtaining an AUC of 0.931. Those results highly correlated with the ones acquired by LC-MS/MS PRM.

14. We demonstrated the current issues related to endometrial biopsies. Specifically, between 14-23% of the endometrial biopsies obtained with pipelle biopsy failed due to technical reasons or insufficient material at diagnosis, and women had to undergo an invasive hysteroscopy. Additionally, histological type from the endometrial biopsy and grade were not correctly determined in the pre-operatory assessment in relation to the final diagnosis in 14-15% and 35-41% of cases, respectively.

15. Finally, the results contained in this thesis brings forward the proteomic research of biomarkers in gynecological fluids to improve diagnosis and risk prediction of EC. These fluidic-based signatures are expected to be the basis for the development of a non-invasive tool to accurately diagnose EC, and/or to be combined with the current pathological examination of pipelle biopsies to improve the pre-operative risk assessment of EC. This will ultimately preclude the need of an invasive EC diagnosis, and will help to guide an optimal surgical treatment for EC patients, which will impact on the quality of life of women undergoing the diagnostic process, and importantly, also in the life expectancy for EC patients

Bibliography

1. Gershenson, D., Lentz, G., Valea, F. & Lobo, R. *Comprehensive Gynecology*. *Elsevier* (Elsevier, 2022). doi:10.1016/C2017-0-04207-6.

2. Gerard J. Tortora, B. H. D. *Prinicples of Anatomy & Physiology*. (Wiley, 2020).

3. Kunz, G., Beil, D., Huppert, P. & Leyendecker, G. Oxytocin – a stimulator of directed sperm transport in humans. *Reprod. Biomed. Online* **14**, 32–39 (2007).

4. Gasner, A. & P A, A. *Physiology, Uterus*. *StatPearls* (StatPearls Publishing, 2020).

5. Monis, C. N. & Tetrokalashvili, M. *Menstrual Cycle Proliferative And Follicular Phase*. *StatPearls* (StatPearls Publishing, 2019).

6. Cabero, L. & Cabrillo, E. *Tratado de Ginecología y Obstetricia. Ginecología y medicina de la reproducción. Editorial médica Panamericana* (2013).

7. Santoro, N. *et al.* Helping midlife women predict the onset of the final menses: SWAN, the Study of Women's Health Across the Nation. *Menopause* **14**, 415–424 (2007).

8. Potter, B., Schrager, S., Dalby, J., Torell, E. & Hampton, A. Menopause. *Prim. Care - Clin. Off. Pract.* **45**, 625–641 (2018).

9. 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).

10. Buckley, H. & Fox, H. *Biopsy Pathology of the Endometrium*. (CRC Press, 2002).

11. Mutter, G. L., Zaino, R. J., Baak, J. P. A., Bentley, R. C. & Robboy, S. J. Benign endometrial hyperplasia sequence and endometrial intraepithelial neoplasia. *International Journal of Gynecological Pathology* vol. 26 103–114 (2007).

12. *Female Genital Tumours*. (2020).

13. MH, A. *et al.* Guideline No. 390-Classification and Management of Endometrial Hyperplasia. *J. Obstet. Gynaecol. Can.* **41**, 1789–1800 (2019).

14. Mills, A. M. & Longacre, T. A. Endometrial hyperplasia. *Semin. Diagn. Pathol.* **27**, 199–214 (2010).

15. SG, V. *et al.* Endometrial polyps. An evidence-based diagnosis and management guide. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **260**, 70–77 (2021).

16. Ferlay, J. *et al.* Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953 (2019).

17. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA. Cancer J. Clin.* **72**, 7–33 (2022).

18. Ferlay J *et al.* Global Cancer Observatory: Cancer Tomorrow. *Lyon, France: International Agency for Research on Cancer.* https://gco.iarc.fr/tomorrow (2020).

19. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, C. K. (eds). SEER Cancer Statistics Review, 1975-2018, National Cancer Institute. *Bethesda, MD*.

20. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer Statistics, 2021. *CA. Cancer J. Clin.* **71**, 7–33 (2021).

21. Jemal, A. *et al.* Annual Report to the Nation on the Status of Cancer, 1975-2014, Featuring Survival. *J. Natl. Cancer Inst.* **109**, (2017).

22. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2018. *CA. Cancer J. Clin.* **68**, 7–30 (2018).

23. SEGO. Guía de Asistencia Práctica. Oncoguía SEGO: Cáncer de Endometrio 2016. *Progresos Obstet. y Ginecol. Rev. Of. la Soc. Española Ginecol. y Obstet.* **60**, (2016).

24. Tempfer, C. B., Hilal, Z., Kern, P., Juhasz-Boess, I. & Rezniczek, G. A. Menopausal hormone therapy and risk of endometrial cancer: A systematic review. *Cancers* vol. 12 1–18 (2020).

25. Pinkerton, J. A. V. *et al.* The 2017 hormone therapy position statement of the North American Menopause Society. *Menopause* vol. 24 728–753 (2017).

26. Ignatov, A. & Ortmann, O. Endocrine risk factors of endometrial cancer: Polycystic ovary syndrome, oral contraceptives, infertility, tamoxifen. *Cancers* vol. 12 1–10 (2020).

27. Avgerinos, K. I., Spyrou, N., Mantzoros, C. S. & Dalamaga, M. Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism: Clinical and Experimental* vol. 92 121–135 (2019).

28. Renehan, A. G., Mackintosh, M. L. & Crosbie, E. J. Obesity and endometrial cancer: Unanswered epidemiological questions. *BJOG: An International Journal of Obstetrics and Gynaecology* vol. 123 175–178 (2016).

29. Lundberg, F. E., Iliadou, A. N., Rodriguez-Wallberg, K., Gemzell-Danielsson, K. & Johansson, A. L. V. The risk of breast and gynecological cancer in women with a diagnosis of infertility: a nationwide population-based study. *Eur. J. Epidemiol.* **34**, 499–507 (2019).

30. Murugappan, G., Li, S., Lathi, R. B., Baker, V. L. & Eisenberg, M. L. Risk of cancer in infertile women: analysis of US claims data. *Hum. Reprod.* **34**, 894–902 (2019).

31. 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. *Human Reproduction Update* vol. 20 748–758 (2014).

32. Saed, L. *et al.* The effect of diabetes on the risk of endometrial Cancer: An updated a systematic review and meta-analysis. *BMC Cancer* vol. 19 (2019).

33. Wang, Y., Zhou, R. & Wang, J. Relationship between hypothyroidism and endometrial cancer. *Aging and Disease* vol. 10 190–196 (2019).

34. Aune, D., Sen, A. & Vatten, L. J. Hypertension and the risk of endometrial cancer: A systematic review and meta-analysis of case-control and cohort studies. *Sci. Rep.* **7**, (2017).

35. Cai, D. *et al.* Insulin-Like Growth Factor 1/Mammalian Target of Rapamycin and AMP-Activated Protein Kinase Signaling Involved in the Effects of Metformin in the Human Endometrial Cancer. *Int. J. Gynecol. Cancer* **26**, 1667–1672 (2016).

36. Hampel, H. *et al.* Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res.* **66**, 7810–7817 (2006).

37. Ryan, N. A. J. *et al.* The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genet. Med.* **21**, 2167–2180 (2019).

38. Meyer, L. A., Broaddus, R. R. & Lu, K. H. Endometrial cancer and lynch syndrome: Clinical and pathologic considerations. *Cancer Control* **16**, 14–22 (2009).

39. Iversen, L., Sivasubramaniam, S., Lee, A. J., Fielding, S. & Hannaford, P. C. Lifetime cancer risk and combined oral contraceptives: the Royal College of General Practitioners' Oral Contraception Study. *Am. J. Obstet. Gynecol.* **216**, 580.e1-580.e9 (2017).

40. Troisi, R. *et al.* The role of pregnancy, perinatal factors and hormones in maternal cancer risk: a review of the evidence. *J. Intern. Med.* **283**, 430–445 (2018).

41. Baron, J. A., Nichols, H. B., Anderson, C. & Safe, S. Cigarette Smoking and Estrogen-Related Cancer. *Cancer Epidemiol. Biomarkers Prev.* cebp.1803.2020 (2021) doi:10.1158/1055-9965.epi-20-1803.

42. Moore, S. C., Gierach, G. L., Schatzkin, A. & Matthews, C. E. Physical activity, sedentary behaviours, and the prevention of endometrial cancer. *British Journal of Cancer* vol. 103 933–938 (2010).

43. Bokhman JV. Two pathogenetic types of endometrial carcinoma. *Gynecol. Oncol.* **15**, 10–17 (1983).

44. Getz, G. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).

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

46. Colombo, N. *et al.* Endometrial cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **24**, vi33–vi38 (2013).

47. 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).

48. Joshi, A., Miller, C., Baker, S. J. & Ellenson, L. H. Activated mutant p110α causes endometrial carcinoma in the setting of biallelic pten deletion. *Am. J. Pathol.* **185**, 1104–1113 (2015).

49. Okuda, T. *et al.* Genetics of Endometrial Cancers. *Obstet. Gynecol. Int.* **2010**, 1–8 (2010).

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

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

52. Sarmadi, S., Izadi-Mood, N., Mansourzadeh, N. & Motevalli, D. Evaluation of her2/neu expression in high-grade endometrial carcinoma and its clinicopathological correlation. *Iran. J. Pathol.* **14**, 322–328 (2019).

53. DeLair, D. F. *et al.* The genetic landscape of endometrial clear cell carcinomas. *J. Pathol.* **243**, 230–241 (2017).

54. Bell, D. W. & Ellenson, L. H. Molecular Genetics of Endometrial Carcinoma. *Annu. Rev. Pathol. Mech. Dis.* **14**, 339–367 (2019).

55. 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.* **457**, 27–34 (2010).

56. Murali, R., Soslow, R. A. & Weigelt, B. Classification of endometrial carcinoma: More than two types. *Lancet Oncol.* **15**, e268–e278 (2014).

57. Coll-de la Rubia, E. *et al.* Prognostic Biomarkers in Endometrial Cancer: A Systematic Review and Meta-Analysis. *J. Clin. Med.* **9**, 1900 (2020).

58. Stelloo, E. *et al.* Improved Risk Assessment by Integrating Molecular and Clinicopathological Factors in Early-stage Endometrial Cancer—Combined Analysis of the PORTEC Cohorts. *Clin. Cancer Res.* **22**, 4215–4224 (2016).

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

60. McConechy, M. K. *et al.* Endometrial carcinomas with POLE exonuclease domain mutations have a favorable prognosis. *Clin. Cancer Res.* **22**, 2865–2873 (2016).

61. León-Castillo, A. *et al.* Interpretation of somatic POLE mutations in endometrial carcinoma. *J. Pathol.* **250**, 323–335 (2020).

62. van Gool, I. C. *et al.* POLE Proofreading Mutations Elicit an Antitumor Immune Response in Endometrial Cancer. *Clin. Cancer Res.* **21**, 3347–3355 (2015).

63. AM, S. *et al.* TP53 Mutational Spectrum in Endometrioid and Serous Endometrial Cancers. *Int. J. Gynecol. Pathol.* **35**, 289–300 (2016).

64. Concin, N. *et al.* ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *Int. J. Gynecol. Cancer* **31**, 12–39 (2021).

65. Cao, W. *et al.* Immunotherapy in endometrial cancer: rationale, practice and perspectives. *Biomark. Res.* **9**, 49 (2021).

66. Rosa-Rosa, J. M. *et al.* Molecular genetic heterogeneity in undifferentiated endometrial carcinomas. *Mod.*

Pathol. **29**, 1390–1398 (2016).

67. Köbel, M. *et al.* Frequent Mismatch Repair Protein Deficiency in Mixed Endometrioid and Clear Cell Carcinoma of the Endometrium. *Int. J. Gynecol. Pathol.* **36**, 555–561 (2017).

68. Liccardo, R., De Rosa, M., Izzo, P. & Duraturo, F. Novel implications in molecular diagnosis of lynch syndrome. *Gastroenterology Research and Practice* vol. 2017 (2017).

69. Campbell, B. B. *et al.* Comprehensive Analysis of Hypermutation in Human Cancer. *Cell* **171**, 1042-1056.e10 (2017).

70. Ott, P. A. *et al.* Safety and Antitumor Activity of Pembrolizumab in Advanced Programmed Death Ligand 1–Positive Endometrial Cancer: Results From the KEYNOTE-028 Study. *J. Clin. Oncol.* **35**, 2535–2541 (2017).

71. Marabelle, A. *et al.* Efficacy of Pembrolizumab in Patients With Noncolorectal High Microsatellite Instability/Mismatch Repair–Deficient Cancer: Results From the Phase II KEYNOTE-158 Study. *J. Clin. Oncol.* **38**, 1–10 (2020).

72. Oaknin, A. *et al.* Clinical Activity and Safety of the Anti–Programmed Death 1 Monoclonal Antibody Dostarlimab for Patients With Recurrent or Advanced Mismatch Repair–Deficient Endometrial Cancer. *JAMA Oncol.* **6**, 1766 (2020).

73. Köbel, M. *et al.* Interpretation of P53 Immunohistochemistry in Endometrial Carcinomas. *Int. J. Gynecol. Pathol.* **38**, S123–S131 (2019).

74. León-Castillo, A. *et al.* Clinicopathological and molecular characterisation of 'multiple-classifier' endometrial carcinomas. *J. Pathol.* **250**, 312–322 (2020).

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

76. 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).

77. Kommoss, S. *et al.* Final validation of the ProMisE molecular classifier for endometrial carcinoma in a large population-based case series. *Ann. Oncol.* **29**, 1180–1188 (2018).

78. Bosse, T. *et al.* Molecular Classification of Grade 3 Endometrioid Endometrial Cancers Identifies Distinct Prognostic Subgroups. *Am. J. Surg. Pathol.* **42**, 561–568 (2018).

79. CM, C. *et al.* An NRG Oncology/GOG study of molecular classification for risk prediction in endometrioid endometrial cancer. *Gynecol. Oncol.* **148**, 174–180 (2018).

80. León-Castillo, A. *et al.* Molecular Classification of the PORTEC-3 Trial for High-Risk Endometrial Cancer: Impact on Prognosis and Benefit From Adjuvant Therapy. *J. Clin. Oncol.* **38**, 3388–3397 (2020).

81. Singh, N. *et al.* p53 immunohistochemistry is an accurate surrogate for TP53 mutational analysis in endometrial carcinoma biopsies. *J. Pathol.* **250**, 336–345 (2020).

82.   Murali, R., Delair, D. F., Bean, S. M., Abu-Rustum, N. R. & Soslow, R. A. Evolving roles of histologic evaluation and molecular/genomic profiling in the management of endometrial cancer. *JNCCN J. Natl. Compr. Cancer Netw.* **16**, 201–209 (2018).

83.   National Comprehensive Cancer Network (NCCN). NCCN Clinical Practice Guideliness in Oncology: Uterine Neoplasms. Version 1.2022 - November 4, 2021. www.nccn.org/patients.

84.   Masood, M. & Singh, N. Endometrial carcinoma: changes to classification (WHO 2020). *Diagnostic Histopathol.* **27**, 493–499 (2021).

85.   Vermij, L., Smit, V., Nout, R. & Bosse, T. Incorporation of molecular characteristics into endometrial cancer management. *Histopathology* **76**, 52–63 (2020).

86.   Murali, R. *et al.* High-grade Endometrial Carcinomas: Morphologic and Immunohistochemical Features, Diagnostic Challenges and Recommendations. *Int. J. Gynecol. Pathol.* **38**, S40–S63 (2019).

87.   Pecorelli, S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* vol. 105 103–104 (2009).

88.   Zaino, R. J., Kurman, R. J., Diana, K. L. & Paul Morrow, C. 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).

89.   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.* **22**, 71–74 (2003).

90.   Guan, H. *et al.* Prognosis and reproducibility of new and existing binary grading systems for endometrial carcinoma comparedto FIGO grading in hysterectomy specimens. *Int. J. Gynecol. Cancer* **21**, 654–660 (2011).

91.   MM, A. *et al.* Risk-scoring models for individualized prediction of overall survival in low-grade and high-grade endometrial cancer. *Gynecol. Oncol.* **133**, 485–493 (2014).

92.   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).

93.   Amant, F., Mirza, M. R., Koskas, M. & Creutzberg, C. L. Cancer of the corpus uteri. *Int. J. Gynecol. Obstet.* **143**, 37–50 (2018).

94.   Abal, M. *et al.* Molecular determinants of invasion in endometrial cancer. *Clin. Transl. Oncol.* **9**, 272–277 (2007).

95.   Dane, C. & Bakir, S. The effect of myometrial invasion on prognostic factors and survival analysis in endometrial carcinoma. *Afr. Health Sci.* **19**, 3235–3241 (2019).

96.   Boothe, D. *et al.* Lymphovascular invasion in endometrial cancer: Prognostic value and implications on adjuvant radiation therapy use. *Am. J. Clin. Oncol. Cancer Clin. Trials* **42**, 549–554 (2019).

97. Winer, I. *et al.* Significance of lymphovascular space invasion in uterine serous carcinoma: What matters more; Extent or presence? *Int. J. Gynecol. Pathol.* **34**, 47–56 (2015).

98. Ayhan, A. *et al.* Prognostic significance of lymphovascular space invasion in low-risk endometrial cancer. *Int. J. Gynecol. Cancer* **29**, 505–512 (2019).

99. Clarke, M. A. *et al.* Association of endometrial cancer risk with postmenopausal bleeding in women a systematic review and meta-analysis. *JAMA Intern. Med.* **178**, 1201–1208 (2018).

100. Moodley, M. & Roberts, C. Clinical pathway for the evaluation of postmenopausal bleeding with an emphasis on endometrial cancer detection. *Journal of Obstetrics and Gynaecology* vol. 24 736–741 (2004).

101. Gredmark, T., Kvint, S., Havel, G. & Mattsson, L.-Å. Histopathological findings in women with postmenopausal bleeding. *BJOG An Int. J. Obstet. Gynaecol.* **102**, 133–136 (1995).

102. SR, G. Modern evaluation of the endometrium. *Obstet. Gynecol.* **116**, 168–176 (2010).

103. A, K. & SR, G. Abnormal Uterine Bleeding. *Obstet. Gynecol. Clin. North Am.* **46**, 595–605 (2019).

104. Baggish, M. S., Valle, R. F. & Guedj, H. *Hysteroscopy : visual perspectives of uterine anatomy, physiology and pathology*. (Wolters Kluwer Health/Lippincott Williams & Wilkins, 2007).

105. D, E., S, G., A, L. & AD, A. No. 385-Indications for Pelvic Examination. *J. Obstet. Gynaecol. Can.* **41**, 1221–1234 (2019).

106. C, Y. *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.* **294**, 311–316 (2016).

107. Van Den Bosch, T. *et al.* Typical ultrasound features of various endometrial pathologies described using International Endometrial Tumor Analysis (IETA) terminology in women with abnormal uterine bleeding. *Ultrasound Obstet. Gynecol.* **57**, 164–172 (2021).

108. Long, B., Clarke, M. A., Morillo, A. D. M., Wentzensen, N. & Bakkum-Gamez, J. N. Ultrasound detection of endometrial cancer in women with postmenopausal bleeding: Systematic review and meta-analysis. *Gynecol. Oncol.* **157**, 624–633 (2020).

109. Alcázar, J. L. *et al.* Risk of endometrial cancer and endometrial hyperplasia with atypia in asymptomatic postmenopausal women with endometrial thickness ≥11 mm: A systematic review and meta-analysis. *J. Clin. Ultrasound* **46**, 565–570 (2018).

110. Morrison, J. *et al.* British Gynaecological Cancer Society (BGCS) Uterine Cancer Guidelines: Recommendations for Practice. Version 2.1. (2021).

111. German Guideline Program in Oncology (German Cancer Society, German Cancer Aid, AWMF): Guideline on the Diagnosis, Treatment, and Follow-up of Patients with Endometrial cancer, short version 1.0, 2018, AWMF Registernummer:032/034-OL.

112.   Associazione Italiana di Oncologia Medica (AIOM). Linee guida: Neoplasie dell'utero: endometrio e cervice. Edizione 2021.

113.   SL, A. *et al.* HE4 and CA125 levels in the preoperative assessment of endometrial cancer patients: a prospective multicenter study (ENDOMET). *Acta Obstet. Gynecol. Scand.* **92**, 1313–1322 (2013).

114.   Brennan, D. J. *et al.* Serum HE4 as a prognostic marker in endometrial cancer — A population based study. *Gynecol. Oncol.* **132**, 159–165 (2014).

115.   Wang, Y. *et al.* Predictive value of serum HE4 and CA125 concentrations for lymphatic metastasis of endometrial cancer. *Int. J. Gynaecol. Obstet.* **136**, 58–63 (2017).

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

117.   Rižner, T. L. Discovery of biomarkers for endometrial cancer: current status and prospects. *Expert Rev. Mol. Diagn.* **16**, 1315–1336 (2016).

118.   Soja, M. *et al.* Analysis of the results of invasive diagnostic procedures in patients referred to gynecologic department due to abnormal uterine bleeding. *Prz. Menopauzalny* **19**, 155–159 (2021).

119.   Yi, Y. *et al.* Cost-effectiveness analysis of biopsy strategies for endometrial cancer diagnosis in women with postmenopausal bleeding: Pipelle sampling curette versus dilatation & curettage. *Gynecol. Oncol.* **150**, 112–118 (2018).

120.   Narice, B. F., Delaney, B. & Dickson, J. M. Endometrial sampling in low-risk patients with abnormal uterine bleeding: a systematic review and meta-synthesis. *BMC Fam. Pract.* **19**, 135 (2018).

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

122.   Dijkhuizen, F. P. H. L. J., Mol, B. W. J., Brolmann, H. A. M. & Heintz, A. P. M. The accuracy of endometrial sampling in the diagnosis of patients with endometrial carcinoma and hyperplasia. *Cancer* **89**, 1765–1772 (2000).

123.   Laban, M., Nassar, S., Elsayed, J. & Hassanin, A. S. Correlation between pre-operative diagnosis and final pathological diagnosis of endometrial malignancies; impact on primary surgical treatment. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **263**, 100–105 (2021).

124.   Xie, B. *et al.* Risk Factors for Unsuccessful Office-Based Endometrial Biopsy: A Comparative Study of Office-Based Endometrial Biopsy (Pipelle) and Diagnostic Dilation and Curettage. *J. Minim. Invasive Gynecol.* **25**, 724–729 (2018).

125.   Sanam, M. & Majid, M. M. K. Comparison the Diagnostic Value of Dilatation and Curettage Versus Endometrial Biopsy by Pipelle - a Clinical Trial. *Asian Pacific J. Cancer Prev.* **16**, 4971–4975 (2015).

126.   Utida, G. M. & Kulak, J. Hysteroscopic and aspiration biopsies in the histologic evaluation of the endometrium, a comparative study. *Medicine (Baltimore).* **98**, e17183 (2019).

127. Piriyev, E., Mellin, W. & Römer, T. Comparison of aspirating pipettes and hysteroscopy with curettage. *Arch. Gynecol. Obstet.* **301**, 1485–1492 (2020).

128. Gungorduk, K. *et al.* Comparison of the histopathological diagnoses of preoperative dilatation and curettage and Pipelle biopsy. *Eur. J. Gynaecol. Oncol.* **35**, 539–43 (2014).

129. Abdelazim, I. A., Aboelezz, A. & Abdulkareem, A. F. Pipelle endometrial sampling versus conventional dilatation & curettage in patients with abnormal uterine bleeding. *J. Turkish Ger. Gynecol. Assoc.* **14**, 1–5 (2013).

130. Abdelazim, I. A., Abdelrazak, K. M., Elbiaa, A. A. M., Al-Kadi, M. & Yehia, A. H. Accuracy of endometrial sampling compared to conventional dilatation and curettage in women with abnormal uterine bleeding. *Arch. Gynecol. Obstet.* **291**, 1121–1126 (2015).

131. Demirkiran, F. *et al.* Which is the best technique for endometrial sampling? Aspiration (pipelle) versus dilatation and curettage (D&C). *Arch. Gynecol. Obstet.* **286**, 1277–1282 (2012).

132. Ilavarasi, Cr., Jyothi, G. & Alva, N. Study of the efficacy of pipelle biopsy technique to diagnose endometrial diseases in abnormal uterine bleeding. *J. Midlife. Health* **10**, 75 (2019).

133. Reijnen, C. *et al.* Diagnostic accuracy of endometrial biopsy in relation to the amount of tissue. *J. Clin. Pathol.* **70**, 941–946 (2017).

134. Taraboanta, C. *et al.* Performance Characteristics of Endometrial Sampling in Diagnosis of Endometrial Carcinoma. *Int. J. Gynecol. Pathol.* **39**, 19–25 (2020).

135. van Hanegem, N. *et al.* The accuracy of endometrial sampling in women with postmenopausal bleeding: a systematic review and meta-analysis. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **197**, 147–155 (2016).

136. Visser, N. C. M. *et al.* Accuracy of endometrial sampling in endometrial carcinoma: A systematic review and meta-analysis. *Obstet. Gynecol.* **130**, 803–813 (2017).

137. Issat, T., Beta, J., Nowicka, M. A. & Jakimiuk, A. J. Accuracy and diagnostic value of outpatient hysteroscopy for malign and benign disease. *Eur. J. Gynaecol. Oncol.* **35**, 52–5 (2014).

138. Trojano, G. *et al.* The Role of Hysteroscopy in Evaluating Postmenopausal Asymptomatic Women with Thickened Endometrium. *Gynecol. Minim. invasive Ther.* **7**, 6–9.

139. 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).

140. Clark, T. J. *et al.* Accuracy of Hysteroscopy in the Diagnosis of Endometrial Cancer and Hyperplasia. *JAMA* **288**, 1610 (2002).

141. Van Dongen, H., De Kroon, C., Jacobi, C., Trimbos, J. & Jansen, F. Diagnostic hysteroscopy in abnormal uterine bleeding: a systematic review and meta-analysis. *BJOG An Int. J. Obstet. Gynaecol.* **114**, 664–675 (2007).

142. Phelippeau, J. *et al.* Preoperative diagnosis of tumor grade and type in endometrial cancer by pipelle sampling and hysteroscopy: Results of a French study. *Surg. Oncol.* **25**, 370–377 (2016).

143. Di Cello, A. *et al.* Failure to recognize preoperatively high-risk endometrial carcinoma is associated with a poor outcome. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **194**, 153–160 (2015).

144. Manchanda, R. & Sharma, R. Complications of Hysteroscopy and Management. *Hysteroscopy Simpl. by Masters* 211–222 (2021) doi:10.1007/978-981-15-2505-6_26.

145. Propst, A. Complications of hysteroscopic surgery: predicting patients at risk. *Obstet. Gynecol.* **96**, 517–520 (2000).

146. 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.e2 (2011).

147. H, D., Y, W., M, Z., M, S. & Y, Y. Whether preoperative hysteroscopy increases the dissemination of endometrial cancer cells: A systematic review and meta-analysis. *J. Obstet. Gynaecol. Res.* (2021) doi:10.1111/JOG.14897.

148. Alcázar, J. L. *et al.* Transvaginal ultrasound versus magnetic resonance imaging for preoperative assessment of myometrial infiltration in patients with endometrial cancer: a systematic review and meta-analysis. *J. Gynecol. Oncol.* **28**, 86 (2017).

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

150. Q, B. *et al.* The Diagnostic Value of MRI for Preoperative Staging in Patients with Endometrial Cancer: A Meta-Analysis. *Acad. Radiol.* **27**, 960–968 (2020).

151. Epstein, E. & Blomqvist, L. Imaging in endometrial cancer. *Best Pract. Res. Clin. Obstet. Gynaecol.* **28**, 721–739 (2014).

152. K, K., K, M., E, Y., Y, K. & K, S. Accuracy of integrated FDG-PET/contrast-enhanced CT in detecting pelvic and paraaortic lymph node metastasis in patients with uterine cancer. *Eur. Radiol.* **19**, 1529–1536 (2009).

153. 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).

154. Dutta, S. W., Trifiletti, D. M., Grover, S., Boimel, P. & Showalter, T. N. Management of elderly patients with early-stage medically inoperable endometrial cancer: Systematic review and National Cancer Database analysis. *Brachytherapy* **16**, 526–533 (2017).

155. 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).

156. Sany, O., Singh, K. & Jha, S. Correlation between preoperative endometrial sampling and final endometrial cancer histology. *Eur. J. Gynaecol. Oncol.* **33**, 142–4 (2012).

157. 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).

158. Abdulfatah, E. *et al.* Molecular classification of endometrial carcinoma applied to endometrial biopsy specimens: Towards early personalized patient management. *Gynecol. Oncol.* **154**, 467–474 (2019).

159. de Boer, S. M. *et al.* Adjuvant chemoradiotherapy versus radiotherapy alone in women with high-risk endometrial cancer (PORTEC-3): patterns of recurrence and post-hoc survival analysis of a randomised phase 3 trial. *Lancet Oncol.* **20**, 1273–1285 (2019).

160. ASTEC study group; Kitchener, H; Swart, AMC; Qian, Q; Amos, C; Parmar, M. Efficacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. *Lancet* **373**, 125–136 (2009).

161. Multinu, F. *et al.* Ultrastaging of negative pelvic lymph nodes to decrease the true prevalence of isolated paraaortic dissemination in endometrial cancer. *Gynecol. Oncol.* **154**, 60–64 (2019).

162. 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).

163. Kennard, J. A. *et al.* Sentinel lymph nodes (SLN) in endometrial cancer: The relationship between primary tumor histology, SLN metastasis size, and non-sentinel node metastasis. *Gynecol. Oncol.* **154**, 53–59 (2019).

164. Deura, I. *et al.* Comparison of laparoscopic surgery and conventional laparotomy for surgical staging of patients with presumed low-risk endometrial cancer: The current state of Japan. *Taiwan. J. Obstet. Gynecol.* **58**, 99–104 (2019).

165. Stelloo, E. *et al.* Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod. Pathol.* **28**, 836–844 (2015).

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

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

168. Florkowski, C. M. *Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests*. *Clin Biochem Rev* vol. 29 (2008).

169. Christenson, R. H. *Evidence-based laboratory medicine-a guide for critical evaluation of in vitro laboratory testing*. *Ann Clin Biochem* vol. 44 http://www.cochrane.org (2007).

170. Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J. & Bossuyt, P. M. M. The diagnostic odds ratio: a single indicator of test performance. *J. Clin. Epidemiol.* **56**, 1129–1135 (2003).

171. Alberg, A. J., Park, J. W., Hager, B. W., Brock, M. V. & Diener-West, M. The use of 'overall accuracy' to

evaluate the validity of screening or diagnostic tests. *Journal of General Internal Medicine* vol. 19 460–465 (2004).

172.   Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**, 861–874 (2006).

173.   Ray, P., Manach, Y. Le, Riou, B. & Houle, T. T. Statistical evaluation of a biomarker. *Anesthesiology* vol. 112 1023–1040 (2010).

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

175.   Legrain, P. *et al.* The Human Proteome Project: Current State and Future Direction. *Mol. Cell. Proteomics* **10**, M111.009993 (2011).

176.   Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).

177.   ON, J. Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **8**, 33–41 (2004).

178.   Bose, U., Wijffels, G., Howitt, C. A. & Colgrave, M. L. Proteomics: Tools of the Trade. in *Advances in Experimental Medicine and Biology* vol. 1073 1–22 (Springer New York LLC, 2019).

179.   Anderson, N. L. & Anderson, N. G. The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* **1**, 845–867 (2002).

180.   Geyer, P. E., Holdt, L. M., Teupser, D. & Mann, M. Revisiting biomarker discovery by plasma proteomics. *Mol. Syst. Biol.* **13**, 942 (2017).

181.   NL, H. & DF, H. Cancer biomarkers. *Mol. Oncol.* **6**, 140–146 (2012).

182.   Nakayasu, E. S. *et al.* Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **16**, 3737–3760 (2021).

183.   Ioannidis, J. P. A. & Bossuyt, P. M. M. Waste, leaks, and failures in the biomarker pipeline. *Clin. Chem.* **63**, 963–972 (2017).

184.   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).

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

186.   Ryu, J. & Thomas, S. N. Quantitative Mass Spectrometry-Based Proteomics for Biomarker Development in Ovarian Cancer. *Molecules* **26**, 2674 (2021).

187.   Biomarkers — Early Detection Research Network. https://edrn.nci.nih.gov/data-and-resources/biomarkers#c0=20&b_start=0.

188. Diamandis, E. P. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? *BMC Med.* **10**, 87 (2012).

189. Kearney, P., Boniface, J. J., Price, N. D. & Hood, L. The building blocks of successful translation of proteomics to the clinic. *Curr. Opin. Biotechnol.* **51**, 123–129 (2018).

190. Martinez-Garcia, E. *et al.* Advances in endometrial cancer protein biomarkers for use in the clinic. *Expert Rev. Proteomics* **15**, 81–99 (2018).

191. Tonry, C. *et al.* The Role of Proteomics in Biomarker Development for Improved Patient Diagnosis and Clinical Decision Making in Prostate Cancer. *Diagnostics* **6**, 27 (2016).

192. David Piehowski, P. *et al.* Protein Extraction From FFPE Kidney Tissue Samples: A Review of the Literature and Characterization of Techniques. *Front. Med. | www.frontiersin.org* **1**, 657313 (2021).

193. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2**, 185–195 (2016).

194. Geyer, P. E. *et al.* Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol. Med.* **11**, 1–12 (2019).

195. Njoku, K. *et al.* Urinary Biomarkers and Their Potential for the Non-Invasive Detection of Endometrial Cancer. doi:10.3389/fonc.2020.559016.

196. LI, Y. *et al.* Deep Dive on the Proteome of Human Body Fluids: A Valuable Data Resource for Biomarker Discovery. *Cancer Genomics - Proteomics* **18**, 549–568 (2021).

197. Jones, E. R., O'Flynn, H., Njoku, K. & Crosbie, E. J. Detecting endometrial cancer. *Obstet. Gynaecol.* **23**, 103–112 (2021).

198. 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. Invest.* **76**, 109–16 (1997).

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

200. Martinez-Garcia, E. *et al.* 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. *Clin. Cancer Res.* **23**, 6458–6467 (2017).

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

202. Fernández-González, S. *et al.* Cost-effective improvement in endometrial cancer diagnosis by the incorporation of molecular analysis. *Progresos Obstet. y Ginecol.* **62**, 247–253 (2019).

203. Martinez-Garcia, E. *et al.* Development of a sequential workflow based on LC-PRM for the verification of

endometrial cancer protein biomarkers in uterine aspirate samples. *Oncotarget* **7**, 53102–53115 (2016).

204. Fambrini, M. *et al.* Endometrial carcinoma in high-risk populations: is it time to consider a screening policy? *Cytopathology* **25**, 71–77 (2014).

205. Wang, Y. *et al.* Evaluation of liquid from the Papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Sci. Transl. Med.* **10**, 463–464 (2018).

206. Bakkum-Gamez, J. N. *et al.* Detection of endometrial cancer via molecular analysis of DNA collected with vaginal tampons. *Gynecol. Oncol.* **137**, 14–22 (2015).

207. Frias-Gomez, J. *et al.* Sensitivity of cervico-vaginal cytology in endometrial carcinoma: A systematic review and meta-analysis. *Cancer Cytopathol.* **128**, 792–802 (2020).

208. Doufekas, K. *et al.* DNA Methylation Signatures in Vaginal Fluid Samples for Detection of Cervical and Endometrial Cancer. *Int. J. Gynecol. Cancer* **00**, 1 (2016).

209. Calis, P. *et al.* Assessment of Cervicovaginal Cancer Antigen 125 Levels: A Preliminary Study for Endometrial Cancer Screening. *Gynecol. Obstet. Invest.* **81**, 518–522 (2016).

210. Cheng, S.-C. *et al.* Metabolomic biomarkers in cervicovaginal fluid for detecting endometrial cancer through nuclear magnetic resonance spectroscopy. *Metabolomics* **15**, 146 (2019).

211. Su, M. *et al.* Proteomics, Personalized Medicine and Cancer. *Cancers (Basel).* **13**, 2512 (2021).

212. JC, T. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–258 (2011).

213. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

214. Ignjatovic, V. *et al.* Mass Spectrometry-Based Plasma Proteomics: Considerations from Sample Collection to Achieving Translational Data. *J. Proteome Res.* **18**, 4085–4097 (2019).

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

216. Pitt, J. J. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin. Biochem. Rev.* **30**, 19–34 (2009).

217. Shi, Y., Xiang, R., Horváth, C. & Wilkins, J. A. The role of liquid chromatography in proteomics. *J. Chromatogr. A* **1053**, 27–36 (2004).

218. Mitulović, G. New HPLC Techniques for Proteomics Analysis: A Short Overview of Latest Developments. *http://dx.doi.org/10.1080/10826076.2014.941266* **38**, 390–403 (2014).

219. Bian, Y. *et al.* Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS. *Nat. Commun. 2020 111* **11**, 1–12 (2020).

220. Žuvela, P. *et al.* Column Characterization and Selection Systems in Reversed-Phase High-Performance Liquid Chromatography. *Chem. Rev.* **119**, 3674–3729 (2019).

221. Muntel, J. *et al.* Abundance-based Classifier for the Prediction of Mass Spectrometric Peptide Detectability Upon Enrichment (PPA). *Mol. Cell. Proteomics* **14**, 430 (2015).

222. Haag, A. M. Mass analyzers and mass spectrometers. *Adv. Exp. Med. Biol.* **919**, 157–169 (2016).

223. B, D. & R, A. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).

224. JR, O. An Overview of Mass Spectrometry-Based Methods for Functional Proteomics. *Methods Mol. Biol.* **1871**, 179–196 (2019).

225. He, F., Emmett, M. R., Håkansson, K., Hendrickson, C. L. & Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement. *J. Proteome Res.* **3**, 61–67 (2004).

226. M, M. & NL, K. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18132–18138 (2008).

227. Ong, S.-E. & Mann, M. Mass spectrometry–based proteomics turns quantitative. *Nat. Chem. Biol. 2005 15* **1**, 252–262 (2005).

228. Wal, L. van der & Demmers, J. A. A. Quantitative Mass Spectrometry-based Proteomics. *Recent Adv. Proteomics Res.* (2015) doi:10.5772/61756.

229. Gallien, S. & Domon, B. Detection and quantification of proteins in clinical samples using high resolution mass spectrometry. *Methods* **81**, 15–23 (2015).

230. Calderón-Celis, F., Encinar, J. R. & Sanz-Medel, A. Standardization approaches in absolute quantitative proteomics with mass spectrometry. *Mass Spectrom. Rev.* **37**, 715–737 (2018).

231. Anand, S., Samuel, M., Ang, C. S., Keerthikumar, S. & Mathivanan, S. Label-based and label-free strategies for protein quantitation. *Methods Mol. Biol.* **1549**, 31–43 (2017).

232. Rami, M. *et al.* Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. (2017) doi:10.1021/acs.jproteome.6b00645.

233. DC, T. *et al.* Comparative evaluation of label-free SINQ normalized spectral index quantitation in the central proteomics facilities pipeline. *Proteomics* **11**, 2790–2797 (2011).

234. Lesur, A. & Domon, B. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. *Proteomics* **15**, 880–890 (2015).

235. B, U. *et al.* Two Dimensional-Difference in Gel Electrophoresis (2D-DIGE) Proteomic Approach for the Identification of Biomarkers in Endometrial Cancer Serum. *Cancers (Basel).* **13**, (2021).

236. Rabilloud, T. Two-dimensional gel electrophoresis in proteomics: old, old fashioned, but it still climbs up

the mountains. *Proteomics* **2**, 3–10 (2002).

237. Nilsson, T. *et al.* Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods 2010 79* **7**, 681–685 (2010).

238. Cox, J. *et al.* Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **10**, 1794–1805 (2011).

239. Issaq, H. J. & Veenstra, T. D. *Proteomic and metabolomic approaches to biomarker discovery*. (Academic Press Inc., 2020).

240. GL, A., BL, S., JB, Y., AM, H. & DC, M. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* **83**, 5442–5446 (2011).

241. Zubarev, R. A. & Makarov, A. Orbitrap Mass Spectrometry. *Anal. Chem.* **85**, 5288–5296 (2013).

242. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc. 2016 1112* **11**, 2301–2319 (2016).

243. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods 2016 139* **13**, 731–740 (2016).

244. DL, T. *et al.* Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **9**, 761–776 (2010).

245. 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* **11**, (2012).

246. BC, C. *et al.* Quantifying protein interaction dynamics by SWATH mass spectrometry: application to the 14-3-3 system. *Nat. Methods* **10**, 1246–1253 (2013).

247. G, R. *et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. data* **1**, (2014).

248. Y, L. *et al.* Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* **11**, 786 (2015).

249. Ludwig, C. *et al.* Data-independent acquisition-based SWATH - MS for quantitative proteomics: a tutorial . *Mol. Syst. Biol.* **14**, 1–23 (2018).

250. T, G., J, C. & M, M. Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Mol. Cell. Proteomics* **9**, 2252–2261 (2010).

251. RS, P. *et al.* UPLC/MS(E); a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Commun. Mass Spectrom.* **20**, 1989–1994 (2006).

252. L, R., MJ, S. & A, M. Data-Independent Acquisition for the Orbitrap Q Exactive HF: A Tutorial. *J. Proteome Res.* **18**, 803–813 (2019).

253. Sidoli, S., Fujiwara, R. & Garcia, B. A. Multiplexed data independent acquisition (MSX-DIA) applied by high

resolution mass spectrometry improves quantification quality for the analysis of histone peptides. *Proteomics* **16**, 2095 (2016).

254. Barkovits, K. *et al.* Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition*. *Mol. Cell. Proteomics* **19**, 181–197 (2020).

255. Yang, Y. *et al.* In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. *Nat. Commun. 2020 111* **11**, 1–11 (2020).

256. F, Z., W, G., G, R., X, C. & T, G. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **20**, (2020).

257. T, G. *et al.* Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **21**, 407–413 (2015).

258. I, O., A, R.-A., E, C.-G., MS, A. V. & B, J. G. 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).

259. Y, L. *et al.* Glycoproteomic analysis of prostate cancer tissues by SWATH mass spectrometry discovers N-acylethanolamine acid amidase and protein tyrosine kinase 7 as signatures for tumor aggressiveness. *Mol. Cell. Proteomics* **13**, 1753–1768 (2014).

260. Bouchal, P. *et al.* Breast Cancer Classification Based on Proteotypes Obtained by SWATH Mass Spectrometry. *Cell Rep.* **28**, 832-843.e7 (2019).

261. Aboulouard, S. *et al.* In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer. *Cell Reports Med.* **2**, (2021).

262. Niu, L., Geyer, P. E. & Mann, M. Proteomics in the study of liver diseases. *Hum. Gut-Liver-Axis Heal. Dis.* 165–193 (2018) doi:10.1007/978-3-319-98890-0_11.

263. Picotti, P. & Aebersold, R. Selected reaction monitoring–based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods 2012 96* **9**, 555–566 (2012).

264. 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* **11**, 1475–1488 (2012).

265. LK, P. *et al.* The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom. Rev.* **39**, 229–244 (2020).

266. Method of the Year 2012. *Nat. Methods 2013 101* **10**, 1–1 (2013).

267. JD, W., J, S., Y, Y. & C, J. Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. *Nat. Methods* **11**, 1041–1044 (2014).

268. Chiva, C. & Sabidó, E. On peptide selection for targeted protein quantitation.

269.   Gallien, S., Duriez, E. & Domon, B. Selected reaction monitoring applied to proteomics. *J. Mass Spectrom.* **46**, 298–312 (2011).

270.   Hüttenhain, R., Malmström, J., Picotti, P. & Aebersold, R. Perspectives of targeted mass spectrometry for protein biomarker verification. *Curr. Opin. Chem. Biol.* **13**, 518 (2009).

271.   P, P., B, B., LN, M., B, D. & R, A. Full dynamic range proteome analysis of S. cerevisiae by targeted proteomics. *Cell* **138**, 795–806 (2009).

272.   V, L., P, P., B, D. & R, A. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, (2008).

273.   Kim, B. *et al.* Affinity Enrichment for MS: Improving the yield of low abundance biomarkers. *Expert Rev. Proteomics* **15**, 353 (2018).

274.   Domon, B. & Gallien, S. Recent advances in targeted proteomics for clinical applications. *Proteomics - Clin. Appl.* **9**, 423–431 (2015).

275.   Borràs, E. & Sabidó, E. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics* **17**, 17–18 (2017).

276.   CM, H. *et al.* Quantification by nano liquid chromatography parallel reaction monitoring mass spectrometry of human apolipoprotein A-I, apolipoprotein B, and hemoglobin A1c in dried blood spots. *Proteomics. Clin. Appl.* **11**, (2017).

277.   V, B. *et al.* New Quantitative Mass Spectrometry Approaches Reveal Different ADP-ribosylation Phases Dependent On the Levels of Oxidative Stress. *Mol. Cell. Proteomics* **16**, 949–958 (2017).

278.   Bourmaud, A., Gallien, S. & Domon, B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. *Proteomics* **16**, 2146–2159 (2016).

279.   C, C., H, H. & CH, W. Protein Bioinformatics Databases and Resources. *Methods Mol. Biol.* **1558**, 3–39 (2017).

280.   Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2**, 401–404 (2012).

281.   J, G. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, (2013).

282.   EW, D. *et al.* The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **48**, D1145–D1152 (2020).

283.   Samaras, P. *et al.* ProteomicsDB: a multi-omics and multi-organism resource for life science research. *Nucleic Acids Res.* **48**, D1153–D1163 (2020).

284.   DP, Z. *et al.* ProteomeTools: Systematic Characterization of 21 Post-translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Using Synthetic Peptides. *Mol. Cell.*

Proteomics **17**, 1850–1863 (2018).

285. Cox, K. L. *et al. Immunoassay Methods*. *Assay Guidance Manual* (2004).

286. Cohen, L. & Walt, D. R. Highly Sensitive and Multiplexed Protein Measurements. *Chem. Rev.* **119**, 293–321 (2019).

287. Baker, M. Blame it on the antibodies. *Nature* **521**, 274–276 (2015).

288. Carlyle, B., Trombetta, B. & Arnold, S. Proteomic Approaches for the Discovery of Biofluid Biomarkers of Neurodegenerative Dementias. *Proteomes* **6**, 32 (2018).

289. Cross, T. G. & Hornshaw, M. P. Can LC and LC-MS ever replace immunoassays? *J. Appl. Bioanal.* **2**, 108–116 (2016).

290. TB, M. Development of internal controls for the Luminex instrument as part of a multiplex seven-analyte viral respiratory antibody profile. *Clin. Diagn. Lab. Immunol.* **9**, 41–45 (2002).

291. Astrup, K. & Olivarius, N. D. F. Frequency of spontaneously occurring postmenopausal bleeding in the general population. *Acta Obstet. Gynecol. Scand.* **83**, 203–207 (2004).

292. 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 (2016).

293. Martinez-Garcia, E. *et al.* Development of a sequential workflow based on LC-PRM for the verification of endometrial cancer protein biomarkers in uterine aspirate samples. *Oncotarget* **7**, 53102–53115 (2016).

294. Cheng, S. C. *et al.* Metabolomic biomarkers in cervicovaginal fluid for detecting endometrial cancer through nuclear magnetic resonance spectroscopy. *Metabolomics* **15**, (2019).

295. Shao, X. *et al.* Screening and verifying endometrial carcinoma diagnostic biomarkers based on a urine metabolomic profiling study using UPLC-Q-TOF/MS. *Clin. Chim. Acta* **463**, 200–206 (2016).

296. Paraskevaidi, M. *et al.* Potential of mid-infrared spectroscopy as a non-invasive diagnostic test in urine for endometrial or ovarian cancer. *Analyst* **143**, 3156–3163 (2018).

297. O'Flynn, H. *et al.* Diagnostic accuracy of cytology for the detection of endometrial cancer in urine and vaginal samples. *Nat. Commun.* **12**, 1–8 (2021).

298. Dou, Y. *et al.* Proteogenomic Characterization of Endometrial Carcinoma. *Cell* **180**, 729-748.e26 (2020).

299. McShane, L. M. *et al.* REporting recommendations for tumour MARKer prognostic studies (REMARK). *Br. J. Cancer* **93**, 387–391 (2005).

300. Sauerbrei, W., Taube, S. E., McShane, L. M., Cavenagh, M. M. & Altman, D. G. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An abridged explanation and elaboration. *J. Natl. Cancer Inst.* **110**, 803–811 (2018).

301. Carbon, S. *et al.* Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. *Nucleic Acids Res.* **45**, D331–D338 (2017).

302. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* **47**, D590–D595 (2019).

303. Nazarov, P. V. *et al.* Deconvolution of transcriptomes and miRNomes by independent component analysis provides insights into biological processes and clinical outcomes of melanoma patients. *BMC Med. Genomics* **12**, (2019).

304. Thomas, P. D. *et al.* PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res.* **13**, 2129–2141 (2003).

305. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

306. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).

307. Ochoa, D. *et al.* Open Targets Platform: Supporting systematic drug-target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).

308. Kosmas, K. *et al.* Expression of ki-67 as proliferation biomarker in imprint smears of endometrial carcinoma. *Diagn. Cytopathol.* **41**, 212–217 (2013).

309. Apostolou, G., Apostolou, N., Moulos, P. & Chatzipantelis, P. Endometrial cytopathology. An image analysis approach using the Ki-67 biomarker. *Cytopathology* **28**, 385–390 (2017).

310. Kosmas, K., Mitropoulou, G., Provatas, I., Stamoulas, M. & Marouga, A. Expression of phosphatase and tensin homologue in imprint smears of endometrial carcinoma. *Cytopathology* **29**, 558–564 (2018).

311. Zanotti, L. *et al.* Human epididymis protein 4 as a serum marker for diagnosis of endometrial carcinoma and prediction of clinical outcome. *Clin. Chem. Lab. Med.* **50**, 2189–2198 (2012).

312. Chao, A. *et al.* Potential of an age-stratified CA125 cut-off value to improve the prognostic classification of patients with endometrial cancer. *Gynecol. Oncol.* **129**, 500–504 (2013).

313. Stiekema, A. *et al.* Serum HE4 is correlated to prognostic factors and survival in patients with endometrial cancer. *Virchows Arch.* **470**, 655–664 (2017).

314. Tangen, I. L. *et al.* Expression of L1CAM in curettage or high L1CAM level in preoperative blood samples predicts lymph node metastases and poor outcome in endometrial cancer patients. *Br. J. Cancer* **117**, 840–847 (2017).

315. Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

316. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951 (2019).

317. Jiang, T. *et al.* MiR-29a-5p inhibits proliferation and invasion and induces apoptosis in endometrial

carcinoma via targeting TPX2. *Cell Cycle* **17**, (2018).

318.   Zhang, W. *et al.* Combining Bioinformatics and Experiments to Identify and Verify Key Genes with Prognostic Values in Endometrial Carcinoma. *J. Cancer* **11**, 716–732 (2020).

319.   Besso, M. *et al.* Identification of early stage recurrence endometrial cancer biomarkers using bioinformatics tools. *Oncol. Rep.* **44**, 873–886 (2020).

320.   Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84 (2019).

321.   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.* **17**, 1736–1736 (1999).

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

323.   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).

324.   Guan, J. *et al.* The prognostic significance of estrogen and progesterone receptors in grade I and II endometrioid endometrial adenocarcinoma: hormone receptors in risk stratification. *J. Gynecol. Oncol.* **30**, (2019).

325.   Smith, D. *et al.* ER and PR expression and survival after endometrial cancer. *Gynecol. Oncol.* **148**, 258–266 (2018).

326.   Expression of TPX2 in cancer - Summary - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000088325-TPX2/pathology.

327.   Jiang, Y., Liu, Y., Tan, X., Yu, S. & Luo, J. TPX2 as a Novel Prognostic Indicator and Promising Therapeutic Target in Triple-negative Breast Cancer. *Clin. Breast Cancer* **19**, 450–455 (2019).

328.   KY, C.-G. *et al.* RAN nucleo-cytoplasmic transport and mitotic spindle assembly partners XPO7 and TPX2 are new prognostic biomarkers in serous epithelial ovarian cancer. *PLoS One* **9**, (2014).

329.   Shen, L., Liu, M., Liu, W., Cui, J. & Li, C. Bioinformatics analysis of RNA sequencing data reveals multiple key genes in uterine corpus endometrial carcinoma. *Oncol. Lett.* **15**, 205–212 (2017).

330.   Liu, Y. *et al.* Comprehensive Analysis of the Control of Cancer Stem Cell Characteristics in Endometrial Cancer by Network Analysis. *Comput. Math. Methods Med.* **2021**, 1–11 (2021).

331.   Liu, J. *et al.* Using mRNAsi to identify prognostic-related genes in endometrial carcinoma based on WGCNA. *Life Sci.* **258**, 118231 (2020).

332.   Kommoss, F. K. *et al.* L1CAM further stratifies endometrial carcinoma patients with no specific molecular risk profile. *Br. J. Cancer* **119**, 480–486 (2018).

333.   R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

334.   Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).

335.   Coll-de la Rubia, E. *et al.* In silico Approach for Validating and Unveiling New Applications for Prognostic Biomarkers of Endometrial Cancer. *Cancers (Basel).* **13**, 5052 (2021).

336.   Lu, K. H. & Broaddus, R. R. Endometrial Cancer. *N. Engl. J. Med.* **383**, 2053–2064 (2020).

337.   Li, L. *et al.* Human epididymis protein 4 in endometrial cancer: A meta-analysis. *Clin. Chim. Acta* **482**, 215–223 (2018).

338.   Paraskevaidi, M. *et al.* Potential of mid-infrared spectroscopy as a non-invasive diagnostic test in urine for endometrial or ovarian cancer. *Analyst* **143**, 3156–3163 (2018).

339.   Paraskevaidi, M. *et al.* Detecting Endometrial Cancer by Blood Spectroscopy: A Diagnostic Cross-Sectional Study. *Cancers (Basel).* **12**, 1256 (2020).

340.   Njoku, K. *et al.* Comprehensive Library Generation for Identification and Quantification of Endometrial Cancer Protein Biomarkers in Cervico-Vaginal Fluid Citation: Njoku, K. *Cancers (Basel).* **13**, 3804 (2021).

341.   Guerra, E. & Matias-Guiu, X. Relevance of pathologic features in risk stratification for early-stage endometrial cancer. *J. Gynecol. Oncol.* **32**, 67 (2021).

342.   Bollineni, V. R., Ytre-Hauge, S., Bollineni-Balabay, O., Salvesen, H. B. & Haldorsen, I. S. High Diagnostic Value of 18F-FDG PET/CT in Endometrial Cancer: Systematic Review and Meta-Analysis of the Literature. *J. Nucl. Med.* **57**, 879–885 (2016).

343.   Fasmer, K. E. *et al.* Preoperative 18F-FDG PET/CT tumor markers outperform MRI-based markers for the prediction of lymph node metastases in primary endometrial cancer. *Eur. Radiol.* **30**, 2443–2453 (2020).

344.   Imboden, S. *et al.* Implementation of the 2021 molecular ESGO/ESTRO/ESP risk groups in endometrial cancer. *Gynecol. Oncol.* **162**, 394–400 (2021).

345.   Stasenko, M. *et al.* Clinical outcomes of patients with POLE mutated endometrioid endometrial cancer. *Gynecol. Oncol.* **156**, 194–202 (2020).

346.   Dhib, D. *et al.* 212 Multiplex qPCR hotspot testing of pathogenic POLE mutations: a rapid, simple and reliable approach for POLE assessment in endometrial cancer. *Int. J. Gynecol. Cancer* **31**, A367.2-A368 (2021).

347.   McMeekin, D. S. *et al.* Clinicopathologic Significance of Mismatch Repair Defects in Endometrial Cancer: An NRG Oncology/Gynecologic Oncology Group Study. *J. Clin. Oncol.* **34**, 3062–3068 (2016).

348.   Auguste, A. *et al.* Refinement of high-risk endometrial cancer classification using DNA damage response biomarkers: a TransPORTEC initiative. *Mod. Pathol.* **31**, 1851–1861 (2018).

349. Vrede, S. W. *et al.* Immunohistochemical biomarkers are prognostic relevant in addition to the ESMO-ESGO-ESTRO risk classification in endometrial cancer. *Gynecol. Oncol.* **161**, 787–794 (2021).

350. Vermij, L. *et al.* HER2 Status in High-Risk Endometrial Cancers (PORTEC-3): Relationship with Histotype, Molecular Classification, and Clinical Outcomes. *Cancers 2021, Vol. 13, Page 44* **13**, 44 (2020).

351. Reijnen, C. *et al.* Diagnostic Accuracy of Clinical Biomarkers for Preoperative Prediction of Lymph Node Metastasis in Endometrial Carcinoma: A Systematic Review and Meta-Analysis. *Oncologist* **24**, e880–e890 (2019).

352. Hoang, L. N. *et al.* Interobserver Agreement in Endometrial Carcinoma Histotype Diagnosis Varies Depending on The Cancer Genome Atlas (TCGA)-based Molecular Subgroup. *Am. J. Surg. Pathol.* **41**, 245–252 (2017).

353. Santoro, A. *et al.* New Pathological and Clinical Insights in Endometrial Cancer in View of the Updated ESGO/ESTRO/ESP Guidelines. (2021) doi:10.3390/cancers13112623.

354. Todo, Y. *et al.* A validation study of a scoring system to estimate the risk of lymph node metastasis for patients with endometrial cancer for tailoring the indication of lymphadenectomy. *Gynecol. Oncol.* **104**, 623–628 (2007).

355. Abu-Rustum, N. R. *et al.* The incidence of isolated paraaortic nodal metastasis in surgically staged endometrial cancer patients with negative pelvic lymph nodes. *Gynecol. Oncol.* **115**, 236–238 (2009).

356. Daix, M. *et al.* Concordance between preoperative ESMO-ESGO-ESTRO risk classification and final histology in early-stage endometrial cancer. *J. Gynecol. Oncol.* **32**, (2021).

357. Rossi, E. C. *et al.* A comparison of sentinel lymph node biopsy to lymphadenectomy for endometrial cancer staging (FIRES trial): a multicentre, prospective, cohort study. *Lancet Oncol.* **18**, 384–392 (2017).

358. Persson, J. *et al.* Pelvic Sentinel lymph node detection in High-Risk Endometrial Cancer (SHREC-trial)-the final step towards a paradigm shift in surgical staging. *Eur. J. Cancer* **116**, 77–85 (2019).

359. Cusimano, M. C. *et al.* Assessment of Sentinel Lymph Node Biopsy vs Lymphadenectomy for Intermediate- and High-Grade Endometrial Cancer Staging. *JAMA Surg.* **156**, 1 (2021).

360. Vowinckel, J. *et al.* The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000Research* **2**, 272 (2014).

361. 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).

362. Bratulic, S., Gatto, F. & Nielsen, J. The Translational Status of Cancer Liquid Biopsies. *Regen. Eng. Transl. Med.* **7**, 312–352 (2021).

363. Eriksson, L. S. E. *et al.* Combination of Proactive Molecular Risk Classifier for Endometrial cancer (ProMisE) with sonographic and demographic characteristics in preoperative prediction of recurrence or progression of endometrial cancer. *Ultrasound Obstet. Gynecol.* **58**, 457–468 (2021).

364. Dimitraki, M. *et al.* Clinical evaluation of women with PMB. Is it always necessary an endometrial biopsy to be performed? A review of the literature. *Arch. Gynecol. Obstet.* **283**, 261–266 (2011).

365. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol. 2008 2612* **26**, 1367–1372 (2008).

366. Cox, J. *et al.* Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics* **13**, 2513–2526 (2014).

367. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).

368. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

369. Boylan, K. L. M. *et al.* A feasibility study to identify proteins in the residual Pap test fluid of women with normal cytology by mass spectrometry-based proteomics. *Clin. Proteomics* **11**, 1–15 (2014).

370. Njoku, K., Chiasserini, D., Whetton, A. D. & Crosbie, E. J. Proteomic biomarkers for the detection of endometrial cancer. *Cancers (Basel).* **11**, 1–25 (2019).

371. Costas, L. *et al.* New perspectives on screening and early detection of endometrial cancer. *Int. J. Cancer* **145**, 3194–3206 (2019).

372. Reijnen, C. *et al.* Mutational analysis of cervical cytology improves diagnosis of endometrial cancer: A prospective multicentre cohort study. *Int. J. Cancer* **146**, 2628–2635 (2020).

373. Liew, P.-L. *et al.* Combined genetic mutations and DNA-methylated genes as biomarkers for endometrial cancer detection from cervical scrapings. *Clin. Epigenetics 2019 111* **11**, 1–10 (2019).

374. A, S. *et al.* Combining copy number, methylation markers, and mutations as a panel for endometrial cancer detection via intravaginal tampon collection. *Gynecol. Oncol.* **156**, 387–392 (2020).

375. Wang, Y. *et al.* Evaluation of liquid from the papanicolaou test and other liquid biopsies for the detection of endometrial and ovarian cancers. *Obstet. Gynecol. Surv.* **73**, 463–464 (2018).

376. RL, H. *et al.* Integrated Epigenomics Analysis Reveals a DNA Methylation Panel for Endometrial Cancer Detection Using Cervical Scrapings. *Clin. Cancer Res.* **23**, 263–272 (2017).

377. A, T. *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).

378. Martinez-Garcia, E. *et al.* Advances in endometrial cancer protein biomarkers for use in the clinic. *Expert Rev. Proteomics* **15**, (2018).

379. Frantzi, M., Bhat, A. & Latosinska, A. Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development. (2014).

380. A, D. M. *et al.* Proteomic Profiling of the Human Tissue and Biological Fluid Proteome. *J. Proteome Res.* **20**,

444–452 (2021).

381. Zegels, G., Van Raemdonck, G. A. A., Tjalma, W. A. A. & Van Ostade, X. W. M. Use of cervicovaginal fluid for the identification of biomarkers for pathologies of the female genital tract. *Proteome Sci.* **8**, 63 (2010).

382. K, N., CJ, S., AD, W. & EJ, C. Metabolomic Biomarkers for Detection, Prognosis and Identifying Recurrence in Endometrial Cancer. *Metabolites* **10**, 1–28 (2020).

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

384. Cancer Today. https://gco.iarc.fr/today/online-analysis-table?v=2020&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1.

385. Wild, C. P., Weiderpass, E. & Stewart, B. W. *World Cancer Report: Cancer Research for Cancer Prevention*. *Cancer Control* vol. 199 (2020).

386. L, B.-M. *et al.* Translational biomarkers in the era of precision medicine. *Adv. Clin. Chem.* **102**, 191–232 (2021).

387. Hartl, D. *et al.* Translational precision medicine: an industry perspective. *J Transl Med* **19**, 245 (2021).

388. Califf, R. M. Biomarker definitions and their applications. *Exp. Biol. Med.* **243**, 213–221 (2018).

389. Frantzi, M., Latosinska, A., Kontostathi, G. & Mischak, H. Clinical Proteomics: Closing the Gap from Discovery to Implementation. *Proteomics* **18**, (2018).

390. Tumor Markers in Common Use - National Cancer Institute. https://www.cancer.gov/about-cancer/diagnosis-staging/diagnosis/tumor-markers-list.

391. van Hanegem, N. *et al.* Diagnostic workup for postmenopausal bleeding: a randomised controlled trial. *BJOG An Int. J. Obstet. Gynaecol.* **124**, 231–240 (2017).

392. Campoy, I. *et al.* Exosome-like vesicles in uterine aspirates: a comparison of ultracentrifugation-based isolation protocols. *J. Transl. Med.* **14**, 180 (2016).

393. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (80-.).* **347**, (2015).

394. Dai, C. *et al.* A proteomics sample metadata representation for multiomics integration and big data analysis. *Nat. Commun.* **12**, 5854 (2021).

395. Searle, B. C. *et al.* Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. *Nat. Commun. 2018 91* **9**, 1–12 (2018).

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

Bibliography

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

398.   Suri, A., Perumal, V., Ammalli, P., Suryan, V. & Bansal, S. K. Diagnostic measures comparison for ovarian malignancy risk in Epithelial ovarian cancer patients: a meta-analysis. *Sci. Rep.* **11**, 17308 (2021).

399.   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).

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

401.   Berglund, L. *et al.* A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies. *Mol. Cell. Proteomics* **7**, 2019–2027 (2008).

402.   Weller, M. G. Quality Issues of Research Antibodies. *Anal. Chem. Insights* **11**, 21 (2016).

403.   Uhlen, M. *et al.* A proposal for validation of antibodies. *Nat. Methods* **13**, 823–827 (2016).

404.   Björling, E. & Uhlén, M. Antibodypedia, a portal for sharing antibody and antigen validation data. *Mol. Cell. Proteomics* **7**, 2028–2037 (2008).

405.   Bradshaw, R. A., Hondermarck, H. & Rodriguez, H. Cancer Proteomics and the Elusive Diagnostic Biomarkers. *Proteomics* **19**, 1800445 (2019).

406.   Uhlén, M. *et al.* The human secretome. *Sci. Signal.* **12**, (2019).

407.   United Nations Educational Scientific and Cultural Organization (UNESCO);, Statistics, T. U. I. for & (UIS). Global Investments in R&D. A snapshot of R&D expenditure. **59**, (2020).

408.   Rubio, D. M. *et al.* Defining Translational Research: Implications for Training. *Acad. Med.* **85**, 470 (2010).

409.   G, L., AM, S., L, R.-M., P, B. & G, B. Standardizing in vitro diagnostics tasks in clinical trials: a call for action. *Ann. Transl. Med.* **4**, (2016).

410.   Ilan, Y. Why scientists, academic institutions, and investors fail in bringing more products to the bedside: the Active Compass model for overcoming the innovation paradox. *J. Transl. Med.* **19**, 55 (2021).

411.   Council, E. P. and of the. *Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (Text with EEA relevance.)*. 176–332 (Official Journal of the European Union, 2017).

412.   EvaluateMedTech - World Preview 2018, Outlook to 2024. https://www.evaluate.com/sites/default/files/media/download-files/WPMT2018_0.pdf.

413.   In-Vitro Diagnostics Market Size, Share and Trends forecast to 2025 by Product & Service, Technology, Application, End User | COVID-19 Impact Analysis | MarketsandMarkets™ https://www.marketsandmarkets.com/Market-Reports/ivd-in-vitro-diagnostics-market-703.html.

414. The World Health Organization. In vitro diagnostics - Global. https://www.who.int/health-topics/in-vitro-diagnostics#tab=tab_1.

415. von Lode, Pila; Daussin, Valéria; Raths, Susan; Padurariu, S. Best Practices and Pitfalls in Commercializing IVD-Applicable Biomarkers. *Zenodo* (2020) doi:https://doi.org/10.5281/zenodo.3928239.

416. Vogeser, M. & Zhang, Y. V. Understanding the strategic landscape surrounding the implementation of mass spectrometry in the clinical laboratory: A SWOT analysis. *Clin. Mass Spectrom.* **9**, 1–6 (2018).

417. Jannetto, P. J. & Fitzgerald, R. L. Effective Use of Mass Spectrometry in the Clinical Laboratory. (2015) doi:10.1373/clinchem.2015.248146.

418. Maurer, H. H. Mass spectrometry for research and application in therapeutic drug monitoring or clinical and forensic toxicology. *Ther. Drug Monit.* **40**, 389–393 (2018).

419. Peersman, N. *et al.* UPLC-MS/MS method for determination of retinol and α-tocopherol in serum using a simple sample pretreatment and UniSpray as ionization technique to reduce matrix effects. *Clin. Chem. Lab. Med.* **58**, 769–779 (2020).

420. Danese, E., Montagnana, M., Brentegani, C. & Lippi, G. Short-term stability of free metanephrines in plasma and whole blood. *Clin. Chem. Lab. Med.* **58**, 753–757 (2020).

421. Hawley, J. M., Adaway, J. E., Owen, L. J. & Keevil, B. G. Development of a total serum testosterone, androstenedione, 17-hydroxyprogesterone, 11β-hydroxyandrostenedione and 11-ketotestosterone LC-MS/MS assay and its application to evaluate pre-analytical sample stability. *Clin. Chem. Lab. Med.* **58**, 741–752 (2020).

422. Wiesinger, T. *et al.* Investigating the suitability of high-resolution mass spectrometry for newborn screening: Identification of hemoglobinopathies and β-thalassemias in dried blood spots. *Clin. Chem. Lab. Med.* **58**, 810–816 (2020).

423. Seger, C. & Salzmann, L. After another decade: LC-MS/MS became routine in clinical diagnostics. *Clin. Biochem.* **82**, 2–11 (2020).

424. Nedelkov, D. Mass spectrometry protein tests: ready for prime time (or not). *Expert Rev. Proteomics* **14**, 1–7 (2017).

425. Li, X. J. *et al.* A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci. Transl. Med.* **5**, (2013).

426. Vachani, A. *et al.* Clinical Utility of a Plasma Protein Classifier for Indeterminate Lung Nodules. *Lung* **193**, 1023–1027 (2015).

427. Saade, G. R. *et al.* Development and validation of a spontaneous preterm delivery predictor in asymptomatic women. *Am. J. Obstet. Gynecol.* **214**, 633.e1-633.e24 (2016).

428. Bradford, C. *et al.* Analytical validation of protein biomarkers for risk of spontaneous preterm birth. *Clin. Mass Spectrom.* **3**, 25–38 (2017).

429.   Taussig, M. J., Fonseca, C. & Trimmer, J. S. Antibody validation: a view from the mountains. doi:10.1016/j.nbt.2018.08.002.

430.   Medicines Agency, E. ICH guideline M10 on bioanalytical method validation 4 Step 2b 5. (2019).

431.   Fda & Cder. Bioanalytical Method Validation Guidance for Industry Biopharmaceutics Bioanalytical Method Validation Guidance for Industry Biopharmaceutics Contains Nonbinding Recommendations. (2018).

432.   Overview of IVD Regulation | FDA. https://www.fda.gov/medical-devices/ivd-regulatory-assistance/overview-ivd-regulation.

433.   What Are In Vitro Diagnostic Tests, and How Are They Regulated? | The Pew Charitable Trusts. https://www.pewtrusts.org/es/research-and-analysis/issue-briefs/2019/05/what-are-in-vitro-diagnostic-tests-and-how-are-they-regulated.

434.   Population of WORLD 2019 - PopulationPyramid.net. https://www.populationpyramid.net/.

435.   Proteomics Market | By Product, Reagent, Application | Global Forecast to 2026 | MarketsandMarkets™ https://www.marketsandmarkets.com/Market-Reports/proteomics-market-731.html.

436.   Immunoassay Market Size & Share Report, 2021-2028. https://www.grandviewresearch.com/industry-analysis/immunoassay-market.

437.   Njoku, K. *et al.* Comprehensive Library Generation for Identification and Quantification of Endometrial Cancer Protein Biomarkers in Cervico-Vaginal Fluid. *Cancers (Basel).* **13**, 3804 (2021).

438.   Costas, L. *et al.* Defining a mutational signature for endometrial cancer screening and early detection. *Cancer Epidemiol.* **61**, 129–132 (2019).

439.   Gontero, P. *et al.* Comparison of the performances of the ADXBLADDER test and urinary cytology in the follow-up of non-muscle-invasive bladder cancer: a blinded prospective multicentric study. *BJU Int.* **127**, 198–204 (2021).

440.   Stockley, J. *et al.* Detection of MCM5 as a novel non-invasive aid for the diagnosis of endometrial and ovarian tumours. *BMC Cancer* **20**, 1000 (2020).

441.   Introducción, C. D. E. & Investigación, A. L. A. Capítulo 8: Pruebas diagnósticas. Concordancia. **33**, (2007).

442.   Kinde, I. *et al.* Evaluation of DNA from the Papanicolaou Test to Detect Ovarian and Endometrial Cancers. *Sci. Transl. Med.* **5**, 167ra4-167ra4 (2013).

443.   Consortium, T. U. *et al.* UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

444.   Vasaikar, S. V. *et al.* EMTome: a resource for pan-cancer analysis of epithelial-mesenchymal transition genes and signatures. *Br. J. Cancer 2020 1241* **124**, 259–269 (2020).

Annexes

ANNEX 1

Table A1. 1 List of 255 proteins associated to prognostic factors in EC based on the literature review described in Chapter 1. Highlighted in blue, the 30 validated proteins identified from the *in-silico* analysis explained in Chapter 1. In bold, 158 proteins identified by DDA in uterine fluid samples and included in the spectral library generated in Chapter 2. Highlighted in pink, the 15 proteins analyzed by DIA showing prognostic significance (either for histological type, histological grade and/or recurrence) in uterine fluids, explained in Chapter 2. Highlighted in green, 3 proteins that were validated for some prognostic factor in both: in the *in-silico* analysis in Chapter 1 and in uterine fluids in Chapter 2.

Figure A1. 1 Supplementary Boxplots of the described validated biomarkers for histological type representing different NEEC as different entities (n=271 EEC, n=62 NEEC including n=10 mixed carcinomas, and n=52 SEC).
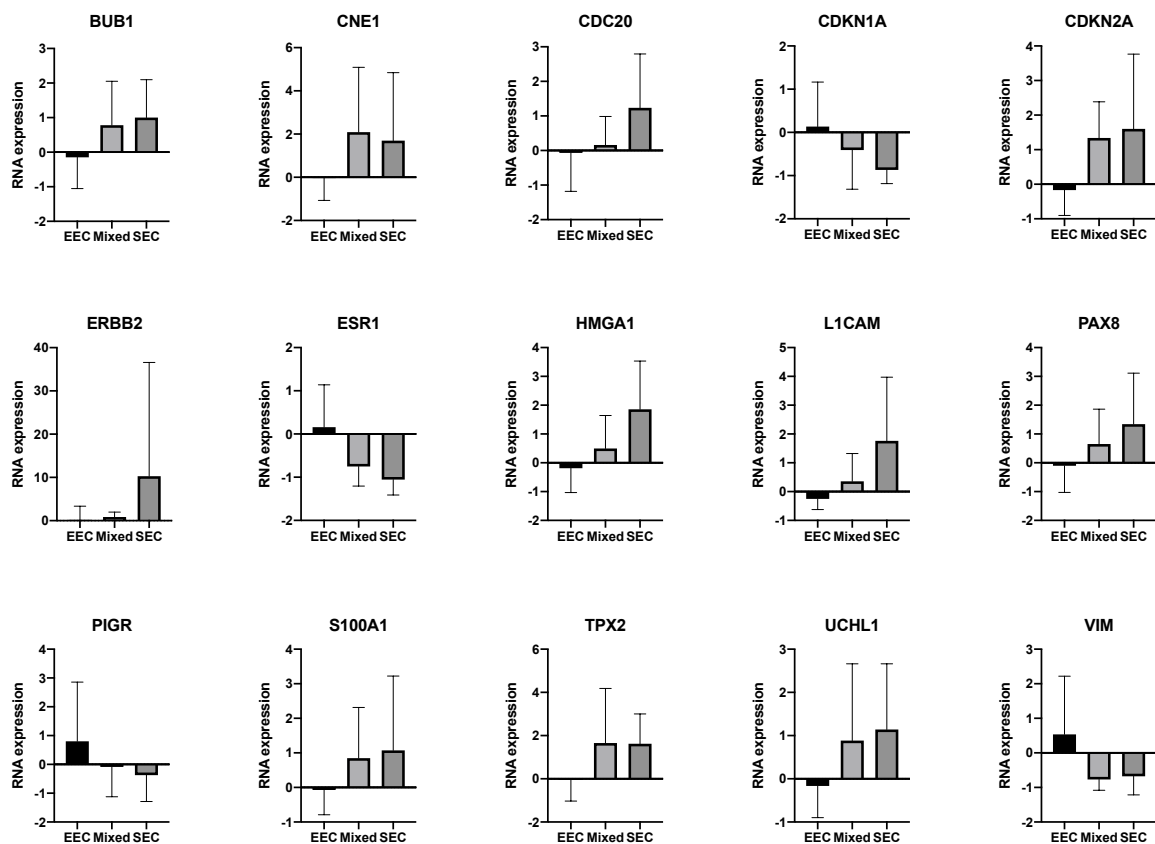
Figure A1. 2 Biomarkers related to FIGO stage and molecular classification. (**A**) Table of the proteins that were found differentially expressed between any of the FIGO stages in any of the tested cohorts. Highlighted in yellow, the specific cohort in which that protein was found to be differentially expressed between FIGO stages. Proteins highlighted in blue are those validated in more than one cohort, and therefore, the ones that we considered as validated biomarkers. (**B**) Table of genes/proteins that were found differentially expressed between any of the molecular subgroups in any of the tested cohorts (in grey). Highlighted in yellow, the specific cohort in which each protein was found to be differentially expressed between subgroups and indicated the specific comparison. The numbers indicate the specific comparison where each biomarker was found differential. Proteins highlighted in blue are those validated in more than one cohort. MC, Molecular classification defined by The Cancer Genome Atlas Network; 1, Comparison between POLE mutated vs. microsatellite instability (MSI); 2, Comparison between POLE mutated vs. copy number low (CN-Low); 3, Comparison between POLE mutated vs. copy number high (CN-High); 4, Comparison between microsatellite instability (MSI) vs. copy number low (CN-Low); 5, Comparison between microsatellite instability (MSI) vs. copy number high (CN-High); 6, copy number low (CN-Low) vs. copy number high (CN-High). T_RNAseq, RNA-Seq data of the cohort of the TCGA; C_RNAseq, RNA-Seq data of the cohort of the CPTAC; C_prot, proteomic data of the cohort of the CPTAC. (**C**) Boxplots showing the expression of the significant biomarkers for FIGO stage in each cohort of patients: TCGA RNA-Seq data, CPTAC RNA-Seq data, and CPTAC proteomic data, respectively. (**D**) Boxplots showing the expression of the significant biomarkers for molecular classification in each cohort of patients: TCGA RNA-Seq data, CPTAC RNA-Seq data, and CPTAC proteomic data, respectively. Literature: literature revision from Coll-de la Rubia E et al. 2020 [57]; T_RNAseq: RNA-

Seq data of the TCGA's cohort; C_RNAseq: RNA-Seq data of the CPTAC's cohort; C_prot: proteomic data of the CPTAC's cohort.

Figure A1. 3 Description of the functions of the proteins described as validated biomarkers in our study, their relation to the epithelial-mesenchymal transition, and their prognostic behavior in other types of cancer. Highlighted in bold are gynecological cancers in which a prognostic association of that specific protein has been described. Source: Uniprot [443], EMTome [444], and The Human Protein Atlas (www.proteinatlas.org).

Gene Name	Function [CC] - Uniprot Function	EMT - EMTome EMT-related proteins	Cancer prognostic summary - The Human Protein Atlas Favorable	Unfavorable
ASRGL1	Has both L-asparaginase and beta-aspartyl peptidase activity.	No	colorectal, renal	liver
ATAD2	Involved in the estrogen-induced cell proliferation and cell cycle progression of breast cancer cells.	No		lung, renal
BUB1	Serine/threonine-protein kinase that performs 2 crucial functions during mitosis: it is essential for spindle-assembly checkpoint signaling and for correct chromosome alignment.	No		liver, pancreatic
CAPG	Calcium-sensitive protein which reversibly blocks the barbed ends of actin filaments but does not sever preformed actin filaments.	Yes	renal	liver
CCNE1	Essential for the control of the cell cycle at the G1/S (start) transition.	No		liver, **ovarian**
CDC20	Required for full ubiquitin ligase activity of the anaphase promoting complex/cyclosome (APC/C).	No		liver, pancreatic, renal
CDKN1A	Binds to and inhibits cyclin-dependent kinase activity, preventing phosphorylation of critical cyclin-dependent kinase substrates and blocking cell cycle progression.	Yes	renal	lung
CDKN2A	Acts as a negative regulator of the proliferation of normal cells by interacting strongly with CDK4 and CDK6.	Yes	head and neck	liver, renal
ERBB2	Protein tyrosine kinase that is part of several cell surface receptor complexes, but that apparently needs a coreceptor for ligand binding. In the nucleus is involved in transcriptional regulation. Involved in the transcription of rRNA genes by RNA Pol I and enhances protein synthesis and cell growth.	Yes	renal	pancreatic
ESR1	Nuclear hormone receptor. The steroid hormones and their receptors are involved in the regulation of eukaryotic gene expression and affect cellular proliferation and differentiation in target tissues. Also mediates membrane-initiated estrogen signaling involving various kinase cascades.	Yes		
FASN	Fatty acid synthetase is a multifunctional enzyme that catalyzes the de novo biosynthesis of long-chain saturated fatty acids starting from acetyl-CoA and malonyl-CoA in the presence of NADPH.	No		**cervical**, renal
HDGF	[Isoform 1]: Acts as a transcriptional repressor. Has mitogenic activity for fibroblasts. Heparin-binding protein.	Yes	**ovarian**	liver
HMGA1	HMG-I/Y bind preferentially to the minor groove of A+T rich regions in double-stranded DNA.They are also involved in the transcription regulation of genes containing, or in close proximity to A+T-rich regions.	Yes		liver, lung, pancreatic
L1CAM	Neural cell adhesion molecule involved in the dynamics of cell adhesion and in the generation of transmembrane signals at tyrosine kinase receptors.	Yes		head and neck, lung, renal
MACC1	Acts as a transcription activator for MET and as a key regulator of HGF-MET signaling. Promotes cell motility, proliferation and hepatocyte growth factor (HGF)-dependent.	Yes	renal	
MCM6	Acts as component of the MCM2-7 complex (MCM complex) which is the putative replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation in eukaryotic cells.	No		liver, melanoma, renal
MCM7	Acts as component of the MCM2-7 complex (MCM complex) which is the putative replicative helicase essential for 'once per cell cycle' DNA replication initiation and elongation in eukaryotic cells. Required for S-phase checkpoint activation upon UV-induced damage.	Yes	**cervical**	liver
MSH2	Component of the post-replicative DNA mismatch repair system (MMR).	No		liver, pancreatic
MSH6	Component of the post-replicative DNA mismatch repair system (MMR).	No		liver, renal
PAX8	Transcription factor for the thyroid-specific expression of the genes exclusively expressed in the thyroid cell type, maintaining the functional differentiation of such cells.	No		
PGR	The steroid hormones and their receptors are involved in the regulation of eukaryotic gene expression and affect cellular proliferation and differentiation in target tissues.	Yes		
PIGR	Mediates selective transcytosis of polymeric IgA and IgM across mucosal epithelial cells. Binds polymeric IgA and IgM at the basolateral surface of epithelial cells.	No	**breast**, renal	
PTK2	Non-receptor protein-tyrosine kinase that plays an essential role in regulating cell migration, adhesion, spreading, reorganization of the actin cytoskeleton, formation and disassembly of focal adhesions and cell protrusions, cell cycle progression, cell proliferation and apoptosis.	Yes		**breast**
S100A1	Small calcium binding protein that plays important roles in several biological processes such as Ca(2+) homeostasis, chondrocyte biology and cardiomyocyte regulation.	No		renal
SCGB2A1	May bind androgens and other steroids, may also bind estramustine, a chemotherapeutic agent used for prostate cancer. May be under transcriptional regulation of steroid hormones.	No	colorectal, renal	
TMEFF2	May be a survival factor for hippocampal and mesencephalic neurons. The shedded form up-regulates cancer cell proliferation, probably by promoting ERK1/2 phosphorylation.	No	Gene product is not prognostic	
TPX2	Spindle assembly factor required for normal assembly of mitotic spindles.	No		liver, lung, pancreatic, renal
TRA2B	Sequence-specific RNA-binding protein which participates in the control of pre-mRNA splicing. Can either activate or suppress exon inclusion.	No	**ovarian**	liver
UCHL1	Ubiquitin-protein hydrolase involved both in the processing of ubiquitin precursors and of ubiquitinated proteins (Probable). This enzyme is a thiol protease that recognizes and hydrolyzes a peptide bond at the C-terminal glycine of ubiquitin. Also binds to free monoubiquitin and may prevent its degradation in lysosomes (By similarity).	Yes		urothelial
VIM	Vimentins are class-III intermediate filaments found in various non-epithelial cells, especially mesenchymal cells. Vimentin is attached to the nucleus, endoplasmic reticulum, and mitochondria, either laterally or terminally.	Yes		renal

ANNEX 2

Table A2. 1 Performance of proteins to discriminate between different EC histological subtypes: endometrioid adenocarcinomas (EEC) and NEEC, including serous adenocarcinoma (SEC) and other NEEC histologies (Others). Statistical results of the proteins showing an *Adjusted P* value <0.05 to discriminate between EEC (n=104) *vs* NEEC (n=45) and/or EEC (n=104) *vs* SEC (n=16) and/or EEC (n=104) *vs* Others (n=29) and/or SEC (n=16) *vs* Others (n=29). Highlighted in grey significant proteins showing an *Adjusted P* value <0.05, *Fold-Change* (FC) >|2| and AUC value > 0.7 for the specific comparison, and AUC values > 0.8 are indicated in bold. Highlighted in pink proteins contained in the combinations.

Gene Name	FC (EEC vs NEEC)	Adjusted P value (EEC vs NEEC)	AUC (EEC vs NEEC)	FC (EEC vs SEC)	Adjusted P value (EEC vs SEC)	AUC (EEC vs SEC)	FC (EEC vs Others)	Adjusted P value (EEC vs Others)	AUC (EEC vs Others)	FC (SEC vs Others)	Adjusted P value (SEC vs Others)	AUC (SEC vs Others)
	9.58	4E-11	**0.85**	3.82	0.00	**0.88**	6.48	0.00	**0.83**	-2.18	0.38	0.65
	-6.00	2E-06	**0.80**	-2.92	0.00	**0.87**	-4.75	0.00	0.77	1.60	0.61	0.59
	-4.28	2E-05	0.78	-2.21	0.01	**0.80**	-3.95	0.00	0.77	1.18	0.88	0.51
	-2.40	3E-04	0.76	-0.84	0.24	0.68	-3.21	0.00	**0.80**	-1.79	0.34	0.65
	-2.24	2E-05	0.76	-1.20	0.01	0.77	-2.19	0.00	0.75	1.05	0.94	0.56
	-3.06	3E-04	0.76	-1.41	0.09	0.76	-3.52	0.00	0.76	-1.33	0.73	0.58
	-2.52	2E-03	0.76	-0.69	0.46	0.63	-3.93	0.00	**0.82**	-2.43	0.28	0.73
	-6.00	3E-04	0.76	-2.95	0.02	**0.81**	-4.65	0.01	0.73	1.66	0.69	0.56
	-3.12	6E-05	0.75	-1.65	0.03	0.78	-3.10	0.00	0.74	1.01	0.99	0.52
	-3.71	5E-05	0.75	-1.56	0.08	0.72	-4.65	0.00	0.76	-1.57	0.56	0.58
	-3.31	2E-03	0.75	-1.17	0.32	0.66	-4.87	0.00	0.79	-2.17	0.39	0.69
	-3.66	7E-04	0.74	-1.67	0.11	0.71	-4.22	0.00	0.76	-1.33	0.78	0.58
	-3.64	7E-04	0.74	-1.31	0.23	0.69	-5.31	0.00	0.76	-2.14	0.39	0.63
	-2.45	5E-04	0.74	-1.54	0.02	**0.84**	-2.07	0.02	0.69	1.40	0.58	0.57
	-2.02	2E-03	0.74	-0.99	0.10	0.73	-2.06	0.00	0.74	-1.03	0.97	0.53
	-2.09	3E-03	0.74	-0.74	0.33	0.64	-2.62	0.00	0.78	-1.57	0.44	0.67
	-5.13	4E-05	0.74	-2.82	0.00	**0.82**	-3.72	0.00	0.70	1.90	0.50	0.61
	-1.84	3E-03	0.73	-0.69	0.24	0.70	-2.11	0.00	0.75	-1.31	0.56	0.59
	-1.67	5E-03	0.73	-0.40	0.49	0.68	-2.11	0.00	0.75	-1.59	0.30	0.58
	2.92	8E-03	0.73	1.49	0.18	0.73	3.03	0.02	0.72	1.08	0.95	0.51
	-2.64	3E-03	0.73	-1.35	0.13	0.72	-2.72	0.01	0.73	-1.07	0.95	0.52
	-3.37	3E-03	0.73	-1.27	0.28	0.69	-4.69	0.00	0.74	-1.94	0.47	0.59
	-2.04	5E-03	0.72	-1.08	0.11	0.74	-1.97	0.02	0.72	1.07	0.93	0.58
	-2.09	3E-03	0.72	-1.01	0.13	0.74	-2.16	0.00	0.71	-1.07	0.92	0.51
	-2.13	3E-03	0.72	-1.12	0.09	0.73	-2.09	0.01	0.72	1.04	0.96	0.55
ST6GALNAC1	-2.27	2E-03	0.72	-1.17	0.10	0.72	-2.28	0.01	0.72	-1.02	0.99	0.51
	-2.26	3E-03	0.72	-1.08	0.14	0.69	-2.40	0.00	0.73	-1.14	0.87	0.57
	-1.68	4E-02	0.72	-0.46	0.56	0.65	-2.07	0.01	0.75	-1.51	0.45	0.59
	-2.05	3E-03	0.72	-0.99	0.14	0.71	-2.12	0.01	0.72	-1.07	0.93	0.51
	-3.52	1E-03	0.72	-2.23	0.02	**0.80**	-2.64	0.03	0.68	1.78	0.52	0.59
	-1.85	1E-02	0.71	-0.59	0.43	0.68	-2.28	0.00	0.73	-1.52	0.44	0.64
	-1.81	3E-02	0.71	-0.42	0.63	0.65	-2.46	0.00	0.74	-1.83	0.34	0.67
	-2.14	4E-03	0.71	-0.88	0.25	0.67	-2.50	0.00	0.73	-1.36	0.63	0.58
	-2.25	4E-03	0.71	-1.06	0.17	0.69	-2.42	0.00	0.72	-1.16	0.85	0.53
	-2.17	9E-03	0.71	-1.04	0.21	0.69	-2.28	0.01	0.72	-1.11	0.90	0.52
	-2.39	2E-03	0.71	-1.60	0.02	0.80	-1.88	0.06	0.67	1.61	0.45	0.59
	-2.51	2E-02	0.71	-0.72	0.56	0.63	-3.83	0.00	0.75	-2.33	0.36	0.67
	-1.80	5E-03	0.71	-0.31	0.66	0.61	-2.60	0.00	0.75	-2.10	0.21	0.69
	-2.06	1E-02	0.70	-0.53	0.56	0.57	-2.94	0.00	0.77	-2.04	0.30	0.69
	-1.99	8E-02	0.70	-0.41	0.74	0.60	-2.97	0.00	0.75	-2.23	0.34	0.67
	-2.53	8E-03	0.70	-1.42	0.13	0.72	-2.40	0.03	0.69	1.12	0.92	0.51
TMEM132A	-2.13	9E-03	0.70	-0.86	0.32	0.65	-2.52	0.00	0.72	-1.39	0.63	0.60
	-1.44	3E-01	0.70	0.15	0.89	0.53	-2.31	0.01	0.78	-2.57	0.21	**0.82**
	-2.16	1E-02	0.69	-0.79	0.41	0.64	-2.71	0.00	0.72	-1.57	0.52	0.58
	-2.35	6E-03	0.69	-0.98	0.29	0.65	-2.81	0.00	0.70	-1.43	0.63	0.56
	-1.67	1E-01	0.69	-0.20	0.85	0.58	-2.41	0.00	0.74	-2.10	0.29	0.69
	3.07	8E-04	0.69	2.41	0.00	**0.80**	1.77	0.23	0.63	-3.00	0.22	0.75
	-1.68	4E-02	0.68	-0.35	0.66	0.58	-2.22	0.00	0.73	-1.74	0.34	0.67
	2.48	3E-04	0.68	2.18	0.00	**0.88**	1.36	0.49	0.58	-3.35	0.08	**0.84**
	-1.74	5E-02	0.67	-0.53	0.52	0.60	-2.09	0.01	0.71	-1.45	0.53	0.64
	-1.57	2E-01	0.67	-0.07	0.95	0.54	-2.33	0.00	0.74	-2.22	0.27	0.71
	-1.61	7E-02	0.67	-0.17	0.85	0.52	-2.31	0.00	0.74	-2.05	0.25	0.73
	-1.40	4E-01	0.67	0.35	0.72	0.51	-2.50	0.00	0.75	-3.19	0.09	0.78
	-1.53	7E-02	0.67	-0.14	0.87	0.58	-2.11	0.00	0.71	-1.91	0.25	0.66
	-1.44	4E-01	0.64	0.23	0.84	0.51	-2.44	0.01	0.72	-2.87	0.21	0.70
	-1.45	3E-01	0.64	0.01	0.99	0.52	-2.11	0.02	0.70	-2.12	0.29	0.67
	2.03	2E-05	0.64	1.97	0.00	**0.87**	1.06	0.87	0.53	-3.70	0.00	**0.86**
	-1.53	3E-01	0.64	-0.01	0.99	0.53	-2.33	0.01	0.72	-2.31	0.26	0.75
	-1.36	3E-01	0.63	0.21	0.79	0.52	-2.13	0.00	0.70	-2.47	0.12	0.73
	-1.48	3E-01	0.63	0.31	0.75	0.57	-2.72	0.00	0.72	-3.36	0.08	0.77
	-2.03	5E-01	0.55	-2.00	0.00	0.77	-1.03	0.96	0.56	3.89	0.02	0.79
	1.94	8E-03	0.67	1.58	0.01	**0.88**	1.26	0.59	0.56	-2.37	0.21	**0.86**
	-1.32	2E-01	0.63	-0.27	0.63	0.59	-1.45	0.07	0.65	-1.20	0.65	0.55

Table A2. 2 Performance of proteins to discriminate between different EC histological grades in endometrioid tumors: grade 1 (G1) and grade 2 (G2) defined as low-grade EEC (n=69) *vs* grade 3 defined as high-grade EEC (n=35). Statistical results of the 14 proteins showing an *Adjusted P* value <0.05 to discriminate between low-grade EEC *vs* high-grade EEC. Highlighted in grey significant proteins showing an *Adjusted P* value <0.05, *Fold-Change* (FC) >|2| and AUC value > 0.7, and AUC values > 0.8 are indicated in bold. Proteins contained in the protein panel are highlighted in pink.

	Histological grade (EEC)	
FC (low *vs* high)	*Adjusted P* value (low vs high)	**AUC** (low vs high)
-4.97	2E-05	**0.82**
-5.22	7E-05	**0.80**
-2.68	3E-04	0.78
-2.52	8E-04	0.78
-3.45	8E-04	0.78
-6.99	8E-04	0.77
-3.33	8E-04	0.77
-3.06	8E-04	0.76
-2.00	7E-04	0.76
-3.18	8E-04	0.76
-3.28	9E-04	0.75
-2.14	9E-03	0.75
-3.61	1E-03	0.75
-3.04	1E-03	0.75
-2.22	1E-02	0.74
-3.09	3E-03	0.74
-3.62	3E-03	0.74
-3.06	5E-03	0.74
-3.59	3E-03	0.74
-2.35	3E-03	0.74
-2.52	7E-03	0.74
-2.78	2E-03	0.73
-2.26	2E-02	0.73
-2.10	6E-03	0.73
4.99	8E-03	0.73
-2.61	5E-03	0.73
-2.80	2E-02	0.73
-2.22	1E-02	0.72
-2.62	8E-03	0.72
-3.28	8E-03	0.72
-3.09	2E-02	0.72
-2.68	9E-03	0.72
3.39	2E-02	0.71
-2.05	3E-02	0.70
-2.13	6E-03	0.70
-2.30	7E-03	0.70
-2.28	2E-02	0.70

Table A2. 3 Performance of proteins to discriminate between recurrent (n=42) and non-recurrent (n=107) endometrial adenocarcinomas, between recurrent endometrioid adenocarcinomas (EEC) (n=23) and non-recurrent EEC (n=81), and/or recurrent NEEC (n=19) and non-recurrent NEEC (n=26). Statistical results of the 25 proteins showing an *Adjusted P* value <0.05 to discriminate between any of the comparisons and/or AUC value > 0.7 for the comparison NEEC (Rec *vs* No-Rec). Highlighted in grey significant proteins showing an *Adjusted P* value <0.05, *Fold-Change* (FC) >|2| and AUC value > 0.7, and AUC values > 0.8 are indicated in bold. Highlighted in pink proteins contained in the combinations.

Gene Name	Recurrence								
	FC (Rec *vs* No-Rec)	*Adjusted P* value (Rec *vs* No-Rec)	AUC (Rec *vs* No-Rec)	FC (EEC: Rec *vs* No-Rec)	*Adjusted P* value (EEC: Rec *vs* No-Rec)	AUC (EEC: Rec *vs* No-Rec)	FC (NEEC: Rec *vs* No-Rec)	*Adjusted P* value (NEEC: Rec *vs* No-Rec)	AUC (NEEC: Rec *vs* No-Rec)
	-2.05	0.00	0.78	-2.37	0.00	**0.81**	-1.37	0.93	0.67
	-1.85	0.00	0.77	-1.54	0.02	0.70	-2.02	0.31	0.77
	-1.76	0.00	0.74	-1.61	0.01	0.71	-1.54	0.93	0.69
	-2.69	0.00	0.73	-1.84	0.03	0.70	-2.74	0.93	0.72
	-3.61	0.00	0.73	-3.76	0.00	0.75	-1.60	0.93	0.59
	-1.52	0.00	0.73	-1.52	0.01	0.74	-1.27	0.93	0.65
	-2.33	0.00	0.73	-2.75	0.00	0.79	-1.73	0.93	0.64
	-1.83	0.01	0.73	-2.01	0.01	0.74	-1.38	0.93	0.66
	-3.05	0.00	0.72	-2.13	0.02	0.67	-3.00	0.93	0.72
	-1.59	0.00	0.72	-1.71	0.00	0.77	-1.16	0.93	0.59
	1.84	0.01	0.72	2.10	0.00	0.76	1.33	0.93	0.60
	-1.50	0.01	0.72	-1.51	0.01	0.72	-1.13	0.93	0.58
	-2.05	0.01	0.71	-2.25	0.01	0.71	-1.30	0.93	0.64
	-2.45	0.01	0.71	-1.98	0.06	0.65	-2.20	0.93	0.74
	-1.61	0.01	0.71	-1.60	0.01	0.71	-1.56	0.93	0.68
	2.12	0.01	0.71	2.87	0.00	0.77	-1.00	1.00	0.51
	-1.81	0.01	0.70	-2.17	0.00	0.75	-1.16	0.94	0.59
	1.61	0.02	0.70	1.90	0.00	0.77	1.17	0.93	0.57
	-1.92	0.00	0.70	-1.49	0.07	0.63	-2.11	0.84	0.74
	-1.68	0.02	0.70	-2.08	0.00	0.77	-1.01	0.99	0.51
	2.25	0.02	0.70	2.67	0.01	0.76	1.61	0.93	0.58
	-2.07	0.02	0.70	-3.16	0.00	0.77	-1.02	0.99	0.53
	2.11	0.01	0.69	2.70	0.00	0.77	1.22	0.93	0.53
	-1.87	0.02	0.68	-2.34	0.00	0.75	-1.18	0.94	0.55
	-1.94	0.03	0.68	-2.35	0.01	0.72	-1.07	0.98	0.53
	-1.54	0.06	0.68	-2.01	0.00	0.72	1.18	0.93	0.55
	-1.70	0.02	0.68	-2.31	0.00	0.76	1.05	0.98	0.51
	-1.77	0.05	0.67	-2.38	0.00	0.76	-1.06	0.98	0.52
	-1.77	0.05	0.67	-2.11	0.01	0.71	-1.27	0.93	0.59
	-1.65	0.02	0.67	-2.09	0.00	0.75	-1.24	0.93	0.59
	1.50	0.07	0.67	2.06	0.00	0.77	-1.25	0.93	0.56
	-1.92	0.03	0.66	-2.42	0.01	0.71	-1.25	0.93	0.56
	1.74	0.05	0.66	2.33	0.01	0.72	-1.06	0.98	0.52
	-1.81	0.01	0.66	-2.39	0.00	0.75	-1.17	0.93	0.53
	-1.46	0.06	0.66	-2.01	0.00	0.75	1.18	0.93	0.54
	-1.58	0.10	0.66	-2.58	0.00	0.78	1.23	0.93	0.50
	-2.63	0.03	0.66	-1.24	0.60	0.55	-4.18	0.58	0.74
	-1.69	0.02	0.66	-2.23	0.00	0.72	-1.26	0.93	0.57
	1.66	0.06	0.66	1.07	0.80	0.51	2.68	0.58	0.79
	1.59	0.07	0.66	2.21	0.01	0.74	-1.08	0.98	0.52
	-1.99	0.03	0.65	-3.10	0.00	0.75	-1.01	0.99	0.51
	1.73	0.06	0.65	2.35	0.01	0.72	-1.02	0.99	0.53
	1.62	0.08	0.65	2.04	0.01	0.75	1.20	0.93	0.53
	2.37	0.06	0.65	1.54	0.25	0.54	2.68	0.93	0.70
	1.80	0.06	0.65	2.38	0.01	0.71	1.09	0.98	0.52
	-1.45	0.10	0.65	-2.33	0.00	**0.80**	1.29	0.93	0.55
	-1.66	0.10	0.65	-2.41	0.01	0.75	-1.08	0.98	0.52
	2.85	0.04	0.65	4.26	0.01	0.71	1.52	0.93	0.54
	1.58	0.07	0.65	2.03	0.01	0.72	-1.01	0.99	0.51
	-1.55	0.06	0.65	-2.02	0.00	0.74	1.01	1.00	0.50

Gene Name	FC (Rec vs No-Rec)	Adjusted P value (Rec vs No-Rec)	AUC (Rec vs No-Rec)	FC (EEC: Rec vs No-Rec)	Adjusted P value (EEC: Rec vs No-Rec)	AUC (EEC: Rec vs No-Rec)	FC (NEEC: Rec vs No-Rec)	Adjusted P value (NEEC: Rec vs No-Rec)	AUC (NEEC: Rec vs No-Rec)
	1.59	0.08	0.65	2.09	0.01	0.71	1.01	0.99	0.53
	-1.47	0.05	0.64	-2.09	0.00	0.79	1.11	0.94	0.55
	1.70	0.09	0.64	2.28	0.01	0.71	1.03	0.98	0.52
	-2.19	0.09	0.64	-1.16	0.74	0.54	-3.01	0.93	0.72
	1.52	0.15	0.64	2.37	0.00	0.75	-1.41	0.93	0.55
	-1.58	0.07	0.64	-2.18	0.00	0.77	-1.05	0.98	0.55
	1.64	0.09	0.64	2.22	0.01	0.72	1.02	0.99	0.51
	1.54	0.09	0.63	2.04	0.01	0.71	-1.11	0.97	0.53
	2.28	0.19	0.63	2.74	0.10	0.61	2.61	0.93	0.71
	-1.48	0.10	0.63	-2.03	0.75		1.05	0.98	0.53
	1.58	0.07	0.63	1.18	0.42	0.58	3.16	0.58	0.75
	1.52	0.10	0.63	2.08	0.01	0.71	-1.16	0.93	0.54
	-1.57	0.16	0.63	-2.58	0.00	0.73	1.68	0.93	0.59
	1.26	0.41	0.63	-1.19	0.51	0.55	2.66	0.58	0.76
	-1.44	0.12	0.63	-2.22	0.00	0.74	1.31	0.93	0.55
	1.63	0.12	0.63	2.54	0.01	0.73	-1.18	0.93	0.53
	1.47	0.13	0.63	2.01	0.01	0.72	-1.09	0.98	0.51
	-1.67	0.12	0.63	1.03	0.94	0.51	-2.79	0.93	0.72
	1.65	0.14	0.62	2.44	0.01	0.71	-1.12	0.98	0.53
	1.46	0.13	0.62	2.06	0.01	0.72	-1.29	0.93	0.58
	1.68	0.17	0.62	2.52	0.01	0.72	-1.24	0.94	0.55
	-1.75	0.10	0.62	-2.76	0.01	0.72	1.26	0.93	0.56
	-1.55	0.13	0.62	-2.55	0.00	0.72	1.31	0.93	0.55
	-1.57	0.15	0.62	1.01	0.97	0.51	-2.10	0.93	0.74
	-1.66	0.07	0.62	-2.33	0.00	0.72	1.15	0.94	0.57
	-1.73	0.16	0.62	-2.93	0.01	0.71	1.53	0.93	0.58
	1.49	0.19	0.62	2.17	0.01	0.72	-1.35	0.93	0.58
	1.62	0.17	0.62	2.65	0.01	0.73	-1.59	0.93	0.62
	-1.38	0.22	0.61	-2.30	0.00	0.75	1.54	0.93	0.59
	1.44	0.15	0.61	2.02	0.01	0.71	-1.20	0.93	0.55
	-1.37	0.18	0.61	-2.02	0.01	0.72	1.45	0.93	0.58
	-1.57	0.11	0.61	-2.63	0.00	0.74	1.35	0.93	0.57
	-1.57	0.13	0.61	-2.19	0.01	0.71	-1.08	0.98	0.50
	-1.46	0.12	0.61	-2.07	0.01	0.73	1.16	0.93	0.57
	-1.61	0.20	0.61	-2.78	0.01	0.74	1.45	0.93	0.59
	1.61	0.21	0.61	2.76	0.01	0.72	-1.59	0.93	0.58
	-1.87	0.17	0.60	1.02	0.96	0.53	-3.32	0.93	0.74
	1.71	0.21	0.60	1.14	0.77	0.50	3.32	0.58	0.74
	1.49	0.24	0.60	2.61	0.01	0.74	-1.66	0.93	0.62
	1.39	0.18	0.60	2.01	0.01	0.72	-1.19	0.93	0.55
	-1.36	0.23	0.60	-2.05	0.01	0.71	1.33	0.93	0.56
	-1.50	0.18	0.60	-2.24	0.01	0.70	-1.28	0.93	0.56
	-1.31	0.25	0.59	-2.04	0.00	0.76	1.55	0.93	0.65
	-1.42	0.20	0.58	-2.27	0.00	0.73	1.40	0.93	0.61
	-1.27	0.44	0.57	-2.03	0.02	0.72	1.32	0.93	0.57
	-1.31	0.27	0.57	-2.05	0.00	0.72	1.30	0.93	0.60
	-1.37	0.34	0.57	-2.15	0.02	0.70	1.46	0.93	0.60
	1.17	0.57	0.55	-1.36	0.16	0.62	2.33	0.93	0.73
	-1.33	0.21	0.55	-2.13	0.00	0.70	1.27	0.93	0.60
	-1.50	0.11	0.54	-2.44	0.00	0.71	1.16	0.93	0.64
	1.10	0.77	0.54	-1.42	0.16	0.59	2.32	0.58	0.79
	-1.29	0.38	0.54	-2.63	0.00	0.73	2.38	0.84	0.72
	-1.26	0.42	0.53	-2.30	0.01	0.72	1.55	0.93	0.69
	-1.11	0.82	0.51	-2.32	0.02	0.68	3.07	0.65	0.76
	1.20	0.12	0.65	1.09	0.44	0.58	1.24	0.93	0.66
	-1.41	0.01	0.71	-1.30	0.06	0.66	-1.40	0.93	0.73
	-1.17	0.05	0.67	-1.28	0.01	0.76	-1.03	0.98	0.56

Table A2. 4 Summary of the 71 proteins showing great performance in discriminating between different EC histological types, grades and/or predict recurrence. The following information is compiled for each protein: Uniprot, Gene, Gene description, biological process, molecular function, subcellular location, secretome location, pathology prognostics in EC, gynecological cancers (including breast, cervix and ovarian cancers), and other cancers, endometrial RNA tissue [nTPM] (Endo.Tissue in the table), and blood concentration of the protein [pg/L] measured in MS (Blood conc. MS in the table). Additionally, column named as significance shows in which parameter the protein was identified (HT, histological type; HG, histological grade; REC, recurrence) when measuring their levels in the uterine fluids of 149 EC patients included in the clinical study.

ANNEX 3

Figure A3. 1 LC-PRM data quality control. Coefficient of variation (CV) of the peptide signals quantified in the triplicates of the pool of uterine fluid samples, M1 samples and M3 samples.
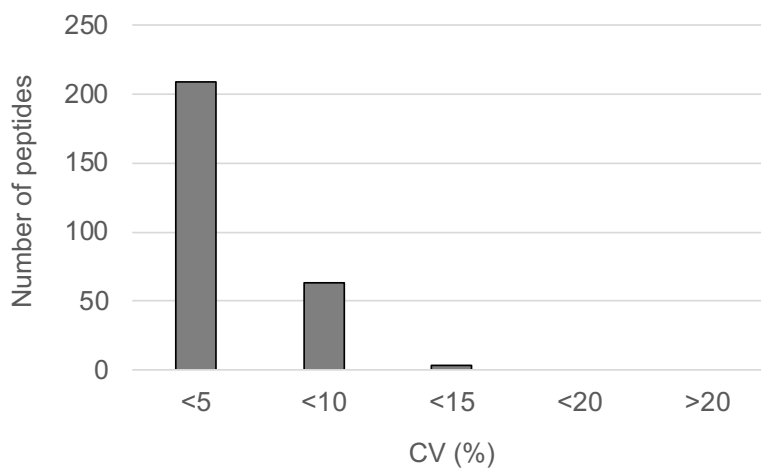
Table A3. 1 Statistical results of the 94 peptides analyzed in the verification study in 41 patients for the three different matrices (raw uterine fluids, cervical samples M1 collected with the Rovers Cervex Brush®, cervical samples M3 collected using the endocervical swab (HC2 DNA collection device Digene)). In bold are highlighted the *Fold Change* values > |2|, *Adjusted P value* < 0.05, and AUC values higher than 0.7. Sn, sensitivity; Sp, specificity.

UTERINE FLUIDS					M1 CERVICAL SAMPLES					M3 CERVICAL SAMPLES				
Fold Change	Adjusted P value	AUC	Sn (%)	Sp (%)	Fold Change	Adjusted P value	AUC	Sn (%)	Sp (%)	Fold Change	Adjusted P value	AUC	Sn (%)	Sp (%)
6.15	0.01	0.77	100	40	0.62	0.48	0.59	50	81	1.26	0.43	0.58	53	61
6.76	0.01	0.77	95	50	0.70	0.48	0.59	44	86	1.12	0.96	0.51	58	61
8.10	0.01	0.80	100	45	0.68	0.66	0.56	50	81	1.35	0.03	0.75	58	83
8.89	0.07	0.68	100	50	0.72	0.83	0.53	44	86	1.11	0.10	0.68	58	78
4.04	0.02	0.76	100	45	2.08	0.13	0.73	94	52	1.58	0.08	0.69	53	94
4.37	0.00	0.83	95	65	1.36	0.31	0.62	50	71	1.39	0.07	0.71	47	89
3.81	0.00	0.81	95	60	1.40	0.27	0.64	50	81	1.40	0.03	0.76	58	89
3.67	0.02	0.75	100	45	2.35	0.21	0.68	56	76	1.41	0.44	0.58	47	83
3.32	0.02	0.74	95	45	2.14	0.25	0.67	56	81	1.63	0.30	0.61	47	83
6.43	0.00	0.90	89	85	1.41	0.25	0.66	61	81	1.25	0.06	0.71	37	94
3.20	0.01	0.81	95	65	1.71	0.13	0.73	72	57	1.62	0.03	0.75	74	61
4.50	0.00	0.87	79	75	0.77	0.62	0.57	44	76	1.83	0.02	0.79	53	89
3.74	0.00	0.84	89	65	0.90	0.89	0.52	33	90	1.82	0.02	0.81	63	89
3.37	0.00	0.89	89	80	1.12	0.30	0.63	61	76	1.69	0.02	0.80	68	83
3.29	0.00	0.89	84	85	1.24	0.25	0.66	67	71	1.42	0.02	0.77	53	89
5.64	0.00	0.89	100	65	2.01	0.11	0.76	83	67	2.25	0.02	0.83	89	72
5.91	0.00	0.89	89	75	2.17	0.09	0.79	83	81	2.53	0.02	0.82	68	89
0.92	0.68	0.55	79	45	1.30	0.78	0.53	94	29	1.33	0.29	0.61	68	44
4.88	0.00	0.85	100	60	2.03	0.11	0.75	72	86	2.13	0.02	0.82	84	78
4.48	0.00	0.81	100	55	2.07	0.21	0.68	61	81	2.32	0.02	0.78	74	94
1.58	0.07	0.69	79	55	1.48	0.27	0.64	33	90	1.17	0.10	0.68	47	83
1.60	0.07	0.69	89	40	1.22	0.64	0.57	50	71	1.10	0.23	0.63	42	83
5.09	0.00	0.95	100	80	1.21	0.21	0.69	56	95	1.69	0.02	0.81	68	89
4.75	0.00	0.94	89	85	1.32	0.15	0.72	61	86	1.56	0.02	0.79	63	89
2.73	0.00	0.89	95	80	0.38	0.74	0.55	61	67	1.10	0.72	0.54	42	78
2.54	0.00	0.88	95	80	0.46	0.77	0.54	61	67	1.00	0.97	0.50	47	78
0.36	0.63	0.56	26	95	0.56	0.77	0.54	50	62	0.53	0.28	0.62	47	83
0.40	0.76	0.53	26	100	0.59	0.97	0.50	44	67	0.61	0.35	0.60	47	83
1.80	0.08	0.68	68	70	1.46	0.30	0.63	56	76	2.19	0.03	0.75	63	83
1.96	0.12	0.66	68	70	1.87	0.15	0.71	56	86	2.12	0.05	0.72	63	78
2.95	0.01	0.78	100	55	1.27	0.27	0.65	56	86	1.29	0.06	0.71	42	89
2.91	0.01	0.77	100	50	1.28	0.27	0.65	56	86	1.37	0.05	0.73	63	78
2.21	0.02	0.75	53	85	1.41	0.27	0.66	72	71	1.54	0.06	0.71	63	72
2.71	0.05	0.71	58	75	1.34	0.28	0.64	72	67	1.99	0.02	0.77	47	89
5.09	0.01	0.79	84	60	1.53	0.27	0.65	56	71	1.33	0.02	0.82	79	72
5.63	0.01	0.78	100	45	1.74	0.26	0.66	50	76	1.23	0.02	0.81	79	72
5.02	0.00	0.85	95	80	1.99	0.13	0.73	44	95	1.99	0.02	0.78	47	94
5.45	0.00	0.83	95	55	1.74	0.21	0.68	39	90	1.87	0.03	0.76	47	94
12.50	0.05	0.71	63	75	5.55	0.11	0.76	67	81	2.61	0.04	0.74	63	83
13.17	0.12	0.66	63	60	6.44	0.11	0.74	78	76	2.40	0.05	0.73	58	83
2.54	0.04	0.72	68	70	0.66	0.27	0.64	67	71	0.65	0.19	0.65	79	67
2.05	0.07	0.69	68	70	0.63	0.28	0.64	72	67	0.67	0.23	0.63	68	72
7.13	0.00	0.96	100	80	1.49	0.11	0.76	78	81	1.57	0.03	0.75	79	61
6.18	0.00	0.94	84	95	1.47	0.11	0.76	78	81	1.50	0.04	0.74	53	83
1.37	0.18	0.64	47	80	2.07	0.11	0.75	56	90	2.03	0.04	0.73	79	56
1.25	0.36	0.60	21	100	2.02	0.13	0.73	56	90	2.10	0.03	0.76	79	56
2.54	0.05	0.71	84	50	1.24	0.30	0.63	61	67	1.32	0.02	0.80	68	83

260

UTERINE FLUIDS					M1 CERVICAL SAMPLES					M3 CERVICAL SAMPLES				
Fold Change	Adjusted P value	AUC	Sn (%)	Sp (%)	Fold Change	Adjusted P value	AUC	Sn (%)	Sp (%)	Fold Change	Adjusted P value	AUC	Sn (%)	Sp (%)
2.61	0.06	0.70	100	35	1.27	0.28	0.64	50	81	1.28	0.02	0.79	68	83
3.67	0.00	0.89	84	80	1.23	0.27	0.65	50	90	1.63	0.02	0.80	53	89
8.42	0.01	0.76	100	60	1.34	0.21	0.68	67	71	1.52	0.05	0.72	58	78
0.73	0.12	0.66	68	65	1.66	0.40	0.61	61	67	0.72	0.09	0.69	53	83
3.55	0.00	0.82	89	65	0.67	0.47	0.59	94	33	0.44	0.27	0.62	63	78
3.67	0.01	0.80	89	60	0.65	0.50	0.59	67	57	0.58	0.42	0.59	74	61
9.23	0.01	0.80	89	60	1.36	0.70	0.55	56	67	1.08	0.35	0.60	58	72
8.57	0.01	0.79	89	60	1.50	0.66	0.56	56	67	0.95	0.59	0.56	53	72
1.75	0.13	0.66	79	60	0.90	0.92	0.51	50	67	1.68	0.08	0.69	37	94
1.78	0.18	0.64	89	45	1.02	0.66	0.57	33	81	1.41	0.19	0.65	63	61
10.48	0.00	0.89	95	75	1.66	0.15	0.71	50	86	1.18	0.26	0.63	21	100
9.57	0.00	0.87	74	90	1.74	0.11	0.75	61	81	1.07	0.32	0.61	21	100
3.26	0.00	0.82	100	55	0.91	0.30	0.63	44	90	1.25	0.34	0.60	21	94
2.74	0.02	0.76	100	45	0.78	0.78	0.54	72	48	1.12	0.57	0.56	21	100
0.04	0.05	0.70	63	90	0.01	0.25	0.66	44	95	0.00	0.05	0.71	47	89
1.97	0.04	0.72	100	45	0.77	0.84	0.52	44	81	1.39	0.07	0.70	63	78
2.00	0.04	0.72	100	35	0.94	0.72	0.55	22	95	1.41	0.08	0.70	63	78
1.13	0.47	0.58	58	70	0.26	0.21	0.68	44	90	0.66	0.29	0.62	79	50
1.15	0.74	0.54	42	75	0.55	0.47	0.59	61	48	0.67	0.31	0.61	63	61
5.41	0.00	0.93	89	90	1.32	0.21	0.68	61	86	1.37	0.02	0.78	84	61
1.93	0.18	0.64	89	45	4.23	0.27	0.65	89	48	1.49	0.21	0.64	26	100
2.83	0.00	0.84	100	50	1.14	0.31	0.62	50	86	1.58	0.02	0.78	53	89
3.06	0.00	0.83	100	55	1.18	0.29	0.63	56	86	1.84	0.02	0.83	79	78
5.11	0.00	0.89	100	65	1.34	0.25	0.67	61	86	1.56	0.02	0.81	68	83
4.58	0.00	0.84	100	60	1.45	0.21	0.69	61	86	1.59	0.02	0.82	68	89
2.83	0.01	0.78	100	55	1.23	0.30	0.63	50	81	1.64	0.03	0.75	58	89
3.28	0.01	0.76	100	55	1.33	0.27	0.65	61	81	1.43	0.08	0.70	42	89
2.48	0.93	0.51	63	55	0.19	0.83	0.53	67	57	0.13	0.83	0.52	53	72
2.41	0.03	0.74	74	65	5.89	0.06	0.83	83	81	3.98	0.04	0.74	84	56
2.67	0.04	0.72	58	85	5.43	0.21	0.66	83	57	4.22	0.05	0.72	63	72
1.63	0.18	0.64	100	30	0.94	0.83	0.53	50	86	1.38	0.09	0.69	58	78
1.62	0.19	0.64	100	30	0.89	0.70	0.55	56	86	1.38	0.11	0.67	58	78
4.42	0.00	0.88	95	65	0.96	0.67	0.56	67	52	1.53	0.05	0.73	63	78
4.06	0.00	0.86	95	65	0.76	0.67	0.56	61	57	1.53	0.07	0.70	63	78
8.45	0.00	0.81	79	65	1.96	0.11	0.75	61	90	1.40	0.59	0.56	42	83
3.96	0.02	0.75	89	60	0.85	0.23	0.67	78	67	0.58	0.77	0.53	42	83
4.65	0.00	0.90	89	75	1.49	0.15	0.71	56	95	2.32	0.02	0.77	47	94
4.63	0.00	0.89	89	75	1.58	0.15	0.71	50	95	2.32	0.02	0.77	58	72
2.05	0.04	0.72	68	70	3.94	0.27	0.65	83	52	6.35	0.02	0.84	79	61
1.88	0.11	0.67	63	70	5.25	0.27	0.64	83	57	5.80	0.02	0.77	84	89
0.88	0.80	0.53	16	100	0.56	0.28	0.64	72	57	0.66	0.08	0.70	58	56
3.32	0.00	0.83	100	60	1.27	0.28	0.64	50	95	1.66	0.02	0.78	84	89
3.40	0.01	0.79	100	55	1.37	0.21	0.69	56	95	1.51	0.02	0.76	63	78
0.99	0.76	0.54	53	55	1.19	0.27	0.65	28	100	1.49	0.08	0.70	79	61
0.97	0.86	0.52	74	50	2.65	0.09	0.80	78	81	2.72	0.02	0.78	100	44
0.90	0.61	0.56	79	45	0.51	0.66	0.56	89	29	0.52	0.10	0.68	47	89
1.12	0.96	0.50	63	55	0.52	0.77	0.54	61	52	0.53	0.30	0.61	63	67

ANNEX 4

Figure A4. 1 Characterization of the proteome of the cervical fluids. (**A**) Venn diagram comparing the proteome obtained to the described EC proteome by CPTAC [298] and proteins related to diagnosis of EC [190]; (**B**) Gene ontological analysis (biological function) of the proteome of the cervical fluids; (**C**) Gene ontological analysis of the significant proteins between non-EC and EC groups of the cervical fluids.

Table A4. 1 Performance of the 75 proteins measured in the verification phase to discriminate between non-EC and EC patients in the cervical fluids. Results obtained in the discovery phase (n=59) for those proteins are also given for the relative quantification (RQ) analysis and the absence/presence (AP) analysis, and the ones also belonging in the pilot study (PS) are indicated. The proteins that were significant for the comparison of the cervical pathology (CP) with EC and CNT in the RQ analysis (*Adjusted P* value < 0.05) are highlighted in grey in the corresponding columns. Bars on the presence/absence analysis represent the number of patients in which each protein was identified. Furthermore, results from the verification phase (n=241, n=128 EC, n=106 non-EC, 7 atypical hyperplasia) are shown for the comparisons EC *vs* CNT or early-stage IA EC *vs* CNT). FC, *Fold Change*; EC, endometrial cancer; CNT, control, or non-EC cases; CP, cervical pathology cases.

Note: The AP analysis of the discovery phase also displays bar charts (PC, CNT, EC – number of patients in which each protein was identified) and grey highlight boxes for the CP vs EC / CP vs CNT comparisons, which are graphical and not reproduced numerically below.

Discovery phase (n=60)					Verification phase (n=241)						
RQ analysis			AP analysis			EC vs CNT			EEC-Stage IA vs CNT		
FC (EC vs CNT)	Adjusted P value (EC vs CNT)	AUC (EC vs CNT)	P value	AUC (EC vs CNT)	Method of Choice	FC (EC vs CNT)	Adjusted P value (EC vs CNT)	AUC (EC vs CNT)	FC (EEC-Stage IA vs CNT)	Adjusted P value (EEC-Stage IA vs CNT)	AUC (EEC-Stage IA vs CNT)
			***	0.79	AP	7.01	2E-26	0.92	7.81	8E-16	0.90
			*	0.71	AP	4.26	3E-16	0.83	3.79	1E-08	0.79
			***	0.81	AP	3.31	2E-15	0.81	2.35	7E-07	0.74
2.52	0.05	0.84			RQ	2.97	2E-14	0.80	2.36	5E-07	0.76
			***	0.79	AP	2.12	2E-14	0.80	1.75	1E-06	0.75
			**	0.76	AP; PS	4.28	1E-14	0.80	4.16	9E-09	0.79
			***	0.71	AP	2.91	6E-14	0.79	2.46	2E-06	0.73
1.48	0.09	0.77			RQ	2.47	9E-14	0.79	1.96	2E-06	0.74
1.52	0.11	0.78			RQ	2.27	2E-13	0.79	1.73	1E-05	0.72
			*	0.80	AP	2.24	9E-13	0.78	2.26	3E-07	0.76
			***	0.69	AP	2.49	1E-13	0.78	3.00	4E-10	0.79
1.75	0.05	0.86			RQ	1.93	6E-11	0.76	1.67	5E-05	0.71
			*	0.79	AP	2.16	8E-11	0.76	1.67	2E-04	0.69
			*	0.71	AP	1.92	8E-11	0.76	1.87	2E-05	0.72
			**	0.79	AP	2.54	8E-11	0.76	1.76	4E-05	0.71
1.46	0.07	0.81			RQ	1.85	2E-10	0.75	1.55	2E-04	0.69
			*	0.80	AP	2.11	2E-10	0.75	1.57	4E-04	0.68
-1.48	0.03	0.80			RQ; PS	-2.04	4E-10	0.75	-1.63	2E-03	0.66
			***	0.72	AP	1.86	4E-10	0.74	1.46	7E-04	0.67
3.11	0.00	0.88			RQ	1.87	5E-10	0.74	1.45	1E-03	0.66
			***	0.87	AP	-2.52	4E-10	0.74	-2.19	4E-05	0.71
			***	0.71	AP	3.90	8E-11	0.74	2.94	1E-04	0.68
1.09	0.11	0.73			RQ	2.14	1E-09	0.74	1.46	2E-03	0.65
-1.95	0.15	0.67			RQ	-2.99	1E-08	0.72	-2.48	7E-04	0.67
-1.87	0.07	0.76			RQ	-2.97	1E-08	0.72	-2.40	7E-04	0.67
			*	0.68	AP	-1.93	1E-08	0.72	-2.46	5E-08	0.78
			*	0.66	AP	-2.50	1E-08	0.72	-2.09	1E-04	0.69
			*	0.67	AP; PS	2.30	2E-08	0.72	2.16	3E-05	0.71
			*	0.63	AP	2.33	1E-08	0.71	1.63	3E-03	0.64
			***	0.74	AP; PS	-1.77	8E-08	0.71	-1.52	4E-03	0.64
			***	0.80	AP	1.66	9E-08	0.71	1.46	2E-03	0.66
			*	0.74	AP	-2.83	1E-07	0.71	-2.40	3E-03	0.65
-1.64	0.12	0.70			RQ	-2.26	1E-07	0.71	-2.08	3E-04	0.68
			***	0.75	AP	1.61	1E-07	0.71	1.28	7E-03	0.64
			*	0.73	AP	1.45	1E-07	0.70	1.38	4E-04	0.68
				0.66	AP	-1.82	2E-07	0.70	-1.81	2E-02	0.62
-1.20	0.03	0.80			RQ	-1.81	8E-07	0.69	-1.59	1E-02	0.63

 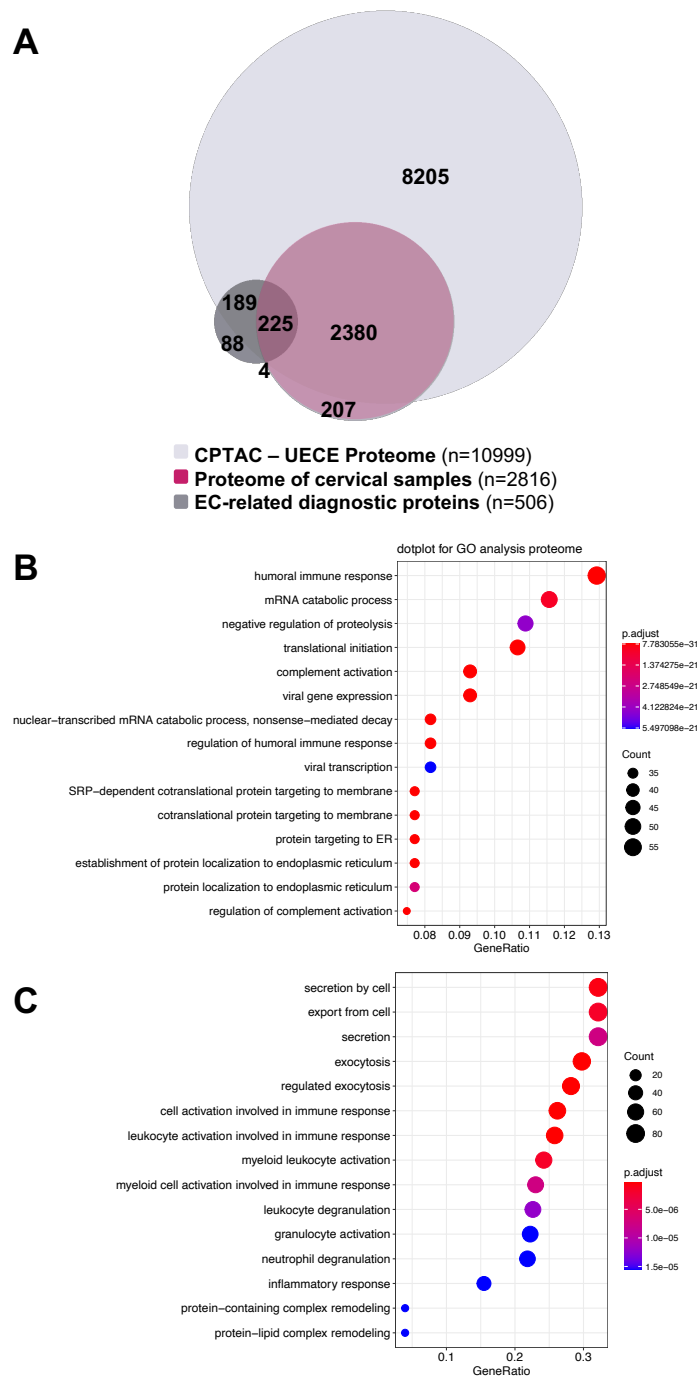

Discovery phase (n=60)									Verification phase (n=241)						
RQ analysis				AP analysis											
FC (EC *vs* CNT)	***Adjusted P* value** (EC *vs* CNT)	**AUC** (EC *vs* CNT)	**CP *vs* EC CP *vs* CNT**	**PC**	**CNT**	**EC**	***P* value**	**AUC** (EC vs CNT)	**Method of Choice**	**FC** (EC *vs* CNT)	***Adjusted P* value** (EC *vs* CNT)	**AUC** (EC *vs* CNT)	**FC** (EEC-Stage IA *vs* CNT)	***Adjusted P* value** (EEC-Stage IA *vs* CNT)	**AUC** (EEC-Stage IA *vs* CNT)
							**	0.76	AP	-1.70	3E-06	0.68	-1.62	3E-03	0.65
2.58	0.07	0.76	▨						RQ	1.85	4E-06	0.68	1.45	3E-02	0.61
-1.97	0.04	0.75	▨						RQ	-2.84	5E-06	0.68	-2.21	4E-02	0.61
							**	0.76	AP	-2.01	7E-06	0.67	-1.82	5E-03	0.64
							***	0.73	AP	-1.85	9E-06	0.67	-1.77	2E-03	0.66
-1.96	0.00	0.86	▨						RQ	-1.78	3E-05	0.66	-1.55	2E-02	0.61
							***	0.84	AP	-1.41	6E-05	0.66	-1.43	2E-03	0.65
							*	0.69	AP	-1.40	6E-05	0.66	-1.44	2E-03	0.66
							*	0.73	AP	-1.73	1E-04	0.65	-1.85	6E-04	0.67
									PS	1.69	2E-04	0.65	1.35	1E-01	0.58
-1.93	0.12	0.68	▨						RQ	-3.12	3E-04	0.64	-2.52	2E-02	0.62
-0.61	0.21	0.65	▨						PS	1.66	8E-04	0.63	1.66	2E-01	0.57
								0.64	PS	1.48	2E-03	0.62	1.63	4E-02	0.61
-1.83	0.11	0.66	▨						RQ	-2.30	1E-04	0.62	-1.94	5E-02	0.59
							*	0.62	AP	-1.53	4E-03	0.61	-1.82	9E-04	0.67
-	-	-							PS	1.07	5E-03	0.61	-1.24	1E-01	0.58
2.02	0.11	0.74							RQ	-1.16	1E-02	0.60	-1.73	3E-01	0.56
							*	0.64	AP	1.09	1E-02	0.60	-1.05	5E-01	0.54
								0.53	PS	1.10	1E-02	0.60	-1.12	8E-01	0.51
							*	0.63	AP	-1.43	1E-02	0.60	-1.73	4E-03	0.64
							*	0.76	AP	-1.47	2E-02	0.59	-1.42	3E-01	0.55
0.19	0.69	0.56							PS	1.13	4E-02	0.58	-1.10	5E-01	0.54
							**	0.72	AP	-1.24	6E-02	0.57	-1.38	3E-02	0.61
							*	0.72	AP	-1.37	7E-02	0.57	-1.26	6E-01	0.53
0.02	0.96	0.53							PS	-1.18	8E-02	0.57	-1.65	6E-01	0.53
-1.91	0.00	0.84	▨						RQ	-1.29	1E-01	0.56	-1.28	5E-01	0.54
							*	0.68	AP	-1.40	2E-01	0.55	-1.35	1E+00	0.50
-0.12	0.85	0.54							PS	-1.07	3E-01	0.54	-1.35	5E-01	0.54
							**	0.71	AP	-1.06	3E-01	0.54	-1.28	8E-01	0.51
								0.52	PS	-1.02	4E-01	0.54	-1.18	9E-01	0.51
							***	0.77	AP	-1.25	4E-01	0.53	-1.25	8E-01	0.51
0.02	0.96	0.53							PS	-1.15	4E-01	0.53	-1.26	2E-01	0.56
-0.59	0.45	0.56							PS	1.15	5E-01	0.53	1.65	6E-02	0.60
							*	0.74	AP	-1.18	5E-01	0.53	-1.31	2E-01	0.56
-1.46	0.11	0.70	▨						RQ	-1.21	6E-01	0.52	-1.06	6E-01	0.53
0.05	0.91	0.53							PS	-1.06	7E-01	0.51	-1.22	6E-01	0.53
								0.58	PS	1.09	9E-01	0.50	-1.15	6E-01	0.53
-0.39	0.33	0.64							PS	-1.10	9E-01	0.50	-1.22	3E-01	0.56

Table A4. 2 Performance of the 75 proteins measured in the verification phase to discriminate between prognostic factors in the cervical fluids of EC patients (n=128). Table showing results obtained from the assessment of the selected proteins to between histological types and/or tumor grades (EEC *vs* NEEC and/or EEC-G1G2 *vs* EEC-G3 and/or EEC-G1 *vs* EEC-G3). FC, *Fold Change.*

Verification phase (n=128)								
Histological type			Histological grade (only EEC)					
FC (EEC vs NEEC)	*Adjusted P* value (EEC vs NEEC)	AUC (EEC vs NEEC)	FC (G1G2 vs G3)	*Adjusted P* value (G1G2 vs G3)	AUC (G1G2 vs G3)	FC (G1 vs G3)	*Adjusted P* value (G1 vs G3)	AUC (G1 vs G3)
-5.02	5E-06	0.85	-3.26	3E-02	0.77	-4.36	5E-07	0.84
2.64	9E-05	0.80	-1.20	2E-01	0.54	1.04	1E-03	0.52
1.78	2E-03	0.75	2.50	3E-03	0.72	3.93	3E-04	0.80
2.25	3E-03	0.74	-1.20	2E-01	0.56	1.02	8E-03	0.52
2.74	1E-02	0.72	1.24	2E-01	0.60	1.48	9E-03	0.63
1.95	1E-02	0.71	1.30	1E-01	0.63	1.61	2E-03	0.70
1.62	2E-02	0.70	1.47	1E-01	0.62	1.81	2E-03	0.68
1.65	2E-02	0.70	1.43	3E-02	0.68	2.23	2E-04	0.77
1.50	2E-02	0.69	-1.57	3E-02	0.67	-1.34	1E-01	0.61
-8.71	1E-02	0.69	-1.26	2E-01	0.53	-1.50	7E-04	0.64
1.75	2E-02	0.69	-1.35	2E-01	0.60	-1.12	4E-02	0.52
1.33	2E-02	0.69	1.37	1E-01	0.63	2.06	1E-04	0.75
1.80	2E-02	0.69	1.60	1E-01	0.61	2.38	3E-04	0.73
2.12	2E-02	0.69	-1.18	2E-01	0.55	1.01	4E-02	0.51
-4.18	2E-02	0.68	-1.27	1E-01	0.61	-1.75	3E-03	0.66
1.60	2E-02	0.68	1.32	1E-01	0.60	1.77	3E-03	0.66
1.54	2E-02	0.68	-1.30	1E-01	0.61	-1.22	1E-01	0.56
-2.11	2E-02	0.68	-1.58	3E-02	0.68	-1.92	6E-04	0.76
1.71	2E-02	0.68	1.62	3E-02	0.67	2.39	6E-04	0.75
-1.84	2E-02	0.68	-1.96	3E-02	0.77	-2.16	5E-04	0.82
-2.77	2E-02	0.68	1.07	2E-01	0.53	-1.10	3E-03	0.64
1.58	2E-02	0.67	-1.59	1E-01	0.61	-1.39	6E-02	0.50
1.48	3E-02	0.67	1.17	2E-01	0.59	1.54	2E-03	0.68
1.47	3E-02	0.67	1.14	1E-01	0.60	1.43	2E-03	0.70
1.75	4E-02	0.66	1.30	1E-01	0.64	1.50	2E-02	0.66
1.68	4E-02	0.65	-1.13	2E-01	0.52	1.05	7E-02	0.55
-2.13	4E-02	0.65	-1.09	3E-01	0.51	-1.42	9E-02	0.53
1.41	5E-02	0.65	1.21	2E-01	0.58	1.58	1E-03	0.68
1.30	5E-02	0.65	-1.14	2E-01	0.53	1.15	8E-03	0.58
-2.00	5E-02	0.65	-3.65	3E-02	0.75	-3.93	4E-03	0.79
1.25	6E-02	0.64	1.30	1E-01	0.60	1.83	1E-03	0.71
1.34	6E-02	0.64	-1.42	1E-01	0.62	-1.26	3E-01	0.57
1.25	7E-02	0.64	-1.55	2E-01	0.59	-1.36	5E-02	0.50
1.34	8E-02	0.63	1.29	3E-02	0.67	1.72	6E-04	0.77
1.30	1E-01	0.63	1.27	1E-01	0.62	1.56	9E-03	0.68
-1.55	1E-01	0.63	-1.16	2E-01	0.51	-1.42	2E-01	0.56
-1.45	1E-01	0.62	-1.40	1E-01	0.63	-1.78	1E-03	0.74
1.53	1E-01	0.62	-1.45	1E-01	0.61	-1.24	3E-01	0.55
1.28	1E-01	0.62	-1.66	2E-01	0.58	-1.56	5E-02	0.54
1.51	1E-01	0.61	-1.05	2E-01	0.51	-1.03	3E-01	0.51
1.73	1E-01	0.61	-1.25	2E-01	0.55	-1.10	2E-01	0.51
-1.46	2E-01	0.60	-2.75	1E-01	0.61	-2.94	7E-01	0.52
1.29	2E-01	0.60	-1.34	1E-01	0.61	1.07	9E-02	0.52
1.23	2E-01	0.59	1.35	2E-01	0.59	1.65	3E-02	0.64
1.62	2E-01	0.59	1.31	1E-01	0.61	1.35	2E-01	0.62
-3.39	2E-01	0.59	-3.47	3E-01	0.58	-6.34	7E-02	0.64
1.37	3E-01	0.59	-1.27	2E-01	0.57	-1.10	4E-01	0.51
1.23	3E-01	0.58	1.07	2E-01	0.54	1.28	5E-02	0.63
-1.17	3E-01	0.58	-1.47	3E-02	0.70	-1.90	2E-02	0.74
1.64	3E-01	0.58	-1.02	2E-01	0.53	1.21	3E-01	0.57
-1.32	3E-01	0.57	-1.68	3E-02	0.75	-1.77	2E-02	0.75
-1.94	4E-01	0.57	1.64	1E-01	0.61	1.23	3E-02	0.65
1.36	4E-01	0.57	-1.27	2E-01	0.53	-1.18	4E-01	0.52
-1.01	4E-01	0.57	-1.50	2E-01	0.55	1.04	1E-01	0.55
1.10	4E-01	0.56	-1.23	2E-01	0.59	-1.22	9E-01	0.61
1.32	5E-01	0.56	-1.81	3E-01	0.50	-1.87	6E-01	0.50
1.22	5E-01	0.56	-1.41	1E-01	0.64	-1.45	6E-01	0.67
-1.49	5E-01	0.56	-1.26	2E-01	0.51	1.09	6E-01	0.56
1.15	5E-01	0.56	-1.24	2E-01	0.54	-1.17	4E-01	0.52
1.11	5E-01	0.55	-1.30	1E-01	0.62	-1.35	9E-01	0.61
-1.19	5E-01	0.55	-1.98	3E-02	0.73	-2.22	5E-02	0.77
-1.13	5E-01	0.55	-1.31	1E-01	0.61	-1.33	3E-01	0.62
1.18	6E-01	0.54	-1.63	1E-01	0.62	-1.11	4E-01	0.55
-1.91	6E-01	0.54	-1.82	3E-01	0.51	1.08	4E-01	0.59
1.09	6E-01	0.54	-1.16	2E-01	0.57	-1.07	8E-01	0.53
1.11	7E-01	0.53	-1.93	3E-02	0.73	-2.06	4E-01	0.74
1.02	7E-01	0.53	1.13	2E-01	0.53	1.26	3E-01	0.56
1.27	8E-01	0.52	-2.74	3E-02	0.79	-2.96	2E-01	0.81
-1.24	7E-01	0.52	-1.02	3E-01	0.50	-1.38	4E-02	0.59
1.19	9E-01	0.51	-1.14	2E-01	0.51	1.10	9E-01	0.55
-1.13	9E-01	0.51	-1.31	1E-01	0.64	-1.44	3E-01	0.64
1.00	9E-01	0.51	-2.60	3E-02	0.77	-3.14	9E-02	0.79
-1.03	9E-01	0.51	-1.55	3E-02	0.66	-1.58	6E-01	0.65
1.09	9E-01	0.51	-1.39	2E-01	0.59	-1.35	9E-01	0.56
-1.14	9E-01	0.51	2.39	2E-01	0.53	3.88	4E-01	0.54

Table A4. 3 Summary of the 75 proteins measured in the verification phase. The following information is compiled for each protein: Uniprot, Gene, Gene description, biological process, molecular function, subcellular location, secretome location, pathology prognostics in EC, gynecological cancers (including breast, cervix and ovarian cancers), and other cancers, endometrial RNA tissue [nTPM] (Endo.Tissue in the table), and blood concentration of the protein [pg/L] measured in MS (Blood conc. MS in the table). Additionally, *Adjusted P values* of the different comparisons resulted from the analysis of the levels of proteins measured in the cervical fluids of 241 patients included in the verification phase are shown: Dx, diagnostic (non-EC *vs* EC); HT, histological type (EEC *vs* NEEC); HG, histological grade (G1G2 *vs* G3).

Figure A4. 2 Predictive biomarkers for EC. (**A**) Table showing proteins able to significantly discriminate (*Adjusted P* value < 0.05) between the different groups of the molecular classification (MSI *vs* CN-Low and/or MSI *vs* CN-High and/or CN-Low *vs* CN-high). (**B**) Violin plots of the distribution of the light/heavy (L/H) ratios of the 3 proteins discriminating across all different molecular subgroups, exempting POLE ultramutated subgroup assessed in 92 EC patients included in the verification phase. (**C**) Violin plots of the L/H rations of the 3 proteins suggesting different distribution when comparing POLE ultramutated *vs* the other molecular subtypes. FC, *Fold Change;* POLE, POLE ultramutated; MSI, microsatellite instable; CN-Low, Copy Number Low; CN-High, Copy Number High.

Verification phase (n=92)					
Molecular classification					
FC (MMRd vs NSMP)	Adjusted P value (MMRd vs NSMP)	FC (MMRd vs p53mut)	Adjusted P value (MMRd vs p53mut)	FC (NSMP vs p53mut)	Adjusted P value (NSMP vs p53mut)
-1.64	0.04	1.33	0.85	2.19	0.02
-1.77	0.00	1.03	0.94	1.82	0.01
-1.71	0.00	-1.05	0.72	1.63	0.05
-1.64	0.01	1.14	0.98	1.86	0.07
-1.87	0.00	-1.16	0.88	1.61	0.07
-1.72	0.01	-1.08	0.80	1.58	0.12
3.04	0.02	-1.32	0.88	-4.00	0.05
-1.78	0.00	-1.14	0.88	1.56	0.07
-1.51	0.04	1.10	0.80	1.66	0.06
3.47	0.00	1.45	0.97	-2.40	0.07
-1.72	0.00	-1.25	0.69	1.37	0.10
-1.96	0.00	-1.32	0.61	1.48	0.13
2.84	0.01	1.40	0.90	-2.03	0.09
-1.52	0.02	-1.19	0.70	1.29	0.13
1.74	0.04	1.03	0.94	-1.70	0.14
-1.36	0.01	-1.07	0.72	1.28	0.10
2.10	0.01	-1.65	0.84	-3.47	0.10
-1.33	0.02	-1.16	0.61	1.15	0.19
-1.61	0.00	-1.24	0.59	1.30	0.13
1.54	0.04	1.02	0.94	-1.51	0.12
-1.57	0.01	-1.11	0.84	1.41	0.12
2.20	0.00	1.44	0.35	-1.53	0.12
2.15	0.00	1.44	0.41	-1.49	0.18
1.64	0.04	1.00	0.99	-1.64	0.18
1.84	0.04	1.60	0.72	-1.15	0.33
1.11	0.41	-3.16	0.18	-3.51	0.01
-1.08	0.32	1.99	0.14	2.14	0.00
-1.32	0.05	1.04	0.89	1.37	0.05

ANNEX 5

Publications

"**Prognostic biomarkers in endometrial cancer: a systematic review and meta-analysis**" *(partially described in Chapter 1).* Coll-de la Rubia E, Martinez-Garcia E, Dittmar G, Gil-Moreno A, Cabrera S, Colas E. **Journal of Clinical Medicine. 2020**. DOI: 10.3390/jcm9061900.

"***In Silico* Approach for Validating and Unveiling New Applications for Prognostic Biomarkers of Endometrial Cancer**" *(partially described in Chapter 1).* Coll-de la Rubia E*, Martinez-Garcia E, Dittmar G, Nazarov PV, Bebia V, Cabrera S, Gil-Moreno A, Colas E*. **Cancers. 2021**. DOI:10.3390/cancers13205052.

"**Protein biomarkers in uterine liquid biopsy for an objective and accurate pre-operative risk assessment in endometrial cancer**" *(described in Chapter 2)*. Coll-de la Rubia E*, Martinez-Garcia E*, Lesur A, Reques A, Bebia V, Cabrera S, Gil-Moreno A, Dittmar G*, Colas E*. Article in preparation.

"**Potential of five different cervical fluids as a source of protein biomarkers for endometrial cancer diagnosis**" *(described in Chapter 3)*. Coll-de la Rubia E*, Martinez-Garcia E*, Lesur A, Cabrera S, Gil-Moreno A, Colas E. Article in preparation.

"**Protein signatures in cervical fluids permit a highly accurate and non-invasive diagnosis of endometrial cancer**" *(described in Chapter 4)*. Coll-de la Rubia E, Martínez-García E, Lesur A, Gómez-Tato A, Casares de Cal MA, Reques A, Sabidó E, Borràs E, Bebia V, Peiró R, Dittmar G, Gil-Moreno A, Cabrera S, Colas E. Article in preparation.

Patents

"**Markers of endometrial cancer**"

Inventors: Coll-de la Rubia E, Gil-Moreno A, Cabrera S, Colas E.
Holding Institution: Vall d´Hebrón Research Institute (Spain) and Luxembourg Institute of Health (Luxembourg).
Priority application number: EP21382680.3
Priority application date: 23rd July 2021
Deposit office: OEPM

Internships

During a total of 13 months, I worked in the Quantitative Biology Unit at the Luxembourg Institute of Health (Luxembourg), under the supervision of Dr. Gunnar Dittmar. Here I performed the sample preparation and LC-MS analysis of the studies described in Chapters 2 to 4.

Awards

Award for the best poster presentation. VI Simposium 2021 Liquid Biopsy. Organized by Universidad Franciso de Vitoria (UFV) Madrid. 25th-28th January 2021. Online.

Award for the best poster presentation. Mass Spectrometry: Application in Clinical Lab (MSACL) 2018 EU – 5th Annual European Congress & Exhibition. Given by The Association of MSACL. Salzburg, Austria. September 2018.

Grants awarded to continue with the development of the molecular diagnostic tools

- ERA Permed (National Institute of Health Carlos III and European Comission). Clinical validation and development of the asset described in Chapter 4. October 2021.
- Programa d'Emprenedoria Avançada de l'AGAUR (Generalitat de Catalunya and La Salle Campus Barcelona). Training of Lean Launchpad, entrepreneurship, and leadership. January-April 2021.
- Wildcard 2020 (EIT Health). Training, networking and mentoring to validate business model. September 2020-November 2020.
- BioEnterpreneur Bootcamp 2018 - BioM (EIT Health). Coaching to design and validate business concept. June 2018-July 2018.

Publications in collaboration

"Advances in endometrial cancer protein biomarkers for the use in clinic". Martínez-Garcia E, Lopez-Gil C, Campoy I, Vallve J, Coll E, Cabrera S, Ramon y Cajal S, Matias-Guiu X, Van Oostrum J, Reventos J, Gil-Moreno A, Colas E. **Expert Review of Proteomics. 2018**. DOI: 10.1080/14789450.2018.1410061.

"Metabolic and Lipidomic Profiling Identifies The Role of the RNA Editing Pathway in Endometrial Carcinogenesis". Altadill T, Dowdy TM, Gill K, Reques A, Menon SS, Moiola CP, Lopez-Gil C, Coll E, Matias-Guiu X, Cabrera S, Garcia A, Reventos J, Byers SW, Gil-Moreno A, Cheema AK, Colas E. **Scientific Reports. 2017**. DOI:10.1038/s41598-017-09169-2.

"Uterine-derived exosome biomarkers for diagnosis and stratification of endometrial cancer patients" Campoy I, Coll-de la Rubia E, Pfammatter S, Moiola CP, Hirschfeld M, Cabrera S, Matias-Guiu X, Sabidó E, Gil-Moreno A, Thibault P, Colas E. Article in preparation.

Acknowledgements

Si no vols ser aturat, no hi ha res que pugui aturar-te. I quan dic res, és res. Ni suposant una impossible pandèmia sanitària a nivell mundial.

Hi ha dies bons, i d'altres molt bons. Els bons són aquells en què res surt bé, en què no arribes, en què plores, t'enfades i dramatitzes, i perds la perspectiva de la realitat. I són bons perquè n'aprens, tornes al lloc on toca i, des d'aleshores, ja saps on no has de tornar. Els molt bons són aquells d'eufòria, els de nervits, els de pessigolles a l'estómac, els de l'alegria quan tens bons resultats, publiques un article, fas un viatge (rere altre), tens una festa, és nadal, o és carnaval i ets de Vilanova.

I fer-ho sempre de la mà d'aquells qui estimes.

Si no quieres ser parado, no hay nada que pueda detenerte. Y cuando digo nada, es nada. Ni suponiendo una imposible pandemia sanitaria a nivel mundial.

Hay días buenos y otros muy buenos. Los buenos son aquellos en los que nada sale bien, en los que no llegas, en los que lloras, te enfadas y dramatizas, y pierdes la perspectiva de la realidad. Y son buenos porque aprendes, vuelves al lugar donde toca y, desde entonces, ya sabes dónde no tienes que volver. Los muy buenos son aquellos de euforia, los de nervios, los de cosquillas en el estómago, los de la alegría cuando tienes buenos resultados, publicas un artículo, haces un viaje (tras otro), tienes una fiesta, es navidad, o es carnaval y eres de Vilanova.

Y hacerlo siempre de la mano de aquellos a quien quieres.

If you don't want to be stopped then nothing can stop you. And when I say nothing, there's nothing, not even a seemingly impossible global health pandemic.

Some days are good, and some are very good. The good days are the ones where nothing goes well, when you don't get there, when you cry, get angry and overreact, and lose sight of reality. And they're good because you learn, you get back on track, and from then, you know where you shouldn't go back. The very good days are those of euphoria, those of nerves, those of butterflies in the stomach, those of joy when you have good results, you publish a paper, you take a trip (after another), you have a party, it's Christmas, or it's carnival and you're from Vilanova.

Doing it always with the help of the ones you love.

Gràcies a tots. Per cadascun dels moment viscuts al llarg d'aquests quatre anys i escaig, en què el que semblava impossible ha esdevingut realitat i, tot i així, ho hem *superat i* hem *fet això possible.*

Gracias a todos. Por cada uno de los momentos vividos a lo largo de estos cuatro años y pico, en los que lo que parecia imposible se ha convertido en realidad y, sin embargo, lo hemos *superado y* hemos *hecho esto posible.*

Thank you all. For each of the moments we have experienced over these (a bit more than) four years, in which the seemingly impossible became a reality. However, we *made it through.*