Approach to personalized medicine in endometrial cancer for the control of recurrent disease

Carlos López Gil

PhD thesis 2022

Approach to personalized medicine in endometrial cancer for the control of recurrent disease

Doctoral thesis presented by

Carlos López Gil

to obtain the degree of

Doctor for the Universitat Autònoma de Barcelona (UAB)

Doctoral thesis performed at Vall d'Hebron Research Institute (VHIR), in the Group of Biomedical Research in Gynecology, under the supervision of

Dr. Eva Colás and Dr. Antonio Gil-Moreno

Doctoral study in Biochemistry, Molecular Biology and Biomedicine, Faculty of Medicine at the UAB, under the supervision of

Dr. Xavier Comella

Universitat Autònoma de Barcelona, September 2022

Dr. Eva Colás (director) Dr. Antonio Gil-Moreno (director)

Dr. Xavier Comella (tutor) Carlos López (student)

"A mi familia"

ABSTRACT

Endometrial cancer (EC) is the second most frequent gynecological cancer worldwide. In general, most patients are diagnosed at early stages of the disease with an overall 5-years survival rate of 95%. However, this percentage drastically decreases when patients are diagnosed with regional or distant metastases to 69% and 17%, respectively. Although surgery as a primary treatment is very effective in those cases of localized tumor, it is inefficient when the tumor has spread the uterus. Additionally, the current adjuvant treatments are not specific for EC and are not well working. Management of EC patients is decided based on the clinicopathological and molecular features of the tumor, which are the bases of the risk stratification system. However, this classification is not perfect, and consequently, there is a percentage of cases that recur independently of the risk group that was identified to the patient. Overall, there is a clinical unmet in the identification of patients that will recur after primary and adjuvant treatment, and in the identification of new treatments for those high-risk cases where chemotherapy, the most aggressive therapy, is not able to inhibit tumor evolution.

This thesis has been divided in two parts. On the one hand, we have elucidated protein biomarkers in primary tissues to predict recurrence for intermediate to high-risk EC. In a discovery phase, we identified 439 proteins in endometrioid EC and 56 proteins in serous EC to potentially predict recurrence using non-targeted proteomics. *In silico* analysis using online databases confirmed the potential use of two of those biomarkers predicting recurrence. Among the identified proteins, we prioritized a list of 58 proteins that were analyzed in primary tissues from an independent cohort of 129 patients using targeted proteomics. Altogether, we verified the use of 4 proteins as potential EC biomarkers predicting recurrence. On the other hand, we have performed a preclinical study using a comprehensive cohort of 15 patient-derived xenograft (PDX) EC models to assess the toxicity and efficacy of SYD985 and niraparib, in monotherapy and in combination, as new treatments for high-risk ECs. Targeted-treatment against HER2 (SYD985) was able to reduce tumor growth in 33% of the PDX models analyzed, whilst its combination with PPAR pathway (niraparib) resulted in an excellent combination for high-risk EC patients, specifically for patients bearing HER2-expressing tumors. Especially, 60% of patients achieved a complete response.

The results of this thesis are expected to improve the detection of high-risk of recurrence EC patients and provide a new alternative for the treatment of these patients. The improved management of high-risk EC patients is expected to decrease mortality and morbidity associated with this disease.

RESUMEN

El cáncer de endometrio (CE) es el segundo cáncer ginecológico más frecuente a nivel mundial. En general, la mayoría de las pacientes son diagnosticadas en etapas tempranas de la enfermedad con una tasa de supervivencia a los 5 años del 95 %. Sin embargo, este porcentaje desciende drásticamente cuando las pacientes son diagnosticadas con metástasis regional hasta el 69%, o con metástasis a distancia hasta un 17% de supervivencia. Aunque la cirugía, el tratamiento primario, es muy eficaz en aquellos casos en los que el tumor está localizado, es ineficaz cuando el tumor se ha diseminado por el útero. Además, los tratamientos adyuvantes actuales no son específicos para el CE y no funcionan bien. El manejo de las pacientes con CE se decide en base a las características clínico-patológicas y moleculares del tumor, que son la base del sistema de estratificación de riesgo. Sin embargo, esta clasificación no es perfecta y, consecuentemente, hay un porcentaje de casos que recurren independientemente del grupo de riesgo que se identifica para la paciente. En general, existe un desajuste clínico en la identificación de pacientes que tendrán una recurrencia después del tratamiento primario y adyuvante, y también hay una falta en la identificación de nuevos tratamientos para aquellos casos de riesgo elevado donde la quimioterapia, la terapia más agresiva, no es capaz de inhibir la evolución tumoral.

Esta tesis se ha dividido en dos partes. Por un lado, hemos identificado biomarcadores proteicos en el tejido primario para predecir la recurrencia en pacientes con CE con riesgo intermedio y elevado. En la fase de descubrimiento hemos identificado 439 proteínas en la histología endometrioide y 56 proteínas en la histología serosa para predecir la recurrencia utilizando proteómica no dirigida. El análisis *in silico* utilizando bases de datos online confirmó el potencial uso de dos de estos biomarcadores para predecir la recurrencia. Entre las proteínas identificadas, generamos un listado de 58 proteínas que fueron analizas en el tejido primaria de una cohorte independiente formada por 129 pacientes usando proteómica dirigida. En total, verificamos el uso de 4 proteínas como biomarcadores potenciales en CE para predecir la recurrencia. Por otro lado, hemos realizado un estudio preclínico utilizando una cohorte de 15 xenoinjertos derivados de pacientes (PDX) con CE para evaluar la toxicidad y eficacia de SYD985 y niraparib, en monoterapia y en combinación, como nuevos tratamientos para pacientes de CE con riesgo elevado. El tratamiento dirigido contra HER2 (SYD985) fue capaz de reducir el crecimiento tumoral en un 33% de los PDX analizados, mientras que la combinación con la vía PPAR (niraparib) resultó en una excelente combinación para las pacientes de CE con riesgo elevado, sobre todo para aquellas pacientes con tumores

que expresan HER2. Específicamente, un 60% de las pacientes consiguió una respuesta completa.

En conclusión, se espera que los resultados de esta tesis mejoren la detección de pacientes con CE con elevado riesgo de recurrencia y proporcionen nuevas alternativas para el tratamiento de estas pacientes. Además, la mejora en el manejo de pacientes con CE de riesgo elevado, se espera que disminuya la mortalidad y la morbilidad asociadas a esta enfermedad.

Palabras clave: cáncer de endometrio, biomarcador proteico, recurrencia, proteómica, estudio preclínico, medicina personalizada, SYD985, niraparib, modelo animal

RESUM

El càncer d'endometri (CE) és el segon càncer ginecològic més freqüent a tot el món. En general, la majoria de les pacients es diagnostiquen en les primeres etapes de la malaltia amb una taxa de supervivència global a 5 anys del 95%. Tanmateix, aquest percentatge disminueix dràsticament quan les pacients són diagnosticades amb metàstasis regionals fins al 69%, o amb metàstasis distants fins el 17% de supervivència. Encara que la cirurgia com a tractament primari és molt eficaç en aquells casos en els que el tumor està localitzat, és ineficaç quan el tumor es troba fora de l'úter. A més, els tractaments adjuvants actuals no són específics pel CE i no funcionen bé. El maneig de les pacients amb CE es decideix en base a les característiques clinicopatològiques i moleculars del tumor, que són la base del sistema d'estratificació del risc. Tot i això, aquesta classificació no és perfecte i, conseqüentment, hi ha un percentatge de casos que pateixen recurrència independentment del grup de risc identificat al pacient. En conjunt, hi ha un desconeixement clínic en la identificació de pacients que patiran una recurrència després del tractament primari i adjuvant, i també hi ha una falta en la identificació de nous tractaments per aquells casos amb risc elevat on la quimioteràpia, la teràpia més agressiva, no és capaç d'inhibir l'evolució del tumor.

Aquesta tesi s'ha dividit en dues parts. D'una banda, hem identificat biomarcadors proteics en el teixit primari per predir la recurrència en pacients de CE amb risc intermig i elevat. En la fase de descobriment hem identificat 439 proteïnes en la histologia endometrioide i 56 proteïnes en la histologia serosa per predir la recurrència a través de proteòmica no dirigida. L'anàlisi *in silico* mitjançant bases de dades online ens ha confirmat l'ús potencial de 2 d'aquests biomarcadors per a predir la recurrència. Entre les proteïnes identificades, hem generat una llista de 58 proteïnes que hem analitzat en el teixit primari en una cohort independent de 129 pacients a través de proteòmica dirigida. En total, hem verificat l'ús de 4 proteïnes com a biomarcadors potencials en CE per a predir la recurrència. D'altra banda, hem realitzat un estudi preclínic utilitzant una cohort de 15 xenografts derivats de pacients (PDX) de CE per avaluar la toxicitat i l'eficàcia de SYD985 i niraparib, en monoteràpia i en combinació, com a nous tractaments per a pacients de CE amb risc elevat. El tractament dirigit contra HER2 (SYD985) ha estat capaç de reduir el creixement tumoral en un 33% dels PDX analitzats, mentre que la combinació amb la via PPAR (niraparib) ha resultat en una excel·lent combinació per les pacients de CE amb risc elevat, sobretot les pacients que expressen HER2 al tumor. Específicament, un 60% de les pacients ha aconseguit una resposta completa.

En conclusió, s'espera que els resultats d'aquesta tesis millorin la detecció de les pacients amb CE amb elevat risc de recurrència i que proporcionin noves alternatives pel tractament d'aquestes pacients. A més, també s'espera que la millora en el maneig de les pacients amb CE de risc elevat, redueixi la mortalitat i la morbiditat associades a aquesta malaltia.

Paraules clau: càncer d'endometri, biomarcador proteic, recurrència, proteòmica, estudi preclínic, medicina personalitzada, SYD985, niraparib, model animal

INDEX

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A
ABC: Ammonium bicarbonate
ACN: Acetonitrile
ADC: Antibody-drug conjugate
AGC: Auto gain control
ANXA1: Annexin 1
AP: Absence-Presence
ASPG: Asparthylglucosaminidase
AUB: Abnormal vaginal bleeding
AUC: Area under the ROC curve
AWD: Alive with disease
AWTD: Alive without disease

B
BER: Base excision repair
BRCA: Breast cancer susceptibility protein

C
CAM: Cell adhesion molecule
CCC: Clear cell carcinoma
CEIC: Ethical committee for clinical investigation
CH$_2$O$_2$: Formic acid
CID: Collision-induced dissociation
CIN: Chromosomal instability
CPTAC: Clinical proteomic tumor analysis consortium
CS: Carcinosarcoma
CTNNB1: Beta-catenin

D
DDA: Data dependent acquisition
DIA: Data independent acquisition
DOD: Dead of disease
DSB: Double-strand break

E
EB: Elution buffer
EC: Endometrial cancer
EEC: Endometrioid endometrial cancer
ELISA: Enzyme-linked immunosorbent assay

EN: Elastic net

ER: Estrogen receptor

ESGO: European Society of Gynaecological Oncology

ESI: Electrospray ionization

ESP: European Society of Pathology

ESTRO: European Society of Radiotherapy & Oncology

F

FA: Formic acid

FC: Fold change

FDA: Food and Drug Administration

FDR: False discovery rate

FF: Fresh-frozen

FFPE: Formalin-fixed paraffin-embedded

FIGO: Federation International of Gynecology and Obstetrics

FRET: Fluorescence resonant energy transfer

G

GBP: Guanine-based purine

GDP: Guanosine diphosphate

GMP: Guanosine 5'-monophosphate

GSEA: Gene set enrichment analysis

GUA: Nucleobase guanine

GUAD: Guanine deaminase

GUO: Nucleoside guanosine

H

HCD: High-energy collision dissociation

HER2: Human epidermal growth factor receptor 2

H&E: Hematoxylin and eosin

HGSOC: High-grade serous ovarian carcinoma

HR: Homologous recombination

I

IHC: Immunohistochemistry

ILC: Invasive lobular carcinoma

K

KGUA: Guanylate kinase

L

LC: Liquid chromatography

LVSI: Lymphovascular space invasion

L1CAM: L1 cell adhesion molecule

M

MALDI: Matrix-assisted laser desorption ionization

MARS: Multiple adaptive regression splines

MELF: Microcystic, elongated and fragmented

MI: Myometrial invasion

mIHC/IF: Multiplex immunohistochemistry/immunofluorescence

MMR: Mismatch repair

MS: Mass spectrometry

MSI: Microsatellite instability

MYH9: Myosin 9

N

NEEC: Non-endometrioid endometrial cancer

NHEJ: Non-homologous end joining

NH$_4$HCO$_3$: Ammonium bicarbonate

NOS: Not otherwise specified

NSMP: No-specific molecular profile

O

OCT: Optimal cutting temperature

OS: Overall survival

OT: Orbitrap

P

PARPi: Poly (ADP-ribose) polymerase inhibitor

PBS: Phosphate-buffered saline

PCA: Principal component analysis

PCR: Polymerase chain reaction

PDTO: Patient-derived tumor organoid

PDX: Patient-derived xenograft

PEG: Polyethylene glycol

PEN: Polyethylene naphthalate

PFS: Progression-free survival

PMVK: Phosphomevalonate kinase

POLE: DNA polymerase epsilon catalytic subunit

ProMisE: Proactive Molecular Risk Classifier for Endometrial Cancer

PRM: Parallel reaction monitoring

PSA: Prostate-specific antigen

PTM: Posttranslational modification

p53abn: P53 abnormal

Q

QA: Quantitative analysis

QIT: Quadrupole ion trap

Q-OT: Quadrupole orbitrap

QqTOF: Quadrupole time of flight

R

RAINBO: Refining adjuvant treatment in endometrial cancer based on molecular profile

RD: Residual disease

RF: Random Forest

ROC curve: Receiver operating characteristic curve

ROS: Reactive oxygen species

RP-HPLC: Reverse phase high performance liquid chromatography

RPPA: Reverse phase protein array

S

SCC: Squamous cell carcinoma

SCNA: Somatic copy number alterations

SCX: Strong cation exchange

SDS: Sodium dodecyl sulfate

SEC: Serous endometrial cancer

SILAC: Stable isotope labeling by amino acids in cell culture

SRM: Selected reaction monitoring

SSB: Single-strand break

T

TCGA: The Cancer Genome Atlas

T-DM1: Ado-trastuzumab emtansine

TIL: Tumor-infiltrating lymphocyte

TOF: Time of flight

Tris-HCl: Tris-hydrochloride

2D-DIGE: Two-dimensional difference gel electrophoresis

2D-PAGE: Two-dimensional polyacrylamide gel electrophoresis

U

UA: Uterine aspirate

UPSC: Uterine papillary serous carcinoma

W

WES: Whole exome sequencing

WGS: Whole genome sequencing

WHO: World Health Organization

X

XIC: Extracted ion chromatogram

INTRODUCTION

1. THE ENDOMETRIUM

1.1. ANATOMY AND FUNCTION OF THE UTERUS

The uterus, also called womb, is the organ of the female reproductive system responsible for the development of the fetus during pregnancy. It is located within the pelvic cavity, between the bladder and the rectum. It is an inverted pear-shaped muscular organ that measures on average 7 cm long, 5 cm wide and 2.5 cm in diameter. The uterus is divided in four major regions (Figure 1):

- The **fundus** is the broad curved upper area of the uterus, opposite from the cervix which connects the fallopian tubes to the uterus.
- The **body** or corpus is the main part of the uterus superior to the cervix.
- The **isthmus** is a thin region that separates the body of the uterus and the cervix.
- The **cervix** is the narrow inferior region which connects the uterus to the vagina.
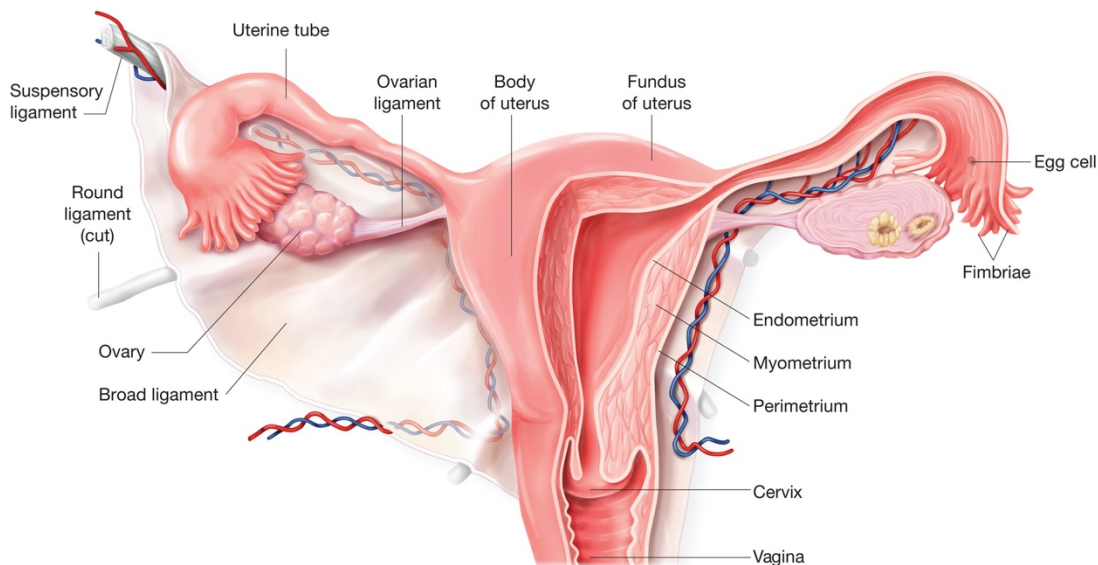
Figura 1. *Anatomy of the uterus (Image from www.anatomy-medicine.com)*

The uterus wall is composed by three tissues layers [1] (Figure 1):

The **perimetrium** is the outer layer that forms the external skin of the uterus. It is a serosal layer that protects the uterus from friction by forming a smooth layer of simple squamous epithelial along its surface and by secreting watery serous fluid to lubricate its surface.

The **myometrium** is the middle layer (1-2 cm thick) composed of smooth muscle cells supported by an underlying connective tissue. It provides to the uterus the capacity to expand during pregnancy and to contract during childbirth.

The **endometrium** is a smooth layer highly vascularized that forms the inner part of the uterus. It consists of a connective tissue where glands and other cell types (like immune cells) are also present, and simple columnar epithelial cells.

1.2. HISTOLOGY OF THE ENDOMETRIUM

The endometrium is composed by three main constitutive elements (Figure 2): i) a simple columnar epithelial cells and secretory cells in the surface; ii) an underlying matrix of connective tissue stroma which contain a supply of blood vessels; and iii) uterine glands formed by invagination of the epithelium that are extended through the stroma. Around 90% of endometrial carcinomas originate from the epithelial glands in the endometrium.
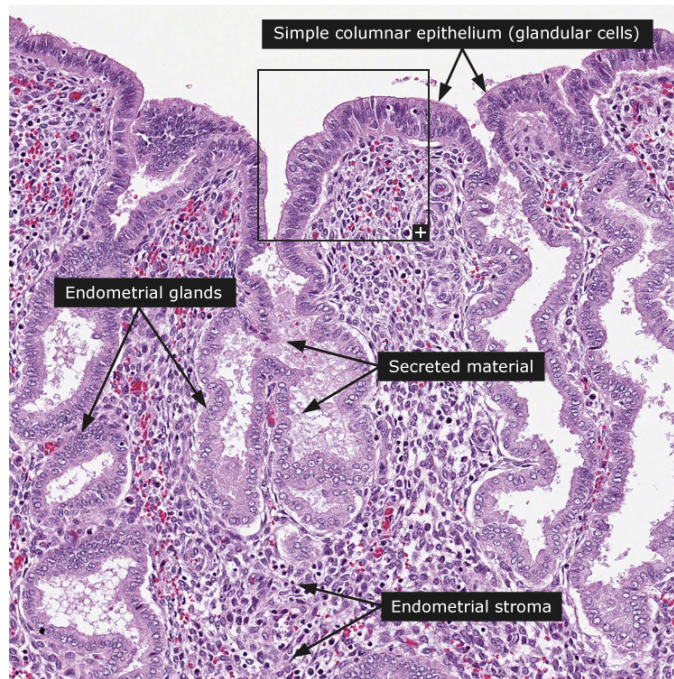
Figura 2. *Endometrial histology (Image from www.proteinatlas.org)*

2. ENDOMETRIAL CANCER

2.1. EPIDEMIOLOGY

The incidence and mortality rate of endometrial cancer differ between the considered "developed" (Europe, North America, Japan, and Australia/New Zealand) and "developing" regions (remaining areas and countries). Endometrial cancer (EC) is the 6th most common cancer worldwide (417,000 new cases that represents 4.5% of all cancers in women) and the 13th cause of death due to cancer (97,000 deaths that represents 2% of all cancers in women)[2]. In contrast, in developed countries, EC is the 4th most common cancer and the 6th cause of death and is the most frequent tumor of the female genital tract[3]. Recent data from the United States estimates that 65,950 new cases will be diagnosed in 2022 (7% of all cancers in women) and 12,550 patients will die because of the disease (4% of all cancers in women) (Figure 3A).

A

Incidence		Estimated new cases in 2022	
	Breast	287,850	31%
	Lung & bronchus	118,830	13%
	Colon & rectum	70,340	8%
	Uterine corpus	**65,950**	**7%**
	Melanoma of the skin	42,600	5%
	Non-Hodgkin lymphoma	36,350	4%
	Thyroid	31,940	3%
	Pancreas	29,240	3%
	Kidney & renal pelvis	28,710	3%
	Leukemia	24,840	3%
	All sites	934,870	100%

Mortality		Estimated deaths in 2022	
	Lung & bronchus	61,360	21%
	Breast	43,250	15%
	Colon & rectum	24,180	8%
	Pancreas	23,860	8%
	Ovary	12,810	4%
	Uterine corpus	**12,550**	**4%**
	Liver & intrahepatic bile duct	10,100	4%
	Leukemia	9,980	3%
	Non-Hodgkin lymphoma	8,550	3%
	Brain & other nervous system	7,570	3%
	All sites	287,270	100%

B Diagnosis

C Survival

Figura 3. 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 for endometrial cancer cases. Data from the United States for 2014-2018. C. Five-year survival rate by stage distribution at diagnosis for endometrial cancer cases. Data from the United States for 2014-2018. Adapted from Siegel et al.[3]

In Europe, EC represented 6.6% of all the cancer cases in women in 2018 [4]. In Spain, EC was the 5th most common cancer diagnosed in women (6,923 new cases in 2021). Unfortunately, its incidence has grown progressively in the last years due to the association of EC with risk factors that are continuously increasing in our population, such as age and obesity [5]. As a consequence, cancer death rate has been rising since the mid-1990s, with an increase of 1% annually [6].

Despite of the elevated incidence, only 19% of cases die because of endometrial cancer. Thanks to the early diagnosis, EC is mostly detected in the initial stages where the tumor is still confined to the uterus and presents a 5-years survival rate of 95%. The 5-year survival drops to 69% or 18% for EC patients diagnosed at an advanced stage of the disease presenting with regional dissemination (18% EC patients) or distant dissemination (9% EC patients), respectively (Figure 3B-C).

2.2. PATIENT PATHWAY

The most common symptom of EC is abnormal uterine bleeding (AUB), which is present in around 90% of patients [7]. The detection of this symptom, especially in post-menopausal patients, should be considered suspicious for EC. Just 10% of post-menopausal women with AUB will be diagnosed with EC [8]. The remaining cases are endometriosis and vaginitis, presence of polyps, hyperplasia, and others [9]. There are other recurrent symptoms of EC: abdominal pain, abdominal distension, changes in bladder functions, alterations in vaginal discharge, and anemia. Nevertheless, these symptoms are less frequent and/or are associated with advanced stages [1].

Therefore, a woman with AUB will initiate the process of EC diagnosis, which first is to perform a **pelvic examination** and a **transvaginal ultrasonography**. The pelvic examination is done by the gynecologist to localize the source of the AUB [10]. In early stages, the results are frequently normal, whereas in advanced stages, there can be changes in the size, shape, or consistency of the uterus and surroundings. The transvaginal ultrasonography is a diagnostic imaging technique to assess the thickness of the endometrium. To determine the presence of EC, it has been set a cut-off of more than 3-5 mm, with sensitivity ranging between 93% and 97% and specificity ranging between 45% and 74% [11,12]. The low specificity is the weakness of this technique because other benign conditions can also be the cause of the thickening. Hence, the definitive diagnosis of EC always requires the histopathological examination of an **endometrial biopsy** (Figure 4).

DIAGNOSIS

Abnormal uterine bleeding

Pelvic examination
&
Transvaginal ultrasonography

Endometrial thickness <5 mm

Endometrial thickness >5 mm or persistent symptoms

EC suspicion discarded

Endometrial biopsy

Positive

Negative

EC confirmation

EC suspicion discarded

Histological type and grade | Molecular classification

TREATMENT

Surgery – Primary treatment

Lymphovascular Space Invasion (LVSI) | Histological type and grade
Myometrial invasion | FIGO stage
Molecular classification | Pelvic and aortic lymph nodes involvement

Adjuvant treatment

Recurrence

Positive

Negative

Novel treatment
&
Clinical follow-up

Clinical follow-up

Figura 4. Diagnosis and therapy/treatment procedure for EC patients

There are two main procedures to obtain an endometrial biopsy: by aspiration using a pipelle or guided by hysteroscopy. The endometrial biopsy by aspiration is the first method of choice because is fast, cost-effective, does not require anesthesia, well tolerated by patients, and it can be performed in the office of the clinician. However, as it is performed blindly, endometrial sampling fails in up to 42% of cases [13,14], and the subjectivity of the pathological examinations causes discrepancies in 11.4% of cases regarding tumor histology and 27% of disagreement in grade staging [15,16]. Consequently,

endometrial biopsy is also obtained by hysteroscopy. In this technique, the hysteroscope is placed into the vagina to visualize the uterine cavity and then, a catheter is used to collect a small piece from the endometrial lesion. The main problems of hysteroscopy are more expensive, sometimes requires anesthesia, and is a more invasive procedure that can generate complications for the patients.

The endometrial biopsy is analyzed in the pathology department to evaluate the presence or absence of tumor cells. In case of EC confirmation, it is further studied to determine the histological type, grade, and if available, the molecular classification is also performed (see section 2.3.). These factors, combined with imaging techniques to measure the spread of the disease, define the extension of the **surgery**, which is the primary treatment for EC. However, high discordance rates have been reported in these clinicopathologic features between the preoperative endometrial biopsy and the final surgical pathology. Different studies have reported that about 10% to 60% were upgrade, downgrade or even a different histology in the final evaluation [17–20]. These discordances may be explained by the small tissue available for the examination in endometrial biopsies, the intratumor heterogeneity [21–23] and the interobserver variability [24,25].

Once the patients undergo surgery, the resected tumor is analyzed to obtain a final diagnosis, which includes the assessment of histological type and grade, FIGO stage, lymphovascular space invasion (LVSI), myometrial invasion, pelvic and aortic lymph nodes involvement, and if available, molecular classification. All these items are the bases of the risk stratification system that determines the **adjuvant treatment** (including no adjuvant treatment, brachytherapy, radiotherapy, chemotherapy, hormonotherapy, palliative treatment, alternative treatments such as immunotherapy, or combination of the aforementioned) [26].

In general, EC is considered a cancer with good prognosis, mainly because of the early diagnosis, but still, approximately 10-20% of tumors recur. The majority (80-90%) of **recurrences** take place within 2-3 years [27,28], and therefore, there is an important clinical follow-up for EC patients during this time after surgery. In cases of recurrence, there is no gold-standard treatment. The few options available are to remove as many macro-metastases as possible by surgery and treat the rest of macro- and micro-metastases with the same adjuvant treatment, and/or with few **novel treatments** that have recently appeared, but those need to be further validated [29–31].

2.3. ENDOMETRIAL CANCER CLASSIFICATION

EC presents a great intertumor heterogeneity with different clinical outcomes. To reflect this diversity, EC is characterized based on clinical features such as histological type, FIGO stage, and on molecular features. This information permits to allocate tumors in two major classification systems: the dualistic model and the molecular classification.

2.3.1. Dualistic model

In 1983, Bokhman et al. [32] proposed the dualistic model of EC based on clinicopathological features. According to this model, EC can be divided in two categories: type I or endometrioid EC (EEC), and type II or non-endometrioid EC (NEEC) (Table 1).

Table 1. *Clinicopathological classification of EC (OS = Overall survival). Adapted from Morice et al.* [33]

Features	Type I or EEC	Type II or NEEC
Age	Pre-perimenopausal	Postmenopausal
Specific subtype	Endometrioid and mucinous	Papillary serous and clear cell
Prevalence	80-90%	10-20%
Diagnosis	Early stages	Advanced stages
Grade	Mostly low (G1 and G2), but also G3	High (G3)
Hormone dependence	Yes	No
TP53 mutation	Rare	90% Serous, 35% Clear cell
PTEN mutation	75-85%	11% Serous, 80% Clear cell
PIK3CA mutation	50-60%	35-45% Serous, 18% Clear cell
ERBB alterations	None	25-30% Serous, 12-16% Clear cell
Evolution	Slow and stable	Aggressive
Prognosis	Good (85% OS at 5 years)	Poor (55% OS at 5 years)
Recurrence	Low rate	High rate

EEC or endometrioid endometrial carcinomas are the most common subtype and account for about 80-90% of ECs. This type of tumor is usually developed in pre- or perimenopausal women. It also includes others histologies such as mucinous, villoglandular, squamous and secretory variants. They express progesterone and estrogen receptors. They are commonly diagnosed at early stages when the tumor is still confined to the uterus and mostly at low grade (grade 1-2). Type I tumor*s* present a

slow and stable evolution. Hence, they used to present a good prognosis with an overall survival of 85% at 5 years and a low rate of recurrence. Regarding molecular alterations, it is characterized by a high mutational frequency and microsatellite instability. The PI3K pathway is altered in the majority cases, being PTEN and PIK3CA the most mutated. Otherwise, TP53 is very rare and there are no ERRB alterations [33–35].

NEEC or non-endometrioid endometrial carcinomas represent 10-20% of ECs an occur normally in postmenopausal women. The papillary serous is the most common histology followed by clear cell and other subtypes less prevalent (undifferentiated, dedifferentiated, carcinosarcomas, mixed cell adenocarcinoma, etc.). Type II tumors are generally diagnosed at advances stages of the disease and all of them are high grade. Therefore, they present an aggressive evolution, a poor prognosis with an overall survival of 55% at 5 years, and a high rate of recurrence. Regarding molecular alterations, type II tumors are characterized to bear TP53 mutation and chromosomal instability. However, each subtype presents different features. Papillary serous used to present mutations in the PIK3CA, whereas clear cell is prompt to show PTEN mutations, and a small proportion of both subtypes also display mutations and/or amplifications of ERBB [33,36,37].

2.3.2. Histological Classification

EC should be classified in different histological types according to the 2020 WHO classification. The main difference from 2014 WHO classification is the incorporation of molecular diagnostic features to avoid discrepancies and to provide a uniform nomenclature [38].

Endometrioid EC (EEC) is the most common type of EC. It is a malignant epithelial neoplasm displaying varying proportion of glandular, papillary, and solid architecture, with the neoplastic cells showing endometrioid differentiation (Figure 5A). It is usually diagnosed as a well-differentiated tumor and low grade, but it also presents several complexities and can be observed in advanced stages and high grade.

Serous EC (SEC) is a carcinoma with nuclear pleomorphism, and typically exhibiting papillary and/or glandular growth patterns (Figure 5B). It is the most common subtype of the NEEC histology, and it represents approximately 10% of EC but accounts up to 40% of EC deaths. SEC patients are usually diagnosed at advanced stages and are associated with a poor prognosis.

Among the rest of non-endometrioid endometrial carcinoma (NEEC), **clear cell carcinomas (CCC)** represent around 2% of all EC. These tumors are characterized by

papillary, tubulocystic, and/or solid architectural patterns (Figure 5C). Other NEECs are **undifferentiated** and **dedifferentiated** carcinomas. Undifferentiated carcinomas are a malignant epithelial neoplasm with no cell lineage differentiation, while dedifferentiated carcinomas are composed of an undifferentiated component and a differentiated component (Figure 5D). **Mixed carcinomas** apply when two histological types are identified in the tumor and one of which is serous or clear cell, even if it is in a small proportion (Figure 5E). There is evidence that any percentage of serous or clear cell confers a poor prognosis. **Carcinosarcomas** are a biphasic tumor composed of high-grade carcinomatous and sarcomatous component (Figure 5F).
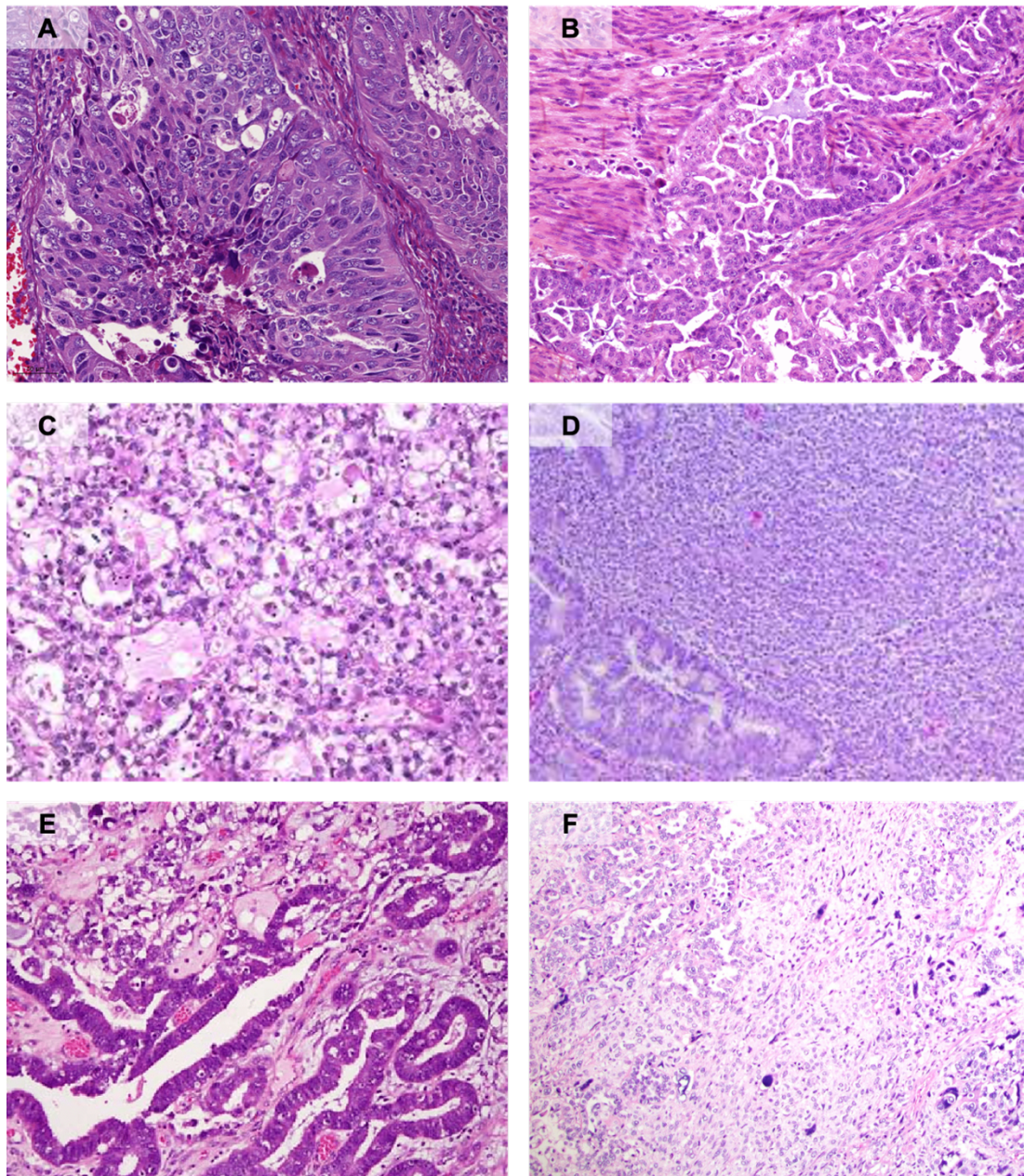
Figura 5. Histology of EC. A. *Endometrioid endometrial carcinoma (EEC).* ***B.*** *Serous endometrial carcinoma (SEC).* ***C.*** *Clear cell carcinoma (CCC).* ***D.*** *Undifferentiated and dedifferentiated carcinoma.* ***E.*** *Mixed clear cell and endometrial carcinoma.* ***F.*** *Carcinosarcoma. Adapted from www.webofpathology.com.*

2.3.3. FIGO Staging

EC is classified in different stages according to the International Federation of Gynecology and Obstetrics (FIGO). This system was updated in 2009 from the previous classification in 1988 to solve problems of reproducibility and to be more accurate [39].

The FIGO staging considers the extent of tumor with the following parameters: tumor size and location, myometrial invasion, lymphovascular space invasion, cervical involvement, and lymph nodes affection. This system classifies EC into four stages, which are summarized in Table 2 and depictured in Figure 6.

Table 2. FIGO staging for EC. Adapted from Pecorelli et al. [40]

Stage I*		Tumor confined to the corpus uteri
	IA	0-50% myometrial invasion
	IB	≥50% myometrial invasion
Stage II*		**Tumor invades cervical stroma, but does not extend beyond the uterus****
Stage III*		**Local and/or regional spread of the uterus**
	IIIA	Tumor invades the serosa of the corpus uteri and/or adnexa[#]
	IIIB	Vaginal and/or parametrial involvement[#]
	IIIC	Metastases to pelvic and/or para-aortic lymph nodes[#]
	IIIC1	Positive pelvic nodes
	IIIC2	Positive para-aortic nodes with or without positive pelvic nodes
Stage IV*		**Tumor invades bladder and/or bowel mucosa, and/or distant metastases**
	IVA	Tumor invasion of bladder and/or bowel mucosa
	IVB	Distant metastases, including intra-abdominal metastases and/or inguinal lymph nodes

*Either grade 1, grade 2, or grade 3
**Endocervical glandular involvement only should be considered as Stage I, and no longer as Stage II
#Positive cytology must be reported separately without changing the stage.

2.3.4. Differentiation grade

The grade describes how much cancer cells resemble healthy cells. When the tumor tissue appears like healthy tissue it is called "differentiated" or "low-grade" (grade 1-2); whereas when the tumor looks very different to healthy tissue it is called "undifferentiated" or "high-grade" (grade 3). While all the NEEC are considered as high-grade, EEC are classified in 3 grades according to the differentiation (Table 3).

Table 3. Grade classification for EEC

Grade	Description	% Differentiation
Grade 1	Well differentiated	<5%
Grade2	Moderately differentiated	6-50%
Grade 3	Poorly differentiated	>50%

2.3.5. Other prognostic factors

There are other independent prognostic factors important to consider for the EC classification: **myometrial invasion**, **lymphovascular space invasion (LVSI)**, and **residual disease**.

Deep myometrial invasion correlates with the risk of metastasis to pelvic and/or paraaortic lymph nodes and is mainly treated as <50% or >50% [41]. LVSI is defined as the presence or absence of tumor cells within endothelial-lined spaces within the uterine wall outside the main tumor. It is a poor prognostic factor and is associated with nodal metastases and recurrences [42]. Finally, residual disease is characterized by the presence or absence of cancer cells after surgery. It is divided into R0 (no residual disease), R1 (microscopic residual disease), and R2 (macroscopic residual disease). Complete resection is significantly associated with improved overall survival [43].
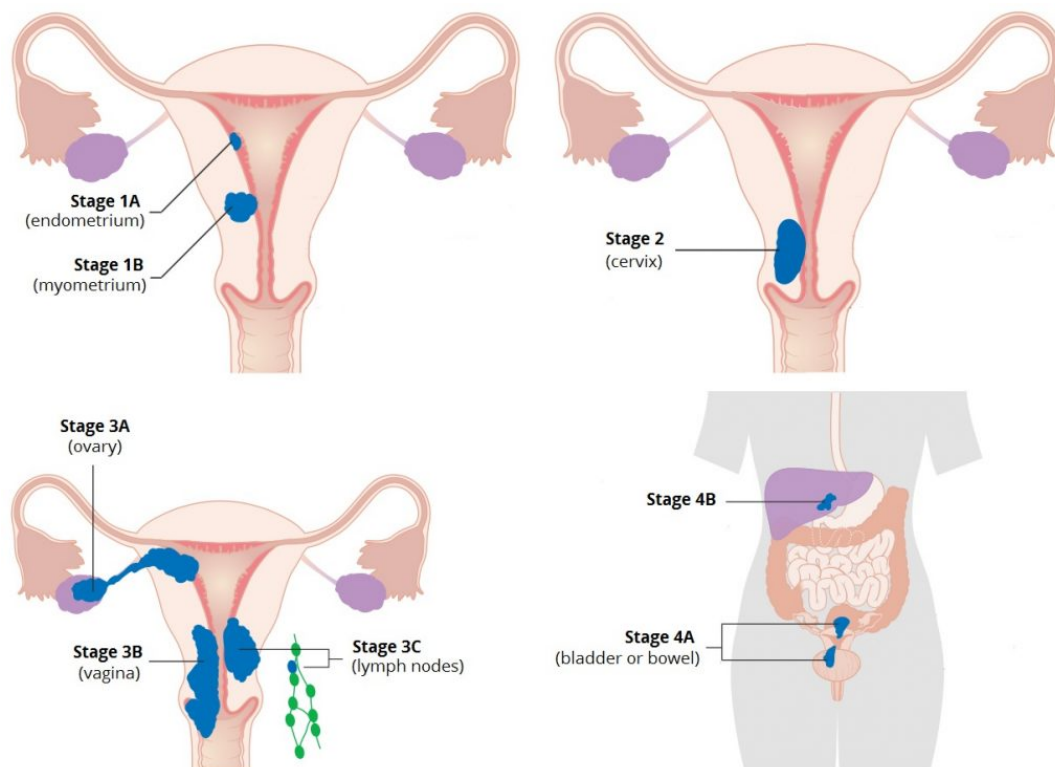
Figura 6. FIGO staging for EC. Adapted from www.teachmeobgyn.com

2.3.6. Molecular Classification

Although the dualistic classification has been broadly used, its prognostic value remains limited because in 30% of cases there is major disagreement among expert pathologists about the subtype diagnosis [24]. Moreover, around 20% of type I ECs relapse, whereas 50% of type II EC do not. This exemplifies the problems of the dualistic model and the necessity of a new classification based on objectives features.

The Cancer Genome Atlas (TCGA) performed and integrated genomic, transcriptomic, and proteomic characterization of endometrial cancer to search a new more accurate classification. This analysis resulted in a new classification with four subgroups: POLE ultramutated, microsatellite instability hypermutated (MSI), low-copy number (microsatellite stable), and high-copy number (serous-like) [44] (Table 4).

Table 4. *Molecular classification of ECs combining information of mutations, microsatellite stability and copy-number alterations. Adapted from Murali et al.* [45]

	POLE ultramutated	MSI hypermutated	Low-copy number	High-copy number
Mutation rate	Very high	High	Low	Low
Microsatellite status	Mixed	Instable	Stable	Stable
Copy-number alterations	Low	Low	Low	High
Genes commonly mutated	POLE (100%) PTEN (94%) FBXW7 (82%) ARID1A (76%) PIK3CA (71%) PIK3R1 (65%) KRAS (53%) ARID5B (47%)	PTEN (88%) PIK3CA (54%) RPL22 (37%) ARID1A (37%) KRAS (35%)	PTEN (77%) PIK3CA (53%) CTNNB1 (52%) ARID1A (42%) PIK3R1 (33%)	TP53 (92%) PIK3CA (47%) PPP2R1A (22%)
Histological type	Endometrioid	Endometrioid	Endometrioid	Endometrioid and Non-Endometrioid
Tumor grade	Grades 1-3	Grades 1-3	Grades 1 and 2	Grade 3
Prognostic	Good	Intermediate	Intermediate	Poor
Surrogate marker	POLE sequencing Ex. 9-11-13-14	IHC MLH1, PMS2, MSH2, MSH6	-	IHC TP53

POLE ultramutated: around 10% of ECs fall in this category and defines a group with an unusual high mutation rate (>100 mut/Mb). POLE is a catalytic component of the DNA polymerase epsilon complex and participates in chromosomal DNA replication. It has 3'-

5' proofreading exonuclease activity that corrects errors arising during DNA replication. This subset of patients is characterized by a unique mutation in the POLE gene, which generates a loss of proofreading function. About 50% of POLE ECs are high-grade endometrioid, they comprise few copy-number alterations and increased frequency of C to A and C to T conversions.

To identify pathogenic somatic mutations in POLE, it has been proposed the sequencing of exons 9, 11, 13, and 14. These exons correspond to the catalytic subunit of POLE. Specifically, it has been defined 11 hotspots as pathogenic mutations, being 5 of them (P286R, S297F, V411L, A456P, and S459F) the most frequent mutations (Table 5) [46].

Table 5. Pathogenic POLE exonuclease domain mutation

Protein Change	Nucleotide Substitution	Exon
P286R	857C>G	9
M295R	884T>G	9
S297F	890C>T	9
F367S	1100T>C	11
D368Y	1102G>T	11
V411L	1231G>T/C	13
L424I	1270C>A	13
P436R	1307C>G	13
M444K	1331T>A	13
A456P	1366G>C	14
S459F	1376C>T	14

Microsatellite instability (MSI) hypermutated or DNA mismatch repair (MMR) deficient: this group represents around 30% of ECs, and it is hypermutated (>10 mut/Mb). This set is 100% microsatellite unstable, which can be sporadic (97%) or germline (3% - Lynch Syndrome). It involves mostly EECs and it is characterized by MLH1 promoter methylation or somatic mutations in MLH1, PMS2, MSH2, and MSH6. These genes participate in the non-homologous end joining (NHEJ) DNA reparation. Therefore, mutations in these genes induce an increment in the mutation rate.

To identify this subgroup, the surrogate marker is the immunohistochemistry (IHQ) of MLH1, PMS2, MSH2, and MSH6. However, IHC for only PMS2 and MSH6 has been demonstrated to be both sensitive and specific [47]. Another possible option is the detection of MSI by polymerase chain reaction (PCR) of BAT25, BAT26, NR21, NR24 MONO27 [48].

Low-copy number (low-CN) or no specific molecular profile (NSMP): most ECs fall in this group (45%) and are characterized to be microsatellite stable (MSS) and do not present POLE or TP53 mutations. Most patients are EEC with low grade (G1 or G2). They present a heterogeneous mutational profile: PTEN (80%), PIK3CA (50%), KRAS (30%). However, there is no surrogate marker for this group.

High-copy number (high-CN) or p53 abnormal (p53abn): they represent 15% of ECs, they are also known as serous-like because the most frequent mutation occurs in TP53 and accumulate the highest number of EC serous histologies. Around 2/3 patients are SECs and 1/3 are EECs high grade. This group is characterized by low mutation rate, and high levels of somatic copy number alterations (SCNA). To identify these patients, the IHC of p53 is a good surrogate marker, and it can also be further validated by PCR to detect mutations in TP53 [49,50].

This novel molecular diagnostic provides more accurate and objective prognostic information. While most ECs can be classified based on a single classifier, a small group of tumors (3-6%) harbor more than one molecular classifier known as multiple-classifier EC [51]. There can be different situations:

- POLE-p53abn or MSI-p53abn: mutational burden and SCNA in these cases are like patients classified as single classification POLE or MSI, respectively. This strongly suggest that TP53 mutations occurs as passenger events without affecting the molecular landscape of the tumor.
- POLE-MSI: POLE with hotspots mutations and MSI have genomic architecture like POLE tumors, suggesting their classification as POLE. Otherwise, patients with non-pathogenic POLE and MSI are more closely resembled to MSI.
- POLE-MSI-p53abn: the presence of p53 subclonal staining suggest that TP53 mutation occurs as a secondary event. However, in these "triple-classifier" is difficult to assess whether POLE or MSI are the driving events in these patients.

More investigation should be performed in a larger number of POLE-MSI and POLE-MSI-p53abn cases to understand the biological behavior of these tumors. From now, considering the results from León-Castillo et al. [46,51] and Vermij et al. [52], multiple-classifier tumors with pathogenic POLE mutations are classified as POLE ultramutated (Figure 7).
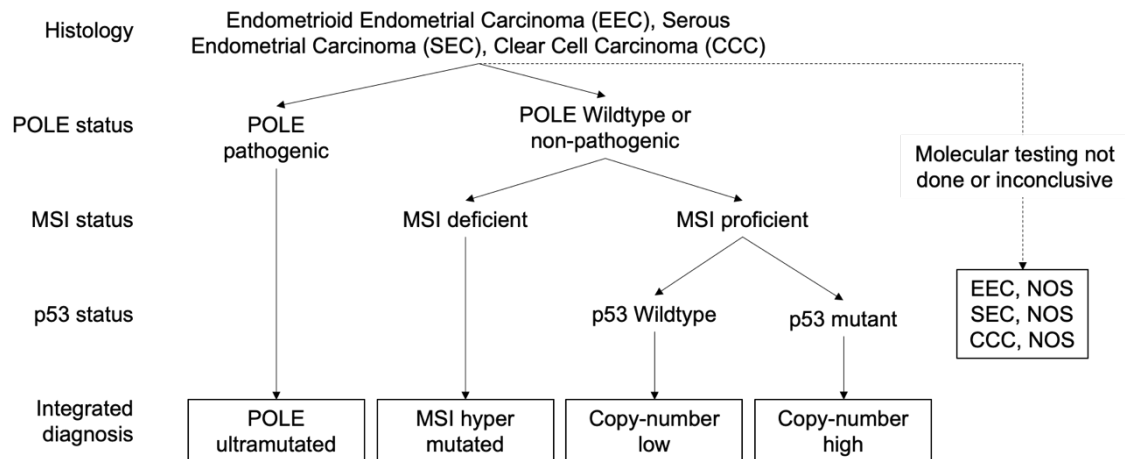
Figura 7. Molecular classification of Endometrial Cancer. POLE pathogenic include: P286R, M295R, S297F, F367S, D368Y, V411L, L424I, P436R, M444K, A456P, and S459F. MSI deficient is caused by the loss of one or more MMR proteins (MLH1, PMS2, MSH2, and MSH6). NOS: Not otherwise specific. Adapted from Vermij et al. [52]

2.4. RISK STRATIFICATION SYSTEM

Generally, EC is detected at early stages and presents a good prognosis. However, some patients recur even when the tumor is still confined to the uterus [53]. Therefore, it is important to establish predictive and prognostic factors to identify this subgroup of patients and select the best surgery as a primary treatment and the best subsequent adjuvant treatment.

EC prognostic factors has been usually divided into uterine and extra-uterine factors. On the one hand, uterine factors include histological type, histological grade, depth of myometrial invasion, presence of atypical endometrial hyperplasia, vascular invasion, cervical affection, DNA ploidy and S-phase fraction, and hormone receptor status. On the other hand, extra-uterine factors include positive peritoneal cytology, adnexal involvement, pelvic and paraaortic lymph node metastasis, and peritoneal metastasis.

The most important prognostic factors are histological type, histological grade, myometrial invasion, and the affection of lymph nodes. Hence, FIGO stage is also an important prognostic factor, as it includes some of these items, and is an important factor for the risk stratification system. Advanced FIGO stages are associated with poor prognosis and higher percentages of recurrences. Moreover, the presence of positive pelvic and/or paraaortic lymph nodes is also an important component. However, lymphadenectomy remains controversially in patients at early stages of the disease because it has not demonstrated benefit for this subgroup [54].

In the last ESGO/ESTRO/ESP guideline, molecular classification was included as a prognostic factor in the risk stratification system [26]. As it is a novel method to classify EC patients and requires specific infrastructure that is not available in all hospitals, the risk groups have been defined including/not including molecular classification status.

All the prognostic factors are evaluated in the tumor in order to classify the risk of recurrence of each EC patient, and this occurs before and after surgery. This classification before surgery guides the extent of the surgical treatment, while after surgery helps to guide the most optimal adjuvant treatment (Table 6). The final staging of the tumor is always determined after surgery.

Table 6. Risk stratification system. Classification based on histology type, histology grade, FIGO staging, LVSI (lymphovascular space invasion), MI (myometrial invasion), RD (residual disease), and molecular classification. Adapted from Concin et al. [26]

	Molecular Classification Unknown	Molecular Classification Known
Low risk	• Stage IA, G1-2 with no or focal LVSI (EEC)	• Stage I-II POLE with no RD (EEC) • Stage IA G1-2 MMRd/NSMP with no or focal LVSI (EEC)
Intermediate risk	• Stage IB G1-2 with no or focal LVSI (EEC) • Stage IA G3 with no or focal LVSI (EEC) • Stage IA without MI (NEEC)	• Stage IB G1-G2 MMRd/NSMP with no or focal LVSI (EEC) • Stage IA G3 MMRd/NSMP with no or focal LVSI (EEC) • Stage IA p53abn without MI (NEEC)
High-intermediate risk	• Stage I with LVSI positive (EEC) • Stage IB G3 (EEC) • Stage II	• Stage I MMRd/NSMP with LVSI positive (EEC) • Stage IB G3 MMRd/NSMP (EEC) • Stage II MMRd/NSMP (EEC)
High risk	• Stage III-IVA with no RD (EEC) • Stage I-IVA with MI and with no RD (NEEC)	• Stage III-IVA MMRd/NSMP with no RD (EEC) • Stage I-IVA p53abn with MI and with no RD (EEC) • Stage I-IVA MMRd/NSMP with MI and with no RD (NEEC)
Advanced metastatic	• Stage III-IVA with RD • Stage IVB	• Stage III-IVA with RD • Stage IVB

2.5. PRIMARY TREATMENT

Surgery is the primary treatment for EC patients. Total hysterectomy (removal of uterus) and bilateral adnexectomy (removal of Fallopian tubes and ovarian) is the standard treatment, and it can be extended according to the staging of the tumor. Lymphadenectomy (removal of lymph nodes) to assess lymphatic dissemination has

been in controversies. Several studies conclude that lymphadenectomy do not show benefit for EC patients at early stages [55,56]. Additionally, it can be associated with morbidities such as lymphedema. Hence, lymphadenectomy is usually recommended for advanced stages of the disease [57,58] Otherwise, sentinel node biopsy is an alternative to lymphadenectomy. Multiple studies confirmed high sensitivity of sentinel lymph node in patients at early stages of the disease [59–61]. The surgical procedure that will be followed according to the risk of recurrence of each patient is described in Table 7.

Table 7. *Surgical treatment based on risk stratification system. Adapted from Concin et al.* [26]

Risk group	Recommended surgical procedure
Low risk	Hysterectomy with bilateral adnexectomy ± sentinel lymph node
Intermediate risk	Hysterectomy with bilateral adnexectomy ± sentinel lymph node
High-intermediate risk	Hysterectomy with bilateral adnexectomy ± bilateral pelvic-paraaortic lymphadenectomy
High risk	Maximal surgery cytoreduction + bilateral pelvic-paraaortic lymphadenectomy
Advanced metastatic	No candidates for radical surgery; systemic therapeutic approach with palliative surgery

2.6. ADJUVANT TREATMENT

The optimal adjuvant treatment is based on the risk stratification system. Most EC patients are diagnosed with low risk and are only treated with surgery. For all the other patients, Figure 8 shows the recommended adjuvant treatment for each risk group.
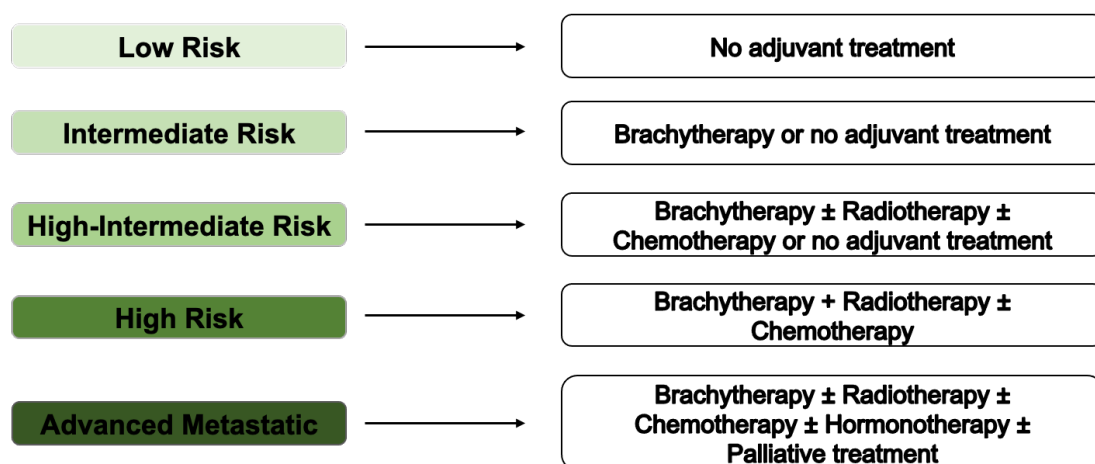
Figura 8. *Adjuvant treatment based on risk stratification system. Adapted from Concin et al.* [26]

Brachytherapy is a type of radiotherapy where the radiation source is placed specifically in the required area, hence, in the pelvic zone. The main advantage of this therapy is the higher radiation that can be used while reducing side-effects to the surrounding areas.

Therefore, it is usually chosen in patients diagnosed at early stages of the disease [62,63]. Otherwise, **radiotherapy** is recommended for EC patients with tumors that have infiltrated the myometrium and present substantial LVSI. It is also indicated in patients who cannot go under surgery as a primary treatment [64,65].

Unlike brachytherapy and radiotherapy, which are localized treatments, **chemotherapy** is a systemic treatment. It is based on intravenous administration of carboplatin and paclitaxel and, as the drugs travel through circulatory system, can generate more side-effects. Consequently, it is normally administered in different cycles (between 4 and 6) and using different drug concentration according to the risk of recurrence [66].

Type I or EEC are characterized to be hormone-dependent tumors; hence, **hormone therapy** is a good treatment for these tumor subtypes. There are different progestational and estrogen agents in the clinics, such as medroxyprogesterone acetate, megestrol acetate, hydroxyprogesterone caproate, and tamoxifen. This kind of treatment is recommended for patients who cannot undergo surgical treatment or patients at advanced stages of the disease that express progesterone and/or estrogen receptors [67,68].

2.7. RECURRENCE

Although the risk stratification system allows clinicians to classify patients according to the risk of recurrence, there is still a limited accuracy to predict recurrence, which is the most important cause of death in EC patients. There is a percentage of cases that recur independently of the risk group that was identified to the patient (Figure 9). Therefore, research efforts should be directed to improve this system in order to identify recurrent EC patients.

Low Risk	→	5% of recurrence
Intermediate Risk	→	5-15% of recurrence
High-Intermediate Risk	→	5-15% of recurrence
High Risk	→	30-40% of recurrence
Advanced Metastatic	→	>40% of recurrence
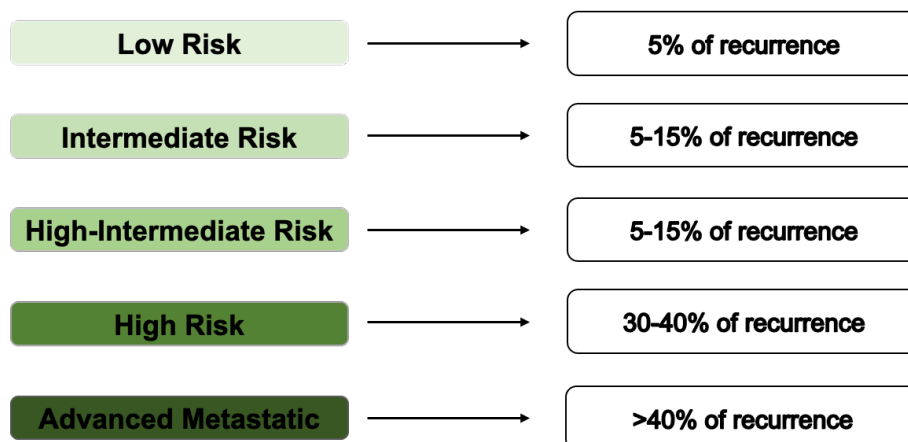
Figura 9. Percentage of recurrence based on risk stratification system

Around 5% of patients classified as **low risk** develop locoregional or distant recurrence. These patients have significantly reduced 5-year survival rates and therapy success rates of only 40% [69]. Several studies identify a subset of low-risk EC patients that behaves as high-risk and should be suspected when having positive molecular markers (such as CTNNB1, TP53, L1CAM, etc.), large tumor size, and/or lymphovascular invasion [70–73]. Low-risk EC patients are treated only with surgery and no adjuvant treatment. Hence, the identification of this subset of patients will allow them to get profit from a most appropriate adjuvant therapy.

Intermediate risk and **high-intermediate risk** are the more controversial group since they can be treated from no adjuvant treatment to a combination of brachytherapy, radiotherapy, and chemotherapy. Five to 15% of patients from these groups will recur. Some studies have identified a differential expression of gene signatures in recurrent patients compared to non-recurrent patients [74,75]. However, the major limitation of these studies is the small sample size analyzed. Moreover, the PORTEC-2 trial has demonstrated a decrease in the percentage of recurrence in those patients treated with radiotherapy compared to vaginal brachytherapy; but it does not reflect an increase in the overall survival [76,77]. Therefore, further research is needed to provide information and facilitate the decision regarding adjuvant treatment for these subsets of patients.

Finally, **high risk** and **advanced metastatic** have 30-40% or more than 40% of risk of recurrence, respectively. Most of EC patients falling in these groups are NEEC (Type II) characterized to be very aggressive or EEC (Type I) with extra-uterine invasion. There are some markers that have been studied in high-risk EC patients, such as L1CAM [78], TGF-β1 [79] or circulating tumor DNA [80]. Nevertheless, many studies have been focused on analyzing the best current adjuvant therapy, which are non-specific for EC, instead of the identification of the subset of EC patients that will get more profit from new specific adjuvant treatments [81–83].

In order to improve the risk stratification system, there exist few advances, which are described in the following sections.

2.7.1. Molecular classification

Clinico-pathological features have been used as standard of care for EC classification. Specifically, histology type, FIGO stage and grade have an impact on recurrence and survival [84]. However, there is still a need to better understand the mechanisms underlying recurrence. At this point, molecular classification has changed the paradigm of EC and

is adding information to further identify, at a limited extent, those EC patients with higher risk of recurrence.

POLE ultramutated: there are only few cases that recur (5%), and the overall survival is almost 100% [85,86]. Around 50% of EC patients are high-grade (G3). Therefore, considering the previous risk stratification system, these patients would have been classified as high risk and treated with aggressive adjuvant therapy. However, molecular classification led us to identify this subset of patients and do not over treat them.

Microsatellite instability (MSI) hypermutated: around 10-15% of these patients recur. The favorable prognosis of this group can also result in the overtreatment of some EC patients [87,88].

Low-copy number (low-CN) or no specific molecular profile (NSMP): this group generates controversies since it gathers the rest of EC patients without a specific molecular profile. The percentage of recurrence vary between 25 to 50% [87,88]. Further studies are needed to understand this subset of patients.

High-copy number (high-CN) or p53 abnormal (p53abn): around 50% of EC patients from this group will recur. This is the worst prognostic group, and it comprises all the NEEC (Type II). Nevertheless, it also includes a small percentage of EEC that can get profit with these new considerations and adapt the adjuvant treatment to not under treat them [87,88].

2.7.2. Other molecular markers

Research on molecular markers to improve prediction of EC recurrence has been ongoing for the last years. In here, we described the main molecular biomarkers identified in multiple studies that seems to increment the ability to classify patients into the risk groups. However, further research is needed to reach clinical implementation.

On the one hand, **L1 cell adhesion molecule (L1CAM)** promotes epithelial to mesenchymal transition and formation of cancer initiating cells in EC [89]. It has been associated with advanced stage, high tumor grade, non-endometrioid histology, nodal involvement, lymphovascular space invasion (LVSI), and distant recurrences. Moreover, it is an independent predictor of poor survival in the endometrioid histology, but not in the non-endometrioid [90–92]. Regarding the new molecular classification, expression of L1CAM is more frequent in p53 abnormal tumors (80%) and is predictive of worse outcome in the low-CN group [93].

One the other hand, **β-catenin (CTNNB1)** is involved in the regulation of cell adhesion and is a key downstream component of the canonical Wnt signaling pathway. Activating mutations of CTNNB1 are likely early drivers in endometrial carcinogenesis and are identified in a significant proportion of low-CN tumors. This cases are associated to a worse prognosis and, therefore, it is proposed that CTNNB1-mutated ECs patients could be the fifth molecular group [94–96].

L1CAM and CTNNB1 are two molecular markers to predict prognosis that are being introduced to the molecular classification. Like them, there are many other molecular markers that can help to predict patient outcome and determine which patients have a higher percentage of recurrence.

2.8. NOVEL TREATMENTS

The current adjuvant treatment is based on brachytherapy, radiotherapy, and chemotherapy, but it is not efficient for metastatic or recurrent patients. Specifically, in the advanced metastatic risk group there is a median survival rate of less than one year, and a median progression-free survival rate of four months [97]. These treatments are not specific for EC patients. Fortunately, in the last years have been appearing several molecules to target specific pathways against EC, such as mTOR inhibitors, PI3K inhibitors, EGFR inhibitors…, but with a limited efficacy [98]. Consequently, additional efforts are needed to improve personalized medicine in EC.

2.8.1. Treatments based on molecular classification

Molecular classification has added extra information to a better risk stratification system, but it has also provided different new strategies for adjuvant treatment.

POLE ultramutated: considering the higher percentage of progression-free survival and overall survival, POLE EC patients do not need any adjuvant treatment. Consequently, it will reduce the comorbidities and costs associated to the treatment and hospitalization [99,100].

Microsatellite instability (MSI) hypermutated: recent studies has questioned the effect of chemotherapy in addition to radiotherapy in this group. There is no extra benefit for these EC patients. This molecular group is characterized by a high mutational load and high levels of tumor-infiltrating lymphocytes (TILs). Therefore, immune checkpoint blockade therapy is a recommended approach. There are many inhibitors with high

response rates in other cancers such as pembrolizumab, and now, they are also being tested in EC [88,101].

Low-copy number (low-CN) or no specific molecular profile (NSMP): this is the most heterogeneous molecular subgroup and there is not a specific treatment. However, as CTNNB1 seems to be a marker of worse prognosis, the Wnt-pathway inhibition could be a good approach for these EC patients. Another possible option is to explore hormonal therapy since most of EC classified in NSMP are low grade, and thus, are probably estrogen and progesterone receptor positive [101].

High-copy number (high-CN) or p53 abnormal (p53abn): it is the most aggressive and lethal molecular subtype. Consequently, there are many efforts to understand the mechanisms underlying this subgroup and to find better treatment options. Nowadays, there are two main strategies based on targeting the homologous recombination deficiency (PARP inhibitors) and the human epidermal growth factor 2 (HER2) [101].

In order to determine the efficacy of specific treatments based on the molecular classification, the TransPORTEC consortium is developing the RAINBO Umbrella Trial (Figure 10). This study is an international prospective study in which patients classified in the POLE group do not receive adjuvant treatment and are only observed after surgery; patients with MSI receive a combination of radiotherapy and immune checkpoint inhibitor (dostarlimab); patients with NSMP receive a combination of radiotherapy and hormonal therapy; and patients with p53abn are treated with a combination of chemoradiation and PARP inhibitors (niraparib) [101]. This is the first trial to implement target therapy based on the molecular classification.
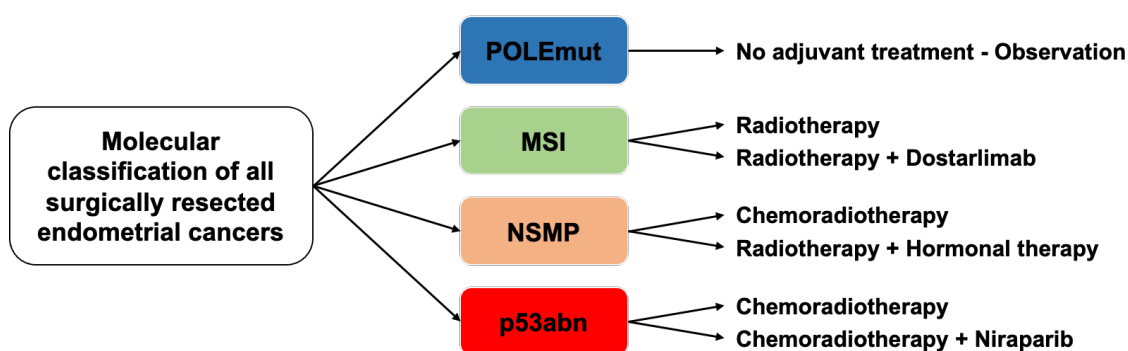
Figura 10. *The planned treatment arms for the TransPORTEC RAINBO program of clinical trials (RAINBO, refining adjuvant treatment in endometrial cancer based on molecular profile). Adapted from Jamieson et al.* [101]

2.8.2. Treatments based on homologous recombination (Niraparib)

Poly (ADP-ribose) polymerase (PARP) inhibitors (PARPi) are a novel class of anti-cancer therapies which have shown to be effective in the treatment of homologous recombination (HR) deficient tumors [102]. PARPi mechanism of action is based on hampering the repair of single-strand break (SSB), inducing the conversion of SSB to double-strand break (DSB). Consequently, cells lacking a compensatory repair system due to tumor suppressor BRCA1 and/or BRCA2 gene alteration, result in the accumulation of DBS, promotion of replication fork collapse, and ultimately leading to synthetic lethality cell death. Otherwise, PARPi can also bind and trap the PARP1 enzyme on the chromatin, creating a lesion in the DNA, and finally inducing cell death [103] (Figure 11).
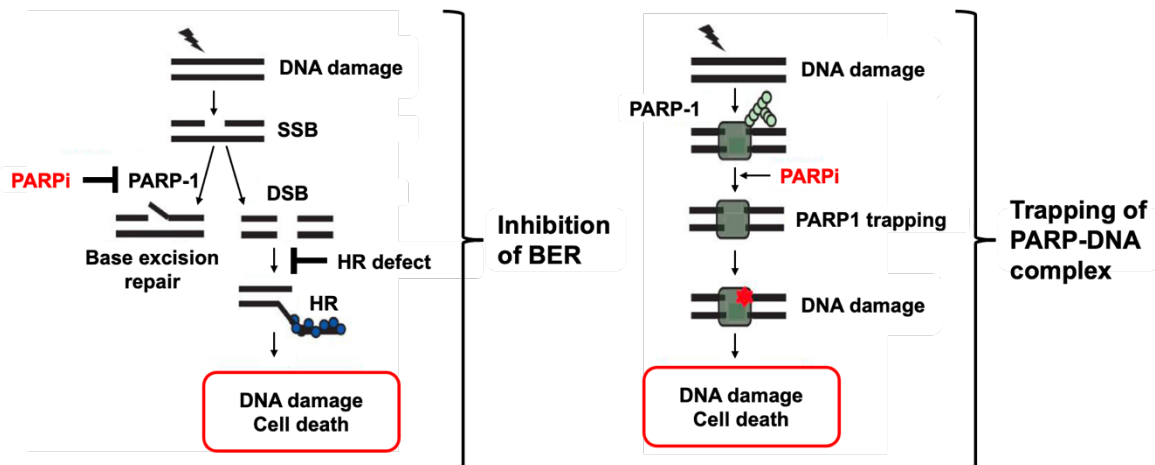
Figura 11. *Mechanism of action of PARP inhibitors (BER, Base Excision Repair). Adapted from D'Andrea et al.* [103]

PARPi have shown promising results in both experimental and clinical trials in ovarian, breast, prostate, and pancreatic cancer [104–106]. Moreover, some studies have also established the significant antitumor benefits of utilizing PARPi in combination with other anti-cancer agents to induce tumor regression [107].

Specifically, niraparib, an inhibitor of poly (ADP-ribose) polymerase (PARP) enzymes, PARP-1 and PARP-2, has shown *in vitro* cytotoxicity by increasing formation of PARP-DNA complexes resulting in DNA damage, apoptosis and cell death. Surprisingly, niraparib-induced cytotoxicity was observed in tumor cell lines with or without deficiencies in BRCA1/2 [108,109]. They conclude that cells that show a response to PARPi may also have a defect in homologous recombination repair genes, not only in BRCA1/2. In addition, niraparib decrease tumor growth in mouse xenograft models of human cancer cell lines with deficiencies in BRCA1/2 and in human patient-derived xenograft

tumor models with homologous recombination deficiency that had either mutated or wild type BRCA1/2 [110,111].

Only few studies have been conducted to explore the correlation between the BRCA mutational status and EC [112]. Some of this studies suggest that BRCA mutated patients have a higher risk to develop uterine papillary serous carcinoma (UPSC) [113,114]. It is now clear that other genetic alterations may play important role in the homologous repair system: ARID1A, ATM, RAD51, etc. For instance, ARID1A is mutated in around 40% of EEC, 20-35% of carcinosarcomas, and it is rare in serous carcinoma [115]. Moreover, PTEN is extremely mutated in EC, and it has been demonstrated that PTEN contributes to maintain the stability of RAD51, a protein involved in repair of DNA double strand breaks [116]. Otherwise, homologous recombination deficiency has also been associated with the non-endometrioid histologies and p53 abnormal patients [117]. Consequently, EC patients are good candidates to test the efficacy of PARPi (Figure 12).
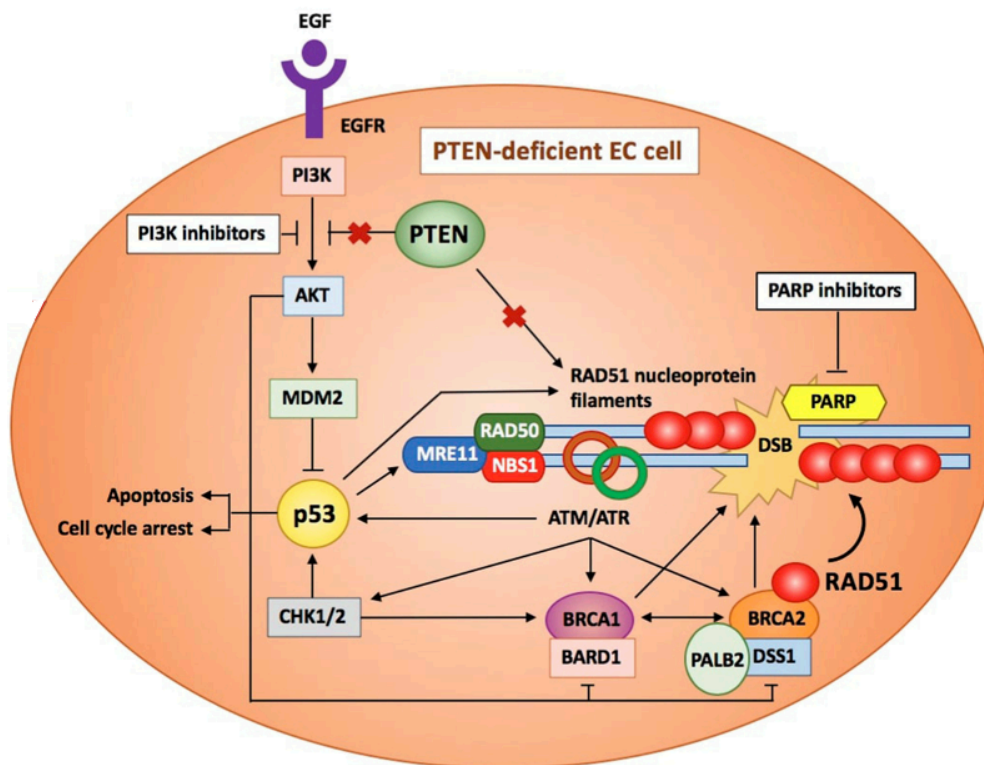
Figura 12. Overview of the HR pathway in PTEN-deficient EC. *Mutant PTEN inhibits p53 and RAD51 expression, thus impairing the HR system. Similarly, mutant p53 downregulate the MRN complex (MRE11, RAD50, and NBS) and RAD51 levels. Adapted from Musacchio et al.* [112]

2.8.3. Treatments anti-HER2 (SYD985)

ERBB2 is a protein coding gene responsible for the expression of human epidermal growth factor 2 (HER2), a membrane receptor with tyrosine kinase activity that induce signal transduction through the PI3K/Akt pathway, the RAS/MAPK pathway, and the JAK-STAT pathway [118]. This downstream activation leads the induction of genes that

promote oncogenic transformation via cell survival, proliferation, angiogenesis, and metastasis [36,119].

HER2-positivity varies considerable in EC (17-80%). The rates of HER2 overexpression and/or amplification differ in the literature [118,120]. However, existing evidence correlate HER2 protein overexpression and/or ERBB2 amplification in EC with the serous histology, with frequencies around 35% [121]. It is also detected in other subtypes of EC such as clear cell carcinoma (30%), or carcinosarcoma (25%). In contrast, there is a low expression of HER2 in EEC (10-20%) [122]. Considering that NEEC tumors are responsible of more than 50% of EC death, targeting HER2 in these cases can be a very good approach.

Trastuzumab is the main monoclonal anti-HER2 antibody used to treat HER2-positive breast, gastroesophageal, and gastric cancers [123,124]. Trastuzumab binds to the extracellular binding domain of HER2 and inhibit the intracellular signaling cascades [125] (Figure 13). Although it is considered one of the most effective treatments in oncology, a significant number of HER2-positive patients do not benefit from it. There are some mechanisms of resistance: i) obstacles preventing trastuzumab binding to HER2 such as the increased expression of a constitutively active p95HER2; ii) upregulation of downstream signaling pathway like PIK3CA mutation, loss of PTEN, or increased AKT; iii) signaling through alternative pathway including other HER family members such as EGFR and HER3, and unrelated pathways such as Notch; and iv) failure to trigger the mechanism to kill tumor cells [126,127].

Antibody-drug conjugate (ADC) provides a solution to solve some of these problems related with resistance. ADC is composed of three components: payload (or cytotoxic drug), monoclonal antibody, and linker. The monoclonal antibody targets the antigen-expressing tumor cells and internalizes the payload inducing cell toxicity [128]. Ado-trastuzumab emtansine (T-DM1) is an example of first-generation ADC. Despite its efficacy, primary or acquired resistance frequently develops. Therefore, second-generation ADC are meant to supersede them by using a more potent payload with a different mechanism of action.

SYD985 is a second-generation ADC consisting of trastuzumab bound to duocarmycin payload via a cleavable linker. Duocarmycin alkylates DNA, inducing DNA damage, impaired DNA transcription, mitochondrial stress, and apoptosis [129]. It has been tested through *in vitro* and *in vivo* studies and compared against T-DM1. Both ADCs showed similar binding affinity to HER2, internalization and cytotoxicity. However, T-DM1 only presented efficacy in HER2 3+ cell lines, whereas SYD985 also induce cytotoxicity in

Introduction

HER2 negative and HER2 1+, 2+, and 3+ cell lines [129,130]. Specifically, SYD985 has also been investigated in SEC and uterine carcinosarcoma (CS). The *in vitro* studies using cell lines and the *in vivo* studies through patient-derived-xenograft (PDX) models showed high efficacy of SYD985 in strong (3+) as well as in low to moderate (1+/2+) HER2 positive tumors [131,132]. Consequently, targeting HER2 positive EC tumors using SYD985 could be a very good approach that should be further investigated.
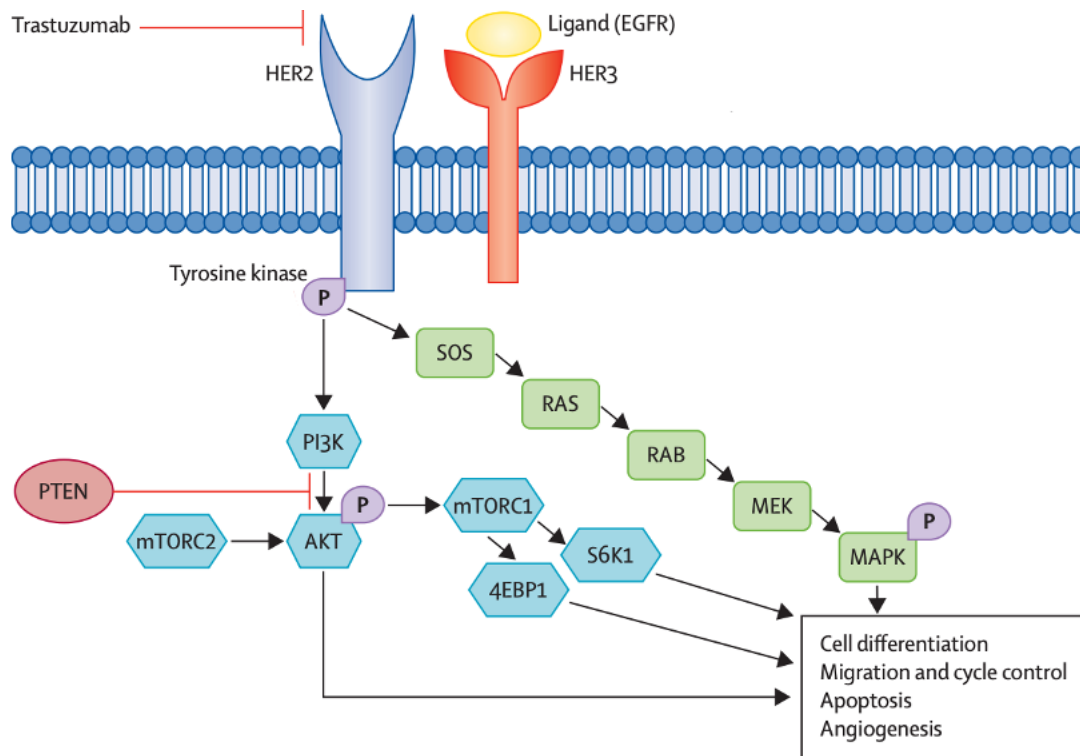
Figura 13. *Mechanism of action of trastuzumab in HER2 positive cells. Adapted from Loibl et al.* [125]

3. ENDOMETRIAL CANCER BIOMARKERS

3.1. BIOMARKER DEFINITION

The group of the National Institute of Health define a **biomarker** as "a biological molecule that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic response to a therapeutic intervention" [133]. An ideal biomarker should be easily obtained by a non-invasive, reliable, robust, and reproducible technique. It should have high sensitivity (low rate of false negatives) and high specificity (low rate of false positive), and it should be simple, cheap, and accessible to all the population that require it [134].

Biomarkers have several applications, including screening, diagnosis, prognosis, prediction, and monitoring. Each biomarker should discriminate patients with different status of the disease:

- **Screening biomarkers:** predict the probability of developing a specific disease. In cancer, they must detect asymptomatic patients at a curable stage. One of the most used screening biomarker is prostate-specific antigen (PSA) [135].
- **Diagnostic biomarkers:** they are applied in patients presenting symptoms and should indicate whether an individual has or not a specific disease. In EC, many research efforts have been undertaken to provide clinically useful biomarkers, but none of them have still reached clinical practice [136]. Importantly, recent studies have demonstrated that uterine aspirate are a valuable sample to detect diagnostic biomarkers of EC [137].
- **Prognostic biomarkers:** provide information about the disease course in a patient already diagnosed with a specific disease. These biomarkers are used to classify patients into different risk groups and are very useful when there are different treatment options. In EC, L1CAM and CTNNB1 are some examples of these type of biomarkers [138] (described in section 2.7.2.).
- **Predictive biomarkers:** help determining which patients are more likely to respond to a specific treatment. They are the bases of personalized medicine.
- **Monitoring biomarkers:** assess that a patient remains disease free after the treatment. These biomarkers should be very sensitive and specific to detect any change of the disease.

3.2. PROTEIN BIOMARKER

Biomarkers may include any class of biological molecule, including DNA, RNA, proteins, and metabolites, among others, which are investigated thanks to a broad spectrum of strategies such as genomics, transcriptomics, proteomics, and metabolomics, respectively. Among them, proteins and proteomics present remarkable advantages as biomarkers. Firstly, proteins are the biological end products that determine normal or disease physiology and are the target of most drugs. Secondly, the diversity of proteins is much higher than that of DNA or RNA, because alternative splicing and post-translational modifications generate different proteins from the same gene. Humans have an estimated number of 20,300 genes, 100,000 mRNAs, and 114,000 metabolites [139,140]. In contrast, humans can produce up to 1,8 million different proteins [141]. This diversity increases the probability to identify a specific protein, or a group of proteins associated to a specific condition, such as cancer. From an analytical perspective, protein biomarkers can be detected, analyzed, and quantified easily in clinical laboratories through techniques already well implemented such as immunoassays or immunohistochemistry.

Otherwise, proteins present several challenges regarding the measurement. Due to the high number of proteins, there is also a huge range of concentration that vary over six to seven orders of magnitude in tissues and up to twelve in blood [142,143]. Therefore, low-abundance proteins are hard to detect because they can be cover by high-abundance proteins. Moreover, the physiology of the disease could be explained by the accumulation of a protein in a specific subcellular component or due to post-translational modifications, while the levels remain equal in a healthy person. Fortunately, proteomics has evolved in the last years to provide more sensitivity, specificity, and reproducibility, adding new tools for biomarker discovery [144,145].

3.3. BIOMARKER SOURCES: CLINICAL SAMPLES

One of the most important challenges to succeed in biomarker research include the selection of the best suitable sample. There is a variety of clinical samples, such as tissue, blood/serum/plasma, or proximal fluids.

3.3.1. Tissue samples

Tissue samples collected by biopsy or surgery are the most used clinical samples for biomarker research. The main advantage of these samples is the higher concentration

of biomarkers in the affected tissue compared to other biological samples, where biomarkers are found in lower concentrations. However, collection of tissue specimens requires invasive procedures for the patient and its study may produce a bias due to the tumoral heterogeneity [146].

Among tissue samples, formalin-fixed paraffin-embedded (FFPE) tissue is the most available sample but the less used in biomarker research. The principal advantage of FFPE tissue samples is that they have routinely collected and stored in pathology departments for many years so there are huge repositories of these type of samples. In contrast, their application in biomarker research is limited because of cross-linking and fragmentation of nucleic acids, as well as loss of enzymatic activity [147]. The time between sample acquisition and fixation, fixation time, and sample thickness are key elements to guarantee the preservation of nucleic acids and proteins [148]. Up to date, it is considered that working with FFPE in biomarker studies is challenging.

3.3.2. Blood / Serum / Plasma

The most common liquid biopsy and the preferred clinical sample for biomarker research has always been blood, serum and/or plasma. They are collected routinely, easily, and minimally invasive from the patient. Blood is the most complete human proteome because it is in contact with all the body organs, and it includes molecules secreted by all the tissues. Therefore, it can capture the heterogeneity of a specific disease. Moreover, total protein content in blood is very similar among the population, facilitating the comparison of a specific biomarker between patients.

However, blood is the most challenging clinical sample because it contains a high number of proteins with concentrations that vary from ten to twelve orders of magnitude [142]. Consequently, it reduces the ability to identify biomarkers at low concentration. For instance, albumin is detected in levels of mg/ml whereas cytokines are present in levels of pg/ml.

3.3.3. Proximal fluids

Proximal fluids such as urine, saliva, cerebrospinal fluid or uterine aspirate (UA), can also be used for the identification of biomarkers. In contrast to blood, proximal fluids are in direct contact or close to the site of the disease. It has been demonstrated that these fluids are highly enriched of proteins from the disease tissue and microenvironment [149]. If those fluids are easily and rapidly obtained, they might be an attractive source of biomarkers.

Nevertheless, there are some limitations to consider. Invasiveness to collect each specific proximal fluid vary between samples. For instance, cerebrospinal fluid requires a very invasive technique for its collection and the volume obtained is low, whereas urine can be obtained through non-invasive techniques in large quantities. Moreover, proximal fluids are less characterized than blood and thus, they require great efforts to standardize preanalytics and analytics. Finally, similar to blood, there are some proximal fluids presenting high number of proteins with different magnitudes of orders, hindering the detection of low-concentrated biomarkers [150].

3.4. BIOMARKER PIPELINE

Generally, the phases of biomarker research consist of discovery, verification, validation, and clinical evaluation prior to the final implementation in the clinics. The number of samples and analytes expected in each phase is described in Figure 14.
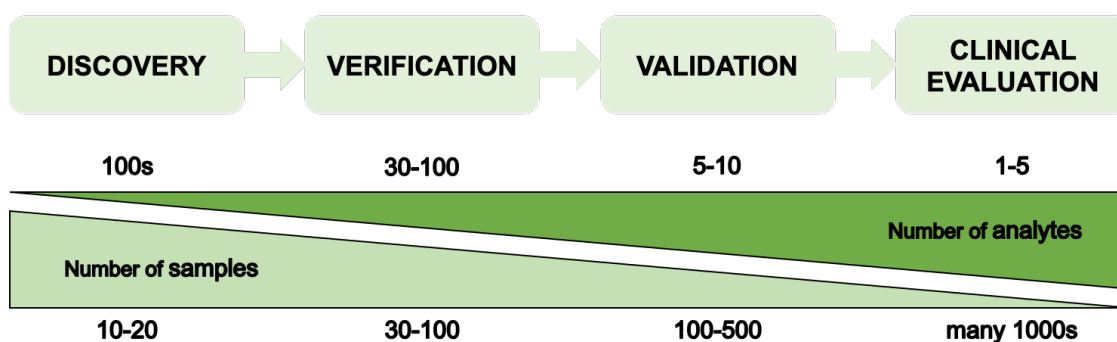
Figura 14. *Biomarker development pipeline*

The **discovery phase** is an untargeted process whose main goal is to look for differential proteins between two groups (cases versus controls). The number of patients included in this phase is low, approximately between 10-20 patients per group, and the number of differentially proteins analyzed are high. The false discovery rate (FDR) is expected to be high in this phase because of the limited number of patients analyzed and the heterogeneity among samples. Therefore, they are considered candidate biomarkers [151].

The **verification phase** is a crucial step and is considered the bottleneck in the biomarker pipeline. Candidate biomarkers from the discovery phase are analyzed in a higher number of samples. Indeed, increasing the number of patients per group enables to increase the sensitivity and specificity of the biomarkers. This phase is crucial in order to prioritize the most robust biomarkers to a subsequent validation phase.

The **validation phase** requires high efforts and investment; hence, it is assessed with the most promising biomarkers from the verification phase. This phase is characterized

to use a large cohort of patients, which represents the intrinsic heterogeneity of the target population.

Finally, the **clinical evaluation** is needed before the regulatory approval. During this phase, the final biomarker, or biomarker panels, is analyzed in a biggest and independent cohort of patients. At this stage, it is preferably to use clinical samples obtained with a non-invasive technique and easy to be applied in the clinical practice [152].

Despite the efforts and investment in biomarker research in the last decades and that there are some promising biomarkers for EC diagnosis and prognosis, none of them are already implemented into the clinics [137].

4. PROTEOMICS IN BIOMARKER RESEARCH

Proteomics is defined as the large-scale study of proteins. In comparison to other omics (genomics, transcriptomics, and metabolomics), the study of proteins is more complex because concentration vary in several orders of magnitude and they cannot be amplified, contrary to acid nucleic where polymerase chain reaction (PCR) enable to amplify those genes poorly expressed.

Regarding the biomarker pipeline, mass spectrometry (MS) has been the proteomic approach most widely used in the first's steps (discovery and verification phase), whereas antibody approaches are the gold standard for validation and clinical evaluation since it is broadly implemented in clinical routine.

4.1. MASS SPECTROMETRY

4.1.1. Mass Spectrometry Basics

The general workflow of MS consists of three steps: sample preparation, sample separation, and MS analysis.

❖ **Sample preparation**

Proteins must be purified and separated from the rest of molecules of the sample. In tissue samples, cells should be lysed, and proteins solubilized, whereas in biofluids such as plasma or urine, proteins are already soluble. Proteins are then denatured by heat or using denaturation reagents such as urea or SDS. This step is necessary to break the tridimensional structure of proteins (reduction and alkylation) and allow the access of proteolytic enzymes [153,154]. Trypsin is the most used protease, and it specifically fragments proteins into peptides at the carboxyl side of the amino acids arginine (R) and lysine (K) (Figure 15). These peptides have an average of 14 amino acids that can be predicted *in-silico*, quantified, and analyzed by MS. There are others proteolytic enzymes such as metalloproteases or pepsins that break proteins in specific amino acids. They can be combined in order to increase the protein sequence coverage [155].

MS-approach can identify the whole proteome with a good sensitivity, but a lack of specificity. Therefore, there are different methodologies to increase the specificity according to the aim of the study. For instance, depletion of abundant protein, such as

albumin or immunoglobulin G, allow the detection of proteins at low-concentration, or glycoprotein enrichment using specific affinity columns [156,157].
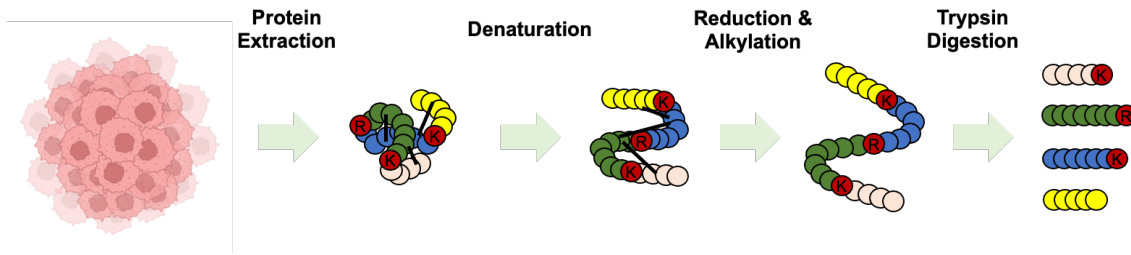
Figura 15. Sample preparation. *Proteins are extracted from clinical sample, denatured by heat, reduced, and alkylated by reagents, and digested by trypsin. This process fragment proteins into peptides ready to be analyzed in the MS.*

❖ **Sample separation**

Proteins should be separated because the proteome is very complex, and the huge number of proteins can collapse the matrix before going inside the mass analyzer. There are different techniques to separate proteins that increase the sensitivity and the proteome coverage:

1. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): it is the first method used for protein separation of complex biological mixtures. This technique is based on the ability to separate proteins according to physiochemical properties such as isoelectric point and molecular weight. A modified version, the two-dimensional difference gel electrophoresis (2D-DIGE), offered a greater sensitivity and reproducibility because samples are labeled with fluorescent dyes and co-resolved on a single 2D gel for direct quantitation [158].

2. Strong cation exchange (SCX): the stationary phase contains negative charges, so acidified peptides with a positive charge stick to them. A solvent with increasing amounts of salt competes with them and induce peptide elution. This technique has been mainly used for the study of posttranslational modifications (PTMs), specifically phosphorylation and N-terminal acetylation [159,160].

3. Reverse phase high performance liquid chromatography (RP-HPLC): this separation method is based on the hydrophobicity of protein. The stationary phase is composed by silica gel or synthetic molecules where peptides can be bound. The elution occurs by increasing the concentration of an organic solvent, such as acetonitrile or methanol. Nowadays, this is the approach most used because of its compatibility with MS [161].

The main advantage of SCX and RP-HPLC in front of 2D-PAGE and 2D-DIGE is the number of proteins that can separate, and the capacity to be connected directly to the mass spectrometer.

❖ MS analysis

Mass spectrometers measure the ratio between mass and charge (m/z). Consequently, samples must be ionized before going inside the mass analyzer. There are different methods to produce ions from biological materials: Electrospray Ionization (ESI) or Matrix-Assisted Laser Desorption Ionization (MALDI). In ESI, samples are injected in a conductive capillary with a high voltage and result the emission of aerosols of charged droplets of samples, whereas, in MALDI, a laser impact a matrix containing the samples and induces the desorption of proteins (Figure 16. The primary advantage of ESI over MALDI is that ionization is initiated directly from the bulk solution, so it can be directly coupled with the chromatography from the sample separation [162].
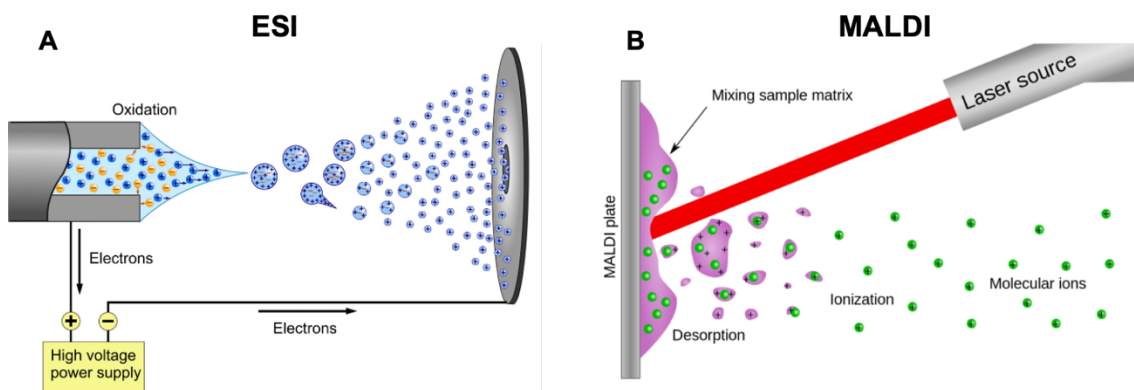
*Figura 16. Ionization methods. **A.** Electrospray Ionization (ESI). **B.** Matrix-Assisted Laser Desorption Ionization (MALDI). Image from www.bioexcel.eu*

Once peptides are ionized, they are introduced and analyzed by one or several mass analyzers, depending on the mass spectrometer configuration [163]. There are different MS approximations (Figure 17):

1. Time Of Flight (TOF): ions are generated in pulses, typically from MALDI, and are accelerated with the same kinetic energy. Hence, smallest ions fly most quickly. The time taken to traverse the flight tube can be converted to m/z value.
2. Quadrupole Ion Trap (QIT): it is filled with ions, trapping them. After, QIT scans by selectively destabilizing ions and ejecting them to the detector.

3. Orbitrap (OT): ions orbit in a spindle from the beginning to the end at a characteristic frequency. The Fourier transformation infers these frequencies and enable m/z value computation.
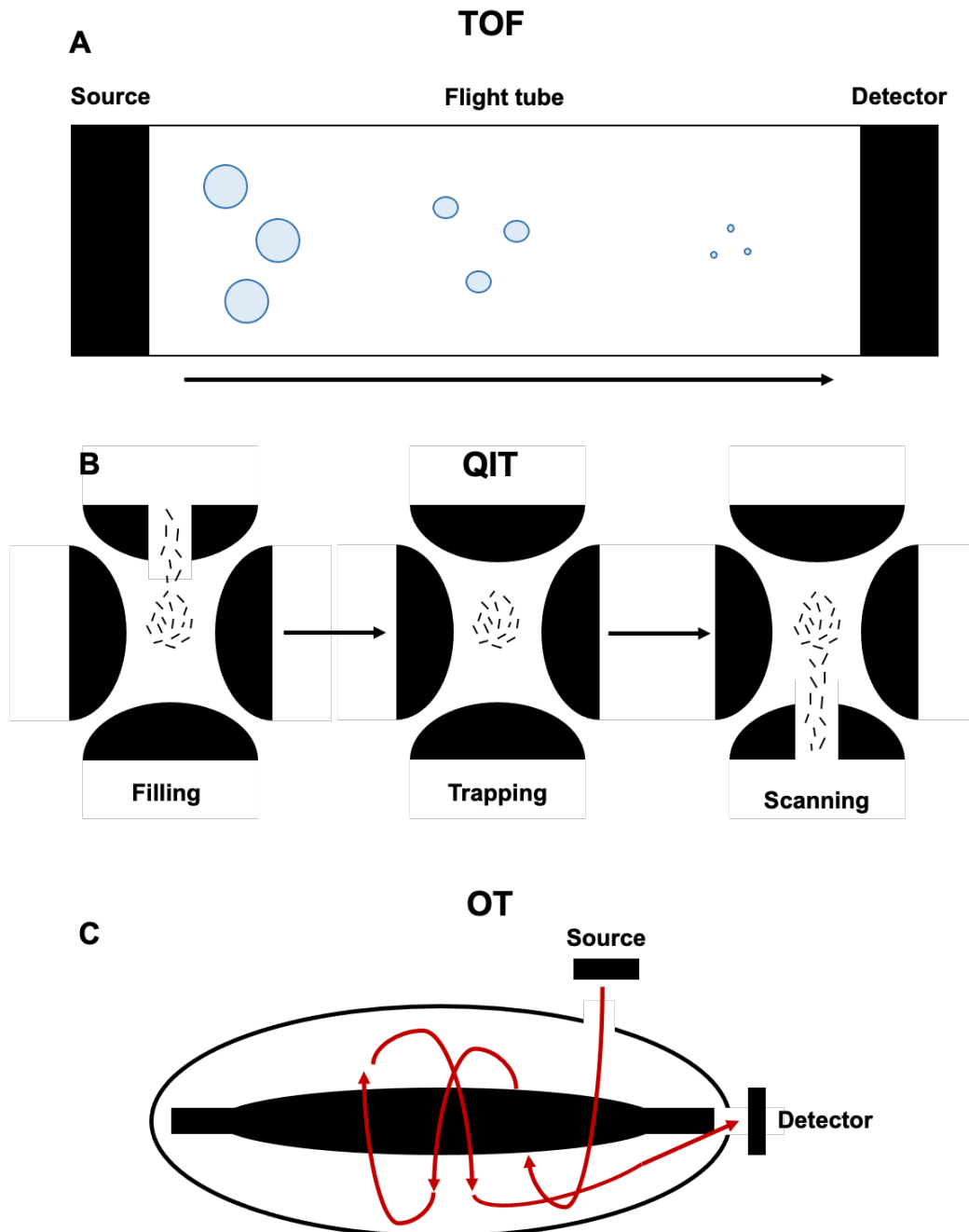
Figura 17. Mass analyzers. A. *Time Of Flight (TOF). The time to traverse the flight tube can be converted to m/z value.* **B.** *Quadrupole Ion Trap (QIT). Ions fill the MS, are trapped, and are scanned.* **C.** *Orbitrap (OT). Ion orbit around the spindle and by Fourier transformation the m/z value is computed.*

Nowadays, most proteomic studies use tandem mass spectrometers, which are based on combination of mass analyzers. Common tandem mass spectrometers are the quadrupole time of flight (QqTOF), the quadrupole orbitrap (Q-OT), and the ion trap

orbitrap. In these approximations, peptides are selected in the first mass analyzer (MS1), fragmented in the collision-induced dissociation (CID), and the resulting fragments are analyzed in the second mass analyzer (MS2) [154,163]. Finally, the detector records the intensity of ions from the mass analyzer and convert them to m/z values (Figure 18).
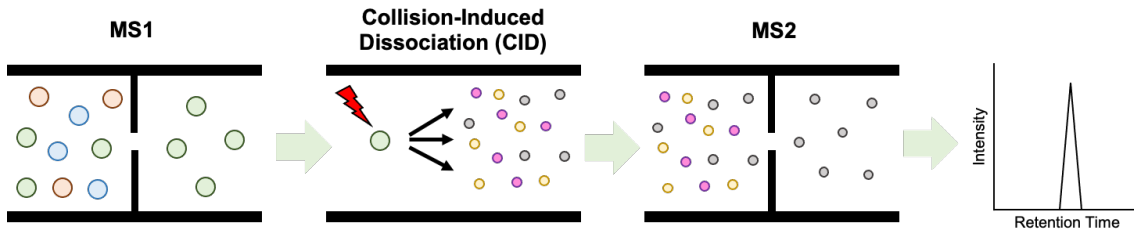
Figura 18. Tandem Mass Spectrometers. Peptides are selected in the first MS (MS1), they are fragmented in CID cell, and resulting fragments are analyzed in the second MS (MS2).

4.1.2. Protein Quantification

In the last years, proteomics has evolved from a qualitative to a more quantitative approach [164]. Although mass analyzers are quantitative approximations, the resulting ion signals is subjected to variations. For instance, during sample preparation, peptides can be degraded in the chromatographic column; or, even during ionization, it can suppress or enhance a specific ion specie. There are new strategies to solve this problem that can be divided into: label-free and label-based approaches.

❖ **Label-free approach**

In this approach, the same amount of each sample is analyzed in the MS to detect the relative abundance of protein. Therefore, it is very important to control all the variables to minimize variations and obtain an optimal result [165].

Label-free approach is not an accurate quantitative method, but it is easy and cost-effective. Consequently, it is appropriate in the first steps of biomarker pipeline (discovery phase). The main disadvantage of this technology is the limited accuracy and precision; hence, it is recommended to consider only those differences between the groups of comparison with high significance (for instance, greater than two-fold) [166].

❖ **Label-based approach**

The incorporation of amino acids with stable isotopes (^{13}C, ^{15}N and/or ^{18}O) allows to control internal variations during sample preparation, sample separation and MS analysis. Peptides containing these isotopes, called heavy peptides, behave identically

as the endogenous peptides from the sample, called light peptides. The only difference between them is the increased mass, which can be detected and differentiated by MS. Heavy peptides are added to the sample and, as they are going to be fragmented equally as light peptides, and they share the similar physic-chemical characteristics during sample separation, they will be detected at the same time in the MS analysis. Later, the MS signal of light peptides can be normalized by the MS signal of heavy peptides [167,168]. Otherwise, the addition of heavy peptides in the firsts steps of processing is crucial to control the maximum number variables during the whole process. In addition, label-based approaches increase the confidence and reliability of the results because of the co-elution of heavy and light peptides.

There are different label-based methods, being the isotopically labeled proteins and the isotopically labelled synthetic peptides the most important.

Isotopically labeled proteins are the ideal approximation for quantitative proteomics because they can be added during sample preparation and control the variations occurring during enzyme proteolysis and the next steps. However, the chemical synthesis of proteins is still unavailable because of the size, and the difficulties to reproduce the specific folding and the tridimensional structure.

Another strategy is the incorporation of amino acids with stable isotope, such as ^{13}C or ^{15}N labeled to arginine or lysine, into a cell culture (this is called SILAC). In this approach, there are two population of cells growing in two different culture media, one with the "light" medium containing amino acids with the natural isotope, and another with the "heavy" medium containing amino acids with the stable isotope. After several cell divisions, at least 5 in mammalian cells, all proteins from the "heavy" medium incorporate the amino acids into the proteome. Therefore, mixing both cell lines allow the comparison of relative quantification between two different conditions [169,170]. Nevertheless, SILAC is only applicable in cell culture, single-cell organism (e.g., yeast and bacteria) or, at most, in small organisms like mice [171]; and it is hard to compare multiple samples.

The super-SILAC approach has emerged as a variant that can be used in tissue and biofluid samples. This technique uses a mixture of SILAC-labeled cells as a spike-in standard for accurate quantification [172]. The main challenge is the selection of the super-SILAC mix that better represents the proteome of the clinical sample [173]. In addition, mixing the sample of interest by an exogenous proteome increase the complexity, and may decrease the sensitivity and specificity of the approach.

Both techniques, SILAC and super-SILAC, are appropriate methods for discovery studies because they allow the relative quantification of the complete proteome of many samples.

Isotopically labelled synthetic peptides can be chemically synthetized by several companies. It can incorporate post-translational modifications, such as phosphorylation and acetylation. The main advantage of this approach is that it introduces less complexity to the sample compared to SILAC and super-SILAC [174]. This approach is mainly used in advanced steps in the biomarker pipeline, such as verification and validation phases, when it is known which are the proteins and peptides to study. However, there are some limitations since synthetic peptides do not control the first stages of sample preparation (e.g., enzymatic digestion), the preparation is hard, and they have a low stability [175].

4.1.3. Mass Spectrometry Acquisition Strategies

There are two different approaches for the identification of proteins through MS: untargeted MS and targeted MS. The selection of the most appropriate strategy is dependent on the phase of the biomarker pipeline.

❖ **Untargeted MS approaches**

Typically, MS has been used in proteomics to characterize the proteome of clinical samples and in discovery phases to identify proteins differentially expressed between two conditions. Untargeted MS is a label-free approach, so it is more qualitative rather than quantitative method. There are two main platforms for untargeted MS: Data Dependent Acquisition (DDA) and Data Independent Acquisition (DIA).

Data Dependent Acquisition (DDA) is a method for protein identification without the need of any knowledge of the analyzed sample. Therefore, it is a good approximation for the first steps of biomarker pipeline (discovery phase). In this approach, the most intense peptides are selected in the first MS (MS1), fragmented in the CID cell, and detected and analyzed in the second MS (MS2). This process is repeated for all the peptides resulting in a list of several thousand of MS spectra [176] (Figure 19A). After that, this MS data is compared with the theoretic MS data generated from sequenced protein databases (such as Uniprot). All the proteins from the database are *in silico* digested using the same proteolytic enzymes that the ones used during sample preparation. Finally, searching algorithm, like Mascot or Sequest, match the best theoretic peptide with each MS spectra [177,178].

The correlation of peptides that are specific and unique allows the identification of protein, whereas the correlation of peptides that are common in many proteins leads to the identification of protein groups. Importantly, DDA selects the most intense peptides in the MS1 and could generate a bias towards the most abundant proteins. Consequently, it generates a limited reproducibility between sample replicates with an overlap between 35 to 60% [179].
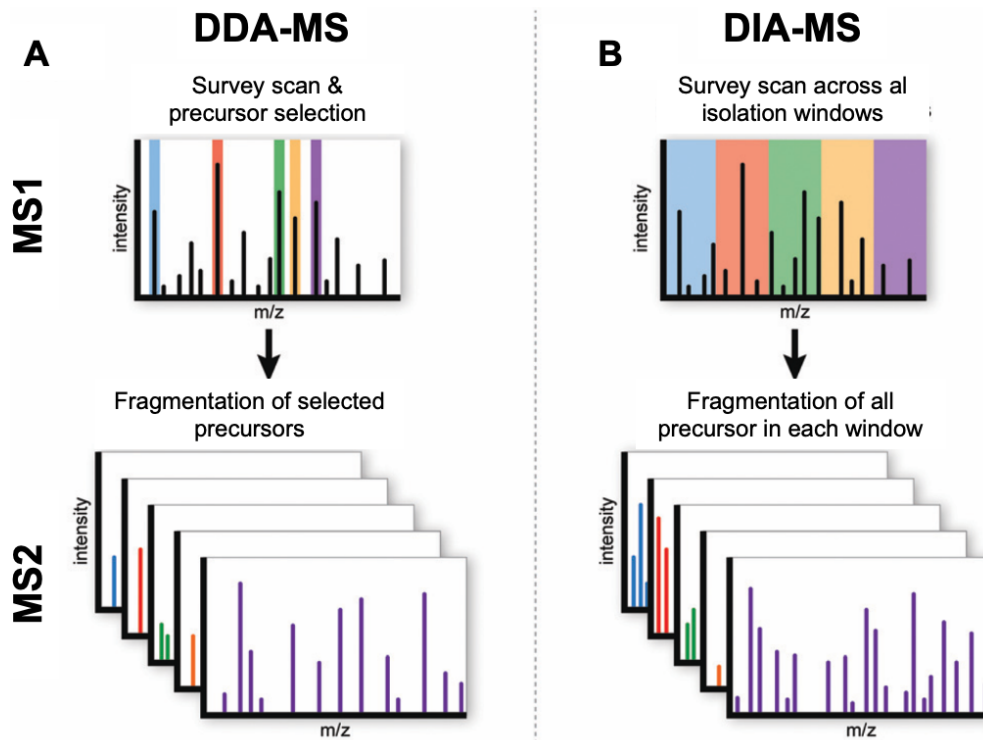
Figura 19. Schematic overview of the DDA-MS and DIA-MS. A. *In DDA-MS, the top n most abundant peptides are selected based on the survey scan (MS1) and fragmented in MS2.* **B.** *In DIA-MS, the survey scan provides snapshot of the peptides (MS1). Pre-defined wide isolation windows cover the whole MS1 m/z range, and all peptides are fragmented in MS2. Image adapted from Krasny et al.* [176]

Data Independent Acquisition (DIA) is a method that acquire data from MS1 and MS2 without bias to the most intense peptides, overcoming the inherent limitation and reproducibility of DDA. The principle of DIA is the generation of the whole proteome performing the fragmentation of all the peptides. Specifically, peptides are identified in the MS1, and the resulting fragments are accumulated in a different windows of mass-to-charge (m/z) range and analyzed in the MS2 [176,180] (Figure 19B). For this reason, DIA is also a good approach for the first steps of biomarker pipeline (discovery).

The major strength of DIA is the exceptional reproducibility in protein identification across multiple experiments. However, the analysis of all the fragments generates complex MS spectra and, hence, the data processing is very challenging. It implies the necessity to

generate specific libraries. Therefore, in situations where a reference library is not available, the generation of it requires high number of samples, instrument time and costs [176].

❖ **Targeted MS approaches**

Generally, antibody-based assays, such as IHQ or ELISA, has been used for targeted studies. However, their low-throughput capacity makes them more suitable for advanced steps of the biomarker pipeline (validation phase and clinical evaluation). Targeted MS approaches appears as a very good configuration for intermediate steps of the biomarker pipeline (verification and/or validation phases). Targeted MS approaches are able to evaluate a large number of potential biomarkers from the discovery and select the most promising biomarker for validation phases [154].

Before performing targeted MS, it is necessary the selection and generation of peptides representing the proteins that will be analyzed (heavy peptides). There is a limited number of peptides that can be analyzed at the same time in each sample (around 100). During sample separation, generally in liquid chromatography, the retention time of each peptide should be monitored to determine the specific time window of the resulting fragments. Therefore, to maximize the number of analyzed peptides, each one should be eluted at different times. This procedure is called scheduled acquisition [181,182].

The main advantage of targeted MS approach is the capacity to generate replicates that can be quantitatively compared and avoids producing missing data. There are two different targeted MS approaches: Selected Reaction Monitoring (SRM) and Parallel Reaction Monitoring (PRM).

Selected Reaction Monitoring (SRM) is performed by the selection of peptides from the MS1, the generation of fragments in the collision cell, and the selection and analysis of fragments in the MS2. The fragment associated to a peptide is called transition, and usually three to five transitions are measured from each peptide to increase the confidence. Finally, the quantification results on the integration of the area under the peak of elution of each transition [183] (Figure 20A).

SRM is an approach that allows the quantification of several tens of peptides in a single MS analysis. The obtained results are reliable and robust because of the two MS, and with high reproducibility among samples. Moreover, internal controls, such as isotopically labeled proteins or isotopically labeled synthetic peptides, increase the confidence of peptide identification since both co-elute at the same time and present the similar

fragmentation pattern, and enables the possibility to perform an absolute quantification when it is known the concentration of the heavy peptides [184,185].

However, SRM also presents some limitations: i) the development of this approach is time consuming because the transition list should be defined before MS analysis; ii) the analysis of only a limited number of transitions per peptide can generate bias and some samples should be reanalyzed; iii) all the transitions must be validated in a representative number of samples before the analysis of the full set [186].
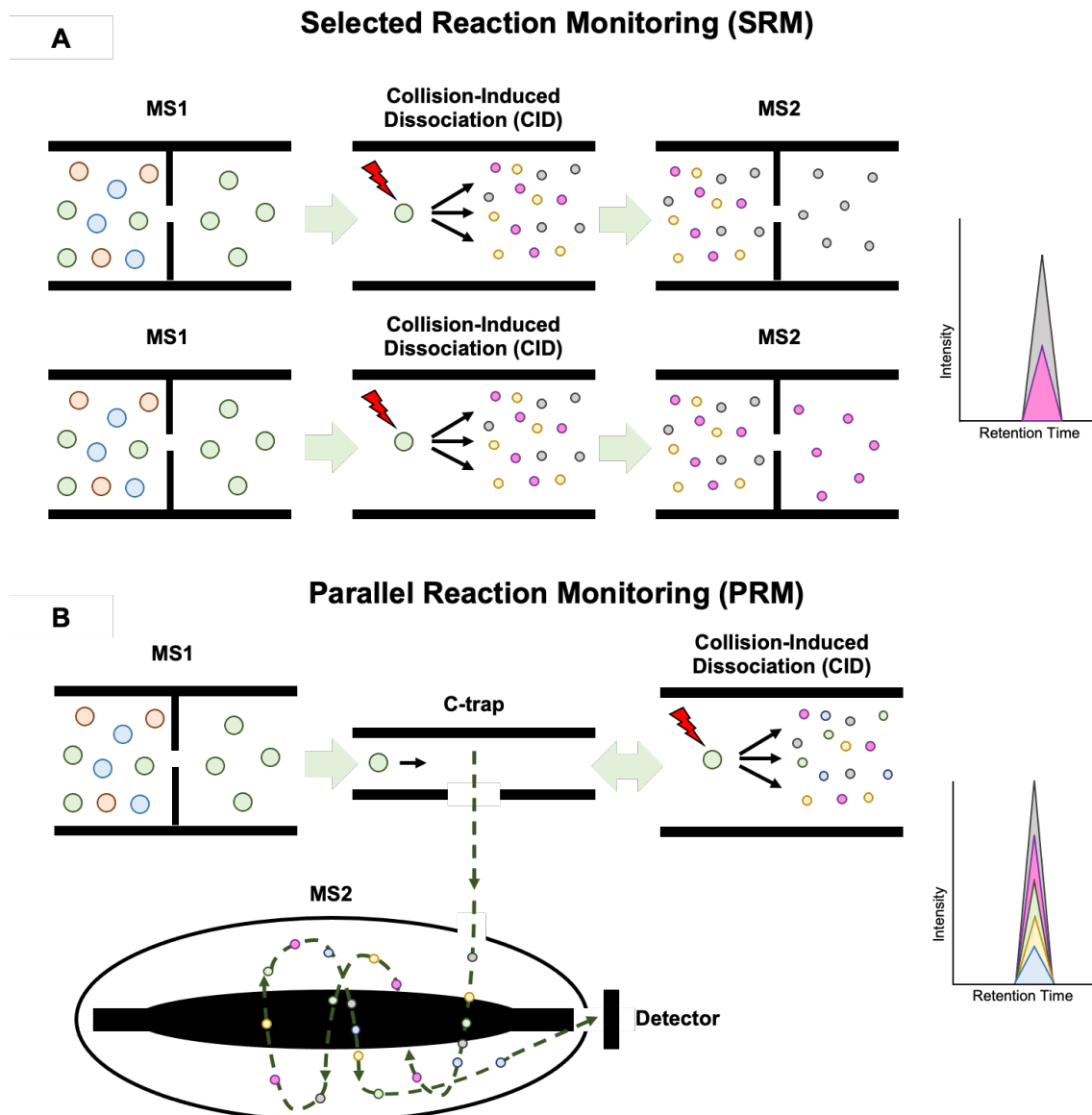
Figura 20. Targeted MS approaches. A. *Selected Reaction Monitoring. Each MS analysis quantify a single fragment.* **B.** *Parallel Reaction Monitoring (PRM). Only one MS analysis allow the quantification of all fragments.*

Parallel Reaction Monitoring (PRM) performs like the SRM, but the triple quadrupole (typically the MS2) is replaced by a high-resolution MS (TOF or Orbitrap). In this

approach, peptides are selected in the MS1 and transferred to the collision cell through the C-trap. The resulted fragments go back through to the C-trap to the MS2 where are analyzed at high resolution. Finally, the quantification can be performed integrating the areas under the peaks of all the fragments [187] (Figure 20B).

The main advantage of PRM over SRM is the capacity to analyze all the fragments from a protein in a single MS analysis, whereas SRM requires one MS analysis for each fragment. Therefore, it increases the resolution, and reduce the possibility of background and the necessity to reanalyze [182,188]. Moreover, PRM approach is also compatible with isotopically labeled proteins or isotopically labeled synthetic peptides with the advantages previously described.

4.2. ANTIBODY-BASED TECHNIQUES

MS-based approaches have evolved in the last years from qualitative to quantitative methods to measure protein expression, but still, there are not technologies broadly implemented in clinical practice. On contrary, antibody-based approaches, such as IHQ or enzyme-linked immunosorbent assays (ELISA), are widely employed in the clinical environment because they are cost-effective and simple. These methods are based on specific antibodies that target the protein of interest. Later, a second antibody linked with a reporter (biotin or fluorophore) recognized the first antibody. Finally, a substrate is added and is converted into a colorimetric signal that can be quantified. This signal is proportional to the amount of protein contained in the sample [189].

Among the most available techniques, **IHQ** is performed in FFPE tissues, whilst **ELISA** is mainly used to assess protein expression in biofluids. These approximations are highly sensitive, and the absolute quantification facilitates the comparison of samples even between laboratories. However, the main challenge of antibody-based approaches is the limited availability of antibodies [190,191]. Commercial antibodies are often not very specific and could induce a lack of reproducibility between assays.

5. PDX MODELS

Establishing suitable cancer models is challenging. In the last years, mouse models have emerged as a relevant preclinical tool since they can mimic different steps of a disease and can be used for biomarker identification or therapy drug screening [192,193]. Still, the first murine models based on implantation of cancer cell lines or transgenic mice failed to recapitulate some aspects of the disease and could not adequately predict drug effect [194]. At that moment, it was clear the need to generate animal models that represent the disease, including the intra-tumor diversity of the disease (i.e., clinico-pathological features including histology, grade, FIGO, etc., and molecular features).

Patient-derived xenograft (PDX) models are a good alternative to overcome some of these weaknesses. PDX generation is based on the implantation of fresh tumor tissue from the patient directly into immunocompromised mice [195]. Specifically, tumor tissue obtained in the surgery room from the patient is sliced in small fragments or disaggregated in a cell suspension and transplanted to an immunocompromised mouse. There are two different strategies for implantation: heterotopic implantation occurs when the tumor is implanted in a different area to the original tumor site (generally subcutaneously); orthotopic implantation occurs when the tumor is implanted in the same organ as the original tumor site. The main advantage of heterotopic PDX is the easy implantation and growing follow-up, but they rarely develop metastases and they do not mimic the real tumor microenvironment. In contrast, orthotopic PDX reproduce the same environment and metastatic process, but they can be hard to generate depending to the anatomical site of implantation and present difficulties to do a clinical follow-up [196].

PDX models present several advantages: i) preserve the molecular profile of the initial tumor along several passages; ii) maintain the clinic-pathological features of the initial tumor, including tissue architecture, cell-cell interactions, and spatial distribution; iii) amplification of tumor facilitate cancer research, such as biomarker identification or tumor characterization; iv) generation of a mice cohort from the same tumor can be used for preclinical studies in drug screening [196,197].

5.1. EC PDX MODELS GENERATION

The first orthotopic EC PDX model was described by Cabrera et al. [198]. Firstly, the tumor was grown subcutaneously before the orthotopic implantation. Concretely, once the subcutaneous tumor was engrafted, it was removed and mechanically fragmented to be injected transvaginally or transmyometrially into a nude mouse. Transmyometrial

implantation show a higher rate of engraftment. These orthotopic tumors present myometrial infiltration, lymph-vascular space invasion, and dissemination to the pelvic cavity, mimicking the same behavior of EC. Moreover, tumors maintain the same molecular histopathological characteristics as the original tumor.

Likewise, Haldorsen et al. [199] develop an orthotopic EC PDX by mechanical disruption of the original tissue in a cell suspension and injected them to the uterine horn of an immunocompromised mouse. While PDX from Cabrera et al. [198] were euthanized after 63 days because of tumor invasion; PDX from Haldorsen et al. [199] took almost 10 months to grow. Nevertheless, they also demonstrated that orthotopic PDX tumors can be removed, mechanically disrupted, and reinjected again to another mouse, and develop a large cohort of orthotopic EC PDX models.

Apart from that, Depreeuw et al. [200] was the first to develop and characterize a big cohort of 24 subcutaneous EC PDX. They demonstrated that EC PDX could be established from the original tumor, but also from metastasis and recurrence with and engraftment ratio of 60%. Moreover, as they included different histologies and genetic subtypes, they confirmed that PDX models resemble the histologic and molecular features as the original tumor. The same was demonstrated by Unno et al. [201] in a cohort of 4 EC PDX models. Importantly, they unveiled that the that human EC stroma was replaced by mouse stroma after engraftment. Similarly, Unno et al. [201] corroborated that PDX also retain the characteristics that are unique for EEC and NEEC.

Lastly, Pauli et al. [202] described the development of PDX from organoids, called patient-derived tumor organoids (PDTO). The authors established culture organoids from two EC patients and characterized them by cytology and histology. Afterwards, the organoids were subcutaneously injected into immunocompromised mice. PDTOs reported an engraftment ratio of 86.4%. Furthermore, they found that PDXs and PDTOs have similar clinicopathological features to the original tumor.

5.2. USE OF EC PDX MODELS IN PRECLINICAL STUDIES

Personalized medicine is a practice of medicine that uses the genetic profile of an individual to guide decisions regarding prevention, diagnosis, and treatment of the disease. Consequently, it is expected to increase the efficacy of treatments while reducing possible side-effects. At this point, PDX is an emerging tool for personalized medicine because it retains clinic-pathological and molecular features from the original

tumor. In addition, several studies already demonstrated that drug responses in PDX correlates with those observed in the clinical practice with patients [203,204].

In EC, the PI3K/AKT pathway is one of the most active due to mutations and it has been attractive for personalized medicine in different EC preclinical studies. Winder et al. [205] tested the efficacy of MK2206, an inhibitor of AKT, in three different EC PDX models. They observed that MK2206 inhibit the tumoral growth in all three models (2 EEC and 1 NEEC). Similarly, Yu et al. [206] studied the combinative effect of ARQ092 (AKT inhibitor) and ARQ087 (FGFR1/2 inhibitor) in two EC PDX models. The combination of two different drugs tries to overcome AKT resistance and loss of efficacy. Although they saw synergistic effect *in vitro*, they only found a combinative effect in one of the models comparing with the single-agent treatments. However, the authors suggested that is necessary to define molecular signatures to overcome the possible resistance and increase drug efficacy.

Likewise, Depreeuw et al. [200] tested the efficacy of NVP-BEZ235 (PI3K/mTOR inhibitor) in combination with AZD6244 (MEK1/2 inhibitor) in an EC PDX with high risk of recurrence and carrying KRAS, PIK3CA, and PTEN mutations. The saw that single-agent treatment significantly reduced tumor growth compared with controls. Importantly, the combination of both therapies resulted in no increment of tumor growth and disease stabilization.

All the above-mentioned studies employ EC PDX models to evaluate the efficacy of drugs targeting EC altered pathways. However, EC PDX models could also be used to identify those pathways responsible of resistance and to identify new mechanism to overcome them. For instance, sorafenib is an antiangiogenic drug that was proposed for EC, but a multicenter phase II clinical trial only showed moderate response. Recently, Eritja et al. [207] studied the resistance mechanism of sorafenib *in vitro* and *in vivo*, and they observed that autophagy acted as a protective mechanism against sorafenib. Specifically, they used three orthotopic EC PDX and observed that sorafenib could be potentiated by the simultaneous inhibition of autophagy using chloroquine.

Otherwise, PDX models are also important in the identification of biomarkers that predict treatment-response. Groeneweg et al. [208] tested the efficacy of lapatinib as a single-treatment and in combination with trastuzumab. The *in vitro* and *in vivo* studies allowed the authors to demonstrate that single-treatment and combination of anti-HER2 only affect those tumors harboring HER2 amplification. Importantly, no response was seen in those tumors without HER2 amplification. Hence, HER2 amplification can be used as a predictor biomarker of response for anti-HER2 drugs.

These data reflect that EC PDX models are currently an important tool in personalized medicine defining new therapeutic options for the different EC subtypes and identifying those patients most likely to be sensitive to a specific agent. However, most of these studies are only based in one or few EC PDX models, so these results are hard to translate to the general population.

5.3. NEW PERSPECTIVES OF EC PDX MODELS

There are new perspectives for the exploitation of PDX models, such as the co-clinical trials. This is a concept like personalized medicine, which is based on the PDX generation from patients that are already enrolled in a clinical trial. Therefore, it could help in the identification of biomarkers that predict treatment-response and assess the therapeutic efficacy of each drug [209,210]. However, this method implies some limitations: i) the ratio of tumor engraftment for each model; ii) the time for PDX development can be very extensive, and it might impair the evaluation of PDX response for all patients from the clinical trial; iii) the heterogeneity of the original tumor can induce a bias in the PDX response.

Another emerging approach for PDX models is the preclinical studies. In contrast to traditional methods that were based on the analysis of drugs in few PDX models, this new method uses each PDX model as if they were patients participating in a clinical trial. Concretely, a cohort of different PDX models including different clinic-pathological and/or molecular features of the disease are used to test the efficacy of several treatment. The results of these preclinical studies are based on the population response, not only in the individual response of each mouse/tumor [211,212]. Therefore, it is very important to capture the heterogeneity of the specific disease and it is necessary a large cohort of PDX models. This type of approach is currently being used by pharmaceutical companies to increase the success ratio of drugs before the application in clinical trials with patients.

5.4. PDX-RELATED CHALLENGES

Although EC PDX have been widely used in the last years, there are still some challenges to overcome:

- Tumors are implanted in immunocompromised mice to avoid the rejection of human tumor. Hence, the immune system response associated to the tumor cannot be studied or is partially studied thanks to the use of humanized murine models. Humanized animal models have been successfully established in many cancer types

[213,214], and some studies have just developed humanized EC models to analyze the effect of anti-PD1 treatments [215,216].

- Human stroma is replaced by mouse stroma after several passages, and it is known that stroma is an important part of the tumor microenvironment implicated in cancer progression [217].

- Only small pieces from the tumor are used to be implanted in mice, so it can generate a loss of information because of tumor heterogeneity.

- Despite most of molecular and clinic-pathological features are maintained, tumors partially undergo mouse-specific evolution. Concretely, Ben-David et al. [218] monitored the SCNA of 1110 PDX of different cancers. They observed that some SCNA from the original tumor disappear, and others were acquired in the PDX model. However, other studies suggest that most of these changes are not in oncogenic genes and do not affect tumor evolution [200,219].

OBJECTIVES

BACKGROUND

Endometrial cancer (EC) is the most common cancer of the female genital tract and the fourth most frequent cancer in women in the United States [3]. Among cancers, EC incidence is increasing about 1% per year. This rising has been related with the increment of body weight in western countries; however, a recent study suggested that this trend is driven by non-endometrioid subtypes, which are not directly associated with obesity [5]. In general, patients diagnosed at early stages of the diseases are associated with an overall 5-years survival rate of 95%. Nevertheless, it decreases to 69% for patients diagnosed at regional metastasis, and a drastic reduction to only 17% in cases with distant metastasis [3].

Cancer survival has improved in the last decades for most cancers except uterine cervix and uterine corpus, reflecting the absence of new advances treatment for these cancers [3]. The most efficient treatment for EC is surgery followed by brachytherapy, radiotherapy and/or chemotherapy, according to the risk stratification of EC patients [26]. Whilst treatment is very effective and it cures most cases having a localized tumor, this is very inefficient in cases of metastases. This means that primary treatment based on uterus removal is efficacious, whereas adjuvant treatment when the tumor has spread the uterus is not well working. Consequently, the increasing incidence and the absence of new treatments explain the rise of death rates. From 1997 to 2008, EC deaths increased 0.3% per year, whereas from 2009 to 2018, it has increased 1.9% per year [6].

Clinic-pathological features are the basis for the risk stratification of EC patients and consequently, this classification determines the surgery and the subsequent therapy for each patient. However, risk stratification system is not accurate because 5% of patients classified as low risk suffer a recurrence [69]. Likewise, around 5-15% of patients classified between intermediate to high-intermediate [74,75], and around 30-40% of patients classified from high to advanced risk will suffer a recurrence [78–80]. The prediction of recurrence on those patients may allow to adjust a more aggressive therapy and/or specific follow-up, which ultimately will increase their overall survival.

In this context, there is a clinical unmeet in the identification of patients that will recur after primary and adjuvant treatment, and the availability of new treatments for those patients at high risk of recurrence, where chemotherapy, the most aggressive therapy, is not able to inhibit tumor evolution.

GENERAL OBJECTIVES

This thesis has been divided in two main objectives. Firstly, the identification of a set of predictive biomarkers of recurrence for different subtypes of endometrial cancer, specifically, for endometrioid EC (EEC) and for serous EC (SEC). The work done in the frame of this objective is summarized from chapter 1 to chapter 4. Secondly, we assessed the toxicity and efficacy of niraparib and SYD985, as monotherapy and as combinatory treatment, in PDX animal models for the personalized treatment of EC patients. This objective is shown in chapter 5.

These approaches are expected to improve the management of EC patients and decrease their mortality and morbidity, thanks to first, the identification of predictive biomarkers of recurrence which will permit to better monitor and adjust the treatment for high-risk EC patients; and second, to provide evidence of new and efficient treatments for specific subsets of high-risk EC.

SPECIFIC OBJECTIVES

To achieve these general objectives, this thesis has been divided in five chapters with the following specific objectives:

CHAPTER 1. Protocol Optimization. Protein extraction from Formalin-Fixed Paraffin-Embedded (FFPE) tissue specimens for Mass Spectrometry analysis.

1.1. Develop a protocol for FFPE tissue processing and evaluate whether the protein extract can be measured by MS-based approaches

1.2. Assess if storage time of FFPE samples can affect protein recovery and protein identification.

1.3. Determine whether macrodissection can improve the enrichment of tumor-specific proteins.

CHAPTER 2. Biomarker Discovery. Identification of EC biomarkers predicting recurrence in FFPE primary tissues using a non-targeted proteomic approach.

2.1. Evaluate the performance of Data-Dependent Acquisition (DDA) in clinical samples and assess the potential of FFPE tissue as a source of EC biomarker in endometrioid and serous histology.

2.2. Identify potential protein biomarkers predicting recurrence in the endometrioid and serous histology, as well as in molecular EC groups.

2.3. Assess the relation of the identified protein biomarkers with clinic-pathological and molecular features.

2.4. Unveil the proteomic landscape of recurrent EC and identify the most relevant pathways involved in recurrence.

CHAPTER 3. Bioinformatics Analysis. In-silico analysis of the TCGA and CPTAC data to identify EC biomarkers predicting recurrence.

3.1. Analyze online databases (TCGA and CPTAC) to discovery potential biomarkers predicting recurrence.

3.2. Assess the relation of the identified protein biomarkers with clinic-pathological and molecular features.

3.3. Compare potential biomarkers from the discovery phase in Chapter 2 and the in-silico analysis in Chapter 3.

CHAPTER 4. Biomarker Verification. Targeted proteomics in FFPE primary EC tumors to verify EC biomarkers predicting recurrence in the endometrioid histology.

4.1. Prioritize a list of potential biomarkers identified in Chapter 2 and 3 for the verification phase.

4.2. Assess the potential to predict recurrence of the prioritized potential biomarkers in an independent cohort of EEC patients using targeted proteomics.

4.3. Assess the relation of the identified protein biomarkers with clinic-pathological and molecular features.

4.4. Develop protein panels to improve the individual performance of the biomarkers predicting recurrence.

CHAPTER 5. Novel Treatments in EC. Preclinical study to evaluate the efficacy of niraparib and SYD985 in HER2-positive endometrial cancer PDX models.

5.1. Assess toxicity of SYD985 and niraparib in EC PDX models.

5.2. Analyze the efficacy of SYD985 and niraparib in monotherapy or in combination in EC PDX models with aberrant expression of *erbb2* (HER2).

RESULTS

CHAPTER 1

PROTOCOL OPTIMIZATION

PROTEIN EXTRACTION FROM FORMALIN-FIXED

PARAFFIN-EMBEDDED (FFPE) TISSUE SPECIMENS

FOR MASS SPECTROMETRY ANALYSIS

SPECIFIC BACKGROUND

In the recent years, MS-based proteomics have emerged as a powerful tool for the identification, verification, and validation of biomarkers. However, it has been limited by the lack of available samples. Proteomics have been conducted in fresh-frozen samples (i.e., tissue, plasma, or proximal fluids), but they are not always available. At this point, FFPE tissue appears as a good alternative because the storage can be at room temperature, is a routinely procedure in all hospitals, and consequently, there are large repositories of accessible samples. Importantly, recurrence in EC is not a frequent event and there is a limited number of patients. Therefore, FFPE specimens are optimal clinical samples for the study of biomarkers in recurrent EC.

Gámez-Pozo et al. [220] were the first to describe a global procedure to evaluate phosphoproteomics from FFPE tissues. For this thesis, this protocol was used and adapted in order to perform proteomics on FFPE EC primary tissues. The work presented in this chapter aimed to i) develop a protocol for FFPE tissue processing and evaluate whether the protein extract can be measured by MS-based approaches; ii) assess if storage time of FFPE samples can affect protein recovery and protein identification; and iii) determine whether macrodissection can improve the enrichment of tumor-specific proteins.

MATERIAL AND METHODS

Patient recruitment

This project was approved by the Ethical Committee for Clinical Evaluation (CEIC) at Vall Hebron (PRAG 446-2020). A total of 3 FFPE tissue were selected from three different EC patients diagnosed in the Vall Hebron Hospital (Barcelona, Spain) in 2001, 2006, and 2012, respectively. The selected FFPE blocks were enriched in tumor content, i.e., tumor cells represented more than 50% of the tissue specimen.

Protein identification by LC-MS analysis

❖ Tissue deparaffinization, protein extraction and quantification

The protocol describe in this section is the one from Gámez-Pozo et al., but the final protocol used for this thesis is explained in Annex 1.

Five to ten sections of 5 μm thickness from each FFPE block were cut according to the percentage of tumor content and were placed into a microcentrifuge tube. Samples were deparaffinized adding 1 mL of xylene, vortex and centrifuged at maximum speed for 2 minutes. Xylene was discarded without disturbing the pellet and the process was repeated. After, 1 mL of ethanol was added to the sample, vortex and centrifuged at maximum speed for 2 minutes. Supernatant was discarded and pellet was dried at room temperature for 15 minutes.

For protein extraction, pellets were resuspended in 100 to 500 μL of buffer extraction (Tris-HCl 40mM – 2% SDS, pH 8.2). The added volume was approximately 1:1 of sample-extraction buffer. This buffer is crucial to extract proteins from cells, and do the denaturation, alkylation, and reduction of proteins during sample preparation. Tubes were heated at 99ºC for 30 minutes with agitation at 750rpm and then, at 80ºC for 2 hours with agitation at 750rpm. Afterwards, samples were centrifuged at room temperature for 20 minutes at 15.000G. Supernatant containing the protein extract was transferred to a new collection tube.

For protein quantification, the DC Protein Assay method (Bio-Rad, reference #5000112) was used following manufacturer instructions. Protein extracts were used directly in the next steps, stored at -20ºC for short-time, or at -80ºC for long-time storage.

❖ Sample preparation for MS analysis

A total of 10 μg of protein extract was diluted 20 times (dilution 1:20) with 50 mM ammonium bicarbonate (NH_4HCO_3). Then, 1 mg/mL of Trypsin Gold (Promega, reference #V5111) was added to the samples and incubated overnight at 37ºC with agitation at 750rpm. Later, HiPPR^{TM} Detergent Removal Resin (Thermo Scientific, reference #88305) was used to eliminate the SDS of the protein extract following manufacturer instructions. Briefly, detergent removal resin was added to the column and centrifuge at 1500G for 1 minute. The flow-through was discarded and PBS was added to the column to clean it. Columns were centrifuged again at 1500G for 1 minute, and this process was repeated 3 times. Next, the protein extract was added to the column, mixed with the resin, and incubated at room temperature for 10 minutes. Finally, columns were centrifuged at 1500G for 2 minutes and the flow-through was transferred to a new collection tube containing the protein extract without SDS.

Proteins must be acidified before LC-MS, so it was added a 10% of volume of 100% concentrated formic acid (CH_2O_2). After that, Ultra MicroSpin^{TM} columns (The Nest

Group, reference SUM SS18V) were used for the desalting. Five different steps were used for this process:

1. Conditioning: 400 μL of methanol were added to the column, were centrifuged 5 minutes at 200G, and the flow-through was discarded.

2. Equilibration: 300 μL of formic acid at 5% were added to the column, were centrifuged 5 minutes at 200G, and the flow-through was discarded. This step was repeated twice.

3. Loading: the protein extract was added to the column and was centrifuged 10 minutes at 100G. Later, the flow-through was charged again to the column, was centrifuged 10 minutes at 100G, and the resulting flow-through was discarded.

4. Washing: 300 μL of formic acid at 5% were added to the column, were centrifuged 5 minutes at 100G, and the flow-through was discarded. This step was repeated twice.

5. Elution: 300 μL of elution buffer (50% acetonitrile and 5% formic acid) were added to the column and were centrifuged 5 minutes at 100G. Later, 300 μL of elution buffer were added again to the column and were centrifuged 5 minutes at 100G (total volume of 600 μL).

At the end of sample preparation: i) proteins have been through denaturation, reduction, and alkylation; ii) proteins are digested by trypsin, an enzymatic protease that cuts proteins in the arginine (R) or lysine (K) amino acid; iii) SDS is extracted from the sample to avoid interferences during MS analysis; and iv) desalting is performed to eliminate possible salts from the sample and increase protein purity.

❖ **LC-MS and data analysis**

An amount of 1 μg of each sample was analyzed by LC-MS using a 60 minutes gradient in the Orbitrap XL (Thermo Fisher Scientific). All data were acquired with Xcalibur™ software v4.1.31.9 (Thermo Fisher Scientific). As a quality control, bovine serum albumin (New England Biolabs, reference P8108S) was digested in parallel and ran between each sample to avoid carryover and to assure stability of the instrument.

Acquired spectra were analyzed using the Proteome Discoverer software suite (v2.0, Thermo Fisher Scientific) and the Mascot search engine (v2.5.1, Matrix Science). The data were searched against a Swiss-Proteome human database (as in October 2018, 20.408 entries). Peptides were filtered based on FDR, and only peptides showing an FDR lower than 5% were retained. The detection and identification of 500 proteins was

used as a reference to consider that the protocol is useful for FFPE processing and LC-MS analysis.

Online data source

CPTAC (Uterine Corpus Endometrial Carcinoma) data was obtained from LinkedOmics database (http://www.linkedomics.org/login.php, accessed on 21 January 2021). The proteome from 95 EC patients, including 12.525 proteins, was used (CPTAC-proteome) to analyze the percentage of tumor-related proteins.

RESULTS

Optimization of 5 sections per sample without restrictions on the time storage of FFPE blocks

Three FFPE blocks were selected from three different EC patients diagnosed in 2001, 2006, and 2012. Their corresponding blocks were obtained from the archives of the Pathology Department of Vall Hebron Hospital and five, ten, and twenty sections were cut from each block. All samples were deparaffinized, protein was extracted, quantified and analyzed by MS. The results from protein quantification are summarized in Table 8.

Table 8. *Protein quantification of 9 samples from 3 different EC patients. EB volume was proportional to the amount of pellet of each sample (EB: Elution Buffer)*

Sample	# Sections	EB Volume (µL)	Protein concentration (µg/µL)	Total protein (µg)
2001	5	250	0.147	366
	10	200	0.242	483
	20	250	0.409	1022
2006	5	200	0.015	30
	10	100	0.026	26
	20	100	0.051	51
2012	5	200	0.733	1466
	10	100	0.511	511
	20	300	1.243	3727

Considering that it is needed 10 μg of total protein for the LC-MS analysis, all conditions provide with sufficient protein amount. The minimum number of sections, 5 sections, already yield enough protein and consequently, this was selected as the preferred condition.

Next, we investigated if there were differences in the number of proteins detected using FFPE tissues generated at different times. For this, we continued the protocol and performed MS analysis only with the samples from 2001 (5 sections) and 2012 (5 sections), which yield similar number of identifications: 452 proteins (sample 2001) and 534 proteins (sample 2012).

Following the results of this study, the protocol for protein extraction was optimized to use 5 sections per sample, and without restrictions on the time storage of the FFPE blocks.

Macrodissection to increase tumor-related proteins

Generally, FFPE blocks from EC contain a percentage of tumor cells but also a percentage of normal cells. This percentage is variable between blocks. The use of the whole FFPE section in biomarker research could mask the identification of tumor biomarkers if the percentage of tumor cells is not controlled. To solve this problem, we tested if an initial step of macrodissection, prior to protein extraction, might enrich on the tumor fraction and avoid the normal component.

To assess this, sample 2012 was prepared in two conditions: a) 5 sections of the whole block; and b) 10 sections mounted on polyethylene naphthalate (PEN) membranes (ZEISS, reference #415190-9041-000) and macrodissection following this procedure:

1. Membranes were sterilized:
    a. Chemistry sterilization: abundant RNase Zap was added to membranes, washed for 30 seconds in miliQ water three times, and incubated at 65ºC for 2 hours.
    b. UV sterilization: crosslinker chamber was used at 1 Joule for 30 minutes.
2. Five to ten sections of each FFPE block were cut and placed into membranes. The number of sections depends on the percentage of tumor present in the FFPE block.
3. Membranes were heated at 65ºC overnight.
4. Deparaffinization: 20 seconds in xylene tray (3 times), 30 seconds in 100% ethanol tray (2 times), 30 seconds in 90% ethanol tray (1 time), and 30 seconds in 70% ethanol tray (1 time).
5. Membranes were dried at room temperature for 20 minutes.

6. Using the H&E from the same block as a template, tumor tissue was cut from membranes using a scalpel, and placed into a collection tube.

After these steps, both conditions followed the protocol previously described with the deparaffinization, protein extraction, protein quantification, and LC-MS analysis. The macrodissected tumor area yielded a total of 869 proteins, whereas the whole area of the FFPE block yield 534 proteins (Figure 21).
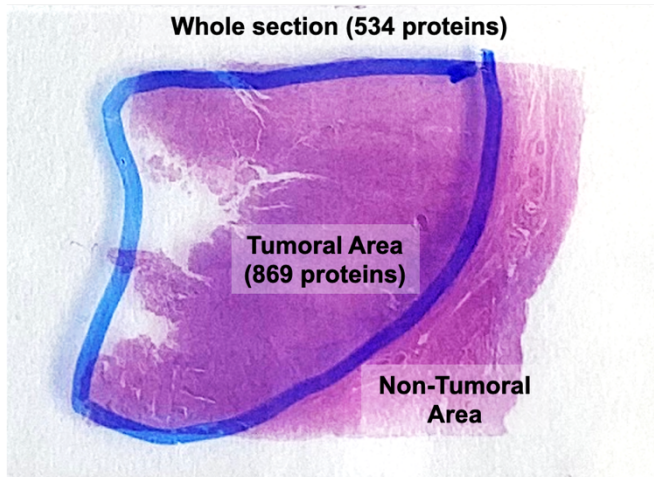
Figura 21. Number of proteins detected through MS analysis using the whole section (534 proteins) or using macrodissection of the tumoral area (869 proteins).

Finally, we compared the proteins detected from both conditions with CPTAC (Clinical Proteomic Tumor Analysis Consortium), an online EC proteomic database. We observed that most proteins from both conditions were also identified in CPTAC (Figure 22). Importantly, macrodissection was able to increase proteins detected in EC patients from 93% to 97%, suggesting that most proteins detected and identified using this technique were tumor-related proteins.
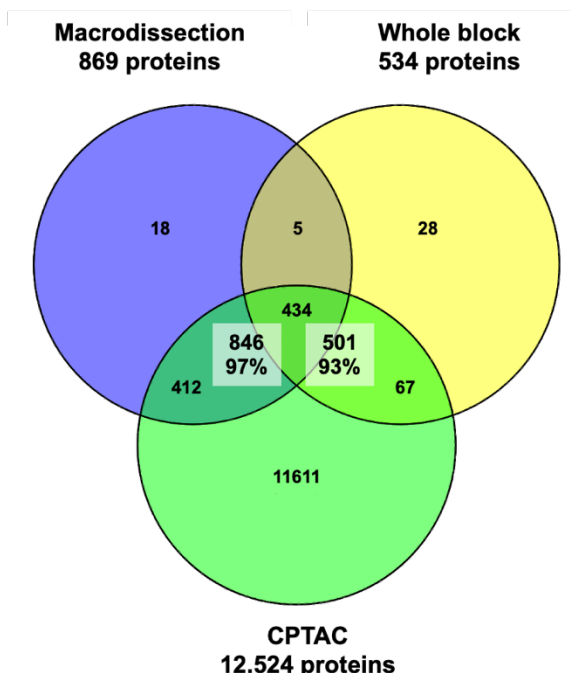
Figura 22. Venn diagram between proteins identified using macrodissection, the whole block, and the EC proteomic database CPTAC. *In the whole block, 501 (93%) proteins were also detected in CPTAC, whereas in the macrodissection process, 846 (97%) proteins were also detected in CPTAC.*

Overview of the modifications done in the FFPE protein extraction protocol to allow LC-MS analysis

The main modifications on the protocol described by Gámez-Pozo et al. are:

- Addition of a macrodissection step prior to deparaffinization. This permitted to increase the identification of tumor-related proteins.

- Reduction of the SDS percentage in the elution buffer from 2% to 1%. SDS is necessary for protein extraction and protein denaturation, but it is important to eliminate from the protein extract before MS because it is a large molecule that could interfere during the analysis.

- Substitution of PBS during detergent elimination for ammonium bicarbonate (ABC). Manufacturer instructions determine the use of PBS for washing during detergent elimination, but it increases the presence of salts in the MS analysis. Therefore, we changed it for ABC, which is the same buffer used to dilute protein extract in the first step of the sample preparation.

Importantly, detergent elimination is a critical step to remove SDS from the protein extract, so it must be done adequately. Commercial kits are based on columns and a resin that retain SDS and allow proteins to be eluted SDS-free. The most important step during this process is the homogenization of the protein extract with the resin. Samples from 2001 and 2012 were processed in duplicates, in order to split protein extracts and perform detergent elimination with and without homogenization. Strikingly, the MS results showed peaks well-defined for homogenized samples, whereas there was no protein detected in the chromatogram of samples without homogenization (Figure 23).
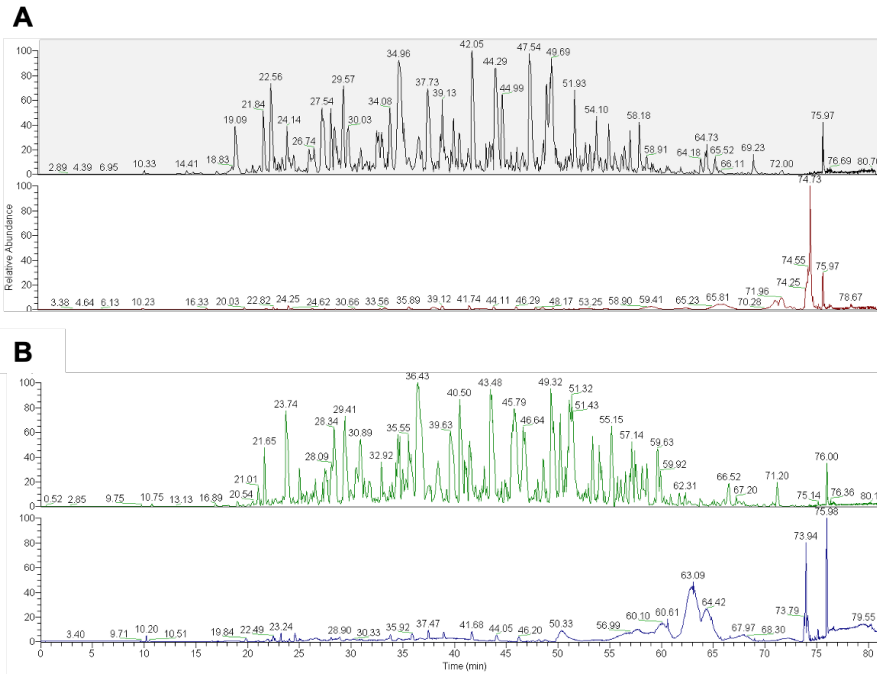
Figura 23. Chromatogram extracted from Xcalibur^{TM} software v4.1.31.9. A. *Sample from 2001. In the top, homogenized sample with well-defined peaks; in the bottom, sample without homogenization and with no protein identification.* *B.* *Sample from 2012. In the top, homogenized sample with well-defined peaks; in the bottom, sample without homogenization and with no protein identification.*

Taking all these considerations, we have described a final protocol in Annex 1 and summarized in Figure 24:
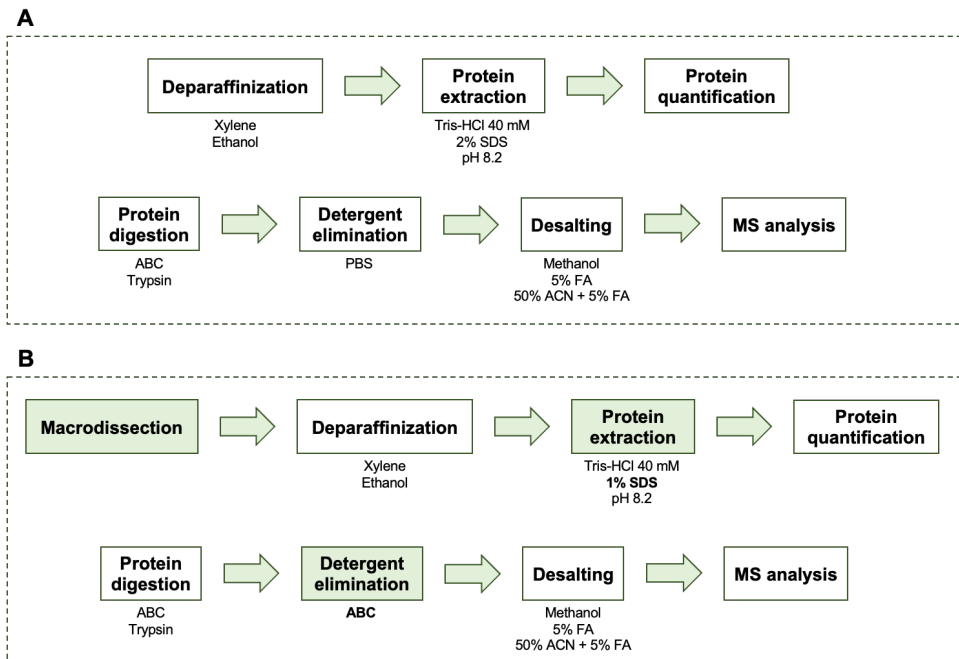
Figura 24. Protocol for MS analysis from FFPE tissues. *A.* *Protocol described by Gámez-Pozo et al. for MS analysis from FFPE tissue using the whole section.* *B.* *Final protocol applied for MS analysis from FFPE tissue using macrodissection as a first-step and changes during protein extraction and detergent elimination (ABC: ammonium bicarbonate; FA: formic acid; ACN: acetonitrile).*

DISCUSSION

Although the prognosis for endometrial cancer is good because of the early diagnosis, approximately 10-20% of tumors recur, and most recurrences take place within 2-3 years [27,28] without means of properly identifying and managing these patients. Researchers working on EC dedicate efforts to find potential EC predictive biomarkers, but they face difficulties to compile large number of cases due to the low ratio of recurrences. At this point, FFPE tissues appear as a suitable alternative. They represent an inexpensive tissue storage system in which samples are stable at room temperature and is a procedure routinely used in hospitals worldwide.

Typically, FFPE tissues have been used in IHQ for determining the distribution of an antigen, or in oncology for making diagnosis, prognosis, and predicting response to therapy. In addition, due to technological and methodological advances, FFPE specimens are becoming a valuable source of DNA, RNA, and proteins for research and therapeutic applications. They can be used for i) transcriptome and gene expression analysis [221]; ii) multiplex immunofluorescence for spatial signatures [222]; (iii) epigenetic profiling [223]; or iv) tumor profiling [224]. Remarkably, MS has emerged as a new tool for the identification and validation of potential biomarkers, and the optimization of protein extraction from FFPE tissues for MS analysis would be beneficial for the identification of new biomarkers at a protein level.

Similar to FFPE, optimal cutting temperature (OCT) compound is an alternative to use long-term preserved samples, although OCT might not be as extended as the FFPE storage. Recently, Valdés et al. [225] demonstrated that a higher number of proteins could be analyzed in the OCT rather than in FFPE samples through MS. However, data from OCT was more dispersed and the data from FFPE had better and more consistent clustering of samples. Importantly, we should consider that formalin induces protein cross-linking and degradation in FFPE samples [226]. Consequently, not all proteins can be studied using this source of samples.

Gámez-Pozo et al. [220] were the first to describe the global procedure for evaluating proteins from FFPE tissues. This protocol was specific to analyze phosphorylation sites and it required specific steps to enrich samples by using immobilized metal affinity chromatography or metal oxide affinity chromatography. Hence, we readapted this protocol to our necessities, and we modified some steps to increase the detection of proteins through the mass analyzer.

Results

In parallel to our optimization, other authors have published protocols for protein extraction in FFPE specimens. Regarding protein extraction, there have been many studies comparing different detergents for the identification of proteins from FFPE. The main detergents used are: Rapigest, PPS, ProteaseMax, PEG 20 000, SDS, urea, or a combination of them. Some investigations have not observed any difference between them [227,228], while other researchers have detected increased number of proteins using a specific detergent. For instance, Dapic et al. [229] compared Rapigest, PPS and ProteaseMax and concluded that PPS was the best method with the highest number of proteins identified and a similar number comparing FFPE with fresh-frozen (FF) samples; Davalieva et al. [230] detected more proteins using Rapigest in comparison to FAST; and Rossouw et al. [231] identified an increased number of proteins excluding PEG 20 000 from the protein extraction buffer and concluded that the method was more reproducible compared to samples processed with the presence of PEG 20 000. These discrepancies in the literature enhanced the necessity of further studies about the composition of the protein extraction buffer.

In our protocol, we used SDS following the protocol from Gámez-Pozo. This molecule was added at 2% in the protein extraction buffer to induce cell and protein denaturation. Nevertheless, its concentration must be reduced because it interferes during mass analysis. Therefore, the reduction from 2% to 1% in the protein extraction buffer and the posterior SDS removal improve the identification of proteins. It has been tested different methods for protein purification including detergent removal, acetone precipitation and formic acid resolubilization, and on-bead digestion, being detergent removal the most efficient and reproducible [232]. Another possibility could be the used of guanidine hydrochloride or a combination of acetonitrile and ammonium bicarbonate for protein extraction. These methods avoid the necessity of SDS removal reducing the number of steps, but the protein extraction yield decreases in comparison to SDS method [233].

In this chapter, we demonstrated that our protocol permits to extract sufficient amount of protein from FFPE EC primary tumors in a quality that is suitable for MS. Regarding the MS analysis, we used a 60 minutes gradient in the Orbitrap XL. The sensitivity and selectivity of this mass spectrometer is limited. It can identify around 1,000 proteins in comparison to new mass spectrometer that can detect up to 5,000 proteins per sample [234]. The acquisition method, DDA or DIA, is also related with the number of proteins identified. However, in this chapter we were focused on the capability to analyze FFPE tissues in MS instead of identifying and quantifying proteins from different samples. Therefore, we performed the analysis in a mass spectrometer that can show results easily and fast.

Importantly, our results showed that time storage did not affect the number of protein identifications, in line with the results of Coscia et al. [227] and Rossouw et al. [232]. They observed that the amount of protein extracted was dependent on block age, but the final number of proteins detected by the mass analyzer was not perturbed. Notably, Vall Hebron Hospital has been collecting FFPE since 1960 with the associated clinical data, solving the problem of the limited number of samples needed to study EC recurrence.

Pathologists perform the final diagnosis of EC using FFPE tissue because this has been proved as the best method for maintaining histology structures, letting them assess FIGO stage, grade, and histology, among other parameters. Biomarker research studies should control the variable amounts of tumor:normal components since this can generate bias in protein identification because proteins from the normal content can mask the expression of proteins from the tumor content. We proved that macrodissection was beneficial as a first step to increase tumor-related proteins. In addition, macrodissection might permit the study of specific zones of the tumor such as MELF (microcystic, elongated and fragmented) pattern of invasion. Another alternative is to perform microdissection, a technique very useful both in the research setting and for molecular testing in FFPE tissue samples. The available technique ranges from simple and inexpensive (manual microdissection) to complex and expensive (laser-capture microdissection). It has been demonstrated that microdissection could enrich of tumor material from mixed tumor-normal samples by up to 67% [235]. Importantly, combined with appropriate settings for LC-MS/MS, it allows the identification of proteins from a single microdissected section [236,237].

In conclusion, in this chapter we have described an optimized protocol to use FFPE tissue in MS and we have shown that FFPE specimens are an unvaluable resource for clinical and biomarker researcher.

Results

CHAPTER 2

BIOMARKER DISCOVERY

IDENTIFICATION OF EC BIOMARKERS PREDICTING
RECURRENCE IN FFPE PRIMARY TISSUES USING A
NON-TARGETED PROTEOMIC APPROACH

SPECIFIC BACKGROUND

The most important cause of death in EC patients is the event of recurrence and thus, tumor dissemination. The risk stratification system allows clinicians to classify patients according to the risk of recurrence in order to adequate surgical and adjuvant treatment, but it has a limited accuracy. Recently, molecular classification has been included as a new parameter to improve the predictive capacity. However, there is still a percentage of cases that recur independently of the risk group that was identified to the patient. Therefore, translational researchers, such as our group, is devoting efforts in the identification of biomarkers to predict EC recurrence.

Proteins are key players in biological processes and variations in their levels can be associated with different pathologies like cancer. Importantly, protein biomarkers are more easily implemented into clinics because they are measured using simple and fast methods which are widely implemented in clinical or pathological laboratories, such as IHQ or ELISA. As shown in chapter 1, we are now able to perform MS-based approaches using FFPE tissues, and this opens an avenue to investigate in novel protein biomarkers to predict EC recurrence.

The work presented in this chapter aimed to i) evaluate the performance of Data-Dependent Acquisition (DDA) in clinical samples and assess the potential of FFPE tissue as a source of EC biomarker in endometrioid and serous histology; ii) identify potential protein biomarkers predicting recurrence in the endometrioid and serous histology, as well as in molecular EC groups; iii) assess the relation of the identified protein biomarkers with clinic-pathological and molecular features; iv) unveil the proteomic landscape of recurrent EC and identify the most relevant pathways involved in recurrence;

MATERIAL AND METHODS

Patient recruitment

A total of 96 patients diagnosed with EC were selected from the Vall Hebron Hospital (Barcelona, Spain) and Arnau de Vilanova Hospital (Lleida, Spain). All patients signed informed consent forms approved by the Ethical Committee for Clinical Investigation (CEIC) at Vall Hebron Hospital and Arnau de Vilanova Hospital. Inclusion criteria was: i) patients diagnosed with endometrioid or serous histology; ii) patients at FIGO stage IA, IB, or II; iii) in case of no recurrence, a clinical follow-up of at least 3 years. Exclusion

criteria was: i) patients diagnosed with other cancer type at the same time; ii) patients without all clinical data information.

❖ Molecular classification

To molecularly classify EC patients, we followed the Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) surrogate system, which is based on evaluation of MMR proteins and P53 expression by IHQ, and analysis of somatic mutations of POLE by DNA sequencing.

We performed IHQ on FFPE tissues from EC patients. We used the same specific block that was used for protein extraction and MS analysis. Hematoxylin-eosin (H&E) and IHQ for MMR proteins and P53 were performed and analyzed blinded. Concretely, IHQ staining was performed in 5 µm FFPE tissue slides using a fully automated system, the Benchmark ULTRA slide Stainer (Ventana Medical Systems, Tucson, AZ). Slides were deparaffinized (EZ prep TM (10x), Ventana Medical Systems), and antigen retrieval was performed incubating slides with cell conditioning (CC1 solution pH 8, Ventana Medical Systems). Afterwards, slides were blocked with hydrogen peroxide solution (3%) and rinsed with reaction buffer (10X, Ventana Medical Systems). Next, slides were incubated with the primary antibody P53 (DO-7, reference #800-2912) at 37ºC for 44 minutes; MLH1 (M1, reference #760-5091), MSH6 (SP93, reference #760-5092) and MSH2 (G219-1129, reference #760—5093) at 37ºC for 40 minutes; and PMS2 (A16-4, reference #760-5094) at 37ºC for 92 minutes. All antibodies were from Ventana Medical Systems. Amplification was done using an Ultra View Polymer Detection Kit (Ventana Medical Systems) with diaminobenzidine as the chromogen. Finally, slides were counterstained with hematoxylin and deparaffinized. In addition, those patients with discordant MMR proteins profile were validated by testing amplifications of the microsatellite biomarkers BAT25, BAT26, NR-21, NR-24, and MONO-27 through the MSI Analysis System Kit (Promega, reference MD3140) following manufacturer instructions.

The interpretation of IHQ staining was performed by an expert pathologist at the Department of Pathology at Vall Hebron Hospital. MMR proteins were identified as aberrant expression when no nuclear staining was detected. Aberrant P53 expression was indicated either by overexpression (nuclei of tumor cell stained with an intensity higher than 75%) or no expression of P53 in the tumor cell nuclei.

DNA extraction was performed in 5 sections of 5 µm from FFPE specimens. Tissues were deparaffinized using xylene and ethanol. Then, DNA was extracted using the

QIAamp® DNA FFPE Tissue Kit (QIAGEN, reference #56404) following manufacturer instructions. POLE mutations were determined by Polymerase Chain Reaction (PCR) amplification of POLE gene (exons 9, 11, 13, and 14). Primers sets used to amplify and sequence exonuclease domain regions are described in Table 9.

Table 9. Primers used to amplify exons 9, 11, 13, and 14 for POLE gene and primers used to sequence exonuclease domain regions.

Primer	Sequence
Exon 9 Fw	5'-TGTAAAACGACGGCCAGTGCCTAATGGGGAGTTTAGAGC-3'
Exon 9 Rv	5'-CAGGAAACAGCTATGACCTACTTCCCAGAAGCCACCTG-3'
Exon 11 Fw	5'-TGTAAAACGACGGCCAGTGAGAAAGAGCAGACCTCTGAC-3'
Exon 11 Rv	5'-CAGGAAACAGCTATGACCCCAGTTACTCATAGAGAAGACACAGA-3'
Exon 13 Fw	5'-TGTAAAACGACGGCCAGTTCTGTTCTCATTCTCCTTCCAG-3'
Exon 13 Rv	5'-CAGGAAACAGCTATGACCCGGGATGTGGCTTACGTG-3'
Exon 14 Fw	5'-TGTAAAACGACGGCCAGTTCTGGCGTTCTCTCCTCAG-3'
Exon 14 Rv	5'-CAGGAAACAGCTATGACCCGACAGGACAGATAATGCTCA-3'
Sequence Fw	5'-TGTAAAACGACGGCCAGT-3'
Sequence Rv	5'-CAGGAAACAGCTATGACC-3'

Additionally, we further validated POLE mutations using KASP technology V4.0 2x. Specifically, we analyzed the five most common hotspots mutation sites: P286R, S297F, V411L, A456P, and S459 (GC Genomics/Bioscience, reference KBS-1016-021).

Protein identification by LC-MS analysis

❖ Sample preparation

The global procedure for macrodissection, tissue deparaffinization, protein extraction and quantification, and sample preparation for LC-MS analysis was performed as explained in Chapter 1 and described in Annex 1.

❖ LC-MS analysis

A total amount of 1 µg of each sample was analyzed by LC-MS using a 90 minutes gradient in the LTQ-Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-nLC 100 (Thermo Fisher Scientific). Peptides were loaded directly into 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 Fisher Scientific).

Chromatographic gradients started at 95% buffer A and 5% buffer B with a flow rate of 300 mL/min for 5 minutes, gradually increased to 22% buffer B and 78% buffer A in 79 minutes, and to 35% buffer B and 65% buffer A in 11 minutes. After each analysis, the column was washed for 10 minutes of 10% buffer A and 90% buffer B. Concretely, buffer A was 0.1% formic acid in water, and buffer B was 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 275ºC. Ultramark 1621 was used for external calibration of the mass analyzer prior the analysis, 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 mass analyzer. Auto gain control (AGC) was set to 1E5, and charge state filtering disqualifying singly charged peptides was activated. In each cycle of DDA analysis, following each survey scan, the most intense ions above a threshold ion count of 10,000 were selected for fragmentation. 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 were produced via high-energy collision dissociation (HCD) at normalized collision energy of 28% and they were acquired in the ion trap mass analyzer. All data were acquired with Xcalibur™ software v4.1.31.9 (Thermo Fisher Scientific).

Digested bovine serum albumin (New England Biolabs, reference P8108S) was analyzed between each sample to avoid sample carryover and to assure stability of the instrument, and QCloud was used to control instrument longitudinal performance during the project.

❖ **Data analysis**

Acquired spectra were analyzed using the Proteome Discoverer software suite (v2.0, Thermo Fisher Scientific) and the Mascot search engine (v2.6, Matrix Science). The data were searched against a Swiss-Proteome human database (as in October 2018, 20,408 entries) plus a list of common contaminants and all the corresponding decoy entries. For peptide identification, a precursor ion mass tolerance of 7 ppm was used for MS1 level, trypsin was chosen as enzyme, and up to three missed cleavages were allowed. The fragment ion mass tolerance was set to 0.5 Da for MS2 spectra. Oxidation of methionine and N-terminal protein acetylation were used as variable modifications, whereas carbamidomethylation on cysteines was set as a fixed modification. False discovery rate (FDR) in peptide identification was set to a maximum of 5%.

Peptide quantification data were retrieved from the "Precursor ion area detector" node from Proteome Discoverer (v2.0) using 2 ppm mass tolerance for the peptide extracted ion chromatogram (XIC). Protein abundance in each condition was estimated using the average of the three most intense peptides per protein group. The obtained values were used to calculate protein fold-changes and their corresponding adjusted p-values.

Statistical analysis

The statistical analysis was performed in R software (v.4.2.0) and Graph Pad Prism (v.8.2.1) (GraphPad Software, La Jolla, CA, USA). Heatmap was done to classify patients in differential clusters to determine the relation between the proteomic landscape and clinic-pathological and molecular features. The correlation between clinic-pathologic characteristics and recurrence was calculated using Fisher's exact test for 2 variables and Chi-square for 3 or more variables. Two different approaches were used for the analysis of significantly altered proteins:

- Quantitative analysis was performed only in those protein detected in more than 80% of patients in both groups. Due to the non-normality of the data, comparison of the abundance protein between recurrence and non-recurrence was done using the non-parametric Mann-Whitney U test. Adjusted p-value lower than 0.05 and/or log fold change (FC) higher than 1 were considered statistically significant.
- The absence/presence analysis was done using Fisher's exact test. P-values lower than 0.05 were considered statistically significant.

Principal component analysis (PCA) was carried out to determine general differences between recurrence and non-recurrence population. Receiver operating characteristic (ROC) curves were used to calculate the relationship between sensitivity and specificity for each EC biomarker candidate. Significant proteins were used for an enrichment analysis to identify relevant pathways involved in recurrence. It has been performed 4 different enrichment analysis using different platforms: Gene Set Enrichment Analysis (GSEA), G-Profiler, Reactome (https:///www.reactome.org/), and DAVID (https://david.ncifcrf.gov/).

RESULTS

Proteomic landscape of EC patients in the discovery phase

A total of 96 patients diagnosed with EC were selected from the Vall Hebron Hospital (Barcelona, Spain) and Arnau de Vilanova Hospital (Lleida, Spain). Precisely, 48 EC patients with recurrence and 48 EC patients without recurrence. These patients were selected retrospectively to carry out a differential study between EEC and SEC. All patients were classified according to histological features (histology, grade, FIGO stage, myometrial invasion, and LVSI), molecular features (POLE, MSI, low-CN, and high-CN), and using this parameters, they were also classified into the risk stratification system.

A specific FFPE block with high tumoral content was chosen for every EC patient and, following the protocol explained in Chapter 1, protein was extracted and prepared for MS analysis. Proteomic analysis allows us to identify 107 proteins differentially expressed between EEC and SEC with an adjusted p-value < 0.05, and they were used to classify patients in different clusters through a heatmap (Figure 25).
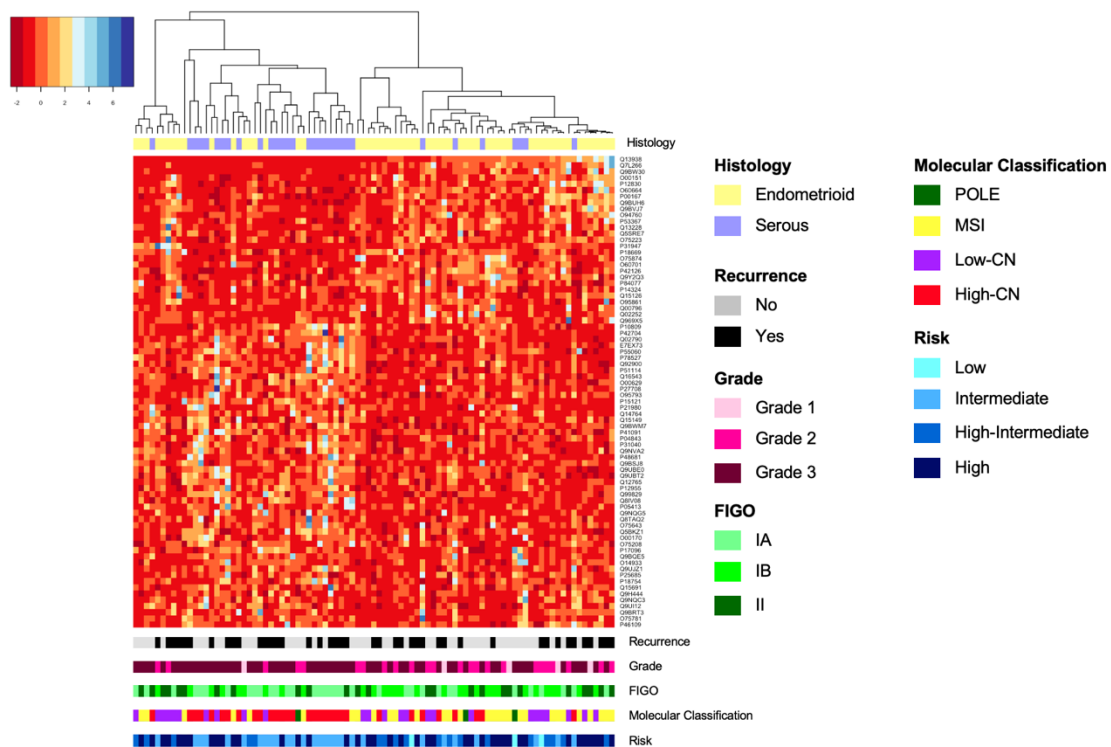
Figura 25. Heatmap of 96 patients enrolled in the discovery phase. *Patients are classified according to the histology (Endometrioid or Serous), recurrence (Yes or No), grade (1, 2, or 3), FIGO stage (IA, IB, or II), molecular classification (POLE, MSI, Low-CN, or High-CN), and risk stratification system (Low, Intermediate, High-Intermediate, or High).*

As expected, the most important feature to differentiate EC patients was histology. There was a clear difference between endometrioid and serous histology into two separated clusters. It was also related with the grade because of the inherent classification of serous histology as a grade 3. Importantly, molecular classification was able to distinguish high-CN from the other groups, but not between POLE, MSI and low-CN.

Due to the great molecular differences between EEC and SEC histologies, the subsequent sections to identify biomarkers predicting recurrence was performed independently for each histology.

Endometrioid EC patients of the discovery phase

A total of 63 patients were selected for the identification of biomarkers predicting recurrence in the endometrioid histology. The cohort was age-balanced and divided in non-recurrent and recurrent EEC patients. Specifically, 32 EEC patients were non-recurrent after at least 3 years from the primary surgery, with a mean follow-up of 68 months (45 – 151); whilst 31 EEC patients had at least one recurrence event. Recurrences were mainly located at regional or distant organs. The mean time to recur was 33 months (1 – 94) and the mean number of recurrences per patient was 1.2 (1 – 2). At the time that this study was performed, 45% of recurrent patients died due to the disease, with a mean overall survival of 60 months (7 – 152). There were no other statistical differences in any of the clinic-pathological features (grade, FIGO stage, myometrial invasion, LVSI, molecular classification, and risk classification) in this patient cohort (Table 10).

Biomarkers predicting recurrence in Endometrioid EC patients

One FFPE tissue from each EEC patient was used for protein extraction and MS analysis following data-dependent acquisition (DDA) strategy. A total of 4.569 proteins were detected comparing recurrence versus non-recurrence patients. For the identification of differentially expressed proteins it was followed two different approaches as explained in the methods section.

On the one side, the quantitative analysis was performed with the subset of 1.646 proteins that were detected in more than 80% of patients in both groups. A total of 252 proteins were differentially expressed with an adjusted p-value lower than 0.25. Among them, there were 21 proteins with an adjusted p-value lower than 0.05 and 4 proteins with a log fold change (FC) higher than 1. All proteins from quantitative analysis were

used for the Principal Component Analysis (PCA) and the volcano plot (Figure 26). From the PCA we were able to distinguish the two groups: recurrent EEC patients (red dots) and non-recurrent EEC patients (blue dots).

Table 10. *Clinic-pathological characteristics of women diagnosed with endometrioid EC enrolled in the discovery phase.*

	NO Recurrence	Recurrence	P-value
Total Number	32	31	
Recruitment Hospital			>0.99
Vall Hebron	24	24	
Arnau de Vilanova	8	7	
Age (years)	68 (38-85)	69 (52-85)	0.94
Grade			0.90
1	2	3	
2	14	12	
3	16	16	
FIGO Stage			0.72
IA	8	7	
IB	14	13	
II	10	11	
Myometrial Invasion			0.11
<50%	15	8	
>50%	17	23	
LVSI			0.79
NO	21	19	
YES	10	11	
Molecular Classification			0.25
POLE	2	0	
MSI	15	15	
Low-CN	13	11	
High-CN	2	5	
Risk Classification			0.60
Low	2	1	
Intermediate	11	10	
High-Intermediate	18	19	
High	1	1	
Recurrence Site			N/A
Local	-	1	
Regional	-	10	
Distant	-	20	
Time to Recurrence			N/A
Early (<24 months)	-	20	
Late (>24 months)	-	11	
Status			<0.0001
Alive without disease	30	5	
Alive with disease	0	8	
Dead of disease	0	14	
Dead other causes	2	2	
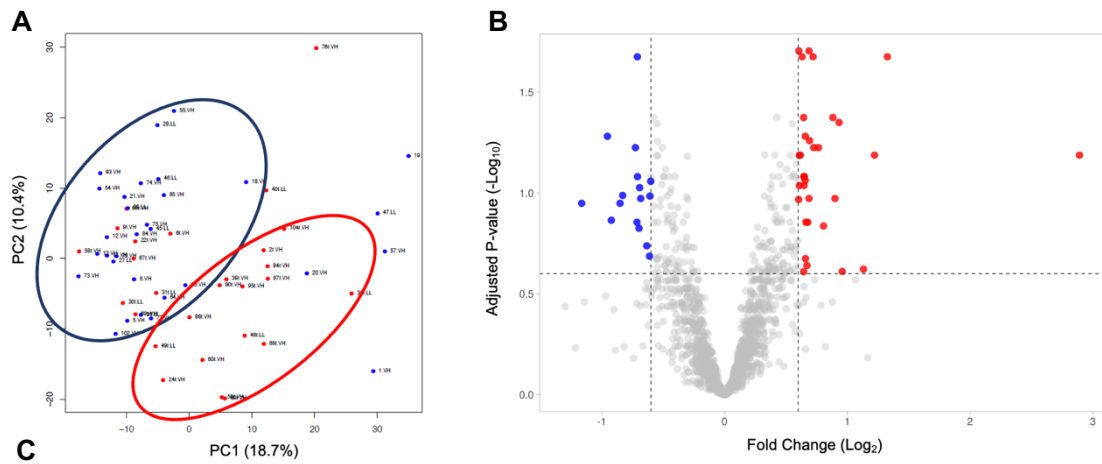
Protein name	Gene name | Protein ID	logFC	P.Value	adj.P.Val	AUC
Myosin-9	MYH9 | MYH9	-0,71	1,95E-06	0,001	0,859
Thiosulfate:glutathione sulfurtransferase	TSTD1 | TSTD1	1,39	1,31E-06	0,001	0,883
40S ribosomal protein S14	RPS14 | RPS14	0,65	1,68E-05	0,008	0,821
Small nuclear ribonucleoprotein F	SNRPF | RUXF	0,62	2,15E-05	0,008	0,755
Alpha-actin-4	ACTN4 | ACTN4	-0,61	0,0001	0,033	0,768
Polymeric immunoglobulin receptor	PIGR | PIGR	2,77	0,0001	0,033	0,668
Serine/arginine-rich splicing factor 2	SRSF2 | SRSF2	0,67	0,0001	0,033	0,804
Chloride intracellular channel protein 4	CLIC4 | CLIC4	-0,60	0,0002	0,033	0,780
Dipeptidyl peptidase 2	DPP7 | DPP2	1,03	0,0002	0,033	0,686
Lysosomal alpha-glucosidase	GAA | LYAG	0,80	0,0002	0,033	0,700
Nuclear ubiquitous casein and cyclin-dependent kinase 1	NUCKS1 | NUCKS	0,68	0,0002	0,033	0,814
Phosphomevalonate kinase	PMVK | PMVK	0,68	0,0002	0,033	0,752
Transcriptional activator protein Pur-beta	PURB | PURB	0,62	0,0002	0,033	0,769
DNA-directed RNA polymerases I, II, and III	POLR2H | RPAB3	0,61	0,0003	0,034	0,808
Talin-1	TLN1 | TLN1	-0,57	0,0003	0,038	0,791
Protein-glutamine gamma-glutamyltransferase 2	TGM2 | TGM2	-0,78	0,0004	0,040	0,761
Alpha-actin-1	ACTN1 | ACTN1	-0,63	0,0005	0,046	0,796
Calponin-3	CNN3 | CNN3	-0,67	0,0005	0,046	0,733
Collagen alpha-1(XVIII) chain	COL18A1 | COIA1	0,77	0,0005	0,046	0,729
Serine/arginine-rich splicing factor 7	SRSF7 | SRSF7	0,63	0,0005	0,046	0,747
Guanylate kinase	GUK1 | KGUA	0,52	0,0006	0,046	0,762
26 proteasome non-ATPase regulatory subunit 3	PSMD3 | PSMD3	-1,05	0,0009	0,058	0,759
Guanylate-binding protein 1	GBP1 | GBP1	-1,02	0,0018	0,075	0,830
Clusterin	CLU | CLUS	1,09	0,0046	0,096	0,668
Carbonic anhydrase 1	CA1 | CAH1	-1,28	0,011	0,148	0,754

Figura 26. Results from the quantitative analysis in EEC histology. A. Principal Component Analysis (blue dots are non-recurrent EC patients and red dots are recurrent EC patients) from the quantitative analysis. B. Volcano plot from the quantitative analysis (red dots are proteins upregulated in recurrent EC patients and blue dots are proteins downregulated in recurrent EC patients). In the x axis, Log$_2$ FC $\pm$ 0.6 symbolize FC $\pm$ 1.5; in the y axis, Log$_{10}$ Adjusted p-value > 0.6 symbolize adjusted p-value < 0.25. C. Twenty-one proteins with an adjusted p-value < 0.05 and four proteins with a FC > 4

On the other side, the absence/presence analysis was performed with all the 4.569 proteins. A total of 168 proteins were differentially expressed with a p-value lower than 0.05 and, among them, 37 proteins were differentially expressed with a p-value lower than 0.01 (Table 11).

Table 11. Results from the absence/presence analysis in EEC histology. Thirty-seven proteins differentially expressed with a p-value lower than 0.01

Protein name	Gene name | Protein ID	NO REC (32)	REC (31)	P.Value
Tropomyosin alpha-1 chain	TPM1 | TPM1	72% (23)	35% (11)	< 0,01
Pyruvate kinase	PKM | KPYM	34% (11)	3% (1)	< 0,01
Putative adenosyl homocysteinase 2	AHCYL1 | SAHH2	66% (21)	29% (9)	< 0,01
Retinal dehydrogenase 2	ALDH1A2 | AL1A2	69% (22)	29% (9)	< 0,01
26S proteasome non-ATPase regulatory subunit 7	PSMD7 | PSMD7	84% (27)	48% (15)	< 0,01
Serine/threonine-protein kinase 4	STK4 | STK4	63% (20)	23% (7)	< 0,01
Beta-2-syntrophin	SNTB2 | SNTB2	53% (17)	16% (5)	< 0,01
Pro-interleykin-16; Interleukin-16	IL16 | IL16	78% (25)	42% (13)	< 0,01
Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL | NSDHL	56% (18)	19% (6)	< 0,01
Fermitin family homolog 3	FERMT3 | URP2	81% (26)	42% (13)	< 0,01
Septin 6	SEPT6 | SEPT6	38% (12)	6% (2)	0,01
FERM, RhoGEF and pleckstrin domain-containing protein 1	FARP1 | FARP1	59% (19)	26% (8)	0,01
NADH dehydrogenase flavoprotein 2	NDUFV2 | NDUFV2	100% (32)	81% (25)	0,01
28S ribosomal protein S22	MRPS22 | MRPS22	91% (29)	61% (19)	0,01
Ubiquitin domain-containing protein UBFD1	UBFD1 | UBFD1	72% (23)	39% (12)	0,01
Lethal giant larvae protein homolog 2	LLGL2 | LLGL2	44% (14)	13% (4)	0,01
Lon protease homolog	LONP1 | LONP1	63% (20)	26% (8)	0,01
Lysosomal alpha-mannosidase	MAN2B1 | MA2B1	100% (32)	81% (25)	0,01
Laminin subunit alpha-5	LAMA5 | LAMA5	3% (1)	29% (9)	0,01
Probable 18S rRNA methyltransferase	WBSCR22 | BUD23	34% (11)	71% (22)	0,01
Hemopexin	HPX | HEMO	97% (31)	74% (23)	0,01
Alcohol dehydrogenase 1A	ADH1A | ADH1A	47% (15)	13% (4)	0,01
Biglycan	BGN | PGS1	97% (31)	74% (23)	0,01
Leiomodin-1	LMOD1 | LMOD1	59% (19)	23% (7)	0,01
Alpha-2-macroglobulin receptor-associated protein	LRPAP1 | AMRP	97% (31)	74% (23)	0,01
Radixin	RDX | RADI	97% (31)	74% (23)	0,01
Dynamin-2	DNM2 | DYN2	91% (29)	61% (19)	0,01
Transcription factor A	TFAM | TFAM	100% (32)	81% (25)	0,01
Bcl-2-like protein 1	BCL2L1 | BCL2L	84% (27)	52% (16)	0,01
26S proteasome non-ATPase regulatory subunit 6	PSMD6 | PSMD6	88% (28)	58% (18)	0,01
ATP-binding cassette sub-family F member 1	ABCF1 | ABCF1	91% (29)	61% (19)	0,01
Zinc finger CCCH domain-containing protein 15	ZC3H15 | ZC3HF	94% (30)	68% (21)	0,01
CCA tRNA nucleotidyltransferase 1	TRNT1 | TRNT1	66% (21)	29% (9)	0,01
Programmed cell death protein 10	PDCD10 | PDC10	88% (28)	58% (18)	0,01
BAG family molecular chaperone regulator 5	BAG5 | BAG5	22% (7)	0% (0)	0,01
UPF0568 protein	C14orf166 | RTRAF	97% (31)	71% (22)	0,01
Nck-associated protein 1	NCKAP1 | NCKP1	59% (19)	26% (8)	0,01

Apart from the general analysis, we also did sub-analysis considering clinic-pathologic features from EC patients. We compared recurrent EEC patients versus non-recurrent EEC patients considering stage, grade, LVSI and risk classification. For instance, we specifically selected those patients classified with FIGO Stage IA and compared 8 EEC patients without recurrence versus 7 EEC patients with recurrence. In addition, we also compared recurrent EEC patients between them considering the time to recur and the final status for the EC patient. The number of proteins differentially expressed for each analysis is summarized in Table 12 and all the comparisons are shown in Annex 2.

Table 12. *Number of proteins differentially expressed with an adjusted p-value < 0.25 and < 0.05 for each sub-analysis and percentage of proteins from each sub-analysis that was also identified in the general comparison (NO REC: no recurrence; REC: recurrence; Early REC: recurrence before 24 months; Late REC: recurrence after 24 months; DOD: dead of disease; AWD: alive with disease; AWTD: alive without disease).*

		Group 1	Group 2	Adjusted p-value < 0.25 (% in GENERAL)	Adjusted p-value < 0.05 (% in GENERAL)
GENERAL	NO REC – REC	32 NO REC	31 REC	252	21
STAGE	Stage IA	8 NO REC	7 REC	0	0
	Stage IB	14 NO REC	13 REC	0	0
	Stage II	10 NO REC	11 REC	295 (58%)	31 (90%)
GRADE	Grade 1-2	16 NO REC	15 REC	29 (86%)	0
	Grade 3	16 NO REC	16 REC	5 (100%)	0
LVSI	LVSI positive	21 NO REC	19 REC	74 (51%)	3 (33%)
	LVSI negative	10 NO REC	11 REC	4 (100%)	0
RISK	High-Intermediate risk	12 NO REC	8 REC	0	0
	High risk	15 NO REC	17 REC	55 (95%)	2 (100%)
TIME TO REC	Early REC – Late REC	20 early	11 late	31 (42%)	1 (100%)
	NO REC – Early REC	32 NO REC	20 early	0	0
	NO REC – Late REC	32 NO REC	11 late	438 (65%)	154 (75%)
REC SITE	Regional – Distant REC	10 regional	20 distant	0	0
STATUS	NO REC – DOD	32 NO REC	14 DOD	91 (16%)	0
	AWD – DOD	8 AWD	14 DOD	2 (0%)	1 (0%)
	NO REC – AWD	32 NO REC	8 AWD	1 (0%)	0
	AWTD – DOD	35 AWTD	14 DOD	29 (97%)	2 (100%)

Interestingly, the highest number of differentially expressed proteins with an adjusted p-value < 0.05 were identified in the comparisons with more aggressive clinic-pathological characteristics (stage II, LVSI positive, and high risk), and between EC patients without

recurrence and EC patients with late recurrence (more than 24 months). Finally, we compared differentially expressed proteins from each sub-analysis with the general comparison (NOREC vs REC) and we observed that most proteins were also statistically significant for the general comparison.

Altered pathways in Endometrioid EC recurrent patients

We selected all proteins statistically significant (adjusted p-value < 0.25) from the general comparison between recurrence and non-recurrence group to do an enrichment analysis and determine the most relevant pathways involved in recurrence. A total of 252 proteins were included in the analysis using different software: David, Reactome, G-Profiler, and Gene Set Enrichment Analysis (GSEA). Importantly, David, Reactome, and G-Profiler do not consider protein expression level, while Gene Set Enrichment Analysis (GSEA) consider p-value / adjusted p-value and fold change.

From the GSEA analysis, we observed only 16 pathways activated with an adjusted p-value < 0.05, meaning the low significance of these group. In contrast, we found 195 pathways suppressed with an adjusted p-value < 0.05, including 51 pathways related with adhesion and binding. Next, we compared the 10 most important activated and the 10 most important suppressed pathways in the recurrence group from the GSEA analysis (Figure 27) with all the pathways statistically significant from the three others analysis (David, Reactome, and G-Profiler). On the one hand, from the activated there were 5 pathways represented at least in two out of the four analyses: mRNA processing (GO:0006397), mRNA splicing via spliceosome (GO:0000398), structural constituent of ribosome (GO:0003735), RNA splicing (GO:0008380), and U2-type precatalytic spliceosome (GO:0071005). On the other hand, from the suppressed there were 5 pathways represented at least in two out of the four analyses: nucleotide binding (GO:0000166), nucleoside phosphatase binding (GO:1901265), actin binding (GO:0003779), cell adhesion (GO:0007155), and purine nucleotide binding (GO:0017076).
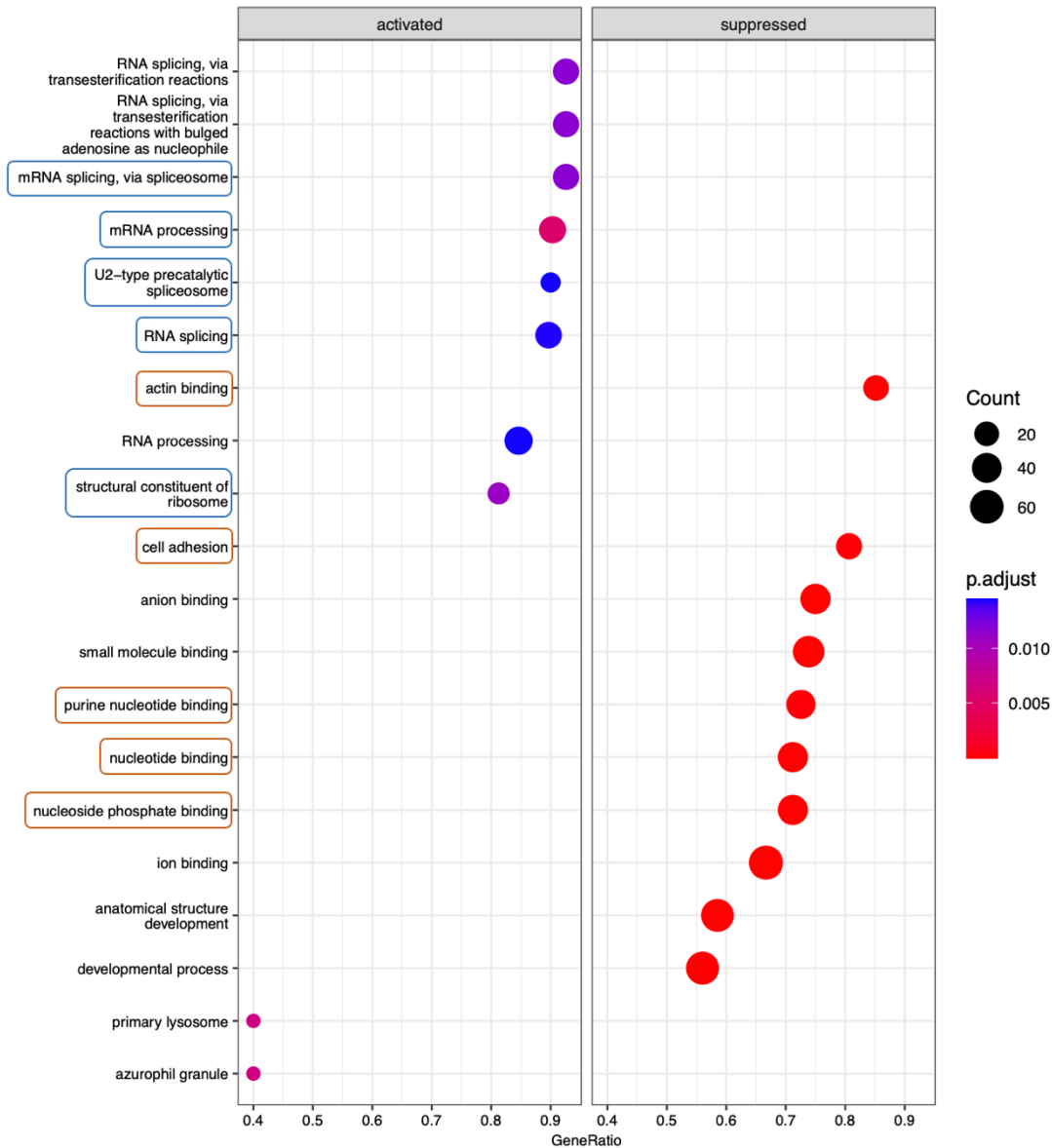
Figura 27. Gene Set Enrichment Analysis (GSEA) in Endometrioid Endometrial Cancer (EEC) patients. *Dot plot representing the top 10 activated and suppressed pathways in the recurrence group. Activated pathways in blue were also detected in the others enrichment analysis, while suppressed pathways in red were also detected in the others enrichment analysis*

In summary, in recurrent EEC patients we detected an activation of pathways related with RNA processing and splicing, and a suppression of pathways related to adhesion and binding**.** These results highlighted that adhesion and binding pathways were suppressed in the primary tumor of EEC patients that will recur.

Serous EC patients of the discovery phase

A total of 33 patients were selected for the identification of biomarkers predicting recurrence in the serous histology. The cohort was divided in non-recurrent and recurrent

SEC patients. Specifically, 16 SEC patients were non-recurrent after at least 3 years from the primary surgery, with a mean follow-up of 93 months (45 – 213); whilst 17 SEC patients had at least one recurrence event. Recurrences were mainly located at regional or distant organs. The mean time to recur was 19 months (1 – 32) and the mean number of recurrences per patient was 1.3 (1 – 3). At the time that this study was performed, 82% of recurrent patients died due to the disease, with a mean overall survival of 42 months (9 – 148). Importantly, 64% of patients classified as high-CN were in the recurrent group, while 86% of patients classified as POLE, MSI or low-CN were in the non-recurrent group. There were no other statistical differences in any of the clinic-pathological features (grade, FIGO stage, myometrial invasion, LVSI, and risk classification) in this patient cohort (Table 13).

Biomarkers predicting recurrence in Serous EC patients

One FFPE tissue from each SEC patient was used for protein extraction and MS analysis following data-dependent acquisition (DDA) strategy. A total of 5.747 proteins were detected comparing recurrence versus non-recurrence patients. For the identification of differentially expressed proteins it was followed two different approaches as explained in the methods section.

On the one side, the quantitative analysis was performed with the subset of 1.710 proteins that were detected in more than 80% of patients in both groups. Any protein was differentially expressed with an adjusted p-value lower than 0.25. All proteins from quantitative analysis were used for the Principal Component Analysis (PCA) and the volcano plot (Figure 28A-B). From the PCA we were not able to distinguish recurrent SEC patients (red dots) from non-recurrent SEC patients (blue dots), representing these few numbers of proteins differentially expressed between both groups.

On the other side, the absence/presence analysis was performed with all the 5.747 proteins. A total of 27 proteins were differentially expressed with a p-value lower than 0.05 and, among them, only 1 protein was differentially expressed with a p-value lower than 0.01 (Figure 28C). This was the most promising biomarker candidate, guanidine deaminase (GUAD) for the serous histology. It was detected in 12 SEC patients without recurrence (75%) and only in 3 SEC patients with recurrence (18%).

Table 13. Clinic-pathological characteristics of women diagnosed with serous EC enrolled in the discovery phase.

	NO Recurrence	Recurrence	P-value
Total Number	16	17	
Recruitment Hospital			0.25
Vall Hebron	13	10	
Arnau de Vilanova	3	7	
Age (years)	69 (48-91)	75 (57-88)	0.07
Grade			N/A
1	0	0	
2	0	0	
3	16	17	
FIGO Stage			0.72
IA	10	10	
IB	2	5	
II	4	2	
Myometrial Invasion			0.43
<50%	13	11	
>50%	3	6	
LVSI			>0.99
NO	12	13	
YES	3	4	
Molecular Classification			0.03
POLE	1	0	
MSI	3	1	
Low-CN	2	0	
High-CN	9	16	
Risk Classification			0.59
Low	1	0	
Intermediate	9	10	
High-Intermediate	0	0	
High	6	7	
Recurrence Site			N/A
Local	-	0	
Regional	-	10	
Distant	-	6	
Time to Recurrence			N/A
Early (<24 months)	-	13	
Late (>24 months)	-	4	
Status			0.008
Alive without disease	12	1	
Alive with disease	0	2	
Dead of disease	0	14	
Dead other causes	4	0	
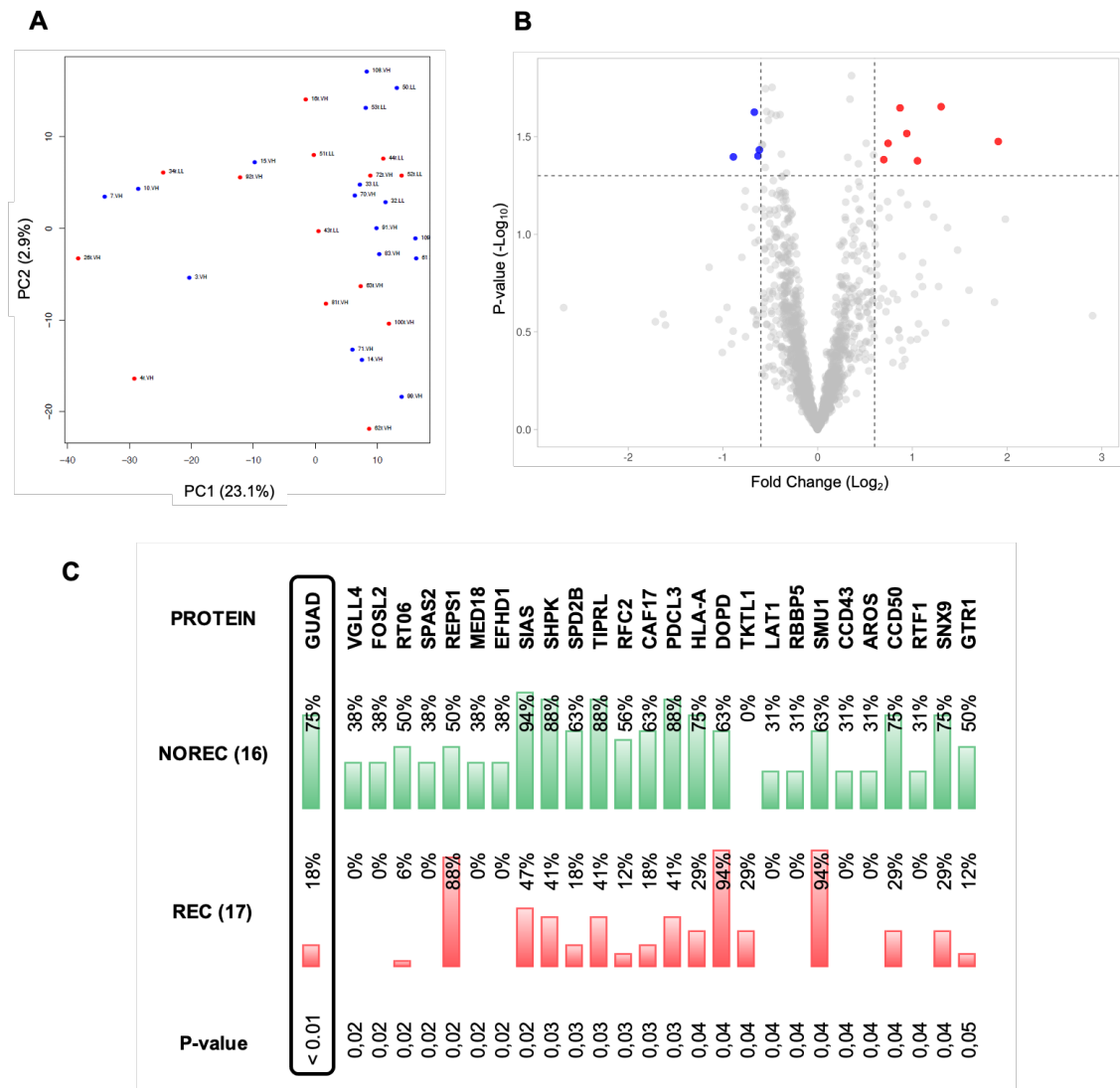
Figura 28. Results from the discovery in SEC histology. A. Principal Component Analysis (blue dots are non-recurrent EC patients and red dots are recurrent EC patients) from the quantitative analysis. B. Volcano plot from the quantitative analysis (red dots are proteins upregulated in recurrent EC patients and blue dots are protein downregulated in recurrent EC patients). In the x axis, Log2 FC ± 0.6 symbolize FC ± 1.5; in the y axis, Log10 P-value > 1.3 symbolize p-value < 0.05. C. Differentially expressed proteins from the absence/presence analysis.

Altered pathways in Serous EC recurrent patients

We selected all proteins statistically significant (in this case a p-value < 0.05) from the general comparison between recurrence and non-recurrence group to do an enrichment analysis and determine the most relevant pathways involved in recurrence. A total of 47 proteins were included in the analysis using different software: David, Reactome, G-Profiler, and GSEA.

From the GSEA analysis, we observed only 6 pathways activated and 18 pathways suppressed with an adjusted p-value < 0.05. Next, we compared the 6 activated and the

10 most important suppressed pathways in the recurrence group from the GSEA analysis (Figure 29), with all the pathways statistically significant from the three others analysis (David, Reactome, and G-Profiler). On the one hand, from the activated pathways, all were represented at least in two out of the four analyses: extracellular region (GO:0005576), extracellular space (GO:0005615), extracellular organelle (GO:0043230), extracellular membrane-bounded organelle (GO:0065010), extracellular exosome (GO:0070062), extracellular vesicle (GO:1903561). On the other hand, from the suppressed there were 3 pathways represented at least in two out of the four analyses: nucleus (GO:0005634), cytosol (GO:0005829), and RNA binding (GO:0003723).
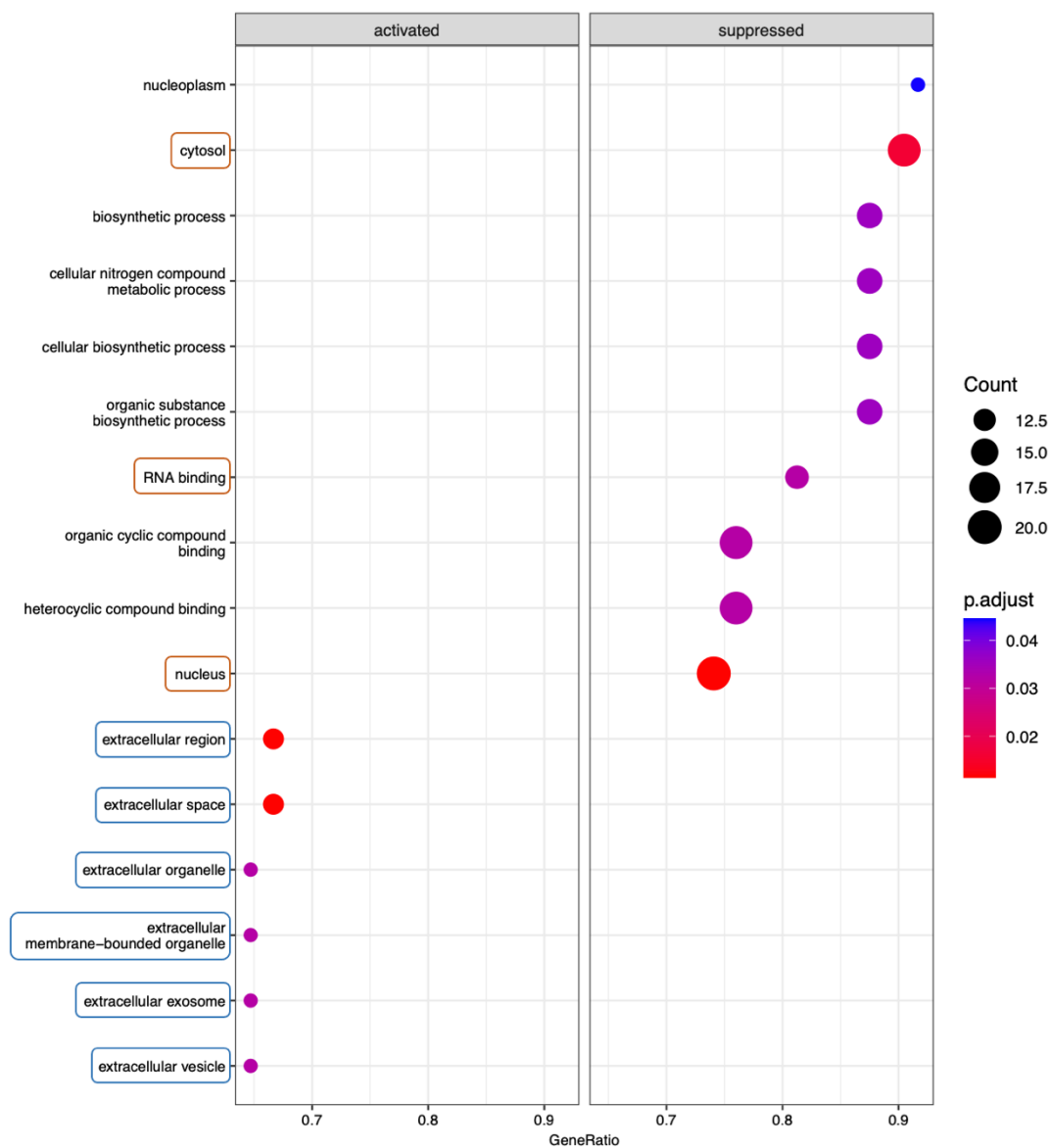
***Figura 29. Gene Set Enrichment Analysis (GSEA) in Serous Endometrial Cancer (SEC) patients.*** *Dot plot representing the top 6 activated and top 10 suppressed pathways in the recurrence group. Activated pathways in blue were also detected in the others enrichment analysis, while suppressed pathways in red were also detected in the others enrichment analysis.*

Overall, recurrent SEC patients presented in the primary tumor a higher number of proteins in the extracellular component, and fewer proteins in the intracellular component (cytosol and nucleus). These results were not very robust because of the low number of proteins selected in the analysis, and the pathways extracted from the analysis were mainly related to cellular component.

Biomarkers predicting recurrence in molecularly classified EC patients

By the time that this study was initiated, the dualistic classification was the most standardized and used EC classification in clinical routine. This is the main reason why the study design focused on the identification of biomarkers predicting recurrence in the most frequent histologies: endometrioid and serous ECs. However, in the recently published clinical guidelines, molecular classification has been incorporated as an optional method for EC classification contributing to the risk classification system.

In this study, we also molecularly classified all EC patients using the ProMisE surrogate system. Briefly, we performed the IHQ of MMR proteins (MSH2, MSH6, MLH1, and PMS2) and P53, and we analyze the POLE gene by KASP genotyping of the top 5 hotspots mutations (P286R, S297F, V411L, A456P, and S459) and DNA sequencing of exons 9, 11, 13, and 14 (Figure 30). We were able to classify 99% (95/96) of EC patients, being 3% POLE, 36% MSI, 27% low-CN, and 34% high-CN (Table 14). We could not classify one patient of the non-recurrence group due to the limited FFPE material.
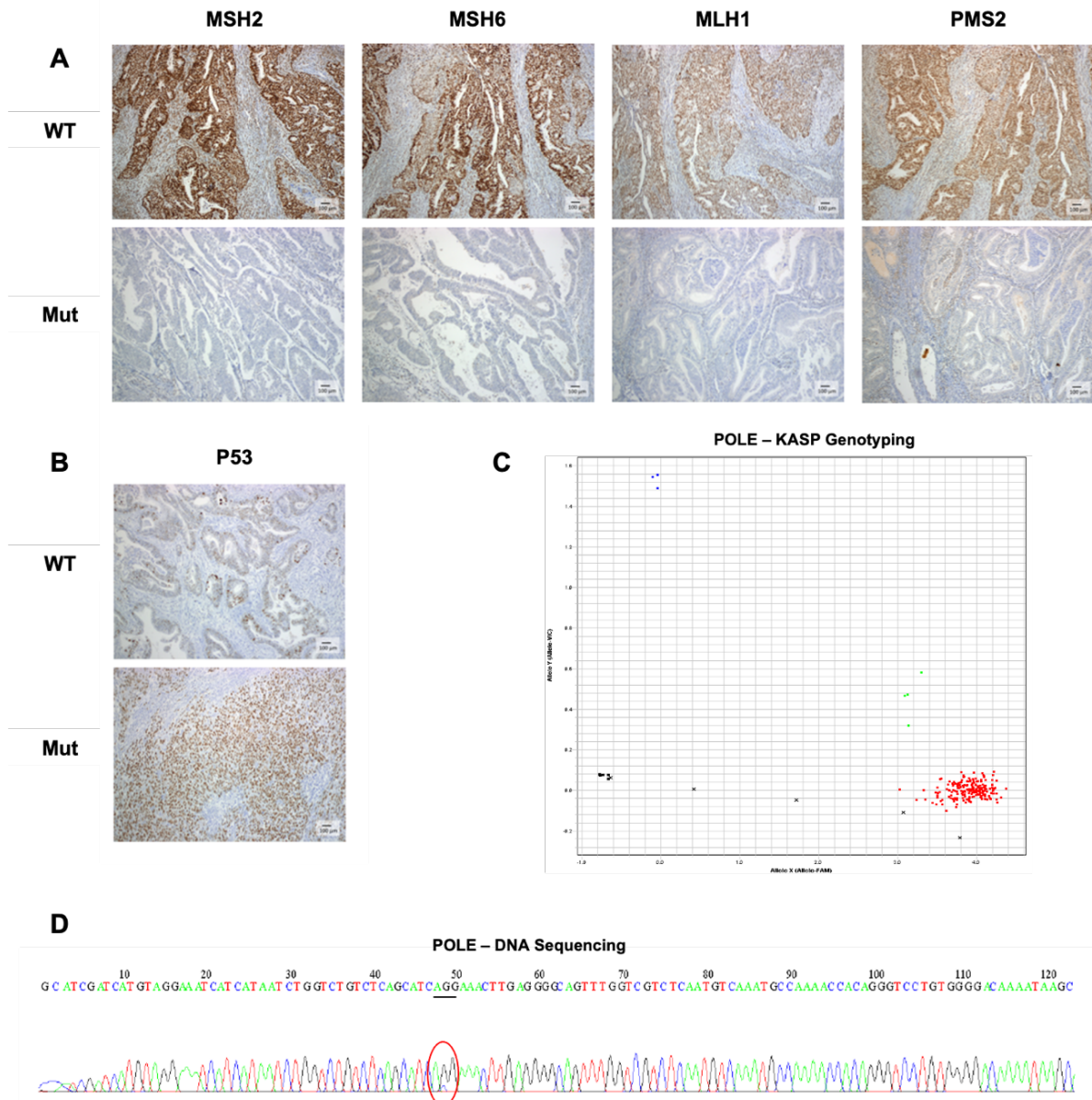
Figura 30. Surrogate markers for molecular classification of EC patients. A. Immunohistochemistry of MMR proteins (MSH2, MHS6, MLH1, PMS2). B. Immunohistochemistry of P53. C. POLE analysis using KASP Genotyping Technology. Black dots are negative control without expression; red dots are POLE WT patients with only expression of FAM; blue dots are control positive for POLE with only expression of VIC; green dots are POLE patients with both expression of FAM and VIC. D. Sequence of exon 9 and detection of P286R mutation from an EC patient.

As expected, the 3 EC patients classified in the POLE mutated group did not recur and were alive without disease. Patients classified in the microsatellite instability or low-CN group show a similar ratio of recurrence and non-recurrence. In contrast, we observed a higher percentage of patients classified in the high-CN group with recurrence. This highlights the usefulness of molecular classification to improve the accuracy to predict recurrence for POLE mutated and the high-CN group. However, molecular classification

Results

is not useful to distinguish recurrence in the microsatellite instability or in the low-CN group, where falls 75% of all EC patients.

Although we selected a balanced cohort of SEC patients (16 NOREC vs 17 REC), we observed a clear disbalance in the high-CN (11 NOREC vs 21 REC). This was caused by the inclusion of some EEC in the recurrence group, and the change of some SEC to MSI group because of mutations in the homologous recombination system. Additionally, we compared several clinicopathological features (histology, grade, FIGO stage, etc.) between recurrence and non-recurrence according to the molecular classification (Table 14). We only detected statistically differences in the final status of EC patients classified as low-CN or high-CN where most non-recurrent EC patients were alive and most recurrent EC patients were dead.

Table 14. Molecular classification for EC patients enrolled in the discovery phase and associated clinico-pathological features

		POLE		MSI		Low-CN		High-CN	
		NO REC	REC	NO REC	REC	NO REC	REC	NO REC	REC
PATIENTS		3	0	18	16	15	11	11	21
AGE	Years	72	N/A	66	70	70	70	69	73
HISTOLOGY	Endometrioid	2	N/A	15	15	13	11	2	5
	Serous	1	N/A	3	1	2	0	9	16
GRADE	1	0	N/A	1	3	1	0	0	0
	2	1	N/A	6	5	6	6	1	1
	3	2	N/A	11	8	8	5	10	20
STAGE	IA	1	N/A	5	1	4	2	7	14
	IB	1	N/A	8	9	5	4	2	5
	II	1	N/A	5	6	6	5	2	2
MYOMETRIAL INVASION	<50%	1	N/A	8	2	9	2	9	15
	>50%	2	N/A	10	14	6	9	2	6
LVSI	NO	3	N/A	7	10	12	8	10	14
	YES	0	N/A	10	5	2	3	1	7
RISK	Low	3	N/A	0	0	0	1	0	0
	Intermediate	0	N/A	5	3	7	2	7	14
	High-Intermediate	0	N/A	11	12	7	7	0	0
	High	0	N/A	2	1	1	0	4	7
REC SITE	Local	-	N/A	-	0	-	1	-	0
	Regional	-	N/A	-	5	-	4	-	11
	Distant	-	N/A	-	11	-	6	-	9
TIME TO REC	Early (<24 months)	-	N/A	-	11	-	6	-	15
	Late (>24 months)	-	N/A	-	5	-	5	-	6
STATUS	Alive without disease	2	N/A	14	6	14*	2*	11*	0*
	Alive with disease	0	N/A	0	2	0*	4*	0*	2*
	Dead of disease	0	N/A	0	7	0*	4*	0*	17*
	Dead other causes	1	N/A	4	1	1*	1*	0*	0*

* statistically significant differences between non-recurrence and recurrence

A quantitative analysis and an absence/presence analysis was conducted to compare recurrent vs non-recurrent patients from each molecular EC group, except for the POLE mutated group due to the absence of recurrent EC cases (Table 15).

Table 15. Number of significantly altered proteins between recurrence and non-recurrence according to molecular classification in the quantitative and the absence/presence analysis.

	QUANTITATIVE ANALYSIS				ABSENCE / PRESENCE ANALYSIS		
	MSI	Low-CN	High-CN		MSI	Low-CN	High-CN
P-value < 0.05	75	280	37	**P-value < 0.05**	101	91	78
Adjusted p-value < 0.25	0	251	0	**P-value < 0.01**	8	14	10
Adjusted p-value < 0.05	0	1	0				

From the quantitative analysis, we did not detect any significant protein (adjusted p-value < 0.25) neither in the MSI or the high-CN comparison. Results from the high-CN comparison were in concordance with the ones obtained from the serous histology. Interestingly, we found 251 significant proteins in the low-CN ECs, 34 proteins with logFC higher than ±1 (13 proteins were downregulated and 21 proteins were upregulated in the recurrence group) (see Annex 3). Importantly, we identified one protein with an adjusted p-value < 0.05, VAMP8, with a logFC<1 and AUC value of 0.94 (Figure 31A). Additionally, comparing the 251 differentially expressed proteins obtained in low-CN with the 252 significant proteins identified between recurrent and non-recurrent EEC patients, there was an overlap of 132 proteins (53%) (Figure 31B).

From the absence/presence analysis, we detected a higher number of significant proteins (see Annex 3). Specifically, we identified 8 proteins in the MSI group, 14 proteins in the low-CN group, and 10 proteins in the high-CN group (p-value < 0.01) (Figure 31C). In addition, comparing the results between high-CN and serous histology we found an overlap of 15 proteins (55%), including the best candidate (GUAD) to predict recurrence on the SEC (Figure 31D). Similarly, comparing the best protein candidates from the MSI and low-CN versus the proteins from the endometrioid histology, we found an overlap of 72 proteins (43%) (Figure 31E).

This highlights that the most significant proteins detected in the endometrioid, and serous histology were also important in the specific comparisons of MSI, low-CN and high-CN. Specifically, there was a correlation between the MSI, low-CN and endometrioid histology, and between high-CN and serous histology.
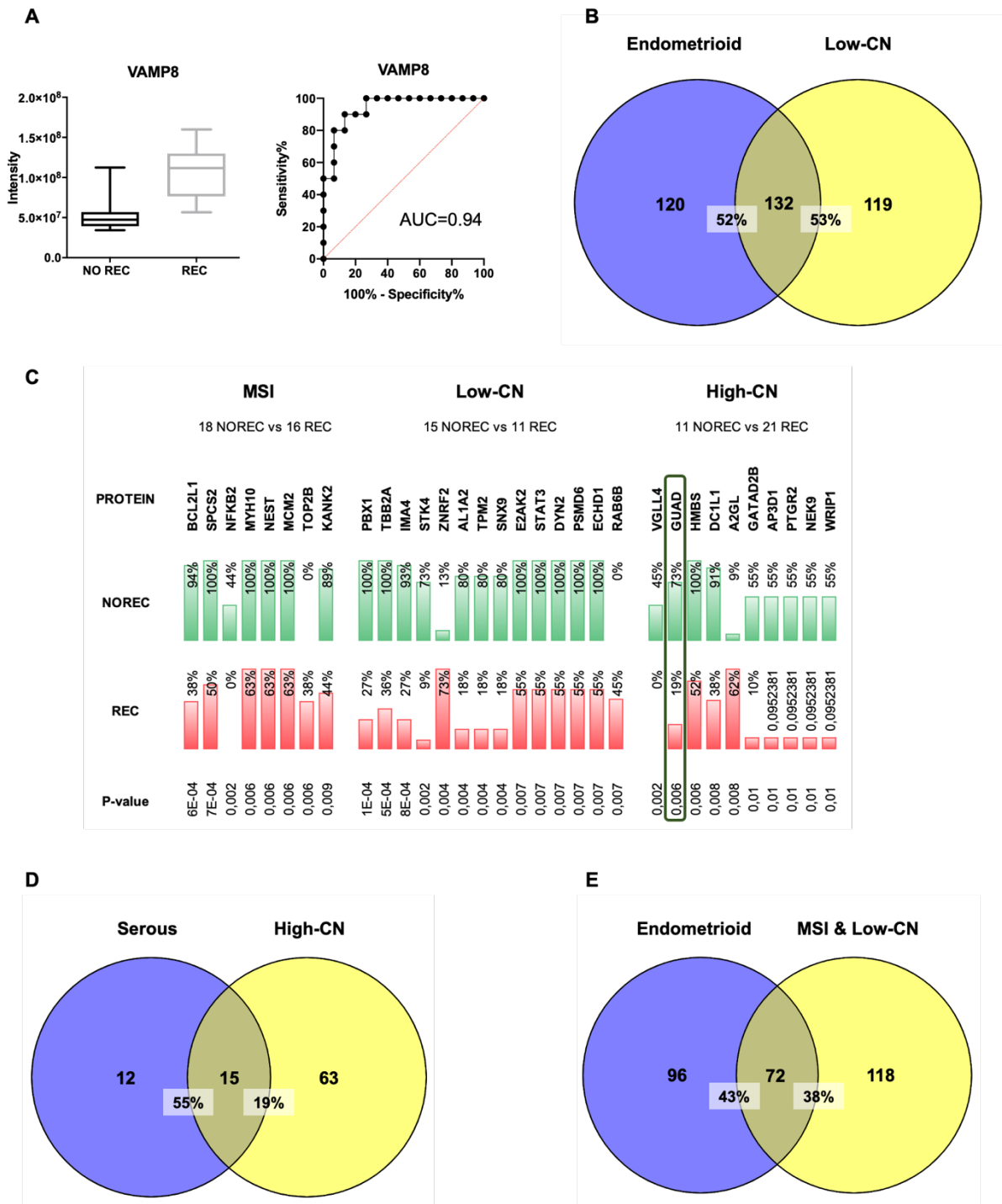
Results

Figura 31. Results from the Discovery in molecularly classified EC patients. A. Box plot and ROC curve of VAMP8. B. Venn diagram comparing 252 differentially expressed proteins from the endometrioid histology and 252 differentially expressed proteins from the low-CN group (quantitative analysis). C. Differentially expressed proteins from the absence/presence analysis in the MSI, low-CN, and high-CN group. D. Venn diagram comparing 27 differentially expressed proteins from the serous histology and 78 differentially expressed proteins from the high-CN group (absence/presence analysis). E. Venn diagram comparing 168 differentially expressed proteins from the endometrioid histology and 190 differentially expressed proteins from the MSI and low-CN group (absence/presence analysis).

DISCUSSION

Although the prognosis for endometrial cancer is good thanks to the early diagnosis, approximately 10-20% of tumors recur, and most recurrences take place within 2-3 years [27,28]. Recurrence in endometrial cancer is one of the main factors associated to mortality. The risk stratification system allows clinician to classify patients according to a risk group and adapt the adjuvant treatment and follow-up. However, there is a percentage of cases that recur independently of the risk group that was identified to the patient. Clinic-pathological features have been the standard of care for the risk stratification. Histological type, FIGO stage, grade, LVSI, and residual disease were the main hallmarks to predict recurrence. Importantly, in the last ESGO/ESTRO/ESP guidelines, molecular classification has been included as a new prognostic factor in the risk stratification system [26]. Nevertheless, because of the specific infrastructure required for POLE detection, not all hospitals are available to perform it. Therefore, the current guidelines include two different risk group classification depending on the availability to have the molecular classification.

It has been demonstrated that the incorporation of molecular features reclassifies EC patients to a new risk group. Imboden et al. [238] classified 594 EC patients using the prior 2016 clinicopathologic system and the current system including molecular classification. A total of 39 patients (7%) were reclassified in a new group risk, mainly because of P53 or POLE mutations. The switch from high risk to low risk and vice versa were the groups that was most affected by a change in the adjuvant treatment. Specifically, in their cohort, 24 patients would have benefit from a molecular classification. Many studies suggest that the incorporation of molecular classification into the risk stratification increases the potential to predict recurrence [88]. In our cohort, 6 patients have been reclassified, and 5 of them would have get profit because of a change in the adjuvant treatment. Specifically, two patients who experienced recurrence increased to a higher risk group due to the presence of a P53 mutation, whilst three patients who did not recur would have been allocated in a lower risk group thanks to the presence of a POLE mutation. These changes were mainly caused by P53 and POLE mutations, reflecting the importance of these two molecular markers. These results are in concordance with most studies where POLE patients do not recur, and high-CN patients have a higher percentage of recurrences [239,240]. However, there are still some limitations of the molecular classification to be investigated: studies should allocate the question on how the FIGO stage contributes to prognosis in each molecular subtype, and if there are additional biomarkers that might permit to identify 100% of recurrent patients. Although this new

Results

stratification system allows a better understanding of recurrence, it is not enough to detect it in all patients.

In this chapter, we have analyzed the proteome of 96 patients to identify biomarkers predicting recurrence. We began the study with a discovery phase, following the typical phases of a biomarker pipeline, and face with the inherent challenges that limit the successful translation of candidate biomarkers into clinical use. Important aspects to emphasize at this phase was experimental design, appropriate statistical methodologies, and quality assessment of results [241]. As we showed in Chapter 1, MS has emerged as a new tool for the identification and validation of potential biomarkers. Currently, Data Dependent Acquisition (DDA) and Data Independent Acquisition (DIA) are two different approaches for the identification of biomarkers in a discovery phase. On the one hand, DDA has been the preferable method for the identification of proteins without the need of any knowledge of the analyzed sample, but it generates bias towards the most abundant proteins and has a limited reproducibility. On the other hand, DIA has emerged in the last years solving the problem of reproducibility, but data processing is challenging, and it could imply the necessity of library generation [176,180]. DIA was not available by the time this project was initiated, and thus DDA was selected to conduct the discovery phase. Despite that, Coscia et al. [227] performed a similar study with FFPE specimens from different tissue types (ovarian cancer, glioma, colorectal adenoma and urachal carcinoma) and using DDA for MS analysis, and they identified and average of 4,933 proteins per cancer type. Additionally, they also performed DIA and obtained an average of 4,983 proteins per sample. In our case, we obtained 4,569 proteins in EEC and 5,747 proteins in SEC, suggesting that both techniques detect a similar number of proteins.

The proteomic study permitted to understand the proteomic landscape of our patients and how proteins are able to differentiate between clinical and molecular features. Interestingly, we observed that the proteome of our cohort was clearly different from endometrioid or serous histologies, which is in line whit the observation done in other studies [242,243]. Additionally, we have also detected a different pattern in the high-CN patients in comparison to the other three molecular subgroups. It has been shown that high-CN group presents the worst progression-free survival (PFS) and overall survival (OS) [44]. Similarly, this pattern was related to grade 3, but it resulted from the inherent classification of serous histology as high grade.

For the endometrioid histology, we used 32 patients in the non-recurrence group and 31 patients in the recurrence group. We detected a total of 4.569 proteins, with 252 proteins differentially expressed from the quantitative analysis (adjusted p-value < 0.25) and 168 proteins differentially detected from the absence/presence analysis (p-value < 0.05). We

also identified proteins in the sub-comparisons of stage II, LVSI positive, high risk, and time to recurrence (NO REC vs late REC). Interestingly, from the 252 proteins from the general comparison (NO REC vs REC), 213 proteins (85%) were also differentially expressed in the different sub-comparisons. Altogether, combining the quantitative and the absence/presence analysis, we generated a list of 420 potential biomarkers predicting recurrence in EEC. Most EC patients included in this analysis have an intermediate to high risk of recurrence, and in the general population 15-40% of patients from these group risk will recur. Thanks to this study, the identified protein biomarkers are expected to detect those patients that will recur with higher accuracy.

Differential proteins identified in the proteomic study of EEC primary tumors was used to conduct a Gene Set Enrichment Analysis (GSEA) to understand the major pathways controlling EC recurrence. The activation of RNA processing and spliceosome pathway, and the suppression of adhesion and binding pathways were the most altered pathways in the primary tumors of recurrent patients. On the one hand, the spliceosome ribonucleoprotein complex catalyzes the removal of introns and exons ligation, a fundamental post-transcriptional process that generates mature mRNA. It has been shown that some mutations in spliceosome components enhance the activity, induces aberrant splice site selection and, finally, the production of novel isoform variants supports tumorigenesis, allowing tumor progression, invasion, and migration [244,245]. On the other hand, adhesive and binding molecules are responsible for the cell-cell interaction and the surrounding intercellular environment creating normal tissue architecture. Suppression of these pathways disturb epithelium integrity and cell connections and induce cell infiltration to surrounding tissues and the spread of cancer through the formation of metastases [246].

For the serous histology, we used 16 patients in the non-recurrence group and 17 patients in the recurrence group. We detected a total of 5.747 proteins, but only 27 were differentially expressed in the absence/presence analysis (p-value < 0.05). As well, GSEA analysis did not show any relevant pathway, mainly because of the few number of significant proteins included in the analysis. Last, PCA analysis did not show a clear differentiation regarding recurrence in these patients. Unexpectantly, the proteomic landscape of recurrent and non-recurrent SECs seems quite homogeneous, at least for the low number of patients included in the analysis.

The most promising biomarker for the serous histology was **GUAD or Guanine Deaminase**. It is an enzyme responsible for the hydrolytic deamination of guanine, producing xanthine and ammonia. The end-products of guanine-based purines (GBPs), like xanthine, have been associated to reactive oxygen species (ROS) production and

Results

DNA damage. Although GUAD has not been directly studied in cancer, some evidences suggest a possible role of GBPs in cancer, with purine salvage pathway being the fuel of nucleotide pool maintenance and correct cell division [247]. Additionally, several findings support the anti-proliferative effect of nucleoside guanosine (GUO), nucleobase guanine (GUA), and guanosine 5'-monophosphate (GMP) in glioblastoma cells, prostate cancer cells, lung adenocarcinoma cells and myeloid leukemia cells [248–250]. In our results, we observed a diminished expression of GUAD in the recurrence patients, suggesting that it might contribute to nucleotide pool destabilization and enhance tumor aggressivity.

Regarding molecular classification, the analysis of recurrence in the different molecular subgroups did not show additional results from the information obtained in the endometrioid or serous analysis. In general, we detected less differentially expressed proteins per group. Importantly, we observed a correlation between MSI / low-CN and endometrioid histology, and between the high-CN and serous histology. On the one hand, 53% of significant proteins from the quantitative analysis and 43% of significant proteins from the absence/presence analysis in the endometrioid histology were also detected in the MSI and/or low-CN analysis. On the other hand, 55% of significant proteins from the absence/presence analysis in the serous histology were also detected in the high-CN analysis. These results confirm that most SEC patients behaves as high-CN, while most EEC behaves as MSI or low-CN.

The most promising biomarker derived from molecular classification, specifically for the low-CN group, was **VAMP8** or **Vesicle Associated Membrane Protein 8**. The encoded protein is involved in the fusion of vesicles and in autophagy through the direct control of autophagosome membrane fusion with the lysosome membrane [251,252]. Although VAMP8 has not been studied in EC, it has been investigated in several cancers with different roles. Wang et al. [253] observed that VAMP8 drives the secretion of TIMP1 to inhibit tumor metastases in lung cancers. Specifically, lung cancer patients with low VAMP8 showed distant metastasis, poor OS and PFS. Otherwise, Chen et al. [254] saw that overexpression of VAMP8 promoted cell proliferation *in vitro* and *in vivo* in human glioma, whereas knockdown of VAMP8 attenuated glioma growth by arresting cell cycle in the G0/G1 phase. In our results, VAMP8 was upregulated in the recurrence group, suggesting that it might contribute to cell proliferation and migration. However, its role in EC has to be further analyzed.

From our knowledge, this is the first study which aim is to identify protein biomarkers predicting recurrence. Most discovery phases have been focused on the identification of diagnostic and/or prognostic protein biomarkers . Audet-Delage et al. [255] performed a similar study using MS-based in metabolomics. They compared non-recurrent and

recurrent patients and identified that the combination of 2-oleoylglycerol and TAG42:2-FA12:0 allowed the distinction of both groups with an AUC of 0.901 (P-value < 0.001). However, these are pilot studies with preliminary results that need to be further investigated.

In conclusion, in this chapter we have identified protein candidates predicting recurrence in the endometrioid histology, serous histology, and in the different molecular subgroups. Additionally, we have also unveiled the proteomic landscape of recurrent EC and identified the most relevant pathways involved in recurrence.

Results

CHAPTER 3

BIOINFORMATICS ANALYSIS

IN-SILICO ANALYSIS OF THE TCGA AND CPTAC DATA

TO IDENTIFY EC BIOMARKERS PREDICTING

RECURRENCE

SPECIFIC BACKGROUND

Online databases are a valuable resource to identify and/or validate new biomarkers. The most widely repositories to explore the molecular and proteomic landscape of EC are The Cancer Genomic Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC), respectively.

On the one hand, TCGA performed an integrated genomic, transcriptomic, and proteomic characterization of EC patients using array- and sequencing-based technologies [44]. Regarding proteomics, TCGA used reverse phase protein array (RPPA), which is a type of protein microarray derivative of gene expression microarray and immunoassay. Specifically, the RPPA from TCGA analyzed the expression of 208 cancer-related in 440 EC patients, including 201 endometrioid and 58 serous histologies.

On the other hand, CPTAC undertook a comprehensive proteogenomic characterization of 95 endometrial carcinomas, comprising 83 EEC and 12 SEC tumors [242]. Concretely, CPTAC analyzed the expression level of 12.524 proteins in all 95 EC patients.

The work presented in this chapter aimed to i) analyze online databases (TCGA and CPTAC) to discovery potential biomarkers predicting recurrence; ii) assess the relation of the identified protein biomarkers with clinic-pathological and molecular features; iii) compare potential biomarkers from the discovery phase in Chapter 2 and the *in-silico* analysis in Chapter 3.

MATERIAL AND METHODS

Data source

Expression data of EC patients were collected from the TCGA database through cBioPortal (https:///www.cbioportal.org/, accessed on 2 July 2020). The protein expression data of 208 proteins evaluated in a cohort of 440 EC patients from the Uterine Corpus Endometrial Carcinoma (TCGA, Firehose Legacy) study RPPA was used.

CPTAC (Uterine Corpus Endometrial Carcinoma) data was obtained from LinkedOmics database (http://www.linkedomics.org/login.php, accessed on 21 January 2021). The protein expression data of 12.524 proteins evaluated in a cohort of 95 EC patients was used (CPTAC-proteome).

Patient recruitment

A total of 440 EC patients were selected from the TCGA study and 95 EC patients were selected from the CPTAC study. From those, we excluded EC patients without recurrence with a follow-up less than 2 years and EC patients without clinical data information about recurrence, resulting in 271 EC patients from the TCGA study and 87 EC patients from the CPTAC study. Clinic-pathological data of these patients is detailed in Table 16 and Table 19, respectively.

Statistical analysis

The statistical analysis was performed in Graph Pad Prism (v.8.2.1) (GraphPad Software, La Jolla, CA, USA). The correlation between clinic-pathologic characteristics and recurrence was calculated using Fisher's exact test for 2 variables and Chi-square for 3 or more variables. Two different approaches were used for the analysis of differentially expressed proteins:

- Quantitative analysis was performed only in those protein detected in more than 80% of the patients in both groups. Due to the non-normality of the data, comparison of the abundance protein between recurrence and non-recurrence was done using the non-parametric Mann-Whitney U test. Adjusted p-values lower than 0.25 were considered statistically significant.
- The absence/presence analysis was done using Fisher's exact test. P-values lower than 0.05 were considered statistically significant.

Receiver operating characteristic (ROC) curves were used to calculate the relationship between sensitivity and specificity for each EC biomarker candidate.

RESULTS

Potential biomarkers predicting recurrence in the TCGA study

A total of 440 EC patients from the TCGA were selected for this study, but 169 patients were excluded because they were non-recurrent with a follow-up less than 2 years. Therefore, 271 EC patients were included in the final analysis comparing 187 non-recurrent EC patients against 84 recurrent EC patients. The clinical and pathological characteristics of those EC patients are summarized in Table 16 and Figure 32.

Table 16. *Clinic-pathological characteristics of women diagnosed with EC enrolled in the TCGA study.*

	NO Recurrence	Recurrence	P-value
Total Number	187	84	
Age (years)	62 (31-89)	65 (39-87)	0.10
Histology			0.001
Endometrioid	151	50	
Serous	29	29	
Mixed	7	5	
Grade			0.001
1	43	6	
2	39	17	
3	105	61	
Figo Stage			<0.0001
IA	86	18	
IB	42	18	
II	17	6	
III	39	27	
IV	3	15	
Myometrial Invasion			0.005
<50%	98	28	
>50%	70	45	
Molecular Classification			0.0001
POLE	15	0	
MSI	26	7	
Low-CN	44	9	
High-CN	18	19	
Risk Classification			<0.0001
Low	38	2	
Intermediate	37	7	
High-Intermediate	58	15	
High	46	19	
Advanced metastatic	10	15	
Time to Recurrence			N/A
Early (<24 months)	-	59	
Late (>24 months)	-	25	
Status			<0.0001
Living	187	43	
Deceased	0	41	

Our study cohort was age-balanced with a mean age for non-recurrent EC patients of 62 years and for recurrent EC patients of 65 years. As seen in other studies, the event of recurrence was significantly associated to all clinical features, including serous and

mixed histologies, advanced FIGO stage, deep myometrial invasion, and high grade; and to the high-CN group of the molecular classification (Table 16 and Figure 32). As expected, the risk classification system permitted to detect most of the recurrent cases in those patients classified at high-risk of recurrence. Notwithstanding this, 5% and 15% of low and intermediate risk EC patients also end up having a recurrence. Recurrent EC patients presented a PFS of 20 months (1 – 89) and OS of 36 months (185 – 2). Detection of recurrence is of vital importance since this event is highly related to EC mortality. As shown in Table 16, whilst all non-recurrent patients were alive after a follow-up of 2 years (median 54 months [24 – 149]), 49% of recurrent patients deceased. Overall, we observed that all clinic-pathological and molecular features were statistically significant to detect recurrence, however, there is still a percentage of cases that recur, and we are not able to identify.
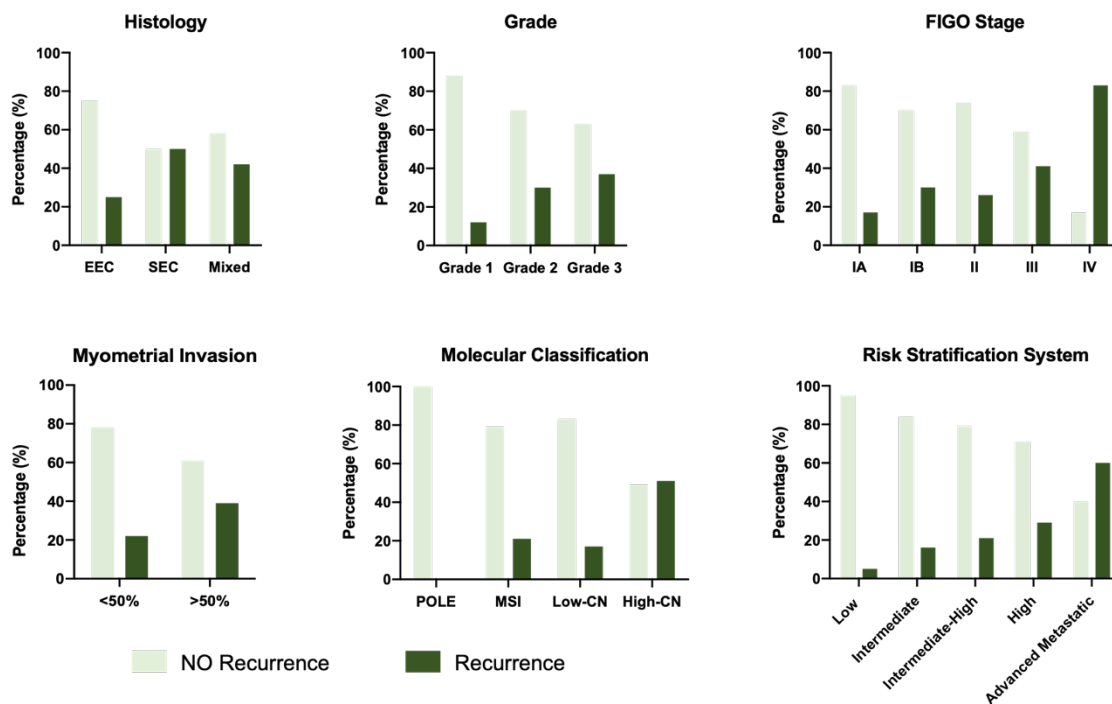
Figura 32. Clinic-pathological and molecular features of patients from the TCGA data. *Comparison of no-recurrence versus recurrence according to histology, grade, FIGO stage, myometrial invasion, molecular classification, and risk stratification system (EEC: endometrioid endometrial cancer; SEC: serous endometrial cancer).*

In order to unveil protein biomarkers predicting recurrence, we compared the differential expression of the 208 cancer-related proteins between recurrence and non-recurrence patients. Five proteins were excluded because they did not reach 80% of detection in patients from each comparison group. Out of the 203 proteins included in the quantitative analysis, 15 proteins were significant with an adjusted p-value < 0.25 (Table 17).

Table 17. *Fifteen proteins differentially expressed between recurrent and non-recurrent EC patients from the TCGA study.*

Uniprot Number	Protein Name	Protein ID	Adjusted p-value	AUC	REC status	Protein Location
P04083	Annexin A1	ANXA1	0.0023	0.675	↓	Nucleus, cytoplasm, extracellular region
O96017	Serine/threonine-protein kinase Chk2	CHK2_pT68	0.0072	0.645	↑	Nucleus
P46937	Transcriptional coactivator YAP1	YAP_pS127	0.0520	0.631	↓	Nucleus, cytoplasm
P04637	Cellular tumor antigen p53	P53	0.0678	0.643	↑	Nucleus, cytoplasm
P43246	DNA mismatch repair protein Msh2	MSH2	0.0678	0.614	↑	Nucleus
P49327	Fatty acid synthase	FAS	0.0778	0.594	↑	Cytoplasm
P35579	Myosin-9	MYH9	0.0778	0.604	↓	Cytoskeleton
P52701	DNA mismatch repair protein Msh6	MSH6	0.0856	0.608	↑	Nucleus
P27986 / O00459	Phosphatidylinositol 3-kinase regulatory subunit alpha / beta	PI3K_p85	0.0856	0.605	↓	Nucleus, cytoplasm
P03372	Estrogen receptor	ESR1	0.0856	0.612	↓	Nucleus, cytoplasm
P35579	Myosin-9	MYH9_pS1943	0.1362	0.601	↓	Cytoskeleton
P31749 / P31751 / Q9Y243	RAC-alpha/beta/gamma serine/threonine-protein kinase	AKT_pT308	0.1549	0.597	↓	Nucleus, cytoplasm
Q13490	Baculoviral IAP repeat-containing protein 2	BIRC2	0.1576	0.600	↓	Nucleus, cytoplasm
P12830	Cadherin-1	CADH1	0.1576	0.623	↓	Plasma membrane
Q13131	5'-AMP-activated protein kinase catalytic subunit alpha-1	AAPK1	0.2090	0.591	↓	Nucleus, cytoplasm

Five of the fifteen proteins were phosphorylated, so we could not compare with the results from our discovery phase. In addition, we found TP53, MSH2 and MSH6 as a potential biomarker predicting recurrence, which are proteins used for the molecular classification. Also, we identified proteins which have been previously related to poor EC prognosis, such as ESR1 and CADH1. Among the 7 significant proteins that we were able to compare with the results obtained in the discovery phase, 2 of the proteins were also differentially expressed: ANXA1 and MYH9. Both proteins appeared downregulated in recurrent patients in the discovery phase and in the *in-silico* study of the TCGA dataset (Figure 33).
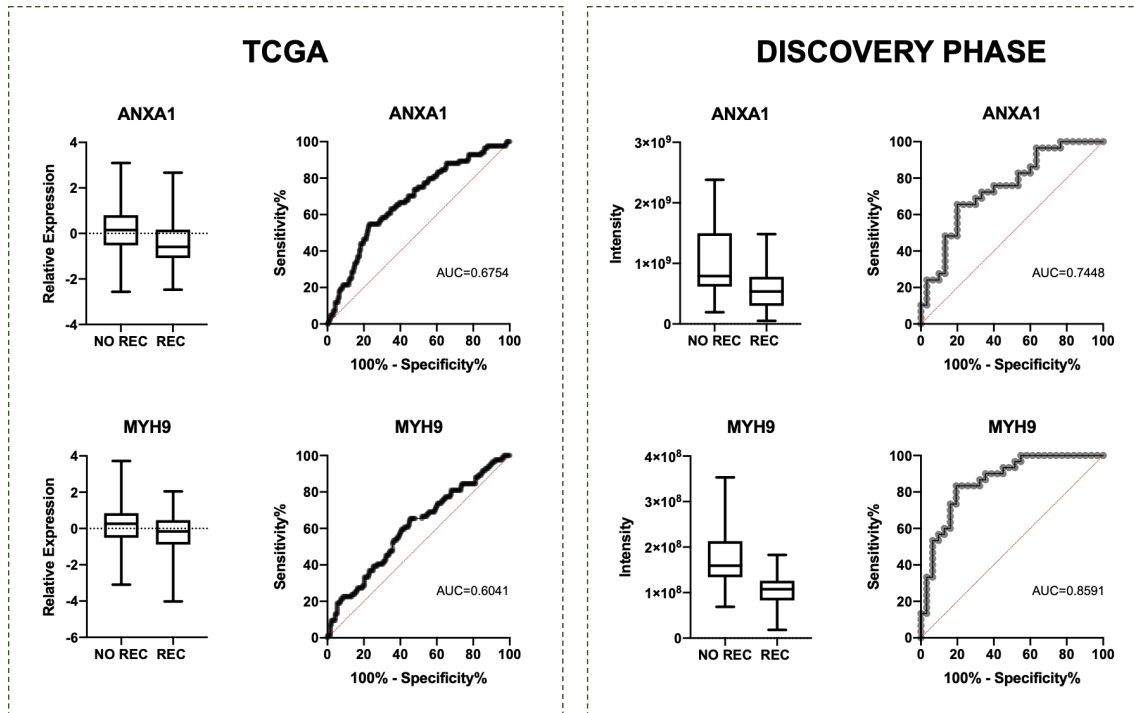
Figura 33. *Comparison of ANXA1 and MYH9 expression between data from the TCGA and the discovery phase.*

Next, we evaluated the protein expression level between recurrence and non-recurrence through different comparisons in which we restricted the dataset according to the histology, the inclusion criteria followed in the discovery phase (FIGO stage IA, IB or II), and the molecular classification. These comparisons are summarized in Table 18.

Table 18. Results from the TCGA analysis in the different comparisons. *Number of patients in the non-recurrence and recurrence group and number of proteins differentially expressed in the comparisons considering histology and molecular classification.*

Comparison	Number of patients		Number of proteins statistically significant	
	NO REC	REC	P-value < 0.05	Adjusted p-value < 0.25
General	187	84	27	15
EEC	151	50	12	2
EEC*	81	24	5	0
SEC	29	29	5	0
SEC*	16	6	7	0
MSI	26	7	6	0
Low-CN	44	9	4	0
High-CN	18	19	11	0

*Only those EC patients following the inclusion criteria of the discovery phase (endometrioid or serous histology, and FIGO stage IA, IB or II)

On the one hand, in the endometrioid histology, considering all patients (EEC), we were able to detect 2 proteins with an adjusted p-value < 0.25. These proteins were ANXA1 and YAP_pS127. We observed again ANXA1 as a potential biomarker for the endometrioid histology. However, in the specific comparison using only those patients that fulfill the criteria for the discovery phase (EEC*), we did not detect any protein. On the other hand, in the serous histology, we did not find out any potential biomarker, neither considering all patients (SEC) or the subgroup of patients following the criteria of the discovery phase (SEC*). These results were similar as the ones that we obtained in the discovery phase without the detection of a significant biomarker. Finally, we did not identify differentially expressed proteins in the MSI, low-CN, or high-CN comparison between recurrent and non-recurrent patients.

To summarize, we identified 15 differentially expressed proteins in the general comparison and 2 differentially expressed proteins in the endometrioid histology. Importantly, two of them (ANXA1 and MYH9) were also proteins identified in the discovery phase (Chapter 2), suggesting that they might be potential biomarkers predicting recurrence in EC patients.

Potential biomarkers predicting recurrence in the CPTAC study

A total of 95 EC patients were included in the CPTAC dataset. However, 87 EC patients were included into the analysis since some patients were discarded because they were non-recurrent with a follow-up less than 2 years. The cohort was splitted in 74 non-recurrent and 13 recurrent EC patients. The clinical and pathological characteristics of the EC patients are summarized in Table 19 and Figure 34.

Age was unbalanced between the non-recurrent and recurrent EC patients, with a mean age of 61 and 71 years, respectively. Similar to the TCGA study, the event of recurrence was significantly associated to non-endometrioid histologies (serous and carcinosarcoma), advanced FIGO stage, deep myometrial invasion, LVSI, and high grade (Table 19 and Figure 34). Regarding molecular classification, any POLE EC patient had recurrence and the highest percentage of recurrent was found in the high-CN group, although these difference did not reach significance.

As seen in the TCGA study, the amount of recurrent cases increased with the risk determined by the risk classification system (Low: 0% > Intermediate: 21% > High-Intermediate: 25% > High: 33%). Still, this classification is not perfect and there is a percentage of cases that recur, and we are not able to identify. This is an important

unmet clinical need since 38% of recurrent patients deceased, whilst all non-recurrent patients were alive at least two years after primary surgery.

Table 19. Clinic-pathological characteristics of women diagnosed with EC enrolled in the CPTAC study.

|		NO Recurrence	Recurrence	P-value
Total Number		74	13	
Age (years)		61 (38-78)	71 (48-86)	0.001
Histology				0.03
	Endometrioid	64	9	
	Serous	9	3	
	Clear Cell	1	0	
	Carcinosarcoma	0	1	
Grade				<0.0001
	1	28	0	
	2	31	3	
	3	15	10	
Figo Stage				0.0006
	IA	49	2	
	IB	13	3	
	II	6	5	
	III	7	3	
Myometrial Invasion				0.02
	<50%	47	4	
	>50%	20	8	
LVSI				0.01
	NO	52	4	
	YES	21	8	
Molecular Classification				0.12
	POLE	6	0	
	MSI	17	4	
	Low-CN	35	2	
	High-CN	12	6	
Risk Classification				0.0004
	Low	40	0	
	Intermediate	11	3	
	High-Intermediate	12	4	
	High	10	5	
Status				<0.0001
	Living	74	8	
	Deceased	0	5	
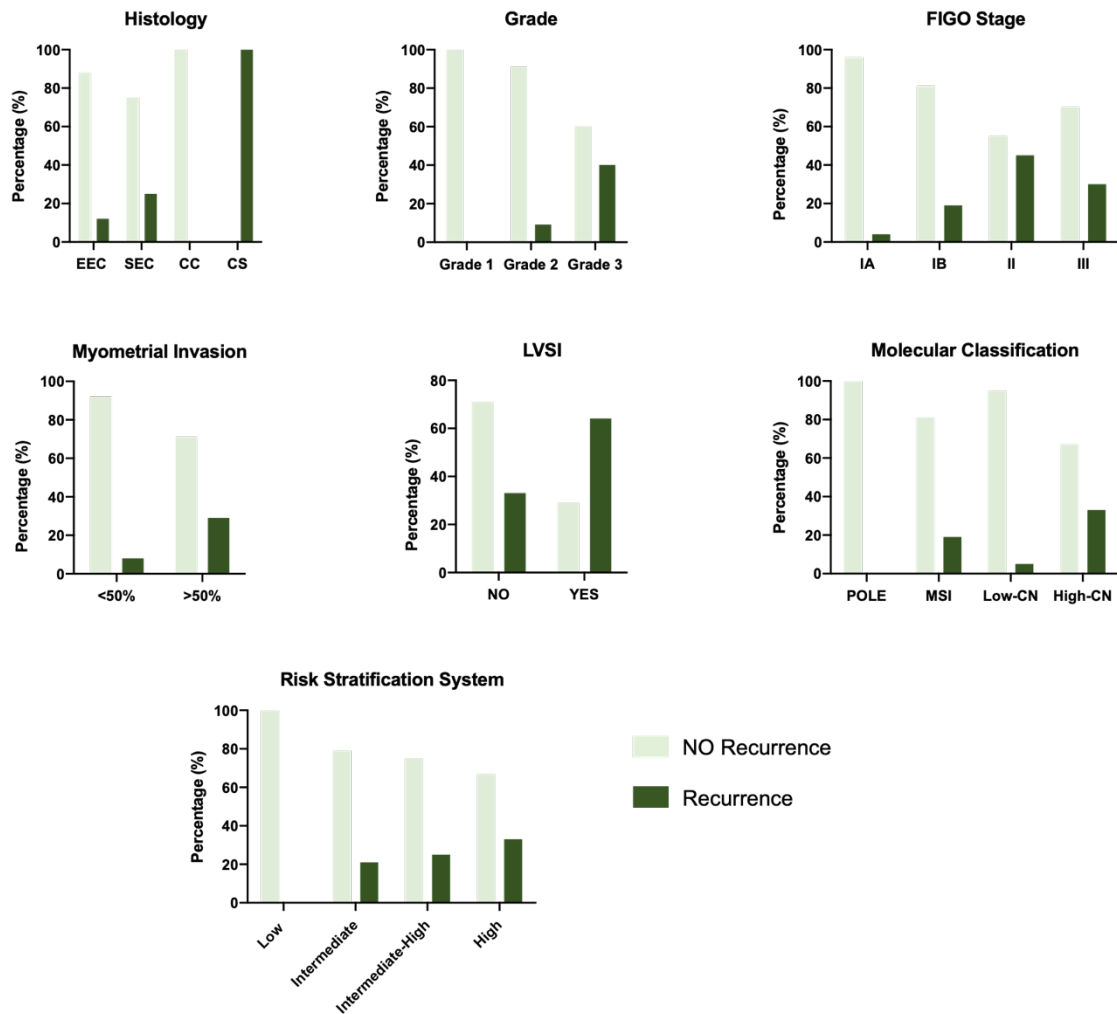
Figura 34. Clinic-pathological and molecular features of patients from the CPTAC data.
*Comparison of no-recurrence versus recurrence according to histology, grade, FIGO stage,
myometrial invasion, LVSI, molecular classification, and risk stratification system (EEC:
endometrioid endometrial cancer; SEC: serous endometrial cancer; CC: clear cell; CS:
carcinosarcoma).*

In order to unveil protein biomarkers predicting recurrence, we compared the expression
level of 12.524 proteins in the 74 non-recurrent and 13 recurrent EC patients using
quantitative and the absence/presence analysis. The data analysis yields 115 proteins
with an adjusted p-value < 0.25 (including 3 proteins with and adjusted p-value < 0.01)
in the quantitative analysis, and 138 proteins with a p-value < 0.05 (including 23 proteins
with a p-value < 0.01) in the absence/presence analysis (Figure 35 and Annex 4). When
comparing this *in-silico* results with the previously performed in the discovery phase
(Chapter 2), none of the best biomarkers predicting recurrence (adjusted p-value < 0.25
in the quantitative analysis or p-value < 0.01 in the absence/presence analysis) appeared
in both comparisons.
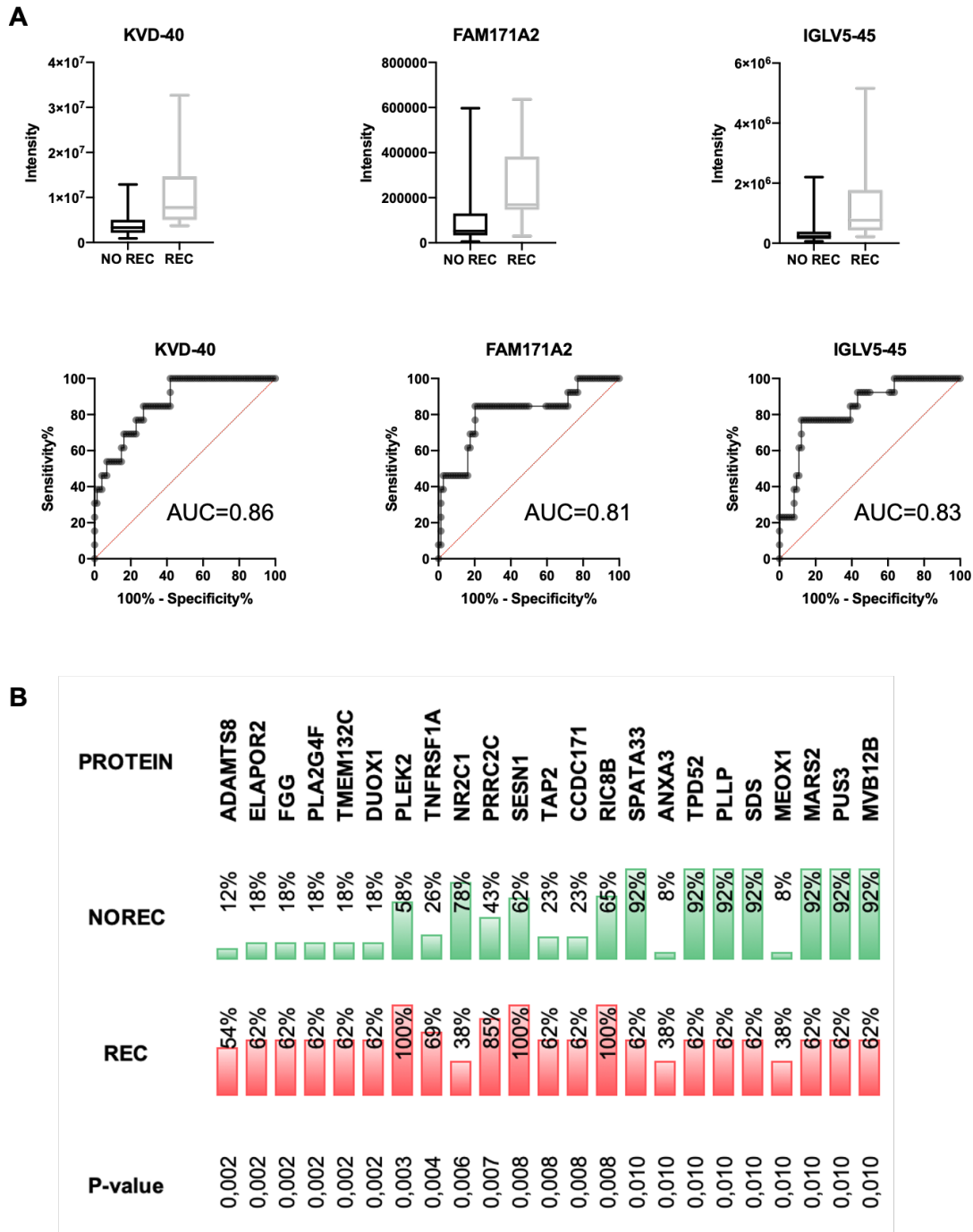
Results

Figura 35. Results from the CPTAC analysis. A. *Boxplot and ROC curve of the top3 biomarkers from the quantitative analysis with an adjusted p-value < 0.05.* **B.** *Differentially expressed proteins from the absence/presence analysis with a p-value < 0.05.*

Due to the low sample size in the recurrent EC patients, we only proceed to split the cohort to make a sub-analysis of the endometrioid histology. Specifically, we compared 64 non-recurrent against 9 recurrent EC patients. From the quantitative analysis, we identified 41 proteins with an adjusted p-value < 0.25, and from the absence/presence analysis, we detected 159 proteins with a p-value < 0.05 (see Annex 4). When comparing this *in-silico* results with the previously performed in the discovery phase (Chapter 2),

none of the best biomarkers predicting recurrence (adjusted p-value < 0.25 in the quantitative analysis or p-value < 0.01 in the absence/presence analysis) appeared in both comparisons.

To summarize, we identified 253 differentially expressed proteins in the general comparison and 200 differentially expressed proteins in the endometrioid histology. Although none of the best biomarkers were also identified in the discovery phase (Chapter 2), they have demonstrated a high accuracy to predict EC recurrence and thus, they should be investigated in subsequent studies.

DISCUSSION

Online databases are a valuable resource to identify and/or validate potential biomarkers. In EC, The Cancer Genome Atlas Research Network was the first to perform an integrated genomic, transcriptomic and proteomic characterization of 373 endometrial carcinomas using array- and sequencing-based technologies [44]. Data from these patients is available free, and it has been uploaded and increased the number of patients through several actualizations (Firehose Legacy and PanCancer Atlas). Most studies have been focused on the genomic and transcriptomic data. Li et al. [256] revealed the possibility of integrating histopathology, molecular pathology and chromosomal instability (CIN) to improve risk prognostication and treatment recommendations in early stage-EC. Bosquet et al. [257] generated a biomarker panel using differential gene-specific mutations and mRNA expressions to differentiate high- and low-risk EC. Regarding recurrence prediction, the four molecular subgroups were related to progression-free survival (PFS). Similar to our observations, Raffone et al. [258] combined data from the TCGA and other studies and concluded that high-CN group had a risk of recurrent 3.5 times than the other groups, while POLE group showed the best patient outcome.

Proteomic data from TCGA is based on reverse phase protein array (RPPA), which is a type of protein microarray derivative of gene expression microarray and immunoassay. The RPPA from TCGA analyzed the expression of 208 cancer-related proteins in 440 EC patients. Lai et al. [259] constructed a prognostic signature based on 9 proteins, which could divide patients into low- and high-risk with distinct prognoses combining the protein expression data and RNA-seq expression from the TCGA. Interestingly, two of these proteins were ANXA1 and MYH9, which are also two of the significant biomarkers that we identified in both the *in-silico* and the previous discovery phase study. ANXA1 and MYH9 were significantly downregulated in high-risk compared to low-risk EC patients,

suggesting that both proteins may serve as a negative biomarker in cancer development and in the progression of EC. In addition, they also identified ESR1 as a favorable prognostic protein and Chk2_pT68 as an unfavorable prognostic factor, in concordance with our results obtained from the TCGA analysis.

A limitation on using the TCGA proteomic data is the low number of proteins analyzed (208), which is related to the availability and reproducibility of antibodies included in the RPPA technology. Among the 208 proteins analyzed in the TCGA study, we were able to study 203 proteins (5 proteins were excluded due to their low detection). However, 54 proteins were phosphorylated biomarkers (27%), and 47 proteins were not detected in our previous discovery phase (23%), which limited to 102 proteins our capacity to validate *in-silico* our previous results. Still, we have been able to validate ANXA1 and MYH9 as potential biomarkers predicting recurrence in EC, both of them downregulated in EC recurrent patients. On the one hand, **ANXA1 (Annexin 1)** is a membrane-localized protein that binds phospholipids, it inhibits phospholipase A2 and has anti-inflammatory activity, and its loss of function or expression has been detected in multiple tumors [260]. Additionally, it belongs to cell adhesion molecules (CAMs) family, which are involved in the binding of a cell to another cell or to the extracellular matrix, and they have roles in cell proliferation, differentiation, motility, trafficking, apoptosis, and tissue architecture. In EC, Voisin et al. [261] found ANXA1 to be overexpressed in tumors in comparison to normal endometrium, and as described earlier, Lai et al. [259] observed ANXA1 downregulation in high-risk compared to low-risk EC patients. Moreover, high expression of ANXA1 has been correlated with recurrence in rectal cancer [262], while low expression of ANXA1 has been related with recurrence in prostate cancer [263]. The dual effect of ANXA1 in recurrence and in cancer reveals that further investigation is needed for each tumor type. On the other hand, **MYH9 (Myosin-9)** is a myosin with several important functions, including cytokinesis, cell motility, and maintenance of cell shape. Variants of MYH9 cause an autosomal-dominant disorder, termed MYH9-related disease, and may be involved in other conditions such as cancer [264]. Some investigations have obtained strongest evidence that MYH9 acts as a tumor suppressor. Schramek et al. [265] found that 24-31% of human skin and head and neck squamous cell carcinomas (SCCs) were characterized by no or very weak MYH9 expression. This study showed that MYH9 deficiency induce defective activation of the p53 protein upon DNA damage. Similarly, Kas et al. [266] found that mutations resulting in heterozygous loss of MYH9 act as a driver event for the development of invasive lobular breast carcinoma (ILC).

Although we only validated ANXA1 and MYH9 from the TCGA analysis, we also identified other potential biomarkers that have been already studied in EC such ESR1

and CDH1. Firstly, **ESR1** or **Estrogen Receptor**, is usually expressed in the endometrioid histology (hormone dependent tumors) and poorly expressed in the serous histology (non-hormone dependent tumors). Considering the high percentage of recurrence in the serous histology, this biomarker seems to be more related to histology than recurrence status. However, Backes et al. [267] analyzed the expression of ESR1 in 320 pure EEC and observed that patients with ESR1 negative were more often diagnosed with higher grade, advanced stage, and LVSI. Secondly, **CDH1** or **Cadherin-1**, is a calcium-dependent cell adhesion protein involved in mechanisms regulating cell-cell adhesions, mobility and proliferation. In EC, it has been associated with the epithelial-mesenchymal transition (EMT) and has demonstrated to have a role in the progression of EC [268]. Importantly, its loss has been associated with worse prognostic factors and poorer survival [269]. In the TCGA analysis, we detected lower expression of ESR1 and CDH1 in recurrent EC patients, suggesting that they might be potential biomarkers predicting EC recurrence.

Recently, a new proteomic repository was created by the Clinical Proteomic Tumor Analysis Consortium (CPTAC), who undertook a comprehensive proteogenomic characterization of 95 prospectively collected endometrial carcinomas, comprising 83 EEC and 12 SEC tumors [242]. Proteins were analyzed by LC-MS in the Orbitrap Fusion Lumos mass spectrometer, similar to our discovery phase, and they detected the expression level of 12.524 proteins. Among the 5.747 proteins identified in our discovery phase, 5.419 proteins (95%) were represented in the CPTAC study, which permitted us to *in-silico* validate our results in parallel to explore novel biomarkers predicting recurrence. The output of the analysis permitted to elucidate 253 novel biomarkers predicting recurrence. Among those, we observed high expression of immunoglobulins in recurrent EC patients, suggesting that these tumors may be more aggressive and induce this immune reactivity in the primary tumor.

Any of the best significant proteins identified in the discovery phase was *in-silico* validated with the CPTAC cohort. One of the limitations of the CPTAC cohort is the lack of recurrent EC patients. Among the 95 patients, 8 patients were discarded because they were non-recurrent with a follow-up less than 2 years. Hence, there were 87 EC patients included in the study, but only 13 patients in the recurrence group. This low and unbalanced number of patients unable us to perform an optimal analysis and inhibit us to conduct different sub-analysis such in the discovery phase. The complexity of proteomic instrumentation for LC-MS introduces many possible sources of variability. It has been demonstrated that technical replicates overlapped by 35-60% [179]. Therefore,

Results

141

the combination of an unbalanced and low sample size and low reproducibility ends up with a restriction on using CPTAC for EC recurrence analysis.

Other studies have used the TCGA and CPTAC cohort to perform *in-silico* studies on EC. Li et al. [270] analyzed the expression of SIX1 in the TCGA and CPTAC and demonstrated that it was overexpressed in EC and was associated with adverse clinicopathological outcomes. Similarly, Liu et al. [271] identified 5 cell cycle-related genes from the TCGA and validated them in the CPTAC, which can distinguish high-risk patients from low-risk patients. Regarding recurrence, Coll-de la Rubia et al. [272] analyzed CPTAC data and developed a model of 5 proteins to predict recurrence-free survival (RFS). One protein from this panel, **TRA2B** or **Transformer-2 protein homolog beta**, is an RNA-binding protein which participates in the control of pre-mRNA splicing. As we have seen in Chapter 2, some mutations in spliceosome components enhance the activity, induces aberrant splice site selection and, finally, the production of novel isoform variants that supports tumorigenesis, allowing tumor progression, invasion and migration [244,245]. Notably, we also identified this biomarker in some sub-analysis of the endometrioid histology in our discovery phase (Chapter 2), indicating that it might be a potential EC biomarker predicting recurrence.

In conclusion, in this chapter we have analyzed the online repositories of TCGA and CPTAC datasets, and we have *in*-silico validated two of the potential biomarkers predicting recurrence identified in the discovery phase: ANXA1 and MYH9. These 2 biomarkers will be further validated in the verification phase.

CHAPTER 4

BIOMARKER VERIFICATION

TARGETED PROTEOMICS IN FFPE PRIMARY EC
TUMORS TO VERIFY EC BIOMARKERS PREDICTING
RECURRENCE IN THE ENDOMETRIOID HISTOLOGY

SPECIFIC BACKGROUND

The ideal biomarker pipeline consists of discovery, verification, and validation phase. Most studies are based on the initial phase identifying many biomarkers using a limited number of samples, without going to further validation phases; and other studies are based on a specific biomarker using many samples, without considering the possibility of developing signatures to increase the individual biomarker performance. In the middle, verification phase is a crucial step to prioritize the list of biomarkers identified in the discovery phase to increase the probability of moving clinically relevant biomarkers in the validation phase. This is of special importance in biomarker research of heterogeneous and multifactorial disease, such as cancer. It is expected that a single biomarker will not display enough power to affect clinical decisions. Therefore, the generation of predictive models through a panel of biomarkers is crucial to provide more accurate information. Importantly, MS targeted approaches have emerged in the last years as an optimal technology to conduct verification phases. Specifically, selected reaction monitoring (SRM) and parallel reaction monitoring (PRM) are ideal because they can reliably quantify numerous proteins in many samples.

The work presented in this chapter aimed to i) prioritize a list of potential biomarkers identified in Chapter 2 and 3 for the verification phase; ii) assess the potential to predict recurrence of the prioritized potential biomarkers in an independent cohort of EEC patients using targeted proteomics; iii) assess the relation of the identified protein biomarkers with clinic-pathological and molecular features; iv) develop protein panels to improve the individual performance of the biomarkers predicting recurrence.

MATERIAL AND METHODS

Patient recruitment

A total of 129 patients diagnosed with EEC were selected from the Haukeland University Hospital (Bergen, Norway), Vall Hebron Hospital (Barcelona, Spain), and Arnau de Vilanova Hospital (Lleida, Spain). All patients signed informed consent forms approved by the Ethical Committee for Clinical Investigation (CEIC) at each institution. Inclusion criteria was: i) patients diagnosed with endometrioid histology; ii) patients at FIGO stage IA, IB or II; iii) patients diagnosed without recurrence with a minimum follow-up of 3 years. Exclusion criteria was: i) patients diagnosed with other cancer type at the same time; ii) patients without all clinical data information.

Patients were divided in two different cohorts for the development and validation of a signature of biomarkers. Specifically, the BERGEN cohort was formed by 70 patients from Haukeland University Hospital, and the CAT cohort was formed by 59 patients from Vall Hebron Hospital (n=38) and Arnau de Vilanova Hospital (n=21). All patients were classified according to their clinic-pathological and molecular features, as described in Table 21 and Table 24. To molecularly classify EC patients, we followed the Proactive Molecular Risk Classifier for Endometrial Cancer (ProMisE) surrogate system as described in Chapter 2.

Protein identification by LC-MS analysis

❖ Sample preparation

The global procedure for macrodissection, tissue deparaffinization, protein extraction and quantification, and sample preparation for LC-MS analysis was performed as described in Chapter 1 and Annex 1.

❖ LC-MS analysis

Up to three unique peptides per protein were selected for targeted protein quantification (153 peptides from the 52 selected proteins in the discovery phase, see Annex 5). For each selected peptide, an isotopically-labeled peptide ($^{13}C_6^{15}N_4$-Arg, and $^{13}C_6^{15}N_2$-Lys) was spiked in the peptide mixtures and used as internal standard for quantification purposes. The amount of internal standard peptide to be spiked in each sample was evaluated using dilutions curves and the final concentration was chosen based on the following criteria: i) to be within the concentration range in which a linear response of the peptide was observed; ii) to have an area as close to the endogenous peptide area as possible.

Once the peptide mixture was optimized and incorporated in all samples, 1 μg of each sample was analyzed by Parallel Reaction Monitoring (PRM) using a 90 minutes gradient in the LTQ-Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-nLC 100 (Thermo Fisher Scientific) with a 50-cm C18 chromatographic column (Easy-Spray Column, PepMap RSLC C18). Peptide mixes were separated with a chromatographic gradient 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% buffer A in 79 minutes and then, to 35% buffer B and 65% buffer A in 11 minutes (Buffer A: 0.1% formic acid in water; Buffer B: 0.1% formic acid in acetonitrile). The

Orbitrap Fusion Lumos was operated in positive ionization mode with an EASY-Spray nanosource at 1.4kV.

A scheduled PRM method with 12 minutes windows was used for data acquisition with a quadrupole isolation window set to 1.4 m/z and MS2 scans over a mass range of m/z 300-2000, with detection in the Orbitrap mass analyzer at a 30K resolutions. MS2 fragmentation was performed using HCD fragmentation at a normalized collision energy of 30%, the auto gain control (AGC) was set at 1E5 and the maximum injection time at 54 ms. All data was acquired with XCalibur software.

❖ Data analysis

Product ion chromatographic traces corresponding to the targeted precursor peptides were evaluated with Skyline software based on i) co-elution of endogenous and internal standard peptides; ii) the number of detected traces; iii) correlation of the trace relative intensities between endogenous and internal standard peptides; iv) expected retention time. Normalization was performed based on the internal standard peptides. Protein abundance estimates was performed with the software package MSstats 3.8.2.

Statistical analysis

The statistical analysis was performed in R software (v.3.5.0) and in Graph Pad Prism (v.8.2.1) (GraphPad Software, La Jolla, CA, USA). The correlation between clinic-pathologic characteristics and recurrence was calculated using Fisher's exact test for 2 variables and Chi-square for 3 or more variables.

For the verification of biomarkers, the lineal model *limma* was used to differentiate non-recurrence and recurrence group. Apart from that, a logistic regression model was adjusted to the data to build predictive models able to classify the group of a new individual. Three different methods were used to build the predictive model: Multiple Adaptive Regression Splines (MARS), Elastic net (combines lasso and ridge regression), and Random Forest (generalization of classification trees). ROC curves were generated for each of these regression models; and the AUC, sensitivity, and specificity for discrimination between recurrence and non-recurrence group were obtained. P-values lower than 0.05 were considered statistically significant.

Results

RESULTS

List of potential biomarkers predicting recurrence

Targeted MS-approach was selected as the technology to use in the verification phase, since this enables us to quantify multiple peptides within a single analysis. Although this is a high-throughput technology, the number of peptides to analyze in each run is limited to 50-200 peptides and consequently, proteins must be selected prior to MS acquisition. Thanks to the targeted-MS approach, the analysis of the peptides is done with a high resolving power and the use of peptide isotopically labeled allow the quantification with a high level of selectivity.

In Chapter 2 and Chapter 3 we identified more than 600 potential biomarkers predicting recurrence in EEC primary tumors (Annex 2, Annex 3, and Annex 4). In order to prioritize those proteins to be analyzed in the verification phase, we used two different rationales:

1. Using the results from the quantitative analysis of the discovery phase, we selected 36 proteins, including 19 proteins from the general comparison with an adjusted p-value < 0.05 (TSTD1, MYH9, RPS14, SNRPF, PIGR, SRSF2, NUCKS1, CLIC4, PMVK, DPP7, GAA, PURB, POLR2H, TLN1, TGM2, CNN3, COL18A1, ACTN1, GUK1); 4 proteins from the general comparison with an adjusted p-value < 0.25 and log FC ±1 (PSMD3, GBP1, CLU, CA1); 11 proteins resulting from the comparison between no recurrence and late recurrence with an adjusted p-value <0.005 and log FC ±1 (SORBS2, AGA, MSN, SEC16A, SUB1, LSM8, DSG2, MIEN1, ANXA6, SEC61B, DDX1); and 2 proteins from 4 comparison with an adjusted p-value < 0.10 (ENAH, ANXA1).

2. Using the results from the absence/presence analysis of the discovery phase, we selected 16 proteins, including 10 proteins with a p-value < 0.01 (TPM1, PKM, AHCYL1, ALDH1A2, PSMD7, STK4, SNTB2, IL16, NSDHL, FERMT3) and 6 proteins with a p-value < 0.05 (BAG5, L1RE1, LONP1, LSR, NOP16, WBSCR22).

To summarize, a list of 52 proteins were prioritized for the verification phase (Table 20).

Table 20. Selection of potential biomarkers for the verification phase. *Thirty-six proteins selected from the quantitative analysis (QA), and sixteen proteins selected from the absence/presence analysis (AP). All statistical analysis shown in the table are from the general comparison in the endometrioid histology (recurrence versus non-recurrence).*

Uniprot Number	Protein Name	Gene name | Protein ID	Method of Choice	QUANTITATIVE ANALYSIS		ABSENCE/PRESENCE ANALYSIS			Protein Location
				Adjusted P-value	AUC	NO REC (32)	REC (31)	P-value	
Q8NFU3	Thiosulfate:glutathione sulfurtransferase	TSTD1 | TSTD1	QA	0,00154	0.883	100%	100%		Perinuclear region
P35579	Myosin-9	MYH9 | MYH9	QA	0,00154	0.859	100%	100%		Cytoskeleton
P20933	N(4)-(beta-N-acetylglucosaminyl)-L-asparaginase	AGA | ASPG	QA	0,10359	0.844	69%	87%		Lysosome
P32455	Guanylate-binding protein 1	GBP1 | GBP1	QA	0,07501	0.830	88%	84%		Cytosol, plasma mebrane, extracellular region or sectreded
P62263	40S ribosomal protein S14	RPS14 | RS14	QA	0,00853	0.821	100%	100%		Nucleus, cytosol, extracellular region or secreted
Q9BRT3	Migration and invasion enhancer 1	MIEN1 | MIEN1	QA	0,07799	0.819	84%	97%		Cytosol, plasma mebrane
Q9H1E3	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	NUCKS1 | NUCKS	QA	0,03302	0.814	100%	97%		Nucleus
P52434	DNA-directed RNA polymerases I, II, and III subunit RPABC3	POLR2H | RPAB3	QA	0,03452	0.808	94%	90%		Nucleus
Q01130	Serine/arginine-rich splicing factor 2	SRSF2 | SRSF2	QA	0,03302	0.804	100%	100%		Nucleus
P12814	Alpha-actinin-1	ACTN1 | ACTN1	QA	0,0461	0.796	100%	100%		Cytoskeleton, plasma membrane
Q9Y490	Talin-1	TLN1 | TLN1	QA	0,03825	0.791	100%	100%		Cytoskeleton, plasma membrane
Q92499	ATP-dependent RNA helicase DDX1	DDX1 | DDX1	QA	0,08709	0.782	97%	100%		Nucleus, cytosol
Q9Y696	Chloride intracellular channel protein 4	CLIC4 | CLIC4	QA	0,03302	0.780	100%	84%	*	Nucleus, mitochondrion, cytoskeleton, plasma membrane
Q96QR8	Transcriptional activator protein Pur-beta	PURB | PURB	QA	0,03302	0.769	100%	97%		Nucleus
P08133	Annexin A6	ANXA6 | ANXA6	QA	0,09785	0.767	100%	97%		Cytosol
Q16774	Guanylate kinase	GUK1 | KGUA	QA	0,04642	0.762	100%	100%		Cytosol
P21980	Protein-glutamine gamma-glutamyltransferase 2	TGM2 | TGM2	QA	0,04075	0.761	100%	97%		Nucleus, cytosol, extracellular region or secreted
O43242	26S proteasome non-ATPase regulatory subunit 3	PSMD3 | PSMD3	QA	0,05893	0.759	84%	77%		Nucleus, cytosol, extracellular region or secreted
O15027	Protein transport protein Sec16A	SEC16A | SC16A	QA	0,07799	0.759	84%	94%		Cytosol
P62306	Small nuclear ribonucleoprotein F	SNRPF | RUXF	QA	0,00853	0.755	91%	84%		Nucleus, cytosol
P00915	Carbonic anhydrase 1	CA1 | CAH1	QA	0,14891	0.754	94%	84%		Cytosol
Q15126	Phosphomevalonate kinase	PMVK | PMVK	QA	0,03302	0.752	100%	94%		Cytosol
O95777	U6 snRNA-associated Sm-like protein LSm8	LSM8 | LSM8	QA	0,07799	0.749	100%	100%		Nucleus
P04083	Annexin A1	ANXA1 | ANXA1	QA	0,07501	0.744	100%	100%		Nucleus, cytosol, plasma membrane, extracellular region or secreted
Q15417	Calponin-3	CNN3 | CNN3	QA	0,0461	0.733	100%	100%		Cytoskeleton, cytosol
P39060	Collagen alpha-1(XVIII) chain	COL18A1 | COIA1	QA	0,0461	0.729	100%	94%		Extracellular region or secreted
P60468	Protein transport protein Sec61 subunit beta	SEC61B | SC61B	QA	0,09644	0.715	100%	100%		Endoplasmic reticulum
Q14126	Desmoglein-2	DSG2 | DSG2	QA	0,17428	0.710	81%	84%		Plasma membrane
Q8N8S7	Protein enabled homolog	ENAH | ENAH	QA	0,06413	0.703	100%	100%		Cytoskeleton, cytosol
P10253	Lysosomal alpha-glucosidase	GAA | LYAG	QA	0,03302	0.700	100%	97%		Lysosome
P26038	Moesin	MSN | MOES	QA	0,08709	0.693	100%	100%		Cytoskeleton, plasma membrane
Q9UHL4	Dipeptidyl peptidase 2	DPP7 | DPP2	QA	0,03302	0.686	97%	97%		Lysosome, extracellular region or secreted
P01833	Polymeric immunoglobulin receptor	PIGR | PIGR	QA	0,03302	0.668	97%	87%		Plasma membrane, extracellular region or secreted
P10909	Clusterin	CLU | CLUS	QA	0,09644	0.668	100%	100%		Nucleus, cytosol, extracellular region or secreted
P53999	Activated RNA polymerase II transcriptional coactivator p15	SUB1 | TCP4	QA	0,17428	0.640	100%	100%		Nucleus
O94875	Sorbin and SH3 domain-containing protein 2	SORBS2 | SRBS2	QA	0,19554	0.501	81%	87%		Plasma membrane
P09493	Tropomyosin alpha-1 chain	TPM1 | TPM1	AP	-	-	72%	35%	***	Cytoskeleton
P14618	Pyruvate kinase	PKM | KPYM	AP	-	-	34%	3%	***	Nucleus, cytosol
O43865	Putative adenosylhomocysteinase 2	AHCYL1 | SAHH2	AP	-	-	66%	29%	***	Cytosol
O94788	Retinal dehydrogenase 2	ALDH1A2 | AL1A2	AP	-	-	69%	29%	***	Cytosol
P51665	26S proteasome non-ATPase regulatory subunit 7	PSMD7 | PSMD7	AP	-	-	84%	48%	***	Nucleus, cytosol, extracellular region or secreted
Q13043	Serine/threonine-protein kinase 4	STK4 | STK4	AP	-	-	63%	23%	***	Nucleus, cytosol
Q13425	Beta-2-syntrophin	SNTB2 | SNTB2	AP	-	-	53%	16%	***	Cytoskeleton
Q14005	Pro-interleukin-16; Interleukin-16	IL16 | IL16	AP	-	-	78%	42%	***	Nucleus, cytosol, extracellular region or secreted
Q15738	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	NSDHL | NSDHL	AP	-	-	56%	19%	***	Endoplasmic reticulum
Q86UX7	Fermitin family homolog 3	FERMT3 | URP2	AP	-	-	81%	42%	***	Podosome
P36776	Lon protease homolog, mitochondrial	LONP1 | LONM	AP	-	-	63%	26%	**	Mitochondrion
O43709	Probable 18S rRNA (guanine-N(7))-methyltransferase	WBSCR22 | BUD23	AP	-	-	34%	71%	**	Nucleus, cytosol
Q9UL15	BAG family molecular chaperone regulator 5	BAG5 | BAG5	AP	-	-	22%	0%	**	Nucleus, cytosol
Q86X29	Lipolysis-stimulated lipoprotein receptor	LSR | LSR	AP	-	-	47%	81%	*	Plasma membrane
Q9Y3C1	Nucleolar protein 16	NOP16 | NOP16	AP	-	-	31%	65%	*	Nucleus
Q9UN81	LINE-1 retrotransposable element ORF1 protein	L1RE1 | LORF1	AP	-	-	0%	16%	*	Nucleus, cytosol

Results

Although the discovery phase results were the main source to select the biomarkers for the verification phase, ANXA1 and MYH9 proteins that were detected in the TCGA analysis were also included. Additionally, we confirmed that 40 proteins (77%) out of 52 were also good biomarkers considering molecular classification. Specifically, 17, 32, and 5 proteins were significant biomarkers predicting recurrence in the MSI, low-CN and high-CN groups, respectively. Top 10 biomarkers from the quantitative analysis (TSTD1, MYH9, ASPG, GBP1, RS14, MIEN1, NUCKS, RPAB3, SRSF2, and ACTN1) are shown in Figure 36.
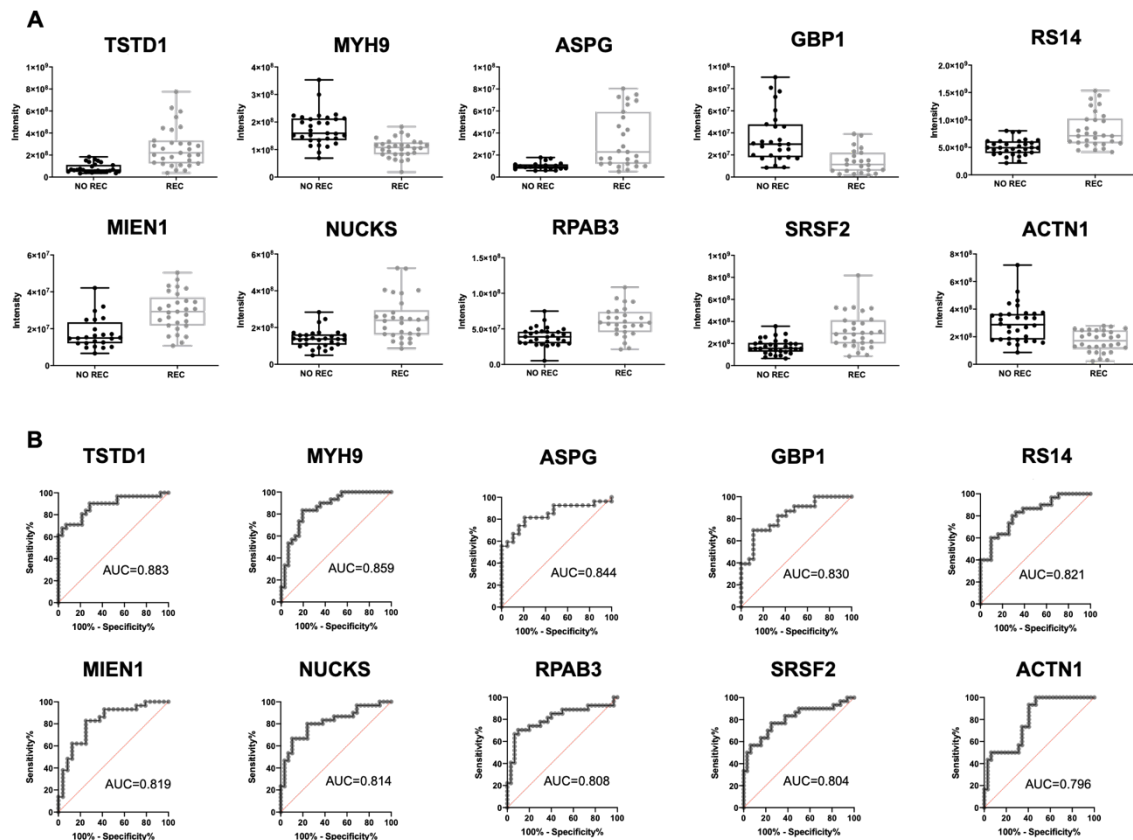
Figura 36. Results for the top 6 biomarkers. A. *Scattering plots of the top 6 biomarkers.* ***B.*** *Receiver Operating Characteristics (ROC) curves of the top 6 biomarkers.*

Targeted MS analysis of potential biomarkers predicting recurrence in 2 different cohorts: BERGEN and CAT

During the study design, we selected a total of 2-3 peptides (153 peptides) to analyze the 52 potential biomarkers. Those peptides were detected in our discovery phase or in other targeted MS analysis, and importantly, are proteotypic peptides (i.e., peptides that uniquely identify each protein and are consistently observed when a sample mixture is interrogated by MS) (see Annex 5). The MS acquisition detected 93 peptides, and this

permitted us to detect and quantify our protein biomarkers with at least 1-2 peptides per protein. The study cohort was 129 EEC patients. Specifically, 70 EEC patients were from Haukeland University Hospital (Bergen, Norway), 21 EEC patients were from Arnau de Vilanova Hospital (Lleida, Spain), and 38 EEC patients were from Vall Hebron Hospital (Barcelona, Spain).

For sample preparation, protein extraction of the FFPE primary EEC tissues of the Haukeland University Hospital was done in Bergen, while samples of the Arnau de Vilanova Hospital and Vall Hebron Hospital was done in Barcelona. Later on, protein digestion, detergent elimination, desalting, and MS analysis was performed in Barcelona for all samples. Unfortunately, protein extracts from Bergen's patients were defrosted during transportation and this affected the stability and expression of some proteins. We observed three different patterns of alterations: proteins with the two peptides degraded (such as ANXA6), proteins without degradation (such as FINC), and proteins with one peptide degraded and not the other one (such as SAHH2) (Figure 37). Overall, MS analysis revealed that 41 out of 93 peptides (44%) were degraded. Since there was not a homogeneous degradation pattern across samples, we could not compensate for this effect during the data analysis and therefore, we decided to perform the statistical analysis separately for each of the cohorts (BERGEN cohort and CAT cohort).
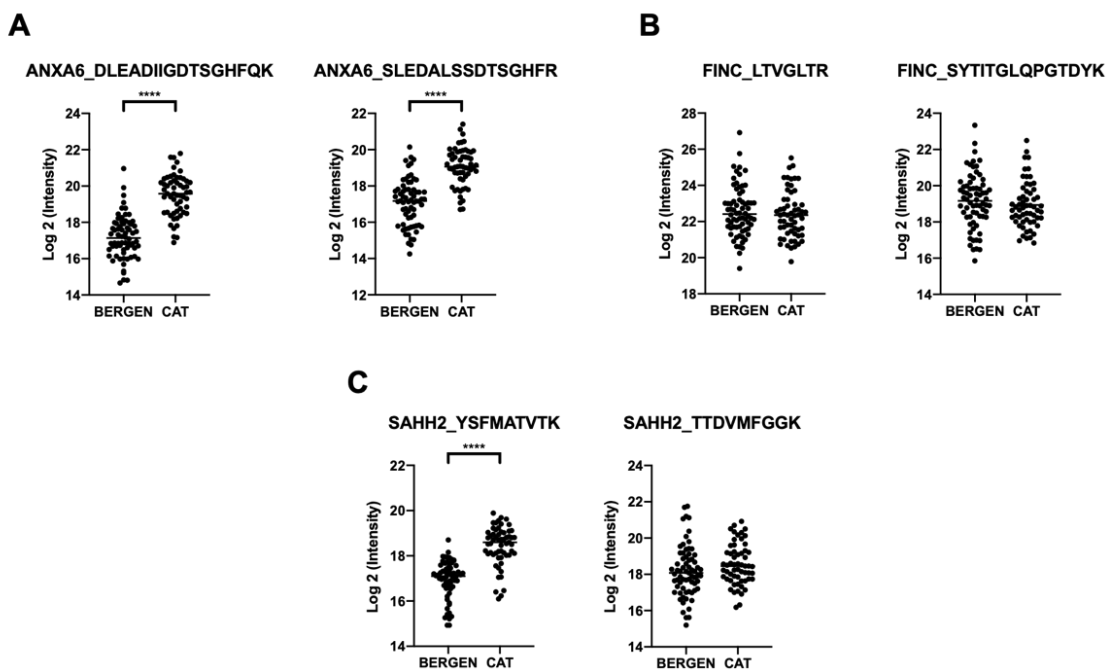
Figura 37. Distribution of the intensities obtained by targeted MS analysis to discriminate between BERGEN cohort and CAT cohort (Arnau de Vilanova Hospital and Vall Hebron Hospital). A. *Example of a protein with the two peptides analyzed degraded.* **B.** *Example of a protein with the two peptides analyzed without degradation.* **C.** *Example of a protein with one peptide degraded and not the other one.*

Verification of potential biomarkers predicting recurrence in the BERGEN cohort

Seventy EEC patients were included in the BERGEN cohort, including 34 non-recurrent EEC patients and 36 recurrent EEC patients. The clinical and pathological characteristics are summarized in Table 21.

Table 21. *Clinic-pathological characteristics of women diagnosed with endometrioid EC enrolled in the verification phase from Haukeland Hospital (BERGEN cohort).*

		NO Recurrence	Recurrence	P-value
Total Number		34	36	
Age (years)		68 (46-91)	69 (46-90)	0.49
Grade				0.07
	1	14	7	
	2	13	18	
	3	7	11	
FIGO Stage				0.46
	IA	2	5	
	IB	26	25	
	II	6	6	
Myometrial Invasion				0.37
	<50%	5	9	
	>50%	29	27	
LVSI				>0.99
	NO	28	25	
	YES	0	0	
Molecular Classification				0.46
	POLE	0	1	
	MSI	11	8	
	Low-CN	17	20	
	High-CN	2	4	
Risk Classification				0.56
	Low	0	1	
	Intermediate	20	16	
	High-Intermediate	8	10	
	High	2	3	
Recurrence Site				N/A
	Local	-	14	
	Regional	-	10	
	Distant	-	12	
Time to Recurrence				N/A
	Early (<24 months)	-	23	
	Late (>24 months)	-	12	
Status				<0.0001
	Alive without disease	29	2	
	Alive with disease	0	8	
	Dead of disease	0	23	
	Dead other causes	5	3	

The BERGEN cohort was age-balanced with a mean age of 68 and 69 years for the non-recurrent and the recurrent group, respectively. As expected, there were no statistical differences in any of the clinic-pathological features (grade, FIGO stage, myometrial invasion, LVSI, molecular classification, and risk classification), except for the status of the patient. The mean time to recur was 24 months (2 – 72). Non-recurrent patients were alive without disease after a follow-up of 92 months (37 – 213) or dead for other causes, and 64% of recurrent patients were death of disease.

To assess the potential of the 52 candidate biomarkers, we compared the abundance of each protein between 34 non-recurrent and 36 recurrent EEC patients. The relative levels (light/heavy ratio) of the 93 peptides were subjected to the lineal model *limma* for their comparison between recurrence and non-recurrence. Twenty-six peptides corresponding to 21 proteins showed significant differences between the two groups with a p-value < 0.05 (Table 22).

Table 22. Peptides showing statistical differences between non-recurrence (n=34) and recurrence (n=36) EEC patients with a p-value < 0.05 (BERGEN cohort). In bold, peptides showing statistical significance with adjusted p-value < 0.05.

Protein Name	Peptide	Log FC	P-value	Adjusted p-value	AUC
ANXA1	**DITSDTSGDFR**	**-0.912**	**0.0003**	**0.0259**	**0.627**
	SEDFGVNEDLADSDAR	**-0.853**	**0.0007**	**0.0259**	**0.612**
LYAG	VTSEGAGLQLQK	0.749	0.0045	0.0784	0.531
	WGYSSTAITR	0.522	0.0338	0.1413	0.533
PMVK	LLDTSTYK	0.535	0.0053	0.0784	0.535
	EAYGAVTQTVR	0.506	0.0244	0.1249	0.551
LSR	LLEEAVR	0.593	0.0053	0.0784	0.584
	SGDLPYDGR	0.528	0.0176	0.1047	0.578
TSTD1	GLQATQLAR	0.701	0.0078	0.0803	0.529
	SLGYTGAR	0.622	0.0119	0.0925	0.551
ASPG	**FLPSYQAVEYMR**	**0.970**	**0.0006**	**0.0259**	**0.547**
CLUS	ELDESLQVAER	1.007	0.0062	0.0801	0.569
ANXA6	SLEDALSSDTSGHFR	-0.622	0.0076	0.0803	0.597
CLIC4	HPESNTAGMDIFAK	-0.636	0.0095	0.0849	0.557
TLN1	ALEATTEHIR	-0.855	0.0099	0.0849	0.523
SNTB2	SPSLGSDLTFATR	-0.659	0.0131	0.0925	0.498
CNN3	GFHTTIDIGVK	-1.414	0.0136	0.0925	0.606
SC16A	QALQSTPLGSSSK	0.628	0.0154	0.0982	0.525
MYH9	IAQLEEQLDNETK	-0.563	0.0184	0.1047	0.554
GBP1	NEIQDLQTK	-0.672	0.0228	0.1225	0.605
DSG2	IHSDLAEER	0.698	0.0299	0.1413	0.544
PSMD3	EQQDLEFAK	-0.384	0.0335	0.1413	0.566
MOES	ALTSELANAR	-0.454	0.0346	0.1413	0.512
SC61B	FYTEDSPGLK	-0.342	0.0458	0.1648	0.549
TPM1	MEIQEIQLK	-0.781	0.0461	0.1648	0.492
KGUA	VAVQAVQAMNR	0.317	0.0468	0.1648	0.487

Nine proteins were upregulated in recurrent EEC patients (LYAG, PMVK, LSR, TSTD1, ASPG, CLUS, SC16A, DSG2, and KGUA), whereas twelve proteins were downregulated in recurrent EEC patients (ANXA1, ANXA6, CLIC4, TLN1, SNTB2, CNN3, MYH9, GBP1, PSMD3, MOES, SC61B, and TPM1) (Figure 38A). Among the significant proteins, we found that both peptides from ANXA1, LYAG, PMVK, LSR, and TSTD1 were statistically significant with similar values of fold change and in the same tendency of expression as in the discovery phase. In particular, ANXA1 showed the highest individual AUC value (0.612 – 0.627) and was also significant with an adjusted p-value < 0.05 (Figure 38B).
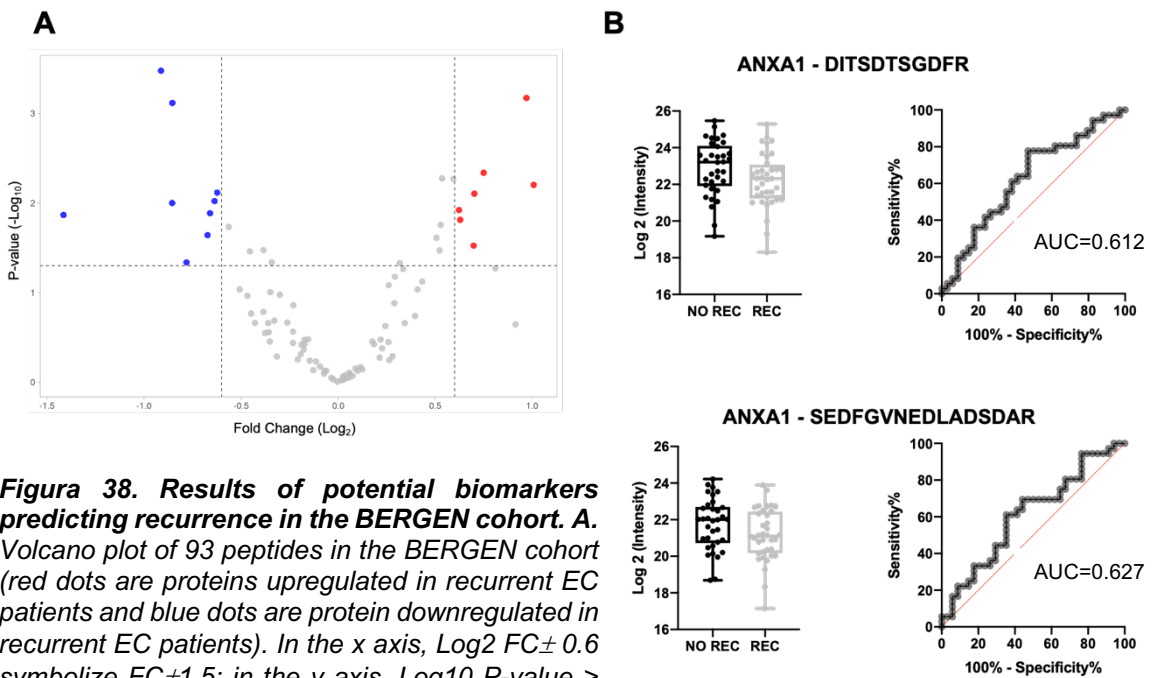
Figura 38. Results of potential biomarkers predicting recurrence in the BERGEN cohort. A. Volcano plot of 93 peptides in the BERGEN cohort (red dots are proteins upregulated in recurrent EC patients and blue dots are protein downregulated in recurrent EC patients). In the x axis, Log2 FC± 0.6 symbolize FC±1.5; in the y axis, Log10 P-value > 1.3 symbolize p-value < 0.05 B. Boxplots and ROC curves of both peptides of ANXA1.

The combination of multiple proteins was evaluated to improve the prediction of recurrence over the use of a single protein, as single biomarkers may not necessarily reflect the heterogeneity of EC. Therefore, we used the BERGEN cohort to develop a predictive model. Specifically, we evaluated three different models: Multiple Adaptive Regression Splines (MARS), Elastic net (EN), and Random Forest (RF) (Table 23)

154

Table 23. Development of predictive models to predict recurrence in the BERGEN cohort.
Three different models were evaluated: Multiple Adaptive Regression Splines (MARS), Elastic Net (EN), and Random Forest (RF). Each analysis shows sensitivity, specificity, AUC value, and the top 5 proteins involved in the predictive model.

Predictive Model	Sensitivity	Specificity	AUC	Top 5 Proteins
MARS	0.616	0.566	0.691	DPP2, PSMD7, ANXA1, KCRB, KGUA
EN	0.625	0.625	0.741	SC61B, KPYM, PMVK, SRSF1, ANXA1
RF	0.616	0.700	0.697	ANXA1, CNN3, DPP2, PMVK, PDLI1

The best predictive model was a 5-protein panel constructed with the Elastic net (EN), which includes SC61B, KPYM, PMVK, SRSF2, and ANXA1. Combining several proteins, we could increase the AUC value up to 0.741, highlighting the development of predictive models.

Verification of potential biomarkers predicting recurrence from the CAT cohort

Fifty-nine EEC patients were included in the CAT cohort, including 40 non-recurrent EEC patients and 19 recurrent EEC patients. The clinical and pathological characteristics are summarized in Table 24.

The CAT cohort was age-balanced with a mean age of 65 and 68 years for the non-recurrent and the recurrent group, respectively. As expected, there were no statistical differences in any of the clinic-pathological features (grade, FIGO stage, myometrial invasion, LVSI, molecular classification, and risk classification), except for the status of the patient. The mean time to recur was 30 months (5 – 58). Non-recurrent patients were alive without disease after a follow-up of 67 months (38 – 147) or dead for other causes, and 42% of recurrent patients were death of disease.

Results

Table 24. *Clinic-pathological characteristics of women diagnosed with endometrioid EC enrolled in the verification phase from Vall Hebron Hospital and Arnau de Vilanova Hospital (CAT cohort).*

	NO Recurrence	Recurrence	P-value
Total Number	40	19	
Recruitment Hospital			0.04
Vall Hebron	22	16	
Arnau de Vilanova	18	3	
Age (years)	65 (49-86)	68 (55-81)	0.15
Grade			0.17
1	8	1	
2	8	4	
3	24	14	
FIGO Stage			0.58
IA	3	3	
IB	24	10	
II	13	6	
Myometrial Invasion			>0.99
<50%	8	4	
>50%	32	15	
LVSI			0.75
NO	27	9	
YES	13	6	
Molecular Classification			0.06
POLE	6	1	
MSI	17	5	
Low-CN	14	10	
High-CN	3	3	
Risk Classification			0.24
Low	6	1	
Intermediate	11	3	
High-Intermediate	20	12	
High	3	1	
Recurrence Site			N/A
Local	-	2	
Regional	-	6	
Distant	-	11	
Time to Recurrence			N/A
Early (<24 months)	-	8	
Late (>24 months)	-	11	
Status			<0.0001
Alive without disease	39	5	
Alive with disease	0	6	
Dead of disease	0	8	
Dead other causes	1	0	

To assess the potential of the 52 candidate biomarkers, we compared the abundance of each protein between 40 non-recurrent and 19 recurrent EEC patients. The relative levels (light/heavy ratio) of the 93 peptides were subjected to the lineal model *limma* for their comparison between recurrence and non-recurrence. Nine peptides corresponding to 7 proteins showed significant differences between the two groups with a p-value < 0.05 (Table 25).

Table 25. *Peptides showing statistical differences between non-recurrence (n=40) and recurrence (n=19) EEC patients with a p-value < 0.05 (CAT cohort).*

Protein Name	Peptide	Log FC	P-value	Adjusted p-value	AUC
ASPG	FLPSYQAVEYMR	0.591	0.0148	0.5513	0.691
	NVIPDPSK	0.617	0.0459	0.5513	0.664
ANXA1	DITSDTSGDFR	-0.564	0.0432	0.5513	0.612
	SEDFGVNEDLADSDAR	-0.582	0.0476	0.5513	0.599
PSMD7	IVGWYHTGPK	-1.529	0.0007	0.0715	0.563
ENAH	LEQEQLER	0.466	0.0294	0.5513	0.696
KGUA	VAVQAVQAMNR	0.351	0.0409	0.5513	0.642
SAHH2	TTDVMFGGK	-0.657	0.0435	0.5513	0.642
PMVK	LLDTSTYK	0.477	0.0486	0.5513	0.666

Four proteins were upregulated in recurrent EEC patients (ASPG, ENAH, KGUA, and PMVK), whereas three proteins were downregulated in recurrent EEC patients (ANXA1, PSMD7, and SAHH2). Importantly, both peptides from ANXA1 and ASPG and one peptide from PMVK and KGUA were significantly differentiated with similar values of fold change and in the same tendency of expression as in the discovery phase and in the verification performed with the BERGEN cohort (Table 26). Thus, we were able to verify in independent studies, the potential of ASPG, ANXA1, PMVK, and KGUA as biomarkers to predict recurrence in EEC patients.

Results

Table 26. Four differentially expressed proteins in the discovery and verification phase (BERGEN and CAT cohort).

	Study	Peptide	Log FC	P-value	Adjusted p-value	AUC
ANXA1	Discovery	-	-0,778	0,0019	0,075	0,744
	BERGEN	DITSDTSGDFR	-0,912	0,0003	0,026	0,627
		SEDFGVNEDLADSDAR	-0,853	0,0007	0,026	0,612
	CAT	DITSDTSGDFR	-0,564	0,0432	0,551	0,612
		SEDFGVNEDLADSDAR	-0,582	0,0476	0,551	0,599
ASPG	Discovery	-	0,710	0,0054	0,104	0,844
	BERGEN	FLPSYQAVEYMR	0,97	0,0006	0,026	0,547
	CAT	FLPSYQAVEYMR	0,591	0,0148	0,551	0,691
		NVIPDPSK	0,617	0,0459	0,551	0,664
PMVK	Discovery	-	0,688	0,0002	0,033	0,752
	BERGEN	LLDTSTYK	0,535	0,0053	0,078	0,535
		EAYGAVTQTVR	0,506	0,0244	0,125	0,551
	CAT	LLDTSTYK	0,477	0,0486	0,551	0,666
KGUA	Discovery	-	0,528	0,0006	0,046	0,762
	BERGEN	VAVQAVQAMNR	0,317	0,0468	0,165	0,487
	CAT	VAVQAVQAMNR	0,351	0,0409	0,551	0,642

Similar to the BERGEN cohort, we used the CAT cohort to develop a predictive model. The same three different models were used for the algorithm development: MARS, EN, and RF (Table 27). The best predictive model was again the Elastic net (EN). A 5-protein panel including KPYM, SC61B, GBP1, PSMD7, and MYH9 yield an accuracy of 0.713, with sensitivity of 80% and specificity of 50%. In this case, combining several proteins did not improve greatly the individual accuracy of our biomarkers, since ASPG had an individual AUC of 0.691.

Table 27. Development of predictive models to predict recurrence in the CAT cohort. *Three different models were evaluated: Multiple Adaptive Regression Splines (MARS), Elastic Net (EN), and Random Forest (RF). Each analysis shows sensitivity, specificity, AUC value, and the top 5 proteins involved in the predictive model.*

Predictive Model	Sensitivity	Specificity	AUC	Top 5 Proteins
MARS	0.775	0.450	0.650	KPYM, SC61B, MYH9, BAG5, SC16A
EN	0.800	0.500	0.713	KPYM, SC61B, GBP1, PSMD7, MYH9
RF	0.975	0.100	0.638	KPYM, PSMD7, ANXA1, SC61B, COIA1

DISCUSSION

The identification of sensitive and specific biomarkers to improve the detection of recurrence is an important clinical need. As discussed in Chapter 2, most efforts and investments have been done to identify diagnostic and/or prognostic EC biomarkers, and few studies have been focused on recurrence. Additionally, there is a poor translation of the results from those studies to the clinical application. The main factor is because of the lack of verification studies that act as a bridge between discovery and validation phases, which has been defined as the bottleneck of the biomarker pipeline [273]. In general, discovery studies generate a large list of differentially expressed proteins, but many of those candidate biomarkers are never validated or are false positive because of the small sample size analyzed. Consequently, there is a need to verify those candidate biomarkers that can enter into large validation phases. In this study, we selected a list of potential biomarkers from the discovery phase and the *in-silico* analysis and verified them by targeted MS analysis in a larger and independent cohort of patients.

Targeted MS approaches are a new technology for biomarker verification because they combine precision, sensitivity, and absence of missing values. Typically, selected reaction monitoring (SRM) method has been chosen for the quantification of peptides [183,274]. However, SRM is time consuming, it only analyzes a limited number of transitions per peptide that can generate bias, and all the transitions must be validated in a representative number of samples before the analysis [186]. Therefore, we implemented parallel reaction monitoring (PRM) method, a new generation of targeted MS approach with high resolution. The main advantage of PRM is the capacity to analyze all fragments in a single MS analysis, increasing the resolution and reducing the risk of inferences due the complexity of the background. In addition, PRM acquisition allow the quantification of 100 pair of peptides in one analysis with excellent precision [188]. Finally, the implementation of liquid chromatography (LC) prior MS analysis is faster and easier with the PRM acquisition because the selection of the fragmented ions is performed post-acquisition [275].

Another strength of this study is the use of FFPE tissue in a targeted MS approach. As discussed in Chapter 1, FFPE tissue represent an inexpensive tissue storage system widely used which permits to overcome the great limitation of sample availability to study EC recurrence. In this thesis, we have confirmed the optimal use FFPE specimens in the proteomic workflow to discover and verify potential biomarkers predicting recurrence using untargeted and targeted MS approaches.

Results

159

In this chapter, we selected 52 potential candidates based on the discovery phase, and the bioinformatic analysis of TCGA and CPTAC. We evaluated the protein abundance in the primary tumor of 129 EEC patients, 74 without recurrence and 55 with recurrence. Unfortunately, protein extracts from Bergen were defrosted during transportation affecting the stability of some proteins. Stability is an inherent characteristic of a protein, meaning that not all proteins behave equally under the same storage conditions. While a protein can tolerate freezing and thawing many times or stay at room temperature for hour or days, other proteins can be denatured and degraded [276]. We detected that 41 out of 93 peptides (44%) were degraded in the BERGEN cohort. Considering this analytical bias, we decided to split up all the patients in two different cohorts according to the hospital location (BERGEN and CAT). Consequently, instead of having a robust study with 129 patients, we performed two independent studies with 70 and 59 EEC patients, respectively. This is the main limitation of our verification phase, which could explain the low significance at the adjusted p-value level in both cohorts reducing the potential of the study.

On the one hand, in the BERGEN cohort we identified 26 peptides corresponding to 21 proteins that could differentiate recurrent and non-recurrent EEC patients. Importantly, we developed a predictive model combining 5-proteins that improve the predictive power of the individual proteins up to an AUC value of 0.741. On the other hand, in the CAT cohort we identified 9 peptides corresponding to 7 proteins that could differentiate recurrent and non-recurrent EEC patients. In this case, the predictive model of a 5-protein panel performed with a similar AUC value (0.713) than the one obtained for the best individual biomarker (0.691). Overall, we highlight the potential of 4 protein biomarkers that were differentially expressed in both studies: ANXA1, ASPG, PMVK, and KGUA.

Firstly, as discussed in Chapter 3, **ANXA1 or Annexin 1** have several roles in cell proliferation, differentiation, motility, trafficking, etc. It is overexpressed in tumors in comparison to normal endometrium [261], but it has been poorly investigated in EC recurrence. Lai et al. [259] constructed a prognostic signature based on 9 proteins, which could divide patients into low- and high-risk with distinct prognoses combining the protein expression data and RNA-seq expression from the TCGA. One of these proteins was ANXA1, which was downregulated in high-risk compared to low-risk EC patients, suggesting that ANXA1 may serve as a negative biomarker in cancer development and in the progression of EC.

Secondly, **ASPG or Asparthylglucosaminidase** is a member of the N-terminal nucleophile hydrolase family and is involved in the catabolism of N-linked

oligosaccharides of glycoproteins. Mutations in this gene are associated with the lysosomal storage disease aspartylglycosaminuria that results in progressive neurodegeneration. ASPG has a potential cytotoxic activity toward leukemia cells that are dependent on their external supply of L-asparagine because it depletes the extracellular and intracellular L-asparagine reservoirs inducing apoptosis [277]. In our results we detected high expression level in EC patients with recurrence. Although ASPG has not been well-related with cancer, in The Human Protein Atlas we observed a tendency that EC patients with high expression of ASPG presented less overall survival, which is in line with our results.

Thirdly, **PMVK or Phosphomevalonate Kinase** is a peroxisomal enzyme that catalyzes the conversion of mevalonate 5-phosphate to mevalonate 5-diphosphate, which is the fifth step in the mevalonate pathway of isoprenoid biosynthesis. Mutations in this gene are linked to certain types of porokeratosis. In our study, PMVK is highly expressed in EEC recurrent patients. To our knowledge, this is the first time studying PMVK in EC. Kim et al. [278] conducted label-free LC-MS in high-grade serous ovarian carcinoma (HGSOC) and identified PMVK as a prognostic biomarker. They validated PMVK via IHQ and observed that patients with high expression levels of PMVK presented better PFS than those with low expression levels. In addition, high expression of PMVK was significantly associated with platinum sensitivity and improved the survival of patients with HGSOC. Similarly, Shen et al. [279] showed that PMVK was positively associated with drug response in ER positive cells from breast cancer. In The Human Protein Atlas, PMVK is also positively related with overall survival in ovarian cancer and renal cancer.

Fourthly, **KGUA or Guanylate Kinase** is an enzyme that catalyzes the transfer of a phosphate group from ATP to guanosine monophosphate (GMP) to form guanosine diphosphate (GDP). This protein is thought to be a good target for cancer chemotherapy. Da Rocha et al. [280] hypothesized that KGUA may act as a tumor suppressor in pituitary tumorigenesis because they found that mRNA expression was down regulated in the metastasis. In contrast, Wolfe et al. [281] demonstrated that the formation of a metabolic compartment for localized purine biosynthesis at the leading edge (including KGUA), may promote nucleotide metabolism for cell migration and metastasis in cancers. This correlates with our results where we found high expression levels of KGUA in recurrent EC patients.

There is few information about these proteins and cancer, except for ANXA1. Therefore, they all merit further validation in order to unveil their role in cancer and specifically in EC recurrence. Also, their predictive power of recurrence should be validated in large

and independent cohorts of patients. If validated, they could be implanted in clinical practice to develop a more accurate risk classification system.

In conclusion, in this chapter we have proved the efficiency of LC-MS operated in PRM acquisition in order to verify a large number of potential biomarkers selected from Chapter 2 and 3. In addition, we have also defined a panel of protein that achieve the best performance to predict recurrence in one of the cohorts, and we have verified 4 biomarkers predicting recurrence: ANXA1, ASPG, PMVK, and KGUA.

CHAPTER 5

NOVEL TREATMENTS IN EC

PRECLINICAL STUDY TO EVALUATE THE EFFICACY
OF NIRAPARIB AND SYD985 IN HER2-POSITIVE
ENDOMETRIAL CANCER PDX MODELS

SPECIFIC BACKGROUND

The gold-standard treatment for EC is surgery followed by adjuvant treatment that can vary from brachytherapy, radiotherapy, chemotherapy, hormone therapy or combination of the aforementioned. Regarding chemotherapy, standard of care is based on paclitaxel and cisplatin [26]. However, this combination is not specific for EC and many patients do not respond to this drug regimen. Consequently, in the last years has appeared targeted treatments for specific subpopulation of EC patients. For instance, in the high-CN group, which is the most aggressive and lethal subtype, targeting the homologous recombination deficiency with PARP inhibitors (such as niraparib) and the human epidermal growth factor 2 (such as SYD985) have demonstrated to be good therapeutic approaches [44].

In order to position novel treatments at the disposal of cancer patients, pharmaceutical companies spent a range of $314 million to $2.8 billion in research and development [282]. Despite this enormous budgets, the approval of novel oncology drugs during past decade continues to be modest and this failure can be attributed, at least in part, to the lack of clinically relevant models to underpin the clinical development [283]. Amongst the large repertoire of *in vivo* systems used to study cancer, mouse models represent the most widely used system. Specially, patient-derived tumor xenografts (PDXs) have been established as a useful tool for translational research [210]. PDX models mimic not only the pathohistological and genetic/epigenetic features of original tumor tissues, but also therapeutic responses to anti-cancer treatments. Mounting evidence demonstrates that PDX models have the potential of effectively predict the efficacy of both conventional and novel anti-cancer therapeutics, suggesting that these models could be employed in pre-clinical or co-clinical trials [284]. Additionally, these models can be used to predict treatment response in particular subpopulations of patients, since molecular data of the models is easily available, and therapies can be tested according to specific molecular profiles.

The work presented in this chapter aimed to i) assess toxicity of SYD985 and niraparib in EC PDX models; and ii) analyze the efficacy of SYD985 and niraparib in monotherapy or in combination in EC PDX models with aberrant expression of *erbb2* (HER2).

MATERIAL AND METHODS

Patient recruitment

All patients included in the study signed an informed consent. The study was approved by the Ethical Committee for Clinical Investigation (CEIC) of Vall d'Hebron Hospital, and it follows national and international guidelines regulations on data protection and confidentiality.

A total of 43 patients diagnosed with EC participated in this study. For each patient, a fragment of the resected primary tumor or metastasis was macroscopically collected by an experienced pathologist. Medical records, as well as clinicopathological data, were also available in a dissociated, pseudo-anonymized manner.

Tissue microarray construction and immunohistochemistry

Tissue microarrays (TMA) were constructed from patients' tumor and from PDX tumors representing the same EC patients. FFPE tumor samples were assembled into TMA using a Tissue Microarrayer (Beecher Instrument) with a core size of 1 mm. Three cores per cases were selected from each tumor after reviewing H&E of the whole tumor slides and were precisely arrayed in a new paraffin block.

Five (5) µm sections were obtained from all TMA blocks. HER2 IHQ was performed in a fully automated system using an automated slide Stainer (Benchmark ULTRA; Ventana Medical Systems, Tucson, AZ). The slides were deparaffinized (EZ prep TM (10x), Ventana Medical Systems), and antigen retrieval was performed at 97ºC for 64 min with CC1 solution (EDTA buffer, Ventana Medical Systems). The slides were then incubated with the primary antibody (Her2/neu (4B5), rabbit monoclonal #790-2991, Ventana Medical Systems) at 37ºC for 32 min followed by amplification using an UltraView Polymer Detection Kit (Ventana Medical Systems) with diaminobenzidine as the chromogen. Finally, slides were counterstained with hematoxylin and deparaffinized. An experienced pathologist evaluated the intensity of the staining of tumoral epithelial cells, ranging from 0 –no expression- to 3 –strongly positive.

Patient-derived tumor xenograft (PDX) generation, maintenance, expansion and randomization method

All procedures involving animals were approved by the Ethical Committee for Animal Research (CEEA) at Vall d'Hebron Institute of Research following national and international guidelines for animal welfare. PDX mice models were developed by subcutaneous inoculation of 50 mm^3 tumor piece obtained directly from patients. Once a tumor reached 1000-2500 mm^3 volume, it was freshly excised from the animal, rinsed with sterile saline solution, divided in pieces and transferred into healthy 6-8 weeks-old mouse/mice. In this study, we amplify the PDX model in 5 SCID mice (CB17/Icr-Prkdscid/IcrIcoCrl) (Phase I), and then, we expanded to 30 Cesc-1 KO SCID mice provided by Charles River Laboratories (Phase II). The cohort of PDX models engrafted in the Phase II were used in the preclinical study.

The inclusion criteria to initiate treatment was tumor volume range 75 to 250 mm^3, normal behavior (locomotion, mobility, grooming) and good aspect in general (skin, weight, fur) of the animal. Exclusion criteria was tumor volume out of range (>60 mm^3 or <250 mm^3), delayed tumor growth compared to other mice in Phase II, and mice with signs of distress, pain, or bad aspect in general (loss of weight, mobility). Each mice surpassing the inclusion criteria was incorporated in a treatment group.

The first set of animals were allocated in each treatment group following a randomization method: tumor volumes were recorded and listed from higher to lower volume. Only the biggest tumor of each animal was considered for the initial step of the randomization process. Once listed, a number was assigned to each tumor in the following order, 1 2 3 4 5 – 5 4 3 2 1, where each number represent a different treatment group. Once each animal was assigned to a group, we included the information of the contralateral tumor volume and calculated the new tumor volume mean. In case of a great imbalance among groups, animals were distributed manually one by one to equalize the mean tumor volume of the treatment group. For the subsequent addition of animals (i.e., animals that do not fulfill the inclusion criteria at the moment of randomization), we considered the following parameters in order to assign each animal to a group (Dynamic balanced randomization):

- Tumor volume shall not affect the mean tumor volume of the group.
- Loss of animals in any specific group due to incidences related or not with treatment.
- Homogenization of groups.

Drug testing

A total of 15 PDX models representing different EC patients were selected based on a positive HER2 expression (i.e., HER2 from 1+ to 3+ expression). Each PDX model was amplified to 30 mice (Phase II) and those fulfilling the inclusion criteria were randomized into 5 treatment groups: placebo (solution 092v12 + methylcellulose [0.5%]), SYD989 (1 mg/kg), SYD985 (1 mg/kg), niraparib (40 or 50 mg/kg), or double treated with SYD985 + niraparib. SYD989 is the trastuzumab antibody alone and is used as an internal control for SYD985. Treatments were given as follows: placebo and niraparib by oral gavage daily for 21 days; and SYD989 and SYD985 intravenously in a single dose the first day of treatment.

Table 28. Dose, route, and schedule administration of each treatment.

Treatment	Dose	Route administration	Schedule administration
Placebo	Solution 092v12 + methylcellulose (0.5%)	Oral gavage	Daily for 21 days
SYD989	1 mg/kg	Intravenously	Single dose
SYD985	1 mg/kg	Intravenously	Single dose
Niraparib	40 – 50 mg/kg	Oral gavage	Daily for 21 days

Tumor size was measured 2-3 times per week with a digital caliper, and tumor volume was calculated using the formula: length x width^2 / 2 = mm^3. Mice were also weighted 2 times per week and monitor their welfare. Criteria for endpoint was: one tumor volume > 1500 mm^3 or two tumor volumes which sum is > 2500 mm^3; body weight loss (>20% body mass loss); distress and pain curve; or 12 weeks from treatment onset. At endpoint, mice were euthanized, and tumors were harvested and collected for further analyses.

Additionally, blood sample was collected from mice 96 hours after the first day of treatment to detect the presence of niraparib, SYD985 and SYD989.

Histopathological analysis of EC patients and PDX tumors

FFPE tissues from EC patients and PDX were stained and analyzed at the Department of Pathology at Vall d'Hebron Hospital. Molecular classification was performed by using the PROMISE surrogate system which requires testing of MMR proteins, p53 and somatic mutation analysis of POLE (exons 9, 11, 13, and 14) as described in Chapter 2.

Statistical analysis

The statistical analysis was performed in Graph Pad Prism (v.8.2.1) (GraphPad Software, La Jolla, CA, USA). The median of normalized tumor volume from each treatment group was used to compare the efficacy against placebo in every time point. In those cases where placebo did not reach 12 weeks of treatment, the last measurement was used for the different comparisons. Turkey's multiple comparisons test was used to compared body weight and tumor growth between all groups in every time point. P-values lower than 0.05 were considered statistically significant.

In order to interpret the efficacy of each treatment, we defined i) Complete Response (CR) when the PDX tumor growth was delayed or repressed between 70-100% in comparison to the placebo group for a particular treatment; ii) Partial Response (PR) when tumor growth rate was inhibited between 30-70% compared to control group; and iii) No Response (NR) when a decrease in tumor growth rate was less than 30% compared to placebo group. Moreover, we introduced the concept of long-term when a response was achieved at least for 75% of experimental time; while intermediate-term and short-term response indicated between 25 to 75% or less than 25% of experimental time, respectively.

RESULTS

Subject characteristics

The primary aim of this study was to investigate the efficacy of a novel compound, SYD985, which is a HER2-targeting antibody drug conjugate, as monotherapy or in combination with the PARP inhibitor niraparib, in a set of EC PDX models with aberrant expression of ERBB2 (HER2). In order to select the positive HER2-EC PDX models, we stain a tissue microarray (TMA), including 43 EC PDX models, with the HER2 antibody. A total of 15 PDX models with HER2 staining levels from 1+ to 3+ were selected for the preclinical study. A representative image on the HER2 staining is depictured in Figure 39.
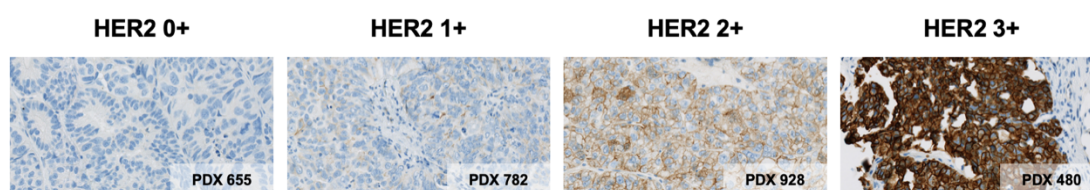
Figura 39. Immunohistochemistry staining of HER2. *Four PDX models (655, 782, 928, and 480) with HER2 expression 0+, 1+, 2+, and 3+, respectively. Magnification 40x.*

Clinicopathological and molecular features of both, EC patients and the corresponding 15 PDX models are summarized in Table 29. Particularly, our EC cohort included EEC (n=6) and NEEC (n=9) patients, represented mainly by serous histology (n=6), undifferentiated tumors (n=2) and a patient with mix histology (n=1). Regarding molecular classification, one patient was classified as POLE, six patients were classified as MSI, one patient was classified as low-CN, and six patients were classified as high-CN. As expected, most PDX were histological and molecular classified as their patient counterpart, recapitulating most of the histologic and molecular features of the original human tissue. Of particular interest is the comparison of HER2 between patient and PDX. Our results indicates that all patients, except three, had HER2 expression.

Table 29. Clinicopathological and molecular features of EC patients (PT) and PDX models

Sample	REC	Histology	Grade	Risk	POLE	MSH6	MSH2	MLH1	PMS2	P53	Mol. Class.	HER2
PT003	-	SEC	3	High	WT	WT	WT	WT	WT	Mut	HCN	0
PDX003	-	SEC	3	-	WT	WT	WT	WT	WT	Mut	HCN	1+
PT440	No	EEC	2	High-Intermediate	WT	WT	WT	Mut	Mut	WT	MSI	1+
PDX440	-	EEC	2	-	WT	WT	WT	Mut	Mut	WT	MSI	2+
PT480	No	SEC	3	High	WT	WT	WT	WT	WT	-	-	2+
PDX480	-	SEC	3	-	-	WT	WT	WT	-	-	-	3+
PT516	Yes	EEC	3	High-Intermediate	WT	WT	WT	Mut	Mut	WT	MSI	1+
PDX516	-	EEC	3	-	WT	WT	WT	Mut	Mut	WT	MSI	1+
PT521	No	EEC	2	High	WT	WT	WT	WT	WT	WT	LCN	1+
PDX521	-	EEC	2	-	WT	WT	WT	WT	Mut	WT	MSI	1+
PT524	No	EEC	3	High-Intermediate	WT	Mut	Mut	WT	WT	WT	MSI	1+
PDX524	-	EEC	3	-	WT	Mut	Mut	WT	WT	WT	MSI	1+
PT526	No	EEC	3	High-Intermediate	WT	WT	WT	WT	Mut	WT	MSI	0
PDX526	-	EEC	3	-	WT	WT	WT	WT	Mut	Mut	MSI	1+
PT558	Yes	UND	3	High-Intermediate	WT	WT	WT	WT	WT	Mut	HCN	0
PDX558	-	UND	3	-	-	WT	WT	WT	WT	Mut	-	1+
PT573	No	UND	3	High	WT	WT	WT	Mut	Mut	WT	MSI	1+
PDX573	-	UND	3	-	-	WT	WT	Mut	Mut	WT	-	1+
PT589	Yes	SEC	3	High-Intermediate	WT	WT	WT	WT	WT	Mut	HCN	1+
PDX589	-	SEC	3	-	WT	WT	WT	WT	WT	Mut	HCN	1+
PT596	Yes	SEC	3	High-Intermediate	WT	WT	WT	WT	WT	Mut	HCN	1+
PDX596	-	SEC	3	-	WT	WT	WT	WT	WT	Mut	HCN	1+
PT782	No	SEC	3	High	WT	WT	WT	WT	WT	Mut	HCN	1+
PDX782	-	SEC	3	-	WT	WT	WT	WT	Mut	Mut	MSI	1+
PT928	Yes	SEC	3	High	WT	WT	WT	WT	WT	Mut	HCN	1+
PDX928	-	SEC	3	-	-	WT	WT	WT	WT	Mut	-	2+
PT959	No	MIX	3	High	WT	Mut	Mut	WT	WT	WT	MSI	1+
PDX959	-	MIX	3	-	-	Mut	Mut	WT	WT	WT	-	1+
PT1083	No	EEC	3	Low	Mut	WT	WT	WT	WT	WT	POLE	1+
PDX1083	-	EEC	3	-	-	WT	WT	WT	WT	WT	-	1+

EEC: Endometrioid endometrial cancer; SEC: Serous endometrial cancer; UND: Undifferentiated; MIX: Mixed; MSI: Microsatellite instability; LCN: Low-copy number; HCN: High-copy number

Toxicity of SYD985 and Niraparib in EC PDX models

The preclinical mice study to assess toxicity of SYD985 and niraparib was evaluated in the 15 previously selected PDX models. In total, we implanted 60 tumors for each PDX model (2 tumors in 30 mice), and we obtained an engraftment of 82% (Table 30). In order to assess the right administration of SYD985, SYD989 and niraparib, blood samples from each animal were collected at day four after treatment onset. As expected, 92% of mice were correctly treated and exhibited high levels of the drug. Mice not presenting the drug on blood were discarded from the study analysis. Table 30 summarized the information regarding number of tumors included in each of the five-branch condition. Specifically, placebo and niraparib were administered daily for 21 days by oral gavage, while SYD989 and SYD985 were administered intravenously in a single dose the first day of treatment. Treatment, frequency, and doses of drug administration are indicated in Figure 40A.

Table 30. *Number of tumors included in each of the five-branch conditions*

Model (% engraftment)	Vehicle	SYD989	SYD985	Niraparib	Niraparib + SYD985
PDX003 (98%)	8	10	6	11	12
PDX440 (95%)	8	9	9	13	11
PDX480 (93%)	6	8	7	13	12
PDX516 (93%)	7	7	8	9	13
PDX521 (93%)	6	4	4	12	12
PDX524 (100%)	8	9	11	14	14
PDX526 (63%)	6	6	6	6	5
PDX558 (52%)	3	4	2	3	4
PDX573 (93%)	8	8	9	12	10
PDX589 (95%)	9	12	11	12	11
PDX596 (93%)	6	6	7	8	13
PDX782 (23%)	2	2	3	3	4
PDX928 (62%)	6	6	7	6	6
PDX959 (87%)	9	8	10	10	12
PDX1083 (97%)	10	9	9	13	13

Toxicity of the drugs was evaluated by monitoring the mouse body weight twice weekly. A loss of body weight between 5-10% compared to initial body weight (start of treatment) suggested toxicity, whereas body weight loss greater than 20% triggered endpoint criteria, resulting mouse withdrawn from study. We only observed toxicity associated to niraparib treatment (Figure 40B-C). Figure 40D-E show the percentage of mice withdrawn from the study and euthanized due to toxicity associated to niraparib

treatment. Niraparib was initially administered at 50 mg/mL daily doses to 4 PDX models. Toxicity was seen in most PDX models, with an average toxicity of 38% and 44% of animals affected in each PDX model for the monotherapy and the combinatory treatment, respectively. Toxicity was greatly reduced when niraparib dosage was adjusted to 40 mg/mL, with an average of 17% and 11% of toxicity in the monotherapy and combinatory treatment, respectively.
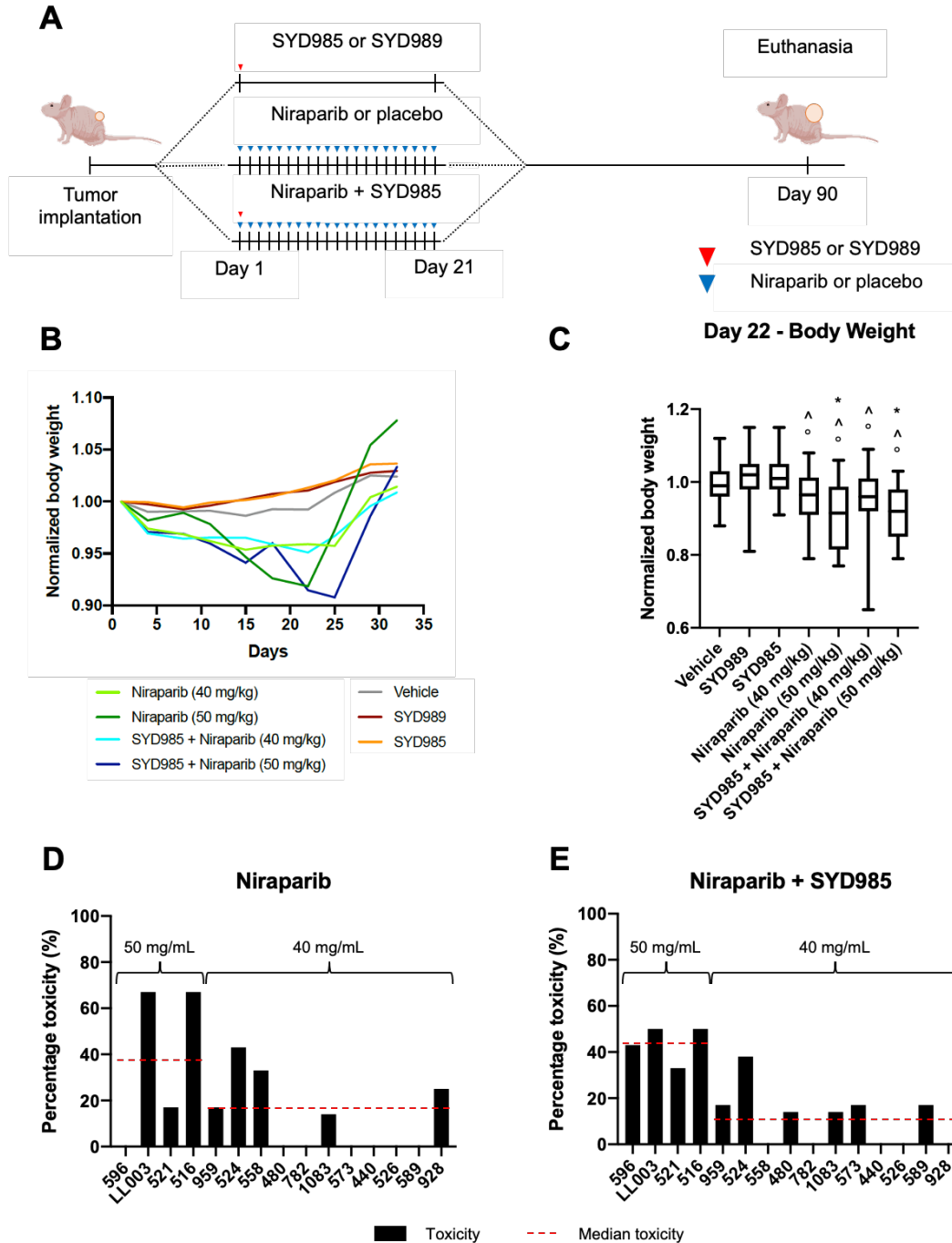
Figura 40. Toxicity assessment. A. *Scheme of drug administration.* **B.** *Percentage of animals euthanized due to toxicity associated to niraparib treatment.* **C.** *Percentage of animals euthanized due to toxicity associated to niraparib and SYD985 treatment.* **D.** *Normalized body weight.* **E.** *Normalized body weight at day 22 (statistically significant versus vehicle (*), statistically significant versus SYD989 (^), and statistically significant versus SYD985 (°).*

Additionally, we compiled body mass weight from all mice and plotted for each condition the ratio variance compared to mice body weight at day 1 (Figure 40C). We observed that mice treated with niraparib as monotherapy or in combination with SYD985 exhibited a continuous decrease of body mass during treatment (day 21), which was recovered once the treatment was finalized. Remarkably, body mass loss was enhanced in mice treated with niraparib 50 mg/mL than in 40 mg/mL.

Altogether, our results showed that niraparib induced toxicity in animals, but their effects could be reduced at concentration of 40 mg/mL and completely counteracted by treatment withdrawal.

Efficacy assessment of SYD985 and Niraparib in EC PDX models

We determined treatment response for each PDX model by evaluating tumor growth rate and mice survival along time. Three thresholds of tumor growth inhibition (<30%; 30-70%; 70-100%) were used to define the efficacy of a specific treatment in comparison to the placebo group as no response (NR), partial response (PR), and complete response (CR), respectively. Additionally, we classified the treatment response in short-, intermediate-, and long-term responses when a response was achieved for less than 25%, between 25 to 75%, or more than 75% of the experimental time, respectively.

As shown in Figure 41, SYD989 and niraparib as monotherapies, showed the lowest response rate. In both treatments, 53% (8/15) of PDX models exhibiting NR, 40% (6/15) of PDX models PR, and only one PDX model (6%) exhibited CR, which was a short-term response for SYD989 (PDX558), and an intermediate-term response for niraparib (PDX003) (Figure 41A-B).

Remarkably, PDX treated with SYD985 as monotherapy or in combination with niraparib exhibited an enhanced response. Regarding SYD985 as monotherapy, we saw that only two out of the 15 PDX models did not respond (13% NR), 53% PDX models (8/15) exhibited a transient PR, being PDX596 a long-term PR, whilst the other PDX models were short to intermediate-term PR. Importantly, SYD985 treatment increased up to 33% the number of PDX models (5/15) with CR. Among those, 4 PDX models exhibited an intermediate-term CR (PDX440, PDX526, PDX480, and PDX928), and PDX573 had long-term CR (Figure 41C). Importantly, most of PDX models (11/15, 73%) exhibited better outcomes when they were treated with SYD985 in comparison to SYD989.

Combination of SYD985 and niraparib resulted the most efficient treatment. Remarkably, 60% of PDX models presented CR (9/15) (Figure 41D). Among those, 4 models had long-term CR (PDX526, PDX573, PDX782, and PDX440), while the remaining PDX

models exhibited intermediate-term CR (PDX516, PDX596, PDX1083, and PDX480) or short-term CR (PDX524). As well, the combinatory treatment exhibited longer periods of PR compared to SYD985 in monotherapy. Two PDX models showed NR (PDX959, and 589). Those PDX models were highly chemoresistant for all treatments.
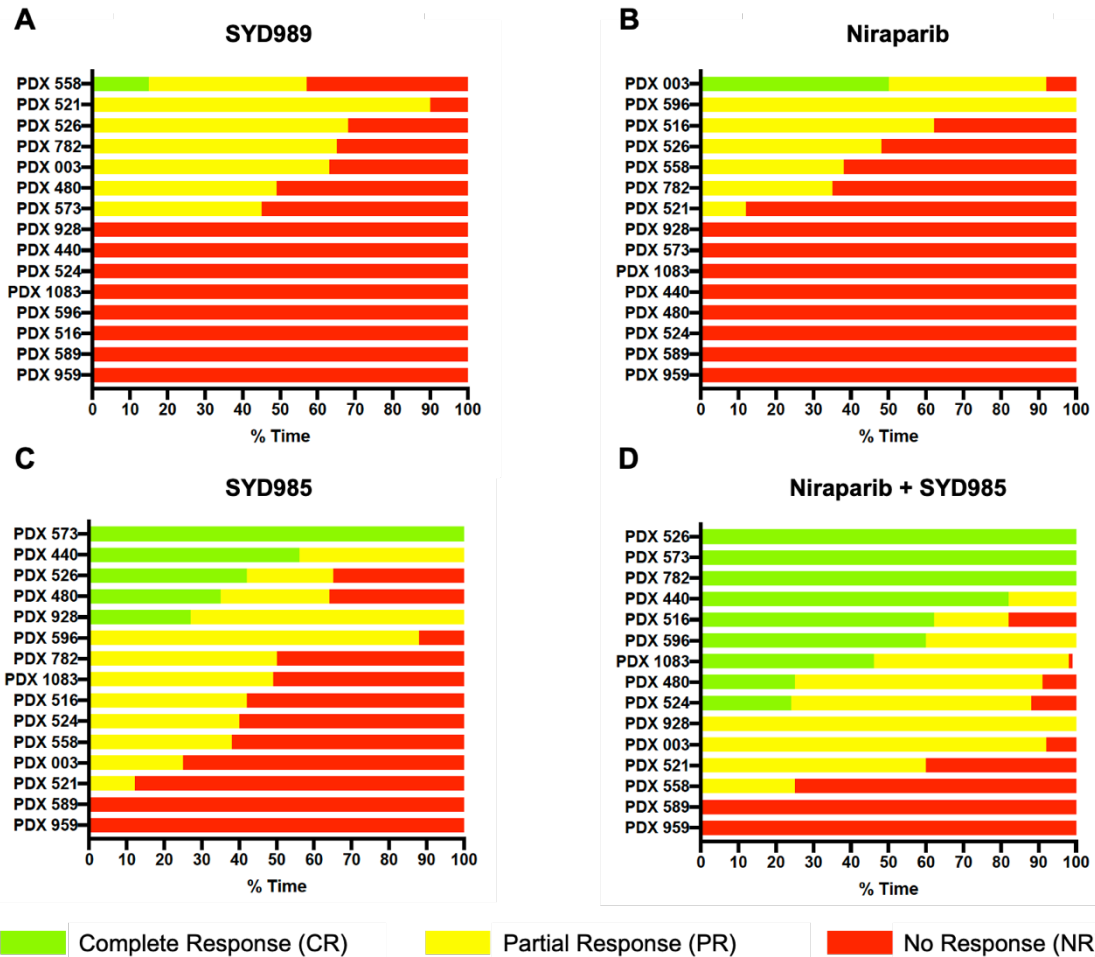
Figura 41. Efficacy assessment. A. *Treatment response for SYD989 in the PDX cohort.* **B.** *Treatment response for Niraparib in the PDX cohort.* **C.** *Treatment response for SYD985 in the PDX cohort.* **D.** *Treatment response for Niraparib + SYD985 in the PDX cohort.*

Based on the treatment efficacy results, we established four different therapy-responding PDX groups. The **non-responding** group (G1), including mice models that did not respond to any treatment; the **partial-responding** group (G2), which contains PDX that partially response to either monotherapy or combinatory treatment; the monotherapy **complete-responding** group (G3); and the combinatory **complete-responding** group (G4). G3 and G4 showed a long-term complete response to monotherapies or combinatory treatment, respectively.

First, we analyzed G1 including PDX that did not response to any treatment, PDX589 and PDX959. As expected, those models did not showed differences in tumor growth

rate among treatments compared to the placebo group. Similarly, there were any differences in overall survival (Figure 42).
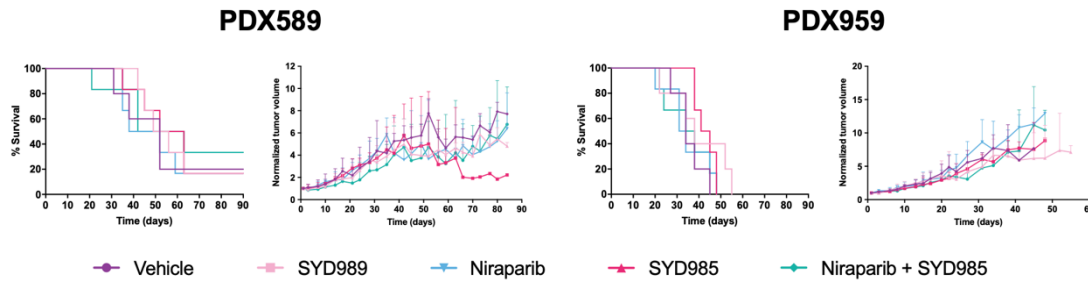
Figura 42. Non-responding group (G1). *Survival curves and normalized tumor volume growth of PDX589 and PDX959*

Next, we analyzed the partial-responding group (G2) represented by PDX521 and PDX558 (Figure 43). Particularly, we observed that PDX521 exhibited an intermediate-term PR to combinatory treatment, resulting in a delayed in tumor growth rate compared to placebo. Similarly, PDX558 also demonstrated an intermediate-term PR to niraparib, SYD985 and their combination. Interestingly, PDX558 showed a short-term CR to SYD989, however the effect observed in tumor growth rate was lost after 25 days. Regarding survival curves, we did not observe significant differences among treatments for neither PDX models (521 and 558) (Figure 43).
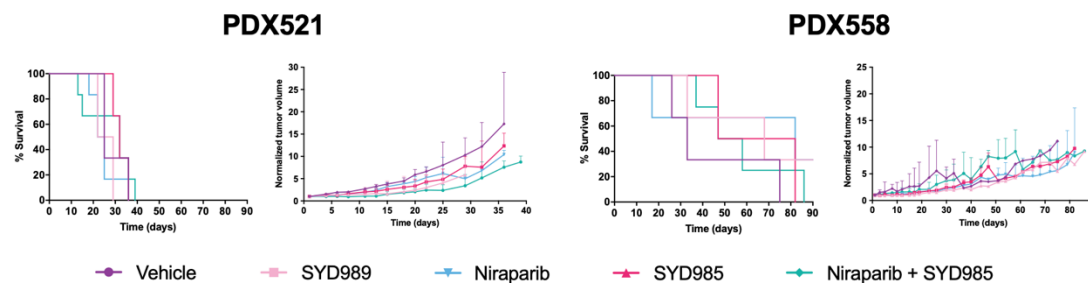
Figura 43. Partial-responding group (G2). *Survival curves and normalized tumor volume growth of PDX521 and PDX558.*

PDX models included in G3 exhibited CR to monotherapy (Figure 44). Among them, we found that three out of the four PDX models showed excellent response to SYD985 (PDX573, PDX480, and PDX928); whilst PDX003 had a CR to niraparib. Tumor growth curves of these models clearly showed that the efficacy was related to the monotherapy since the addition of the combinatory treatment did not improve treatment response (Figure 44).

PDX003

PDX480

PDX573

PDX928

Vehicle — SYD989 — Niraparib — SYD985 — Niraparib + SYD985
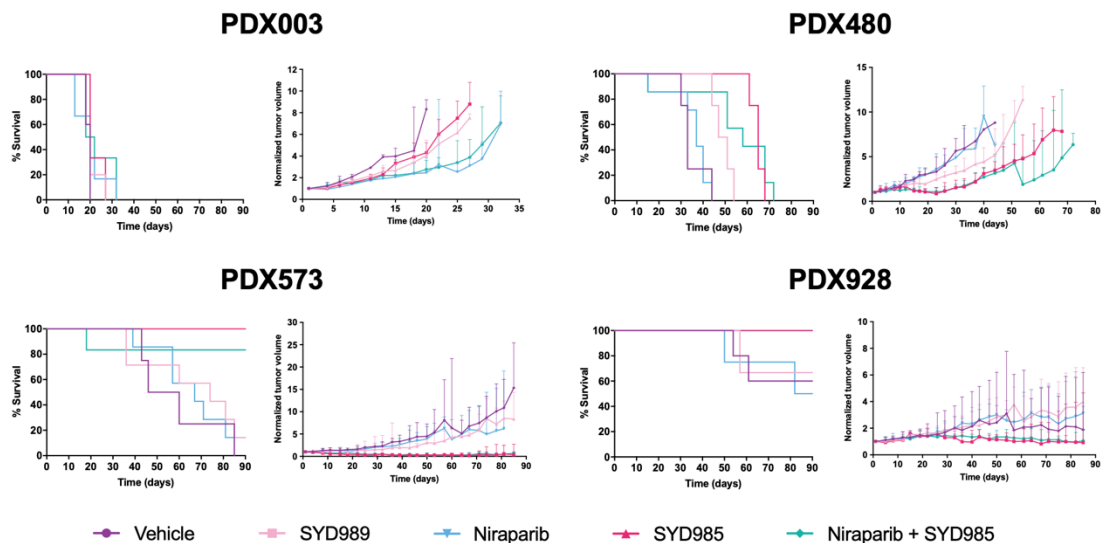
Figura 44. Monotherapy complete-responding group (G3). Survival curves and normalized tumor volume growth of PDX003, PDX480, PDX573, and PDX928.

Lastly, we analyzed G4 including PDX models with long-term CR and higher overall survival animals treated with the combination of SYD985 and niraparib (Figure 45). Despite many of the PDX included in this group showed a marked decrease in tumor growth already with SYD985 monotherapy, the inclusion of niraparib had an additive effect on tumor growth inhibition. The combinatory treatment resulted in a superior tumor growth inhibition compared to other treatments and a significant overall survival rate of PDX models. From the seven PDX models included in G4, PDX440, PDX526, and PDX782 presented the longer-term CR compared to PDX516, PDX524, PDX596, and PDX1083, which exhibited an intermediate-term CR (Figure 45).

Due to the sample size (15 PDX models), we could not divide the cohort between CR, PR, and NR and analyze the respond of each drug with clinico-pathologies and molecular features.

Altogether, our results demonstrated that targeted-treatment against HER2 and PPAR pathway are an excellent combination for EC patients, showing that HER2-expressing EC patients will benefit from this approach. Specifically, we observed that SYD985 alone was able to reduce tumor growth and, importantly, we saw that niraparib added efficacy to SYD985 even in those cases where niraparib alone did not affect tumoral growth.

176
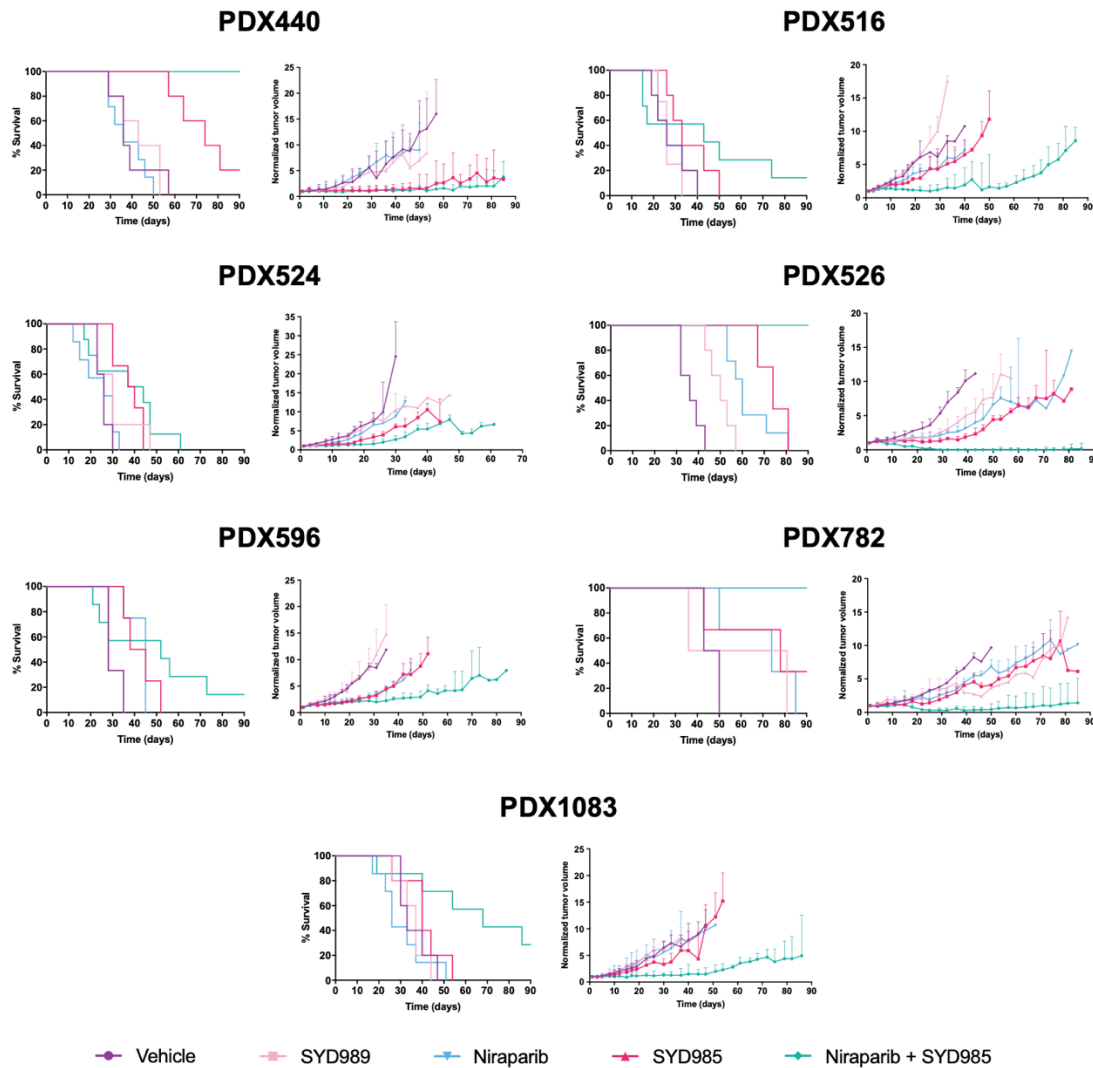
Figura 45. Combinatory complete-responding group (G4). *Survival curves and normalized tumor volume growth of PDX440, PDX516, PDX524, PDX526, PDX596, PDX782, and PDX1083*

DISCUSSION

In general, EC is detected at early stages and surgery, as a primary treatment, is very effective with 95% of patients alive after 5-years. However, when EC is diagnosed at advanced stage of the disease, it presents a worse prognosis with a 5-years survival rate dropping to 69% in cases of regional metastasis and to 18% in cases of distant metastasis [3]. In those patients, adjuvant treatment includes brachytherapy, radiotherapy, chemotherapy, hormone therapy, or a combination of the aforementioned. Despite of this multiple alternatives that have been coexisting for many years, those are not specific for EC and their success is limited to cure the disease. In order to provide with new and more effective alternatives, research on drug development has focused on targeting specific pathways in EC, such as mTOR inhibitors, PI3K inhibitors, EGFR inhibitors [98],

etc. and more recently, immunotherapy; and the development of companion diagnostics in order to identify which patients will benefit from a specific therapy.

Up to 40% of EC deaths are related to the molecular subtype of high-CN or p53 abnormal [87]. According to this profile, the TransPORTEC consortium just launched the RAINBO Umbrella Trial (NCT05255653), where p53 abnormal endometrial cancer patients are assigned to a trial evaluating the performance of chemoradiation as monotherapy or in combination with olaparib, a PARP inhibitor [101]. Patients in the high-CN group present 40% of homologous recombination deficiency, which indicates the use of PARP inhibitors [117]. Additionally, those patients also present HER2 alterations [101]. To our knowledge, this is the first study reporting the efficacy of the combinatory treatment using HER2 and PARP inhibitors on a selected population of HER2-positive PDX EC models.

On the one hand, HER2 is a well-established therapeutic target in a large subset of women with breast cancer. HER2 is also overexpressed in patients with other solid tumors and HER2-targeted therapies are being tested in these types of tumors harboring HER2 overexpression, amplifications or other mutations of the gene encoding HER2 (*erbb2*) [285]. Notably, trastuzumab has improved the overall survival of patients with HER2-positive gastric cancers [286]. However, other HER2-targeted therapies, such as lapatinib or T-DM1, have failed to provide significant improvements in the outcomes of patients with HER2-positive gastric cancer [287,288]. Results in gastric cancer suggests that the success observed in HER2-positive breast cancer might not be replicated in other solid tumors, and thus, efficacy of HER2-targeted therapies should be investigated for each tumor type. Differences in the level of HER2 overexpression and other aspects of the biology might affect the response to HER2-targeted therapies.

The rates of HER2 overexpression and/or amplifications in EC differ in the literature. Many studies report an association between HER2 and uterine serous carcinomas, with the rates of HER2 overexpression ranging from 14% to 80%, and the rates of HER2 amplification ranging from 21% to 47% [122,289]. These differences are caused by discrepancies in HER2 IHQ evaluation. All pathologists agree that 0 expression is negative and 3+ expression is positive, but there is not a clear consensus for those cases with 1+ and 2+ expression where most of EC falls. Additionally, significant heterogeneity of HER2 protein expression has been observed in HER2 positive endometrial serous carcinomas [290], which difficulties the evaluation of HER2 in EC.

In this chapter, we have selected 15 PDX EC models based on HER2 status to evaluate the efficacy of SYD985, a second-generation ADC consisting of trastuzumab bound to duocarmicin payload via a cleavable linker. From all models, 12 EC PDX presented 1+

HER2 expression, 2 EC PDX presented 2+ HER2 expression, and 1 EC PDX presented 3+ HER2 expression. We observed that 34% of the PDX models showed a complete response to SYD985 as monotherapy independently of HER2 status, suggesting that 1+ HER2 expression in EC could be considered as positive. Importantly, SYD985 has been compared with T-DM1 (a first-generation ADC anti-HER2) and both ADC showed similar binding affinity to HER2, internalization and cytotoxicity. However, T-DM1 only presented efficacy in HER2 3+ cell lines, whereas SYD985 also induced cytotoxicity in HER2 negative and HER2 1+, 2+, and 3+ cell lines [129,130]. In line with our results, SYD985 has shown high efficacy to control tumor growth independently of HER2 expression in 9 uterine serous carcinoma (USC) and 4 primary carcinosarcoma (CS) cell lines; as well as in 3 USC and 3 CS PDX models [131,132]. Specifically, from the 9 NEEC patients of our cohort, we observed 33% (3/9) of CR, 45% (4/9) of PR, and 22% (2/9) of NR without any correlation to HER2 expression. Altogether, these results suggests that target EC with SYD985 could be a good therapeutic approach independently of the HER2 status.

On the other hand, PARP inhibitors are a novel class of therapies that have shown to be effective in the treatment of homologous recombination repair deficient tumors [102]. Niraparib, an inhibitor of PARP-1 and PARP-2, has shown *in vitro* cytotoxicity by increasing formation of PARP-DNA complexes resulting in DNA damage, apoptosis and cell death. Surprisingly, niraparib-induced cytotoxicity was observed in tumor cell lines with or without deficiencies in BRCA1/2 [108,109], which is the main gene affected in homologous repair deficiency. Some studies suggests that BRCA mutated patients have a higher risk to develop serous endometrial cancer [113,114] and homologous recombination deficiency has also been associated with non-endometrioid histologies and p53 abnormal patients [117].

In this chapter, we have evaluated the efficacy of niraparib in 15 PDX EC models. Firstly, we observed that niraparib induced toxicity and it could be completely counteracted by treatment withdrawal. Hatch et al. [291] evaluate the toxicity of some PARP inhibitors, including olaparib, niraparib, and rucaparib, in ovarian and endometrial cancer. They saw that 51% of patients required a dose reduction due to toxicities. They conclude that PARP inhibitors are associated with numerous toxicities that are best managed through a multi-modal approach. Regarding efficacy, in our PDX EC cohort we only observed one complete response using niraparib in monotherapy in the PDX003. Strikingly, although niraparib alone was not able to reduce tumor volume in most PDX EC models, it had an important additive effect to SYD985 to enhance the anti-tumoral activity. Specifically, we observed 7% and 27% of CR in niraparib and SYD985 as monotherapy, respectively, while it increased up to 60% of CR in the combination; and it also increased

the treatment response from short- and intermediate-responses to intermediate- and long-responses. Importantly, considering that up to 40% of EC deaths are related to the molecular subtype of high-CN, the combination of niraparib and SYD985 induced a CR in 67% in this molecular subgroup of our PDX cohort, suggesting that these specific EC treatment could increase the overall survival of EC patients.

Animal models have emerged as a good approximation since they can mimic different steps of a disease and can be used for therapy drug testing. Although we have observed some discrepancies in the molecular classification and HER2 status between EC patients and their PDX counterparts, it has been validated that EC PDX reliably recapitulate the majority of histologic and molecular EC features [292], and that the PDX models still represent the subclonal nature of EC tumors. In this study, we have used each PDX model as if they were patients participating in a clinical trial, with the advantage that several treatments were tested in each PDX model. Despite of the inherent challenges of using PDX – such as the lack of immune system, replacement of human stroma by mouse stroma after several passages, among others – this comprehensive study of a large number of PDX models allow us to underpin the clinical development. Currently, a phase I clinical trial (NCT04235101) has been launched to evaluate safety, pharmacokinetics and efficacy of SYD985 and niraparib in patients with HER2-expressing locally advanced or metastatic solid tumors, including endometrial cancer.

In conclusion, in this chapter we demonstrated the potential combinatory effect of niraparib and SYD985 in HER2-positive EC PDX models. As well, we have demonstrated the importance of using extensive cohorts of PDX models in order to underpin the clinical development of novel therapies.

DISCUSSION

Endometrial cancer (EC) is the most common cancer of the female genital tract and the fourth most common cancer in women in the United States [3]. Thanks to the early diagnosis, EC is mostly detected in the initial stages where the tumor is still confined to the uterus and presents a 5-years survival rate of 95%. The 5-year survival drops to 69% or 18% for EC patients diagnosed at an advanced stage of the disease presenting with regional dissemination (18% EC patients) or distant dissemination (9% EC patients) [3]. Although cancer survival has improved in the last decades for most cancers, the incidence of EC is increasing about 1% per year [6]. The standard of care for EC is surgery followed by adjuvant treatment (including no adjuvant treatment, brachytherapy, radiotherapy, chemotherapy, hormonotherapy, palliative treatment, alternative treatments such as immunotherapy, or combination of the aforementioned) according to the risk stratification system [26]. This treatment is effective in those cases of localized tumor but is inefficient in cases where the tumor has spread beyond the uterus. The current adjuvant treatment has a poor effect in cases of metastases and recurrences. Therefore, the increasing incidence and the absence of specific treatments could explain the rise of death rates.

Clinic-pathological and molecular features will be obtained at the time of diagnosis and guide the surgical treatment of the patient. Also, they will be assessed on the resected tissue, used to classify patients in a risk stratification system, and accordingly, define the subsequent adjuvant therapy. However, the risk stratification system is not accurate because 5% of patients classified as low risk suffer a recurrence [69]. Likewise, around 5-15% of patients classified between intermediate to high-intermediate [74,75], and around 30-40% of patients classified from high to advanced risk will suffer a recurrence [78–80]. The identification of these patients prior the recurrence may increase the overall survival by re-adjusting the adjuvant therapy and increasing the follow-up.

In this context, this thesis has been divided in two main objectives. On the one hand, we work on the identification of a set of biomarkers predicting recurrence for EEC and for SEC. On the other hand, we assessed the toxicity and efficacy of a new treatment for EC using 15 preclinical studies on different EC patient-derived xenograft (PDX) models. These approaches are expected to improve the detection of recurrent EC patients using the primary tumor and provide new therapies for the treatment, while decreasing the mortality and morbidity associated with this disease.

RECURRENCE PREDICTION IN FFPE SPECIMENS

As said, the first part of this thesis consisted in the identification of protein biomarkers to predict recurrence in the primary tumor using FFPE specimens. The current stratification

system classifies patients according to the risk of recurrence and define the adjuvant treatment. However, there is a percentage of cases that recur independently of the risk group that was associated to the patient. Hence, we aimed to improve this system by identifying a biomarker or a panel of biomarkers in the primary tumor to improve the accuracy of the current risk stratification system to predict the recurrence. The results derived from these chapters are expected to tailor adjuvant treatment and follow-up of the patient, to ultimately decrease mortality and comorbidities. The workflow of this first objective is described from Chapter 1 to Chapter 4.

In **Chapter 1** we developed a protocol for the identification of proteins using MS from FFPE specimens. MS has emerged as a new tool for the identification and validation of potential biomarkers, but it has been mainly used in fresh-frozen samples. Because of the lack of samples to study EC recurrence, we aimed to use the large biobanks of FFPE samples stored in Pathology Departments as the source of biomarkers. The technical challenge to optimally extract enough quality and quantity of proteins from FFPE samples was achieved thanks to readapting the protocol described by Gámez-Pozo et al. [220]. We demonstrated that our protocol was useful to detect and identify proteins in FFPE blocks with different storage time, and we introduced macrodissection as a first step to increase tumor-related proteins. There are different primary sources to study recurrence (FFPE or OCT) and several protocols for protein extraction (Rapigest, PPF, ProteaseMax, PEG 20.000, SDS, urea, etc.) each one with advantages and limitations. Although our protocol worked for its purpose, further investigation could be performed to compare our protocol to other protein extraction conditions. Overall, we have shown that FFPE specimens are an unvaluable resource for clinical and biomarker researcher.

In **Chapter 2** we analyzed the proteome of 96 EC patients to identify biomarkers predicting recurrence using an untargeted MS approach. The proteomic landscape of our cohort determined a clear differentiation according to the histology in two different groups: EEC and SEC. Consequently, the search for biomarkers predicting recurrence was performed independently for each histology. The proteomic study on EEC primary tissues yield 4.569 proteins. Among those, 252 proteins were differentially expressed in the quantitative analysis and 168 proteins were differentially expressed in the absence/presence analysis when comparing recurrent vs non-recurrent patients. Importantly, we identified 21 proteins with an adjusted p-value lower than 0.05 and 4 proteins with a log FC high than 1 in the quantitative analysis, and 37 proteins with a p-value lower than 0.01 in the absence/presence analysis. These were potential biomarkers since the discovery phase is able to provide with the overall proteomic landscape but includes a large number of false positive biomarkers. Using all the altered

proteins in a GSEA analysis, we understood that the activation of RNA processing and spliceosome pathway, and the suppression of adhesion and binding pathway in the primary tumor are the main altered pathways in those patients that will recur.

The proteomic study on SEC primary tissues permitted to identify 5.757 proteins, being 27 proteins differentially expressed in the absence/presence analysis when comparing recurrent vs non-recurrent patients. The most promising biomarker was GUAD or Guanine Deaminase with a p-value lower than 0.01. The end-products of guanine-based purines (GBPs), like xanthine, have been associated to reactive oxygen species (ROS) production and DNA damage. Although GUAD has not been directly studied in cancer, some evidences suggest a possible role of GBPs in cancer, with purine salvage pathway being the fuel of nucleotide pool maintenance and correct cell division [247]. In our results, we observed a diminished expression of GUAD in the recurrence patients, suggesting that it might contribute to nucleotide pool destabilization and enhance tumor aggressivity. Surprisingly, the identification of potential biomarkers predicting recurrence in the SEC histology was quite limited, and any significant biomarker was identified in the quantitative analysis. As well, the GSEA analysis did not show any relevant pathway for this histology.

The identification of biomarkers predicting recurrence was also performed subdividing the cohort according to the molecular classification groups. We observed a correlation between MSI / low-CN and endometrioid histology, and between the high-CN and serous histology. On the one hand, 53% of significant proteins from the quantitative analysis and 43% of significant proteins from the absence/presence analysis in the endometrioid histology were also detected in the MSI and/or low-CN analysis. On the other hand, 55% of significant proteins from the absence/presence analysis in the serous histology were also detected in the high-CN analysis.

Overall, we identified potential biomarkers predicting recurrence in the endometrioid and serous histology, and in the different molecular subgroups; and we also unveiled the proteomic landscape of recurrent EC and detected the most relevant pathways involved in the recurrence of EEC patients.

In **Chapter 3** we performed an *in-silico* analysis of the TCGA and CPTAC to identify additionally EC biomarkers predicting recurrence, and also validating results obtained in the discovery phase (Chapter 2). On the one hand, we used the TCGA dataset. TCGA consortium was the first to perform an integrated genomic, transcriptomic, and proteomic characterization of EC patients using array- and sequencing-based technologies. Specifically, we used the proteomic data of the RPPA, which was done in 440 EC

patients and analyzed the expression level of 208 proteins. Although only 102 (49%) of these proteins were also detected in the discovery phase (Chapter 2), we could validate the potential use of ANXA1 and MYH9 as biomarkers predicting recurrence. On the other hand, we used the CPTAC dataset. CPTAC consortium undertook a comprehensive proteogenomic characterization of 95 EC patients analyzing the proteome by LC-MS in the Orbitrap Fusion Lumos mass spectrometer. The main limitation of this analysis was the low number of patients with recurrence, only 13 recurrent EC, but the main strength of the study was the great coverage of their proteome, with 12.524 proteins identified in EC primary tissues. The output of the analysis permitted to elucidate 253 novel biomarkers predicting recurrence. Among those, we observed high expression of immunoglobulins in recurrent EC patients, but further investigation is needed in the immunology field in EC recurrence. Overall, we analyzed the online repositories of TCGA and CPTAC, and we identified 15 and 253 potential biomarkers predicting recurrence, respectively, including ANXA1 and MYH9 that were also identified in the discovery phase (Chapter 2) and will be further validated in the verification phase.

In **Chapter 4** we selected 58 potential biomarkers based on the discovery phase (Chapter 2) and the bioinformatic analysis (Chapter 3), specifically for EEC patients. We evaluated protein expression level in the primary tumor of 129 EEC patients using a targeted MS approach. After noticing a problem on protein destabilization in the Bergen's cohort, we proceed to perform two independent verification phases using either the Bergen (n=70) and the CAT (n=59) cohort of patients. In the Bergen cohort we identified 26 peptides corresponding to 21 proteins differentially expressed between recurrent vs non-recurrent patients, and we developed a 5-protein predictive model with an AUC of 0.741. On the other hand, in the CAT cohort we identified 9 peptides corresponding to 7 proteins as significantly altered in recurrence. Importantly, we detected that ANXA1, ASPG, PMVK, and KGUA were proteins differentially expressed in the Bergen's and CAT cohort with similar values of FC and in the same tendency of expression as in the discovery phase. Overall, we proved the efficiency of LC-MS operated in PRM acquisition to perform verification studies. We verified 4 biomarkers predicting recurrence: ANXA1, ASPG, PMVK, and KGUA, and we defined a panel of proteins that achieve the best performance to predict recurrence with an AUC value of 0.741 (62.5% sensitivity – 62.5% specificity).

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

1. **The use of proteins as potential biomarkers**. In biomarker research, there is a broad spectrum of strategies such genomics, transcriptomics, metabolomics and

proteomics. Among them, proteomics shows remarkable advantages for the biomarker pipeline. Firstly, proteins are the biological end products that determine normal or disease physiology. Secondly, the diversity of proteins is much higher than that of DNA or RNA because alternative splicing and post-translation modifications generate different proteins from the same gene. Therefore, this diversity increases the probability to identify a protein, or a group of proteins, associated to a specific condition like recurrence on EC patients. Thirdly, proteins can be easily measured through low-cost methods that are widely available in clinical laboratories. The main techniques are ELISA for fluids and IHQ for tissues. However, these are antibody-based methods that depend on available and specific antibodies. In 2008, Berglund et al. [293] tested 6,100 commercial antibodies obtaining only an average success rate of 49%, representing the poor correlation and reproducibility between assays. In this context, MS appears as a promising tool for protein detection and quantification. Although the application of MS in the routine practice requires further development, over the last years there have been rapid advances in MS-based proteomics. Importantly, it has been developed the first laboratory automated with LC-MS/MS instruments, and new multiplexes tests using SRM-MS have entered in clinical practice [294].

2. **The development of a protocol to use FFPE specimens on MS analysis.** Many proteomic studies are based on fresh-frozen tissues for the identification and validation of biomarkers [278,295]. Nevertheless, those type of samples need specific conditions for collection and storage, being expensive to maintain. This combined with the low ratio of recurrences in EC, it induces the necessity to find alternatives to the fresh-frozen material. At this point, FFPE blocks appear as a suitable alternative because they represent an inexpensive tissue storage system where samples are stable at room temperature. In addition, this is a procedure widely used in hospitals worldwide. Hence, the use FFPE blocks for MS-proteomics drastically increment the number of samples for study not only recurrence in EC, but also in many other diseases where there is a lack of samples. Although FFPE specimens seems a good alternative, they present some limitations to analyze them by MS. The most important is that formalin could induce protein cross-linking and degradation [226]. Hence, not all proteins can be studied. Moreover, the method for FFPE generation has slightly varied during the years and it is not standardized. These differences could impact on protein stabilization and/or degradation and hinder the use of FFPE blocks from hospitals with different protocols. Lastly, sample preparation of FFPE samples is a critical step, in particular protein extraction. In the last years, some researchers have analyzed different protocols with contradictory data [229–231], meaning that there is still a long way to go in this field.

3. **Application of molecular classification.** In 2013, the TCGA consortium performed and integrated genomic, transcriptomic and proteomic characterization of endometrial cancer to search a new more accurate EC classification [44]. This analysis resulted in a new molecular classification with four subgroups: POLE ultramutated, microsatellite instability hypermutated, low-CN (or non-specific molecular profile), and high-CN (or serous-like). This classification was based on whole genome sequencing (WGS) and whole exome sequencing (WES), which hinder the application in hospitals. Consequently, a surrogate classification was constructed, which can be implemented in the clinical routine, based on sequencing POLE gene to classify POLE patients [296,297]; IHQ of MLH1, PMS2, MSH2, and MSH6 to classify MSI patients [47,48]; and IHQ of TP53 to classify high-CN [49,50]. The most complex technique is POLE detection because it requires POLE sequencing. Generally, POLE mutations are detected by PCR and sequencing of exons 9 to 14, which correspond to the exonuclease activity. Recently, many researchers are developing a cancer gene panel including POLE gene, which enables the analysis of the whole gene and the detection of mutations outside the exonuclease [298,299]. Nevertheless, these techniques require DNA extraction, mostly from FFPE specimens, and the DNA quality does not always allow POLE sequencing. Another alternative is KASP Genotyping, which can detect specific mutations with high sensitivity [88,300]. This technology is based on two allele-specific primers harboring a unique tail sequence that corresponds with a universal FRET (fluorescence resonant energy transfer) cassette; one labelled with FAM[TM] dye and the other with HEX[TM] dye. Both primers are exactly the same, with the difference in the last base-pair containing the mutated or wild-type nucleotide. The value of this method is the polymerase that only could amplify when the last base-pair of the primer is linked to the DNA. Hence, polymerase will amplify using one of the primers and we can detect the specific fluorescence. Although one mutation can just be analyzed per assay, considering the top 5 hotspots (S297F, P296R, V411L, P436R, and S459F), this technique permits the high-throughput analysis of POLE status of more than 50 EC patients in one run.

In this thesis, we performed POLE sequencing of exons 9, 11, 13, and 14, and KASP Genotyping analyzing mutations S297F, P286R, V411L, A456P, and S459F. We detected a correlation of 97% between both techniques. The small differences between those techniques were caused by mutations detected by PCR, which were not analyzed by KASP genotypic, such as D301E; and mutations detected by KASP genotyping that were not well-defined by PCR because of the background.

Overall, we used 225 EC patients in this thesis, and we molecularly classify 217 (96%). Specifically, we detected 11 POLE (5%), 75 MSI (35%), 87 low-CN (40%), and 44 high-

CN (20%). Our cohort of patients was selected to balance recurrent and non-recurrent patients, and to include mostly intermediate to high-risk EC patients. However, molecular classification was not included in our selection criteria since this study was designed prior to the incorporation of molecular classification in clinical guidelines. Despite that, we identified VAMP8 as a potential biomarker for the low-CN group in the quantitative analysis, and up to 32 proteins in the absence/presence analysis for the MSI, low-CN and high-CN that should be further investigated in EC recurrence.

Among recurrent patients, we identified 2 POLE patients. Both patients were grade 3, stage IB, and consequently classified as low-risk of recurrence; one patient with local recurrence is alive after 17 years of follow-up, and the other patient with distant recurrence died of disease. POLE patients belong to the group with the best outcome, very low frequency of recurrent events and deaths are very rare. Despite that, other studies have also identified POLE mutant tumors who experienced either recurrence or death [301,302]. Therefore, there is still an unmet clinical need in the detection of recurrence that molecular classification and the current risk stratification system cannot detect, and we have verified some potential biomarkers that can be incorporated in this classification to better understand recurrent events in EC.

LIMITATIONS OF THE STUDY

This study analyzed the protein expression level of FFPE EC primary tumors to detect potential biomarkers predicting recurrence. Although we identified and verified some biomarkers, there are some aspects to consider:

1. **Intratumoral heterogeneity.** We selected one FFPE block from each patient, preferable the one with the highest tumoral content. Although we enriched tumor-related proteins through macrodissection, we only used one FFPE specimen per patient and this creates a bias, i.e., the expression level of a protein could be affected because of intratumor heterogeneity. Most tumors emerge and progress under a strong evolutionary pressure favoring the establishment of a neoplastic microenvironment that exhibits genetic, phenotypic, and behavioral heterogeneity in all its components [303]. This heterogeneity might result in a non-uniform distribution of genetically distinct tumor-cell subpopulations across and within disease sites (spatial heterogeneity) or temporal variations in the molecular makeup of cancer cells (temporal heterogeneity), which is obviously directly related with a broad variety of protein expression patterns [304].

The selection of only one FFPE per patient can generate a bias for some proteins enriched in the specific tumor zone analyzed. It can explain why we did not verify most of potential biomarkers selected from the discovery phase, while in previous studies from

our laboratory we have verified almost all proteins [137]. These previous studies used uterine aspirate to detect proteins in the fluid fraction or in exosomes, and it has been demonstrated that uterine aspirate can reflect the intratumoral heterogeneity in EC [305]. In future studies, we suggest to i) select at least three FFPE specimens from the same patient to represent intratumoral heterogeneity, although this might be expensive and time-consuming; ii) validate those potential biomarkers using IHQ and if possible, including several FFPE blocks from the same patient. This technique will allow to understand protein expression in different cell and tumor sites.

2. **Few potential biomarkers in the serous histology**. Neither in the discovery phase, nor in the TCGA and CPTAC study, we identified many potential biomarkers predicting recurrence in the serous histology. Similarly, we did not detect proteins when analyzing high-CN patients, which mostly contained SEC. It can be caused by the low number of patients included in the analysis. However, we performed several sub-analyses in the endometrioid histology with less patients and we identified many more potential biomarkers. Another possible explanation is the inherent biology of these type of tumors. We saw in the PCA analysis from Chapter 2 that we cannot distinguish recurrent and non-recurrent patients, and it is known that serous is a very aggressive histology [306]. Therefore, we hypothesize that there are slightly differences between recurrent and non-recurrent SEC patients. In order to find this minor differences, we propose future investigations to increase the number of SEC patients analyzed or, even better, since the appearance of molecular classification, focus on high-CN group instead of only serous histology.

FUTURE PERSPECTIVES

The main goal of this part was the identification of a biomarker or a panel of biomarkers that could predict recurrence using the primary tumor from the patient, and that it can be easily implemented into the clinics. MS has a promising future in clinical laboratories because of the high analytical specificity and sensitivity achieved. Additionally, MS would allow the measurement of individual biomarkers or a panel of biomarkers, such as the described in this thesis, in just only one assay. However, the application of MS is still far from its adoption into routine use in clinical laboratories. Therefore, until this technology becomes widely available, immunoassays continue being the gold standard in clinical diagnostics.

At the moment, the most feasible approach to apply our findings in the clinics is IHC. It is one of the most widely available techniques in the clinics because it can be fully automated and fast once the antibodies are available. IHC is used to help on diagnose,

including cancer diagnosis. However, this technique is associated with a number of limitations, including high inter-observer variability and the capacity to label only one marker per tissue section. Thus, while IHC remains a highly practical and cost-effective diagnostic and prognostic method, this single-marker method cannot define the whole complex. Multiplexed techniques have emerged to circumvent these constraints, allowing simultaneous detection of multiple markers on a single tissue section and the comprehensive study of cell composition, cellular functional and cell-cell interactions. Among these techniques, multiplex immunohistochemistry/immunofluorescence (mIHC/IF) has emerged to be particularly promising [307]. This technique allows the detection of multiple markers simultaneously and circumvent the limitation of inter-observer variability by using image analysis systems [308]. There are several mIHC/IF platforms, including Discovery Ultra, metal-based, Vectra, etc. Although mIHC/IF techniques have a great potential in translational medicine, they are not already clinically available, and they present some limitations. For instance, Discovery Ultra has a restriction in the visualization of co-localized biomarkers; or metal-based mIHC/IF is time-consuming, costly and less sensitive than IF alone due to the nature of meta-conjugations.

ASSESS TOXICITY AND EFFICACY OF SYD985 AND NIRAPARIB FOR PERSONALIZED TREATMENT

The second part of this thesis consisted in the evaluation of toxicity and efficacy of an anti-HER2 (SYD985) and a PARP inhibitor (niraparib) in HER2-positive EC PDX models. The current adjuvant treatment based on brachytherapy, radiotherapy, chemotherapy, and/or hormone therapy, are inefficient in those cases of regional and distant metastasis. Hence, we aimed to assess the toxicity and efficacy of new treatments for personalized medicine on EC patients, specifically the ones bearing HER2 alterations. The results of this second objective are described in Chapter 5.

In **Chapter 5** we selected 15 EC PDX models based on HER2 status and with each of those, we performed a preclinical study with 5 branch therapies consisting of placebo, SYD989, SYD985, niraparib, and SYD985+niraparib. We only found toxicity on those animals treated with niraparib alone or in combination with SYD985, but their effects could be reduced at concentration of 40 mg/mL and completely counteracted by treatment withdrawal. Notably, we demonstrated that targeted-treatment with the combination of SYD985 and niraparib were an excellent combination for EC patients. Concretely, SYD985 alone was able to reduce tumor growth in 33% of animals with CR and, importantly, treatment with SYD985 and niraparib had additive effect reaching up

to 60% of CR. The combinatory treatment was highly effective even in those cases where niraparib alone did not affect tumor growth.

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

1. Personalized medicine based on molecular features. In recent years, it has been identified complex and unique biological features associated with carcinogenesis. Tumor and cell-free DNA profiling, immune markers, and proteomic and RNA analyses are used to identify these characteristics for optimization of anticancer therapy in individual patients, and this is defined as companion diagnostics. Hence, clinical trials have evolved demonstrating that matched companion diagnostics with therapy is associated with superior outcomes across tumor types [309]. In EC, the four molecular subtypes (POLE, MSI, low-CN, and high-CN) has been used to stratify patients in clinical trials and develop targeted therapies specific for each molecular subtype. For instance, some investigations indicate that favorable outcomes observed in POLE mutated EC may be independent of treatment, suggesting that these patients may be cured by surgery alone and women could be spared adjuvant therapies and their associated toxicities [310]. Similarly, FDA has approved anti-PD-1/PD-L1 therapy for MSI cases because of the high mutational load that exhibit these patients.

There are many clinical trials for EC testing specific drugs for targeted genes. Among them, the first to direct patients to different clinical trials based on the molecular profile of EC patients was the RAINBO (Refining Adjuvant Treatment IN Endometrial Cancer Based On Molecular Features). This is an umbrella program consisting of four clinical trials investigating new adjuvant therapies in molecularly-classified EC patients, as follows:

- The POLE trial is a single arm where safety de-escalation of adjuvant therapy is investigated. Specifically, no adjuvant therapy for stage I-II and only radiotherapy for stages III.
- The MSI trial is a doble arm wherein radiotherapy combined with and followed by durvalumab (anti PD-L1) for one year is compared to radiotherapy alone.
- The low-CN trial is a doble arm wherein radiotherapy followed by progestogens for two years is compared to chemoradiation.
- The high-CN trial is a doble arm wherein chemoradiation followed by olaparib (PARP inhibitor) for two years is compared to chemoradiation.

Overall, we have demonstrated that targeting-treatment against HER2 in combination with PARP inhibitors are an excellent treatment for those EC patients with HER2

overexpression. Following our results, a preclinical trial evaluating the combination of SYD985 and niraparib is being conducted. As seen, the optimized therapy in personalized medicine may require the utilization of combinations of drugs and/or strategies that attack the tumor from multiple angles.

2. **PDX models for clinical trials.** The estimated clinical approval success rate for cancer compounds is around 15% [311]. This low rate highlights the weak understanding of the complexity of human cancer and the limitations of existing preclinical models. Hence, there is a need for experimental systems that replicate the diversity of tumor in a preclinical setting. At this point, PDX appears as widely recognized and physiologically relevant preclinical models. In EC, PDX has been validated to reliably recapitulate the majority of histologic and molecular features [292,312]. Hence, these preclinical models could potentially improve the evaluation of treatments and enhance the ability to predict responses in clinical trials.

We have demonstrated that EC PDX models are good preclinical models that recapitulates the genomic landscape of the primary tumor and can be used to test and evaluate several drug responses. Gao et al. [313] stablished around 1,000 PDX models to assess the response to 62 different treatments. They demonstrated both the reproducibility and the clinical translatability of this approach by identifying associations between a genotype and drug response, and established mechanisms of resistance. Similarly, Pauli et al. [202] used high throughput drug screening to identify effective drugs that were subsequently validated in PDX models. Altogether, PDX models provides personalized therapeutic options for individual patients where standard clinical options have been exhausted and could enhance the success ratio of new cancer compounds in clinical trials.

LIMITATIONS OF THE STUDY

This study assesses the toxicity and efficacy of SYD985 and niraparib in EC PDX. Although we demonstrated the combinatory effect of both drugs, there are some aspects to consider:

1. **Number of patients/PDX models included in the study**. From our knowledge, this is the largest EC PDX cohort analyzed for drug testing. While most researchers analyzed only 1-2 PDX per drug [30,314], we have used up to 15 models. However, we have not been able to correlate the response to SYD985 and/or niraparib to histological and molecular features neither to HER2 status because of the low number of patients analyzed. There are only 2 PDX without response to any treatment, while the others 13 models present partial or complete response at least to some treatments. Additionally, as discussed in

the previous section, we selected a specific area from the primary tumor to implant in a mouse, and it could generate bias in the results because of the intratumor heterogeneity. Ideally, we should choose different areas from the primary tumor and perform the drug testing using all of them to have the general overview of tumor behavior as well as to increase the number of PDX models analyzed.

2. **Chemotherapy as the current adjuvant treatment**. We have tested the efficacy of SYD989, SYD985, niraparib, the combination of SYD985 plus niraparib, and we have used placebo to control the normal growth of the tumor (without treatment). Nevertheless, we did not include the standard treatment of EC patients, which is a combination of platin and taxanes (chemotherapy) to treat our PDX models. The inclusion of the standard treatment would have let us to compare the toxicity and efficacy of this novel approach to the current standard of care. This was not possible as there was already a large number of animals required per model and the inclusion of a new branch-therapy would have increased the cost and the time of the study. Alternatively, we could compare with previous experiments where we have used these PDX models with chemotherapy, but we only used PDX521 and we obtained a partial response to chemotherapy [30], similar to the combination of SYD985 and niraparib.

FUTURE PERSPECTIVES

The main goal of this part was to assess the toxicity and efficacy of SYD985 and niraparib in EC PDX, but the final aim is the application of this new therapy in EC patients. To do that, there are several steps to consider. On the one hand, we have used immunocompromised mice that lack an efficient immune system. In the last years, the interaction between the immune system, cancer cells, and tumoral microenvironment has been studied, but it is far from being fully understood. New insights into the human immune system as well as the mechanisms by which tumors evade immune control have led to the new and innovative therapeutic strategies [315]. Hence, our EC PDX models should be grown in humanized animal models to face these problems. They have been successfully established in many cancer types [213,214], and some studies have just developed humanized EC models to analyze the effect of anti-PD1 treatments [215,216].

On the other hand, once we have demonstrated the efficacy of this new treatment in EC PDX models, the next step would be the development of a clinical trial with patients. Currently, there are 9 clinical trials ongoing in EC testing niraparib. Most of them are recruiting patients and there is only one clinical trial (NCT04178460) that has finished, but they have not published their results. Regarding SYD985, the results derived from our study were the basis to launch two preclinical trials. Firstly, NCT04235101 has been

launched to evaluate safety, pharmacokinetics and efficacy of SYD985 and niraparib in patients with HER2-expressing locally advanced or metastatic solid tumors, including endometrial cancer. Secondly, NCT04205630 is evaluating the safety and efficacy of SYD985 recurrent, advanced or metastatic EC patients. Both clinical trials are active but still not recruiting patients.

To implement this treatment into the clinics, the primary tumor should be analyzed to evaluate HER2 status. However, SYD985 also induce cytotoxicity in HER2-negative cell lines [129,130]. We have only selected HER2-positive EC PDX and thus, SYD985 as monotherapy as well as in combination with niraparib should be tested in HER2-negative EC PDX. Therefore, instead of treating 35% of SEC patients harboring HER2 mutations [121], we could use this treatment for all EC patients independently of HER2 status.

CONCLUSIONS

The main conclusions derived from this thesis are:

1) Proteomics in formalin-fixed paraffin-embedded (FFPE) tissue is an unvaluable resource for clinical and biomarker researcher. We develop a protocol for processing FFPE specimens and preparing proteins for MS approach. It can be applied independently for untargeted proteomics (DDA) or targeted proteomics (PRM).

2) The proteomic study of 96 primary endometrial cancer tissues unveiled the proteomic landscape of recurrent endometrial cancer with the identification of more than 5,000 proteins and showed that histology continues being an important prognostic feature.

3) Two histological subtypes of endometrial cancer were investigated. The study on endometrioid endometrial cancer yield 439 significant protein biomarkers predicting recurrence, whilst only 56 proteins were identified as biomarkers on serous endometrial cancer. Molecular classification revealed that most of these potential biomarkers were also valid for EC patients with microsatellite instability, low-copy number, and high-copy number group.

4) The Gene Ontology analysis from the endometrioid EC patients showed that the activation of the RNA processing and spliceosome and the suppression of the adhesion and binding pathways were the most important altered pathways in recurrent endometrioid EC patients.

5) *In silico* studies are a potential tool to discover new biomarkers and/or validate promising biomarkers. We validate ANXA1 and MYH9 as a potential biomarkers predicting recurrence thanks to the proteomic dataset published by the TCGA, while CPTAC detected 253 novel potential biomarkers.

6) The use of LC-PRM in the verification phase enabled the accurate detection and quantification of the 52 most relevant biomarkers predicting recurrence identified in the discovery phase and *in silico* studies. Those proteins were analyzed in two independent cohorts of patients of 70 and 59 endometrioid EC patients coming from Bergen and Catalonian (CAT) hospitals, respectively.

7) In the Bergen cohort, we verified the potential of 26 peptides corresponding to 21 proteins as biomarkers predicting recurrence; while in the CAT cohort, we verified 9 peptides corresponding to 7 proteins. Four proteins: **ANXA1**, **ASPG**, **PMVK**, and **KGUA** were verified in both studies, and a 5-protein panel was developed to reach an AUC of 0.741.

8) In the assessment of novel treatments for EC, we demonstrated that niraparib induce toxicity in animals, but their effects could be reduced at concentration of 40 mg/mL and completely counteracted by treatment withdrawal.

9) Targeted-treatment against HER2 and PPAR pathway with SYD985 and niraparib resulted in a highly efficient treatment for HER2-positive EC patients. Our preclinical studies performed with 15 different HER2-positive EC PDX models showed that SYD985 alone is able to reduce tumor growth (CR of 33%) but, importantly, the combinatory treatment had an additive effect and increase CR up to 60%.

BIBLIOGRAPHY

BIBLIOGRAPHY

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

2. Sung, H. et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin 71, 209–249 (2021).

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

4. Ferlay, J. et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. European Journal of Cancer vol. 103 356–387 Preprint at https://doi.org/10.1016/j.ejca.2018.07.005 (2018).

5. Arem, H. & Irwin, M. L. Obesity and endometrial cancer survival: A systematic review. International Journal of Obesity vol. 37 634–639 Preprint at https://doi.org/10.1038/ijo.2012.94 (2013).

6. American Cancer Society - 2022 - Cancer facts and figures.

7. Khafaga, A. & Goldstein, S. R. Abnormal Uterine Bleeding. Obstetrics and Gynecology Clinics of North America vol. 46 595–605 Preprint at https://doi.org/10.1016/j.ogc.2019.07.001 (2019).

8. Soleymani, E. et al. Histopathological findings of endometrial specimens in abnormal uterine bleeding. Arch Gynecol Obstet 289, 845–849 (2014).

9. Munro, M. G. Classification of menstrual bleeding disorders. Rev Endocr Metab Disord 13, 225–234 (2012).

10. Denschlag, D., Ulrich, U. & Emons, G. Diagnostik und therapie des endometriumkarzinoms: Fortschritt und kontroversen. Deutsches Arzteblatt vol. 108 571–577 Preprint at https://doi.org/10.3238/arztebl.2011.0571 (2011).

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

12. Breijer, M. C. & Mol, B. W. J. Transvaginal ultrasound measurement of the endometrium remains the first line test for investigating postmenopausal bleeding but integration of patient characteristics into testing may further improve diagnostic algorithms. BJOG: An International Journal of Obstetrics and Gynaecology vol. 123 447 Preprint at https://doi.org/10.1111/1471-0528.13438 (2016).

13. van Hanegem, N. et al. The accuracy of endometrial sampling in women with postmenopausal bleeding: A systematic review and meta-analysis. European Journal of Obstetrics and Gynecology and Reproductive Biology vol. 197 147–155 Preprint at https://doi.org/10.1016/j.ejogrb.2015.12.008 (2016).

14. Reijnen, C. et al. Diagnostic accuracy of endometrial biopsy in relation to the amount of tissue. J Clin Pathol 70, 941–946 (2017).

15. Clark, T. J. et al. Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial cancer: a systematic quantitative review. www.bjog-elsevier.com.

16. Visser, N. C. M. et al. Accuracy of endometrial sampling in endometrial carcinoma: A systematic review and meta-analysis. Obstetrics and Gynecology vol. 130 803–813 Preprint at https://doi.org/10.1097/AOG.0000000000002261 (2017).

17. Francis, J. A., Weir, M. M., Ettler, H. C., Qiu, F. & Kwon, J. S. Should preoperative pathology be used to select patients for surgical staging in endometrial cancer? International Journal of Gynecological Cancer 19, 380–384 (2009).

18. 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).

19. Batista, T. P., Cavalcanti, C. L. C., Tejo, A. A. G. & Bezerra, A. L. R. Accuracy of preoperative endometrial sampling diagnosis for predicting the final pathology grading in uterine endometrioid carcinoma. European Journal of Surgical Oncology 42, 1367–1371 (2016).

20. Visser, N. C. M. et al. Accuracy of endometrial sampling in endometrial carcinoma: A systematic review and meta-analysis. Obstetrics and Gynecology vol. 130 803–813 Preprint at https://doi.org/10.1097/AOG.0000000000002261 (2017).

21. Silva, C., Pires-Luís, A. S., Rocha, E., Bartosch, C. & Lopes, J. M. Phenotypic Intratumoral Heterogeneity of Endometrial Carcinomas. International Journal of Gynecological Pathology 37, 154–166 (2018).

22. Gatius, S. et al. Tumor Heterogeneity in Endometrial Carcinoma: Practical Consequences. Pathobiology 85, 35–40 (2018).

23. Leskela, S. et al. Molecular basis of tumor heterogeneity in endometrial carcinosarcoma. Cancers vol. 11 Preprint at https://doi.org/10.3390/cancers11070964 (2019).

24. Gilks, C. B., Oliva, E. & Soslow, R. A. Poor Interobserver Reproducibility in the Diagnosis of High-grade Endometrial Carcinoma. http://www.gpec.ubc.ca/in- (2013).

25. Nofech-Mozes, S. et al. Interobserver Agreement for Endometrial Cancer Characteristics Evaluated on Biopsy Material. Obstet Gynecol Int 2012, 1–6 (2012).

26. Concin, N. et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. International Journal of Gynecological Cancer vol. 31 12–39 Preprint at https://doi.org/10.1136/ijgc-2020-002230 (2021).

27. Sorbe, B., Juresta, C. & Ahlin, C. Natural history of recurrences in endometrial carcinoma. Oncol Lett 8, 1800–1806 (2014).

28. A. N. J. Huijgens - 2013 - Factors predicting recurrent endometrial cancer.

29. Gehrig, P. A. & Bae-Jump, V. L. Promising novel therapies for the treatment of endometrial cancer. Gynecologic Oncology vol. 116 187–194 Preprint at https://doi.org/10.1016/j.ygyno.2009.10.041 (2010).

30. Felip, I. et al. Therapeutic potential of the new TRIB3-mediated cell autophagy anticancer drug ABTL0812 in endometrial cancer. Gynecol Oncol 153, 425–435 (2019).

31. di Tucci, C. et al. Immunotherapy in endometrial cancer: New scenarios on the horizon. J Gynecol Oncol 30, (2019).

32. Bokhman, J. v. Two Pathogenetic Types of Endometrial Carcinoma. GYNECOLOGK ONCOLOGY vol. 15 (1983).

33. Morice, P., Leary, A., Creutzberg, C., Abu-Rustum, N. & Darai, E. Endometrial cancer. in The Lancet vol. 387 1094–1108 (Lancet Publishing Group, 2016).

34. Markowska, A., Pawałowska, M., Lubin, J. & Markowska, J. Signalling pathways in endometrial cancer. Wspolczesna Onkologia vol. 18 143–148 Preprint at https://doi.org/10.5114/wo.2014.43154 (2014).

35. Ryan, A. J., Susil, B., Jobling, T. W. & Oehler, M. K. Endometrial cancer. Cell and Tissue Research vol. 322 53–61 Preprint at https://doi.org/10.1007/s00441-005-1109-5 (2005).

36. English, D. P., Roque, D. M., Buza, N. & Santin, A. D. HER2 as Biomarker for Endometrial Cancer. in Biomarkers in Cancer 1–16 (Springer Netherlands, 2014). doi:10.1007/978-94-007-7744-6_26-1.

37. Lobo, F. & Thomas, E. Type II endometrial cancers: A case series. J Midlife Health 7, 69–72 (2016).

38. Masood, M. & Singh, N. Endometrial carcinoma: changes to classification (WHO 2020).

39. Haltia, U. M., Bützow, R., Leminen, A. & Loukovaara, M. FIGO 1988 versus 2009 staging for endometrial carcinoma: A comparative study on prediction of survival and stage distribution according to histologic subtype. J Gynecol Oncol 25, 30–35 (2014).

40. 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 Preprint at https://doi.org/10.1016/j.ijgo.2009.02.012 (2009).

41. Espinosa, I. et al. Myometrial Invasion and Lymph Node Metastasis in Endometrioid Carcinomas: Tumor-associated Macrophages, Microvessel Density, and HIF1A Have a Crucial Role. www.ajsp.com (2010).

42. Moatasim, A., Hameed, Z. & Ahmad, I. Assessment of lymphovascular invasion in early stage endometrial carcinoma -a retrospective study. Surgical and Experimental Pathology 4, (2021).

43. Rütten, H. et al. Recurrent endometrial cancer: Local and systemic treatment options. Cancers vol. 13 Preprint at https://doi.org/10.3390/cancers13246275 (2021).

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

45. Murali, R., Soslow, R. A. & Weigelt, B. Classification of endometrial carcinoma: More than two types. The Lancet Oncology vol. 15 Preprint at https://doi.org/10.1016/S1470-2045(13)70591-6 (2014).

46. León-Castillo, A. et al. Interpretation of somatic POLE mutations in endometrial carcinoma. Journal of Pathology 250, 323–335 (2020).

47. Stelloo, E. et al. Practical guidance for mismatch repair-deficiency testing in endometrial cancer. Annals of Oncology 28, 96–102 (2017).

48. Kanopiene, D., Smailyte, G., Vidugiriene, J. & Bacher, J. Impact of microsatellite instability on survival of endometrial cancer patients. Medicina (Lithuania) 50, 216–221 (2014).

49. Köbel, M. et al. Interpretation of P53 Immunohistochemistry in Endometrial Carcinomas: Toward Increased Reproducibility. International Journal of Gynecological Pathology 38, S123–S131 (2019).

50. Singh, N. et al. p53 immunohistochemistry is an accurate surrogate for TP53 mutational analysis in endometrial carcinoma biopsies. Journal of Pathology 250, 336–345 (2020).

51. León-Castillo, A. et al. Clinicopathological and molecular characterisation of 'multiple-classifier' endometrial carcinomas. Journal of Pathology 250, 312–322 (2020).

52. Vermij, L., Smit, V., Nout, R. & Bosse, T. Incorporation of molecular characteristics into endometrial cancer management. Histopathology vol. 76 52–63 Preprint at https://doi.org/10.1111/his.14015 (2020).

53. Fung-Kee-Fung, M. et al. Follow-up after primary therapy for endometrial cancer: A systematic review. Gynecologic Oncology vol. 101 520–529 Preprint at https://doi.org/10.1016/j.ygyno.2006.02.011 (2006).

54. York, N. et al. Contemporary management of endometrial cancer. The Lancet 379, 1352–1360 (2012).

55. writing committee on behalf of the ASTEC study group, T. Effi cacy of systematic pelvic lymphadenectomy in endometrial cancer (MRC ASTEC trial): a randomised study. The Lancet 373, 125–136.

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

57. Frost, J. A., Webster, K. E., Bryant, A. & Morrison, J. Lymphadenectomy for the management of endometrial cancer. Cochrane Database of Systematic Reviews vol. 2017 Preprint at https://doi.org/10.1002/14651858.CD007585.pub4 (2017).

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

59. Tschernichovsky, R., Diver, E. J., Schorge, J. O. & Goodman, A. The Role of Lymphadenectomy Versus Sentinel Lymph Node Biopsy in Early-stage Endometrial Cancer. American Journal of Clinical Oncology: Cancer Clinical Trials vol. 39 516–521 Preprint at https://doi.org/10.1097/COC.0000000000000302 (2016).

60. Renz, M. et al. Immediate intraoperative sentinel lymph node analysis by frozen section is predictive of lymph node metastasis in endometrial cancer. J Robot Surg 14, 35–40 (2020).

61. Bogani, G., Murgia, F., Ditto, A. & Raspagliesi, F. Sentinel node mapping vs. lymphadenectomy in endometrial cancer: A systematic review and meta-analysis. Gynecologic Oncology vol. 153 676–683 Preprint at https://doi.org/10.1016/j.ygyno.2019.03.254 (2019).

62. Hass, P. et al. Vaginal brachytherapy for endometrial cancer. J Cancer Res Clin Oncol 144, 1523–1530 (2018).

63. Harkenrider, M. M., Block, A. M., Siddiqui, Z. A. & Small, W. The role of vaginal cuff brachytherapy in endometrial cancer. Gynecologic Oncology vol. 136 365–372 Preprint at https://doi.org/10.1016/j.ygyno.2014.12.036 (2015).

64. Kong, A., Johnson, N., Kitchener, H. C. & Lawrie, T. A. Adjuvant radiotherapy for stage i endometrial cancer: An updated cochrane systematic review and meta-

analysis. Journal of the National Cancer Institute vol. 104 1625–1634 Preprint at https://doi.org/10.1093/jnci/djs374 (2012).

65. Creutzberg, C. L. & Nout, R. A. The role of radiotherapy in endometrial cancer: Current evidence and trends. Curr Oncol Rep 13, 472–478 (2011).

66. Bestvina, C. M. & Fleming, G. F. Chemotherapy for Endometrial Cancer in Adjuvant and Advanced Disease Settings. Oncologist 21, 1250–1259 (2016).

67. Lee, W. L. et al. Hormone therapy for patients with advanced or recurrent endometrial cancer. Journal of the Chinese Medical Association vol. 77 221–226 Preprint at https://doi.org/10.1016/j.jcma.2014.02.007 (2014).

68. Sjögren, L. L., Mørch, L. S. & Løkkegaard, E. Hormone replacement therapy and the risk of endometrial cancer: A systematic review. Maturitas vol. 91 25–35 Preprint at https://doi.org/10.1016/j.maturitas.2016.05.013 (2016).

69. Laban, M. et al. The Prediction of Recurrence in Low-Risk Endometrial Cancer: Is It Time for a Paradigm Shift in Adjuvant Therapy? Reproductive Sciences vol. 29 1068–1085 Preprint at https://doi.org/10.1007/s43032-021-00565-8 (2022).

70. Moroney, M. R. et al. Molecular markers in recurrent stage I, grade 1 endometrioid endometrial cancers. Gynecol Oncol 153, 517–520 (2019).

71. Ureyen, I. et al. Factors predicting recurrence in patients with stage IA endometrioid endometrial cancer: what is the importance of LVSI? Arch Gynecol Obstet 301, 737–744 (2020).

72. Sozzi, G. et al. Tumor Size, an Additional Risk Factor of Local Recurrence in Low-Risk Endometrial Cancer: A Large Multicentric Retrospective Study. International Journal of Gynecological Cancer 28, 684–691 (2018).

73. Kommoss, F. et al. L1CAM: amending the "low-risk" category in endometrial carcinoma. J Cancer Res Clin Oncol 143, 255–262 (2017).

74. Arend, R. C. et al. Identifying a molecular profile to predict the risk of recurrence in high-intermediate risk endometrial cancer. Cancer Med 10, 8238–8250 (2021).

75. de Andrade, D. A. P. et al. A 4-Gene Signature Associated With Recurrence in Low- and Intermediate-Risk Endometrial Cancer. Front Oncol 11, (2021).

76. Wortman, B. G. et al. Ten-year results of the PORTEC-2 trial for high-intermediate risk endometrial carcinoma: improving patient selection for adjuvant therapy. Br J Cancer 119, 1067–1074 (2018).

77. Nout, R. A. et al. Vaginal brachytherapy versus pelvic external beam radiotherapy for patients with endometrial cancer of high-intermediate risk (PORTEC-2): an open-label, non-inferiority, randomised trial. The Lancet 375, 816–823 (2010).

78.   van Gool, I. C. et al. Prognostic significance of L1CAM expression and its association with mutant p53 expression in high-risk endometrial cancer. Modern Pathology 29, 174–181 (2016).

79.   Muinelo-Romay, L. et al. High-risk endometrial carcinoma profiling identifies TGF-β1 as a key factor in the initiation of tumor invasion. Mol Cancer Ther 10, 1357–1366 (2011).

80.   Feng, W. et al. Circulating tumor DNA as a prognostic marker in high-risk endometrial cancer. J Transl Med 19, (2021).

81.   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).

82.   de Boer, S. M., Nout, R. A., Bosse, T. & Creutzberg, C. L. Adjuvant therapy for high-risk endometrial cancer: recent evidence and future directions. Expert Review of Anticancer Therapy vol. 19 51–60 Preprint at https://doi.org/10.1080/14737140.2019.1531708 (2019).

83.   Bendifallah, S., Ilenko, A. & Daraï, E. High risk endometrial cancer: Clues towards a revision of the therapeutic paradigm. Journal of Gynecology Obstetrics and Human Reproduction vol. 48 863–871 Preprint at https://doi.org/10.1016/j.jogoh.2019.06.003 (2019).

84.   Altman, A. D. et al. Canadian high risk endometrial cancer (CHREC) consortium: Analyzing the clinical behavior of high risk endometrial cancers. Gynecol Oncol 139, 268–274 (2015).

85.   McAlpine, J. N. et al. Evaluation of treatment effects in patients with endometrial cancer and POLE mutations: An individual patient data meta-analysis. Cancer 127, 2409–2422 (2021).

86.   Travaglino, A. et al. Prognostic value of the TCGA molecular classification in uterine carcinosarcoma. International Journal of Gynecology and Obstetrics Preprint at https://doi.org/10.1002/ijgo.13937 (2021).

87.   Stelloo, E. et al. Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; A TransPORTEC initiative. Modern Pathology 28, 836–844 (2015).

88.   Lé On-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).

89.   Chen, J., Gao, F. & Liu, N. L1CAM promotes epithelial to mesenchymal transition and formation of cancer initiating cells in human endometrial cancer. Exp Ther Med 15, 2792–2797 (2018).

90.   Dellinger, T. H. et al. L1CAM is an independent predictor of poor survival in endometrial cancer - An analysis of The Cancer Genome Atlas (TCGA). Gynecol Oncol 141, 336–340 (2016).

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

92.   Guo, M., Gong, H., Nie, D. & Li, Z. High L1CAM expression predicts poor prognosis of patients with endometrial cancer: A systematic review and meta-analysis. Medicine 100, e25330 (2021).

93.   Kommoss, F. K. F. et al. L1cam further stratifies endometrial carcinoma patients with no specific molecular risk profile. Br J Cancer 119, 480–486 (2018).

94.   Liu, Y. et al. Clinical significance of CTNNB1 mutation and Wnt pathway activation in endometrioid endometrial carcinoma. J Natl Cancer Inst 106, (2014).

95.   Kurnit, K. C. et al. CTNNB1 (beta-catenin) mutation identifies low grade, early stage endometrial cancer patients at increased risk of recurrence. Modern Pathology 30, 1032–1041 (2017).

96.   de Leo, A. et al. Arid1a and ctnnb1/β-catenin molecular status affects the clinicopathologic features and prognosis of endometrial carcinoma: Implications for an improved surrogate molecular classification. Cancers (Basel) 13, 1–22 (2021).

97.   Fleming, G. F. Systemic chemotherapy for uterine carcinoma: Metastatic and adjuvant. Journal of Clinical Oncology vol. 25 2983–2990 Preprint at https://doi.org/10.1200/JCO.2007.10.8431 (2007).

98.   Megino-Luque, C. et al. Small-molecule inhibitors (Smis) as an effective therapeutic strategy for endometrial cancer. Cancers vol. 12 1–26 Preprint at https://doi.org/10.3390/cancers12102751 (2020).

99.   Cosgrove, C. M., Cohn, D. E. & Goodfellow, P. J. Primum non nocere: Are we ready for POLE testing in endometrial cancer? Gynecologic Oncology vol. 147 240–242 Preprint at https://doi.org/10.1016/j.ygyno.2017.09.015 (2017).

100. Nero, C. et al. Adjuvant Treatment Recommendations in Early-Stage Endometrial Cancer: What Changes With the Introduction of The Integrated Molecular-Based Risk Assessment. Frontiers in Oncology vol. 11 Preprint at https://doi.org/10.3389/fonc.2021.612450 (2021).

101. Jamieson, A., Bosse, T. & McAlpine, J. N. The emerging role of molecular pathology in directing the systemic treatment of endometrial cancer. Therapeutic

Advances in Medical Oncology vol. 13 Preprint at https://doi.org/10.1177/17588359211035959 (2021).

102. Fong, P. C. et al. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors from BRCA Mutation Carriers . New England Journal of Medicine 361, 123–134 (2009).

103. D'Andrea, A. D. Mechanisms of PARP inhibitor sensitivity and resistance. DNA Repair vol. 71 172–176 Preprint at https://doi.org/10.1016/j.dnarep.2018.08.021 (2018).

104. Risdon, E. N., Chau, C. H., Price, D. K., Sartor, O. & Figg, W. D. PARP Inhibitors and Prostate Cancer: To Infinity and Beyond BRCA . Oncologist 26, e115–e129 (2021).

105. Zhu, H. et al. PARP inhibitors in pancreatic cancer: Molecular mechanisms and clinical applications. Molecular Cancer vol. 19 Preprint at https://doi.org/10.1186/s12943-020-01167-9 (2020).

106. Mittica, G. et al. PARP Inhibitors in Ovarian Cancer. Recent Pat Anticancer Drug Discov 13, 392–410 (2018).

107. Pilié, P. G., Gay, C. M., Byers, L. A., O'Connor, M. J. & Yap, T. A. PARP inhibitors: Extending benefit beyond BRCA-mutant cancers. Clinical Cancer Research vol. 25 3759–3771 Preprint at https://doi.org/10.1158/1078-0432.CCR-18-0968 (2019).

108. Genther Williams, S. M. et al. Treatment with the PARP inhibitor, niraparib, sensitizes colorectal cancer cell lines to irinotecan regardless of MSI/MSS status. Cancer Cell Int 15, (2015).

109. Keung, M. Y., Wu, Y., Badar, F. & Vadgama, J. v. Response of breast cancer cells to PARP inhibitors is independent of BRCA status. J Clin Med 9, (2020).

110. Zhi, W., Li, S., Wan, Y., Wu, F. & Hong, L. Short-term starvation synergistically enhances cytotoxicity of Niraparib via Akt/mTOR signaling pathway in ovarian cancer therapy. Cancer Cell Int 22, (2022).

111. Sun, K. et al. A comparative pharmacokinetic study of PARP inhibitors demonstrates favorable properties for niraparib efficacy in preclinical tumor models. www.oncotarget.com (2018).

112. Musacchio, L. et al. PARP inhibitors in endometrial cancer: Current status and perspectives. Cancer Management and Research vol. 12 6123–6135 Preprint at https://doi.org/10.2147/CMAR.S221001 (2020).

113. Shu, C. A. et al. Uterine cancer after risk-reducing salpingo-oophorectomy without hysterectomy in women with BRCA Mutations. JAMA Oncol 2, 1434–1440 (2016).

114. Lavie, O. et al. BRCA germline mutations in Jewish women with uterine serous papillary carcinoma. Gynecol Oncol 92, 521–524 (2004).

115. Toumpeki, C. et al. The role of ARID1A in endometrial cancer and the molecular pathways associated with pathogenesis and cancer progression. In Vivo vol. 33 659–667 Preprint at https://doi.org/10.21873/invivo.11524 (2019).

116. Shen, W. H. et al. Essential Role for Nuclear PTEN in Maintaining Chromosomal Integrity. Cell 128, 157–170 (2007).

117. de Jonge, M. M. et al. Frequent homologous recombination deficiency in high-grade endometrial carcinomas. Clinical Cancer Research 25, 1087–1097 (2019).

118. Erickson, B. K., Zeybek, B., Santin, A. D. & Fader, A. N. Targeting human epidermal growth factor receptor 2 (HER2) in gynecologic malignancies. Current Opinion in Obstetrics and Gynecology vol. 32 57–64 Preprint at https://doi.org/10.1097/GCO.0000000000000599 (2020).

119. Growdon, W. B. et al. HER2 over-expressing high grade endometrial cancer expresses high levels of p95HER2 variant. Gynecol Oncol 137, 160–166 (2015).

120. Vermij, L. et al. Her2 status in high-risk endometrial cancers (Portec-3): Relationship with histotype, molecular classification, and clinical outcomes. Cancers (Basel) 13, 1–14 (2021).

121. Lopez, S., Zeybek, B. & Santin, A. D. Targeting Her2/neu in uterine serous carcinoma: A paradigm shift in management. Oncotarget vol. 9 http://www.jnccn.org (2018).

122. Halle, M. K. et al. HER2 expression patterns in paired primary and metastatic endometrial cancer lesions. Br J Cancer 118, 378–387 (2018).

123. Boku, N. HER2-positive gastric cancer. Gastric Cancer vol. 17 1–12 Preprint at https://doi.org/10.1007/s10120-013-0252-z (2014).

124. Maximiano, S., Magalhães, P., Guerreiro, M. P. & Morgado, M. Trastuzumab in the Treatment of Breast Cancer. BioDrugs vol. 30 75–86 Preprint at https://doi.org/10.1007/s40259-016-0162-9 (2016).

125. Loibl, S. & Gianni, L. HER2-positive breast cancer. The Lancet vol. 389 2415–2429 Preprint at https://doi.org/10.1016/S0140-6736(16)32417-5 (2017).

126. Diver, E. J., Foster, R., Rueda, B. R. & Growdon, W. B. The Therapeutic Challenge of Targeting HER2 in Endometrial Cancer. Oncologist 20, 1058–1068 (2015).

127. Pohlmann, P. R., Mayer, I. A. & Mernaugh, R. Resistance to trastuzumab in breast cancer. Clinical Cancer Research vol. 15 7479–7491 Preprint at https://doi.org/10.1158/1078-0432.CCR-09-0636 (2009).

128. Xu, Z. et al. Novel HER2-Targeting Antibody-Drug Conjugates of Trastuzumab Beyond T-DM1 in Breast Cancer: Trastuzumab Deruxtecan(DS-8201a) and (Vic-)Trastuzumab Duocarmazine (SYD985). European Journal of Medicinal

Chemistry vol. 183 Preprint at https://doi.org/10.1016/j.ejmech.2019.111682 (2019).

129. Nadal-Serrano, M. et al. The second generation antibody-drug conjugate syd985 overcomes resistances to T-DM1. Cancers (Basel) 12, (2020).

130. van der Lee, M. M. C. et al. The preclinical profile of the duocarmycin-based HER2-targeting ADC SYD985 predicts for clinical benefit in low HER2-expressing breast cancers. Mol Cancer Ther 14, 692–703 (2015).

131. Menderes, G. et al. SYD985, a novel duocarmycin-based her2-targeting antibody–drug conjugate, shows antitumor activity in uterine and ovarian carcinosarcoma with HER2/Neu expression. Clinical Cancer Research 23, 5836–5845 (2017).

132. Black, J. et al. Syd985, a novel duocarmycin-based her2-targeting antibody-drug conjugate, shows antitumor activity in uterine serous carcinoma with her2/neu expression. Mol Cancer Ther 15, 1900–1909 (2016).

133. Atkinson, A. J. et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology and Therapeutics vol. 69 89–95 Preprint at https://doi.org/10.1067/mcp.2001.113989 (2001).

134. Strimbu, K. & Tavel, J. A. What are biomarkers? Current Opinion in HIV and AIDS vol. 5 463–466 Preprint at https://doi.org/10.1097/COH.0b013e32833ed177 (2010).

135. Tosoian, J. & Loeb, S. PSA and beyond: The past, present, and future of investigative biomarkers for prostate cancer. ScientificWorldJournal 10, 1919–1931 (2010).

136. Martinez-Garcia, E. et al. Advances in endometrial cancer protein biomarkers for use in the clinic. Expert Review of Proteomics vol. 15 81–99 Preprint at https://doi.org/10.1080/14789450.2018.1410061 (2018).

137. E Martinez - 2016 - Development of a sequential workflow based on LC PRM for the verification of endometrial cancer protein biomarkers in uterine aspirate samples.

138. la Rubia, E. C. de et al. Prognostic biomarkers in endometrial cancer: A systematic review and meta-analysis. Journal of Clinical Medicine vol. 9 1–20 Preprint at https://doi.org/10.3390/jcm9061900 (2020).

139. Wishart, D. S. et al. HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res 46, D608–D617 (2018).

140. Legrain, P. et al. The human proteome project: Current state and future direction. Molecular and Cellular Proteomics 10, (2011).

141. Jensen, O. N. Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. Current Opinion in Chemical Biology vol. 8 33–41 Preprint at https://doi.org/10.1016/j.cbpa.2003.12.009 (2004).

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

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

144. Aslam, B., Basit, M., Nisar, M. A., Khurshid, M. & Rasool, M. H. Proteomics: Technologies and their applications. Journal of Chromatographic Science vol. 55 182–196 Preprint at https://doi.org/10.1093/chromsci/bmw167 (2017).

145. Li, X., Wang, W. & Chen, J. Recent progress in mass spectrometry proteomics for biomedical research. Science China Life Sciences vol. 60 1093–1113 Preprint at https://doi.org/10.1007/s11427-017-9175-2 (2017).

146. Yin, F. F. et al. Intra-tumor heterogeneity for endometrial cancer and its clinical significance. Chin Med J (Engl) 132, 1550–1562 (2019).

147. Lou, J. J. et al. A review of room temperature storage of biospecimen tissue and nucleic acids for anatomic pathology laboratories and biorepositories. Clin Biochem 47, 267–273 (2014).

148. Granato, A. et al. DNA and RNA isolation from canine oncologic formalin-fixed, paraffin-embedded tissues for downstream '-omic' analyses: Possible or not? Journal of Veterinary Diagnostic Investigation 26, 117–124 (2014).

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

150. Teng, P. N., Bateman, N. W., Hood, B. L. & Conrads, T. P. Advances in proximal fluid proteomics for disease biomarker discovery. Journal of Proteome Research vol. 9 6091–6100 Preprint at https://doi.org/10.1021/pr100904q (2010).

151. Ioannidis, J. P. A. & Bossuyt, P. M. M. Waste, leaks, and failures in the biomarker pipeline. Clinical Chemistry vol. 63 963–972 Preprint at https://doi.org/10.1373/clinchem.2016.254649 (2017).

152. Drabovich, A. P., Martínez-Morillo, E. & Diamandis, E. P. Toward an integrated pipeline for protein biomarker development. Biochimica et Biophysica Acta - Proteins and Proteomics vol. 1854 677–686 Preprint at https://doi.org/10.1016/j.bbapap.2014.09.006 (2015).

153. Feist, P. & Hummon, A. B. Proteomic challenges: Sample preparation techniques for Microgram-Quantity protein analysis from biological samples. International Journal of Molecular Sciences vol. 16 3537–3563 Preprint at https://doi.org/10.3390/ijms16023537 (2015).

154. Matthiesen, R. & Bunkenborg, J. Introduction to mass spectrometry-based proteomics. Methods in Molecular Biology vol. 1007 1–45 Preprint at https://doi.org/10.1007/978-1-62703-392-3_1 (2013).

155. Elsässer, B. & Goettig, P. Mechanisms of proteolytic enzymes and their inhibition in Qm/mm studies. International Journal of Molecular Sciences vol. 22 1–26 Preprint at https://doi.org/10.3390/ijms22063232 (2021).

156. Batschkus, S. et al. A new albumin-depletion strategy improves proteomic research of gingival crevicular fluid from periodontitis patients. Clin Oral Investig 22, 1375–1384 (2018).

157. Song, E. & Mechref, Y. Defining glycoprotein cancer biomarkers by MS in conjunction with glycoprotein enrichment. Biomarkers in Medicine vol. 9 835–844 Preprint at https://doi.org/10.2217/bmm.15.55 (2015).

158. Meleady, P. Two-dimensional gel electrophoresis and 2D-DIGE. in Methods in Molecular Biology vol. 1664 3–14 (Humana Press Inc., 2018).

159. Mohammed, S. & Heck, A. J. R. Strong cation exchange (SCX) based analytical methods for the targeted analysis of protein post-translational modifications. Current Opinion in Biotechnology vol. 22 9–16 Preprint at https://doi.org/10.1016/j.copbio.2010.09.005 (2011).

160. Edelmann, M. J. Strong cation exchange chromatography in analysis of posttranslational modifications: Innovations and perspectives. Journal of Biomedicine and Biotechnology vol. 2011 Preprint at https://doi.org/10.1155/2011/936508 (2011).

161. Josic, D. & Kovac, S. Reversed-phase high performance liquid chromatography of proteins. Curr Protoc Protein Sci 2010, (2010).

162. Liuni, P. & Wilson, D. J. Understanding and optimizing electrospray ionization techniques for proteomic analysis. Expert Review of Proteomics vol. 8 197–209 Preprint at https://doi.org/10.1586/epr.10.111 (2011).

163. Aebersold, R. insight review articles. www.nature.com/nature (2003).

164. Ong, S. E. & Mann, M. Mass Spectrometry–Based Proteomics Turns Quantitative. Nat Chem Biol 1, 252–262 (2005).

165. Neilson, K. A. et al. Less label, more free: Approaches in label-free quantitative mass spectrometry. Proteomics vol. 11 535–553 Preprint at https://doi.org/10.1002/pmic.201000553 (2011).

166. Zhu, W., Smith, J. W. & Huang, C. M. Mass spectrometry-based label-free quantitative proteomics. Journal of Biomedicine and Biotechnology vol. 2010 Preprint at https://doi.org/10.1155/2010/840518 (2010).

167. Calderón-Celis, F., Encinar, J. R. & Sanz-Medel, A. Standardization approaches in absolute quantitative proteomics with mass spectrometry. Mass Spectrometry Reviews vol. 37 715–737 Preprint at https://doi.org/10.1002/mas.21542 (2018).

168. Pappireddi, N., Martin, L. & Wühr, M. A Review on Quantitative Multiplexed Proteomics. ChemBioChem vol. 20 1210–1224 Preprint at https://doi.org/10.1002/cbic.201800650 (2019).

169. Pino, L. K., Rose, J., O'Broin, A., Shah, S. & Schilling, B. Emerging mass spectrometry-based proteomics methodologies for novel biomedical applications. Biochemical Society Transactions vol. 48 1953–1966 Preprint at https://doi.org/10.1042/BST20191091 (2020).

170. Chen, X., Wei, S., Ji, Y., Guo, X. & Yang, F. Quantitative proteomics using SILAC: Principles, applications, and developments. Proteomics vol. 15 3175–3192 Preprint at https://doi.org/10.1002/pmic.201500108 (2015).

171. Konzer, A., Ruhs, A., Braun, T. & Krüger, M. Global protein quantification of mouse heart tissue based on the silac mouse. Methods in Molecular Biology 1005, 39–52 (2013).

172. Neubert, T. A. & Tempst, P. Super-SILAC for tumors and tissues. Nature Methods vol. 7 361–362 Preprint at https://doi.org/10.1038/nmeth0510-361 (2010).

173. Shenoy, A. & Geiger, T. Super-SILAC: Current trends and future perspectives. Expert Rev Proteomics 12, 13–19 (2014).

174. Marx, H. et al. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. Nat Biotechnol 31, 557–564 (2013).

175. Jagannadham, M. V. et al. Mass Spectral Analysis of Synthetic Peptides: Implications in Proteomics. J Biomol Tech jbt.2020-3201-001 (2021) doi:10.7171/jbt.2020-3201-001.

176. Krasny, L. & Huang, P. H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. Molecular Omics vol. 17 29–42 Preprint at https://doi.org/10.1039/d0mo00072h (2021).

177. Defossez, E., Bourquin, J., von Reuss, S., Rasmann, S. & Glauser, G. Eight key rules for successful data-dependent acquisition in mass spectrometry-based metabolomics. Mass Spectrometry Reviews Preprint at https://doi.org/10.1002/mas.21715 (2021).

178. Meyer, J. G. Fast proteome identification and quantification from data-dependent acquisition–tandem mass spectrometry (DDA MS/MS) using free software tools. Methods Protoc 2, 1–17 (2019).

179. Tabb, D. L. et al. Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J Proteome Res 9, 761–776 (2010).

180. Zhang, F., Ge, W., Ruan, G., Cai, X. & Guo, T. Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. Proteomics vol. 20 Preprint at https://doi.org/10.1002/pmic.201900276 (2020).

181. Wenzel, C., Drozdzik, M. & Oswald, S. Mass spectrometry-based targeted proteomics method for the quantification of clinically relevant drug metabolizing enzymes in human specimens. J Chromatogr B Analyt Technol Biomed Life Sci 1180, (2021).

182. Gallien, S., Kim, S. Y. & Domon, B. Large-scale targeted proteomics using internal standard triggered-parallel reaction monitoring (IS-PRM). Molecular and Cellular Proteomics 14, 1630–1644 (2015).

183. Gallien, S., Duriez, E. & Domon, B. Selected reaction monitoring applied to proteomics. Journal of Mass Spectrometry vol. 46 298–312 Preprint at https://doi.org/10.1002/jms.1895 (2011).

184. Calvo, E., Camafeita, E., Fernández-Gutiéérrez, B. & López, J. A. Applying selected reaction monitoring to targeted proteomics. Expert Review of Proteomics vol. 8 165–173 Preprint at https://doi.org/10.1586/epr.11.11 (2011).

185. Holman, S. W., Sims, P. F. G. & Eyers, C. E. The use of selected reaction monitoring in quantitative proteomics. Bioanalysis vol. 4 1763–1786 Preprint at https://doi.org/10.4155/bio.12.126 (2012).

186. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: Workflows, potential, pitfalls and future directions. Nature Methods vol. 9 555–566 Preprint at https://doi.org/10.1038/nmeth.2015 (2012).

187. Bourmaud, A., Gallien, S. & Domon, B. Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications. Proteomics vol. 16 2146–2159 Preprint at https://doi.org/10.1002/pmic.201500543 (2016).

188. 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. Molecular and Cellular Proteomics 11, 1475–1488 (2012).

189. Hofman, F. M. & Taylor, C. R. Immunohistochemistry. Curr Protoc Immunol (2013) doi:10.1002/0471142735.im2104s103.

190. Solier, C. & Langen, H. Antibody-based proteomics and biomarker research-current status and limitations. Proteomics vol. 14 774–783 Preprint at https://doi.org/10.1002/pmic.201300334 (2014).

191. Wingren, C. Antibody-based proteomics. in Advances in Experimental Medicine and Biology vol. 926 163–179 (Springer New York LLC, 2016).

192. Cheon, D. J. & Orsulic, S. Mouse models of cancer. Annual Review of Pathology: Mechanisms of Disease 6, 95–119 (2011).

193. Talmadge, J. E., Singh, R. K., Fidler, I. J. & Raz, A. Murine models to evaluate novel and conventional therapeutic strategies for cancer. American Journal of Pathology vol. 170 793–804 Preprint at https://doi.org/10.2353/ajpath.2007.060929 (2007).

194. Shoemaker, R. Timeline | Major events in the development, implementation and use of the NCI60 cell lines. www.nature.com/reviews/cancer (2006).

195. Cook, N., Jodrell, D. I. & Tuveson, D. A. Predictive in vivo animal models and translation to clinical trials. Drug Discovery Today vol. 17 253–260 Preprint at https://doi.org/10.1016/j.drudis.2012.02.003 (2012).

196. Jung, J., Seol, H. S. & Chang, S. The generation and application of patient-derived xenograft model for cancer research. Cancer Research and Treatment vol. 50 1–10 Preprint at https://doi.org/10.4143/crt.2017.307 (2018).

197. Jung, J. Human tumor xenograft models for preclinical assessment of anticancer drug development. Toxicol Res 30, 1–5 (2014).

198. Cabrera, S. et al. Generation and characterization of orthotopic murine models for endometrial cancer. Clin Exp Metastasis 29, 217–227 (2012).

199. Haldorsen, I. S. et al. Multimodal imaging of orthotopic mouse model of endometrial carcinoma. PLoS One 10, (2015).

200. Depreeuw, J. et al. Characterization of patient-derived tumor xenograft models of endometrial cancer for preclinical evaluation of targeted therapies. Gynecol Oncol 139, 118–126 (2015).

201. Unno, K. et al. Establishment of human patient-derived endometrial cancer xenografts in NOD scid gamma mice for the study of invasion and metastasis. PLoS One 9, (2014).

202. Pauli, C. et al. Personalized in vitro and in vivo cancer models to guide precision medicine. Cancer Discov 7, 462–477 (2017).

203. Topp, M. D. et al. Molecular correlates of platinum response in human high-grade serous ovarian cancer patient-derived xenografts. Mol Oncol 8, 656–668 (2014).

204. Weroha, S. J. et al. Tumorgrafts as in vivo surrogates for women with ovarian cancer. Clinical Cancer Research 20, 1288–1297 (2014).

205. Winder, A., Unno, K., Yu, Y., Lurain, J. & Kim, J. J. The allosteric AKT inhibitor, MK2206, decreases tumor growth and invasion in patient derived xenografts of endometrial cancer. Cancer Biol Ther 18, 958–964 (2017).

206. Yu, Y. et al. In-vitro and in-vivo combined effect of ARQ 092, an AKT inhibitor, with ARQ 087, a FGFR inhibitor. Anticancer Drugs 28, 503–513 (2017).

207. Eritja, N. et al. Autophagy orchestrates adaptive responses to targeted therapy in endometrial cancer. Autophagy 13, 608–624 (2017).

208. Groeneweg, J. W. et al. Dual HER2 targeting impedes growth of HER2 gene-amplified uterine serous carcinoma xenografts. Clinical Cancer Research 20, 6517–6528 (2014).

209. Siolas, D. & Hannon, G. J. Patient-derived tumor xenografts: Transforming clinical samples into mouse models. Cancer Research vol. 73 5315–5319 Preprint at https://doi.org/10.1158/0008-5472.CAN-13-1069 (2013).

210. Hidalgo, M. et al. Patient-derived Xenograft models: An emerging platform for translational cancer research. Cancer Discov 4, 998–1013 (2014).

211. Bertotti, A. et al. A molecularly annotated platform of patient- derived xenografts ('xenopatients') identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. Cancer Discov 1, 508–523 (2011).

212. Migliardi, G. et al. Inhibition of MEK and PI3K/mTOR suppresses tumor growth but does not cause tumor regression in patient-derived xenografts of RAS-mutant colorectal carcinomas. Clinical Cancer Research 18, 2515–2525 (2012).

213. Morton, J. J., Bird, G., Refaeli, Y. & Jimeno, A. Humanized mouse xenograft models: Narrowing the tumor-microenvironment gap. Cancer Research vol. 76 6153–6158 Preprint at https://doi.org/10.1158/0008-5472.CAN-16-1260 (2016).

214. Shultz, L. D., Brehm, M. A., Victor Garcia-Martinez, J. & Greiner, D. L. Humanized mice for immune system investigation: Progress, promise and challenges. Nature Reviews Immunology vol. 12 786–798 Preprint at https://doi.org/10.1038/nri3311 (2012).

215. An, L. et al. KIF2C Is a Novel Prognostic Biomarker and Correlated with Immune Infiltration in Endometrial Cancer. Stem Cells Int 2021, (2021).

216. Kumar, S. et al. Preclinical characterization of dostarlimab, a therapeutic anti-PD-1 antibody with potent activity to enhance immune function in in vitro cellular assays and in vivo animal models. MAbs 13, (2021).

217. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. Nature Medicine vol. 19 1423–1437 Preprint at https://doi.org/10.1038/nm.3394 (2013).

218. Ben-David, U. et al. Patient-derived xenografts undergo mouse-specific tumor evolution. Nat Genet 49, 1567–1575 (2017).

219. Bruna, A. et al. A Biobank of Breast Cancer Explants with Preserved Intra-tumor Heterogeneity to Screen Anticancer Compounds. Cell 167, 260-274.e22 (2016).

220. Gámez-Pozo, A. et al. High-Throughput Phosphoproteomics from Formalin-Fixed, Paraffin-Embedded Tissues. Curr Protoc Chem Biol 4, 161–175 (2012).

221. Trejo, C. L. et al. Extraction-free whole transcriptome gene expression analysis of FFPE sections and histology-directed subareas of tissue. PLoS One 14, (2019).

222. Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nat Biotechnol 38, 586–599 (2020).

223. Oliveira, D. V. N. P. et al. DNA Methylation in Ovarian Tumors-a Comparison Between Fresh Tissue and FFPE Samples. Gynecol Oncol doi:10.1007/s43032-021-00589-0/Published.

224. Tsoulos, N. et al. Tumor molecular profiling of NSCLC patients using next generation sequencing. Oncol Rep 38, 3419–3429 (2017).

225. Valdés, A. et al. Proteomic comparison between different tissue preservation methods for identification of promising biomarkers of urothelial bladder cancer. Sci Rep 11, (2021).

226. Guo, H. et al. An efficient procedure for protein extraction from formalin-fixed, Paraffin-embedded tissues for reverse phase protein arrays. Proteome Sci 10, (2012).

227. Coscia, F. et al. A streamlined mass spectrometry–based proteomics workflow for large-scale FFPE tissue analysis. Journal of Pathology 251, 100–112 (2020).

228. Weke, K. et al. DIA-MS proteome analysis of formalin-fixed paraffin-embedded glioblastoma tissues. Anal Chim Acta 1204, (2022).

229. Dapic, I. et al. Evaluation of Fast and Sensitive Proteome Profiling of FF and FFPE Kidney Patient Tissues. Molecules 27, (2022).

230. Davalieva, K., Kiprijanovska, S., Dimovski, A., Rosoklija, G. & Dwork, A. J. Comparative evaluation of two methods for LC-MS/MS proteomic analysis of formalin fixed and paraffin embedded tissues. J Proteomics 235, (2021).

231. Rossouw, S., Bendou, H., Bell, L., Rigby, J. & Christoffels, A. African Journal of Laboratory Medicine. 2225–2002 (2225) doi:10.4102/ajlm.

232. Rossouw, S. C. et al. Evaluation of Protein Purification Techniques and Effects of Storage Duration on LC-MS/MS Analysis of Archived FFPE Human CRC Tissues. Pathology and Oncology Research 27, (2021).

233. Gámez-Pozo, A. et al. Protein phosphorylation analysis in archival clinical cancer samples by shotgun and targeted proteomics approaches. Mol Biosyst 7, 2368–2374 (2011).

234. Trapp, J. et al. Digging deeper into the pyriproxyfen-response of the amphipod Gammarus fossarum with a next-generation ultra-high-field orbitrap analyser: New perspectives for environmental toxicoproteomics. Front Environ Sci 6, (2018).

235. Coope, R. J. N. et al. Whole-slide laser microdissection for tumour enrichment. Journal of Pathology 253, 225–233 (2021).

236. Hinneburg, H. et al. Unlocking cancer glycomes from histopathological formalin-fixed and paraffin-embedded (FFPE) tissue microdissections. Molecular and Cellular Proteomics 16, 524–536 (2017).

237. Longuespée, R. et al. A laser microdissection-based workflow for FFPE tissue microproteomics: Important considerations for small sample processing. Methods 104, 154–162 (2016).

238. Imboden, S. et al. Implementation of the 2021 molecular ESGO/ESTRO/ESP risk groups in endometrial cancer. Gynecol Oncol 162, 394–400 (2021).

239. Kasius, J. C. et al. Risk stratification of endometrial cancer patients: FIGO stage, biomarkers and molecular classification. Cancers vol. 13 Preprint at https://doi.org/10.3390/cancers13225848 (2021).

240. Bosse, T. et al. Molecular Classification of Grade 3 Endometrioid Endometrial Cancers Identifies Distinct Prognostic Subgroups. American Journal of Surgical Pathology 42, 561–568 (2018).

241. Cohen Freue, G. v. et al. Computational Biomarker Pipeline from Discovery to Clinical Implementation: Plasma Proteomic Biomarkers for Cardiac Transplantation. PLoS Comput Biol 9, (2013).

242. Dou, Y. et al. Proteogenomic Characterization of Endometrial Carcinoma. Cell 180, 729-748.e26 (2020).

243. 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. Clinical Cancer Research 23, 6458–6467 (2017).

244. Niño, C. A., di Perrotolo, R. S. & Polo, S. Recurrent Spliceosome Mutations in Cancer: Mechanisms and Consequences of Aberrant Splice Site Selection. Cancers vol. 14 Preprint at https://doi.org/10.3390/cancers14020281 (2022).

245. Sette, C. & Paronetto, M. P. Somatic Mutations in Core Spliceosome Components Promote Tumorigenesis and Generate an Exploitable Vulnerability in Human

Cancer. Cancers vol. 14 Preprint at https://doi.org/10.3390/cancers14071827 (2022).

246. Lewczuk, Ł., Pryczynicz, A. & Guzińska-Ustymowicz, K. Cell adhesion molecules in endometrial cancer – A systematic review. Advances in Medical Sciences vol. 64 423–429 Preprint at https://doi.org/10.1016/j.advms.2019.08.003 (2019).

247. di Iorio, P. et al. Unfolding New Roles for Guanine-Based Purines and Their Metabolizing Enzymes in Cancer and Aging Disorders. Frontiers in Pharmacology vol. 12 Preprint at https://doi.org/10.3389/fphar.2021.653549 (2021).

248. Oliveira, K. A., Dal-Cim, T. A., Lopes, F. G., Nedel, C. B. & Tasca, C. I. Guanosine promotes cytotoxicity via adenosine receptors and induces apoptosis in temozolomide-treated A172 glioma cells. Purinergic Signal 13, 305–318 (2017).

249. Zhang, N. et al. Cytotoxicity of guanine-based degradation products contributes to the antiproliferative activity of guanine-rich oligonucleotides. Chem Sci 6, 3831–3838 (2015).

250. Garozzo, R., Sortino, M. A., Vancheri, C. & Condorelli, D. F. Antiproliferative effects induced by guanine-based purines require hypoxanthine-guanine phosphoribosyltransferase activity. Biol Chem 391, 1079–1089 (2010).

251. Itakura, E., Kishi-Itakura, C. & Mizushima, N. The hairpin-type tail-anchored SNARE syntaxin 17 targets to autophagosomes for fusion with endosomes/lysosomes. Cell 151, 1256–1269 (2012).

252. Diao, J. et al. ATG14 promotes membrane tethering and fusion of autophagosomes to endolysosomes. Nature 520, 563–566 (2015).

253. Wang, Y. S. et al. VAMP8, a vesicle-SNARE required for RAB37-mediated exocytosis, possesses a tumor metastasis suppressor function. Cancer Lett 437, 79–88 (2018).

254. Chen, Y. et al. VAMP8 facilitates cellular proliferation and temozolomide resistance in human glioma cells. Neuro Oncol 17, 407–418 (2015).

255. Audet-Delage, Y., Villeneuve, L., Grégoire, J., Plante, M. & Guillemette, C. Identification of metabolomic biomarkers for endometrial cancer and its recurrence after surgery in postmenopausal women. Front Endocrinol (Lausanne) 9, (2018).

256. Li, Y. et al. Integrating pathology, chromosomal instability and mutations for risk stratification in early-stage endometrioid endometrial carcinoma. Cell Biosci 10, (2020).

257. Bosquet, J. G. et al. Association of a novel endometrial cancer biomarker panel with prognostic risk, platinum insensitivity, and targetable therapeutic options. PLoS One 16, (2021).

258. Raffone, A. et al. TCGA molecular groups of endometrial cancer: Pooled data about prognosis. Gynecologic Oncology vol. 155 374–383 Preprint at https://doi.org/10.1016/j.ygyno.2019.08.019 (2019).

259. Lai, J., Xu, T. & Yang, H. Protein-based prognostic signature for predicting the survival and immunotherapeutic efficiency of endometrial carcinoma. BMC Cancer 22, (2022).

260. Guo, C., Liu, S. & Sun, M. Z. Potential role of Anxa1 in cancer. Future Oncology vol. 9 1773–1793 Preprint at https://doi.org/10.2217/fon.13.114 (2013).

261. Voisin, S. N. et al. Identification of novel molecular targets for endometrial cancer using a drill-down LC-MS/MS Approach with iTRAQ. PLoS One 6, (2011).

262. Sheu, M. J. et al. Overexpression of ANXA1 confers independent negative prognostic impact in rectal cancers receiving concurrent chemoradiotherapy. Tumor Biology 35, 7755–7763 (2014).

263. Long, Q. et al. Protein-coding and microRNA biomarkers of recurrence of prostate cancer following radical prostatectomy. American Journal of Pathology 179, 46–54 (2011).

264. Pecci, A., Ma, X., Savoia, A. & Adelstein, R. S. MYH9: Structure, functions and role of non-muscle myosin IIA in human disease. Gene vol. 664 152–167 Preprint at https://doi.org/10.1016/j.gene.2018.04.048 (2018).

265. Schramek, D. et al. Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. Science (1979) 343, 309–313 (2014).

266. Kas, S. M. et al. Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. Nat Genet 49, 1219–1230 (2017).

267. Backes, F. J. et al. Estrogen receptor-alpha as a predictive biomarker in endometrioid endometrial cancer. Gynecol Oncol 141, 312–317 (2016).

268. Schneider, J. et al. The E-Cadherin Expression vs. Tumor Cell Proliferation Paradox in Endometrial Cancer.

269. Stefansson, I. M., Salvesen, H. B. & Akslen, L. A. Prognostic impact of alterations in P-cadherin expression and related cell adhesion markers in endometrial cancer. Journal of Clinical Oncology 22, 1242–1252 (2004).

270. Li, W., Qin, Y., Zhou, R., Liu, Y. & Zhang, G. High expression of SIX1 is an independent predictor of poor prognosis in endometrial cancer. Am J Transl Res vol. 13 www.ajtr.org (2021).

271. Liu, J., Mei, J., Li, S., Wu, Z. & Zhang, Y. Establishment of a novel cell cycle-related prognostic signature predicting prognosis in patients with endometrial cancer. Cancer Cell Int 20, (2020).

272. la Rubia, E. C. de et al. In silico approach for validating and unveiling new applications for prognostic biomarkers of endometrial cancer. Cancers (Basel) 13, (2021).

273. Paulovich, A. G., Whiteaker, J. R., Hoofnagle, A. N. & Wang, P. The interface between biomarker discovery and clinical validation: The tar pit of the protein biomarker pipeline. Proteomics - Clinical Applications vol. 2 1386–1402 Preprint at https://doi.org/10.1002/prca.200780174 (2008).

274. Lopez, M. et al. Mass spectrometric discovery and selective reaction monitoring (SRM) of putative protein biomarker candidates in first trimester trisomy 21 maternal serum. J Proteome Res 10, 133–142 (2011).

275. Lesur, A. & Domon, B. Advances in high-resolution accurate mass spectrometry application to targeted proteomics. Proteomics vol. 15 880–890 Preprint at https://doi.org/10.1002/pmic.201400450 (2015).

276. Cao, E., Chen, Y., Cui, Z. & Foster, P. R. Effect of freezing and thawing rates on denaturation of proteins in aqueous solutions. Biotechnol Bioeng 82, 684–690 (2003).

277. Kelo, E., Noronkoski, T. & Mononen, I. Depletion of L-asparagine supply and apoptosis of leukemia cells induced by human glycosylasparaginase. Leukemia 23, 1167–1171 (2009).

278. Kim, S. I. et al. Proteomic discovery of biomarkers to predict prognosis of high-grade serous ovarian carcinoma. Cancers (Basel) 12, (2020).

279. Shen, K. et al. Distinct genes related to drug response identified in ER positive and ER negative breast cancer cell lines. PLoS One 7, (2012).

280. da Rocha, A. A. et al. Hepatocyte growth factor-regulated tyrosine kinase substrate (HGS) and guanylate kinase 1 (GUK1) are differentially expressed in GH-secreting adenomas. Pituitary 9, 83–92 (2006).

281. Wolfe, K. et al. Dynamic compartmentalization of purine nucleotide metabolic enzymes at leading edge in highly motile renal cell carcinoma. Biochem Biophys Res Commun 516, 50–56 (2019).

282. Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA - Journal of the American Medical Association vol. 323 844–853 Preprint at https://doi.org/10.1001/jama.2020.1166 (2020).

283. Ireson, C. R., Alavijeh, M. S., Palmer, A. M., Fowler, E. R. & Jones, H. J. The role of mouse tumour models in the discovery and development of anticancer drugs. British Journal of Cancer vol. 121 101–108 Preprint at https://doi.org/10.1038/s41416-019-0495-5 (2019).

284. Clohessy, J. G. & Pandolfi, P. P. Mouse hospital and co-clinical trial project-from bench to bedside. Nature Reviews Clinical Oncology vol. 12 491–498 Preprint at https://doi.org/10.1038/nrclinonc.2015.62 (2015).

285. Oh, D. Y. & Bang, Y. J. HER2-targeted therapies — a role beyond breast cancer. Nature Reviews Clinical Oncology vol. 17 33–48 Preprint at https://doi.org/10.1038/s41571-019-0268-3 (2020).

286. Fujimoto-Ouchi, K. et al. Antitumor activity of trastuzumab in combination with chemotherapy in human gastric cancer xenograft models. Cancer Chemother Pharmacol 59, 795–805 (2007).

287. Moasser, M. M. Two dimensions in targeting HER2. Journal of Clinical Oncology 32, 2074–2077 (2014).

288. Thuss-Patience, P. C. et al. Trastuzumab emtansine versus taxane use for previously treated HER2-positive locally advanced or metastatic gastric or gastro-oesophageal junction adenocarcinoma (GATSBY): an international randomised, open-label, adaptive, phase 2/3 study. Lancet Oncol 18, 640–653 (2017).

289. Ito, F., Furukawa, N. & Nakai, T. Evaluation of TOP2A as a Predictive Marker for Endometrial Cancer with Taxane-Containing Adjuvant Chemotherapy. International Journal of Gynecological Cancer 26, 325–330 (2016).

290. Buza, N. & Hui, P. Marked heterogeneity of HER2/NEU gene amplification in endometrial serous carcinoma. Genes Chromosomes Cancer 52, 1178–1186 (2013).

291. Hatch, R. v., Patel, S. U., Cambareri, C., Uritsky, T. & Martin, L. P. Evaluation of the management of PARP inhibitor toxicities in ovarian and endometrial cancer within a multi-institution health-system. Journal of Oncology Pharmacy Practice (2021) doi:10.1177/10781552211024728.

292. Villafranca-Magdalena, B. et al. Genomic Validation of Endometrial Cancer Patient-Derived Xenograft Models as a Preclinical Tool. (2022) doi:10.3390/ijms.

293. Berglund, L. et al. A genecentric human protein atlas for expression profiles based on antibodies. Molecular and Cellular Proteomics vol. 7 2019–2027 Preprint at https://doi.org/10.1074/mcp.R800013-MCP200 (2008).

294. Percy, A. J. et al. Clinical translation of MS-based, quantitative plasma proteomics: status, challenges, requirements, and potential. Expert Review of Proteomics vol. 13 673–684 Preprint at https://doi.org/10.1080/14789450.2016.1205950 (2016).

295. Macur, K. et al. A targeted mass spectrometry immunoassay to quantify osteopontin in fresh-frozen breast tumors and adjacent normal breast tissues. J Proteomics 208, (2019).

296. Meng, B. et al. POLE exonuclease domain mutation predicts long progression-free survival in grade 3 endometrioid carcinoma of the endometrium. Gynecol Oncol 134, 15–19 (2014).

297. Church, D. N. et al. Prognostic significance of POLE proofreading mutations in endometrial cancer. J Natl Cancer Inst 107, (2015).

298. da Cruz Paula, A. et al. Genetic and molecular subtype heterogeneity in newly diagnosed early- and advanced-stage endometrial cancer. Gynecol Oncol 161, 535–544 (2021).

299. Ying, J. et al. Additive effects of variants of unknown significance in replication repair-associated DNA polymerase genes on mutational burden and prognosis across diverse cancers. J Immunother Cancer 9, (2021).

300. Stelloo, E. et al. Improved risk assessment by integrating molecular and clinicopathological factors in early-stage endometrial cancer-combined analysis of the PORTEC cohorts. Clinical Cancer Research 22, 4215–4224 (2016).

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

302. Billingsley, C. C., Cohn, D. E., Mutch, D. G., Hade, E. M. & Goodfellow, P. J. Prognostic Significance of POLE Exonuclease Domain Mutations in High-Grade Endometrioid Endometrial Cancer on Survival and Recurrence: A Subanalysis. International Journal of Gynecological Cancer 26, 933–938 (2016).

303. Vitale, I., Shema, E., Loi, S. & Galluzzi, L. Intratumoral heterogeneity in cancer progression and response to immunotherapy. Nature Medicine vol. 27 212–224 Preprint at https://doi.org/10.1038/s41591-021-01233-9 (2021).

304. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies. Nature Reviews Clinical Oncology vol. 15 81–94 Preprint at https://doi.org/10.1038/nrclinonc.2017.166 (2018).

305. Mota, A. et al. Genetic analysis of uterine aspirates improves the diagnostic value and captures the intra-tumor heterogeneity of endometrial cancers. Modern Pathology 30, 134–145 (2017).

306. Jamieson, A., Thompson, E. F., Huvila, J., Gilks, C. B. & McAlpine, J. N. P53abn Endometrial Cancer: Understanding the most aggressive endometrial cancers in the era of molecular classification. International Journal of Gynecological Cancer vol. 31 907–913 Preprint at https://doi.org/10.1136/ijgc-2020-002256 (2021).

307. Tan, W. C. C. et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer

immunotherapy. Cancer Communications vol. 40 135–153 Preprint at https://doi.org/10.1002/cac2.12023 (2020).

308. Fiore, C. et al. Utility of multispectral imaging in automated quantitative scoring of immunohistochemistry. J Clin Pathol 65, 496–502 (2012).

309. Tsimberidou, A. M., Fountzilas, E., Nikanjam, M. & Kurzrock, R. Review of precision cancer medicine: Evolution of the treatment paradigm. Cancer Treatment Reviews vol. 86 Preprint at https://doi.org/10.1016/j.ctrv.2020.102019 (2020).

310. McAlpine, J. N. & Gilks, C. B. Precision medicine in endometrial cancer. Gynecologic Oncology vol. 154 451–453 Preprint at https://doi.org/10.1016/j.ygyno.2019.08.001 (2019).

311. Dimasi, J. A., Reichert, J. M., Feldman, L. & Malins, A. Clinical approval success rates for investigational cancer drugs. Clin Pharmacol Ther 94, 329–335 (2013).

312. Bonazzi, V. F. et al. Patient-derived xenograft models capture genomic heterogeneity in endometrial cancer. Genome Med 14, (2022).

313. Gao, H. et al. High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med 21, 1318–1325 (2015).

314. Gui, T. et al. TCF3 is epigenetically silenced by EZH2 and DNMT3B and functions as a tumor suppressor in endometrial cancer. Cell Death Differ 28, 3316–3328 (2021).

315. Muenst, S. et al. The immune system and cancer evasion strategies: Therapeutic concepts. Journal of Internal Medicine vol. 279 541–562 Preprint at https://doi.org/10.1111/joim.12470 (2016).

ANNEX

ANNEX 1

Protocol of protein extraction and sample preparation from FFPE tissue for LC-MS analysis

Reagents

FFPE block

RNase Zap (Thermo Fisher Scientific, reference #AM9780)

100 xylene (PanReac AppliChem, reference #251769.2714)

100% ethanol (VWR, reference #20821.330)

Tris-HCl, $C_4H_{11}NO_3$ (Sigma-Aldrich, reference #T1503) for protein extraction buffer

SDS (PanReac AppliChem, reference #A2572) for protein extraction buffer

DC Protein Assay method (Bio-Rad, reference #5000112)

Bovine Serum Albumin, BSA (Sigma-Aldrich, reference #A2153)

Ammonium bicarbonate, NH_4HCO_3 (Sigma-Aldrich, reference #O9830)

Trypsin Gold (Promega, reference #V5111)

HiPPR™ Detergent Removal Resin (Thermo Scientific, reference #88305)

Concentrated formic acid, CH_2O_2 (Merck, reference #1.00264)

Methanol gradient grade for liquid chromatography (Merck, reference #1.06007)

Acetonitrile hypergrade for LC-MS (Merck, reference #1.00029)

Equipment

Microtome for tissue sectioning

Polyethylene naphthalate (PEN) membranes (ZEISS, reference #415190-9041-000)

Crosslinker chamber

Oven up to 65ºC for deparaffinization

Safe-Lock Tubes 1.5 mL (Eppendorf, reference #120.086)

Pipet tips 1000 µL (VWR, reference #613-0341)

Pipet tips 200 µL (VWR, reference #613-0241)

Pipet tips 10 µL (VWR, reference #613-0260)

Scalpels

Vortex

Centrifuge capable of 20.000G

Shaker-incubates capable of incubating at 99ºC and shaking at 1000rpm

Plate reader (Epoch Biotek, program Gen5 2.09)

Ultra MicroSpin™ columns (The Nest Group, reference SUM SS18V)

Speed-vacuum

❖ **Macrodissection**

1. Chemistry sterilization: add abundant RNase Zap to membranes, wash 30 seconds in miliQ water, and incubate at 65ºC for 2 hours.

2. UV sterilization: use a crosslinker chamber at 1 Joule for 30 minutes

3. Cut 3-6 sections of each FFPE block and place them into membranes. The number of sections depends on the percentage of tumor present in the FFPE blocks (i.e., 3 sections for samples with 100% of tumor content and 6 sections for samples with >30 of tumor content)

4. Heat membranes at 65ºC overnight.

5. Deparaffinization:
    a. 20 seconds in xylene
    b. 20 seconds in xylene
    c. 20 seconds in xylene
    d. 30 seconds in 100% ethanol
    e. 30 seconds in 90% ethanol
    f. 30 seconds in 70% ethanol

6. Dry membranes at room temperature for 20 minutes.

7. Using the H&E from the same FFPE block as a template, cut only the tumor tissue from the membrane using a scalpel, and place them into a collection tube.

❖ **Tissue deparaffinization, protein extraction and protein quantification**

1. Add 1 mL of xylene, vortex, and centrifuge at maximum speed for 2 minutes. Discard supernatant without disturbing the pellet.

2. Repeat the process with xylene.

3. Add 1 mL of ethanol, vortex, and centrifuge at maximum speed for 2 minutes. Discard supernatant without disturbing the pellet.

4. Dry pellet at room temperature for 15 minutes.

5. Resuspend the pellet in 100-200 μL of extraction buffer. The volume is approximately 1:1 of sample:extraction buffer.

6. Heat tubes at 99ºC for 30 minutes with agitation at 750 rpm.

7. Change tubes to a new heating block, and heat samples at 80ºC for 2 hours with agitation at 750rpm. In case of using only one heating block, keep samples at room temperature while cooling the heating block.

8. Centrifuge samples at room temperature for 20 minutes at 15.000G. Transfer the supernatant containing the protein extract to a new collection tube.

9. Do protein quantification with the DC Protein Assay method:

    a. Vortex samples.

    b. Prepare the standard curve using extraction buffer.

    c. Add 5 μL of sample to well (triplicate for each sample).

    d. Add 25 μL of MIX (Reagent A + Reagent S at dilution 1:50) and 200 μL of Reagent B to well.

    e. Shake and leave at room temperature for 15 minutes. Avoid from light.

    f. Read at 620 nm.

Table X. Standard curve for protein quantification

Sample	BSA (mg/mL)	BSA (µL)	EB (µL)
A	2.00	450 (stock)	-
B	1.75	350 from A	50
C	1.50	300 from B	50
D	1.25	250 from C	50
E	1.00	200 from D	50
F	0.75	150 from E	50
G	0.50	100 from F	50
H	0.25	50 from G	50
I	0.125	50 from H	50
J	0.063	50 from I	50
K	0.031	50 from J	50
L	0	-	50

❖ **Sample preparation**

1. Take 10 μg of protein extract and dilute it 20 times (dilution 1:20) with 50 mM ammonium bicarbonate (NH_4HCO_3).

2. Add 1 μL of Trypsin Gold at 1 mg/mL and incubate overnight at 37ºC with agitation at 750rpm.

3. Detergent elimination:

    a. Add 25-200 μL of Detergent Removal Resin to the column, centrifuge 1 minute at 1500G and discard the flow-through.

    b. Add 25-200 μL of 50 mM ammonium bicarbonate to the column, centrifuge 1 minute at 1500G and discard the flow-through.

    c. Repeat three times the previous step.

    d. Add 25-200 μL of protein extract to the column, vortex and incubate at room temperature for 10 minutes.

    e. Centrifuge the column 2 minutes at 1500G. Save the flow-through for the next steps, which contain protein extract without SDS.

4. Desalting:

    a. Add 10% of volume of 100% concentrated formic acid (CH_2O_2).

    b. Conditioning: add 400 μL of methanol to the column, centrifuge 5 minutes at 200G and discard the flow-through.

    c. Equilibration: add 300 μL of formic acid at 5%, centrifuge 5 minutes at 200G and discard the flow-through. Repeat this step twice.

    d. Loading: add the protein extract and centrifuge 10 minutes at 100G. Charge again the flow-through to the column, centrifuge 10 minutes at 100G and discard the flow-through.

    e. Washing: add 300 μL of formic acid at 5%, centrifuge 5 minutes at 100G and discard the flow-through. Repeat this step twice.

    f. Elution: add 300 μL of elution buffer (50% acetonitrile and 5% formic acid) and centrifuge 5 minutes at 100G. Add again 300 μL of elution buffer and centrifuge 5 minutes at 100G (total volume of 600 μL).

5. Evaporation: use a speed-vacuum to evaporate samples and resuspend the pellet with formic acid (CH_2O_2).

SAMPLES ARE READY TO BE USED FOR LC-MS ANALYSIS

Annexes

ANNEX 2

Comparisons between recurrent and non-recurrent EC patients in the endometrioid histology

Comparisons 1 to 18 summarize the statistics for the most important proteins from each comparison in the quantitative analysis (dark green represents proteins with and adjusted p-value < 0.05; light green represents proteins with an adjusted p-value < 0.25); and grey represents proteins selected for the verification phase).

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				Comparison 1 – 27 REC vs 31 NO REC					
1	MYH9	-0,714	1,95E-06	0,00155	56	TPM4	-0,442	0,00290	0,08710
2	TSTD1	1,393	1,31E-06	0,00155	57	HSPA9	-0,433	0,00323	0,08877
3	RPS14	0,657	1,68E-05	0,00854	58	SRP14	0,319	0,00326	0,08877
4	SNRPF	0,629	2,15E-05	0,00854	59	UFC1	0,609	0,00329	0,08877
5	ACTN4	-0,615	0,00015	0,03302	60	SRSF1	0,478	0,00337	0,08932
6	CLIC4	-0,604	0,00023	0,03302	61	PGM5	-0,952	0,00351	0,09160
7	DPP7	1,032	0,00026	0,03302	62	TRAP1	-0,663	0,00375	0,09633
8	GAA	0,805	0,00027	0,03302	63	ACBD3	0,379	0,00467	0,09645
9	NUCKS1	0,686	0,00021	0,03302	64	BAG2	-0,708	0,00443	0,09645
10	PIGR	2,776	0,00011	0,03302	65	C9orf142	0,509	0,00448	0,09645
11	PMVK	0,688	0,00024	0,03302	66	CLU	1,092	0,00465	0,09645
12	PURB	0,620	0,00027	0,03302	67	DDB1	-0,652	0,00438	0,09645
13	SRSF2	0,679	0,00014	0,03302	68	GGCT	0,473	0,00417	0,09645
14	POLR2H	0,612	0,00030	0,03452	69	GNAI2	-0,552	0,00458	0,09645
15	TLN1	-0,579	0,00036	0,03826	70	GSN	-0,314	0,00399	0,09645
16	TGM2	-0,783	0,00041	0,04076	71	HSPD1	-0,567	0,00386	0,09645
17	ACTN1	-0,634	0,00058	0,04611	72	PSMC1	-0,499	0,00433	0,09645
18	CNN3	-0,672	0,00054	0,04611	73	RPS28	0,921	0,00441	0,09645
19	COL18A1	0,771	0,00055	0,04611	74	SEC61B	0,554	0,00448	0,09645
20	SRSF7	0,634	0,00058	0,04611	75	SF3B4	0,410	0,00398	0,09645
21	GUK1	0,528	0,00061	0,04642	76	STIP1	-0,341	0,00451	0,09645
22	LIN7C	0,452	0,00085	0,05708	77	TCEB2	0,407	0,00441	0,09645
23	MRPL14	0,531	0,00082	0,05708	78	ANXA6	-0,987	0,00486	0,09785
24	MTHFD1	-0,562	0,00086	0,05708	79	LRPPRC	-0,772	0,00491	0,09785
25	IDH2	-0,509	0,00093	0,05893	80	SLTM	0,398	0,00492	0,09785
26	ILK	-0,572	0,00100	0,05893	81	HBD	-0,985	0,00506	0,09939
27	PSMD3	-1,055	0,00098	0,05893	82	IDH3A	-0,346	0,00518	0,10045
28	MRPS21	0,547	0,00112	0,06221	83	AGA	0,710	0,00540	0,10359
29	RPS18	0,409	0,00113	0,06221	84	OGDH	-0,458	0,00561	0,10432
30	ENAH	0,569	0,00121	0,06413	85	RPL8	0,627	0,00570	0,10432
31	ACO2	-0,499	0,00129	0,06641	86	SDHA	-0,698	0,00570	0,10432
32	RPS17	0,439	0,00154	0,07441	87	SRSF3	0,488	0,00551	0,10432
33	VCL	-0,528	0,00154	0,07441	88	RPL22	0,578	0,00582	0,10458
34	ANXA1	-0,778	0,00192	0,07501	89	SERPINH1	-0,596	0,00585	0,10458
35	CORO1C	-0,567	0,00191	0,07501	90	CBR1	-0,502	0,00602	0,10526
36	GBP1	-1,021	0,00189	0,07501	91	DUSP23	0,570	0,00609	0,10526
37	IQGAP1	-0,498	0,00192	0,07501	92	GUSB	0,545	0,00605	0,10526
38	PRKDC	-0,924	0,00193	0,07501	93	LGALS3BP	0,674	0,00621	0,10618
39	UFL1	-0,428	0,00175	0,07501	94	CAP1	-0,399	0,00673	0,11311
40	VAMP8	0,584	0,00169	0,07501	95	DNM2	-0,439	0,00685	0,11311
41	VIM	-0,602	0,00186	0,07501	96	HAX1	0,445	0,00689	0,11311
42	LSM8	0,566	0,00214	0,07799	97	NENF	0,590	0,00690	0,11311
43	MIEN1	0,595	0,00217	0,07799	98	CCAR1	0,386	0,00703	0,11396
44	RPL31	0,474	0,00208	0,07799	99	ELAVL1	0,344	0,00709	0,11396
45	SEC16A	0,573	0,00221	0,07799	100	LSM5	0,561	0,00780	0,12403
46	BPNT1	0,579	0,00231	0,07809	101	EIF3L	-0,565	0,00790	0,12449
47	NAP1L4	-0,312	0,00231	0,07809	102	EML4	0,619	0,00805	0,12557
48	KARS	-0,346	0,00249	0,08138	103	CPSF6	0,367	0,00821	0,12686
49	LSM4	0,439	0,00251	0,08138	104	SNX12	0,425	0,00835	0,12768
50	THRAP3	0,617	0,00269	0,08549	105	U2AF2	0,357	0,00857	0,12986
51	DDX1	-0,858	0,00280	0,08710	106	TCOF1	0,537	0,00869	0,13049
52	LGALS1	-0,698	0,00300	0,08710	107	DBI	0,581	0,00900	0,13381
53	MSN	-0,524	0,00302	0,08710	108	BOLA2;BOLA2B	0,456	0,00915	0,13460
54	PSMC5	-0,605	0,00307	0,08710	109	RPS20	0,396	0,00925	0,13460
55	TMEM43	-0,456	0,00298	0,08710	110	PSMD12	-0,470	0,00932	0,13460

| Comparison 1 – 27 REC vs 31 NO REC ||||||||
Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
111	RAB6A	0,433	0,00939	0,13460	182	RPL18	0,549	0,02271	0,19555
112	SIAE	0,536	0,00989	0,13994	183	BRK1	0,403	0,02279	0,19555
113	ACAT1	-0,604	0,00994	0,13994	184	SORBS2	0,498	0,02281	0,19555
114	SRSF9	0,481	0,01033	0,14415	185	HDDC2	0,427	0,02291	0,19555
115	CALD1	-0,652	0,01094	0,14891	186	ALDH2	-0,498	0,02291	0,19555
116	S100A11	0,554	0,01098	0,14891	187	APOA1	-0,737	0,02305	0,19555
117	CA1	-1,281	0,01104	0,14891	188	DYNLT1	0,350	0,02311	0,19555
118	CHTOP	0,426	0,01104	0,14891	189	CST3	0,653	0,02339	0,19688
119	RPL19	0,543	0,01117	0,14923	190	SGSH	0,448	0,02370	0,19801
120	NUDT5	0,348	0,01126	0,14923	191	SRSF10	0,379	0,02377	0,19801
121	IRF2BP2	0,319	0,01144	0,15036	192	CSE1L	-0,498	0,02409	0,19965
122	RBMX	0,428	0,01169	0,15131	193	ADH5	-0,581	0,02482	0,20411
123	ADRM1	0,381	0,01170	0,15131	194	CKB	-0,816	0,02489	0,20411
124	DDX19A	-0,365	0,01191	0,15277	195	GNG12	0,475	0,02510	0,20482
125	PAICS	-0,487	0,01217	0,15436	196	OXCT1	-0,545	0,02533	0,20555
126	ABCE1	-0,497	0,01222	0,15436	197	SDF4	0,448	0,02545	0,20555
127	SEPT2	-0,427	0,01278	0,15630	198	TCEB1	0,268	0,02604	0,20599
128	PSMC2	-0,348	0,01278	0,15630	199	RPS5	0,235	0,02606	0,20599
129	WIBG	0,405	0,01279	0,15630	200	ETF1	-0,294	0,02611	0,20599
130	HDLBP	-0,402	0,01283	0,15630	201	UBA1	-0,277	0,02614	0,20599
131	HBB	-0,892	0,01287	0,15630	202	FLNA	-0,437	0,02618	0,20599
132	POLR2C	0,308	0,01306	0,15746	203	SEPT7	-0,548	0,02628	0,20599
133	WDR1	-0,423	0,01361	0,16182	204	PSMC4	-0,315	0,02651	0,20678
134	HADHA	-0,702	0,01363	0,16182	205	EIF4G1	-0,303	0,02670	0,20725
135	CORO1A	-0,579	0,01393	0,16370	206	PSMC3	-0,242	0,02691	0,20780
136	COA3	0,482	0,01403	0,16370	207	C21orf33	-0,379	0,02706	0,20797
137	SUMF2	0,518	0,01419	0,16370	208	TXN	0,427	0,02725	0,20841
138	LAD1	0,733	0,01420	0,16370	209	CIRBP	0,379	0,02776	0,21086
139	NUDT21	0,320	0,01440	0,16480	210	PSMD14	-0,370	0,02788	0,21086
140	TES	-0,437	0,01469	0,16694	211	EMD	0,300	0,02823	0,21086
141	RBM12	0,284	0,01545	0,17429	212	RSU1	-0,488	0,02824	0,21086
142	SCPEP1	0,424	0,01566	0,17429	213	ACTR3	-0,407	0,02836	0,21086
143	SNRPD3	0,573	0,01585	0,17429	214	S100A4	-0,741	0,02836	0,21086
144	PSMD13	-0,445	0,01601	0,17429	215	GFPT1	-0,625	0,02856	0,21138
145	SUCLA2	-0,440	0,01602	0,17429	216	ASNA1	-0,270	0,02954	0,21760
146	HIST1H3A;HIST2H3A;HIST3H3	0,672	0,01615	0,17429	217	MLF2	0,468	0,03001	0,21898
147	SUB1	0,431	0,01628	0,17429	218	FN3K	0,336	0,03012	0,21898
148	EIF3J	-0,284	0,01636	0,17429	219	ATP5H	-0,275	0,03025	0,21898
149	FAM136A	0,532	0,01653	0,17429	220	SAP18	0,319	0,03028	0,21898
150	DSG2	0,590	0,01665	0,17429	221	SNAPIN	0,344	0,03068	0,22087
151	RPS21	0,384	0,01673	0,17429	222	RPS4X	0,343	0,03095	0,22141
152	MAGOHB	0,434	0,01684	0,17429	223	RAB11B;RAB11A	0,235	0,03103	0,22141
153	PDLIM7	-0,663	0,01692	0,17429	224	EHD2	-0,605	0,03136	0,22242
154	EIF3C;EIF3CL	-0,308	0,01695	0,17429	225	UBE2V1	0,343	0,03150	0,22242
155	PCBD1	0,505	0,01705	0,17429	226	PALLD	-0,427	0,03168	0,22242
156	CP	0,661	0,01709	0,17429	227	STAT1	-0,784	0,03184	0,22242
157	HNRNPF	0,309	0,01746	0,17695	228	ETFA	-0,287	0,03187	0,22242
158	GSR	-0,557	0,01802	0,18080	229	SAMHD1	-0,665	0,03209	0,22295
159	ARHGAP1	-0,257	0,01807	0,18080	230	PLS3	-0,402	0,03252	0,22423
160	CNPY2	0,440	0,01821	0,18106	231	LSM3	0,566	0,03256	0,22423
161	TSTA3	0,370	0,01839	0,18122	232	HEXB	0,371	0,03315	0,22643
162	TACSTD2	0,554	0,01855	0,18122	233	FGG	-0,698	0,03316	0,22643
163	APOA4	-0,771	0,01869	0,18122	234	PRDX6	0,324	0,03446	0,23389
164	USMG5	0,320	0,01870	0,18122	235	DCTN3	0,262	0,03455	0,23389
165	ABI1	0,336	0,01888	0,18122	236	VTI1B	0,367	0,03511	0,23601
166	SRSF6	0,472	0,01891	0,18122	237	ZNF326	-0,369	0,03516	0,23601
167	HSP90AA1	-0,441	0,01923	0,18298	238	PAPSS1	-0,513	0,03561	0,23804
168	PSMC6	-0,348	0,01932	0,18298	239	TOMM6	0,632	0,03697	0,24499
169	PTRF	-0,798	0,01948	0,18306	240	ACY1;ABHD14A	0,396	0,03698	0,24499
170	FARSA	0,273	0,01956	0,18306	241	ATP2B4	0,350	0,03742	0,24499
171	PSMD11	-0,332	0,02005	0,18653	242	SEPHS1	0,339	0,03746	0,24499
172	HSP90AB1	-0,421	0,02032	0,18681	243	LSM6	0,336	0,03755	0,24499
173	ACADVL	-0,789	0,02036	0,18681	244	ARHGDIB	-0,426	0,03776	0,24499
174	EIF3B	-0,286	0,02051	0,18681	245	GNS	0,341	0,03802	0,24499
175	HBA1	-0,840	0,02059	0,18681	246	UBAP2L	0,344	0,03823	0,24499
176	SELH	0,273	0,02067	0,18681	247	SRI	0,380	0,03827	0,24499
177	FDPS	0,432	0,02117	0,19033	248	HDGF	0,373	0,03830	0,24499
178	RPL11	0,261	0,02159	0,19294	249	HNRNPU	-0,335	0,03834	0,24499
179	EVL	-0,631	0,02225	0,19555	250	SRRM2	0,494	0,03870	0,24631
180	NME1	0,412	0,02233	0,19555	251	H2AFX	-1,294	0,03884	0,24631
181	SON	0,450	0,02247	0,19555	252	ALDH1B1	-1,038	0,03896	0,24631

236

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				Comparison 2 – Stage IA (6 REC vs 8 NOREC)					
1	CBX5	-1,027	0,00096	0,99165	51	RPS15A	0,810	0,03672	0,99165
2	ABCE1	-1,264	0,00129	0,99165	52	FBP1	1,137	0,03698	0,99165
3	DENR	-0,817	0,00374	0,99165	53	GBP1	-1,404	0,03740	0,99165
4	OXSR1	-0,712	0,00410	0,99165	54	EIF4G2	-0,606	0,03819	0,99165
5	NAP1L4	-0,587	0,00417	0,99165	55	ENO2	-0,964	0,03859	0,99165
6	FXR1	-0,893	0,00460	0,99165	56	COPS8	0,723	0,03860	0,99165
7	SDF4	1,140	0,00487	0,99165	57	MYDGF	1,100	0,03909	0,99165
8	COL6A2	1,425	0,00624	0,99165	58	DBI	0,872	0,03944	0,99165
9	BPGM	-1,379	0,00974	0,99165	59	GLUD1	0,759	0,04252	0,99165
10	PPP5C	-0,756	0,01050	0,99165	60	STRAP	-0,491	0,04335	0,99165
11	G3BP2	-0,614	0,01053	0,99165	61	IDH2	-0,632	0,04437	0,99165
12	BPNT1	0,987	0,01123	0,99165	62	BCKDHA	0,489	0,04517	0,99165
13	PRPSAP2	-0,699	0,01332	0,99165	63	PIGR	2,888	0,04547	0,99165
14	CTTN	-0,918	0,01502	0,99165	64	PRKCSH	-0,675	0,04560	0,99165
15	SUB1	0,830	0,01721	0,99165	65	HBD	-1,456	0,04608	0,99165
16	QDPR	0,740	0,01755	0,99165	66	HTRA2	0,592	0,04635	0,99165
17	APOC3	-1,250	0,01881	0,99165	67	ILKAP;ILKAP3	-0,563	0,04669	0,99165
18	RPS17	0,655	0,01883	0,99165	68	BOLA2;BOLA2B	0,697	0,04826	0,99165
19	RPSA	0,578	0,02017	0,99165	69	CHMP1B	-0,560	0,04832	0,99165
20	HNRNPD	-0,641	0,02067	0,99165	70	DBN1	-0,732	0,04851	0,99165
21	KARS	-0,535	0,02125	0,99165	71	PSMC5	-0,807	0,05026	0,99165
22	ABI1	0,647	0,02159	0,99165	72	SEPT11	-0,646	0,05054	0,99165
23	CHD4	-0,662	0,02205	0,99165	73	TMEM43	-0,604	0,05061	0,99165
24	EIF3B	-0,566	0,02261	0,99165	74	STK24	-0,432	0,05063	0,99165
25	PDCD4	0,779	0,02285	0,99165	75	NASP	-0,595	0,05131	0,99165
26	PRDX3	0,807	0,02339	0,99165	76	SAE1	-0,530	0,05276	0,99165
27	ARF1;ARF3	1,128	0,02387	0,99165	77	PAK2	-0,484	0,05331	0,99165
28	DDX42	-0,500	0,02410	0,99165	78	SUCLA2	-0,722	0,05414	0,99165
29	PGM5	-1,511	0,02434	0,99165	79	RAB14	0,741	0,05440	0,99165
30	DDX1	-1,260	0,02481	0,99165	80	STMN1	-0,816	0,05508	0,99165
31	ATP5H	-0,578	0,02550	0,99165	81	ACTR1A	-0,583	0,05713	0,99165
32	CASP3	-0,541	0,02573	0,99165	82	C21orf33	-0,648	0,05954	0,99165
33	SUMF2	0,969	0,02621	0,99165	83	SERPINB6	0,723	0,06041	0,99165
34	UBA2	-0,497	0,02633	0,99165	84	SEC23A	-0,665	0,06152	0,99165
35	DDX19A	-0,607	0,02968	0,99165	85	VAPA	-0,532	0,06184	0,99165
36	DDX3X	-0,750	0,02970	0,99165	86	NENF	0,807	0,06189	0,99165
37	CYCS	0,813	0,02992	0,99165	87	MTX1	0,490	0,06369	0,99165
38	PDIA6	0,682	0,03017	0,99165	88	GANAB	-0,556	0,06474	0,99165
39	GNAI2	-0,854	0,03025	0,99165	89	NANS	0,620	0,06614	0,99165
40	PSMD3	-1,412	0,03042	0,99165	90	MCM3	-1,029	0,06636	0,99165
41	MRPL14	0,675	0,03093	0,99165	91	PDIA4	0,507	0,06768	0,99165
42	PTPN11	-0,565	0,03135	0,99165	92	ZNF326	-0,642	0,06796	0,99165
43	HNRNPU	-0,690	0,03148	0,99165	93	SRSF2	0,639	0,06930	0,99165
44	NNMT	-0,905	0,03249	0,99165	94	PAICS	-0,668	0,07077	0,99165
45	FKBP10	-1,137	0,03295	0,99165	95	SSRP1	-0,538	0,07115	0,99165
46	RPA1	-0,949	0,03375	0,99165	96	THRAP3	0,741	0,07207	0,99165
47	MYH9	-0,613	0,03417	0,99165	97	UPF1	-0,808	0,07528	0,99165
48	IL4I1	1,254	0,03575	0,99165	98	PSMD12	-0,664	0,07566	0,99165
49	RPS28	1,363	0,03628	0,99165	99	FTH1	-0,785	0,07601	0,99165
50	DNAJA1	-0,724	0,03641	0,99165	100	PCBD1	0,761	0,07620	0,99165

Comparison 3 – Stage IB (12 REC cs 13 NOREC)									
Nº	**ID**	**logFC**	**P.Value**	**adj.P.Val**	**Nº**	**ID**	**logFC**	**P.Value**	**adj.P.Val**
1	MZB1	-2,108	0,00077	0,60118	51	PUF60	0,546	0,02674	0,83404
2	RPS14	0,731	0,00105	0,60118	52	RPL19	0,713	0,02762	0,83405
3	MYH9	-0,698	0,00150	0,60118	53	HNRNPA0	0,692	0,02778	0,83405
4	TSTD1	1,297	0,00207	0,60118	54	HSPD1	-0,653	0,02945	0,83543
5	STIP1	-0,566	0,00217	0,60118	55	PURB	0,523	0,03119	0,83543
6	MYO6	-0,764	0,00238	0,60118	56	CBR1	-0,604	0,03134	0,83543
7	SNRPF	0,644	0,00279	0,60118	57	DAZAP1	0,499	0,03166	0,83543
8	U2AF2	0,573	0,00302	0,60118	58	RPS18	0,375	0,03226	0,83543
9	ANXA1	-1,055	0,00415	0,72236	59	RAB1B	0,484	0,03322	0,83543
10	OGDH	-0,698	0,00470	0,72236	60	GNG12	0,699	0,03350	0,83543
11	ACO2	-0,650	0,00597	0,72236	61	HIST1H4A	0,958	0,03356	0,83543
12	POLR2C	0,497	0,00605	0,72236	62	RPL13	0,758	0,03389	0,83543
13	LSM4	0,594	0,00617	0,72236	63	APCS	-1,166	0,03426	0,83543
14	RPL17	0,622	0,00807	0,72236	64	CORO7-PAM16;CORO7	-0,444	0,03440	0,83543
15	ACAA1	-0,847	0,00810	0,72236	65	ACTA2;ACTG2	-1,065	0,03646	0,83543
16	HSPA9	-0,593	0,00886	0,72236	66	CHTOP	0,531	0,03696	0,83543
17	RBMX	0,652	0,00955	0,72236	67	VCL	-0,525	0,03759	0,83543
18	H1FX	0,815	0,00964	0,72236	68	MCCC2	-0,727	0,03785	0,83543
19	TRAP1	-0,900	0,01020	0,72236	69	CPSF7	0,611	0,03851	0,83543
20	SEC61B	0,715	0,01021	0,72236	70	NUDT21	0,416	0,03858	0,83543
21	TLN1	-0,621	0,01022	0,72236	71	CORO1C	-0,574	0,03869	0,83543
22	OXCT1	-0,944	0,01064	0,72236	72	MYLK	-0,946	0,03962	0,83543
23	ACTN4	-0,614	0,01110	0,72236	73	MYL1;MYL3	-0,762	0,03985	0,83543
24	PTBP3	0,696	0,01278	0,72236	74	APOA4	-1,011	0,04090	0,83543
25	NAP1L4	-0,374	0,01293	0,72236	75	USMG5	0,415	0,04091	0,83543
26	SRP14	0,399	0,01407	0,72236	76	NUCKS1	0,558	0,04149	0,83543
27	PMVK	0,673	0,01490	0,72236	77	FLYWCH2	0,459	0,04188	0,83543
28	SNRPD2	0,449	0,01492	0,72236	78	SRSF7	0,532	0,04281	0,83543
29	IQGAP1	-0,588	0,01516	0,72236	79	CPSF6	0,408	0,04315	0,83543
30	LGALS3BP	0,904	0,01544	0,72236	80	LSM8	0,555	0,04345	0,83543
31	PIGR	2,603	0,01582	0,72236	81	FLNA	-0,613	0,04373	0,83543
32	PHYHD1	0,701	0,01672	0,72236	82	VTN	-0,871	0,04378	0,83543
33	UQCRB	-0,575	0,01686	0,72236	83	RDX	-0,476	0,04621	0,83543
34	ESD	0,468	0,01764	0,72236	84	HYOU1	-0,477	0,04633	0,83543
35	ACTN1	-0,646	0,01766	0,72236	85	HNRNPH1	0,434	0,04673	0,83543
36	CCAR1	0,520	0,01775	0,72236	86	ERP29	-0,530	0,04704	0,83543
37	COA3	0,710	0,01796	0,72236	87	ATG3	-0,414	0,04793	0,83543
38	CLIC4	-0,583	0,01830	0,72236	88	RPL35A	0,988	0,04876	0,83543
39	RAB6A	0,584	0,01930	0,72236	89	ELAVL1	0,372	0,04913	0,83543
40	GSN	-0,389	0,01952	0,72236	90	EFHD2	-0,396	0,05035	0,83543
41	SRSF3	0,586	0,01968	0,72236	91	PRKDC	-0,848	0,05113	0,83543
42	CALD1	-0,923	0,01979	0,72236	92	ERO1L	-0,630	0,05136	0,83543
43	HSPB1	-0,767	0,01997	0,72236	93	SRSF2	0,506	0,05255	0,83543
44	HIBADH	-0,689	0,01998	0,72236	94	RPL15	0,728	0,05265	0,83543
45	RPS4X	0,554	0,02149	0,74686	95	MYL9	-0,779	0,05382	0,83543
46	ACAT1	-0,822	0,02159	0,74686	96	PSMC1	-0,513	0,05389	0,83543
47	SELH	0,397	0,02258	0,76172	97	TPM4	-0,435	0,05409	0,83543
48	NANS	-0,573	0,02298	0,76172	98	SNRPE	0,621	0,05424	0,83543
49	CKB	-1,237	0,02509	0,81457	99	RANBP1	-0,502	0,05464	0,83543
50	ACTL6A	0,357	0,02584	0,82236	100	RPS7	0,469	0,05502	0,83543

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
Comparison 4 – Stage II (9 REC vs 10 NOREC)									
1	ENAH	1,465	4,90E-07	0,00078	51	PURB	0,871	0,00222	0,06872
2	DPP7	2,183	4,80E-06	0,00382	52	IRGQ	0,872	0,00225	0,06872
3	SCPEP1	1,225	3,18E-05	0,01464	53	SON	1,025	0,00241	0,07221
4	DNASE2	2,341	3,68E-05	0,01464	54	ARHGEF2	0,908	0,00255	0,07320
5	DYNLT1	0,992	8,14E-05	0,01666	55	LSM5	1,086	0,00256	0,07320
6	GNS	1,102	8,38E-05	0,01666	56	TOMM6	1,546	0,00258	0,07320
7	MLF2	1,422	5,97E-05	0,01666	57	CDH1	1,531	0,00311	0,07622
8	GUK1	1,051	7,15E-05	0,01666	58	THRAP3	1,073	0,00282	0,07622
9	ALDH2	-1,424	0,00015	0,02458	59	MIEN1	1,002	0,00306	0,07622
10	TSTD1	1,869	0,00015	0,02458	60	SLTM	0,731	0,00302	0,07622
11	TGM2	-1,440	0,00019	0,02718	61	LIN7C	0,703	0,00310	0,07622
12	COL18A1	1,420	0,00022	0,02920	62	PAFAH1B3	1,050	0,00283	0,07622
13	SRSF9	1,138	0,00033	0,03059	63	UBA1	-0,644	0,00295	0,07622
14	GAA	1,363	0,00031	0,03059	64	ARHGAP1	-0,555	0,00304	0,07622
15	TBCA	0,723	0,00025	0,03059	65	MRPL14	0,807	0,00292	0,07622
16	ILK	-1,073	0,00032	0,03059	66	AGA	1,289	0,00325	0,07830
17	SNRPF	0,908	0,00031	0,03059	67	NUCB1	0,597	0,00332	0,07887
18	HDDC2	1,133	0,00035	0,03064	68	DDX1	-1,420	0,00342	0,07962
19	SEC16A	1,111	0,00045	0,03806	69	CAP1	-0,750	0,00345	0,07962
20	VIM	-1,117	0,00049	0,03929	70	CA1	-2,540	0,00397	0,08156
21	NUDT5	0,806	0,00058	0,04362	71	GOT2	-1,211	0,00400	0,08156
22	RPS27A;UBC;UBB;UBA52	0,821	0,00083	0,04423	72	RPS12	0,727	0,00367	0,08156
23	TCEB2	0,812	0,00075	0,04423	73	PMVK	0,928	0,00382	0,08156
24	WIBG	0,949	0,00079	0,04423	74	FN3K	0,776	0,00391	0,08156
25	SGSH	1,135	0,00067	0,04423	75	GGCT	0,834	0,00387	0,08156
26	SORBS2	1,274	0,00065	0,04423	76	MAP1B	2,117	0,00365	0,08156
27	LRPPRC	-1,595	0,00082	0,04423	77	GFPT1	-1,457	0,00370	0,08156
28	POLR2H	0,978	0,00067	0,04423	78	TACSTD2	1,199	0,00384	0,08156
29	CLU	2,238	0,00076	0,04423	79	DDB1	-1,144	0,00407	0,08191
30	NUCKS1	1,083	0,00083	0,04423	80	RPS14	0,726	0,00412	0,08202
31	UBAP2L	0,921	0,00086	0,04426	81	PEA15	0,909	0,00419	0,08226
32	RPL31	0,846	0,00103	0,05108	82	TLN1	-0,795	0,00446	0,08646
33	SRRM2	1,326	0,00107	0,05159	83	HAX1	0,812	0,00474	0,09008
34	MYH9	-0,823	0,00115	0,05211	84	SRSF7	0,865	0,00476	0,09008
35	CP	1,537	0,00112	0,05211	85	NSFL1C	0,594	0,00538	0,09100
36	SERPINH1	-1,198	0,00119	0,05276	86	ORM1	1,501	0,00523	0,09100
37	CHMP6	0,787	0,00131	0,05627	87	SRP9	1,109	0,00529	0,09100
38	ANXA6	-1,945	0,00147	0,06136	88	HN1L	1,078	0,00524	0,09100
39	MSN	-0,975	0,00164	0,06668	89	GTF2F1	0,721	0,00533	0,09100
40	ATP2B4	0,927	0,00171	0,06668	90	MRPL49	0,710	0,00491	0,09100
41	DDX19A	-0,756	0,00183	0,06668	91	ACTR3	-0,905	0,00505	0,09100
42	VAMP8	0,983	0,00184	0,06668	92	ACTR2	-1,068	0,00501	0,09100
43	NENF	1,184	0,00175	0,06668	93	RPS28	1,579	0,00499	0,09100
44	SRSF2	0,963	0,00177	0,06668	94	DSG2	1,220	0,00513	0,09100
45	MRPS21	0,879	0,00192	0,06802	95	BAG6;BAT3	0,630	0,00555	0,09288
46	RPL22	1,094	0,00200	0,06872	96	PFDN2	0,642	0,00584	0,09398
47	C9orf142	0,950	0,00223	0,06872	97	CLTA	0,863	0,00585	0,09398
48	RPS18	0,630	0,00214	0,06872	98	ACBD3	0,650	0,00569	0,09398
49	SUMO1	0,942	0,00215	0,06872	99	CRK	0,529	0,00575	0,09398
50	FDX1	1,000	0,00208	0,06872	100	CLIC4	-0,785	0,00611	0,09723

Comparison 5 – Grade 1&2 (14 REC vs 15 NOREC)								
Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
1	TSTD1	2,098	9,00E-05	0,14319	51	AHSA1	-0,610	0,00998	0,28391
2	SF3B4	0,904	0,00033	0,22734	52	TOMM6	1,478	0,01007	0,28391
3	GAA	1,424	0,00060	0,22734	53	HSPD1	-0,922	0,01019	0,28391
4	ANXA6	-2,214	0,00070	0,22734	54	CSTF2	0,893	0,01024	0,28391
5	SRSF2	1,116	0,00073	0,22734	55	SNRPF	0,671	0,01035	0,28391
6	LIN7C	0,854	0,00088	0,22734	56	DUSP23	1,020	0,01041	0,28391
7	PIGR	4,260	0,00115	0,22734	57	RPLP0;RPLP0P6	-0,872	0,01046	0,28391
8	BAG2	-1,508	0,00118	0,22734	58	RPL31	0,738	0,01048	0,28391
9	VIM	-1,152	0,00188	0,22734	59	FDX1	0,912	0,01053	0,28391
10	SEC16A	1,090	0,00190	0,22734	60	DDAH1	1,129	0,01083	0,28391
11	WIBG	0,959	0,00204	0,22734	61	ATP5L	-0,677	0,01091	0,28391
12	ASNA1	-0,695	0,00213	0,22734	62	GRB2	0,717	0,01114	0,28391
13	CYB5R3	-0,843	0,00217	0,22734	63	NUCKS1	0,864	0,01124	0,28391
14	CLIC4	-0,952	0,00218	0,22734	64	PTBP3	0,914	0,01228	0,30103
15	LRPPRC	-1,501	0,00237	0,22734	65	STX7	0,746	0,01243	0,30103
16	PSMC6	-0,815	0,00260	0,22734	66	ILK	-0,815	0,01249	0,30103
17	TCOF1	1,167	0,00265	0,22734	67	TMEM43	-0,705	0,01355	0,31246
18	LSM8	1,039	0,00268	0,22734	68	DYNLT1	0,723	0,01369	0,31246
19	VTI1B	0,989	0,00279	0,22734	69	IRGQ	0,769	0,01373	0,31246
20	C9orf142	1,003	0,00286	0,22734	70	DNASE2	1,591	0,01417	0,31246
21	GUK1	0,862	0,00302	0,22861	71	THUMPD1	0,645	0,01429	0,31246
22	SERPINH1	-1,165	0,00316	0,22861	72	SEC61B	0,909	0,01446	0,31246
23	BRK1	0,969	0,00341	0,23610	73	GTF2F1	0,674	0,01506	0,31246
24	SRI	0,991	0,00382	0,24256	74	COA3	0,895	0,01506	0,31246
25	TRAP1	-1,206	0,00410	0,24256	75	PURB	0,761	0,01517	0,31246
26	CHTF8	1,128	0,00410	0,24256	76	UBL4A	0,836	0,01525	0,31246
27	DNAJB2	0,880	0,00419	0,24256	77	TACSTD2	1,091	0,01538	0,31246
28	LGALS1	-1,272	0,00427	0,24256	78	EIF3A	-1,016	0,01542	0,31246
29	DCD	1,175	0,00442	0,24256	79	GSS	0,580	0,01551	0,31246
30	PPP1R2;PPP1R2P3	0,882	0,00476	0,25226	80	TMA7	0,914	0,01574	0,31306
31	RPS14	0,786	0,00521	0,25286	81	UQCRC2	-0,980	0,01627	0,31935
32	TUBA1A	-0,851	0,00523	0,25286	82	RPS12	0,657	0,01698	0,31935
33	SON	1,036	0,00524	0,25286	83	AHCY	-1,525	0,01707	0,31935
34	GIPC1	0,728	0,00558	0,26089	84	EIF4A1	-1,192	0,01714	0,31935
35	ACTN4	-0,841	0,00574	0,26094	85	ACY1;ABHD14A-ACY1	0,841	0,01733	0,31935
36	MTHFD1	-0,875	0,00591	0,26113	86	SLTM	0,642	0,01767	0,31935
37	ISOC1	-1,435	0,00620	0,26649	87	EIF1AX;EIF1AY	0,658	0,01767	0,31935
38	RAB11B;RAB11A	0,555	0,00697	0,28391	88	MAN2B1	0,863	0,01769	0,31935
39	TUBB4B	-0,890	0,00705	0,28391	89	SUMF2	0,964	0,01793	0,31935
40	EML4	1,183	0,00786	0,28391	90	PTGFRN	0,919	0,01807	0,31935
41	RPS18	0,621	0,00798	0,28391	91	TMED1	0,598	0,01864	0,32094
42	FARSA	0,584	0,00824	0,28391	92	PSMC1	-0,759	0,01954	0,32094
43	ACTN1	-0,930	0,00831	0,28391	93	CTSZ	0,879	0,01971	0,32094
44	ADRM1	0,762	0,00847	0,28391	94	ENAH	0,748	0,01973	0,32094
45	CSRP1	-1,147	0,00895	0,28391	95	TCEB2	0,636	0,01985	0,32094
46	COMT	1,149	0,00939	0,28391	96	SYNCRIP	-0,659	0,01987	0,32094
47	DDX1	-1,424	0,00953	0,28391	97	SNX12	0,704	0,01988	0,32094
48	GSTP1	-1,065	0,00959	0,28391	98	CCT3	-1,329	0,02005	0,32094
49	VAMP8	0,886	0,00988	0,28391	99	DCPS	0,607	0,02083	0,32094
50	SRSF9	0,889	0,00997	0,28391	100	SPTBN1	0,622	0,02100	0,32094

Comparison 6 – Grade 3 (13 REC vs 16 NOREC)									
Nº	**ID**	**logFC**	**P.Value**	**adj.P.Val**	**Nº**	**ID**	**logFC**	**P.Value**	**adj.P.Val**
1	FDPS	1,0263224	8,47E-05	0,1347342	51	MVP	-0,986584	0,0168408	0,4920893
2	BPNT1	0,9580204	0,0002632	0,1916182	52	RPL10	0,5997215	0,0169842	0,4920893
3	NUDT5	0,6597802	0,0004498	0,1916182	53	PRKDC	-1,01107	0,0174108	0,4920893
4	MYH9	-0,724792	0,0005197	0,1916182	54	HNRNPF	0,4233893	0,0174845	0,4920893
5	RPS17	0,6814994	0,0006022	0,1916182	55	LSP1	-0,888823	0,0176506	0,4920893
6	PMVK	0,8431966	0,0017585	0,4095223	56	HBD	-1,176177	0,0181077	0,4920893
7	GBP1	-1,44074	0,0022334	0,4095223	57	PSMD3	-1,05379	0,018464	0,4920893
8	SRSF2	0,7332947	0,0026582	0,4095223	58	SRSF3	0,570469	0,0186011	0,4920893
9	ABCE1	-0,847878	0,0029286	0,4095223	59	APOL2	-0,725569	0,0190388	0,4920893
10	MRPL14	0,6694854	0,0031503	0,4095223	60	SNRPG;SNRPGP15	0,5759946	0,0196139	0,4920893
11	SEPT11	-0,6753	0,0031748	0,4095223	61	RPL9	0,7525847	0,0196348	0,4920893
12	ILF2	0,6103077	0,0032468	0,4095223	62	IDH2	-0,512086	0,0201777	0,4920893
13	SNRPE	0,8670989	0,003863	0,4095223	63	SNRPD3	0,7955377	0,0202624	0,4920893
14	GSN	-0,451145	0,0038807	0,4095223	64	PRDX5	0,6252616	0,0206496	0,4920893
15	ARF1;ARF3	1,0327305	0,0042709	0,4095223	65	RPL23	0,630675	0,0207518	0,4920893
16	PRDX6	0,6253923	0,0042749	0,4095223	66	CD44	-0,705439	0,0209403	0,4920893
17	NES	-1,223074	0,0043758	0,4095223	67	ANXA1	-0,83274	0,0211413	0,4920893
18	PDCD6IP	-0,632373	0,00488	0,4313353	68	GOT1	0,4449537	0,0213671	0,4920893
19	UFL1	-0,534053	0,0055523	0,437998	69	SNRPD1	0,9272117	0,0214452	0,4920893
20	HIST1H4A	1,1831443	0,0056409	0,437998	70	NEDD8-MDP1;NEDD8	0,6288672	0,0216507	0,4920893
21	LSM4	0,5725896	0,006104	0,437998	71	RPS15A	0,6374247	0,0224042	0,5020427
22	TSTD1	1,0541529	0,0063158	0,437998	72	TLN1	-0,522975	0,0228649	0,5052513
23	TMEM43	-0,585391	0,0063318	0,437998	73	APOA4	-1,062466	0,0237338	0,5063862
24	RPSA	0,4708075	0,0068411	0,4384415	74	CORO1A	-0,752922	0,0238494	0,5063862
25	ELAVL1	0,4855251	0,0072562	0,4384415	75	CAPN2	-0,381138	0,0238711	0,5063862
26	CACYBP	0,6295066	0,0073008	0,4384415	76	CORO1C	-0,585307	0,0244251	0,511321
27	EVL	-1,053877	0,0075694	0,4384415	77	SRSF7	0,581134	0,0252165	0,5210325
28	DBI	0,8412677	0,0081051	0,4384415	78	DDB1	-0,695629	0,0264819	0,5401629
29	MAP4	-0,479471	0,0083984	0,4384415	79	PGM3	-0,489064	0,028053	0,5457861
30	DNM2	-0,609759	0,008414	0,4384415	80	EIF3C;EIF3CL	-0,406663	0,0283861	0,5457861
31	GIMAP4	-0,731265	0,0085974	0,4384415	81	CAST	-0,579094	0,0283927	0,5457861
32	TGM2	-0,812401	0,0092019	0,4384415	82	RAB8A	-0,511947	0,0288101	0,5457861
33	IQGAP1	-0,573713	0,0094485	0,4384415	83	VASP	-0,469779	0,0291057	0,5457861
34	NAP1L4	-0,368744	0,0095812	0,4384415	84	RAB14	0,5848522	0,0297946	0,5457861
35	PRDX3	0,6382797	0,0101185	0,4384415	85	HNMT	-0,585585	0,0298151	0,5457861
36	TNKS1BP1	-0,70974	0,0101675	0,4384415	86	ARHGDIA	0,381635	0,0298919	0,5457861
37	RPS14	0,5359252	0,0102922	0,4384415	87	LIMA1	-0,665504	0,0299226	0,5457861
38	GAA	0,7742713	0,0104823	0,4384415	88	BPGM	-0,752896	0,0304414	0,5457861
39	TPM4	-0,541485	0,0112449	0,4384415	89	RPS4X	0,4973194	0,0311052	0,5457861
40	USMG5	0,4990486	0,0112795	0,4384415	90	PIGR	2,0509639	0,0312437	0,5457861
41	SNRPD2	0,4376057	0,0112986	0,4384415	91	SNRNP70	0,4719267	0,0317854	0,5457861
42	IRF2BP2	0,4552992	0,0122784	0,4560239	92	ARPIN	-0,446228	0,0318407	0,5457861
43	POLR2H	0,59406	0,012325	0,4560239	93	LASP1	-0,412041	0,0319033	0,5457861
44	CAT	-0,559379	0,0135533	0,4900752	94	EIF3L	-0,636061	0,0325924	0,5516435
45	CNN3	-0,669286	0,0147006	0,4920893	95	MRE11A	-0,403301	0,0330655	0,5537608
46	SNRPF	0,4727083	0,0151711	0,4920893	96	GNAI2	-0,595151	0,0336233	0,5572355
47	SRSF10	0,5692502	0,0153542	0,4920893	97	PPP1R12A	-0,495423	0,0350295	0,5691978
48	MYO1C	-0,560202	0,0157049	0,4920893	98	ANXA7	0,4691097	0,035169	0,5691978
49	PCBD1	0,7280849	0,0161829	0,4920893	99	DCXR	0,7506043	0,0354183	0,5691978
50	SDHA	-0,864987	0,0165287	0,4920893	100	CLIC4	-0,475609	0,0358823	0,5708877

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				Comparison 7 – LVSI POSITIVE (10 REC vs 9 NOREC)					
1	MYH9	-1,111	1,17E-05	0,01855	51	TXNDC5	-0,858	0,00640	0,19953
2	ARHGEF2	1,163	8,60E-05	0,04637	52	PMVK	0,881	0,00676	0,20298
3	RBM8A	0,845	8,74E-05	0,04637	53	GGCT	0,748	0,00693	0,20298
4	SRP14	0,641	0,00032	0,08997	54	COL18A1	1,035	0,00694	0,20298
5	RBMX	1,056	0,00034	0,08997	55	PABPN1	0,599	0,00702	0,20298
6	TGM2	-1,344	0,00034	0,08997	56	NANS	-0,883	0,00751	0,20482
7	ANXA1	-1,534	0,00047	0,09147	57	CORO1A	-1,080	0,00760	0,20482
8	LSM14A	0,711	0,00051	0,09147	58	MYL12A;MYL12B	-1,241	0,00762	0,20482
9	SRSF1	0,926	0,00056	0,09147	59	S100A11	1,025	0,00770	0,20482
10	ARHGAP1	-0,642	0,00060	0,09147	60	CYB5R1	1,263	0,00776	0,20482
11	SRSF2	0,999	0,00066	0,09147	61	TIAL1	0,750	0,00785	0,20482
12	ATP2B4	0,879	0,00075	0,09147	62	TBCB	0,664	0,00830	0,21117
13	CAP1	-0,855	0,00083	0,09147	63	TOR1AIP1	0,604	0,00836	0,21117
14	YTHDF3	0,929	0,00084	0,09147	64	RBM12	0,527	0,00894	0,21898
15	P4HB	-0,890	0,00086	0,09147	65	ACO2	-0,688	0,00895	0,21898
16	HSPA9	-0,834	0,00099	0,09843	66	GNG12	0,961	0,00933	0,22484
17	SRSF3	0,980	0,00131	0,10903	67	CTBP2	0,942	0,00976	0,23183
18	LCP1	-1,267	0,00135	0,10903	68	FYTTD1	0,686	0,01010	0,23635
19	CPSF6	0,767	0,00138	0,10903	69	PURB	0,758	0,01032	0,23742
20	IQGAP1	-0,889	0,00142	0,10903	70	GBP1	-1,478	0,01045	0,23742
21	SNRNP70	0,823	0,00145	0,10903	71	ARHGDIB	-0,883	0,01112	0,24628
22	MYEF2	1,191	0,00151	0,10903	72	HNRNPA0	0,909	0,01115	0,24628
23	ACTN1	-0,994	0,00180	0,11900	73	SPIN1	0,507	0,01145	0,24904
24	TRA2B	1,547	0,00182	0,11900	74	DPYSL3	-1,102	0,01158	0,24904
25	CORO1C	-0,986	0,00187	0,11900	75	CBR1	-0,811	0,01187	0,25130
26	RBM4;RBM4B	0,789	0,00227	0,13553	76	MARCKSL1	0,967	0,01200	0,25130
27	ELAVL1	0,681	0,00230	0,13553	77	ACTG1	-0,628	0,01217	0,25138
28	CPSF7	1,035	0,00245	0,13940	78	STIP1	-0,528	0,01245	0,25138
29	MYO1C	-0,887	0,00264	0,14503	79	CCAR1	0,618	0,01248	0,25138
30	ACAT1	-1,229	0,00293	0,15506	80	TMPO	0,713	0,01272	0,25214
31	SRSF6	1,007	0,00305	0,15506	81	FAM3C	0,847	0,01315	0,25214
32	LGALS3BP	1,277	0,00319	0,15506	82	HDAC2	0,705	0,01322	0,25214
33	PNN	0,624	0,00332	0,15506	83	CSRP2	0,776	0,01337	0,25214
34	MFAP1	0,685	0,00335	0,15506	84	ANXA3	-1,049	0,01348	0,25214
35	MPO	-2,524	0,00341	0,15506	85	APEX1	0,491	0,01379	0,25214
36	SNRPD2	0,633	0,00379	0,16608	86	UBA1	-0,538	0,01388	0,25214
37	RBMXL1	0,733	0,00386	0,16608	87	RELA	0,590	0,01409	0,25214
38	ACTL6A	0,526	0,00407	0,16828	88	LSP1	-1,122	0,01414	0,25214
39	ACTN4	-0,812	0,00412	0,16828	89	NUCKS1	0,775	0,01415	0,25214
40	TSTD1	1,351	0,00429	0,17030	90	RAB5B	0,656	0,01426	0,25214
41	LMNB1	0,910	0,00449	0,17030	91	SORBS2	0,881	0,01461	0,25547
42	MZB1	-2,068	0,00461	0,17030	92	NHP2L1	0,575	0,01490	0,25763
43	SRSF7	0,892	0,00464	0,17030	93	SNRPF	0,604	0,01520	0,26008
44	DPP7	1,329	0,00471	0,17030	94	SAP18	0,617	0,01584	0,26805
45	CNN3	-0,948	0,00504	0,17826	95	ERO1L	-0,872	0,01612	0,26885
46	LAP3	-0,951	0,00520	0,17983	96	RPS14	0,626	0,01635	0,26885
47	DCTN5	0,658	0,00558	0,18565	97	AGA	1,070	0,01639	0,26885
48	AGR2	-2,179	0,00560	0,18565	98	HSPD1	-0,804	0,01670	0,27114
49	TLN1	-0,753	0,00581	0,18871	99	SRSF10	0,696	0,01725	0,27725
50	LSM6	0,765	0,00622	0,19786	100	TIPRL	0,530	0,01756	0,27944

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				Comparison 8 – LVSI NEGATIVE (16 REC vs 21 NOREC)					
1	VIM	-0,906	0,00016	0,17717	51	ENAH	0,5624	0,01151	0,33631
2	TSTD1	1,277	0,00026	0,17717	52	RPL31	0,4984	0,01168	0,33631
3	PSMD3	-1,475	0,00033	0,17717	53	ADH5	-0,7889	0,01253	0,33631
4	RPS14	0,668	0,00056	0,22227	54	NME1-NME2;NME2;NME1	0,5561	0,01292	0,33631
5	BOLA2;BOLA2B	0,730	0,00099	0,28559	55	SEC16A	0,5937	0,01305	0,33631
6	SNRPF	0,575	0,00164	0,28559	56	EEF1D	0,3728	0,01326	0,33631
7	DDX1	-1,153	0,00179	0,28559	57	TRIM28	-0,3586	0,01350	0,33631
8	SDHA	-0,930	0,00192	0,28559	58	ILK	-0,5403	0,01351	0,33631
9	CLIC4	-0,630	0,00193	0,28559	59	RPS20	0,4821	0,01355	0,33631
10	EIF3L	-0,808	0,00198	0,28559	60	ACTN4	-0,4978	0,01391	0,33631
11	MYH9	-0,534	0,00229	0,28559	61	BAG2	-0,7740	0,01394	0,33631
12	PSMC5	-0,782	0,00231	0,28559	62	NENF	0,6818	0,01442	0,33631
13	PIGR	2,747	0,00255	0,28559	63	RPL11	0,3528	0,01454	0,33631
14	DDX19A	-0,535	0,00279	0,28559	64	SUMF2	0,6414	0,01459	0,33631
15	POLR2H	0,638	0,00304	0,28559	65	FARSA	0,3617	0,01468	0,33631
16	MRPL14	0,596	0,00309	0,28559	66	ZNF326	-0,5506	0,01487	0,33631
17	UFL1	-0,511	0,00321	0,28559	67	OXSR1	-0,3676	0,01510	0,33631
18	TMEM43	-0,544	0,00334	0,28559	68	IDH3A	-0,3877	0,01517	0,33631
19	GAA	0,800	0,00357	0,28559	69	ACBD3	0,4169	0,01538	0,33631
20	PGM5	-1,163	0,00372	0,28559	70	SERPINH1	-0,6487	0,01541	0,33631
21	EIF3H	-0,713	0,00383	0,28559	71	DBI	0,6667	0,01567	0,33631
22	GRHPR	-0,425	0,00427	0,28559	72	IDH2	-0,4631	0,01569	0,33631
23	MIEN1	0,711	0,00428	0,28559	73	PYGL	-0,7509	0,01572	0,33631
24	NUCKS1	0,658	0,00431	0,28559	74	SLC25A24	-0,4978	0,01575	0,33631
25	ASAH1	0,778	0,00456	0,29046	75	SERF2	0,4602	0,01585	0,33631
26	RPS17	0,504	0,00478	0,29271	76	PPIA	0,5797	0,01646	0,34467
27	C9orf142	0,609	0,00527	0,30055	77	EIF3C;EIF3CL	-0,3952	0,01700	0,34504
28	MTHFD1	-0,593	0,00529	0,30055	78	RPS10	0,6709	0,01714	0,34504
29	DDB1	-0,790	0,00563	0,30310	79	RPL8	0,6901	0,01733	0,34504
30	SUCLA2	-0,645	0,00593	0,30310	80	ABCE1	-0,6037	0,01738	0,34504
31	VAMP8	0,645	0,00604	0,30310	81	TLN1	-0,4632	0,01784	0,34504
32	LGALS1	-0,822	0,00615	0,30310	82	GNAI2	-0,5920	0,01792	0,34504
33	MRPS21	0,583	0,00644	0,30310	83	UPF1	-0,6268	0,01800	0,34504
34	DNM2	-0,570	0,00652	0,30310	84	COPS8	0,4997	0,01875	0,35515
35	CP	0,966	0,00667	0,30310	85	COMT	0,7065	0,01910	0,35521
36	SUB1	0,606	0,00769	0,33592	86	CSE1L	-0,6583	0,01920	0,35521
37	SDF4	0,662	0,00788	0,33592	87	LAD1	0,8420	0,01952	0,35686
38	SPR	0,808	0,00802	0,33592	88	PMVK	0,5427	0,01978	0,35686
39	ADD1	-0,446	0,00825	0,33631	89	RPL22	0,6227	0,01996	0,35686
40	VCL	-0,557	0,00898	0,33631	90	RPS28	0,9602	0,02040	0,35710
41	C21orf33	-0,567	0,00909	0,33631	91	SERPINA3	0,8174	0,02043	0,35710
42	NAP1L4	-0,331	0,00912	0,33631	92	PSMC1	-0,5170	0,02094	0,35859
43	GSS	0,419	0,00923	0,33631	93	PRKDC	-0,8677	0,02100	0,35859
44	RPS18	0,414	0,01021	0,33631	94	RPL18	0,7130	0,02119	0,35859
45	TXN	0,628	0,01043	0,33631	95	OTUB1	-0,4899	0,02178	0,35887
46	KARS	-0,358	0,01058	0,33631	96	GUK1	0,4416	0,02196	0,35887
47	PURB	0,545	0,01066	0,33631	97	BPNT1	0,5459	0,02242	0,35887
48	RPS12	0,450	0,01119	0,33631	98	SNX12	0,4661	0,02290	0,35887
49	OXCT1	-0,786	0,01142	0,33631	99	RPL19	0,6167	0,02294	0,35887
50	MYDGF	0,839	0,01142	0,33631	100	PRPSAP2	-0,3897	0,02300	0,35887

243

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
	Comparison 9 – High-Intermediate Risk (7 REC vs 11 NOREC)								
1	ABCE1	-1,169	0,00046	0,45127	51	DDB1	-0,918	0,02171	0,65376
2	DENR	-0,857	0,00059	0,45127	52	PSMD12	-0,760	0,02179	0,65376
3	DBI	1,281	0,00085	0,45127	53	EIF3C;EIF3CL	-0,524	0,02195	0,65376
4	SOD3	1,929	0,00150	0,59512	54	ACTR1A	-0,649	0,02219	0,65376
5	NAP1L4	-0,571	0,00207	0,59598	55	RAB14	0,794	0,02351	0,66483
6	RPS17	0,674	0,00463	0,59598	56	MTHFD1	-0,691	0,02409	0,66483
7	PGM5	-1,591	0,00534	0,59598	57	DARS	-1,180	0,02444	0,66483
8	PAICS	-0,951	0,00537	0,59598	58	SUB1	0,717	0,02473	0,66483
9	HBD	-1,758	0,00603	0,59598	59	DBN1	-0,742	0,02519	0,66483
10	PTPN11	-0,646	0,00624	0,59598	60	MAPRE1	-0,478	0,02540	0,66483
11	RPA1	-1,131	0,00659	0,59598	61	SDF4	0,807	0,02569	0,66483
12	MYH9	-0,695	0,00705	0,59598	62	FKBP10	-1,048	0,02591	0,66483
13	EIF3B	-0,602	0,00714	0,59598	63	RPS28	1,300	0,02699	0,66518
14	TSTD1	1,314	0,00723	0,59598	64	GAA	0,819	0,02702	0,66518
15	BPGM	-1,219	0,00760	0,59598	65	EIF4G1	-0,555	0,02718	0,66518
16	FXR1	-0,725	0,00771	0,59598	66	EIF3L	-0,821	0,02824	0,67107
17	PRPSAP2	-0,662	0,00787	0,59598	67	CBX5	-0,577	0,02826	0,67107
18	COPB2	-0,825	0,00792	0,59598	68	VIM	-0,726	0,02972	0,68904
19	G3BP2	-0,560	0,00797	0,59598	69	SRSF2	0,669	0,03026	0,68904
20	PPP5C	-0,685	0,00838	0,59598	70	OLA1	-0,470	0,03037	0,68904
21	LAD1	1,400	0,00861	0,59598	71	TIMM13	0,887	0,03173	0,68904
22	ZNF326	-0,828	0,00862	0,59598	72	QDPR	0,615	0,03232	0,68904
23	DDX19A	-0,687	0,00870	0,59598	73	NUDT21	0,484	0,03256	0,68904
24	C9orf142	0,843	0,00930	0,59598	74	MRPL14	0,564	0,03323	0,68904
25	SERPINB6	0,872	0,00979	0,59598	75	OXSR1	-0,454	0,03358	0,68904
26	PSMD3	-1,439	0,01001	0,59598	76	MUC1	1,301	0,03388	0,68904
27	EIF5	-0,633	0,01011	0,59598	77	GPS1	-0,540	0,03409	0,68904
28	EIF4G2	-0,654	0,01190	0,65376	78	PDIA6	0,575	0,03464	0,68904
29	COL6A2	1,152	0,01259	0,65376	79	SPCS2	-0,667	0,03479	0,68904
30	ATP5H	-0,562	0,01393	0,65376	80	PFKP	-0,872	0,03504	0,68904
31	LSM4	0,611	0,01426	0,65376	81	NEDD8-MDP1;NEDD8	0,713	0,03508	0,68904
32	SEPT11	-0,710	0,01448	0,65376	82	HIST1H4A	1,159	0,03560	0,69064
33	COPS8	0,751	0,01464	0,65376	83	CASP3	-0,434	0,03646	0,69886
34	SDHA	-1,112	0,01531	0,65376	84	CORO7-PAM16;CORO7	-0,539	0,03973	0,73165
35	NANS	0,717	0,01552	0,65376	85	BPNT1	0,688	0,03994	0,73165
36	THRAP3	0,885	0,01568	0,65376	86	PPP2R2A	-0,487	0,04017	0,73165
37	DDX3X	-0,774	0,01594	0,65376	87	PSMD11	-0,527	0,04120	0,73165
38	DNAJA1	-0,740	0,01603	0,65376	88	NDUFV2	0,735	0,04131	0,73165
39	PRDX3	0,754	0,01604	0,65376	89	AKR1B1	-0,849	0,04143	0,73165
40	CTTN	-0,789	0,01715	0,65376	90	USMG5	0,479	0,04206	0,73165
41	PALM	-0,911	0,01799	0,65376	91	EIF2S1	-0,585	0,04227	0,73165
42	PIGR	3,007	0,01841	0,65376	92	YTHDF2	-0,557	0,04231	0,73165
43	ILF2	0,628	0,01841	0,65376	93	GNAI2	-0,703	0,04283	0,73265
44	PPIA	0,814	0,01978	0,65376	94	HDLBP	-0,591	0,04434	0,74020
45	GLUD1	0,760	0,01999	0,65376	95	HNRNPD	-0,476	0,04450	0,74020
46	PSMC5	-0,857	0,02027	0,65376	96	CLTC	-0,937	0,04499	0,74020
47	ARF1;ARF3	1,073	0,02067	0,65376	97	DDX42	-0,392	0,04514	0,74020
48	LIN7C	0,555	0,02112	0,65376	98	SEPT9	-0,419	0,04610	0,74020
49	MVP	-1,161	0,02134	0,65376	99	RAN	-0,489	0,04634	0,74020
50	IDH2	-0,642	0,02153	0,65376	100	SERPING1	0,911	0,04652	0,74020

Annexes

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
Comparison 10 – High Risk (14 REC vs 15 NO REC)
1	ENAH	1,098	6,57E-06	0,01046	51	ATP2B4	0,646	0,00727	0,22667
2	DPP7	1,603	6,17E-05	0,04904	52	COL18A1	0,852	0,00759	0,23226
3	TSTD1	1,473	0,00020	0,09020	53	CLIC4	-0,611	0,00824	0,24667
4	SNRPF	0,740	0,00026	0,09020	54	TSTA3	0,589	0,00849	0,24667
5	TGM2	-1,124	0,00034	0,09020	55	LIN7C	0,501	0,00853	0,24667
6	GAA	1,071	0,00039	0,09020	56	CORO1C	-0,662	0,00894	0,25033
7	MYH9	-0,728	0,00046	0,09020	57	MSN	-0,647	0,00897	0,25033
8	PMVK	0,946	0,00046	0,09020	58	ARHGAP1	-0,403	0,00921	0,25117
9	POLR2H	0,811	0,00054	0,09020	59	TMEM43	-0,573	0,00933	0,25117
10	NUDT5	0,649	0,00057	0,09020	60	IQGAP1	-0,576	0,00947	0,25117
11	SCPEP1	0,834	0,00075	0,10807	61	CHMP6	0,511	0,00968	0,25241
12	MLF2	0,959	0,00094	0,11932	62	LCP1	-0,837	0,00995	0,25273
13	EVL	-1,268	0,00097	0,11932	63	ACTR2	-0,790	0,01001	0,25273
14	RPS14	0,686	0,00113	0,12314	64	ANXA6	-1,274	0,01018	0,25297
15	IRF2BP2	0,551	0,00116	0,12314	65	GSTM2	0,952	0,01038	0,25411
16	MRPL14	0,682	0,00141	0,14054	66	SON	0,699	0,01142	0,26715
17	PURB	0,740	0,00163	0,14875	67	OGDH	-0,560	0,01170	0,26715
18	GBP1	-1,387	0,00168	0,14875	68	UFL1	-0,497	0,01171	0,26715
19	GUK1	0,675	0,00193	0,16165	69	LAP3	-0,718	0,01209	0,26715
20	ALDH2	-0,950	0,00215	0,16777	70	ORM1	1,094	0,01211	0,26715
21	SGSH	0,811	0,00221	0,16777	71	ACTN4	-0,564	0,01224	0,26715
22	MRPS21	0,714	0,00243	0,17588	72	KARS	-0,404	0,01241	0,26715
23	S100A4	-1,425	0,00264	0,18274	73	SORBS2	0,787	0,01262	0,26715
24	NANS	-0,691	0,00340	0,21168	74	TPM4	-0,534	0,01270	0,26715
25	SRSF3	0,638	0,00352	0,21168	75	DNM2	-0,578	0,01286	0,26715
26	RPS18	0,492	0,00362	0,21168	76	MIF	0,711	0,01343	0,26715
27	PRKDC	-1,217	0,00370	0,21168	77	CHTOP	0,597	0,01359	0,26715
28	TLN1	-0,640	0,00373	0,21168	78	LGALS1	-0,834	0,01393	0,26715
29	MAP1B	1,646	0,00418	0,21348	79	RCN3	-0,857	0,01411	0,26715
30	ILK	-0,704	0,00419	0,21348	80	POLR2C	0,408	0,01437	0,26715
31	SRSF2	0,704	0,00431	0,21348	81	AGA	0,871	0,01458	0,26715
32	GGCT	0,677	0,00435	0,21348	82	NDUFB10	-0,401	0,01463	0,26715
33	SIAE	0,857	0,00448	0,21348	83	LRPPRC	-0,949	0,01471	0,26715
34	CLU	1,557	0,00456	0,21348	84	HLA-DRA	-0,842	0,01478	0,26715
35	SRSF9	0,712	0,00489	0,21555	85	YTHDF3	0,547	0,01487	0,26715
36	LSM5	0,806	0,00514	0,21555	86	PIGR	2,445	0,01499	0,26715
37	VIM	-0,744	0,00521	0,21555	87	IDH3A	-0,430	0,01503	0,26715
38	DYNLT1	0,589	0,00530	0,21555	88	LGALS3BP	0,840	0,01507	0,26715
39	SRP14	0,413	0,00551	0,21555	89	RPL8	0,775	0,01525	0,26715
40	ACTN1	-0,714	0,00554	0,21555	90	HAX1	0,566	0,01537	0,26715
41	ARHGDIB	-0,821	0,00555	0,21555	91	MTHFD1	-0,585	0,01550	0,26715
42	GNS	0,643	0,00616	0,22217	92	ARHGEF2	0,595	0,01553	0,26715
43	RPL31	0,578	0,00640	0,22217	93	CAP1	-0,501	0,01611	0,26715
44	SLTM	0,546	0,00641	0,22217	94	PCBD1	0,718	0,01616	0,26715
45	UBAP2L	0,585	0,00651	0,22217	95	LSM8	0,620	0,01629	0,26715
46	CORO1A	-0,920	0,00655	0,22217	96	GUSB	0,676	0,01632	0,26715
47	CNN3	-0,752	0,00656	0,22217	97	PTGFRN	0,678	0,01635	0,26715
48	TBCA	0,444	0,00693	0,22667	98	BPNT1	0,636	0,01650	0,26715
49	RPL22	0,781	0,00701	0,22667	99	HDDC2	0,637	0,01662	0,26715
50	NUCKS1	0,707	0,00722	0,22667	100	CP	0,921	0,01707	0,27156

Comparison 11 – Time to Recurrence (17 Early vs 10 Late)
Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
1	SORBS2	1,415	2,12E-06	0,00338	51	PURB	0,627	0,00866	0,25618
2	TOM1	0,751	0,00011	0,05391	52	LRPPRC	-1,035	0,00866	0,25618
3	DDI2	0,893	0,00015	0,05391	53	PSME1	-0,882	0,00878	0,25618
4	NDUFB4	0,980	0,00016	0,05391	54	COX5B	0,632	0,00878	0,25618
5	AGA	1,334	0,00019	0,05391	55	HMGB2	-0,742	0,00886	0,25618
6	TJP1	0,729	0,00023	0,05391	56	ARHGDIB	-0,776	0,00948	0,25708
7	SUB1	0,929	0,00024	0,05391	57	SRSF11	0,604	0,00949	0,25708
8	DCD	1,172	0,00033	0,06608	58	GLA	0,632	0,01006	0,25708
9	DSG2	1,214	0,00052	0,09113	59	CTNNB1	0,780	0,01023	0,25708
10	FNTA	0,590	0,00067	0,10735	60	DAD1	0,834	0,01053	0,25708
11	ABI1	0,683	0,00080	0,11483	61	HIBADH	0,760	0,01054	0,25708
12	MSN	-0,828	0,00087	0,11483	62	TMED4	0,529	0,01061	0,25708
13	NDRG2	0,933	0,00124	0,15190	63	FBP1	0,994	0,01063	0,25708
14	NDUFB9	0,729	0,00143	0,15346	64	PSMC4	-0,528	0,01070	0,25708
15	DDAH1	1,069	0,00145	0,15346	65	MRPL11	0,593	0,01089	0,25708
16	SSB	-0,637	0,00173	0,17213	66	EFHD2	-0,550	0,01101	0,25708
17	COX6C	0,685	0,00189	0,17315	67	RAB4A	0,593	0,01109	0,25708
18	CLPP	0,742	0,00196	0,17315	68	CRK	0,408	0,01113	0,25708
19	SSR4	0,671	0,00211	0,17315	69	GUK1	0,554	0,01121	0,25708
20	RBM3	0,726	0,00218	0,17315	70	NDUFA13	0,534	0,01153	0,25708
21	TSTD1	1,157	0,00258	0,19529	71	MYDGF	0,972	0,01183	0,25708
22	SEC16A	0,786	0,00302	0,21654	72	NDUFA7	0,688	0,01183	0,25708
23	EPN1	0,580	0,00313	0,21654	73	HAX1	0,597	0,01186	0,25708
24	CST3	1,219	0,00337	0,21980	74	SERPINH1	-0,784	0,01197	0,25708
25	DNM1L	-0,725	0,00354	0,21980	75	DDRGK1	0,555	0,01212	0,25708
26	ANXA6	-1,446	0,00389	0,21980	76	RALA	0,470	0,01268	0,25847
27	LSM8	0,755	0,00390	0,21980	77	ADSS	0,480	0,01274	0,25847
28	CLIC4	-0,658	0,00397	0,21980	78	HSPD1	-0,702	0,01281	0,25847
29	GPX4	0,682	0,00401	0,21980	79	RAB18	0,534	0,01289	0,25847
30	SEC61B	0,794	0,00431	0,22864	80	TMED1	0,487	0,01313	0,25847
31	MRPL17	0,657	0,00461	0,23663	81	ARPC3	-0,474	0,01316	0,25847
32	RPLP0;RPLP0P6	-0,750	0,00508	0,25089	82	NENF	0,779	0,01355	0,26199
33	RAB3D	0,868	0,00520	0,25089	83	NDUFA2	0,696	0,01367	0,26199
34	BRK1	0,699	0,00631	0,25618	84	HSPA4	-0,709	0,01424	0,26638
35	LSM3	1,049	0,00639	0,25618	85	PKP2	0,878	0,01432	0,26638
36	RAB21	0,603	0,00682	0,25618	86	PSMD13	-0,657	0,01440	0,26638
37	MIEN1	0,747	0,00684	0,25618	87	TUBB3	-1,366	0,01531	0,27082
38	HSPG2	1,109	0,00684	0,25618	88	MDH1	-0,452	0,01556	0,27082
39	SON	0,771	0,00690	0,25618	89	FIP1L1	0,516	0,01561	0,27082
40	ST13;ST13P5	-0,739	0,00695	0,25618	90	NUP35	0,601	0,01570	0,27082
41	TF	-1,039	0,00720	0,25618	91	UQCRQ	0,483	0,01579	0,27082
42	DDX1	-1,101	0,00722	0,25618	92	HSPH1	-0,622	0,01613	0,27082
43	UBL4A	0,710	0,00729	0,25618	93	FGG	-1,153	0,01618	0,27082
44	ASRGL1	1,515	0,00755	0,25618	94	LSM2	0,718	0,01628	0,27082
45	CAD	0,702	0,00765	0,25618	95	RPL31	0,530	0,01631	0,27082
46	PITPNA	0,544	0,00826	0,25618	96	RPS21	0,560	0,01650	0,27082
47	MRPL14	0,593	0,00826	0,25618	97	DSP	0,860	0,01651	0,27082
48	HARS	-0,534	0,00831	0,25618	98	GAA	0,745	0,01680	0,27102
49	H2AFX	-1,329	0,00838	0,25618	99	ARCN1	0,683	0,01686	0,27102
50	PTK7	0,508	0,00865	0,25618	100	TNKS1BP1	0,711	0,01726	0,27460

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
Comparison 12 – Time to Recurrence (17 Early vs 31 NOREC)								
1	MYH9	-0,624	0,00016	0,25714	51	ACTN1	-0,439	0,02830	0,77095
2	TSTD1	0,964	0,00099	0,57170	52	GSTO1	0,421	0,02845	0,77095
3	UFL1	-0,496	0,00158	0,57170	53	SRSF1	0,402	0,02909	0,77095
4	PIGR	2,538	0,00177	0,57170	54	ACAT1	-0,582	0,03007	0,77095
5	SNRPF	0,494	0,00211	0,57170	55	RPS18	0,300	0,03047	0,77095
6	IDH2	-0,535	0,00241	0,57170	56	PURB	0,387	0,03052	0,77095
7	RPS14	0,495	0,00252	0,57170	57	LIN7C	0,320	0,03098	0,77095
8	NUCKS1	0,590	0,00451	0,77095	58	CHTOP	0,412	0,03222	0,77095
9	PRKDC	-0,954	0,00527	0,77095	59	SF3A3	0,342	0,03248	0,77095
10	TOM1	-0,394	0,00558	0,77095	60	RPL22	0,509	0,03264	0,77095
11	EVL	-0,861	0,00639	0,77095	61	MTHFD1	-0,390	0,03407	0,77095
12	SRSF2	0,526	0,00712	0,77095	62	CLIC4	-0,360	0,03443	0,77095
13	ARHGAP1	-0,334	0,00734	0,77095	63	RBM12	0,282	0,03561	0,77095
14	SDHA	-0,770	0,00803	0,77095	64	SEPT11	-0,391	0,03620	0,77095
15	GBP1	-0,990	0,00841	0,77095	65	GUSB	0,472	0,03646	0,77095
16	S100A11	0,656	0,00892	0,77095	66	TPM4	-0,348	0,03655	0,77095
17	ACO2	-0,458	0,00939	0,77095	67	MRPS21	0,384	0,03688	0,77095
18	MYO6	-0,495	0,00940	0,77095	68	TCEB2	0,337	0,03688	0,77095
19	POLR2H	0,482	0,01012	0,77095	69	CLU	0,908	0,03757	0,77095
20	PMVK	0,527	0,01049	0,77095	70	HDLBP	-0,385	0,03785	0,77095
21	TGM2	-0,630	0,01056	0,77095	71	ETFA	-0,318	0,03919	0,77095
22	UFC1	0,606	0,01067	0,77095	72	ELAVL1	0,299	0,03950	0,77095
23	ACTN4	-0,442	0,01168	0,77095	73	PPP2R2A	-0,317	0,04094	0,77095
24	VAMP8	0,528	0,01244	0,77095	74	RPL17	0,362	0,04147	0,77095
25	OXCT1	-0,691	0,01367	0,77095	75	SUCLA2	-0,428	0,04154	0,77095
26	DPP7	0,760	0,01367	0,77095	76	ILK	-0,385	0,04161	0,77095
27	GOT1	0,376	0,01610	0,77095	77	CALD1	-0,590	0,04397	0,77095
28	IQGAP1	-0,437	0,01611	0,77095	78	TCOF1	0,469	0,04409	0,77095
29	KARS	-0,311	0,01663	0,77095	79	GSN	-0,246	0,04418	0,77095
30	BPNT1	0,511	0,01761	0,77095	80	PAPSS1	-0,562	0,04589	0,77095
31	NANS	-0,507	0,01807	0,77095	81	DBI	0,506	0,04597	0,77095
32	LGALS3BP	0,665	0,01849	0,77095	82	UQCRC1	-0,398	0,04681	0,77095
33	LSM4	0,387	0,01876	0,77095	83	ANXA1	-0,549	0,04704	0,77095
34	VIM	-0,510	0,01952	0,77095	84	CBR1	-0,409	0,04753	0,77095
35	FDPS	0,504	0,01965	0,77095	85	DDB1	-0,510	0,04764	0,77095
36	SRP14	0,286	0,02011	0,77095	86	GUK1	0,323	0,04796	0,77095
37	ETF1	-0,352	0,02046	0,77095	87	HIBADH	-0,440	0,04831	0,77095
38	ENAH	0,453	0,02109	0,77095	88	SRSF3	0,391	0,04877	0,77095
39	DNM2	-0,423	0,02284	0,77095	89	ASNA1	-0,280	0,04966	0,77095
40	KRT18	-0,679	0,02310	0,77095	90	RPL8	0,500	0,05059	0,77095
41	TLN1	-0,397	0,02324	0,77095	91	PGM5	-0,711	0,05105	0,77095
42	COL18A1	0,552	0,02352	0,77095	92	EIF3L	-0,467	0,05262	0,77095
43	CNN3	-0,479	0,02394	0,77095	93	HADHA	-0,628	0,05354	0,77095
44	TMEM43	-0,393	0,02398	0,77095	94	OGDH	-0,359	0,05383	0,77095
45	SIAE	0,539	0,02404	0,77095	95	UBE2L3	0,286	0,05420	0,77095
46	LIMA1	-0,544	0,02460	0,77095	96	APOA4	-0,720	0,05547	0,77095
47	GAA	0,529	0,02464	0,77095	97	NDUFB9	-0,322	0,05580	0,77095
48	SRSF7	0,443	0,02701	0,77095	98	NUDT5	0,299	0,05585	0,77095
49	RPS17	0,341	0,02731	0,77095	99	NAP1L4	-0,215	0,05643	0,77095
50	PSMD3	-0,777	0,02765	0,77095	100	RBMX	0,368	0,05645	0,77095

Comparison 13 – Time to Recurrence (10 Late vs 31 NOREC)									
Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
1	TSTD1	2,121	3,43E-08	5,46E-05	51	PSMC5	-0,987	0,00041	0,01259
2	SORBS2	1,389	4,78E-07	0,00038	52	RPS28	1,560	0,00042	0,01259
3	CLIC4	-1,018	3,88E-06	0,00158	53	BRK1	0,843	0,00041	0,01259
4	AGA	1,550	3,96E-06	0,00158	54	SON	0,935	0,00044	0,01288
5	RPS14	0,934	6,41E-06	0,00204	55	U2AF2	0,649	0,00045	0,01288
6	PURB	1,014	1,03E-05	0,00234	56	PSMD3	-1,526	0,00050	0,01434
7	MSN	-1,046	8,95E-06	0,00234	57	RPS18	0,595	0,00055	0,01527
8	MYH9	-0,867	1,91E-05	0,00291	58	ACBD3	0,631	0,00058	0,01592
9	GAA	1,274	2,14E-05	0,00291	59	LSM3	1,226	0,00060	0,01601
10	SEC16A	1,068	2,20E-05	0,00291	60	TGM2	-1,043	0,00060	0,01601
11	SUB1	1,016	1,65E-05	0,00291	61	PSMD13	-0,859	0,00063	0,01631
12	SNRPF	0,858	1,99E-05	0,00291	62	NAP1L4	-0,478	0,00064	0,01650
13	GUK1	0,877	2,70E-05	0,00292	63	ANXA1	-1,167	0,00069	0,01743
14	MRPL14	0,904	2,50E-05	0,00292	64	RPS21	0,736	0,00074	0,01837
15	LSM8	1,041	2,75E-05	0,00292	65	PSMC4	-0,647	0,00077	0,01887
16	DSG2	1,354	3,31E-05	0,00329	66	NUCKS1	0,850	0,00080	0,01934
17	ACTN4	-0,910	3,90E-05	0,00365	67	ARHGDIB	-0,915	0,00099	0,02266
18	TLN1	-0,887	5,86E-05	0,00430	68	LGALS1	-1,065	0,00097	0,02266
19	MIEN1	1,065	4,87E-05	0,00430	69	SRSF9	0,844	0,00100	0,02266
20	ANXA6	-1,898	5,90E-05	0,00430	70	HSP90AB1	-0,814	0,00100	0,02266
21	NDUFB4	0,955	6,41E-05	0,00430	71	DUSP23	0,932	0,00107	0,02397
22	SEC61B	1,053	6,00E-05	0,00430	72	PSMC6	-0,661	0,00111	0,02423
23	DDX1	-1,551	6,48E-05	0,00430	73	ST13;ST13P5	-0,827	0,00111	0,02423
24	ABI1	0,766	5,54E-05	0,00430	74	SSB	-0,602	0,00122	0,02559
25	DPP7	1,495	0,00011	0,00672	75	CPSF6	0,615	0,00121	0,02559
26	SRSF2	0,939	0,00011	0,00673	76	PIGR	3,180	0,00119	0,02559
27	RPL31	0,807	0,00011	0,00673	77	FGG	-1,424	0,00140	0,02874
28	ACTN1	-0,967	0,00012	0,00675	78	TRAP1	-1,003	0,00141	0,02874
29	VCL	-0,871	0,00012	0,00675	79	SF3B4	0,620	0,00152	0,02893
30	LRPPRC	-1,424	0,00013	0,00684	80	ENAH	0,766	0,00151	0,02893
31	SRSF7	0,959	0,00014	0,00710	81	EFHD2	-0,636	0,00152	0,02893
32	HSPD1	-1,009	0,00015	0,00758	82	PSMC3	-0,473	0,00146	0,02893
33	COL18A1	1,143	0,00018	0,00768	83	RPS17	0,606	0,00147	0,02893
34	CNN3	-1,001	0,00016	0,00768	84	MLF2	0,931	0,00153	0,02893
35	CORO1C	-0,935	0,00017	0,00768	85	RSU1	-0,954	0,00164	0,03074
36	PMVK	0,962	0,00017	0,00768	86	CRK	0,466	0,00175	0,03157
37	ILK	-0,891	0,00018	0,00768	87	EIF3C;EIF3CL	-0,553	0,00174	0,03157
38	MTHFD1	-0,855	0,00021	0,00847	88	TJP1	0,557	0,00174	0,03157
39	SERPINH1	-1,090	0,00021	0,00847	89	GLA	0,711	0,00177	0,03157
40	DDAH1	1,148	0,00022	0,00890	90	RBM3	0,675	0,00182	0,03218
41	HAX1	0,821	0,00024	0,00921	91	LSM5	0,901	0,00188	0,03253
42	CST3	1,421	0,00024	0,00927	92	PSMD12	-0,772	0,00187	0,03253
43	LIN7C	0,678	0,00026	0,00948	93	GNG12	0,898	0,00193	0,03266
44	NENF	1,080	0,00027	0,00972	94	BAG2	-1,061	0,00193	0,03266
45	C9orf142	0,887	0,00028	0,00972	95	PSME1	-0,960	0,00200	0,03348
46	THRAP3	1,018	0,00028	0,00979	96	DCD	0,905	0,00205	0,03405
47	HSPA9	-0,721	0,00033	0,01080	97	MYDGF	1,092	0,00222	0,03560
48	POLR2H	0,833	0,00033	0,01080	98	STIP1	-0,503	0,00226	0,03560
49	MRPS21	0,825	0,00033	0,01084	99	SLTM	0,595	0,00222	0,03560
50	GNAI2	-0,943	0,00038	0,01212	100	HBD	-1,478	0,00224	0,03560

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
Comparison 14 – Recurrence Site (6 Regional vs 20 Distant)										
1	GLRX3	-0,854	0,00123	0,94726	51	PARP1	0,778	0,03712	0,98806
2	FYTTD1	-0,739	0,00260	0,94726	52	TSN	0,847	0,03725	0,98806
3	CBX5	0,737	0,00419	0,94726	53	LSP1	-0,888	0,03759	0,98806
4	RAP1B	-0,575	0,00469	0,94726	54	NDUFS1	0,674	0,03762	0,98806
5	ATG3	-0,643	0,00529	0,94726	55	CCDC124	-0,553	0,03767	0,98806
6	TTC38	0,688	0,00584	0,94726	56	GUSB	0,655	0,03936	0,98806
7	GSR	1,033	0,00587	0,94726	57	FKBP4	0,518	0,03989	0,98806
8	TUBB2A	-0,810	0,00637	0,94726	58	CDV3	-0,489	0,04268	0,98806
9	HSPB1	-0,956	0,00701	0,94726	59	SERPINA3	0,914	0,04353	0,98806
10	H2AFY2	1,163	0,00727	0,94726	60	ERGIC1	0,893	0,04397	0,98806
11	TGM2	-0,919	0,00749	0,94726	61	MRPS26	0,522	0,04477	0,98806
12	ACLY	1,234	0,00771	0,94726	62	LCP1	-0,768	0,04486	0,98806
13	TP53I3	-0,816	0,00774	0,94726	63	C12orf57	-0,562	0,04544	0,98806
14	PRKAR1A	-0,622	0,00867	0,98264	64	DDRGK1	-0,491	0,04605	0,98806
15	CHMP4B	-0,847	0,01156	0,98264	65	OSTF1	-0,559	0,04658	0,98806
16	BSG	-0,811	0,01175	0,98264	66	YAP1	-0,542	0,04739	0,98806
17	EFHD2	-0,593	0,01310	0,98264	67	ERP44	-0,470	0,04831	0,98806
18	PSMD9	-0,641	0,01311	0,98264	68	SMARCC2	0,569	0,04964	0,98806
19	ALDH6A1	1,206	0,01337	0,98264	69	GCA	-0,725	0,05064	0,98806
20	GYG1	-0,597	0,01365	0,98264	70	RAB18	-0,465	0,05220	0,98806
21	EVL	-1,090	0,01391	0,98264	71	SEC61B	-0,605	0,05255	0,98806
22	NDUFB10	-0,458	0,01576	0,98264	72	ASL	0,598	0,05382	0,98806
23	F11R	0,675	0,01673	0,98264	73	TWF1	0,568	0,05454	0,98806
24	GANAB	0,583	0,01686	0,98264	74	EIF4A2	0,658	0,05505	0,98806
25	EPHX1	1,460	0,01694	0,98264	75	FDPS	0,582	0,05559	0,98806
26	HSD17B4	1,108	0,01705	0,98264	76	CAP1	-0,452	0,05650	0,98806
27	PRDX3	0,689	0,01735	0,98264	77	FAM3C	0,639	0,05678	0,98806
28	PHYHD1	0,741	0,01818	0,98264	78	CAMP	-1,317	0,05836	0,98806
29	PIGR	2,622	0,01829	0,98264	79	SNX3	-0,459	0,05840	0,98806
30	CNN2	-0,820	0,01865	0,98264	80	QKI	-0,447	0,05852	0,98806
31	H2AFY	0,990	0,01915	0,98264	81	C21orf33	0,525	0,05954	0,98806
32	GAA	0,798	0,02037	0,98806	82	CLTC	0,798	0,06022	0,98806
33	NPM1	0,504	0,02377	0,98806	83	COMT	-0,716	0,06082	0,98806
34	CDC42	-0,517	0,02395	0,98806	84	LASP1	-0,423	0,06161	0,98806
35	TYMP	-0,947	0,02420	0,98806	85	SART3	0,406	0,06192	0,98806
36	HNRNPU	0,588	0,02467	0,98806	86	H1F0	0,953	0,06197	0,98806
37	CORO1A	-0,848	0,02518	0,98806	87	SFXN3	-0,462	0,06230	0,98806
38	BDH2	0,608	0,02577	0,98806	88	RPA1	0,694	0,06366	0,98806
39	GLIPR2	-0,709	0,02688	0,98806	89	SCRN1	-0,663	0,06404	0,98806
40	ABI1	-0,509	0,02736	0,98806	90	UFC1	0,613	0,06441	0,98806
41	SEPT11	0,577	0,02852	0,98806	91	HP1BP3	0,731	0,06520	0,98806
42	APOL2	-0,814	0,03065	0,98806	92	RARS	0,477	0,06605	0,98806
43	ECH1	0,689	0,03068	0,98806	93	PLD3	-0,540	0,06649	0,98806
44	UQCRC2	0,778	0,03081	0,98806	94	HCLS1	-0,782	0,06657	0,98806
45	TPM4	-0,509	0,03110	0,98806	95	CALR	-0,468	0,06673	0,98806
46	PSMD3	1,086	0,03150	0,98806	96	VBP1	0,409	0,06690	0,98806
47	CHMP4A	-0,670	0,03262	0,98806	97	HEXA	0,451	0,06722	0,98806
48	ARHGDIB	-0,706	0,03392	0,98806	98	MAT2B	0,393	0,06829	0,98806
49	ATP1A1	1,058	0,03512	0,98806	99	SNX12	-0,472	0,06967	0,98806
50	ECHS1	0,787	0,03632	0,98806	100	NAGK	-0,366	0,07130	0,98806

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				Comparison 15 – Status (31 NOREC vs 11 Death of Disease)					
1	FBL	1,584	8,94E-05	0,05680	51	HNRNPA0	1,279	0,00601	0,18753
2	CTBP2	1,785	0,00012	0,05680	52	RPL7	1,313	0,00614	0,18792
3	SNRPB;SNRPN	1,688	0,00013	0,05680	53	PUF60	1,000	0,00656	0,19261
4	MYEF2	1,775	0,00014	0,05680	54	TXN	-1,061	0,00671	0,19261
5	PTBP3	1,548	0,00027	0,08584	55	SNRNP70	0,937	0,00693	0,19261
6	RPL27	2,281	0,00037	0,08947	56	SNRPD1	1,727	0,00712	0,19261
7	ANXA1	-1,885	0,00039	0,08947	57	RPL7A	1,442	0,00721	0,19261
8	PRDX1	-0,841	0,00053	0,10161	58	TALDO1	-1,056	0,00725	0,19261
9	CTSD	-1,260	0,00057	0,10161	59	RPL30	1,314	0,00730	0,19261
10	RPL3	2,767	0,00065	0,10360	60	GNB2L1	1,079	0,00734	0,19261
11	MRTO4	1,024	0,00075	0,10870	61	RPS3	1,443	0,00738	0,19261
12	EIF3D	1,060	0,00101	0,13416	62	RAB5A	0,822	0,00809	0,19749
13	SBDS	1,630	0,00127	0,15313	63	HNRNPM	1,002	0,00812	0,19749
14	RPS11	1,839	0,00135	0,15313	64	NONO	1,165	0,00817	0,19749
15	RPS4X	1,135	0,00151	0,15329	65	TARDBP;TDP43	1,174	0,00818	0,19749
16	RAB6A	1,184	0,00154	0,15329	66	CDV3	-0,842	0,00819	0,19749
17	EIF4A3	1,920	0,00188	0,16694	67	NAP1L4	-0,594	0,00859	0,20386
18	CHP1	-1,302	0,00189	0,16694	68	CD44	-1,213	0,00875	0,20412
19	GMPPB	1,268	0,00205	0,17127	69	ASAH1	-1,033	0,00896	0,20412
20	CBR1	-1,209	0,00229	0,17555	70	MRPS26	0,937	0,00909	0,20412
21	RPL13A	1,654	0,00242	0,17555	71	RPS9	1,539	0,00911	0,20412
22	RPL34	1,841	0,00243	0,17555	72	RPS16	1,824	0,00945	0,20799
23	LUC7L2	1,018	0,00279	0,18292	73	RPS27	1,647	0,00954	0,20799
24	CTBP1	1,055	0,00293	0,18292	74	PHB	1,078	0,00983	0,21130
25	RPS7	1,082	0,00303	0,18292	75	SRSF6	1,118	0,01100	0,23334
26	SEPT8	1,041	0,00312	0,18292	76	RPS23	1,504	0,01134	0,23738
27	DRG1	0,789	0,00314	0,18292	77	RPL26	1,561	0,01183	0,24406
28	RPL18A	1,911	0,00331	0,18292	78	ERO1L	-1,208	0,01201	0,24406
29	TIPRL	0,857	0,00333	0,18292	79	CYB5B	-0,795	0,01216	0,24406
30	NIT1	0,987	0,00356	0,18343	80	LRRFIP1	-0,932	0,01263	0,24406
31	RPL19	1,397	0,00365	0,18343	81	TRA2B	1,629	0,01277	0,24406
32	RPL15	1,621	0,00397	0,18343	82	AGR2	-2,475	0,01279	0,24406
33	MZB1	-2,702	0,00398	0,18343	83	EIF3H	1,097	0,01282	0,24406
34	RPL35A	2,184	0,00402	0,18343	84	PDCD4	0,995	0,01289	0,24406
35	ADSL	0,875	0,00418	0,18343	85	TMSB4X	-1,186	0,01330	0,24891
36	CPSF7	1,271	0,00437	0,18343	86	RBM39	0,854	0,01346	0,24902
37	RPS14	0,950	0,00440	0,18343	87	RALY	1,023	0,01380	0,24911
38	HDAC2	1,047	0,00456	0,18343	88	PPCS	1,087	0,01391	0,24911
39	H2AFY2	1,507	0,00472	0,18343	89	RPL28	1,515	0,01407	0,24911
40	RPL8	1,443	0,00493	0,18343	90	S100A4	-1,746	0,01409	0,24911
41	RBMX	1,062	0,00493	0,18343	91	ARPC5	-1,064	0,01427	0,24948
42	RPS2	1,314	0,00501	0,18343	92	PLIN3	-0,825	0,01479	0,25581
43	HIST1H3A;HIST2H3A;HIST3H3	1,775	0,00528	0,18343	93	EEF1B2	-1,002	0,01510	0,25833
44	PMVK	1,100	0,00528	0,18343	94	ATP5H	-0,681	0,01551	0,26126
45	POLR2C	0,773	0,00535	0,18343	95	SNRPD2	0,673	0,01560	0,26126
46	RAB2A	1,271	0,00540	0,18343	96	GORASP2	-0,950	0,01590	0,26357
47	RPL13	1,488	0,00542	0,18343	97	CAPNS1	-0,990	0,01673	0,27109
48	RAB5B	0,983	0,00555	0,18384	98	SNX5	0,695	0,01684	0,27109
49	RPL21	1,230	0,00590	0,18753	99	SFPQ	0,989	0,01687	0,27109
50	FAM107B	-1,127	0,00593	0,18753	100	LCP1	-1,255	0,01719	0,27235

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
Comparison 16 – Status (8 Alive with Disease vs 11 Death of Disease)

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
1	ASAH1	-1,454	1,01E-05	0,01614	51	HSD17B10	0,588	0,03784	0,91209
2	BASP1	-1,474	0,00015	0,12155	52	EPN1	-0,479	0,03867	0,91209
3	S100A11	-1,084	0,00363	0,91209	53	SERPINA1	-1,034	0,03876	0,91209
4	PHB	0,924	0,00512	0,91209	54	CNN3	-0,638	0,03888	0,91209
5	CMPK1	-0,791	0,00557	0,91209	55	CHP1	-0,662	0,03952	0,91209
6	GGCT	-0,770	0,00619	0,91209	56	EIF3I	0,448	0,03954	0,91209
7	PPA2	-0,602	0,00640	0,91209	57	PPP1R7	-0,369	0,03975	0,91209
8	CTSS	-1,317	0,00725	0,91209	58	C1orf123	-0,533	0,04072	0,91209
9	ITIH2	1,523	0,00731	0,91209	59	AIFM1	1,151	0,04156	0,91209
10	FTH1	-0,998	0,00847	0,91209	60	HIBADH	0,724	0,04158	0,91209
11	RPSA	0,547	0,00937	0,91209	61	GTF2I	0,917	0,04251	0,91209
12	COPZ1	0,745	0,00942	0,91209	62	MUC1	-1,212	0,04262	0,91209
13	GPNMB	-1,176	0,01037	0,91209	63	GNB2L1	0,630	0,04343	0,91209
14	HEBP2	-0,906	0,01055	0,91209	64	YARS	0,564	0,04363	0,91209
15	SCP2	-0,694	0,01173	0,91209	65	RHOA	-0,584	0,04365	0,91209
16	MCM2	1,182	0,01268	0,91209	66	PTMA	-0,813	0,04463	0,91209
17	CSTB	-0,987	0,01302	0,91209	67	HINT1	-0,580	0,04496	0,91209
18	HK1	0,847	0,01346	0,91209	68	PSMA4	0,565	0,04497	0,91209
19	CTSD	-0,691	0,01364	0,91209	69	PRELP	-1,173	0,04546	0,91209
20	DPP7	-1,124	0,01567	0,91209	70	COPG1	0,672	0,04696	0,91209
21	DRG1	0,499	0,01608	0,91209	71	SCARB2	-0,525	0,04713	0,91209
22	GLO1	-0,804	0,01616	0,91209	72	LGALS3	-0,875	0,04827	0,91209
23	UQCRFS1;UQCRFS1P1	-0,857	0,01649	0,91209	73	PHB2	1,464	0,04920	0,91209
24	CD44	-0,863	0,01713	0,91209	74	MXRA7	-0,746	0,04942	0,91209
25	CLU	-1,524	0,01840	0,91209	75	CAT	-0,512	0,04979	0,91209
26	PITPNA	-0,580	0,01861	0,91209	76	MRPL40	0,493	0,05036	0,91209
27	TMSB4X	-0,876	0,01975	0,91209	77	CCT4	0,614	0,05136	0,91209
28	CTSZ	-0,809	0,02123	0,91209	78	ATP5A1	0,711	0,05144	0,91209
29	MYL1;MYL3	-0,951	0,02128	0,91209	79	EIF3D	0,478	0,05176	0,91209
30	SELENBP1	-1,194	0,02226	0,91209	80	ERO1L	-0,725	0,05254	0,91209
31	CTBP1	0,626	0,02256	0,91209	81	PGM1	-0,486	0,05308	0,91209
32	SERPING1	-1,019	0,02447	0,91209	82	S100A10	-0,914	0,05319	0,91209
33	GAA	-0,818	0,02530	0,91209	83	PSMB3	0,570	0,05405	0,91209
34	CCT2	0,593	0,02729	0,91209	84	TPP1	-0,662	0,05498	0,91209
35	SOD1	-0,691	0,02992	0,91209	85	PRKDC	0,949	0,05554	0,91209
36	PA2G4	0,508	0,03135	0,91209	86	ATP5C1	0,819	0,05561	0,91209
37	IDH1	-0,855	0,03188	0,91209	87	CAST	-0,584	0,05699	0,91209
38	DUSP23	-0,758	0,03195	0,91209	88	RPS6	0,591	0,05807	0,91209
39	MRE11A	0,501	0,03207	0,91209	89	OGDH	0,546	0,05817	0,91209
40	SFXN1	0,519	0,03313	0,91209	90	RPL10A	0,578	0,05892	0,91209
41	ENSA	-0,632	0,03367	0,91209	91	ANXA1	-0,758	0,05918	0,91209
42	CBR1	-0,647	0,03386	0,91209	92	PGAM1	-0,463	0,05966	0,91209
43	PIGR	-2,535	0,03427	0,91209	93	GLRX3	0,536	0,06097	0,91209
44	CTTN	-0,659	0,03484	0,91209	94	SRRT	0,483	0,06201	0,91209
45	TAGLN2	-0,631	0,03489	0,91209	95	SEPT8	0,502	0,06328	0,91209
46	TSNAX;DISC1	0,765	0,03511	0,91209	96	THYN1	-0,503	0,06525	0,91209
47	ASRGL1	-1,417	0,03586	0,91209	97	ARHGDIB	-0,654	0,06592	0,91209
48	CRYAB	-1,002	0,03627	0,91209	98	SEPT9	0,360	0,06604	0,91209
49	S100A4	-1,154	0,03750	0,91209	99	PSMD11	0,444	0,06794	0,91209
50	SND1	0,650	0,03762	0,91209	100	RPS24	0,656	0,06800	0,91209

Annexes

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
Comparison 17 – Status (31 NOREC vs 8 Alive with Disease)								
1	MYEF2	1,897	0,00011	0,17280	51	ANXA7	0,769	0,03286	0,87896
2	FBL	1,377	0,00091	0,54083	52	APEX1	0,580	0,03321	0,87896
3	SNRPB;SNRPN	1,434	0,00150	0,54083	53	RPL3	1,748	0,03379	0,87896
4	H2AFY2	1,775	0,00164	0,54083	54	SMEK1	0,666	0,03489	0,87896
5	RPL27	2,081	0,00170	0,54083	55	HIST1H3A;HIST2H3A;HIST3H3	1,385	0,03497	0,87896
6	DUSP23	1,475	0,00219	0,54680	56	AK4	0,962	0,03564	0,87896
7	PTBP3	1,326	0,00241	0,54680	57	ANXA1	-1,128	0,03574	0,87896
8	LUC7L2	1,004	0,00470	0,76754	58	EIF4A3	1,322	0,03637	0,87896
9	ACAT1	-1,525	0,00546	0,76754	59	SNRPD1	1,379	0,03773	0,87896
10	LTF	4,112	0,00608	0,76754	60	PITRM1	-1,169	0,03778	0,87896
11	MZB1	-2,673	0,00634	0,76754	61	RPL19	1,026	0,03784	0,87896
12	PMVK	1,125	0,00639	0,76754	62	BPNT1	0,886	0,03896	0,87896
13	HDAC2	1,048	0,00660	0,76754	63	HIST1H4A	1,472	0,04164	0,87896
14	GGCT	1,001	0,00748	0,76754	64	ELAVL1	0,599	0,04164	0,87896
15	SNRNP70	0,971	0,00754	0,76754	65	HSPD1	-0,940	0,04193	0,87896
16	CTBP2	1,237	0,00840	0,76754	66	G6PD	-1,202	0,04292	0,87896
17	AGR2	-2,754	0,00841	0,76754	67	SFPQ	0,866	0,04416	0,87896
18	RAB5B	0,970	0,00876	0,76754	68	GAA	0,975	0,04435	0,87896
19	MRTO4	0,812	0,00924	0,76754	69	ADSL	0,630	0,04468	0,87896
20	RAB5A	0,840	0,00965	0,76754	70	RPL15	1,154	0,04598	0,87896
21	TIPRL	0,768	0,01120	0,82270	71	RPL7	0,978	0,04775	0,87896
22	PRDX1	-0,621	0,01206	0,82270	72	RPS14	0,678	0,04786	0,87896
23	HP1BP3	1,381	0,01260	0,82270	73	ETFA	-0,634	0,04809	0,87896
24	NIT1	0,876	0,01260	0,82270	74	GMPPB	0,830	0,04812	0,87896
25	CLPP	-0,950	0,01293	0,82270	75	RAB2A	0,929	0,04836	0,87896
26	ADK	0,861	0,01470	0,84526	76	GUSB	0,919	0,04908	0,87896
27	SNX5	0,743	0,01486	0,84526	77	ATP5H	-0,574	0,04947	0,87896
28	SRSF3	1,002	0,01503	0,84526	78	MYH9	-0,661	0,04962	0,87896
29	H1F0	1,751	0,01570	0,84526	79	SRSF7	0,812	0,05069	0,87896
30	IGLC6	-1,897	0,01594	0,84526	80	COL18A1	1,009	0,05074	0,87896
31	IGKC	-1,574	0,01744	0,86877	81	SRSF1	0,730	0,05164	0,87896
32	CLU	2,038	0,01802	0,86877	82	PIN1	-0,636	0,05217	0,87896
33	MRPS26	0,879	0,01884	0,87239	83	IMPDH2	0,770	0,05256	0,87896
34	S100A11	1,146	0,01935	0,87239	84	CSRP2	0,810	0,05261	0,87896
35	HNRNPA0	1,126	0,01974	0,87239	85	CCDC6	1,139	0,05272	0,87896
36	RBMX	0,894	0,02222	0,87896	86	VASP	-0,686	0,05295	0,87896
37	HNRNPUL1	1,365	0,02316	0,87896	87	RPS11	1,128	0,05314	0,87896
38	CHTOP	0,866	0,02344	0,87896	88	PTGR1	1,009	0,05324	0,87896
39	PURB	0,891	0,02431	0,87896	89	PDCD4	0,801	0,05330	0,87896
40	SQSTM1	-0,915	0,02459	0,87896	90	S100A6	-1,360	0,05331	0,87896
41	TRA2B	1,521	0,02570	0,87896	91	SPR	-1,054	0,05460	0,87896
42	CPSF7	1,027	0,02585	0,87896	92	CYB5B	-0,629	0,05534	0,87896
43	TXN	-0,902	0,02609	0,87896	93	PIGR	3,032	0,05654	0,87896
44	SRSF6	1,014	0,02665	0,87896	94	C14orf142	-1,004	0,05703	0,87896
45	RAB6A	0,847	0,02689	0,87896	95	CYB5R1	1,249	0,05724	0,87896
46	TST	-1,226	0,02744	0,87896	96	C4B	1,250	0,05768	0,87896
47	GOT1	0,715	0,02747	0,87896	97	UFC1	0,799	0,05912	0,87896
48	GTF2I	-1,301	0,03117	0,87896	98	EFHD2	-0,641	0,06141	0,87896
49	RPS4X	0,786	0,03170	0,87896	99	PRPF31	0,481	0,06273	0,87896
50	SNRPE	1,091	0,03212	0,87896	100	NONO	0,845	0,06291	0,87896

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
1	CNN3	-1,100	1,19E-05	0,01893	51	BAG2	-0,837	0,01090	0,33995
2	ANXA1	-1,296	5,85E-05	0,04655	52	PRELP	-1,141	0,01116	0,34074
3	MYH9	-0,785	0,00010	0,05305	53	LSM8	0,627	0,01135	0,34074
4	SRSF2	0,859	0,00016	0,06488	54	APOA4	-1,137	0,01195	0,34630
5	PMVK	0,879	0,00027	0,07162	55	BPGM	-0,869	0,01197	0,34630
6	ACTN1	-0,891	0,00028	0,07162	56	IDH2	-0,506	0,01227	0,34873
7	RPS14	0,733	0,00032	0,07162	57	HNRNPF	0,427	0,01262	0,34897
8	POLR2H	0,784	0,00055	0,10243	58	NUMA1	0,923	0,01272	0,34897
9	S100A4	-1,499	0,00058	0,10243	59	TMEM43	-0,519	0,01326	0,35744
10	ACTN4	-0,739	0,00065	0,10375	60	IQGAP1	-0,505	0,01366	0,35788
11	TSTD1	1,236	0,00086	0,12419	61	SEC61B	0,627	0,01391	0,35788
12	EML4	1,037	0,00104	0,13780	62	VIM	-0,630	0,01408	0,35788
13	SRSF7	0,783	0,00130	0,15776	63	SPTBN1	0,510	0,01443	0,35788
14	CBR1	-0,755	0,00148	0,15776	64	GRHPR	-0,396	0,01492	0,35788
15	CLIC4	-0,692	0,00149	0,15776	65	LCP1	-0,765	0,01502	0,35788
16	GBP1	-1,392	0,00163	0,16217	66	RPS26;RPS26P11	0,656	0,01508	0,35788
17	ARHGDIB	-0,858	0,00197	0,18447	67	RPS18	0,416	0,01514	0,35788
18	SNRPA	0,450	0,00218	0,19257	68	RBM39	0,498	0,01556	0,35788
19	TCEB2	0,578	0,00247	0,19824	69	POLR2C	0,396	0,01583	0,35788
20	EVL	-1,113	0,00261	0,19824	70	PTGFRN	0,679	0,01606	0,35788
21	ARHGAP1	-0,420	0,00266	0,19824	71	SF3B4	0,464	0,01650	0,35788
22	VCL	-0,676	0,00274	0,19824	72	RPS12	0,425	0,01697	0,35788
23	LGALS1	-0,951	0,00311	0,20921	73	PUF60	0,519	0,01716	0,35788
24	MRPL14	0,628	0,00316	0,20921	74	PHYHD1	0,630	0,01742	0,35788
25	SNRPF	0,572	0,00334	0,21269	75	GUK1	0,489	0,01744	0,35788
26	GSN	-0,425	0,00370	0,22626	76	THRAP3	0,663	0,01750	0,35788
27	HAX1	0,624	0,00406	0,23895	77	RPS4X	0,495	0,01809	0,35788
28	NAP1L4	-0,386	0,00441	0,24255	78	YTHDF3	0,511	0,01815	0,35788
29	CCAR1	0,552	0,00442	0,24255	79	COL18A1	0,700	0,01815	0,35788
30	CAT	-0,568	0,00485	0,25704	80	ALAD	-0,536	0,01819	0,35788
31	CAP1	-0,548	0,00544	0,26730	81	RSU1	-0,714	0,01886	0,35788
32	TCOF1	0,752	0,00547	0,26730	82	NME1	0,557	0,01893	0,35788
33	TLN1	-0,589	0,00557	0,26730	83	PDLIM7	-0,843	0,01899	0,35788
34	CD44	-0,764	0,00580	0,26730	84	TCEB1	0,374	0,01915	0,35788
35	TPM4	-0,542	0,00593	0,26730	85	MTX1	0,402	0,01926	0,35788
36	TGM2	-0,811	0,00605	0,26730	86	RBMX	0,521	0,01948	0,35788
37	RPSA	0,437	0,00640	0,27518	87	FAM136A	0,674	0,01968	0,35788
38	RPL8	0,829	0,00659	0,27598	88	CA1	-1,570	0,01981	0,35788
39	RPL19	0,766	0,00716	0,29207	89	ERO1L	-0,664	0,02002	0,35788
40	CAPZA1	-0,322	0,00749	0,29798	90	BOLA2;BOLA2B	0,544	0,02050	0,36239
41	RBM12	0,423	0,00782	0,30343	91	ATP5H	-0,387	0,02093	0,36312
42	LAP3	-0,698	0,00926	0,33678	92	NCOA5	0,640	0,02129	0,36312
43	HBD	-1,174	0,00931	0,33678	93	PURB	0,519	0,02131	0,36312
44	MRPS21	0,593	0,00931	0,33678	94	KARS	-0,357	0,02145	0,36312
45	HBA1	-1,209	0,00965	0,33867	95	PIGR	2,090	0,02190	0,36554
46	HBB	-1,200	0,00994	0,33867	96	SEC16A	0,574	0,02214	0,36554
47	SRSF6	0,676	0,01002	0,33867	97	CALD1	-0,800	0,02229	0,36554
48	ADRM1	0,500	0,01022	0,33867	98	SNRPD2	0,375	0,02336	0,36872
49	ILK	-0,603	0,01059	0,33995	99	RPS11	0,757	0,02359	0,36872
50	RPL31	0,531	0,01071	0,33995	100	CKB	-1,102	0,02377	0,36872

Comparison 18 – Status (35 Alive without Disease vs 11 Death of Disease)

Annexes

253

Comparison 19 summarize all statistically significant proteins (p-value < 0.05) from the absence/presence analysis (grey represents proteins selected for the verification phase).

Nº	ID	NOREC (32)		REC (31)		P.Value
1	TPM1	23	72%	11	35%	0
2	PKM	11	34%	1	3%	0
3	AHCYL1	21	66%	9	29%	0
4	ALDH1A2	22	69%	9	29%	0
5	PSMD7	27	84%	15	48%	0
6	STK4	20	63%	7	23%	0
7	SNTB2	17	53%	5	16%	0
8	IL16	25	78%	13	42%	0
9	NSDHL	18	56%	6	19%	0
10	FERMT3	26	81%	13	42%	0
11	SEPT6	12	38%	2	6%	0,01
12	FARP1	19	59%	8	26%	0,01
13	NDUFV2	32	100%	25	81%	0,01
14	MRPS22	29	91%	19	61%	0,01
15	UBFD1	23	72%	12	39%	0,01
16	LLGL2	14	44%	4	13%	0,01
17	LONP1	20	63%	8	26%	0,01
18	MAN2B1	32	100%	25	81%	0,01
19	LAMA5	1	3%	9	29%	0,01
20	WBSCR22	11	34%	22	71%	0,01
21	HPX	31	97%	23	74%	0,01
22	ADH1A;ADH1B;HEL-S-117;ADH1C	15	47%	4	13%	0,01
23	BGN	31	97%	23	74%	0,01
24	LMOD1	19	59%	7	23%	0,01
25	LRPAP1	31	97%	23	74%	0,01
26	RDX	31	97%	23	74%	0,01
27	DNM2	29	91%	19	61%	0,01
28	TFAM	32	100%	25	81%	0,01
29	BCL2L1	27	84%	16	52%	0,01
30	PSMD6	28	88%	18	58%	0,01
31	ABCF1	29	91%	19	61%	0,01
32	ZC3H15	30	94%	21	68%	0,01
33	TRNT1	21	66%	9	29%	0,01
34	PDCD10	28	88%	18	58%	0,01
35	BAG5	7	22%	0	0%	0,01
36	C14orf166	31	97%	22	71%	0,01
37	NCKAP1	19	59%	8	26%	0,01
38	GYPA	10	31%	2	6%	0,02
39	NDUFA10	20	63%	10	32%	0,02
40	PBX1	23	72%	13	42%	0,02
41	HLA-DQB1	19	59%	9	29%	0,02
42	IDE	19	59%	9	29%	0,02
43	ALDH3A2	24	75%	14	45%	0,02
44	CAMK2D	23	72%	13	42%	0,02
45	CALD1	22	69%	12	39%	0,02
46	CIAPIN1	30	94%	22	71%	0,02
47	ZNF185	27	84%	17	55%	0,02
48	BANF1	24	75%	14	45%	0,02
49	SOD3	32	100%	26	84%	0,02
50	PLEK	13	41%	4	13%	0,02
51	CTSA	28	88%	19	61%	0,02
52	BCR	3	9%	12	39%	0,02
53	IFI30	28	88%	19	61%	0,02
54	GLUL	32	100%	26	84%	0,02
55	PRKACA;KIN27	21	66%	11	35%	0,02
56	CALB2	6	19%	0	0%	0,02

Table title: Comparison 19. Absence/Presence Analysis (REC vs NOREC)

Nº	ID		NOREC (32)		REC (31)	P.Value
			Comparison 19. Absence/Presence Analysis (REC vs NOREC)			
57	RECQL	18	56%	8	26%	0,02
58	GNPDA1	18	56%	8	26%	0,02
59	NES	30	94%	22	71%	0,02
60	SERPINB9	30	94%	22	71%	0,02
61	FHL3	21	66%	11	35%	0,02
62	KANK2	29	91%	20	65%	0,02
63	ATL3	26	81%	16	52%	0,02
64	LSR	15	47%	25	81%	0,02
65	STX12	30	94%	22	71%	0,02
66	ATPAF2	8	25%	1	3%	0,02
67	SUCLG2	32	100%	26	84%	0,02
68	PSMD1	28	88%	19	61%	0,02
69	KCTD14	12	38%	3	10%	0,02
70	TPPP3	32	100%	26	84%	0,02
71	LZTFL1	26	81%	16	52%	0,02
72	NOP16	10	31%	20	65%	0,02
73	SNX9	17	53%	7	23%	0,02
74	CLIC4	32	100%	26	84%	0,02
75	SCARB2	31	97%	24	77%	0,03
76	B4DLN1	32	100%	26	84%	0,03
77	EML3	1	3%	8	26%	0,03
78	DCTN6	14	44%	5	16%	0,03
79	IPO5	26	81%	17	55%	0,03
80	PFAS	31	97%	24	77%	0,03
81	SMARCA5	27	84%	18	58%	0,03
82	SRP72	27	84%	18	58%	0,03
83	H6PD	29	91%	21	68%	0,03
84	PCCA	25	78%	16	52%	0,03
85	DCN	32	100%	26	84%	0,03
86	EIF2AK2	26	81%	17	55%	0,03
87	MYH10	31	97%	24	77%	0,03
88	HMGCL	16	50%	7	23%	0,03
89	CDKN2C	15	47%	6	19%	0,03
90	PAFAH1B1	27	84%	18	58%	0,03
91	NAMPT	31	97%	24	77%	0,03
92	GTF2I	32	100%	26	84%	0,03
93	NFKB2	11	34%	3	10%	0,03
94	SMU1	26	81%	17	55%	0,03
95	FAM98B	31	97%	24	77%	0,03
96	MISP	15	47%	6	19%	0,03
97	NEK9	8	25%	1	3%	0,03
98	GCN1L1	32	100%	26	84%	0,03
99	COPS5	27	84%	18	58%	0,03
100	USP7	21	66%	11	35%	0,03
101	GALM	31	97%	24	77%	0,03
102	TRMT61A	5	16%	14	45%	0,03
103	ECHDC1	29	91%	21	68%	0,03
104	PDP1	8	25%	1	3%	0,03
105	EVL	31	97%	24	77%	0,03
106	PPP1R8	17	53%	8	26%	0,04
107	HLA-A	23	72%	14	45%	0,04
108	ALDH5A1	21	66%	12	39%	0,04
109	COL6A3	25	78%	16	52%	0,04
110	NDUFB5	17	53%	8	26%	0,04
111	CRMP1	9	28%	2	6%	0,04
112	FKBP11	23	72%	14	45%	0,04

Nº	ID	NOREC (32)		REC (31)		P.Value
113	AIP	30	94%	23	74%	0,04
114	TOMM40	26	81%	17	55%	0,04
115	ATP6V1C1	22	69%	13	42%	0,04
116	TMOD1	21	66%	12	39%	0,04
117	MCM5	21	66%	12	39%	0,04
118	STT3A	25	78%	16	52%	0,04
119	COPB1	17	53%	8	26%	0,04
120	FMR1	25	78%	16	52%	0,04
121	HSPA14	9	28%	2	6%	0,04
122	USP10	24	75%	15	48%	0,04
123	UAP1	9	28%	2	6%	0,04
124	TIMM50	20	63%	11	35%	0,04
125	PITRM1	30	94%	23	74%	0,04
126	EXOSC6	11	34%	20	65%	0,04
127	C14orf159	16	50%	7	23%	0,04
128	SPRYD4	16	50%	7	23%	0,04
129	RNF2	2	6%	9	29%	0,04
130	VAT1	30	94%	23	74%	0,04
131	PLIN2	4	13%	12	39%	0,04
132	POTEKP	21	66%	12	39%	0,04
133	FAM192A;NIP30	19	59%	10	32%	0,04
134	ATXN10	28	88%	20	65%	0,04
135	AP3D1	7	22%	1	3%	0,05
136	ACAT1	31	97%	25	81%	0,05
137	AHSG	31	97%	25	81%	0,05
138	SEPT11	32	100%	27	87%	0,05
139	OPA1	13	41%	5	16%	0,05
140	COPS8	31	97%	25	81%	0,05
141	NDUFS8	22	69%	13	42%	0,05
142	SCYL1	5	16%	0	0%	0,05
143	MSRB3	13	41%	5	16%	0,05
144	COPZ1	32	100%	27	87%	0,05
145	SEH1L	13	41%	5	16%	0,05
146	PSMD14	32	100%	27	87%	0,05
147	DFNA5	5	16%	0	0%	0,05
148	FTL	31	97%	25	81%	0,05
149	MYL1;MYL3	32	100%	27	87%	0,05
150	SRPR	20	63%	11	35%	0,05
151	PEPD	32	100%	27	87%	0,05
152	ANPEP	14	44%	6	19%	0,05
153	GARS	32	100%	27	87%	0,05
154	PSMD8	32	100%	27	87%	0,05
155	DUSP3	30	94%	23	74%	0,05
156	ARHGDIB	32	100%	27	87%	0,05
157	RAD23A	32	100%	27	87%	0,05
158	UFM1	32	100%	27	87%	0,05
159	GFPT1	30	94%	23	74%	0,05
160	FLII	26	81%	18	58%	0,05
161	PLS1	21	66%	12	39%	0,05
162	AARS2	7	22%	1	3%	0,05
163	NT5DC1	21	66%	12	39%	0,05
164	C18orf25	7	22%	1	3%	0,05
165	WRNIP1	5	16%	0	0%	0,05
166	NIF3L1	18	56%	9	29%	0,05
167	PDLIM7	32	100%	27	87%	0,05
168	L1RE1	0	0%	5	16%	0,05

Comparison 19. Absence/Presence Analysis (REC vs NOREC)

ANNEX 3

Comparisons between recurrent and non-recurrent molecularly classified EC patients

Summary of the most important proteins from each comparison in the quantitative analysis (dark green represents proteins with and p-value < 0.01; light green represents proteins with an p-value < 0.05); and grey represents proteins selected for the verification phase).

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				Microsatellite Instability - MSI (16 REC vs 18 NOREC)					
1	MYH9	-0,662	0,0027	0,916	39	LRPPRC	-1,098	0,0248	0,916
2	EIF3B	-0,563	0,0037	0,916	40	TMEM43	-0,583	0,0253	0,916
3	CORO7	-0,787	0,0045	0,916	41	CHTOP	0,489	0,0260	0,916
4	PHYHD1	0,771	0,0047	0,916	42	RAB5A	-0,434	0,0272	0,916
5	NUCKS1	0,782	0,0048	0,916	43	MYO1C	-0,423	0,0273	0,916
6	TSTD1	0,933	0,0056	0,916	44	SERPINA3	0,731	0,0274	0,916
7	COL18A1	0,982	0,0063	0,916	45	MUC1	1,321	0,0282	0,916
8	IPO7	-0,571	0,0071	0,916	46	PDHB	-0,383	0,0291	0,916
9	PRDX6	0,571	0,0073	0,916	47	UFL1	-0,402	0,0299	0,916
10	IDH3B	-0,472	0,0075	0,916	48	SRP14	0,329	0,0323	0,916
11	HSPA9	-0,542	0,0084	0,916	49	OGDH	-0,444	0,0342	0,916
12	SPTBN1	0,532	0,0086	0,916	50	ZNF326	-0,559	0,0343	0,916
13	CORO1C	-0,845	0,0090	0,916	51	PPP3CA	-0,589	0,0346	0,916
14	PAICS	-0,561	0,0095	0,916	52	LAMTOR1	-0,403	0,0348	0,916
15	PSMD11	-0,457	0,0098	0,916	53	SRSF2	0,487	0,0351	0,916
16	TGM2	-0,980	0,0100	0,916	54	RPS14	0,426	0,0363	0,916
17	PIGR	2,272	0,0109	0,916	55	ABI1	0,349	0,0366	0,916
18	STAT1	-1,259	0,0126	0,916	56	EIF4G1	-0,374	0,0367	0,916
19	ABCE1	-0,713	0,0130	0,916	57	ACADVL	-0,782	0,0372	0,916
20	ACTL6A	0,278	0,0131	0,916	58	RBM12	0,400	0,0380	0,916
21	THYN1	0,462	0,0155	0,916	59	PSMD3	-0,688	0,0381	0,916
22	NPM1	0,344	0,0164	0,916	60	DUSP23	0,702	0,0383	0,916
23	DPP7	0,964	0,0168	0,916	61	WDR1	-0,531	0,0388	0,916
24	BASP1	0,761	0,0180	0,916	62	TFG	-0,366	0,0406	0,916
25	THUMPD1	0,532	0,0187	0,916	63	EIF3K	-0,310	0,0410	0,916
26	VARS1	-0,438	0,0188	0,916	64	NSUN2	-0,338	0,0427	0,916
27	S100A11	1,062	0,0190	0,916	65	ATG3	-0,396	0,0433	0,916
28	ARL3	0,476	0,0193	0,916	66	PRKDC	-0,910	0,0445	0,916
29	EIF3J	-0,413	0,0198	0,916	67	RPS17	0,359	0,0446	0,916
30	FDPS	0,700	0,0207	0,916	68	SRSF7	0,498	0,0448	0,916
31	GET3	-0,416	0,0215	0,916	69	SFXN1	-0,322	0,0460	0,916
32	ZNF207	0,325	0,0222	0,916	70	TRAP1	-0,711	0,0463	0,916
33	HMGB1	0,655	0,0229	0,916	71	KPNB1	-0,346	0,0464	0,916
34	MSN	-0,569	0,0231	0,916	72	PARP1	0,786	0,0467	0,916
35	EIF3L	-0,821	0,0234	0,916	73	API5	0,394	0,0481	0,916
36	GGCT	0,512	0,0234	0,916	74	TXNL1	0,252	0,0481	0,916
37	TPM4	-0,505	0,0237	0,916	75	EIF5	-0,300	0,0489	0,916
38	ARPC1B	-0,615	0,0246	0,916					

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
1	VAMP8	1,043	0,00003	0,039	51	KARS1	-0,451	0,00387	0,112
2	ACBD3	0,705	0,00015	0,061	52	MLF2	1,008	0,00393	0,112
3	ACY1	0,925	0,00019	0,061	53	HSPE1	0,843	0,00397	0,112
4	SNRPF	1,034	0,00023	0,061	54	IRF2BP2	0,533	0,00407	0,113
5	RBM3	0,842	0,00028	0,061	55	RPL22	0,720	0,00434	0,116
6	CLIC4	-0,905	0,00029	0,061	56	ACTN1	-0,777	0,00436	0,116
7	RPL31	0,794	0,00029	0,061	57	NME1	0,606	0,00440	0,116
8	ATP5F1D	0,958	0,00036	0,061	58	RPL8	0,877	0,00478	0,124
9	ACTN4	-0,836	0,00037	0,061	59	ASAH1	0,871	0,00492	0,125
10	LAD1	1,667	0,00041	0,061	60	RPS21	0,611	0,00499	0,125
11	LSM8	0,927	0,00045	0,061	61	CCT3	-0,795	0,00528	0,130
12	COA3	1,042	0,00055	0,065	62	SEC16A	1,151	0,00538	0,130
13	MRPL14	1,057	0,00057	0,065	63	RSU1	-1,053	0,00550	0,131
14	POLR2C	0,566	0,00062	0,066	64	HSP90AA1	-0,867	0,00578	0,133
15	TLN1	-0,838	0,00072	0,072	65	JPT1	1,021	0,00579	0,133
16	CCT5	-0,862	0,00083	0,074	66	COMT	1,081	0,00596	0,135
17	RPS18	0,633	0,00093	0,074	67	NAGK	0,395	0,00611	0,137
18	ENAH	0,927	0,00096	0,074	68	FDX1	0,767	0,00629	0,137
19	BPNT1	0,879	0,00101	0,074	69	SRI	0,733	0,00630	0,137
20	ST13	-0,715	0,00103	0,074	70	RPS17	0,426	0,00639	0,137
21	ILK	-0,938	0,00104	0,074	71	LGALS3BP	0,707	0,00671	0,142
22	POLR2H	0,551	0,00112	0,075	72	ANXA6	-1,293	0,00693	0,143
23	GUK1	0,866	0,00115	0,075	73	TRMT112	0,799	0,00699	0,143
24	DUT	1,020	0,00149	0,087	74	FAM136A	0,790	0,00719	0,146
25	PURB	1,041	0,00150	0,087	75	TMED1	0,504	0,00730	0,146
26	DUSP23	1,012	0,00150	0,087	76	RPS20	0,500	0,00749	0,148
27	UBL4A	0,844	0,00189	0,100	77	VTI1B	0,653	0,00786	0,149
28	RPS28	1,130	0,00196	0,100	78	JUP	0,531	0,00794	0,149
29	ENO1	-0,485	0,00199	0,100	79	RAB18	0,525	0,00803	0,149
30	ADSS2	0,459	0,00200	0,100	80	ACTR3	-0,604	0,00819	0,149
31	PAXX	0,876	0,00228	0,103	81	RNH1	-0,465	0,00821	0,149
32	WDR1	-0,503	0,00239	0,103	82	ACAT1	-0,799	0,00824	0,149
33	BRK1	0,733	0,00251	0,103	83	ATP5MK	0,716	0,00832	0,149
34	HSPA1B	-0,591	0,00254	0,103	84	CKB	-1,369	0,00837	0,149
35	CCT7	-0,852	0,00256	0,103	85	LSM5	0,898	0,00868	0,153
36	RAB11B	0,508	0,00258	0,103	86	LSM4	0,701	0,00889	0,155
37	SRP14	0,474	0,00268	0,103	87	CCT8	-0,781	0,00933	0,159
38	TOMM22	0,850	0,00279	0,103	88	TRAPPC3	0,462	0,00945	0,159
39	VCL	-0,732	0,00281	0,103	89	FN3K	0,565	0,00949	0,159
40	GAA	1,367	0,00287	0,103	90	SPR	1,160	0,00955	0,159
41	SEC61B	0,920	0,00290	0,103	91	GIPC1	0,529	0,00990	0,163
42	ADH5	-0,898	0,00295	0,103	92	MTHFD1	-0,723	0,01040	0,167
43	RPS14	0,752	0,00305	0,103	93	GSS	0,641	0,01045	0,167
44	PMVK	1,207	0,00306	0,103	94	RAB21	0,656	0,01052	0,167
45	PYM1	0,712	0,00310	0,103	95	GUSB	0,835	0,01060	0,167
46	SRSF9	0,780	0,00326	0,106	96	PTBP3	0,663	0,01114	0,170
47	TBCA	0,492	0,00348	0,111	97	SSR4	0,595	0,01114	0,170
48	TSTD1	1,668	0,00375	0,112	98	ALDH9A1	0,787	0,01126	0,170
49	MRPL49	0,699	0,00379	0,112	99	BAX	0,475	0,01142	0,170
50	GPI	-0,752	0,00384	0,112	100	NME2	0,596	0,01159	0,170

Low copy-number - LCN (11 REC vs 15 NOREC)

Annexes

Nº	ID	logFC	P.Value	adj.P.Val
	High copy-number - HCN (21 REC vs 11 NOREC)			
1	PYGL	-1,358	0,0009	0,993
2	TUFM	-0,693	0,0032	0,993
3	FAH	0,598	0,0033	0,993
4	TES	-0,822	0,0066	0,993
5	PDCD6IP	-0,657	0,0081	0,993
6	RCC1	-0,652	0,0102	0,993
7	SRSF2	0,765	0,0113	0,993
8	DCXR	1,110	0,0138	0,993
9	RAB11B	0,403	0,0148	0,993
10	SSB	-0,407	0,0148	0,993
11	CTTN	-0,754	0,0175	0,993
12	ENO1	-0,546	0,0176	0,993
13	NASP	-0,513	0,0177	0,993
14	NDUFS1	-0,617	0,0185	0,993
15	PPP1CB	-0,538	0,0190	0,993
16	CHMP4A	-0,492	0,0193	0,993
17	RAB5A	0,381	0,0198	0,993
18	STAT3	-0,523	0,0214	0,993
19	PSMB7	0,685	0,0242	0,993
20	HSPB1	-0,573	0,0242	0,993
21	CRIP2	-0,876	0,0244	0,993
22	CRYZ	-0,624	0,0253	0,993
23	TSTD1	1,252	0,0253	0,993
24	RPA3	0,585	0,0267	0,993
25	G3BP2	-0,409	0,0314	0,993
26	ZYX	-0,553	0,0315	0,993
27	CYB5B	0,508	0,0350	0,993
28	SCCPDH	-0,465	0,0354	0,993
29	UQCRC2	-0,641	0,0360	0,993
30	PRPF31	-0,327	0,0407	0,993
31	CAPRIN1	-0,419	0,0410	0,993
32	LSM2	0,505	0,0425	0,993
33	CPT2	-0,431	0,0428	0,993
34	VPS29	-0,316	0,0439	0,993
35	PPIH	-0,339	0,0479	0,993
36	CLIC4	-0,466	0,0486	0,993
37	FUBP1	-0,488	0,0497	0,993

Annexes

Summary of the most important proteins from each comparison in the absence/presence analysis with a p-value < 0.05 (grey represents proteins selected for the verification phase).

Microsatellite Instability - MSI (18 NOREC vs 16 REC)

Nº	ID	NOREC (18)		REC (16)		P-value	Nº	ID	NOREC (18)		REC (16)		P-value
1	BCL2L1	17	94%	6	38%	0,001	52	CAP2	10	56%	3	19%	0,039
2	SPCS2	18	100%	8	50%	0,001	53	CDKN2C	10	56%	3	19%	0,039
3	NFKB2	8	44%	0	0%	0,002	54	ACY1	18	100%	12	75%	0,039
4	MYH10	18	100%	10	63%	0,006	55	SEPTIN11	18	100%	12	75%	0,039
5	NES	18	100%	10	63%	0,006	56	COPZ1	18	100%	12	75%	0,039
6	MCM2	18	100%	10	63%	0,006	57	CHCHD6	18	100%	12	75%	0,039
7	TOP2B	0	0%	6	38%	0,006	58	MAN2B1	18	100%	12	75%	0,039
8	KANK2	16	89%	7	44%	0,009	59	HSPB6	18	100%	12	75%	0,039
9	BORCS5	10	56%	2	13%	0,013	60	UBE2L6	18	100%	12	75%	0,039
10	MMTAG2	1	6%	7	44%	0,015	61	STAU1	18	100%	12	75%	0,039
11	B4DLN1	18	100%	11	69%	0,016	62	MYL1	18	100%	12	75%	0,039
12	ORM1	18	100%	11	69%	0,016	63	APOA4	18	100%	12	75%	0,039
13	GFPT1	18	100%	11	69%	0,016	64	DERPC	18	100%	12	75%	0,039
14	VAT1	18	100%	11	69%	0,016	65	GARS1	18	100%	12	75%	0,039
15	CFL2	18	100%	11	69%	0,016	66	RAD23A	18	100%	12	75%	0,039
16	SNX5	18	100%	11	69%	0,016	67	TFAM	18	100%	12	75%	0,039
17	AHCYL1	11	61%	3	19%	0,017	68	NAXD	18	100%	12	75%	0,039
18	FHL3	11	61%	3	19%	0,017	69	GCN1	18	100%	12	75%	0,039
19	SERPINA5	7	39%	13	81%	0,017	70	CHP1	18	100%	12	75%	0,039
20	POLR2L	8	44%	1	6%	0,019	71	CLIC4	18	100%	12	75%	0,039
21	EMG1	8	44%	1	6%	0,019	72	SAR1B	18	100%	12	75%	0,039
22	OXA1L	6	33%	0	0%	0,020	73	DMBT1	0	0%	4	25%	0,039
23	SEPTIN6	6	33%	0	0%	0,020	74	IRF6	0	0%	4	25%	0,039
24	CLUH	12	67%	4	25%	0,020	75	GPC4	0	0%	4	25%	0,039
25	STK4	12	67%	4	25%	0,020	76	WDR5	0	0%	4	25%	0,039
26	OPA1	9	50%	2	13%	0,030	77	CDO1	0	0%	4	25%	0,039
27	SSNA1	9	50%	2	13%	0,030	78	SVIP	0	0%	4	25%	0,039
28	RABGGTB	9	50%	2	13%	0,030	79	RNF2	0	0%	4	25%	0,039
29	BORCS6	9	50%	2	13%	0,030	80	RNF213	7	39%	1	6%	0,043
30	RBM26	3	17%	9	56%	0,030	81	ACADSB	7	39%	1	6%	0,043
31	PODXL	3	17%	9	56%	0,030	82	SNTB2	7	39%	1	6%	0,043
32	MRPS18B	17	94%	9	56%	0,030	83	SNX17	7	39%	1	6%	0,043
33	MBNL1	17	94%	10	63%	0,035	84	LONP1	11	61%	4	25%	0,045
34	LRPAP1	17	94%	10	63%	0,035	85	SRPRA	11	61%	4	25%	0,045
35	PPP1R14B	17	94%	10	63%	0,035	86	PRKACA	11	61%	4	25%	0,045
36	RTN4	17	94%	10	63%	0,035	87	ABCF2	11	61%	4	25%	0,045
37	RTRAF	17	94%	10	63%	0,035	88	SCYL1	5	28%	0	0%	0,047
38	GGPS1	1	6%	6	38%	0,035	89	PKM	5	28%	0	0%	0,047
39	ASPN	1	6%	6	38%	0,035	90	PRKAR2B	5	28%	0	0%	0,047
40	DIABLO	14	78%	6	38%	0,035	91	CHI3L1	5	28%	0	0%	0,047
41	CXADR	4	22%	10	63%	0,035	92	GATM	5	28%	0	0%	0,047
42	HLA-DQB1	13	72%	5	31%	0,037	93	HSPA14	5	28%	0	0%	0,047
43	CALD1	13	72%	5	31%	0,037	94	RAB32	5	28%	0	0%	0,047
44	IGF2BP2	13	72%	5	31%	0,037	95	SLC1A5	5	28%	0	0%	0,047
45	ALDH1A2	13	72%	5	31%	0,037	96	SPATS2	5	28%	0	0%	0,047
46	HDHD5	13	72%	5	31%	0,037	97	SMIM20	5	28%	0	0%	0,047
47	LMCD1	13	72%	5	31%	0,037	98	ATPAF2	5	28%	0	0%	0,047
48	GDA	13	72%	5	31%	0,037	99	BPIFB1	5	28%	0	0%	0,047
49	NNT	10	56%	3	19%	0,039	100	PSMG3	5	28%	0	0%	0,047
50	FARP1	10	56%	3	19%	0,039	101	LANCL2	5	28%	0	0%	0,047
51	LMOD1	10	56%	3	19%	0,039							

Nº	ID	NOREC (15)		REC (11)		P-value	Nº	ID	NOREC (15)		REC (11)		P-value
1	PBX1	15	100%	3	27%	0,0001	47	PSMD1	15	100%	7	64%	0,0221
2	TUBB2A	15	100%	4	36%	0,0005	48	CHID1	15	100%	7	64%	0,0221
3	KPNA3	14	93%	3	27%	0,0008	49	ATXN10	15	100%	7	64%	0,0221
4	STK4	11	73%	1	9%	0,0017	50	STX3	0	0%	4	36%	0,0221
5	ZNRF2	2	13%	8	73%	0,0040	51	EML3	0	0%	4	36%	0,0221
6	ALDH1A2	12	80%	2	18%	0,0043	52	EMC7	0	0%	4	36%	0,0221
7	TPM2	12	80%	2	18%	0,0043	53	IMP3	0	0%	4	36%	0,0221
8	SNX9	12	80%	2	18%	0,0043	54	GYPA	6	40%	0	0%	0,0237
9	EIF2AK2	15	100%	6	55%	0,0070	55	CRMP1	6	40%	0	0%	0,0237
10	STAT3	15	100%	6	55%	0,0070	56	RRM1	6	40%	0	0%	0,0237
11	DNM2	15	100%	6	55%	0,0070	57	TMLHE	6	40%	0	0%	0,0237
12	PSMD6	15	100%	6	55%	0,0070	58	ESRP1	8	53%	1	9%	0,0362
13	ECHDC1	15	100%	6	55%	0,0070	59	TPP2	8	53%	1	9%	0,0362
14	RAB6B	0	0%	5	45%	0,0070	60	CRYM	8	53%	1	9%	0,0362
15	KCTD14	7	47%	0	0%	0,0103	61	STIM1	7	47%	10	91%	0,0362
16	UBFD1	13	87%	4	36%	0,0135	62	MRRF	7	47%	10	91%	0,0362
17	GET4	13	87%	4	36%	0,0135	63	SH3PXD2B	13	87%	5	45%	0,0384
18	USP7	13	87%	4	36%	0,0135	64	HNRNPLL	13	87%	5	45%	0,0384
19	ACBD5	2	13%	7	64%	0,0135	65	BANF1	13	87%	5	45%	0,0384
20	DECR2	2	13%	7	64%	0,0135	66	TOMM40	13	87%	5	45%	0,0384
21	C20orf27	9	60%	1	9%	0,0143	67	PPM1A	13	87%	5	45%	0,0384
22	TPM1	12	80%	3	27%	0,0149	68	FLII	13	87%	5	45%	0,0384
23	RECQL	12	80%	3	27%	0,0149	69	EBNA1BP2	13	87%	5	45%	0,0384
24	KIAA1217	3	20%	8	73%	0,0149	70	UBE2Z	13	87%	5	45%	0,0384
25	UMPS	11	73%	2	18%	0,0154	71	COPS7B	13	87%	5	45%	0,0384
26	PSMD10	14	93%	5	45%	0,0208	72	SAR1A	13	87%	5	45%	0,0384
27	DLAT	14	93%	5	45%	0,0208	73	KATNAL2	2	13%	6	55%	0,0384
28	PRMT1	14	93%	5	45%	0,0208	74	MRPL34	2	13%	6	55%	0,0384
29	IPO5	14	93%	5	45%	0,0208	75	HMGN2	2	13%	6	55%	0,0384
30	SMARCA5	14	93%	5	45%	0,0208	76	BCR	2	13%	6	55%	0,0384
31	MARS1	14	93%	5	45%	0,0208	77	IGSF8	2	13%	6	55%	0,0384
32	HMGN3	14	93%	5	45%	0,0208	78	COL6A3	12	80%	4	36%	0,0426
33	COPS5	14	93%	5	45%	0,0208	79	ATP5PB	12	80%	4	36%	0,0426
34	LZTFL1	14	93%	5	45%	0,0208	80	CARS1	12	80%	4	36%	0,0426
35	NECTIN4	1	7%	6	55%	0,0208	81	FERMT3	12	80%	4	36%	0,0426
36	LSR	5	33%	9	82%	0,0214	82	NUP214	3	20%	7	64%	0,0426
37	GNPDA1	10	67%	2	18%	0,0214	83	PTPN18	3	20%	7	64%	0,0426
38	NCBP1	10	67%	2	18%	0,0214	84	CCDC115	3	20%	7	64%	0,0426
39	HLA-DRB1	15	100%	7	64%	0,0221	85	TSC22D4	3	20%	7	64%	0,0426
40	CPNE3	15	100%	7	64%	0,0221	86	TEX264	4	27%	8	73%	0,0447
41	GBP1	15	100%	7	64%	0,0221	87	REPS1	4	27%	8	73%	0,0447
42	RDX	15	100%	7	64%	0,0221	88	IDE	11	73%	3	27%	0,0447
43	TTLL12	15	100%	7	64%	0,0221	89	CAMK2D	11	73%	3	27%	0,0447
44	CCDC6	15	100%	7	64%	0,0221	90	MCM5	11	73%	3	27%	0,0447
45	ZC3H15	15	100%	7	64%	0,0221	91	AGL	11	73%	3	27%	0,0447
46	PRPF31	15	100%	7	64%	0,0221							

Low copy-number - LCN (15 NOREC vs 11 REC)

Nº	ID	NOREC (11)		REC (21)		P-value	Nº	ID	NOREC (11)		REC (21)		P-value
			High copy-number - HCN (11 NOREC vs 21 REC)										
1	VGLL4	5	45%	0	0%	0,002	40	SLC2A1	5	45%	2	10%	0,032
2	GDA	8	73%	4	19%	0,006	41	RTF1	5	45%	2	10%	0,032
3	HMBS	11	100%	11	52%	0,006	42	NOMO3	3	27%	0	0%	0,033
4	DYNC1LI1	10	91%	8	38%	0,008	43	ARVCF	3	27%	0	0%	0,033
5	LRG1	1	9%	13	62%	0,008	44	NOL3	3	27%	0	0%	0,033
6	GATAD2B	6	55%	2	10%	0,010	45	FOSL2	3	27%	0	0%	0,033
7	AP3D1	6	55%	2	10%	0,010	46	JUND	3	27%	0	0%	0,033
8	PTGR2	6	55%	2	10%	0,010	47	RFC3	3	27%	0	0%	0,033
9	NEK9	6	55%	2	10%	0,010	48	SLC7A5	3	27%	0	0%	0,033
10	WRNIP1	6	55%	2	10%	0,010	49	RBBP5	3	27%	0	0%	0,033
11	SHPK	10	91%	9	43%	0,011	50	RPS19BP1	3	27%	0	0%	0,033
12	MRPS6	5	45%	1	5%	0,011	51	DNAJA3	3	27%	0	0%	0,033
13	LAS1L	5	45%	1	5%	0,011	52	BAG5	3	27%	0	0%	0,033
14	RFC2	7	64%	3	14%	0,013	53	AKAP8	6	55%	3	14%	0,035
15	UBE2I	11	100%	12	57%	0,013	54	PGR	6	55%	3	14%	0,035
16	NANS	11	100%	12	57%	0,013	55	MAOA	6	55%	3	14%	0,035
17	OGA	7	64%	4	19%	0,020	56	JAGN1	6	55%	3	14%	0,035
18	RAB13	7	64%	4	19%	0,020	57	C8orf33	4	36%	1	5%	0,037
19	SH3PXD2B	8	73%	5	24%	0,021	58	MYADM	4	36%	1	5%	0,037
20	SMARCC1	8	73%	5	24%	0,021	59	RBM38	4	36%	1	5%	0,037
21	RPRD1A	8	73%	5	24%	0,021	60	CCDC22	4	36%	1	5%	0,037
22	GRSF1	10	91%	10	48%	0,023	61	UBE4B	4	36%	1	5%	0,037
23	CARS2	10	91%	10	48%	0,023	62	SGPL1	4	36%	1	5%	0,037
24	ATP6V1C1	9	82%	7	33%	0,023	63	PEX19	4	36%	1	5%	0,037
25	CRTAP	8	73%	6	29%	0,027	64	SMC1A	4	36%	1	5%	0,037
26	QRICH1	8	73%	6	29%	0,027	65	UQCC2	4	36%	1	5%	0,037
27	ZC3H14	8	73%	6	29%	0,027	66	EFHD1	4	36%	1	5%	0,037
28	NMNAT1	8	73%	6	29%	0,027	67	GID8	4	36%	1	5%	0,037
29	STARD10	8	73%	6	29%	0,027	68	ARFGEF1	4	36%	1	5%	0,037
30	GTPBP1	9	82%	8	38%	0,028	69	DDT	7	64%	20	95%	0,037
31	BIN1	9	82%	8	38%	0,028	70	TNXB	10	91%	11	52%	0,050
32	HTATSF1	9	82%	8	38%	0,028	71	KIF2A	10	91%	11	52%	0,050
33	IPO4	9	82%	8	38%	0,028	72	SCAMP3	10	91%	11	52%	0,050
34	CD2AP	9	82%	8	38%	0,028	73	HEXIM1	10	91%	11	52%	0,050
35	ZNF185	11	100%	13	62%	0,029	74	PSMD7	10	91%	11	52%	0,050
36	PAPSS1	11	100%	13	62%	0,029	75	RRAS2	10	91%	11	52%	0,050
37	FLNC	11	100%	13	62%	0,029	76	MRPS25	10	91%	11	52%	0,050
38	MAP1S	11	100%	13	62%	0,029	77	NPEPL1	10	91%	11	52%	0,050
39	LAMA5	0	0%	8	38%	0,029	78	TARS2	10	91%	11	52%	0,050

ANNEX 4

Comparisons between recurrent and non-recurrent in CPTAC

Summary of the most important proteins from the quantitative analysis (dark green represents proteins with and adjusted p-value < 0.05; light green represents proteins with an adjusted p-value < 0.25).

Nº	ID	logFC	P.Value	adj.P.Val	Nº	ID	logFC	P.Value	adj.P.Val
				GENERAL – 13 REC vs 74 NO REC					
1	IGKV2D-40	1,534	<0,001	0,0002	59	PF4V1	2,005	<0,001	0,1314
2	FAM171A2	1,497	<0,001	0,0072	60	ITM2C	2,243	<0,001	0,1320
3	IGLV5-45	1,893	<0,001	0,0072	61	H2BC13	1,265	<0,001	0,1329
4	FETUB	1,032	<0,001	0,0152	62	CD38	1,451	<0,001	0,1329
5	IGHG1	1,310	<0,001	0,0207	63	PON1	0,869	<0,001	0,1347
6	CPN2	0,887	<0,001	0,0207	64	HBG1	1,639	0,001	0,1414
7	IGLV3-25	1,414	<0,001	0,0207	65	USP38	1,024	0,001	0,1414
8	IGHV1-3	1,429	<0,001	0,0207	66	CA1	1,307	0,001	0,1414
9	IGKV2-30	1,362	<0,001	0,0207	67	PLD3	1,200	0,001	0,1414
10	A2M	0,994	<0,001	0,0246	68	VSIR	0,756	0,001	0,1438
11	IGHV2-26	1,419	<0,001	0,0268	69	IGHV3OR16-9	0,811	0,001	0,1532
12	IGHA1	1,452	<0,001	0,0268	70	IGKC	1,093	0,001	0,1670
13	FERMT3	1,073	<0,001	0,0268	71	ITIH2	1,100	0,001	0,1670
14	TUBB1	1,923	<0,001	0,0268	72	METTL7A	0,722	0,001	0,1670
15	PLEK	1,688	<0,001	0,0268	73	FCER1G	1,446	0,001	0,1670
16	ICAM2	0,692	<0,001	0,0284	74	CD47	1,125	0,001	0,1673
17	IGHV3-9	1,039	<0,001	0,0287	75	CNDP1	0,986	0,001	0,1673
18	F13B	1,136	<0,001	0,0413	76	CDK6	0,844	0,002	0,1819
19	FCGR2A	1,716	<0,001	0,0445	77	CPN1	0,707	0,002	0,1819
20	IGFALS	0,754	<0,001	0,0445	78	OS9	0,632	0,002	0,1851
21	A0A0J9YY99	1,451	<0,001	0,0445	79	APOE	1,109	0,002	0,1851
22	ADD2	1,392	<0,001	0,0451	80	CYBB	1,095	0,002	0,1851
23	ARHGAP4	0,908	<0,001	0,0470	81	THAP4	1,539	0,002	0,1851
24	IGLC1	1,129	<0,001	0,0476	82	IGLV9-49	2,032	0,002	0,1851
25	POFUT1	0,983	<0,001	0,0479	83	IGKV3-15	1,001	0,002	0,1851
26	RTN2	1,178	<0,001	0,0514	84	STON2	1,311	0,002	0,1851
27	HLA-A	1,704	<0,001	0,0514	85	IGHV3-64	0,865	0,002	0,1851
28	MRC1	0,841	<0,001	0,0551	86	OTUB1	0,855	0,002	0,1851
29	BCAM	2,061	<0,001	0,0551	87	DUS2	0,891	0,002	0,1898
30	SLC4A1	1,330	<0,001	0,0556	88	COL2A1	3,773	0,002	0,1898
31	PROZ	1,216	<0,001	0,0746	89	MMRN1	0,633	0,002	0,1898
32	TTC17	0,722	<0,001	0,0746	90	IGLV1-51	2,242	0,002	0,1972
33	SLC8A1	1,669	<0,001	0,0802	91	MEST	1,841	0,002	0,2028
34	ITIH4	0,876	<0,001	0,0811	92	TSPYL5	1,695	0,002	0,2028
35	BLVRB	0,687	<0,001	0,0890	93	F5	1,665	0,002	0,2028
36	MMP12	1,786	<0,001	0,0890	94	AUTS2	1,032	0,002	0,2028
37	SLC35F2	1,212	<0,001	0,0890	95	F13A1	0,713	0,002	0,2093
38	A0A0G2JRQ6	1,457	<0,001	0,0890	96	ALDH1L2	0,977	0,002	0,2113
39	THEMIS2	0,995	<0,001	0,0915	97	FCGR1A	1,264	0,002	0,2153
40	ITGA2B	2,177	<0,001	0,0948	98	TRAM1	1,065	0,002	0,2153
41	SLC22A18	1,194	<0,001	0,0955	99	TMEM43	0,763	0,003	0,2356
42	TCF7L2	1,838	<0,001	0,1010	100	C4A	0,634	0,003	0,2386
43	IGKV3-20	1,120	<0,001	0,1010	101	C17orf80	1,047	0,003	0,2386
44	IGLV4-69	2,296	<0,001	0,1010	102	GC	0,675	0,003	0,2387
45	GMFG	1,065	<0,001	0,1098	103	KIF5C	1,240	0,003	0,2387
46	FGG	0,805	<0,001	0,1127	104	SPTA1	1,151	0,003	0,2499
47	C5	0,786	<0,001	0,1140	105	C2	0,637	0,003	0,2499
48	QRICH1	0,783	<0,001	0,1208	106	CYP2S1	1,169	0,003	0,2499
49	EPB42	1,151	<0,001	0,1208	107	TSC22D3	0,921	0,003	0,2499
50	C4BPB	1,249	<0,001	0,1288	108	FGA	0,947	0,003	0,2499
51	PAG1	1,183	<0,001	0,1298	109	IGHV5-51	1,651	0,003	0,2499
52	PROS1	1,039	<0,001	0,1298	110	SIX1	1,668	0,003	0,2499
53	IGHV2-5	1,711	<0,001	0,1298	111	NELFE	0,628	0,003	0,2499
54	IGLV2-11	2,122	<0,001	0,1298	112	LRG1	0,866	0,003	0,2499
55	POMZP3	0,745	<0,001	0,1314	113	QPCTL	0,966	0,003	0,2499
56	CDKN2A	1,320	<0,001	0,1314	114	SHBG	0,899	0,003	0,2499
57	RAB31	0,886	<0,001	0,1314	115	PRG4	0,983	0,003	0,2499
58	SAA4	1,205	<0,001	0,1314					

EEC – 9 REC vs 64 NO REC
Nº	ID	logFC	P.Value	adj.P.Val
1	IGLV5-45	2,354	<0,001	0,0004
2	IGKV2D-40	1,614	<0,001	0,0004
3	IGLV3-25	1,700	<0,001	0,0074
4	IGHV3-9	1,186	<0,001	0,0508
5	SERPINF1	1,052	<0,001	0,0719
6	IGHG1	1,357	<0,001	0,0719
7	IGHV2-26	1,543	<0,001	0,1202
8	ITIH2	1,066	<0,001	0,1202
9	IGKV2-30	1,422	<0,001	0,1202
10	IGFALS	0,864	<0,001	0,1202
11	IGHA1	1,469	<0,001	0,1202
12	MRC1	1,017	<0,001	0,1202
13	IGLC3	1,320	<0,001	0,1202
14	MMP12	1,596	<0,001	0,1202
15	FETUB	0,907	<0,001	0,1361
16	CPN2	0,922	<0,001	0,1361
17	PECAM1	0,751	<0,001	0,1377
18	A0A0G2JRQ6	1,711	<0,001	0,1463
19	SLC35F2	1,444	<0,001	0,1463
20	THBS1	1,190	<0,001	0,1463
21	RTN2	1,292	<0,001	0,1480
22	VSIR	0,939	<0,001	0,1489
23	PLEK	1,828	<0,001	0,1489
24	IGLV1-51	2,661	<0,001	0,1489
25	SAA4	1,491	<0,001	0,1489
26	COL2A1	4,333	<0,001	0,1520
27	RBP4	1,502	<0,001	0,1520
28	FERMT3	1,124	<0,001	0,1520
29	IGLL5	1,097	<0,001	0,1530
30	C5	0,781	<0,001	0,1707
31	C1QC	1,703	<0,001	0,1805
32	HERC3	1,528	<0,001	0,1931
33	IGHV3-64	1,049	<0,001	0,1931
34	F2	0,873	<0,001	0,2079
35	FCGR2A	1,771	<0,001	0,2281
36	IGLV2-11	2,407	<0,001	0,2317
37	RCN3	0,741	<0,001	0,2369
38	GC	0,856	0,001	0,2369
39	STAB1	0,688	0,001	0,2464
40	FCGR1A	1,548	0,001	0,2464
41	MMRN1	0,705	0,001	0,2464

Summary of the most important proteins from the absence/presence analysis with a p-value < 0.05.

GENERAL - 74 NOREC vs 13 REC

Nº	ID	NOREC (74)		REC (13)		P-value	Nº	ID	NOREC (74)		REC (13)		P-value
1	ADAMTS8	9	12%	7	54%	0,002	51	DEFA6	4	5%	4	31%	0,016
2	ELAPOR2	13	18%	8	62%	0,002	52	CENPT	35	47%	11	85%	0,016
3	FGG	13	18%	8	62%	0,002	53	CBS	50	68%	13	100%	0,016
4	PLA2G4F	13	18%	8	62%	0,002	54	MEIS2	59	80%	6	46%	0,017
5	TMEM132C	13	18%	8	62%	0,002	55	ACY3	59	80%	6	46%	0,017
6	DUOX1	13	18%	8	62%	0,002	56	PCGF6	59	80%	6	46%	0,017
7	PLEK2	43	58%	13	100%	0,003	57	ULBP3	59	80%	6	46%	0,017
8	TNFRSF1A	19	26%	9	69%	0,004	58	PACC1	15	20%	7	54%	0,017
9	NR2C1	58	78%	5	38%	0,006	59	MIEF1	63	85%	7	54%	0,017
10	PRRC2C	32	43%	11	85%	0,007	60	GNG4	63	85%	7	54%	0,017
11	SESN1	46	62%	13	100%	0,008	61	SLC25A42	11	15%	6	46%	0,017
12	TAP2	17	23%	8	62%	0,008	62	MACF1	63	85%	7	54%	0,017
13	CCDC171	17	23%	8	62%	0,008	63	FIGNL1	20	27%	8	62%	0,023
14	RIC8B	48	65%	13	100%	0,008	64	CST6	12	16%	6	46%	0,024
15	SPATA33	68	92%	8	62%	0,010	65	FYB2	12	16%	6	46%	0,024
16	ANXA3	6	8%	5	38%	0,010	66	CEL	69	93%	9	69%	0,026
17	TPD52	68	92%	8	62%	0,010	67	VWA5B2	69	93%	9	69%	0,026
18	PLLP	68	92%	8	62%	0,010	68	UBAP2L	5	7%	4	31%	0,026
19	SDS	68	92%	8	62%	0,010	69	CRLF1	5	7%	4	31%	0,026
20	MEOX1	6	8%	5	38%	0,010	70	ALOX12	5	7%	4	31%	0,026
21	MARS2	68	92%	8	62%	0,010	71	GPC3	5	7%	4	31%	0,026
22	PUS3	68	92%	8	62%	0,010	72	BACH2	5	7%	4	31%	0,026
23	MVB12B	68	92%	8	62%	0,010	73	CCDC85B	43	58%	12	92%	0,026
24	MFSD1	39	53%	12	92%	0,012	74	TNFAIP1	50	68%	4	31%	0,027
25	HMMR	39	53%	12	92%	0,012	75	MAPK10	21	28%	8	62%	0,027
26	MACF1	40	54%	12	92%	0,012	76	B3GALNT2	21	28%	8	62%	0,027
27	SYNE3	64	86%	7	54%	0,012	77	COX7A1	53	72%	5	38%	0,027
28	SSH1	64	86%	7	54%	0,012	78	DTL	44	59%	12	92%	0,027
29	PCSK6	10	14%	6	46%	0,012	79	ARG2	44	59%	12	92%	0,027
30	ZNF197	10	14%	6	46%	0,012	80	ZNF518A	30	41%	1	8%	0,027
31	GAS2	10	14%	6	46%	0,012	81	ZNF552	44	59%	12	92%	0,027
32	MARCHF6	10	14%	6	46%	0,012	82	MFAP5	49	66%	4	31%	0,028
33	LRRC49	27	36%	10	77%	0,013	83	LPAR1	49	66%	4	31%	0,028
34	SLC19A1	23	31%	9	69%	0,013	84	ERN2	49	66%	4	31%	0,028
35	TMEM35A	28	38%	10	77%	0,014	85	PEX16	49	66%	4	31%	0,028
36	FRRS1	28	38%	10	77%	0,014	86	GMNN	45	61%	12	92%	0,030
37	MIGA2	28	38%	10	77%	0,014	87	OPA1	45	61%	12	92%	0,030
38	MAPK8IP1	46	62%	3	23%	0,014	88	ORMDL3	45	61%	12	92%	0,030
39	PCDH7	46	62%	3	23%	0,014	89	ZBTB43	45	61%	12	92%	0,030
40	ZNF408	46	62%	3	23%	0,014	90	ZNF513	26	35%	9	69%	0,031
41	DCAF12L1	34	46%	11	85%	0,014	91	RNF26	26	35%	9	69%	0,031
42	TUBG2	34	46%	11	85%	0,014	92	SHISA5	44	59%	3	23%	0,032
43	POLR3GL	32	43%	1	8%	0,015	93	UNC13B	30	41%	10	77%	0,032
44	NUDT7	7	9%	5	38%	0,015	94	PKDCC	44	59%	3	23%	0,032
45	ULK4	7	9%	5	38%	0,015	95	PRDM2	65	88%	8	62%	0,032
46	KCNN4	67	91%	8	62%	0,015	96	GPAT3	9	12%	5	38%	0,032
47	LGR5	67	91%	8	62%	0,015	97	PWWP2A	65	88%	8	62%	0,032
48	NFYA	67	91%	8	62%	0,015	98	MTRF1	43	58%	3	23%	0,033
49	SPIRE1	67	91%	8	62%	0,015	99	TAF11	43	58%	3	23%	0,033
50	XIRP1	67	91%	8	62%	0,015	100	GFOD1	52	70%	13	100%	0,033

EEC - 64 NOREC vs 9 REC

Nº	ID	NOREC (64)		REC (9)		P-value	Nº	ID	NOREC (64)		REC (9)		P-value
1	RASEF	17	27%	7	78%	0,005	51	DST	31	48%	8	89%	0,032
2	SMPX	17	27%	7	78%	0,005	52	POLM	31	48%	8	89%	0,032
3	MEIS2	51	80%	3	33%	0,008	53	RGS6	8	13%	4	44%	0,035
4	PCGF6	51	80%	3	33%	0,008	54	PCSK6	8	13%	4	44%	0,035
5	MFAP5	44	69%	2	22%	0,011	55	A0A087X1T7	56	88%	5	56%	0,035
6	TTL	36	56%	9	100%	0,011	56	ZNF33A	56	88%	5	56%	0,035
7	EAF1	28	44%	0	0%	0,011	57	ANKRD44	8	13%	4	44%	0,035
8	PLEK2	36	56%	9	100%	0,011	58	NADK2	8	13%	4	44%	0,035
9	DCAF12L1	28	44%	8	89%	0,014	59	OGT	8	13%	4	44%	0,035
10	EME1	49	77%	3	33%	0,014	60	CFAP300	8	13%	4	44%	0,035
11	AMOTL2	49	77%	3	33%	0,014	61	COPZ1	8	13%	4	44%	0,035
12	PHETA2	10	16%	5	56%	0,015	62	MRE11	8	13%	4	44%	0,035
13	SULT1B1	11	17%	5	56%	0,020	63	EIF5A2	8	13%	4	44%	0,035
14	PLA2G4F	11	17%	5	56%	0,020	64	MMADHC	56	88%	5	56%	0,035
15	DUOX1	11	17%	5	56%	0,020	65	HNRNPC	8	13%	4	44%	0,035
16	DTL	37	58%	9	100%	0,022	66	COPRS	8	13%	4	44%	0,035
17	SESN1	37	58%	9	100%	0,022	67	MYO1C	56	88%	5	56%	0,035
18	YIPF2	43	67%	2	22%	0,023	68	YTHDC1	8	13%	4	44%	0,035
19	RNF141	39	61%	9	100%	0,023	69	KLK10	56	88%	5	56%	0,035
20	ATP4A	25	39%	0	0%	0,023	70	CPE	56	88%	5	56%	0,035
21	ASAH1	22	34%	7	78%	0,025	71	GLDC	56	88%	5	56%	0,035
22	NUDT7	7	11%	4	44%	0,025	72	KHK	56	88%	5	56%	0,035
23	CAST	7	11%	4	44%	0,025	73	ZNF362	56	88%	5	56%	0,035
24	GSTA4	57	89%	5	56%	0,025	74	DYNLRB2	8	13%	4	44%	0,035
25	KCNN4	57	89%	5	56%	0,025	75	GEMIN6	56	88%	5	56%	0,035
26	LGR5	57	89%	5	56%	0,025	76	PDGFC	56	88%	5	56%	0,035
27	NFYA	57	89%	5	56%	0,025	77	INPP5E	8	13%	4	44%	0,035
28	SPIRE1	57	89%	5	56%	0,025	78	UBE2J1	56	88%	5	56%	0,035
29	XIRP1	57	89%	5	56%	0,025	79	MACF1	32	50%	8	89%	0,035
30	ULK4	7	11%	4	44%	0,025	80	SPRY4	32	50%	8	89%	0,035
31	ERP27	7	11%	4	44%	0,025	81	MFSD1	32	50%	8	89%	0,035
32	RIC8B	40	63%	9	100%	0,026	82	HMMR	32	50%	8	89%	0,035
33	CLASP1	24	38%	0	0%	0,026	83	PSIP1	32	50%	8	89%	0,035
34	NKAPD1	40	63%	9	100%	0,026	84	ELAPOR2	13	20%	5	56%	0,036
35	DOCK6	12	19%	5	56%	0,027	85	FGG	13	20%	5	56%	0,036
36	ZNF239	12	19%	5	56%	0,027	86	COL5A1	13	20%	5	56%	0,036
37	CCDC40	52	81%	4	44%	0,027	87	NR2C1	51	80%	4	44%	0,036
38	SLC25A19	52	81%	4	44%	0,027	88	NEK4	51	80%	4	44%	0,036
39	KLHL20	52	81%	4	44%	0,027	89	IGHV4-4	60	94%	6	67%	0,036
40	ZNF33B	41	64%	2	22%	0,028	90	A0A087WZT2	60	94%	6	67%	0,036
41	PCDH7	41	64%	2	22%	0,028	91	MTRES1	60	94%	6	67%	0,036
42	SPO11	23	36%	7	78%	0,028	92	ATXN7	60	94%	6	67%	0,036
43	ZUP1	23	36%	7	78%	0,028	93	E7EW20	60	94%	6	67%	0,036
44	HYKK	24	38%	7	78%	0,032	94	DDX55	60	94%	6	67%	0,036
45	ST7	40	63%	2	22%	0,032	95	SLBP	60	94%	6	67%	0,036
46	THBD	24	38%	7	78%	0,032	96	PABIR2	60	94%	6	67%	0,036
47	FGFBP1	40	63%	2	22%	0,032	97	CRISPLD2	60	94%	6	67%	0,036
48	PKDCC	40	63%	2	22%	0,032	98	MAP4	60	94%	6	67%	0,036
49	ANXA8L1	40	63%	2	22%	0,032	99	NFATC1	60	94%	6	67%	0,036
50	DGKB	33	52%	1	11%	0,032	100	MT-ND5	60	94%	6	67%	0,036

ANNEX 5

Peptides from the potential biomarkers in the verification phase

Peptides in bold were detected in the mass spectrometer and used for the subsequent analysis.

Uniprot Number	Gene name | Protein ID	Peptide Priority	Sequence
O15027	SEC16A | SC16A	1	NPSSAAPVQSR
		2	**GLANPEPAPEPK**
		3	**QALQSTPLGSSSK**
O43242	PSMD3 | PSMD3	1	**EQQDLEFAK**
		2	**AVQGFFTSNNATR**
		3	LQLDSPEDAEFIVAK
O43709	WBSCR22 | BUD23	1	EVRPDTQYTGR
		2	**AGFSGGMVVDYPNSAK**
		3	ESVFTNER
O43865	AHCYL1 | SAHH2	1	**TTDVMFGGK**
		2	**YSFMATVTK**
		3	EIEIAEQDMSALISLR
O94788	ALDH1A2 | AL1A2	1	**IFINNEWQNSESGR**
		2	**EMGEFGLR**
		3	LAFSLGSVWR
O94875	SORBS2 | SRBS2	1	TQTYRPLSK
		2	**NEDELELR**
		3	**SFTSSSPSSPSR**
O95777	LSM8 | LSM8	1	**TVAVITSDGR**
		2	**AEPLNSVAH**
		3	GFDQTINLILDESHER
P00915	CA1 | CAH1	1	**GGPFSDSYR**
		2	**YSSLAEAASK**
		3	ESISVSSEQLAQFR
P01833	PIGR | PIGR	1	NADLQVLKPEPELVYEDLR
		2	**TDISMSDFENSR**
		3	DGSFSVVITGLR
P04083	ANXA1 | ANXA1	1	**DITSDTSGDFR**
		2	**SEDFGVNEDLADSDAR**
		3	GLGTDEDTLIEILASR
P08133	ANXA6 | ANXA6	1	GTVRPANDFNPDADAK
		2	**SLEDALSSDTSGHFR**
		3	**DLEADIIGDTSGHFQK**
P10253	GAA | LYAG	1	AGYIIPLQGPGLTTTESR
		2	**VTSEGAGLQLQK**
		3	**WGYSSTAITR**
P10909	CLU | CLUS	1	LFDSDPITVTVPVEVSR
		2	**ELDESLQVAER**
		3	ASSIIDELFQDR
P12814	ACTN1 | ACTN1	1	AGTQIENIEEDFRDGLK
		2	ATLPDADKER
		3	**LAILGIHNEVSK**
P14618	PKM | KPYM	1	**LDIDSPPITAR**
		2	IYVDDGLISLQVK
		3	GSGTAEVELKK
P20933	AGA | ASPG	1	**NVIPDPSK**
		2	**FLPSYQAVEYMR**
		3	TGHIAAGTSTNGIK
P21980	TGM2 | TGM2	1	ALLVEPVINSYLLAER
		2	**TVEIPDPVEAGEEVK**
		3	**NEFGEIQGDK**
P26038	MSN | MOES	1	FYPEDVSEELIQDITQR
		2	**ALTSELANAR**
		3	**ESEAVEWQQK**
P32455	GBP1 | GBP1	1	**NEIQDLQTK**
		2	EAIEVFIR
		3	**GFSLGSTVQSHTK**
P35579	MYH9 | MYH9	1	**LVWVPSDK**
		2	**IAQLEEQLDNETK**
		3	EQADFAIEALAK
P36776	LONP1 | LONM	1	AQAVLEEDHYGMEDVK
		2	**LAQPYVGVFLK**
		3	**FSVGGMTDVAEIK**
P39060	COL18A1 | COIA1	1	**TEAPSATGQASSLLGGR**
		2	ADDILASPPR
		3	**LQDLYSIVR**
P51665	PSMD7 | PSMD7	1	DTTVGTLSQR
		2	**PELAVQK**
		3	**IVGWYHTGPK**
P52434	POLR2H | RPAB3	1	IEGDETSTEAATR
		2	**ADQFEYVMYGK**
		3	AGILFEDIFDVK
P53999	SUB1 | TCP4	1	**EQISDIDDAVR**
		2	**DDNMFQIGK**
		3	GISLNPEQWSQLK
P62306	SNRPF | RUXF	1	**GVEEEEEDGEMR**
		2	WGMEYK

Uniprot Number	Gene name | Protein ID	Peptide Priority	Sequence
P60468	SEC61B | SC61B	1	**FYTEDSPGLK**
		2	**TTSAGTGGMWR**
		3	PGPTPSGTNVGSSGR
P62263	RPS14 | RS14	1	ELGITALHIK
		2	**TPGPGAQSALR**
		3	DESSPYAAMLAAQDVAQR
Q01130	SRSF2 | SRSF2	1	DAEDAMDAMDGAVLDGR
		2	**YGGGGYGR**
		3	VGDVYIPRDR
Q13043	STK4 | STK4	1	ATATQLLQHPFVR
		2	**LGEGSYGSVYK**
		3	**AVGDEMGTVR**
Q13425	SNTB2 | SNTB2	1	**SPSLGSDLTFATR**
		2	**IFPGLAADQSR**
		3	GPAGEAGASPPVR
Q14005	IL16 | IL16	1	**GLPDPALSTQPAPASR**
		2	**LLSTQAEESQGPVLK**
		3	ISSFETFGSSQLPDK
Q14126	DSG2 | DSG2	1	**ILDVNDNIPVVENK**
		2	**IHSDLAEER**
		3	IVSLEPAYPPVFYLNK
Q15126	PMVK | PMVK	1	**EAYGAVTQTVR**
		2	**LLDTSTYK**
		3	VSDIQWFR
Q15417	CNN3 | CNN3	1	**DYQYSDQGIDY**
		2	**GFHTTIDIGVK**
		3	LTLQPVDNSTISLQMGTNK
Q15738	NSDHL | NSDHL	1	**THLTEDTPK**
		2	**QDLYPALK**
		3	AVLGANDPEK
Q16774	GUK1 | KGUA	1	DIAAGDFIEHAEFSGNLYGTSK
		2	**VAVQAVQAMNR**
		3	LAAAQADMESSK
Q86UX7	FERMT3 | URP2	1	VFVGEEDPEAESVTLR
		2	**TASGDYIDSSWELR**
		3	**LTQLYEQAR**
Q86X29	LSR | LSR	1	**LLEEAVR**
		2	**SGDLPYDGR**
		3	SSSAGGQGSYVPLLR
Q8N8S7	ENAH | ENAH	1	**LEQEQLER**
		2	WVPAGGSTGFSR
		3	**ASSTSTPEPTR**
Q8NFU3	TSTD1 | TSTD1	1	**GLQATQLAR**
		2	**SLGYTGAR**
		3	SLLASGR
Q92499	DDX1 | DDX1	1	GHVDILAPTVQELAALEK
		2	**APDGYIVK**
		3	**GIDIHGVPYVINVTLPDEK**
Q96QR8	PURB | PURB	1	YADEMKEIQER
		2	**GGGEQETQELASK**
		3	**IAEVGAGGSK**
Q9BRT3	MIEN1 | MIEN1	1	**EQYPGIEIESR**
		2	DLIEAIR
		3	**LENGGFPYEK**
Q9H1E3	NUCKS1 | NUCKS	1	**ATVTPSPVK**
		2	VGRPTASK
		3	VVDYSQFQESDDADEDYGR
Q9UHL4	DPP7 | DPP2	1	ASHPEDPASVVEAR
		2	**DVTADFEGQSPK**
		3	**LDHFNFER**
Q9UN81	L1RE1 | LORF1	1	**QANVQIQEIQR**
		2	SNYSELR
		3	**VSAMEDEMNEMK**
Q9Y3C1	NOP16 | NOP16	1	**FGYSVNR**
		2	**YMVENHGEDYK**
		3	AMEVDIEERPK
Q9Y490	TLN1 | TLN1	1	LNEAAAGLNQAATELVQASR
		2	**ALEATTEHIR**
		3	**VLVQNAAGSQEK**
Q9Y696	CLIC4 | CLIC4	1	NSRPEANEALER
		2	**HPESNTAGMDIFAK**
		3	**YLTNAYSR**
Q9UL15	BAG5 | BAG5	1	**YLDLEEEADTTK**
		2	**EVVEDINK**
P09493	TPM1 | TPM1	1	**QLEDELVSLQK**
		2	**MEIQEIQLK**

ANNEX 6

Publications in collaboration:

"Metabolomic 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. Sci Rep. 2017 Aug 18;7(1):8803. doi: 10.1038/s41598-017-09169-2. PMID: 28821813

"Advances in endometrial cancer protein biomarkers for use in the clinic". Martinez-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 Rev Proteomics. 2018 Jan;15(1):81-99. doi: 10.1080/14789450.2018.1410061. Epub 2017 Nov 30. PMID: 29183259

"Patient-Derived Xenograft Models for Endometrial Cancer Research". Moiola CP, **Lopez-Gil C**, Cabrera S, Garcia A, Van Nyen T, Annibali D, Fonnes T, Vidal A, Villanueva A, Matias-Guiu X, Krakstad C, Amant F, Gil-Moreno A, Colas E. Int J Mol Sci. 2018 Aug 17;19(8):2431. doi: 10.3390/ijms19082431. PMID: 30126113

"Therapeutic potential of the new TRIB3-mediated cell autophagy anticancer drug ABTL0812 in endometrial cancer". Felip I, Moiola CP, Megino-Luque C, **Lopez-Gil C**, Cabrera S, Solé-Sánchez S, Muñoz-Guardiola P, Megias-Roda E, Pérez-Montoyo H, Alfon J, Yeste-Velasco M, Santacana M, Dolcet X, Reques A, Oaknin A, Rodríguez-Freixinos V, Lizcano JM, Domènech C, Gil-Moreno A, Matias-Guiu X, Colas E, Eritja N. Gynecol Oncol. 2019 May;153(2):425-435. doi: 10.1016/j.ygyno.2019.03.002. Epub 2019 Mar 7. PMID: 30853360

"Endometrial Stromal Cells Circulate in the Bloodstream of Women with Endometriosis: A Pilot Study". Vallvé-Juanico J, **López-Gil C**, Ballesteros A, Santamaria X. Int J Mol Sci. 2019 Jul 31;20(15):3740. doi: 10.3390/ijms20153740. PMID: 31370190

Annexes

"Genomic Profiling of Uterine Aspirates and cfDNA as an Integrative Liquid Biopsy Strategy in Endometrial Cancer". Casas-Arozamena C, Díaz E, Moiola CP, Alonso-Alconada L, Ferreirós A, Abalo A, **Gil CL**, Oltra SS, de Santiago J, Cabrera S, Sampayo V, Bouso M, Arias E, Cueva J, Colas E, Vilar A, Gil-Moreno A, Abal M, Moreno-Bueno G, Muinelo-Romay L. J Clin Med. 2020 Feb 21;9(2):585. doi: 10.3390/jcm9020585. PMID: 32098121

"Small-Molecule Inhibitors (SMIs) as an Effective Therapeutic Strategy for Endometrial Cancer". Megino-Luque C, Moiola CP, Molins-Escuder C, **López-Gil C**, Gil-Moreno A, Matias-Guiu X, Colas E, Eritja N. Cancers (Basel). 2020 Sep 24;12(10):2751. doi: 10.3390/cancers12102751. PMID: 32987790

"External validation of putative biomarkers in eutopic endometrium of women with endometriosis using NanoString technology". Vallvé-Juanico J, **López-Gil C**, Ponomarenko J, Melnychuk T, Castellví J, Ballesteros A, Colás E, Gil-Moreno A, Santamaria Costa X. J Assist Reprod Genet. 2020 Dec;37(12):2981-2987. doi: 10.1007/s10815-020-01965-6. Epub 2020 Oct 9. PMID: 33033989

"Intratumor genetic heterogeneity and clonal evolution to decode endometrial cancer progression". Mota A, Oltra SS, Selenica P, Moiola CP, Casas-Arozamena C, **López-Gil C**, Diaz E, Gatius S, Ruiz-Miro M, Calvo A, Rojo-Sebastián A, Hurtado P, Piñeiro R, Colas E, Gil-Moreno A, Reis-Filho JS, Muinelo-Romay L, Abal M, Matias-Guiu X, Weigelt B, Moreno-Bueno G. Oncogene. 2022 Mar;41(13):1835-1850. doi: 10.1038/s41388-022-02221-0. Epub 2022 Feb 10. PMID: 35145232

"Genomic Validation of Endometrial Cancer Patient-Derived Xenograft Models as a Preclinical Tool". Villafranca-Magdalena B, Masferrer-Ferragutcasas C, **Lopez-Gil C**, Coll-de la Rubia E, Rebull M, Parra G, García Á, Reques A, Cabrera S, Colas E, Gil-Moreno A, Moiola CP. Int J Mol Sci. 2022 Jun 3;23(11):6266. doi: 10.3390/ijms23116266. PMID: 35682944

Annexes

ACKNOWLEDGEMENTS

Tot això de la ciència sempre m'ha agradat i la vaig començar a experimentar des de ben petit quan em van comprar un microscopi i un telescopi (m'agradava molt la astronomia, tot i que no volia ser astronauta perquè les alçades i jo no som massa amics). Però no va ser fins a batxillerat que realment em vaig interessar per la biologia, i tot això gràcies a un gran professor com el **Marc Bosch**. Moltes gràcies per tot el que m'has ensenyat, no només a nivell acadèmic, sino també a nivell personal. Vas fer que els dos anys de batxillerat es fessin més amens amb les teves classes, i sobretot amb les visites esporàdiques a les Illes Medes. En esta época también quiero destacar a todos Los Canteros (**David, Myriam, Kevin, Alca, Juanmi, Víctor y Gil** (el següent any sí o sí que hem de revalidar el títol del duro)). No sé cómo recordaréis vosotros esos dos años, pero para mi la verdad que han sido inolvidables. Es verdad que después con el tiempo cada uno ha hecho un poco su vida y que nos cuesta la vida juntarnos, pero cuando lo hacemos siempre acabamos hablando de las mismas chorradas como si hiciera unos días que hubieramos acabado bachillerato. En aquest punt també vull fer menció especial al **Oriol Asensio**. T'he de reconeixer que em vaig enfadar una mica quan després de batxillerat vas marxar a França a estudiar i desde aleshores ja t'has quedat fora. De totes maneres, sempre que has tornat has intentat fer una forat encara que sigui per veure'ns i fer com si tot seguís igual.

Después del batxillerato llegó la universidad y uno de los mejores años de mi vida donde me encontré a los Guais y Ana, que después con las todas las incorporaciones se acabó ampliando a The Real Group (**Ana Dorrego, David, Marina, Ana García, Cristina, Irene, Aroa, Iris y Nando**). Fueron 4 años bastante intensos ya que prácticamente estábamos casi todo el día juntos, al menos los del equipo biblioteca. Muchas anécdotas y viajes, la tira de viajes: Andorra, Ronda, Dublín, Alicante, la Rioja... E incluso después de haber acabado la carrera a Menorca, Bilbao, Francia... Hace ya 11 años que nos conocemos y aún nos aguantamos (aunque sea en la distancia), así que esto creo que ya va de verdad.

Una vez se acabó la carrera, en tan sólo un año de máster conocí a los Pomelos (**Andrés, Ana Dorrego, Julia, Alba, Ana García, Andrea y Roger**). La de cosas y chorradas que pueden pasar en un año, desde colarte en un centro de investigación en Montpellier hasta disfrazarse de mordisquitos para ir al carnaval. La mayoría ahora estáis lejos por trabajo o si estáis aquí con un mini-pomelito que cuidar y nos vemos poco, pero espero que de tanto en tanto nos sigamos pudiendo ver ¿próximo destino Berlín?

Ahora ya llegamos a estos últimos 5 años en el doctorado. En primer lugar, quiero agradecer tanto a **Eva Colás** como **Antonio Gil** la oportunidad que me dieron de poder

formar parte de este grupo. Todo empezó con un correo que le envié a Eva del todo informal preguntado por como iban las cosas en el laboratorio y que estaba buscando algún grupo con el que hacer el doctorado, y como quien no quiere la cosa aquí estamos. Después de dos embarazos y una pandemia ya se acaba esto. Muchas gracias por todo lo que me habéis enseñado tanto a nivel científico como a nivel personal. También quiero dar las gracias a todos los ginecólogos y patólogos que me han ayudado en estos años, en especial a **Armando** por todas las veces que te he pedido cosas de AP. Después están todas las personas que cuando llegué al grupo sin saber absolutamente nada, me hicieron un hueco y me ayudaron a seguir creciendo (**Nuria, Mireia, Elena, Laura, Irene, Blanca, Lucia, Tati, Alfonso y Gabriel**). També per la millor secretaria, psicóloga i professora que hi pot haver. Moltes gràcies **Eli** per tots els dies que hem estat al laboratori rient i parlant de xorrases. A **Marta** y **Leti**, aunque técnicamente no seais del grupo, en el fondo sabéis que sí. Leti, seguro que lo vas a petar o ya lo has petado con la tesis, i Marta, ja saps que en cas de dubte sempre et pots canviar i venir al nostre laboratori.

Durant tots aquests anys, hi ha dos persones que em van ajudar moltíssim. Per una banda, **Berta**, moltes gràcies per aquests anys que hem compartit al laboratori. Sembla que fos ahir que estavem els dos sols a l'estiu, jo a ratolins i tu fent extraccions d'exosomes mentre miraves la serie de TV3, i tu ara ja estás casada i parlando español a Mèxic !!! Per altra banda, **Julia**, moltes gràcies per alegrar-me tots els dies que vam compartir despatx i inclús després desde gairebé l'altre punta del món. Realment, m'has ensanyat moltíssimes coses i estic molt content d'haver-te conegut. També vull donar les gràcies a tots els estudiants que han anat passant pel laboratori (**Carlitos, Maria, Kaoutar, Paula, Marina, Sergio, Lydia y Sandra**) i també a la **Laia** (la més intel·ligent per deixar la investigació pública i marxar a la farmacèutica). Us desitjo el millor i espero que estigueu tots molt bé. I want also to thanks our international students (**Charlotte and Dana**). It has been only 3 and 6 months with you, but it has been a great pleasure to know you and I wish you the best in Copenhaguen and in Brisbrane. També vull donar les gràcies a dos nois que han passat poc temps pel lab: **Pau**, espero que tot et vagi bé i que trobis un altre lloc en la ciencia on estiguis més bé; **Javi**, ya sabes que cuando te canses de la empresa siempre te puedes venir con nosotros.

Ara arriba tota aquesta gent amb la que continuo treballant al laboratori i que m'han aguantat en aquests últims anys. **Cristian**, el segundo mejor argentino (lo siento pero Messi es el primero), has estado conmigo desde el primer día que empecé en el laboratorio hasta el último. Me has enseñado todo lo que sé en el estabulario, pero también a como se tiene que hacer un buen asado y qué carne hay que coger. **Marta**,