Universitat Autònoma de Barcelona

MOLECULAR ALTERATIONS IN EUTOPIC ENDOMETRIUM OF WOMEN WITH ENDOMETRIOSIS AND IMPLICATIONS IN ITS DIAGNOSTIC

Júlia Vallvé Juanico

MOLECULAR ALTERATIONS IN EUTOPIC ENDOMETRIUM OF WOMEN WITH ENDOMETRIOSIS AND IMPLICATIONS IN ITS DIAGNOSTIC

International PhD Thesis presented by

Júlia Vallvé Juanico

To obtain the degree of

PhD for the Universitat Autònoma de Barcelona (UAB)

PhD thesis carried out at IVI Barcelona S.L. and Biomedical Research Group on Gynecology, at Vall d'Hebron Institute of Research (VHIR), under the supervision of **Dr. Xavier Santamaria Costa and Dr. Antonio Gil Moreno**

Thesis affiliated to the Department of Cell Biology, Physiology and Immunology from the School of Medicine at UAB, in the PhD program of Cell Biology, under the tutoring of **Dr. Joaquima Navarro Ferreté**

Universitat Autònoma de Barcelona, September 18th 2018

Dr. Xavier Santamaria Costa
(Director)

Dr. Antonio Gil Moreno
(Director)

Dr. Joaquima Navarro Ferreté
(Tutor)

Júlia Vallvé Juanico
(Student)

“A la meva mare”

Acknowledgements

Per fi ha arribat el dia! Ha sigut llarg, però ha valgut molt la pena. Em sento orgullosa d'haver arribat fins aquí i si ho he pogut fer ha sigut gràcies a moltes persones.

Primer de tot, vull agrair al **Jaume Reventós** que em guiés en la meva vida professional. Gràcies a tu Jaume vaig aconseguir fer l'entrevista amb el Xavi, em vau acollir al lab 209 i aquí és on va començar tot. Recordo els inicis amb molta emoció. Tenia moltíssimes ganes de començar el doctorat i mai havia treballat a un hospital, cosa que trobava fascinant i m'ha obert un món nou. **Marina i Eva** moltes gràcies per a guiar-me un cop vaig entrar al laboratori. **Gabriel**, moltíssimes gràcies per ser sempre allà des del principi i tenir-me en compte per a tot. Crec que amb el dia a dia va néixer una gran amistat, de la qual estic molt contenta que, encara que no segueixis al lab, seguim tenint. Això ha passat amb vàries persones, la veritat, perquè amb el temps ha passat moltíssima gent pel lab. **Raquel**, tu ets una d'aquestes persones i saps que t'estimo moltíssim. Noies, m'ha encantat conèixer-vos a totes. Hem viscut mil experiències juntes i les recordaré sempre; calçotades, sopars de predocs, cases rurals...**Lucia, Elena, Blanca, Lau, Tati, Ire, Mire, Eli** sou estupendes i sóc molt afortunada d'haver-vos conegut! **Blanca**, moltíssimes gràcies per ajudar-me científicament tantes vegades, per no mencionar totes les estones que hem passat juntes fent coses fora del lab. Ara que comences la teva nova etapa fora d'aquí, se't trobarà a faltar! **Tati**, la meva puteta, amb tu he viscut tantes històries super divertides que seria impossible resumir-les. Quan tornis de fer la ruta de la teva vida segur que seguirem fent mil coses i això em fa feliç. I com no, t'he d'agrair que ens donessis la coseta més bonica del món a mi i al Pol; el nostre estimat Curri-flu. T'estimo puteta. **Berta**, pivon de faldilleta curta, ets preciosa i segur que t'anirà tot super mega ultra bé a partir d'ara en la teva nova etapa fora del lab. Et trobarem a faltar! **Mire**, moltes gràcies per ser com ets. En molts moments m'he sentit identificada amb tu i ens hem donat suport mutu. M'ha encantat compartir el despatx de *les sueltes* amb tu. **Cristian**, también me ha encantado conocerte a ti! Siempre contento y tan optimista. Gracias por aguantar mis palizas, organizar asados y ser tan majo. A tots els nous predocs que heu anat arribant al lab, **Nuria, Alfonso, Leti,**

Melissa, Eva i Carlos, que bé que vau arribar i vau omplir el lab de nou després de la fuga de les noies! Sou tots uns cracks i tots podeu amb la tesi! **Carlos**, a tu t'he d'agrair especialment que m'hagis ajudat amb tota la feina que has fet mentre jo estava a San Francisco. **Alf**, mi ex-compi de poyata, me encanta que pase lo que pase, siempre tengas esa sonrisa tan mona en la cara y que siempre me animes con todo. Eres estupendo!

A les jefes: **Anna**, gràcies per estar allà quan he necessitat consell. **Eva** gràcies per haver-me acollit al grup d'endometri i per sempre ajudar-me amb les presentacions, etc. I a les dues, gràcies per tenir-me sempre en compte i fer-me sentir part del grup. **Xavi**, tu ets la persona que em va brindar la oportunitat de fer el doctorat i per això t'he d'agrair moltíssim. Han passat uns quants anys i hem passat de tot; redacció de projectes, justificacions, congressos...tot i les dificultats que suposava que jo estigués al VHIR i tu a IVI, crec que ens en hem sortit prou bé. A més a més, gràcies a tu vaig poder anar a San Francisco i això m'ha canviat la vida. No només per la gran experiència que he viscut durant aquets mesos, sinó perquè se m'ha obert una nova porta per a la meva carrera investigadora. Gràcies.

A les meves **guais**...no sé ni per on començar. **Irene, Tzabet**, sou el tresor més gran que m'enduc d'aquest lab. Hem compartit tants riures, plors, sessions de psicologia, abraçades de lavabo, etc, etc, etc. En fi, ja ho sabeu, us estimo moltíssim i no hi ha paraules per expressar el que sento per vosaltres. I a més, ara tindrem una nova membre al nostre grup! Benvinguda **mini Júlia** i felicitats **Eli**.

No em podria oblidar de tothom de l'edifici Collserola: responsables de coordinació i cel·ladors, gràcies per fer-nos la vida més fàcil a tots. I com no, menció especial al **Ferran** de la UEB, les vam passat canutes amb tots els resultats de high-seq, però al final ens en vam sortir! A més, el fet de conèixer als cirurgians, ginecòlegs, infermers, patòlegs etc. ha sigut una experiència molt enriquidora; **Elena, Antonio i Taya**, gracias por toda vuestra ayuda, sin la cual esta tesis no se podría haber llevado a cabo. **Josep**, moltíssimes gràcies per estar sempre allà i ser tan eficient.

I, per suposat, vull donar les gràcies a totes les **pacients** que han participat a l'estudi i que sense elles hauria estat impossible realitzar.

People from SF, thank you so much for everything. First of all, **Dr. Giudice**, I would like to thank you that you accepted me in your lab. I have not only learned a lot working with you, but now you give me the opportunity to continue professionally growing in your lab. Thank you for trusting me. **Sahar**, I am very glad to have met you. You are a huge scientist and I have learned a lot from you. **Juan and Evelina**, thank you also for being there always that I needed something in the lab. Finally, **Kim Chi**, you are the best! I cannot imagine the lab without you, you are the joy of the lab and you know it. Please, don't change ever. And of course, my friends from SF: **Sara, Alex, Hector, Helena, Miquel, Pere, Royce, Enrico, Nicole, Jasper, Souhail, Morgane, Carolin, etc.** You are the best friends that I could ever find abroad. I have enjoyed every moment that we have shared, going to the park, to the mountain, road trips...Truly, thank you all of you for all these experiences.

Com no, no em podia oblidar de tots els meus estimats amics de la uni: **Jose, Alba, Ruben, Irene, Cristi, Miquel, Marta...** i en especial **Álvaro i Aurora**. Sempre m'heu animat a seguir endavant i a seguir els meus passos com a investigadora i us ho agraeixo molt. Us estimo moltíssim a tots. També gràcies als **calderins** per a fer l'escriptura de la tesi més amena! **Purora**, què t'he de dir a tu, ho hem compartit tot: la guarderia, primària, institut, batxillerat, carrera, nevodetes, pis durant el doctorat i ara mira'ns, les dues patint amb la tesi, però això ja ho tenim fet! Ets la millor amiga que mai podria haver tingut. T'estimo amb tot el meu cor. **Saidak**, mi niña bonita. A ti también te tengo que agradecer muchísimo. Encara que haguem seguit camins tan diferents, sempre m'has entès i m'has recolzat amb totes les meves decisions. Ets la millor! I ara una super mami! T'estimo molt!

Pol, carinyu, a tu sí que t'he d'agrair coses. Sempre m'has fet triar les bones decisions i mai m'has tallat les ales. Moltes gràcies per recolzar-me sempre amb la meva carrera professional, això no ho aguantaria tothom i tu, sempre tan racional, m'animes a que

marxi perquè saps que és el millor per a mi. Això sí, no t'acostumis a estar sol que quan torni m'hauràs d'aguantar molt de temps! Moltes gràcies per haver-me acompanyat durant aquest camí i per tot amor meu. T'estimo moltíssim.

Gràcies a tota la meva família. Sempre m'heu recolzat amb tot el que he fet i m'heu animat a tirar endavant. Encara que me n'hagi anat lluny i haguem estat separats sempre esteu aquí, no se què faria sense vosaltres. **Papa**, tu em veus amb ulls de papi orgullós i sempre em dius que sóc una crack, t'ho agraeixo. En realitat tots sabem que la crack de la family és l'**Afri**. No em podria imaginar la vida sense tu, sempre juntes. Ets la millor germana que podria haver tingut mai i t'estimo amb bogeria. **Paolo i Mohikano**, m'encanta ser la vostra germana, sempre molestant-me i fent-me riure. Gràcies a vosaltres també per recolzar-me sempre. **Jose y Rocío**, estic contentíssima que formeu part de la nostra família. Jose thank you for being always so optimistic and give me support. I com no, els meus petits tresors, **Maia i Nora**, només d'estar al vostre costat ja sóc feliç. **Maia**, gràcies per entendre que la tieta no pot estar sempre amb tu perquè ha de treballar i perquè marxa a l'altre punta del món a seguir treballant. Sort que ara amb les noves tecnologies podem parlar i veure'ns molt més. Us estimo precioses. **Mama**, a tu sí que t'ho dec tot. Moltes gràcies per insistir sempre tant en que estudiéssim. Gràcies per fer tot el possible per a que pogués anar a la uni, per cuidar-me, escoltar-me, entendre'm, estar al meu costat i animar-me a seguir endavant. Gràcies per tots els teus consells. Gràcies per ser la meva amiga i confident, mai et podré agrair tot el que has fet per tots nosaltres. És per tot això i molt més que aquesta tesi te la dedico a tu. T'estimo moltíssim.

Index

INDEX

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS AND ACRONYMS GLOSSARY

Symbols	
µl	microliters
µm	micrometers
A	
ACTB	Actin Beta
ADA2	Adenosine deaminase 2
Adeno	Adenomyosis
ANG2	Angiotensin 2
Arg1	Arginase 1
aScidea	aScidea Computational S.L.
AUC	Area Under the Curve
B	
BMI	Body Mass Index
BSA	Bovine Serum Albumin
C	
cAMP	Cyclic adenosine monophosphate
CCL1	C-C motif chemokine 1
CCL11	C-C motif chemokine 11
CCL16	C-C motif chemokine 16
CCL17	C-C motif chemokine 17
CCL18	C-C motif chemokine 18
CCL22	C-C motif chemokine 22
CCL24	C-C motif chemokine 24
CCL5	C-C Motif Chemokine Ligand 5
CD10	Common acute lymphoblastic leukemia-associated
CD11b	Integrin alpha M
CD127	Cluster of Differentiation 127
CD133	Cluster of Differentiation 133
CD14	Cluster of Differentiation 14
CD16	Cluster of Differentiation 16
CD163	Cluster of Differentiation 163
CD180	Cluster of Differentiation 180
CD19	Cluster of Differentiation 19
CD1A	Differentiation Antigen CD1-Alpha-1
CD1B	Cluster of Differentiation 1 b
CD1C	Differentiation Antigen CD1-Alpha-3
CD200R1	CD200 Receptor 1
CD209	Cluster of Differentiation 209
CD25	Cluster of Differentiation 25
CD3	Cluster of Differentiation 3
CD300A	Cluster of Differentiation 300A
CD300C	CD300 antigen-like family member C
CD300E	Cluster of Differentiation 300E
CD300LB	CD300 Molecule Like Family Member B
CD300LF	CD300 Molecule Like Family Member F

CD31	Cluster of Differentiation 31
CD33	Cluster of Differentiation 33
CD34	Cluster of Differentiation 34
CD36	Cluster of Differentiation 36
CD4	Cluster of Differentiation 4
CD45	Cluster of Differentiation 45
CD53	Cluster of Differentiation 53
CD56	Cluster of Differentiation 56
CD68	Cluster of Differentiation 68
CD69	Cluster of Differentiation 69
CD74	Cluster of Differentiation 74
CD80	Cluster of Differentiation 80
CD84	Cluster of Differentiation 84
CD86	Cluster of Differentiation 86
CD93	Cluster of Differentiation 93
cDNA	Complementary DNA
CECs	Circulating Endometrial Cells
CK	Cytokeratin
CRG	Center of Genomics Regulation
CRH	Corticotropin Releasing Hormone
CSF-1	Colony Stimulating Factor 1
CTCs	Circulating Tumor Cells
CX3CL1	chemokine (C-X3-C motif) ligand 1
CXCL10	C-X-C motif chemokine 10
CXCL11	C-X-C Motif Chemokine Ligand 11
CXCL12	C-X-C Motif Chemokine Ligand 12
CXCL13	chemokine (C-X-C motif) ligand 13
CXCL16	C-X-C Motif Chemokine Ligand 16
CXCL9	C-X-C Motif Chemokine Ligand 9
CYP19A1	P450aromatase
CYR61	Cysteine Rich Angiogenic Inducer 61
	D
DAPI	6-diamino-2-phenylindole
DEFB103A	Defensin Beta 103A
DEFB4A	Defensin Beta 4A
DEG	Differentially Expressed Genes
DET	Differentially Expressed Transcripts
DIE	Deep Infiltrating Endometriosis
DKK1	Dickkopf-related protein 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	Desoxyribonucleic Aceid
dNTPs	Deoxynucleotidphosphates
DUSP1	Dual Specificity Phosphatase 1

E	
E2	Estrogen
ECAD	E- Cadherin
EDGE	Empowering the Development of Genomics Expertise
EDTA	Ethylenediamine tetra-acetic acid
EMT	Epithelial Mesenchymal Transition
EpCAM	Epithelial Cell Adhesion Molecule
ERG	ETS transcription factor
eSF	Endometrial Stromal Fibroblasts
F	
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal bovine Serum
FC	Fold Change
FDA	Food and Drug Administration
FDR	False Discovery Rate
FFPE	Formalin Fixed Paraffin Embedded
FIM	Fluorescence Intensity Mean
FMO	Fluorocroms Minus One
FOS	Fos Proto-Oncogene
Foxp3	Forkhead Box P3
FPKM	Fragments Per Kilobase Million
FSH	Follicular Stimulatory Hormone
G	
GADD45G	Growth Arrest and DNA Damage Inducible Gamma
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GB	Gigabytes
GC	Guanine/Cytosine
GEO	Gene Expression Omnibus
GFF3	General Feature Format 3
GFP	Green Fluorescence Protein
GGT8P	Gamma-glutamyltransferase 8 pseudogene
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GnRH	Gonadotropin Releasing Hormone
GPR49	G-protein coupled receptor 49
GWAS	Genome Wide Association Studies
H	
H&E	Hematoxylin and Eosin
HSD17B2	Hydroxysteroid 17-Beta Dehydrogenase 2
HSPA1A	Heat Shock Protein Family A Member 1A
ICC	Immunocytochemistry
ID	Identification
I	
IF	Immunofluorescence
IHC	Immunohystochemistry
IL10	Interleukin 10
IL15	Interleukin 15

IL2	Interleukin 2
INFα	Interferon alpha
INHBA	Inhibin A
IPA	Ingenuity Pathway Analysis
IVI	Infertility Valencian Institute
	K
KEGG	Kyoto Encyclopedia of Genes and Genomes
KNN	K-nearest neighbors
	L
LGR5	Leucine-rich Repeat Containing G protein-Coupled Receptor 5
Library-Prep	Library preparation
LIF	Leukemia inhibitory factor
LINC00882	Long Intergenic Non-Protein Coding RNA 882
LOOCV	Leave-one-out-cross validation
LOX	Lysyl Oxidase
LPS	Lipopolysaccharide
LRR	Leucine-rich repeats
	M
miR-483-3p	microRNA-484-3p
MMP11	Matrix Metallopeptidase 11
MMP7	Matrix Metallopeptidase 7
MPA	Medroxiprogesterone
MRPL19	39S ribosomal protein L19
MUC5B	Mucin 5B
Mφ	Macrophages
	N
N	Nitrogen
NCBI	National Center for Biotechnology Information
ND	Non Determined
NFkB	Nuclear Factor Kappa B Subunit 1
ng	nanograms
NGS	Normal Goat Serum
NH4Cl	Ammonium chloride
nt	nucleotides
	O
OCPs	Oral contraceptives
ON	Overnight
	P
P4	Progesterone
PBS	Phosphate Buffered Saline Solution
PCA	Principal Component Analysis
PJA2	Praja Ring Finger Ubiquitin Ligase 2
PLS	Partial Least Square Regresion
PPM1D	Protein phosphatase 1D
PRKCE	Protein kinase C epsilon type
PSMC4	Proteasome 26S Subunit, ATPase 4

	Q
QDA	Quadratic discriminant analysis
	R
rASRM	revised American Society to Reproductive Medicine
RF	Random Forest
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA-Seq	RNA High-Sequencing
ROC	Receiving Operating Curve
RPLP0	Ribosomal Protein Lateral Stalk Subunit P0
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	Room Temperature
RT-qPCR	Real Time-quantitative PCR
	S
SATB1	Special AT-Rich Sequence Binding Protein 1
SCM	Serum Containing Medium
SERPINE1	Serpin Family E Member 1
SF3A1	Splicing Factor 3a Subunit 1
SLPI	Secretory Leukocyte Peptidase Inhibitor
SSC	Somatic Stem Cell
SVM	Support Vector Machine
	T
TAM	Tumor Associated Macrophages
TFF3	Trefoil Factor 3
TGFB1	Luteinizing Hormone
TGFβ	Transforming Growth Factor-beta
TLR4	Toll-like receptor 4
TNF	Tumor Necrosis Factor
TNFα	Tumor Necrosis Factor α
Treg	Regulatory T cells
	U
UAB	Universitat Autònoma de Barcelona
UCSF	University of California San Francisco
UEB	Unit of Bioestatistics and Bioinformatics of Vall d'Hebron
uNK	uterine Natural Killers
	V
VCAN	Versican
VEGF	Vascular Endothelial Growth Factor
VHIO	Vall d'Hebron Institute of Oncology
VHIR	Vall d'Hebron Institure of Research

Introduction

1. INTRODUCTION

1.1. GENERAL CONCEPTS OF ENDOMETRIOSIS

DEFINITION AND EPIDEMIOLOGY

Endometriosis is a chronic estrogen-dependent inflammatory disease that results in acute pelvic pain and/or infertility/subfertility. It affects 10% to 15% of reproductive age women (around 176 million women and teens worldwide)[1,2]. It is characterized by the presence of endometrial-like tissue outside its normal location lining the uterus, typically on pelvic structures but rarely at sites distant from the pelvis[1].

Structurally, human endometrium is composed of two layers: the basalis and the functionalis, the latter of which is shed during menstruation and regenerated in the absence of pregnancy. The endometrium comprises two compartments: epithelium and stroma. Tubular glands reach from the endometrial surface through to the base of the stroma. The stroma is a layer of connective tissue that varies in thickness according to cyclic hormone changes in the menstrual cycle.

The menstrual cycle **(Figure 1)** consists of three phases: proliferative, secretory, and menstrual. The morphology of the endometrium changes throughout the cycle, largely under the influence of ovarian-derived estradiol and progesterone. Early in the cycle (days 1-4), the tissue begins regenerating independent of estradiol. Throughout the proliferative phase, estradiol induces mitosis of all cellular constituents, including luminal and glandular epithelium, stromal fibroblasts, and vascular components. After ovulation, progesterone induces secretory transformation of the epithelial cells and stromal fibroblast differentiation (decidualization), preparing for embryo nidation. In the absence of implantation and prompted by the decline of steroid hormones, desquamation of the tissue ensues. This results in menses, followed by regeneration of the tissue from endometrial stem cells[3].

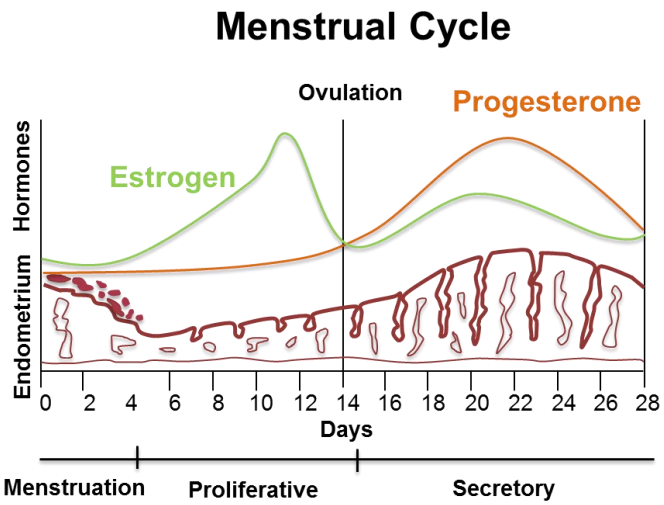


Figure 1. The menstrual cycle. Y-axis: Structure of endometrium and ovarian hormones that acts on endometrium. X-axis: days and phases of the cycle. (*Adapted from Houshdaran, Biol.Reprod., 2016)*

The prevalence of endometriosis in patients with infertility ranges between 25 and 50%, representing around 10-15% of the indications in Assisted Reproduction Techniques (ART). Around 30-50% of women with endometriosis are infertile[4]. Endometriosis is the most common pathology in teenagers who experience pelvic pain, diagnosed in 30% of cases. The prevalence of the disease in teenagers who undergo laparoscopy due to pelvic pain is around 50-60%[5].

Microscopically, endometriosis is characterized by the presence of endometrium, including both epithelial glands and stroma in the ectopic lesion. The epithelial glands are often irregular. It has been observed that sometimes the glands suffer from cyclic changes as they occur in the eutopic endometrium during the secretory phase. Sometimes they look like the basal layer of the endometrium, which does not suffer secretory changes[6].

Classically, adenomyosis endometriosis (internal endometriosis) has been differentiated from the rest of endometriosis (external endometriosis). Different authors affirm that there are three types of endometriosis: peritoneal endometriosis, ovarian endometriosis (endometrioma), and deep infiltrating endometriosis (DIE)[4]. Peritoneal endometriosis can contain typical or atypical lesions. The first are located in serous surfaces such as ovaries

and peritoneum. The formed cysts are small and a brownish or bluish color. The atypical lesions have ruddy lesions and serous vesicles **(Figure 2)**[1]. Microscopically, peritoneal endometriosis is characterized by the absence of inflammatory infiltration, showing fibrosis and mesothelial hyperplasia. Endometriomas are the most representative manifestation of ovarian endometriosis. They are located in the ovaries and, usually, they adhere in the peritoneum **(Figure 2)** [1]. Histologically, it presents as a small portion of glandular epithelium and stroma but a lot of fibrosis. Deep infiltrating endometriosis is characterized by an invasion of more than 5mm under the peritoneum[4]. Microscopically, the lesions are characterized by the presence of glandular epithelium and stroma. In >95% of cases, DIE is associated with very severe pain and is probably a cofactor in infertility.

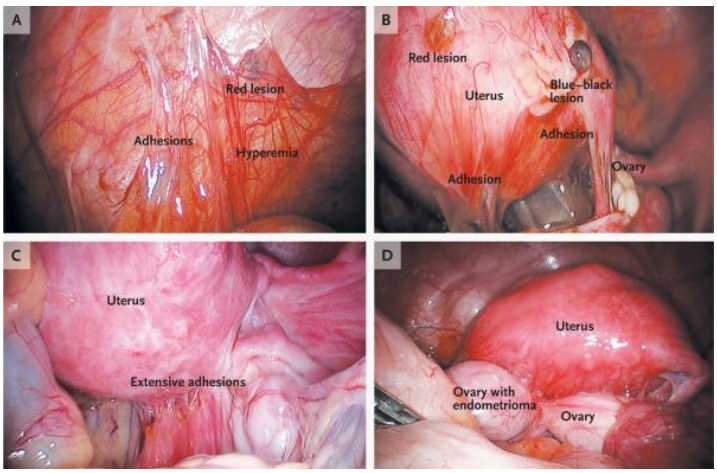


Figure 2. Peritoneal lesions and endometrioma. A: A red lesion, adherences, and hyperemia of the peritoneum. **B:** Red and black peritoneal lesions and adherences. **C:** Extensive adhesions that produce distortion of the normal pelvic anatomy. **D:** Ovary with endometrioma. (*Giudice, 2010*)[1].

On the other hand, the presence of endometrial tissue in the myometrium is diagnosed as adenomyosis. Histologically, it presents as ectopic endometrial glands and stroma surrounded by hypertrophic or hyperplasic myometrium. It is considered as the result of anomalous growth and invagination of the basal endometrium in the sub-endometrial myometrium[4].

The most common locations of endometriosis are the pelvic organs and peritoneum. However, atypical locations have been described, such as the lungs, belly button, eyes, or

brain. The extension of the disease in each location is highly variable. It can be characterized by the presence of small surface lesions, large ovarian cysts, inflammatory infiltration, and intense fibrosis and adhesions. **Table 1** shows the most frequent and infrequent locations of the disease.

Anatomical distribution of endometriosis	
Frequent	**Infrequent**
Cervix	Belly button
Fallopian tubes	Bladder
Intestines	Brain
Ovaries	Kidney
Pelvic linfatic nodes	Lungs
Pelvic peritoneum	Scars
Uterine ligaments	Ureter
Vagina	

Table 1. Anatomical distribution of endometriosis. In the left column, the most frequent sites of endometriotic lesions are listed. The most infrequent sites are listed in the right column.

PATHOPHYSIOLOGY OF ENDOMETRIOSIS

Several theories have been proposed to explain the origin of the disease, although no consensus has been reached in a single theory. Below, the most commonly proposed theories are explained.

Sampson's theory - retrograde menstruation: For many years, the most accepted theory has been the Sampson's hypothesis, which explains that the disease could be caused as a consequence of retrograde menstruation[7,8]. It states that at the time of menses, endometrial cells and tissue fragments reflux through the fallopian tubes, attach to and invade pelvic structures, undergo neuroangiogensis, and elicit a local inflammatory response[9]. Importantly, more than 90% of women have some degree of retrograde menstruation, but only 6-15% have endometriosis[1], suggesting aberrations in the clearance of ectopic endometrial tissue and cells by the immune system in women with the disease versus those without[10]. Despite different lines of evidence favoring this theory, cases of endometriosis in the lungs, eyes, or brain are not explained by this hypothesis. In addition,

the presence of endometriosis has also been reported in pre-menarchal girls, newborns, and males. Thus, there must be an explanation other than retrograde menstruation in those cases[11].

Coleomic metaplasia: Some groups have proposed that endometriosis could be derived from coelomic metaplasia[12]. This theory postulates that endometriosis arises from metaplasia cells lining the visceral and abdominal peritoneum due to multiple stimuli, hormonal or infectious. It explains that there might be a metaplasic process by which the coleomic epithelium develops into endometrial glandular cells. The coelomic epithelium develops into peritoneum (pelvic, thoracic and abdominal), pleura, Mullerian ducts, and the surface of the ovary. Therefore, there is a chance that, by multiple stimuli, these cells could develop the disease[13]. However, metaplasia increases with age in different organs, so this theory does not explain the drastic decrease of endometriosis in postmenopausal women[11].

Embryonic rests: Similar to the previous theory, the embryonic rest theory proposes that the lesions arise from cells remaining from Mullerian duct migration during embryonic development following a specific stimulus, likely hormonal, that could develop endometriosis[13]. This hypothesis could explain the presence of endometriosis in any location along the migration pathway of the embryonic Mullerian system. In addition, this theory could explain the presence of endometriosis in men, because the embryo initially develops female structures that regress with the male genome activation. However, this theory is purely hypothetical[13].

Lymphatic or vascular spread: Sampson suggested that endometrial cells could spread via vascular or lymphatic spread, which would explain the presence of endometriosis in distant locations such as the brain, lymph nodes, or other organs. This theory is supported by the presence of endometrial tissue in uterine blood vessels in patients with adenomyosis[8]. Although this theory could contribute to the pathogenesis of endometriosis, it is not likely to be the principal mechanism because the incidence of endometriosis outside of the peritoneal cavity is low[13].

Stem cells theory: Recent studies suggest that stem cells from bone marrow or from eutopic endometrium could differentiate into endometriotic tissue[14]. Like the lymphatic or hematogenous theory, this hypothesis explain the presence of endometriosis in distant locations as well as the presence of the disease in patients without menstrual endometrium, such as individuals with Rokitansky-Kuster-Hauser syndrome or vaginal agenesis that results in a missing uterus[15].

To date, no theory of pathogenesis can account for all of the described manifestations of endometriosis. The most accepted hypothesis has always been retrograde menstruation. However, interest in the stem cells theory has gained strength in recent years. Furthermore, the exact factors responsible for the implantation and survival of the ectopic endometrium remain unknown. Following Sampson's theory, eutopic endometrial cells could migrate from the uterus to the peritoneal cavity. However, 90% of women suffer from retrograde menstruation and do not develop the disease. It has been postulated that immune cells such as macrophages (Mφ) are responsible for the clearance of endometrial cells and cell debris shed during menstruation are not functional in diseased women, so eutopic endometrial cells are able to migrate to the peritoneal cavity[16]. Furthermore, a defective immune system in the peritoneal cavity would lead to the implantation and development of lesions[15]. The higher survival of those cells could also be due to a decrease in the apoptotic function in the eutopic and ectopic endometrium of women with endometriosis caused by the overexpression of the anti-apoptotic gene B-cell lymphoma (BCL-2) in comparison to healthy women[17]. Furthermore, vascular factors are also related to the development of the disease. To maintain the ectopic lesions, a process of angiogenesis has to occur. It is known that in patients with endometriosis, neutrophils and some molecules related to angiogenesis are increased. Indeed, the vascular endothelial growth factor (VEGF) is the major angiogenic factor present in ectopic lesions or peritoneal fluid[18] and in eutopic endometrium of women with endometriosis[19], and would be involved in the development and maintenance of ectopic lesions.

Regarding the implication of the immune system in ectopic sites, it is not clear if immune populations are responsible or otherwise implicated in the development of the lesions because of promoting a favorable environment for the implantation of endometrial cells or if the inflammation present in the peritoneal fluid and lesions are due to the disease. Either way, inflammation is a necessary characteristic for the development of the disease, but it also justifies much of endometriosis symptomatology.

Finally, it is believed that the hormonal environment is altered in women with endometriosis. It is well known that endometriosis is an estrogen-dependent disease. It was reported that there is an overexpression of P450aromatase (CYP19A1) in the ectopic lesions and a decrease of Hydroxysteroid 17-Beta Dehydrogenase 2 (HSD17B2), which indicates that estradiol would be higher in ectopic sites and could stimulate lesion development[20] **(Figure 3)**.

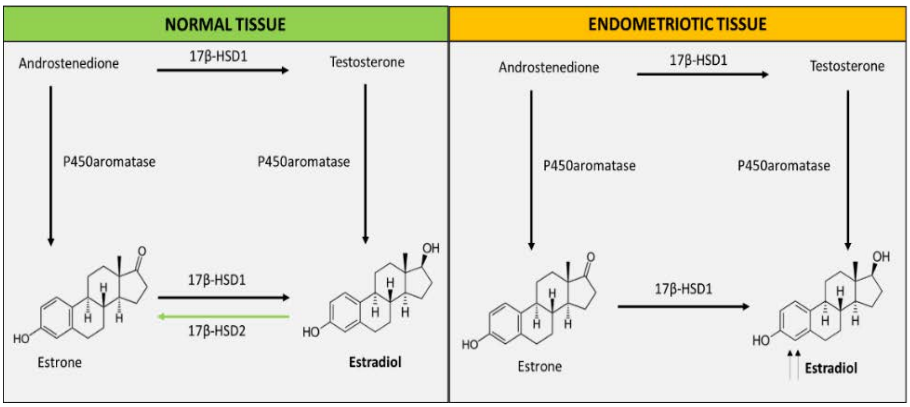


Figure 3. Estradiol formation. As can be observed in the figure, a lack of HSD17B2 increases the production of estradiol in ectopic endometriosis.

ETIOLOGY OF ENDOMETRIOSIS

Endometriosis is considered a multifactorial disease, and many factors are associated with the risk of suffering from it. Here, commonly studied risk factors are briefly explained.

RISK FACTORS

- **AGE:** It is not common to find endometriosis before the menarche, and it tends to decrease after menopause. Therefore, it is a pathology that appears mainly during reproductive age[21].
- **HERITABILITY:** Although the risk of suffering endometriosis is higher in women whose relatives have suffered from the disease, a mendelian pattern (single mutation) has not been yet described. It is postulated that endometriosis could have a multifactorial polygenic heritance[21]. It could be that the heritability of the disease is complex, which means that multiple genetic variants are involved in its manifestation. Eight genome-wide association studies (GWAS) have been done where the whole genome is studied, and it has been found that stages III/IV of endometriosis (the most severe ones) could be differentiated from stages I and II according to genetic variants, but so far only 5% of the risk of suffering endometriosis can be explained by genetic heritability[22].
- **SOCIAL CLASS AND RACE:** A few works have conducted studies in comparable populations, but they did not find differences in terms of social class or race[23]. The high socio-economic level and white race have been considered risk factors, but this could be due to the increase in accessibility to the diagnosis that these women have. Therefore, it is difficult to assess these criteria and determine their impact on the disease risk.
- **MENSTRUAL AND REPRODUCTIVE HISTORY:** Epidemiologic studies suggested that women with premature menarche, short cycles, and hypermenorrhea have an increased risk of suffering endometriosis. Regarding reproductive history, the parity is inversely proportional to the risk of suffering the disease[24,25].
- **ORAL CONTRACEPTIVES (OCPs):** One of the main symptoms of endometriosis is dysmenorrhea (painful periods), and one of the treatments for the pain is to

administrate OCPs. Because OCPs are administrated to women for other purposes, it could be that the symptoms are not present and diagnosis of the disease is underestimated. In any case, total duration of taking oral contraceptives has not been related to the risk of endometriosis[24].

- **LIFE HABITS:** Research suggests that the environment plays an important role in the development of the disease. For example, it is believed that physical exercise could have a protective effect against the development of the disease. In addition, tobacco consumption has been widely studied, and there are controversial results. It seems that tobacco consumption could prevent the disease because it has an anti-estrogenic effect, and endometriosis is an estrogen-dependent disease. However, it has also been found that products from combustion increase the risk of suffering from endometriosis. In addition, a relationship between endometriosis and alcohol and caffeine consumption has been described where the increase of those products would potentiate the risk of endometriosis. However, there are controversial results for this phenomenon as well[26].

- **BODY MASS INDEX (BMI):** BMI has been extensively studied as a risk factor for endometriosis, but there are a lot of controversies. In general, it is believed that overweight women have less endometriosis because their cycles are more irregular and/or anovulatory, both of which are considered protective factors against the development of the disease. Recently, a meta-analysis was performed analyzing more than 3,000 published articles, and it was found that there was an inverse association between endometriosis and BMI. This suggested a possible reduced risk of suffering the disease for women with higher BMI[27]. In addition, it was reported that low childhood body size (between 5-10 years old) also had a correlation with the risk of suffering endometriosis, where an increase in body size decreased the risk of suffering from the disease[28].

- **OTHER FACTORS:** Chemical factors, such as dioxin, have been described as being possibly involved in the development of the disease. For example, in the north of Italy and Belgium, where exposure to dioxin is high, there is a greater incidence of

endometriosis[29]. Moreover, some immunologic alterations, such as lupus, hypo and hyperthyroidism, rheumatoid arthritis, and multiple sclerosis, could be related to endometriosis development[26].

1.2. CLINICAL FEATURES

The disease clinical management is determined depending on the location and severity of the disease. In addition, one of the main problems that physicians face is that diagnosis of the disease is very difficult, and it can be up to 10 years before physicians find that a patient is suffering from endometriosis.

SYMPTOMS OF ENDOMETRIOSIS

The main symptoms of endometriosis are acute pelvic pain and infertility/subfertility, although some patients are asymptomatic. Endometriosis-associated pain is very complex in terms of its characteristics and intensity. Usually, the pain disappears after removal of the cysts, but it is very common for the disease to recur. Another important aspect to consider is that the quality of life of many women suffering from disease is low given the problems that the disease entails. Many women suffer from depression and need professional psychological help. Here, the main symptoms of the disease – pain and infertility – are described.

PAIN: Usually, pelvic pain due to endometriosis is chronic and it is associated with dysmenorrhea (painful periods), acute pelvic pain (cyclic or acyclic), dyspareunia (pain during sexual intercourse), dyschezia (painful defecation), and/or dysuria (painful urination). The pain can be continuous or it can occur unpredictably and intermittently across the menstrual cycle. It can be inexistent, throbbing, or heavy, and it is exacerbated when physical activity is conducted[1]. Normally, patients feel pain in the regions surrounding the endometriotic cysts, and the pain increases when there is menstruation. As the lesions are formed by endometrial tissue, they respond to sex hormones. When menstruation occurs, endometriotic tissue in distant locations also responds to hormones by growing, bleeding, and producing inflammation in the lesion location. This inflammation could

stimulate pelvic end nerves producing chronic pelvic pain[30]. In addition, inflammation and secretion of cytokines and chemokines would also cause pain.

INFERTILITY: The association between infertility and endometriosis and adenomyosis[31] has been widely studied[32]. As mentioned before, it has been described that between 25% and 50% of infertile women suffer from endometriosis, and around 30%-50% of women with endometriosis are infertile[4]. Despite a clear association between endometriosis and infertility, the mechanisms that cause this infertility/subfertility are uncertain[33]. Different mechanisms have been proposed as the cause of infertility, but no single one is conclusive:

- **Pelvic adherences:** Adherences could affect tubal function, oocyte uptake, and/or fecundation. For example, adherences in the fallopian tubes would affect the spermatozoid migration through the tubes.
- **Chronic inflammatory intraperitoneal microenvironment:** Endometriosis is characterized by the presence of chronic inflammation. It has been shown that immune populations are involved in the chronic inflammation present in endometriotic lesions. For example, it has been described that macrophage M2 (Mφ2) is increased in the peritoneal fluid and lesions of women with endometriosis, and that macrophage M1 (Mφ1) is the predominant population in eutopic endometrium of these women[34]. Mφ1 produces pro-inflammatory cytokines and chemokines that induce this inflammatory alteration in eutopic endometrium, which could be related to infertility[35]. More studies are needed to confirm this hypothesis. An extensive explanation of immune populations (Mφ, regulatory T cells, and uterine natural killers) and their described impact in eutopic endometrium of women with endometriosis is shown in *Section 1.5*.
- **Defect in folliculogenesis:** It has been proposed that folliculogenesis, which is the process of the maturation of the ovarian follicle that contains the immature oocytes, has been altered in women with endometriosis. This could be due to a defect in the hypothalamus-hypophysis axis that could affect the folliculogenesis in diseased women[36].

- **Progesterone resistance:** During the secretory phase of the menstrual cycle, progesterone is responsible for producing changes in the stromal cells, which decidualize and prepare an optimal environment for embryo implantation. It has been described that patients with endometriosis could have progesterone resistance[37], which would lead to poorer obstetric outcomes due to the defect in the decidualization of the stromal compartment of the endometrium.
- **Tubal dysfunction:** It has been found that patients with endometriosis have tubal peristalsis, which means that there is poor uterine contractibility. These movements allow sperm to travel through the fallopian tubes to fecundate the oocyte. After the fecundation, the embryo is transported to the uterine cavity for implantation. If these movements are altered in women with endometriosis, it could be another cause of defects in fertility[38].

Complications in embryo implantation are not the only factor that could produce infertility in women with endometriosis. The risk of suffering early miscarriage in women with adenomyosis has also been described[29].

Other symptoms, such as nausea, distension, and early satiety are also typical in women with endometriosis. In addition, some patients have reported symptoms with neuropathic components, such as hypersensitivity[39]. Other symptoms such as migraines have been also reported by patients[40]. Symptoms overlap with other gynecologic diseases (pelvic adhesions, ovarian cysts, myomas, leiomyomas, etc.) but also with other non-gynecological conditions (inflammatory bowel disease, cystitis, myofascial pain, etc.) and factors (depression) that make its diagnosis a major challenge[1].

DIAGNOSTIC OF ENDOMETRIOSIS

Clinical history and pelvic examination help with the diagnosis of the disease, but endometriosis is a very heterogeneous disease in regards to aspects such as the symptoms and a high prevalence of asymptomatic patients. All of this makes the diagnosis very difficult[41]. There is an average diagnosis delay of 6.7 years[28], and most women report having to see a doctor five times or more before being diagnosed or referred[42]. The age at first consultation for symptoms is predominantly between 20-29 years old, which also delays the diagnosis. The percentages of consultation are shown in Figure 4[28].

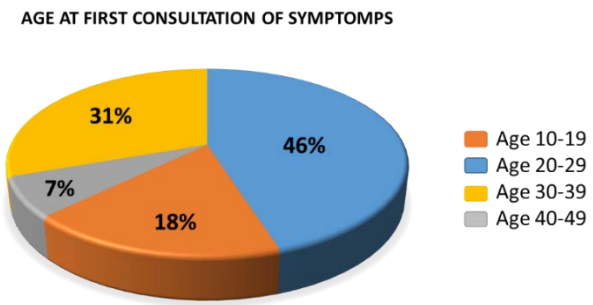


Figure 4. Age at first consultation regarding symptoms of endometriosis. As the figure shows, doctors start to question patients for symptoms of the disease mainly between 20-39 years old.

There are different reasons why some women are excluded from the diagnostic of the disease. The first one is that some of them are asymptomatic. However, when the patient feels pain, there are different paths to be excluded:

- The women feel pain but do not visit the doctor.
- The women feel the pain and present symptoms, but the diagnosis is wrong. It is confused with other pathologies, such as gastric or urinary disorders.
- The women have pain and symptoms, but the doctor administrates OCPs. Therefore, endometriosis symptoms are dismissed.

Currently, no simple non-invasive test for the diagnosis of endometriosis is routinely implemented in clinical practice. Pelvic examination is the first step when there is suspicion

of endometriosis. Non-invasive imaging is also used for the diagnosis of the disease. Transvaginal ultrasounds (TVUSs) and magnetic resonance imaging (MRI) allow doctors to identify the precise anatomical location of some forms of endometriosis, particularly ovarian endometriosis and DIE lesions. Disadvantages include operator-dependency, requirements for specialized training, and failure to detect superficial peritoneal endometriosis[41]. In addition, in TVUSs there is a 40% false negative rate, and in the case of MRIs, a 20% chance[43].

Nowadays, the diagnostic gold standard is visual inspection by laparoscopy and a histologic confirmation of macroscopic lesions. Laparoscopically proven disease has been diagnosed in more than 50% of women with a clinically normal pelvic examination[44]. Disadvantages of laparoscopic surgery include high costs, the need for general anesthesia, and the potential for adhesion formation post procedure[45]. It is estimated that the annual costs of endometriosis have exceed $49 billion in the United States[27].

The accuracy of a diagnostic is determined by the sensitivity and specificity. When there is a high sensitivity, it means that a low number of individuals have a negative test and do have the disease (low numbers of false negative results), and when there is a high specificity, it means that a low number of individuals have a positive test but do not have the disease (few false positive results). The diagnostic accuracy of laparoscopy compared with histological confirmation of endometriosis has been estimated as having a 94% sensitivity and 79% specificity[41]. However, laparoscopy is an invasive diagnostic method. Thus, a less invasive diagnosis is needed for women with disabilities. Nnoaham et al. proposed a clinical prediction model to diagnose type III-IV endometriosis based solely on questioning patients. They developed a symptom-based predictive model as a screening tool. They obtained 84.9% sensitivity and 75.8% specificity for stages III and IV of the disease[46]. Unfortunately, laparoscopy is still the diagnostic with more accuracy, and therefore the most reliable and definitive diagnostic used by the majority of specialists.

CLASSIFICATION OF ENDOMETRIOSIS

Currently, most researchers and clinicians use the revised American Society of Reproductive Medicine (rASRM) classification, which is internationally accepted[47]. It is accepted by the World Endometriosis Society that published an international consensus statement on the classification of the disease through systematic appraisal of evidence and a consensus process that included representatives of international, medical and non-medical societies, patient organizations, and companies in 1996 (updated in 2017)[48]. In this article, they also proposed the use of World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project surgical and clinical data collection tools for research to improve the classification of endometriosis in the future, of particular relevance when surgery is not undertaken[2]. Nowadays, with the emergence of the omics, there is the opportunity to improve the classification of the disease by using new innovative tools from a genetic or proteomic perspective.

The rASRM classification considers appearance, size, and depth of peritoneal or ovarian implants and adhesions visualized during laparoscopy. It classifies the disease as minimal, mild, moderate, or severe using a punctuation system. Unfortunately, it does not correlate clinically with pelvic pain and fertility and does not account for adenomyosis. The punctuation used in this classification is as follows, and the way to analyze it is shown in **Figure 5**[47].

- Stage I / minimal endometriosis: 1-5
- Stage II / mild endometriosis: 6-15
- Stage III / moderate endometriosis: 16-40
- Stage IV / severe endometriosis: >40

Figure 5. The classification method of rASRM (1996)

AMERICAN SOCIETY FOR REPRODUCTIVE MEDICINE
REVISED CLASSIFICATION OF ENDOMETRIOSIS

Patient's Name _____ Date _____

Stage I (Minimal) - 1-5
Stage II (Mild) - 6-15
Stage III (Moderate) - 16-40
Stage IV (Severe) - >40
Total _____

Laparoscopy _____ Laparotomy _____ Photography _____
Recommended Treatment _____
_____
Prognosis _____

PERITONEUM	**ENDOMETRIOSIS**	<1cm	1-3cm	>3cm
	Superficial	1	2	4
	Deep	2	4	6
OVARY	R Superficial	1	2	4
	Deep	4	16	20
	L Superficial	1	2	4
	Deep	4	16	20
	POSTERIOR CULDESAC OBLITERATION	Partial	Complete	
		4	40	
OVARY	**ADHESIONS**	<1/3 Enclosure	1/3-2/3 Enclosure	>2/3 Enclosure
	R Filmy	1	2	4
	Dense	4	8	16
	L Filmy	1	2	4
	Dense	4	8	16
TUBE	R Filmy	1	2	4
	Dense	4*	8*	16
	L Filmy	1	2	4
	Dense	4*	8*	16

*If the fimbriated end of the fallopian tube is completely enclosed, change the point assignment to 16.

Denote appearance of superficial implant types as red [(R), red, red-pink, flamelike, vesicular blobs, clear vesicles], white [(W), opacifications, peritoneal defects, yellow-brown], or black [(B) black, hemosiderin deposits, blue]. Denote percent of total described as R___%, W___% and B___%. Total should equal 100%.

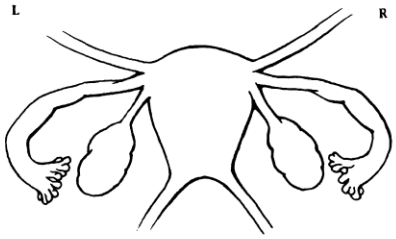
Additional Endometriosis: _____

_____
_____
_____

Associated Pathology: _____

_____
_____
_____

To Be Used with Normal Tubes and Ovaries



To Be Used with Abnormal Tubes and/or Ovaries

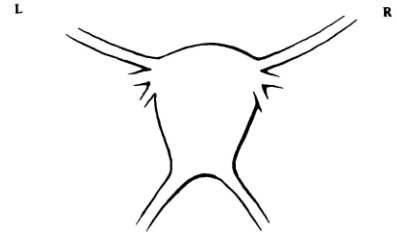


Figure5. The classification method of rASRM (1996) (*Continued*)

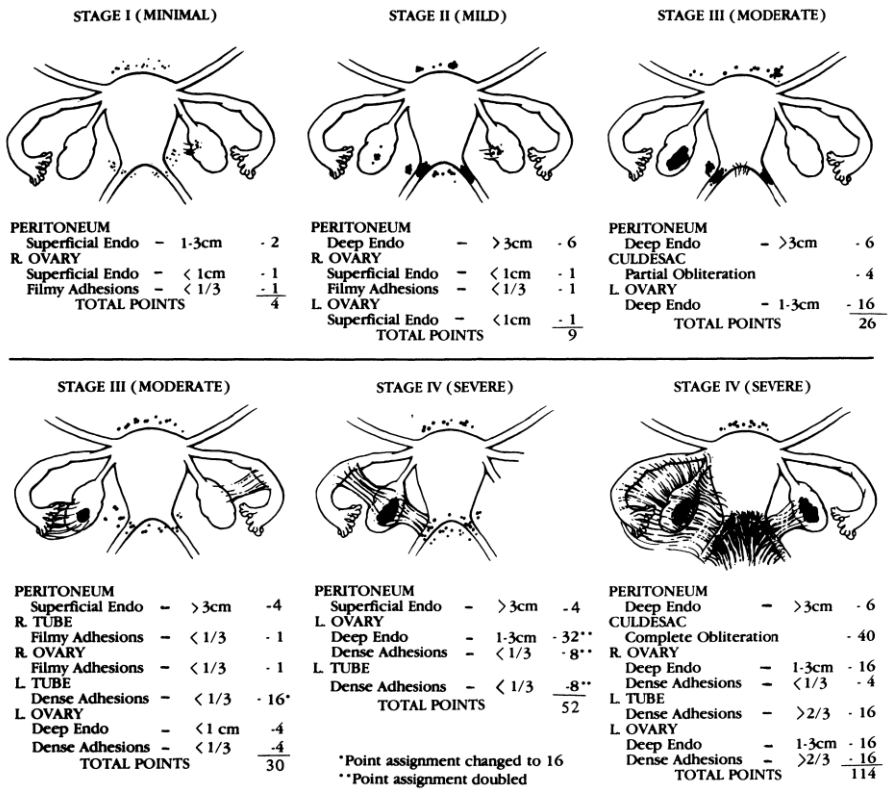


Figure 5. The classification method for rASRM (1996). This classification is based in the aspect, size, and depth of the peritoneal and ovarian implants, the presence, extension, and adherence types of the annexes, and the grade of obliteration of Douglas's pouch.

TREATMENTS IN ENDOMETRIOSIS

No cure for the disease is known. Treatment options include surgery to remove the endometriotic lesions and hormonal therapy[42]. Current hormonal treatments include OCPs, progestogens, and gonadotropin-releasing hormone (GnRH) agonists and antagonists that inhibit or reduce the effect of estrogen on endometrial tissues. This leads to a state of amenarche that decreases the pain. However, taking into account that 30-50% of women with endometriosis are infertile/subfertile, amenarche is not a way to be able to conceive.

In addition, in the long term administration of GnRH, there is the risk of developing osteoporosis[49], thus it is usually a temporary treatment.

To treat the chronic pelvic pain and dysmenorrhea, painkillers can also be administrated. In addition, surgical removal of endometriotic lesions can reduce the pain and improve fertility. However, 40-50% of women have recurrence of lesion growth after surgery[46]. Moreover, there is the discussion of whether is better or worse to do the surgery to remove the lesions. Sometimes, especially in the case of DIE, the surgery can be very complicated, and the post-surgery consequences can be worse than the initial endometriotic lesion symptoms. Therefore, sometimes specialists prefer not to operate on the lesions to avoid having a negative impact on the patient's quality of life.

1.3. DIAGNOSTIC BIOMARKERS

As explained previously, the current diagnostic gold standard for endometriosis is still surgical (laparoscopy). Therefore, the discovery of non- or less-invasive biomarkers is needed. For those women with the symptoms but without endometriosis, the risks of laparoscopy may outweigh the benefits. This is the group of women for whom a biomarker test might be most useful.

The term biomarker, or biological marker, refers to a broad subcategory of medical signs, or objective indications of medical state observed form outside the patient, that can be measured accurately and with reproducibility[50]. A non-invasive biomarker it is described as a procedure that does not involve the penetration of skin or any physical entrance to the body[51]. The hunt for non-invasive biomarkers for endometriosis has been an ongoing, challenging issue. Endometriosis is a very heterogeneous disease regarding symptoms, stages, genetics, and so on. One of the considerations to account for is the phase of the menstrual cycle, because the endometrium is a very dynamic tissue that responds to hormones, and depending on the day of sampling, the biomarkers could change. Besides, when working with human samples, the sample size is always a limitation for studies. There is a need to increase the population of a study to obtain consistent results. To do so, collaborations between research centers are often made. Unfortunately, one of the

challenges is that samples are not always collected the same way, which introduces bias between studies. As mentioned previously, the World Endometriosis Research Foundation has prompted physicians, gynecologists, and researchers to standardize methods of sample collection and analysis of data as a research priority for endometriosis[2].

Non-invasive biomarkers with a high specificity and sensitivity are a challenge for endometriosis, not only because of the heterogeneity of the disease and sample characteristics, but also because of comorbidities suffered by endometriosis patients. It has been described that patients with endometriosis have high incidences of endocrine and autoimmune disorders[52] as well as allergies[42]. Furthermore, gastrointestinal-related abdominal pain and constipation are also common in women with endometriosis[53], and an association has been reported between endometriosis and gynecologic cancers and immunologic diseases[54]. All these conditions may also show a peripheral increase of cytokines commonly associated with endometriosis, introducing bias in studies of immunological markers as biomarkers for endometriosis. Even though it seems that endometriosis could have positive associations with different comorbidities, the only consistent findings have been found with ovarian cancer[2]. With improved characterization of patient history and standardized means of collection, storage, and interpretation of data, tools to identify non-invasive diagnostic biomarkers will be improved.

Many biomarkers have been proposed for endometriosis diagnostics, including several types of molecules such as DNA, RNA, proteins, and cells. They are related to different biological processes and can be obtained from diverse sources[55] but none of them have been translated to clinics. Some authors proposed that it is most probable that the answer lies in the study of a combination of biomarkers and the identification of an ideal panel that can predict the diagnosis and the severity of endometriosis[56].

Below, the main sources of non-invasive biomarkers and the results of the studies conducted using different types of samples (urine, saliva, blood, and endometrial biopsies) are briefly described. The advantages and disadvantages of each sample are explained, along with the Cochrane studies that have been performed on each one.

The **Cochrane Library** (ISSN 1465-1858) is a collection of six databases that contain different types of high quality independent evidence to inform healthcare decision-making and a seventh database that provides information about Cochrane groups.

Different experts have reviewed the biomarker studies from different sources included in the Cochrane Library. They have performed meta-analysis to determine the specificity and sensitivity of the published biomarkers and to elucidate if laparoscopy could be replaced as a diagnostic for the disease. Below, the reported conclusions are explained.

Urine biomarkers: There are advantages to collecting this type of sample. Urine can be easily self-collected and is acceptable for most patients. Moreover, it can be collected in large quantities, can be rapidly analyzed, and it is economic (compared to laparoscopy). However, there are some disadvantages to using the biomarkers found in urine. The relevance to the disease of the molecules excreted in urine is potentially unknown, the urine concentration has to be determined as it varies across individuals, the menstrual cycle might influence the composition of urine, and host cells and/or bacterial cells are a potential source of contamination that can influence the metabolic profile[57]. Liu et al. reviewed eight studies involving 646 participants where urine biomarkers were studied and performed meta-analysis to prove the diagnostic power of the reported biomarkers[58]. All of the biomarkers were assessed in small individual studies and could not be statistically evaluated in a meaningful way. There was not enough evidence to recommend any urinary biomarker for use in clinical practice for the diagnosis of endometriosis. Several of the biomarkers had diagnostic potential, but require further evaluation.

Saliva biomarkers: This sample would be the least invasive way to diagnose the disease. It has a lot of advantages in terms of sample collection and it would be acceptable for all patients. Few genetic markers have been identified through DNA amplification of buccal swabs as well as hormonal markers. This is an area with a lot of potential, but it requires further research and validation[56].

Blood biomarkers: This type of sample is very convenient as a diagnostic source for the disease, as the majority of patients would agree on this type of sampling. Other advantages are that it can be collected in large quantities and a broad range of biomolecules can be analyzed. The latest could also be an inconvenience, because the complex mix of biomolecules reflects many biological processes, some of which may not be relevant to the disease. Different fractions of the sample can be analyzed: whole blood, plasma serum, or cells. Depending on the source (cellular/non-cellular fractions), different samples can be used (RNA, DNA, proteins, metabolites, and cells).

Nisenblat et al. reviewed and analyzed all of the literature in the Cochrane library regarding blood biomarkers. A meta-analysis of 141 studies on 15,141 participants and 122 blood biomarkers was studied[59]. The conclusions of the meta-analysis were the same as for the urine and saliva studies. No marker subjected to meta-analysis consistently met the criteria for a replacement of laparoscopy.

Recently, the identification of circulating cells in blood has been proposed as a new method for the diagnosis of different cancers such as colon, prostate, renal, pancreatic, and lung cancer[60–64]. Circulating tumor cells are rare cells that are shed into bloodstream from primary or metastatic tumors and have the potential to initiate metastasis in distant tissues or organs[65,66]. The Food and Drug Administration (FDA) has approved the CellSearch® platform for the detection of circulating tumor cells (CTCs), and this platform is already being used[67–70]. Therefore, circulating endometrial cells (CECs) are promising biomarkers for endometriosis with great potential for a noninvasive diagnostic assay, but again, further research is needed.

Endometrial biopsy biomarkers: This kind of sample is obtained by either pipelle or curette, and is minimally invasive. Although it can be a bit painful for the patient, endometrial biopsies can be useful not only for testing endometrial receptivity in infertile women with or without endometriosis[71], but also as a diagnostic tool for endometriosis[72]. Regarding Sampson's theory, eutopic endometrium might cause endometriosis. Using the direct source of the disease is perhaps the most logical way to discover biomarkers for the

disease. One thing to account for is that eutopic endometrium varies its gene expression across the menstrual cycle due to cell response to hormones[73]. Thus, studies have to be very well designed and samples have to be collected in the same phase of the cycle to elucidate the impact that different phases could have on biomarker expression. Furthermore, choosing proper control groups has to be considered when designing a study. Many gynecological pathologies could also have repercussions in gene expression or epigenetics in endometrium, and additionally, endometriosis can coexist with myomas (uterine fibroids)[74,75] and/or polyps[76,77].

Recent evidence suggests that significant biological differences exist between eutopic endometrium in women with and without endometriosis[1]. Different approaches have been used to assess these differences, such as DNA[78], transcriptome[78,79], proteomic[80–83], lipidomic[78], and methylome analysis[84]. Moreover, some studies tried to elucidate the differences between stages of the disease to find biomarkers that differentiate its severity[85–87]. None of them have showed potential to replace laparoscopy as a diagnostic.

Grupta et al. performed a meta-analysis of the Cochrane Library to determine the diagnostic accuracy of endometrial biomarkers for pelvic endometriosis using a surgical diagnosis as the reference standard[88]. Fifty-four studies involving 2,729 participants were included. The studies evaluated endometrial biomarkers either in specific phases of the menstrual cycle or outside of it and tested the biomarkers either in menstrual fluid, in whole endometrial tissue, or in separate endometrial components. In these and in the other cases explained above, researchers could not statistically evaluate most of the biomarkers described.

Combination of non-invasive tests: Nisenblat et al. reviewed all published articles combinations of non-invasive tests for the diagnosis of endometriosis[89] and the conclusion was as follows: *"None of the biomarkers evaluated in this review could be evaluated in a meaningful way and there was insufficient or poor-quality evidence. Laparoscopy remains the gold standard for the diagnosis of endometriosis and using any non-invasive tests should only be undertaken in a research setting."*

Despite all of the scientific efforts to discover new biomarkers for this disease, none of the reported studies found methods to replace laparoscopy as the major diagnostic tool for endometriosis. Therefore, there is a need to identify new combinations of biomarkers that have a higher sensitivity and specificity than laparoscopy in order to improve the diagnosis of endometriosis.

1.4. STEM CELL MARKERS IN THE ENDOMETRIUM

The endometrium is a highly dynamic tissue that regenerates monthly. Stem cells are critical to the cyclic renewal of healthy endometrium. Several groups have suggested the existence of a human endometrial somatic stem cell (SSC) population in this tissue[90–92]. The characterization of possible stem/progenitor cells in the endometrium has shed new light on both the origins of ectopic endometrial tissue and the mechanism for the pathogenesis of endometriosis. It was demonstrated that both endometrium-derived and bone marrow-derived stem cells could migrate to ectopic sites and contribute to the development of endometriotic lesions[93]. Evidence based on linage tracing studies suggested that Leucine-rich repeat containing G-protein coupled receptor 5 (LGR5) was an universal SSC, as it was observed in several tissues with different origins such as the small intestine (endodermal origin), hair follicles (ectodermal origin), and kidneys (mesodermal origin)[94–96]. LGR5, also called G-protein coupled receptor 49 (GPR49) is a seven transmembrane G-protein. G-protein coupled receptors belong to one of the largest and most diverse families of membrane proteins. These proteins are important signal transducers that control a wide variety of physiological functions, including hormones, immune responses, enzyme release, smooth muscle contraction, cardiac, neurotransmission, and blood pressure regulation[97]. LGR5 is characterized by the presence of 17 leucine-rich repeats (LRR) within the extracellular domain and is a negative modulator of Wnt signaling. Until very recently it was believed that it was an orphan receptor, but nowadays it is known that R-spondins are high affinity ligands of LGR5[98]. LGR5 positive (LGR5+) cells constitute multipotent stem cells that generate all cell types in the intestinal epithelium[96]. Therefore, it was postulated that LGR5

could be a stem cell marker in the endometrium. Several authors have been studying it in this tissue[92].

1.4.1. LGR5 in eutopic endometrium

In endometrium, it was described in a mice model that LGR5 was mainly expressed in the luminal and glandular epithelia and was downregulated by estrogen and progesterone in the uterus[99]. The expression of LGR5 was uniform across the epithelium, and it was suggested that those cells have the potential to differentiate when necessary for uterine glandular growth[99]. However, mice do not menstruate, and the results have to be analyzed cautiously. Little is known about this protein in human eutopic endometrium, but some authors have found this marker in this tissue[30,92,100–102]. First, it was postulated that LGR5 could be a somatic stem cell (SSC) marker. Somatic stem cells are a restricted subpopulation of quiescent, undifferentiated resident cells with high proliferative capacity, multi-potentiality, self-renewal capability, and the ability to form the tissue from which they originate[103].

Tempest et al. reported that LGR5+ cells are found in the basalis epithelium, where normally the stem cells are located, and in the glandular epithelium in the functionalis layer[30]. They also observed that LGR5 expression correlated with endometrial epithelial proliferation in the stratum functionalis epithelial compartment. Therefore, they hypothesized that human endometrium could have more than one epithelial stem cell/ progenitor pool: one in the basalis responsible for regenerating the functionalis layer after menstruation or parturition, and one supporting the embryo-implantation process that maintains the luminal epithelial cells that are lost daily[30].

LGR5+ cells seem to remain constant throughout the menstrual cycle in healthy human eutopic endometrium at the RNA and protein levels[92,100]. Very recently, Tempest et al., also performed analysis of LGR5 across the menstrual cycle in healthy eutopic endometrium[30]. They described that LGR5 was regulated at RNA level by sex hormones. When medroxiprogesterone (MPA), a synthetic progestin very similar to progesterone, was administered to endometrial primary cell culture, LGR5 was downregulated in the secretory

phase of the menstrual cycle. They also observed this phenomenon when studying endometrial biopsies of women whom had been receiving synthetic progestogen treatment, suggesting that LGR5 is regulated by sex hormones.

In parallel to the performance of this thesis, Cervelló et al. demonstrated that LGR5+ cells from healthy endometrium have a hematopoietic origin[101]. They sorted LGR5+/- cells by Fluorescence activated cell sorting (FACS) of the epithelial and the stromal compartments and a phenotypic analysis of LGR5+ cells were performed. The epithelial fraction contained 1.08 ± 0.73% of LGR5+ cells and the stromal fraction 0.82 ± 0.76% **(Figure 6A)**. They found that LGR5+ cells were negative for endothelial markers (CD31 and CD133) as well as for the hematopoietic stem cell marker CD34. Furthermore, they observed that 62.73% of the epithelial and 57.84% of the stromal LGR5+ cells were also expressing the leukocyte marker CD45, thus concluding that more than the 50% of LGR5+ cells have hematopoietic origin **(Figure 6B)**. Therefore, co-expression with other hematopoietic markers was assessed. They reported that LGR5+CD45+ (49% of epithelial cells and 57% of stromal cells) cells were negative for the T lymphocyte marker CD3, as well as B lymphocyte marker CD19 and the natural killer cells marker CD56. However, they found that they were CD163+ cells, which is a specific marker for Mφ2 **(Figure 7)**.

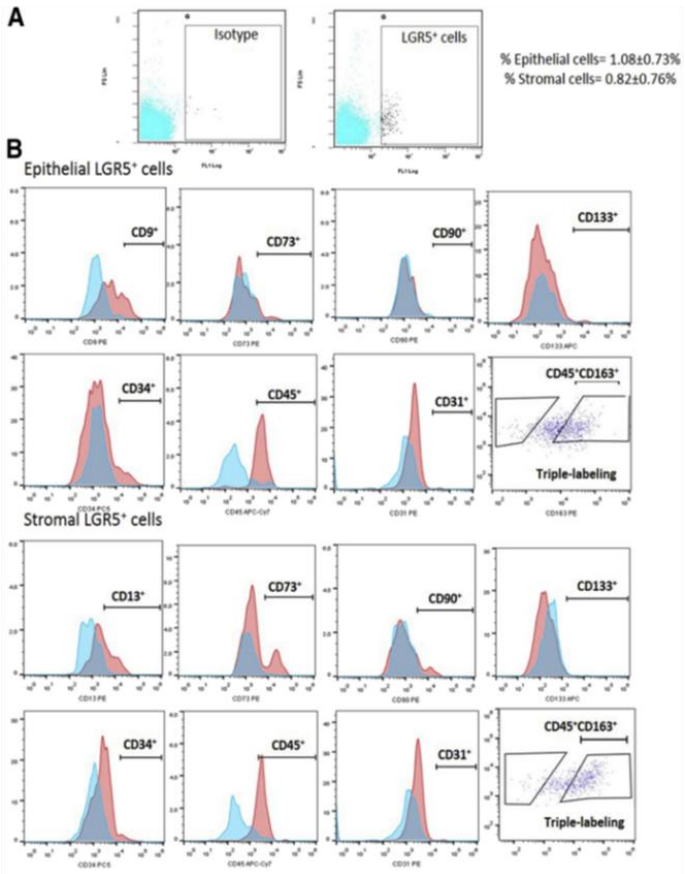


Figure 6. Immunocharacterization of LGR5+ cells in healthy human eutopic endometrium. A) We can observe the proportion of epithelial and stromal cells in FACS analysis. **B)** Different antibodies labeled to discover the origin of LGR5+ cells (*Cervelló I et al., 2017, Fertil Steril.).*

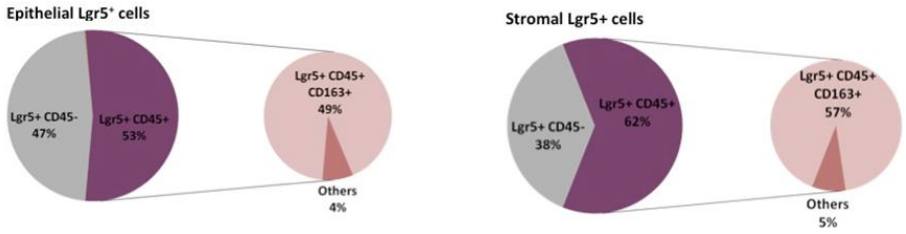


Figure 7. Diagram showing the percentage of LGR5+CD45+CD163+ cells. Of the epithelium and stroma, 49% and 57% respectively co-expressed CD163 *(Cervelló I et al., 2017, Fertil Steril.).*

To elucidate the differences in gene expression between LGR5+CD45+ and LGR5+CD45- cells, RNA-High-Sequencing (RNA-Seq) was performed and no significant differences were found. The discovery of this macrophage-like phenotype opens a door to study LGR5+ cells in the niche of endometrial stem cells. Macrophages (Mφ) are implicated in tissue remodeling and angiogenesis[104] and it is known that they interact with stem cells within their specific niche and modulate their self-renewal and tissue remodeling.

All this data suggests that LGR5+ cells could be involved in the modulation of the endometrial stem cell niche. LGR5 was described to be located in a restricted epithelial and stromal area close to blood vessels in the lower regions of the functionalis layer[92]. Thus, it was proposed that LGR5+ cells might have a role in the modulation of the endometrial stem cell niche rather than be a stem cell per se. Cervelló et al. proposed two hypotheses for the role of LGR5+ cells in endometrium **(Figure 8**). One, after menstruation, perivascular resident Mφ mobilize stem cells to repair and remodel the endometrial layer. Two, these Mφ could be recruited from bone marrow to activate the endogenous niche, contributing to recovery of the functional layer.

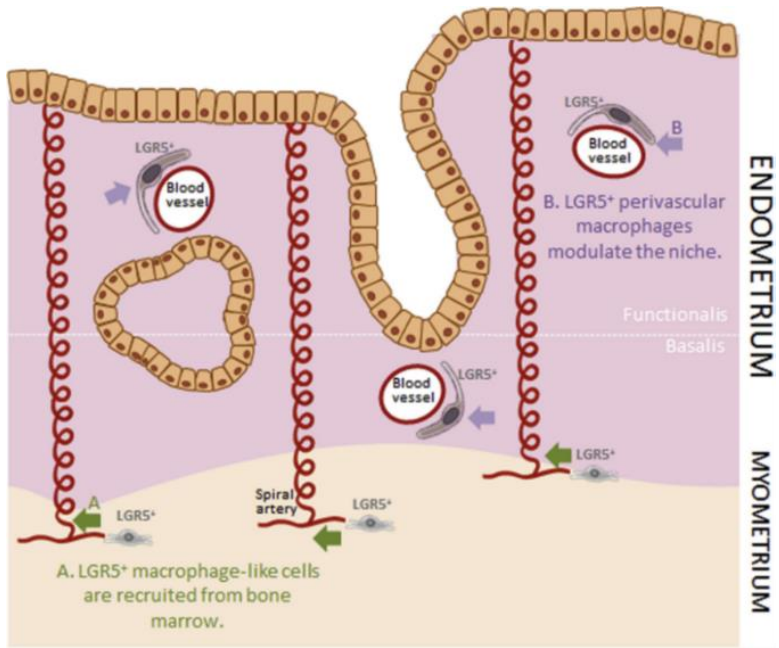


Figure 8. Hypothesis of LGR5+ cells role in eutopic endometrium. A) LGR5+ macrophage-like cells are recruited via bone marrow. **B)** LGR5+ perivascular macrophages modulate the niche *(Cervelló I et al., 2017, Fertil Steril.).*

Bidirectional communication between primary and metastatic tumors has been demonstrated[105] by shedding circulating cells into the blood circulation in certain metastatic cancers. Cells selectively migrate and engraft the original tumor and contribute to the progression of the disease. Additionally, a recent study has observed a substantial reduction in liver metastasis when LGR5+ cells from original colon tumors were depleted[106], and that selective destruction of these cells in the same tumors led to temporary tumor regression. This highlights the important role that LGR5+ cells may have in the progression of the disease[107]. Although endometriosis is a benign disease, it shares some characteristics with cancer. Thus, LGR5 might be involved in endometriotic lesion formation.

Santamaria et al. demonstrated in a rodent model of endometriosis that GFP+ cells from endometriosis lesion migrated specifically to eutopic endometrium, modifying its normal gene expression profile[108]. These migrating cells were mostly located close to blood vessels and aberrantly expressed the epithelial marker cytokeratin (CK) in the stromal compartment. Subsequent gene expression analysis revealed that these cells expressed markers related with cell adhesion, stemness, and epithelial-mesenchymal transition process (EMT), as well as LGR5. These results indicate that LGR5+ cells might migrate from eutopic endometrium through lymphatic or hematogenous spread and could have a role in the development of endometriosis.

1.5. IMMUNE POPULATIONS IN EUTOPIC ENDOMETRIUM

While the etiology of endometriosis is poorly understood, the immune system is believed to play a central role in its pathogenesis, pathophysiology, and symptomatology. Importantly, more than 90% of women have some degree of retrograde menstruation, but only 10% to 15% have endometriosis[1], suggesting aberrations in the clearance of ectopic endometrial tissue and cells by the immune system in women with versus without the disease[4]. What these aberrations are and why they occur is poorly understood, but is increasingly a focus of study to understand endometriosis risk and effective therapies. Both the peripheral immune system and the immune status of endometrium within the uterus are altered in women with

endometriosis and likely contribute to infertility, early pregnancy failure and abnormal tissue homeostasis in affected women[5].

1.5.1. MACROPHAGES

Macrophages (Mφ) are key effector cells in both innate and humoral immunity as they recognize and phagocytose pathogens, act as antigen presenter cells (APC) to activate T cells, and have a role in tissue regeneration[109]. During menstruation, Mφ participate in the process of shedding endometrium from its location within the uterus[104]. They may also be involved in the regeneration of endometrium, as they have a role in angiogenesis and wound healing[104]. In the endometrium, they are approximately 10% of the total immune cell population[110–113], being the second most abundant endometrial leukocyte population after T cells[114]. They change in numbers in different cycle phases, suggesting regulation by the steroid hormones estradiol (E_2) and progesterone (P_4)[16,115–118]. Mφ comprise 1% to 2% of endometrial cells in the proliferative phase, 3% to 5% in the secretory phase, and 6% to 15% in the menstrual phase[113]. Their increased numbers during menses may be attributed to their phagocytic properties and role in clearing cell debris and apoptotic cells during endometrial shedding. Depending on activation state and surface markers, Mφ are classified as either classically activated Mφ (Mφ1) or alternatively activated Mφ (Mφ2)[104]. This plasticity in phenotype is due to environmental cues[119]. For example, tumor associated Mφ (TAM) can make a bidirectional transformation between Mφ1 and Mφ2 phenotypes leading to pro- or anti-inflammatory reactions, depending on their environment[114,120]. Mφ1 secrete pro-inflammatory factors, whereas Mφ2 are involved in angiogenesis, anti-inflammatory processes, and coordination of tissue repair[104,109].

In normal endometrium, the majority of Mφ are CD163+CD14^{Low}, which correspond to alternatively activated macrophages (Mφ2)[109,121]. Mφ have been widely studied in the endometriotic lesions, but little is known about their function in the eutopic endometrium of women with disease. A few studies have characterized their abundance in the eutopic endometrium of women with endometriosis across the menstrual cycle[16,122,123] with inconsistent results. For example, three groups have reported that Mφ are more abundant

in the eutopic endometrium of patients with endometriosis than in controls during all phases of the menstrual cycle[16,123,124]. In contrast, another study observed that the endometrial Mφ content was significantly reduced in the eutopic endometrium of endometriosis patients compared to healthy controls in the proliferative phase, but was not significantly different from the control group in other phases of the cycle[122]. Moreover, while Mφ increases in the menstrual phase in healthy endometrium, this phenomenon does not occur in women with endometriosis[16]. Eutopic endometrial cells in women with endometriosis are less susceptible to apoptosis[122,125,126], and the lower abundance of Mφ during the menses in women with disease, along with aberrant secretion of several pro-inflammatory factors, could enhance survival of shed and refluxed endometrial cells, enabling them to establish endometriosis lesions in the pelvic cavity. Interestingly, eutopic endometrial cells from women with endometriosis have more invasive and adhesive phenotypes[127], consistent with facilitating lesion formation. In addition, nerve fibers are increased in the functionalis layer of women with endometriosis[128,129], and it is known that Mφ play a role in nerve fiber growth, development, and repair[130,131]. Whether Mφ promote dysmenorrhea, the most common symptom in women with endometriosis, is currently unknown and warrants further investigation. Mφ1 contribute to the promotion of tumor development[104], and although endometriosis is a benign disease, it shares some characteristics with cancer[132], with the neoplastic process, including inflammatory processes and tissue invasion. Mφ1 have the ability to alter eutopic endometrial cells of women with endometriosis, potentially leading them to invade the peritoneal cavity and develop tumor-like cysts, suggesting that Mφ could be involved in the development of the disease as well as in its prognosis.

Taking together, in normal conditions, Mφ in eutopic endometrium increase pre-menstrually and during the menstrual phase and likely have a major role in the clearance of shed menstrual cells and cell debris from the tissue in the uterus. In eutopic endometrium in women with endometriosis, however, Mφ do not fluctuate throughout the menstrual cycle, and thus may not effectively eliminate cell debris, promoting the latter's implantation in ectopic sites. Further studies are needed to better understand the molecular differences

between Mφ in normal eutopic endometrium and eutopic endometrium of women with endometriosis.

1.5.2. REGULATORY T CELLS (Treg)

Treg cells are potent suppressors of inflammatory immune responses and have a role in preventing autoimmunity in all tissues[133]. Treg cells have been proposed to regulate the function and proliferation of Mφ, mast cells degranulation, dendritic cells, neutrophils, eosinophils, B cells, T cells, and natural killer cells, all of which have a role in menstruation[134,135].

In normal endometrium, Treg increases during the proliferative phase and decreases at the end of the secretory phase[136]. This increase prior to ovulation could indicate that Tregs are important for successful implantation, and help in providing an immune-tolerant environment by inhibiting cytotoxic activity of other cell types through secretion of immunosuppressive cytokines such as interleukin 10 (IL10) and tumor growth factor beta (TGF-β)[133,136–138]. CD4+ Treg cells are increased in the eutopic endometrium of women with versus without endometriosis[139–142]. There is controversy about the abundance of Treg cells in the eutopic endometrium of women with endometriosis across the menstrual cycle. Some studies have demonstrated higher numbers of Treg cells in the secretory phase compared with the other phases of the cycle, and other studies have reported no differences across the cycle[136,140,143–145]. These controversies in the percentage of Treg have been found in the eutopic and in the ectopic endometrium of women with endometriosis[210–212].

In summary, in normal eutopic endometrium, Tregs increase in abundance in the proliferative phase, unlike in the eutopic endometrium of women with endometriosis, where they increase in the secretory phase or do not fluctuate across the cycle. In normal conditions, Tregs are upregulated prior to ovulation, likely to produce an immune-tolerance environment for impending embryo nidation. Different amounts of these cells in eutopic endometrium in women with endometriosis may indicate an abnormal local immune environment.

1.5.3. NATURAL KILLER CELLS

Natural killer (NK) cells are important components of the innate immune system, acting as the first line of defense against viral infections and tumor growth. They are also important for normal tissue homeostasis[146]. Natural killer cells are found in all tissues of the female reproductive tract, and depending on their location, exhibit different phenotypes, functions, and regulation[83,84]. Natural killer cells found in the female reproductive tract are phenotypically and functionally distinct from those obtained from peripheral blood[111]. Uterine NK (uNK) cells are the predominant leukocyte population in normal human endometrium[147], comprising 30% to 40%[148] of total leukocytes in the proliferative phase and up to 70% in the secretory phase[111,118,148–154]. The dramatic increase of uNK cells in the secretory phase is likely attributable to their critical roles in pregnancy establishment and maintenance[155]. Uterine NK cells can produce and secrete angiogenic factors including VEGF and angiotensin 2 (ANG2), which promote the maturation of blood vessels, making the endometrial spiral arterioles thicker and larger. This is important for successful implantation and throughout pregnancy[156,157]. Therefore, uNK are key players in implantation[158], and dysregulation of their numbers or functionality could lead to a failure in embryo implantation as has been reported in women with endometriosis[158].

In the eutopic endometrium of women with endometriosis, uNK cells increase in the secretory phase, like in women without the disease[159,160]. Uterine NK cells have less cytotoxicity in women with versus without endometriosis[161].

In summary, in normal women, uNK cells have poor cytotoxic activity. In the eutopic endometrium of women with endometriosis, their activity is further reduced, which would explain why uNK do not remove cell debris from menstruation.

Hypothesis and Objectives

2. HYPOTHESIS AND OBJECTIVES

HYPOTHESIS

One of the main problems of endometriosis is its delayed diagnosis. Molecular alterations have been demonstrated in the eutopic endometrium of women with endometriosis. We hypothesized that differences in gene expression in eutopic endometrium of women with endometriosis compared with a control group in combination with phenotypic characteristics of the patients would allow us to discover new biomarkers for the disease and develop a classifier panel as a new diagnostic approach for the disease. Furthermore, we hypothesized that LGR5 and Mφ might be involved in the pathophysiology of endometriosis and that LGR5 could be a diagnostic biomarker.

Therefore, the **main objective** of this thesis is to **develop a new diagnostic approach for endometriosis that is less invasive** than laparoscopy.

SPECIFIC OBJECTIVES

1. To discover new biomarkers in the eutopic endometrium of women with endometriosis

2. To evaluate the potential of LGR5 as a biomarker for endometriosis and the implication of LGR5 and immune cells in the pathophysiology of the disease

Materials and Methods

3. MATERIALS AND METHODS

As previously mentioned, there are two specific objectives in this thesis. To carry them out, different studies were performed:

<u>Specific objective 1:</u> To discover new biomarkers in eutopic endometrium of women with endometriosis

- **Pilot study:**

- To elucidate if eutopic endometrium embedded in paraffin and fresh eutopic endometrium samples are comparable in terms of gene expression

- **Discovery study:**

- To study the differential gene expression in eutopic endometrium of healthy and diseased women and develop a classifier for the disease

- **Validation study:**

- To validate the candidate biomarkers obtained in the discovery study

<u>Specific objective 2:</u> To evaluate the potential of LGR5 as a biomarker for endometriosis and the implication of LGR5 and immune cells in the pathophysiology of the disease

- **Array's validation of the endometriosis mice model:**

- To determine LGR5 co-localization with GFP+ cells in eutopic endometrium

- **Presence of LGR5 in the eutopic endometrium of women with endometriosis:**

- To determine LGR5 presence in the eutopic and ectopic endometrium in women with endometriosis and study its co-localization with epithelial markers

- **LGR5 throughout the menstrual cycle:**

- To elucidate the variation of LGR5 throughout the menstrual cycle in women with and without endometriosis

- **Gene expression analysis of LGR5 positive cells:**

- To determine the percentage of LGR5+ cells in eutopic and ectopic endometriosis endometrium
- ***In silico* study** to discover the genes that vary across the menstrual cycle and in an stimulated endometrium
- **Preliminary study** to evaluate the feasibility of LGR5+ cells to perform RNA-High-Sequencing and to assess the differences in terms of gene expression between the epithelial and stromal compartments
- **Complete study** to study LGR5+ cells' expression between controls and endometriosis patients

- **Pilot study immune populations**

- To assess the differences in gene expression of macrophages (Mφ1 and Mφ2), Treg, and uNK in women with and without endometriosis

The materials and methods section has been organized by techniques. Even though some techniques were common to different studies, they were performed slightly differently, or different kits were used from several commercial providers. Therefore, in order to facilitate the reader's understanding, a subtitle of the specific objective where the technique was used has been introduced in some sections.

In addition, the following diagram **(Figure 9, page 73)** shows the two main projects performed in this thesis with their respective methods. In the diagram, we can observe in parenthesis the section number where the specific technique is explained in materials and methods section.

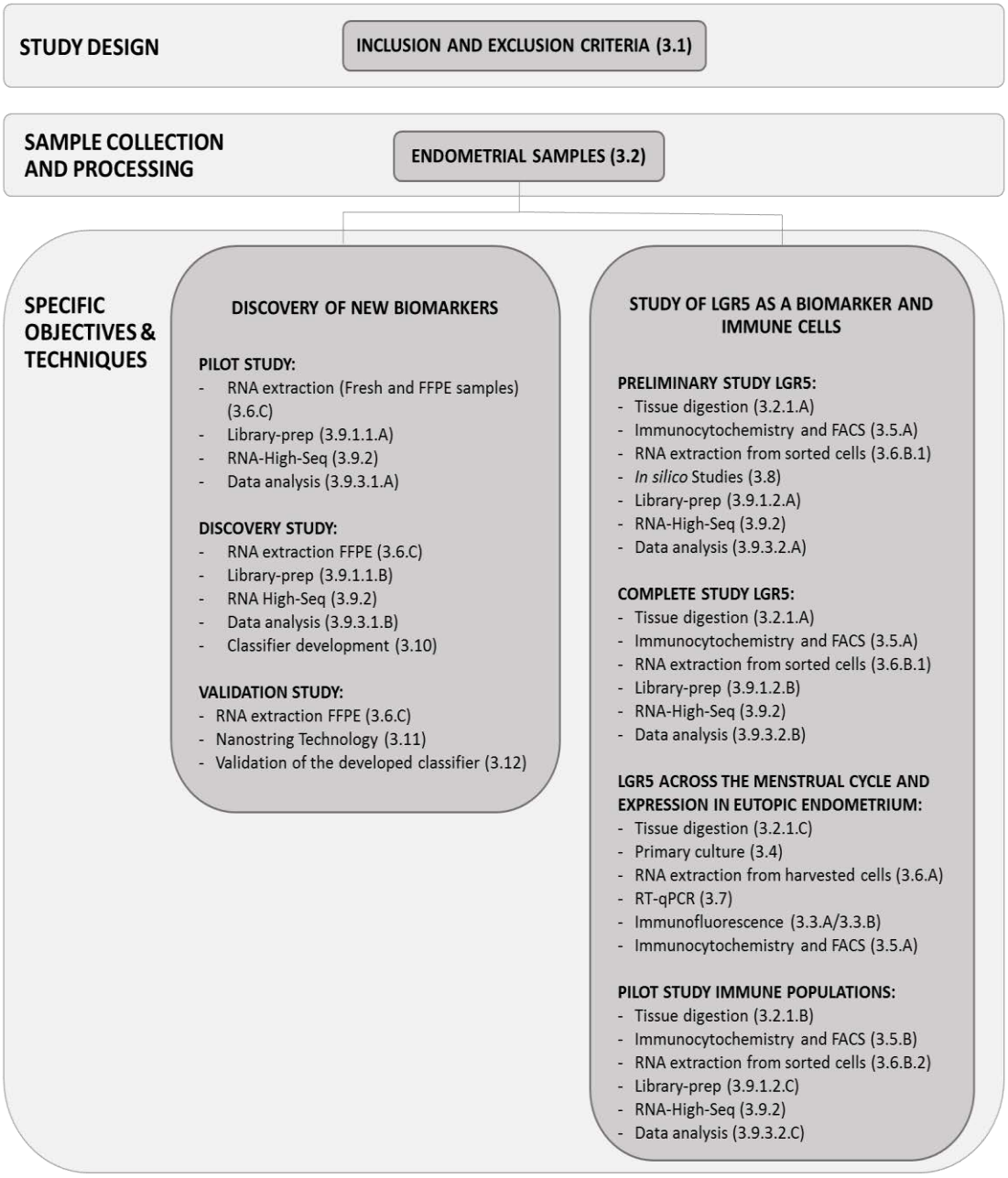


Figure 9. Diagram of materials and methods. This diagram shows the materials and methods used in each specific step of the three objectives of the thesis. In parenthesis, we can observe the section where each specific experiment is explained.

3.1. INCLUSION AND EXCLUSION CRITERIA

Endometriosis is a disease that mainly affects women in reproductive age. Therefore, only women between 18 and 45 years old were included in the studies. A complete clinical history review was done to exclude patients with heritable diseases or with viral infections, such as HIV, VHB, and VHC, and to record their phenotypic variables. Below, the inclusion and exclusion criteria in each objective are explained.

Objective 1: Discovery of new biomarkers in eutopic endometrium

The control and endometriosis groups were women who went to the clinic because they had gynecological symptoms such as pelvic pain, hypermenorrhea, or dysmenorrhea. Thus, an endometrial biopsy was taken for histological analysis. Some of these biopsies resulted in healthy endometrium being reported, and some in endometriosis. Endometriosis cases were proven by laparoscopy. In this case, women who were undergoing hormone treatment, such as oral contraceptives, were not excluded. This choice was made to elucidate if hormones would be an important variable for the diagnostic classification for the disease.

Objective 2: Evaluation of the potential of LGR5 as a biomarker for the disease and the implication of LGR5 and immune cells in the pathophysiology of the disease

In the case of the LGR5 study, the control group was made up of egg donors. In the case of the immune populations study, the control group included patients with benign diseases, such as myomas or polyps. Egg donors were treated with follicular stimulating hormone (FSH) to stimulate the maturation of the oocytes. In these studies, endometriosis patients had not been undergoing hormone treatment for at least three months before the surgery.

3.2. SAMPLE COLLECTION AND PROCESSING

Endometrial biopsies from patients with endometriosis or control patients were obtained from women undergoing hysterectomy or endometrial biopsies. All samples were collected using a Cornier pipelle under an approved protocol after written informed consent. Use of the uterine specimens after surgery was approved by Ethics Committee of Vall d'Hebron

Research Institute, Barcelona (Spain) (PR(AMI)410/2016). Lesion samples were analyzed for diagnosis in the Pathological Anatomy Department of Vall d'Hebron University Hospital. Stage of menstrual cycle was determined for each specimen through histological examination by a pathologist. The collection of samples from egg donors and endometriosis patients was approved by the ethics committee of IVI Barcelona S.L. under the number 1611-BCN-080-XS.

In the case of the immune populations study, samples were obtained from University of California San Francisco (UCSF) under approval of its ethics committee. The number of the approval project was IRB#10-02786.

All patients from all centers received and signed an informed consent agreement.

In total, we obtained 728 endometrial samples embedded in paraffin (FFPE) and 147 fresh endometrial samples from Vall d'Hebron University Hospital. We obtained 39 fresh endometrial biopsies from egg donors from IVI Barcelona S.L. In addition, 17 fresh endometrial samples were obtained from UCSF.

Some of the samples had to be discarded for various reasons. Some of these reason included an insufficient quantity of sample, bad quality or low concentration of RNA after extraction, or because patients did not fit the inclusion criteria, among others.

The following table summarizes the total collected samples and the final number of samples used for each experiment conducted.

	Control			Endometriosis		
	Collected samples	Used samples	Experiment	Collected samples	Used samples	Experiment
FFPE	535	110	Discovery of biomarkers	193	18	Discovery of biomarkers
		19	Validation of biomarkers		175	Validation of biomarkers
		24	IF LGR5 across the menstrual cycle		24	IF LGR5 across the menstrual cycle
Fresh	39	4	LGR5 across the menstrual cycle (*in vitro* study)	159	3	LGR5 across the menstrual cycle (*in vitro* study)
		12	LGR5 across the menstrual cycle (flow cytometry)		23	LGR5 across the menstrual cycle (flow cytometry)
		12	Percentage of LGR5		23 eutopic/10 ectopic	Percentage of LGR5
		2	Preliminary study LGR5		2	Preliminary study LGR5
		5	Complete study LGR5		14	Complete study LGR5
		5	Pilot study immune populations		6	Pilot study immune populations
Total	**574**	**193**		**352**	**298**	

Table 2. Total sample collection and use. This table shows the total number of collected samples and the number of samples used in each experiment.

For all endometrial samples, two procedures were performed: tissue digestion, and fixation and paraffin embedding. Blood from patients was also collected. The processing of these samples and their further usage is explained in the following section.

3.2.1. TISSUE DIGESTION

Endometrial tissue was minced mechanically and digested with DMEM high glucose and collagenase (1mg/ml). After one hour of digestion at 37ºC in rotation or at 4ºC overnight (ON), samples were separated by gravity sedimentation and single cells were filtered in a 40µm mesh and washed with phosphate buffered saline solution (PBS) 1x. In this case, stromal cells were obtained. Epithelial cells remained in the mesh. The upside down mesh was washed with PBS 1x to recover this cell compartment. Afterwards, epithelial cells were incubated with accutase to disaggregate the glands and obtain singe cells. The pellet was washed with PBS 1x and resuspended with serum-containing medium (SCM).

SCM: DMEM high glucose + 10% fetal bovine serum (FBS) + 5% penicillin and streptomycin + 5% sodium pyruvate + 0.05% insulin.

Depending on the intended usage of the samples, we processed the digested tissue differently, as explained below.

A) For the immunocytochemistry of LGR5:

For the preliminary study of LGR5 **(Objective 2)**, we collected 3 endometrial biopsies from egg donors and 3 from endometriomas. For the complete study, 18 uterine samples (n=13 endometriosis and n=5 healthy) were used. We used the freshly digested tissue – both the epithelium and the stromal compartments – to perform the immunocytochemistry (ICC) and fluorescence activated cell sorting (FACS) of the sample on the same day (*see Section 3.5.A*).

B) For the immunocytochemistry of immune populations:

For the study of immune populations **(Objective 2)**, seventeen endometrial biopsies were collected at UCSF: 11 from donors with benign uterine diseases and six from patients with endometriosis. We centrifuged the samples for five minutes at 1200 rpm. The digested media was then removed and the pellet was resupsended with freezing medium and stored at -80ºC. After 24 hours, samples were placed in liquid nitrogen to await further use (*see Section 3.5.B*).

Freezing medium: 80% SCM + 20% dimethyl sulfoxide (DMSO)

C) For the primary cultures:

Primary cultures were established to determine the expression of LGR5 throughout the menstrual cycle **(Objective 2)**. Three endometrial biopsies were obtained from women with endometriosis who had undergone surgery. These samples were provided by Vall d'Hebron University Hospital. We also obtained four fresh eutopic endometrium biopsies from healthy women who were egg donors, provided by IVI Barcelona S.L. After tissue digestion, samples were centrifuged for five minutes at 1,200 rpm and single stromal cells were resuspended with SCM and placed in p100 plates. The next day, the media was changed and cells were grown until confluence. They were frozen at -80ºC with freezing media and placed in liquid nitrogen to await further use (see *Section 3.4*).

3.2.2. ENDOMETRIUM FIXATION AND EMBEDDED IN PARAFFIN

A total of 159 fresh uterine biopsies from women with endometriosis prospectively for diagnosis or treatment of benign gynecological disorders were provided by Vall d'Hebron University Hospital and 39 biopsies from healthy women were provided by IVI Barcelona S.L. As explained above, a piece of the sample was used for tissue digestion. The other piece of each sample was fixed in 4% formaldehyde (FA4%) and embedded in paraffin for further use in the validation phase of the biomarkers discovery study **(Objective 1)** as well as to determine the phase of the menstrual cycle using hematoxylin and eosin (H&E).

We also obtained retrospectively 530 blocks of formalin-fixed paraffin embedded (FFPE) endometrium which corresponded to all types of endometriosis, adenomyosis, and healthy endometrium from the Pathological Anatomy Department of Vall d'Hebron University Hospital to use in the discovery and validation of biomarkers **(Objective1)**.

3.3. IMMUNOFLUORESCENCE

A) Immunofluorescence of LGR5 in mice eutopic endometrium

We performed IF on eight eutopic endometrium samples of the endometriosis mice model described above (Santamaria et al. 2012) to confirm the presence of LGR5 in the GFP+ cells that had migrated from the lesion specifically to the uterus of the mice. Moreover, we analyzed the co-localization of LGR5 with the epithelial marker E-Cadherin (ECAD).

Slides were incubated at 55ºC overnight and treated in a xylene and ethanol circuit. Then, they were treated with ammonium chloride (NH_4Cl) for 15 minutes and for 20 minutes with citrate (pH=6) (Abgent) at 95ºC in a water bath. This was followed by blocking using 0.1% of PBS-Tween detergent + 5% Normal Goat Serum (NGS) (Invitrogen) + 5% Bovine Serum Albumin (BSA) (Sigma-Aldrich) for one hour at room temperature (RT). Primary antibodies were monoclonal rabbit anti-LGR5 antibody (Abgent) in a 1:30 dilution and monoclonal rat anti-GFP antibody (B-Bridge) in a 1:50 dilution. They were diluted with PBS-T 0.1% + 3% NGS + 3% BSA and incubated overnight at 4ºC. Secondary antibodies were goat Alexa647 anti-rabbit (Invitrogen) and goat Alexa488 anti-rat (Invitrogen), all in 1:500 dilutions, also

diluted with PBS-T 0.1% + 3% NGS + 3% BSA, and were incubated for 45 minutes at RT in the dark. We used ProLong Gold anti-fade reagent with 6-diamino-2-phenylindole (DAPI) (Invitrogen) to visualize nuclear DNA. Immunoreaction without primary antibodies and without antibodies was performed to obtain negative controls, and the small intestine was used as a positive control. Visualizations and pictures were done with an OlympusBX61 microscope. Once this co-localization was confirmed, we transferred the results into human.

B) Immunofluorescence in human eutopic endometrium

The process performed for the IF was the same as explained above, but the primary antibodies were monoclonal rabbit anti-LGR5 antibody (Abgent) in a 1:30 dilution, monoclonal mouse anti-E-cadherin antibody (Santa-Cruz) in a 1:150 dilution, and monoclonal mouse anti-pan-Cytokeratin antibody (Santa-Cruz) in a 1:50 dilution. Secondary antibodies were goat Alexa647 anti-rabbit (Invitrogen) and goat Alexa488 anti-mouse (Invitrogen).

We studied LGR5 co-localization with ECAD and CK in eutopic endometrium biopsies from 12 healthy women and 20 diseased patients on 3µm tissue sections. Consecutive cuts were made in order to analyze the expression of LGR5 and epithelial markers in the profundity of the tissue. Thereafter, three slides per patients were obtained in duplicate (three for the co-localization of LGR5 with ECAD and three for the co-staining with CK). The co-localization was visualized and pictured with OlympusBX61 microscope.

In addition, and in order to elucidate LGR5 expression throughout the menstrual cycle, we performed IF of LGR5 in 24 control and 24 endometriosis-affected eutopic endometrium samples. Each group contained eight samples from each of the three menstrual phases: proliferative, secretory, and menstruation. To calculate the protein expression of LGR5 from IF and determine its variation across the menstrual cycle, ImageJ software was used to determine the fluorescence intensity mean (FIM) of the Alexa647 anti-rabbit antibody (red) that corresponded to LGR5 marker and the FMI of DAPI (blue) that corresponded to the number of cells as they stained the nuclei of the cells.

3.4. MENSTRUAL CYCLE *IN VITRO*

Using stromal cells (eSF) primary cell cultures, an assay reproducing menstrual phases was performed. Four samples from egg donors and three from DIE women were cultured. The primary cultures had two-three passages in order to keep the integrity of the eSF and $2x10^5$ cells were cultured in p6 well plates with SCM containing 2% FBS. In order to mimic the different phases of the menstrual cycle (proliferative and secretory), cells were treated with 10^{-8}M estrogen (E_2) β-estradiol (Sigma) for six days and with 10^{-6}M progesterone (P_4) (Sigma) from Day 6 to Day 12. Both hormones were diluted in ethanol and a second group was treated only with ethanol as a vehicle. Every two day, thes medium was changed and the hormones were added. Samples were obtained at Day 0 (control), Day 6 (proliferative phase) and Day 12 (secretory phase). The experiment was performed in duplicate. After the specific treatment, cells were harvested for RNA extraction (*see Section 3.6.A*).

3.5. IMUNOCYTOCHEMISTRY (ICC) AND FLUORESCENCE ACTIVATED CELL SORTING (FACS)

A) ICC and FACS of LGR5 in human eutopic endometrium

In the first place, a preliminary study with nine eutopic endometrium samples from women with endometriosis (five in the proliferative phase and four in the secretory phase) was performed in order to determine the variation of LGR5 throughout the menstrual cycle. Because all of the controls were egg donors, their phase of the menstrual cycle was ovulatory. Therefore, the analysis of LGR5 throughout the menstrual cycle was performed using the results from patients with endometriosis.

After collagenase treatment and separation of the epithelial and stromal compartments, samples were treated with an erythrocyte lysis buffer. Then, they were blocked with 5% BSA (Sigma-Aldrich) for one hour at RT. The primary antibody was monoclonal rabbit anti-LGR5 antibody (BioNova Scientific) (1µl per million of cells) and the secondary was goat Alexa647 anti-rabbit (Invitrogen) in a 1:500 dilution. To discard death cells, they were stained by a DAPI (Invitrogen) probe (5µg/ml). LGR5+/- cells were sorted using BD® FACS ARIA I instrument and were collected separately in Trizol (Invitrogen) and stored at -80ºC.

To confirm that LGR5+ cells had been specifically sorted, a cytospin of 5,000 cells on a slide was done. The cells were stained using the same antibody used in IF (*see Section 3.3.B*). The procedure was almost the same, with the difference that sorted cells were fixed in the slide with FA 4% and, after three washes, with PBS 1x. Otherwise, cells were blocked and stained as explained in *Section 3.3.B*. The cytospin was performed in three endometrial biopsies from healthy women (egg donors).

Then, 35 endometrial biopsies from nine control patients, three pelvic endometriosis patients, nine ovarian endometriosis patients, four adenomyosis patients, and 10 DIE patients were stained following the same procedure. Seventy paired samples were obtained, including LGR5+/- cells from each endometrial biopsy. In addition, ectopic endometrium from one pelvic endometriosis, four ovarian endometriosis, two adenomyosis, and six DIE was also obtained. Analysis of the percentage of LGR5+ cells and statistical analysis were done using FCS Express5.0 and GraphPad software, respectively. T-tests ($p<0.05$) and one-way ANOVAs ($p<0.05$) were used to determine statistical differences between the groups. Comparisons between the percentage of LGR5+ cells in the eutopic endometrium of women with and without endometriosis were performed, as well as between types of endometriosis. In addition, the percentage of LGR5+ cells was compared in 10 ectopic lesions (four ovarian and six DIE) with their respective paired eutopic endometrium as well as between ectopic lesions of different types of endometriosis (four ovarian and six DIE). Not enough pelvic and adenomyosis ectopic tissues were available to perform analysis of the ectopic lesions (only one pelvic endometriosis and two adenomyosis lesions were obtained). Further study of the cytometry results allowed elucidating if there were differences in terms of percentage of LGR5+ cells between the epithelial and the stromal compartments. The comparisons made are shown in the following table **(Table 3)**.

Comparisons
1. Epithelium and Stroma between all groups and Epithelium vs Stoma in each group
2. Total eutopic endometrium (Epithelium + Stroma) between all groups
3. Total eutopic endometrium control vs Total endometriosis
4. Epithelium ectopic vs Epithelium eutopic (patients with endometriosis)
5. Stroma ectopic vs stroma eutopic (patients with endometriosis)
6. Total ectopic vs Total eutopic (4 ovarian and 6 DIE)
7. Ectopic between types of endometriosis (4 ovarian and 6 DIE)

Table 3. Comparisons of LGR5+ cells obtained by flow cytometry. Seven comparisons were made in order to find differences in LGR5 expression in controls and endometriosis samples as well as between epithelium and stroma.

From the sorted samples, only five samples from healthy women and 14 from diseased women were used for RNA-Seq (38 paired samples; LGR5 positive and LGR5- cells). The 14 samples from endometriosis patients included three pelvic endometriosis, three ovarian endometriosis, four adenomyosis, and four DIE.

B) ICC and FACS of immune populations in eutopic endometrium

Panel design:

A cytometry panel of 10 conjugated antibodies able to separate the four immune populations of interest (Mφ1, Mφ2, Treg, and uNK) was performed. Specific membrane markers were needed for each population to be able to obtain a pure sample. Resident and blood infiltrating immune cells were discerned to avoid the counting and sorting of cells coming from blood contamination. For this reason, and because the purity of the sorted populations was important, more than one marker for each population was included, which made building the panel very challenging.

The antigen density of each marker had to be taken into account in order to decide which color would be used. The brightest colors were used for the markers with the lowest antigen density. However, sometimes, because there was overlapping of the light spectrum between colors or because the commercial antibody was not available, the best antibody was not used in all cases. Nevertheless, we successfully achieved no overlapping of the 11

colors (10 conjugated antibodies plus the live/dead dye) using the five lasers that the machine had available. The markers and lasers used for each population and the reasons that each was used are explained below. The panel strategy is shown in **Figure 10**.

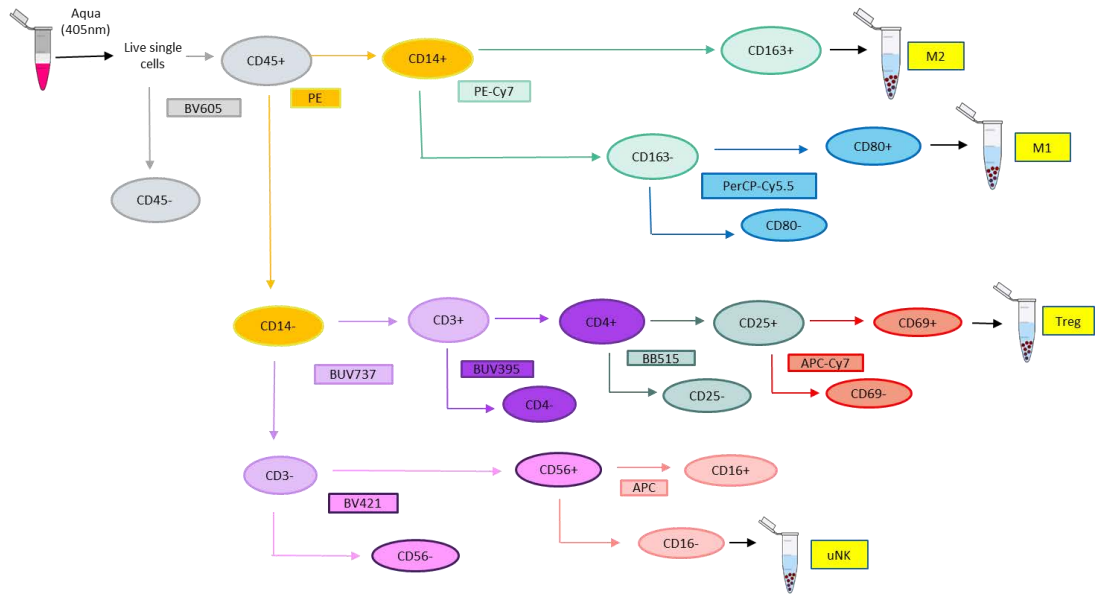


Figure 10. Immune populations FACS strategy. The figure shows the used markers for FACS of Mφ1 and Mφ2, Treg and uNK. It also shows the fluorocroms attached to the antibodies against each marker.

First, the cells were gated with a CD45 marker (leukocyte marker), as all the desired cells express this protein in the plasma membrane. The color conjugated to the CD45 antibody was BV605. BV605 is a very bright dye and was used to identify CD45 because it was the main population to separate from the total endometrial tissue. In the case of Mφ, usually this population is a resident tissue population, so no specific tissue markers were used. It is considerably difficult to differentiate between Mφ1 and Mφ2 subpopulations as they have some common markers. For example, both of them are CD14+. In the case of Mφ2, CD163, which is a specific marker for this population, was used conjugated with PE-Cy7. On the other hand, there are no specific markers for Mφ1. Nonetheless, it has been reported that this type of cell overexpress CD80 when activated. Because it was suspected that the concentration of activated Mφ1 would be low in the endometrium, as it has been described

that Mφ2 are expressed more than Mφ1 in normal endometrium[109], CD80 conjugated with PerCP-Cy5.5 antibody, a bright dye, was used. In addition, separated Treg, or T cells that express CD3 and CD4 markers, was sorted. The most accepted specific marker for Treg is Foxp3, but it is intracellular so it cannot be used for sorting. However, Treg cells also express CD25. To be able to discern between tissue Treg and blood contamination Treg, the marker CD69, which is an activation marker for Treg and it is expressed in tissue, was included in the panel. Thus, CD3+CD4+CD25+CD69+ cells were isolated. The CD3 antibody conjugated with BUV737 was used, as was the CD4+ antibody with BUV395. CD25 has a low antigen density, so BB515 was used, which is one of the brightest dyes on the market. Finally, to collect the resident-tissue Treg, CD69 conjugated with APC-Cy7 was used. The last population to isolate was the uNK cells. Uterine NK are CD56+, and it has been described that blood NK are CD16+ while uNK are CD16 Low/-. For this reason, CD56+CD16- were collected. For the CD56 marker, the brightest color, BV421, was used as it is known that it has a very low antigen density. In the case of CD16, APC conjugated antibody was used. To differentiate between live/dead cells, aqua dye was used.

ICC and FACS:

Frozen samples were thawed at 37ºC. After centrifugation at 1,300rpm for five minutes, the supernatant was discarded and the pellet was washed with PBS 1x. After another centrifugation at 1,300rpm for five minutes, cells were resuspended with PBS 1x + 5% BSA and incubated at RT for 30 minutes. Cells were counted and a minimum of 100,000 cells was separated as a negative control for the sorting of the cells. FACS Aria Jabba the Hutt (BD Biocience) was used to sort the populations.

Moreover, All Fluorocroms Minus One (FMO) was prepared for each antibody to compensate for the channels in the FACS instrument. This step was done three times (with three frozen samples) in order to optimize the machine settings and establish the gates in the software to create a template for further samples.

Ten conjugated antibodies were used to stain the samples: CD45-BV605 (BD), CD14-PE (Biolegend), CD163-PE-Cy7 Biolegend), CD80-PerCP-Cy5.5 (Biolegend), CD3-BUV737 (BD), CD4-BUV395 (BD), CD25-BB515 (BD), CD69-APC-Cy7 (BD), CD56-BV421 (BD), and CD16-APC (Biolegend). 1µl of antibody per million cells was used in all cases except for CD45 and CD4, where 2µl per million cells were needed. After incubation for one hour at 4ºC in PBS 1x + 3% BSA and in the dark, cells were washed with PBS 1x and centrifuged for five minutes at 1,300rpm. The pellet was then resuspended with 500µl of PBS 1x and stained with 1µl of LIVE/DEAD™ Fixable Aqua Dead cell stain dye (ThermoFisher) to be able to discern between living and dead cells.

UltraComp eBeads compensation magnetic beads (ThermoFisher) were stained with each of the 10 antibodies. Each tube containing one drop of beads and 50µl of PBS 1x was incubated separately with 1µl of each antibody for 15 minutes on ice. They were then washed with 1ml of PBS 1x and centrifuged for five minutes at 1,300rpm. The supernatant was discarded and the beads were resuspended with 50µl of PBS 1x. After the gate strategy **(Figure 11)**, each population was sorted and collected in PBS 1x.

Flow cytometry analysis of the sorted cells was performed using FlowJo v10 software, and statistical analysis was done using GraphPad software.

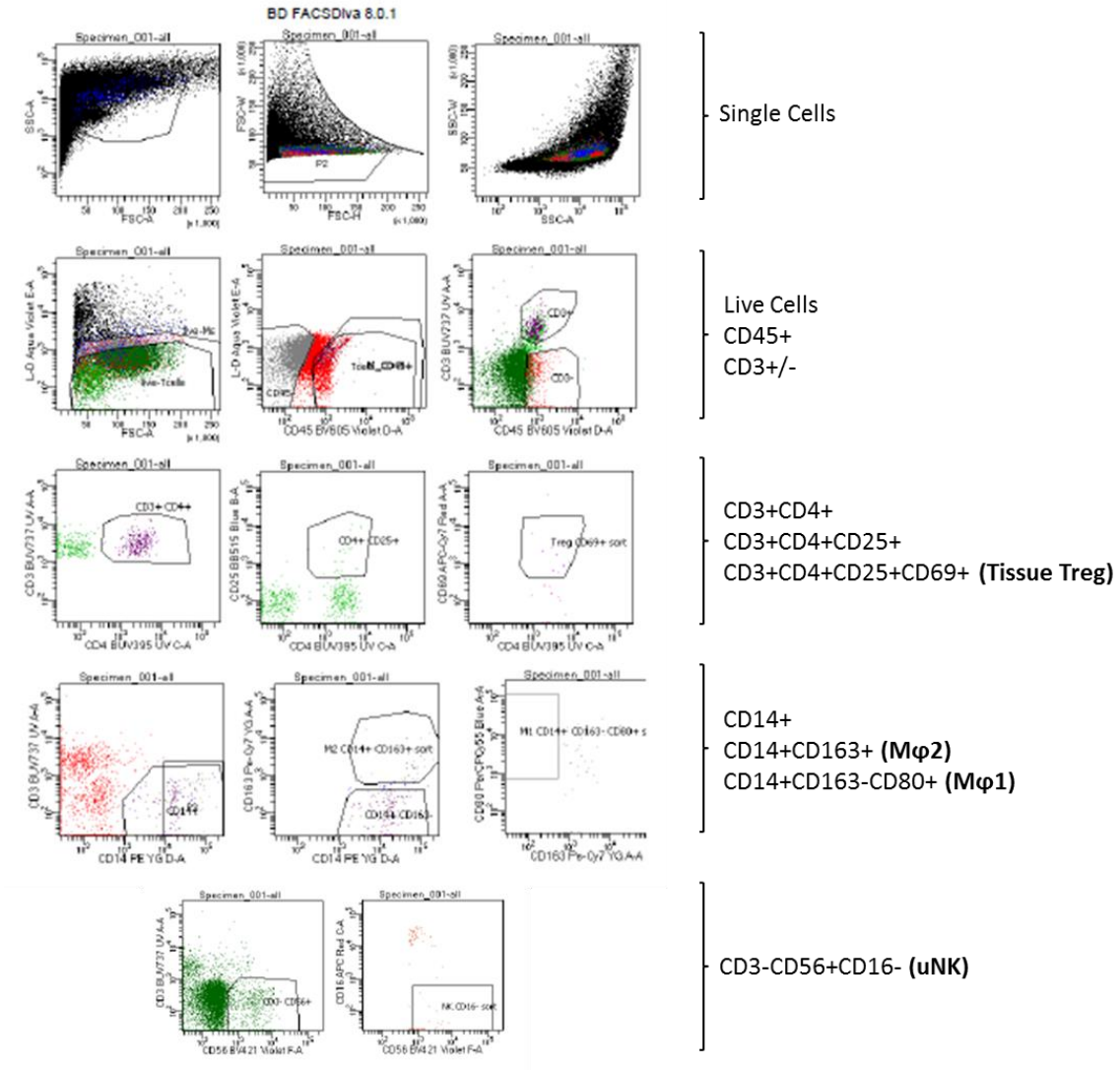


Figure 11. Gating for the immune populations sorting. In the upper panel it can be observed how cells are gated by size (SSC-A/FSC-A), and complexity (FSC-A/FSC-H) to obtain single cells. In the second row, live cells are first separated (Aqua Violet E/FSC-A), then CD45+ cells are gated (CD45 Violet E/Violet D) and CD3 positive and negative cells (CD3 UV A /CD45 Violet D). In the third row, Treg are separated, first by separating CD3+CD4+ cells (CD3 UV a/CD4 UV C). From them, CD25+ cells are gated (CD25 blue B/CD4 UV C). Finally, tissue resident Treg cells are isolated by CD69+ marker (CD69 Red A/ CD25 blue B). In the fourth row, Mφ are gated. First by CD14+ (CD3 UV C/ CD14 YG D), the by CD163 to separate Mφ2 (CD163 YG A/ CD14 YG D) and then by CD80+ to separate Mφ1 (CD80 blue A/ CD163 YG A). Finally, in the lower row, NK are gated by CD56+ (CD3 UV A/ CD56 Violet F) and resident NK (uNK) by CD16- (CD16 Red C/ CD56 Violet F).

3.6. RNA EXTRACTION AND QUANTIFICATION

A) From harvested cells

For the RNA extraction of harvested cells from primary cultures, 10^5 cells were collected in 350µl RLT with 1% β-mercaptoethanol and stored at -80ºC. RNA was isolated using a RNeasy micro kit (Qiagen) following the manufacturer's instructions. The concentration of RNA was measured by Nanodrop® photometer.

B) From sorted cells

1) Sorted LGR5

RNA of 38 samples LGR5+/- was extracted. Total RNA was isolated in two steps. First, cells were lysed using a 1ml syringe. After treatment with chloroform, an aqueous phase was precipitated with 70% ethanol in a volume proportion 1:1. Second, samples were passed through columns (Step 2 of RNAeasy micro kit, Qiagen). Then, the protocol was followed accorded to the manufacturer's suggested conditions. The quality of RNA was determined using Pico-chip by Bioanalyzer Agilent2100 (Agilent). All 38 samples used for RNA-High-Sequencing had a RNA Integrity Number (RIN) ≥7.

2) Sorted immune populations

To obtain the RNA, a RNeasy micro kit (Qiagen) was used. From the 40 sorted samples (11 samples with the four immune populations, except for four samples that were not valid), we extracted the RNA for the library prep from 12 samples. The library prep from the remaining samples was performed directly from cells in PBS 1x. We followed the manufacturer's instructions for RNA extraction. RNA was eluted in 10µl of RNase free-water and the quality and concentration were measured using a Tapestation 4200 System (Agilent). All of the extracted RNA passed the quality controls for the library preparation.

C) From FFPE

Five cuts of 10µm per sample were obtained from 285 FFPE samples. Cuts were deparaffinized using xylene and dehydrated using absolute ethanol. We extracted the RNA

with a High Pure FFPET RNA Isolation Kit (Roche), following the manufacturer's instructions. Nanodrop® was used to quantify concentrations of extracted RNA. Only the samples with more than a minimum of 300ng in 4µl and with a ratio of 260/280 > 1.8 were used for Nanostring® Technology, except for 51 samples that did not reach the required amount of nanograms (ng).

3.7. REAL TIME-QUANTITATIVE PCR (RT-qPCR)

RT-qPCR of specific markers for the proliferative phase (CYR61) and the secretory phase (DKK-1) was performed. LGR5 expression was also studied to determine its variation throughout the menstrual cycle. Primer pairs were created using a primer design bank (http://pga.mgh.harvard.edu/primerbank) and are shown in **Table 4.** All primers had a similar annealing temperature (~57ºC) and have low strength or do not have secondary structures. Eleven microliters of RNA were reverse transcribed into cDNA. Complementary cDNA was synthesized using SuperScript (SSIII) synthesis system. Reverse transcription was performed using 1µl random primers and deoxynucleotidphosphates (dNTPs), after heating for five minutes at 65ºC 1µl of DTT, RNAseOut, and SSIII in 4µl of buffer 5x. The thermic parameters were set as five minutes at 25ºC, one hour at 55ºC, and then 15 minutes at 70ºC thermocycler GS-1 (Ecogen). RT-qPCR was performed using the SYBR-green technique. A PCR reaction was performed using 30µl mix containing 1x faststart universal SYBR-green Master ROX (Roche), 0.3µl of each primer (30µM), 1.2µl DNA template, and 30µl milliQ water. RT-qPCR was done using a LightCycler480 Instrument II (Roche Life Science) and the thermo parameters were set at 10 min at 95ºC, 40 cycles of 15 seconds at 95ºC, and 30 seconds at 57ºC, and ending with 15 seconds at 95ºC, 15 seconds at 60ºC, and 15 seconds at 95ºC. As housekeeping, we used GAPDH and all data was normalized to its levels. Data was normalized using the untreated cells as controls and was analyzed using one-way or two-way ANOVAs followed by Bonferroni comparison tests with a p-value≤0.5.

Gene	Primer Name	Sequence 5'-3'	Tm
CYR61	hCYR61-For-25	CTCGCCTTAGTCGTCACCC	57.6
	hCYR61-Rev-226	CGCCGAAGTTGCATTCCAG	57.1
DKK1	hDKK1-For507	ATAGCACCTTGGATGGGTATTCC	56.6
	hDKK1-Rev-560	CTGATGACCGGAGACAAACAG	55.5
LGR5	hLGR5-For-71	CACCTCCTACCTAGACCTCAGT	57
	hLGR5-Rev-274	CGCAAGACGTAACTCCTCCAG	57.5
GAPDH	hGAPDH-For	CGT CTT CAC CAC CAT GGA GA	61.1
	hGAPDH-Rev	CGG CCA TCA CGC CAC AGT TT	56.7

Table 4. Primers for RT-qPCR. Forward (For) and reverse (Rev) primers used for the three tested genes: Cyr61, Dkk1, LGR5, and GAPDH. The table shows the primer name, its sequence, and the melting temperature (Tm).

3.8. *IN SILICO* STUDIES

In order to minimize the effect of the differences among stimulated (FSH) and non-stimulated cycles and between menstrual phases, we subtracted the differentially expressed genes (DEG) found to be significantly expressed throughout menstrual phases and stimulated a natural cycle according to previous reports in the literature[73,162]. The aim of this analysis was to complement a specific RNA High-Sequencing (RNA-Seq) data analysis by analyzing the DEG from two distinct independent microarray experiments. The first was to contain data from the DEG across the menstrual cycle[73], and the second was to contain the DEG between a natural and FSH stimulated endometrium[162]. Both studies were based on the same arrays' platform (Affymetrix Human Genome U133Plus2.0 Array) and RNA-Seq was used in the LGR5 study. However, several studies have shown that the different platforms (RNA-Seq and arrays) are comparable[163–170]. Two different approaches were completed in order to obtain the lists of DEG. The first one involved varying genes in a stimulated endometrium, downloading microarray raw data files from Gene Expression Omnibus (GEO) GEO Accession GSE19959, and processing the data files with the standard pipeline applied at the Statistics and Bioinformatics Unit of Vall d'Hebron Hospital (UEB). The second one involved varying genes across the menstrual cycle, going to the GEO Dataset GDS2052, and using the National Center for Biotechnology Information (NCBI) data analysis tool provided to directly download the lists. We filtered the genes found to be

significant in the RNA-Seq study with those coming from the comparisons made with GEO data. In addition, in the case of DEG obtained with UEB pipeline (stimulated versus non stimulated endometrium) some pre-analysis of the arrays was performed by arrayQualityMetrics 3.24.0 under R version 3.2.0 to prove that stimulated and non-stimulated arrays were comparable.

3.9. RNA-HIGH-SEQUENCING

Because five different RNA-High-Sequencings were performed for the different sub-objectives of this thesis, the collaboration centers where the library prep was produced and which kit was used to perform it are summarized in **Table 5**. Moreover, it shows the platform, the number of paired-end sequences that were used, and the center where the data analysis was performed. Below, an explanation of each main step of the process (library preparation, RNA-Seq, and bioinformatic analysis) is given in detail. It is worth mentioning that the quality and size of all RNAs was measured by RNA Pico Kit Bioanalyzer (Agilent) or by Tapestation System (Agilent) previous to library preparation.

	Study				
	Pilot study FFPE vs Fresh	Discovery study FFPE	Preliminary study LGR5	Complete study LGR5	Pilot study immune populations
Library prep Center	CRG	aScidea	VHIO	CRG	Genewitz
Library prep kit	TruSeq	TruSeq	TruSeq	TruSeq	Ultra-low input RNA-Seq
HighSeq Platform	Illumina Hi-Seq2000	Illumina Hi-Seq2000	Illumina Hi-Seq2000	Illumina Hi-Seq2000	Illumina Hi-Seq4000
Paired end sequences	2x50nt	2x100nt	2x50nt	2x50nt	2x150nt
Data Analysis Center	UEB	aScidea	UEB	UEB	CRG

Table 5. RNA-High-Sequencing studies. This table shows the five RNA-Seq studies conducted in this thesis. It shows the center where the sequencing was performed, the library preparation kit and platform used in each case, the paired end sequences used in each study and the center where the data was analyzed.

3.9.1. LIBRARY PREPARATION

3.9.1.1. Objective 1

A) Pilot study of FFPE versus fresh samples

To investigate the correlation between FFPE and fresh samples, we obtained RNA from six endometrial biopsies embedded in paraffin with their matched fresh endometrial biopsy. Three controls and three endometriosis samples were obtained.

The library preparation (library-prep) kit used was TruSeq Stranded Total RNA LT, w/Ribo-Zero Gold, Set A kit (Illumina). The manufacturer's instructions were followed. The workflow followed is shown in **Figure 12**.

Total RNA was briefly purified and fragmented. This process was followed by a complementary DNA (cDNA) synthesis. Then, the ends of the fragments were repaired and 3' end were adenylated. The adapters were ligated and PCR amplification was performed. After that, the libraries underwent quality controls and were normalized and pooled for RNA- Seq.

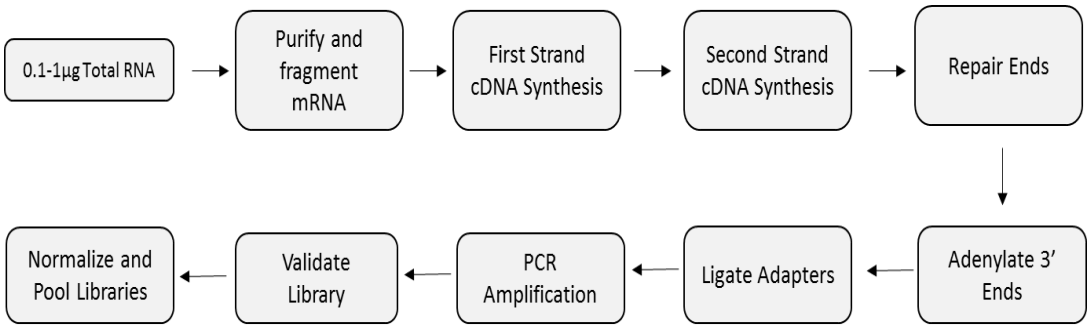


Figure 12. Library-prep workflow. This figure shows the workflow followed to prepare the library to perform RNA-Seq.

B) Discovery study for FFPE samples

In this case, we aimed to sequence 37 samples from the total endometrium (embedded in paraffin): 19 healthy and 18 endometriosis endometrial biopsies. The same kit as in the pilot study was used. The protocol followed was exactly the same as the described above.

3.9.1.2. Objective2

A) Preliminary study LGR5+ cells

A preliminary study to prove the feasibility of the sample to be sequenced was performed due to the scarce amount of LGR5+ cells after sorting and the very low RNA concentrations obtained. Four samples were used for this study (two from patients with endometriosis – specifically two endometriomas and two from egg donors).

After the separation of epithelium and stroma and the cell sorting (LGR5+/- cells), 16 samples were obtained and were submitted to library preparation. As can be observed in **Table 5**, the TruSeq Stranded Total RNA LT, w/Ribo-Zero Gold, Set A kit (Illumina) was also used, but in this case due to the small concentration of RNA, the fragmentation step of RNA was avoided to prevent RNA degradation.

B) Complete study LGR5+ cells

A total of 38 eutopic endometrial biopsies (five healthy donors, three endometriomas, four DIE, three pelvic endometriosis, and four adenomyosis), each one with their respective sorted LGR5+/- cells, were prepared for sequencing as explained in the previous paragraph.

C) Pilot study immune populations

A very small number of cells was obtained after the sorting. Therefore, a SMART-Seq™ v4 Ultra™ low input RNA kit for sequencing (Clontech) was used to perform the library preparation. It allows RNA-Seq to be performed with very small amounts of RNA or the use of whole cells to preserve sample integrity. In total, library preps for 40 samples (17 from healthy endometrium and 23 from endometriosis patients) were performed. RNA was extracted from samples with more than 950 cells (12 samples) using a RNeasy micro kit (Qiagen) and the library preps were performed. Library preps from samples containing up to 950 cells (28 samples) were prepared directly from cells that were in suspension with PBS.

3.9.2. RNA-HIGH-SEQUENCING

Illumina Hi-Seq2000 platform was used for all the studies except for the case of the immune populations' RNA-Seq, where Illumina Hi-Seq4000 was used instead. Hi-Seq4000 is comparable with Hi-Seq2000 with an improved total output and significant decreases in run times. In all cases, paired-end probes were used for the procedure. Different paired-ends were used depending on the study. In the pilot study of FFPE versus fresh samples and in the studies of LGR5+/- cells, 2x50nt was used. In the case of the discovery study for the biomarkers from FFPE samples, 2x100nt was used, and in the case of the immune populations, 2x150nt paired-end was used. Millions of sequences from each amplicon made in the library prep were obtained. The differences in length of nucleotides result in more or fewer millions of reads obtained after the sequencing. The gigabytes (GB) obtained for each sequencing were subjected to bioinformatic and statistical analysis.

3.9.3. BIOINFORMATIC AND STATISTICAL ANALYSIS

The main steps that are usually performed in RNA-Seq data analysis include the following: quality control, alignment and mapping, normalization, differential expression analysis, and pathway analysis. Quality controls after RNA-Seq were performed the same way in all centers. The following steps were performed in all cases and for each sequence obtained by RNA-Seq:

- **Initial quality control:** The quality of the sequences is analyzed after the sequencing of the experiments. The compositions of the sequences are analyzed to define the degree of quality of the overall sequencing process and to eliminate low quality sequences.
- **Data preprocessing:** Data preprocessing is used to eliminate the data that did not undergo the quality control in order to obtain the final data that will be used in further analysis steps.

- **Final quality control:** Once all samples have been tested for quality, quality reports are generated again and the high quality data is further inspected in order to confirm that the data kept is of the highest possible quality.

Because the data was analyzed in different centers for each project, the different methods used in each one are explained below. However, in all cases, pathway analyses were performed in our laboratory using Ingenuity Pathway Analysis® (IPA) (Ingenuity® Systems).

3.9.3.1. Objective 1

A) Pilot study FFPE versus fresh samples

Twenty-four-end FASTQ files (12 samples: six endometriosis (three fresh and three FFPE) and six control (three fresh and three FFPE) endometrial biopsies) were obtained. Data analysis was performed in the Unit of Statistics and Bioinformatics (UEB, Vall d'Hebron, Barcelona, Spain).

Basic quality controls were performed using FASTQC[171], FastX-Toolkit[172], and PRINSEQ[173]. Paired-end (forward-reverse) sample merging, as well as the remaining steps of the bioinformatics analysis, were performed with CLCBio Genomics Workbench® version 8.0.2[174,175] software. Alignment and mapping was done against the current human genome[176]. Counts were normalized with the standard RPKM method[177].

In the first place, the raw data was normalized, dividing each sample by its total counts of the library. A total of 42,599 genes was analyzed, but genes with a value of zero were eliminated. Then, Pearson correlations and a correlation plot of fresh and FFPE expressed genes for each patient were done. Finally, correlations for each gene in the samples from different origins were performed.

B) Discovery study of FFPE samples

The discovery study using FFPE samples was performed by aScidea Computational Biology Solutions, S.L. The quality of the 74 paired-end sequences (37 samples) obtained by HiSeq2000 sequencing was checked with FastQC[178] software. Preprocessing of the reads was performed with fastx-toolkit[179] and aScidea specific perl scripts – property of aScidea

(http://www.ascidea.com) – in order to filter out regions of low quality. Adaptors were trimmed, as were low quality bases at the ends of sequences and reads with undetermined bases or with 80% of their bases having a quality score of less than 20%. Raw reads that passed the quality filter threshold were mapped using Tophat 2.1.1[180] and Bowtie2 2.2.8[181] to identify known and novel splice junctions and to generate read alignments for each sample. The reference genome used was the Homo sapiens reference genome version GRCh38, and genomic annotations were obtained from Illumina iGenomes[182] in general feature format three (GFF3). The inner distance between mate pairs used was 50 bp, and the rest of the parameters were used with the default values. The transcript isoform level and gene level counts were calculated and FPKM normalized using Cufflinks 2.2.1 software[183]. Differential transcript expression was then computed using Cuffdiff. The resulting lists of differentially expressed isoforms were filtered by FC >1 and < -1 and a q-value of 0.05. Differences between each type of endometriosis versus the control were analyzed. In addition, differences in gene expression between types of endometriosis (ovarian, DIE, and adenomyosis) were investigated. Genes selected as being differentially expressed were clustered to look for common patterns of expression. Hierarchical clustering with Jensen-Shannon distance was used to form the groups and heatmaps were used to visualize them. The main statistical analyses were performed using the free statistical language R and the libraries developed for data analysis by the Bioconductor Project[184]. The DEG lists for each endometriosis type and healthy group were used to develop the biomarker classifier (*see Section 3.10*).

3.9.3.2. Objective 2

A) Preliminary study LGR5+ cells

A total of 32 paired-end (2x50) pairs (16 samples) of FASTQ files were obtained. Data analysis was performed by the Unit of Statistics and Bioinformatics of Vall d'Hebron (UEB). Basic quality controls were performed with FASTQC[171], FastX-Toolkit[172], and PRINSEQ[173]. Paired-end (forward-reverse) sample merging, as well as the remaining steps of the bioinformatics analysis, were performed with CLCBio Genomics Workbench® version 8.0.2[174,175] software. Alignment and mapping was done against the current human

genome[176]. Counts were normalized using the standard RPKM method[177]. Significance testing for DEG was implemented using EDGE testing[185] with a significance criteria of $\alpha<0.01$, adjusted for multiple testing with the FDR method[186]. Genes were considered statistically significant if their expression was FDR<0.01 with a FC±2. Different comparisons were studied: ovarian versus control, LGR5+ versus LGR5-, and epithelium versus stroma. Comparisons are shown in **Table 6**. DEG was found to be significantly expressed throughout menstrual phases in stimulated and natural cycles in the *in silico* studies (*see Section 3.8*) and was subtracted from the DEG lists.

Comparisons
1. Ovarian vs Control
2. LGR5+ vs LGR5-
3. Epithelium vs Stroma

Table 6. Comparisons made in the preliminary study of LGR5.

B) Complete study of LGR5+ cells

Bioinformatic analysis was again performed by UEB professionals using the same steps explained in the previous paragraph, and different comparisons were studied **(Table 7)**.

Comparisons
1. LGR5+Control vs LGR5+Endo
2. LGR5-Control vs LGR5-Endo
3. LGR5+Control vs LGR5-Control
4. LGR5+Endo vs LGR5-Endo
5. LGR5+Ovarian vs LGR5+Pelvic
6. LGR5+Ovarian vs LGR5+Adeno
7. LGR5+DIE vs LGR5+Ovarian
8. LGR5+Adeno vs LGR5+Pelvic
9. LGR5+DIE vs LGR5+Pelvic
10. LGR5+DIE vs LGR5+Adeno

Table 7. Comparisons made in the complete study of LGR5

Biological significance for Comparisons 2, 4, 7, 9, and 10 was done. In comparisons 2 and 4, biological significance analysis was performed searching for gene set enrichment analysis against KEGG[187] and GO databases[188] with GAGE[189] and Pathview[190] Bioconductor packages.

In order to determine unique genes expressed in DIE, common DEG and molecular functions in Comparisons 7, 9, and 10 were determined and their biological significance analyses were conducted using Ingenuity Pathway Analysis (IPA) (Ingenuity® Systems). The ideal set size for IPA core analysis from gene expression data is typically between 200 and 3,000. Therefore, in Comparisons 7, 9, and 10, a total of 3,000 codifying DEGs with the highest FC and FDR<0.01 in each comparison were analyzed. Subsequently, in order to minimize the effect of the differences between stimulated and non-stimulated cycles, DEG was found to be significantly expressed throughout menstrual phases in stimulated and natural cycles in the *in silico* studies (*see Section 3.8*) were subtracted from the DEG lists.

C) Pilot study immune populations

The quality of the fastq files was tested using the FastQC (v0.11.5)[171] and the Qualimap (rnaseq module – v2.2.1) softwares[191]. An estimation of ribosomal RNA in the raw data was obtained with riboPicker (v0.4.3)[192]. Reads were aligned with the STAR mapper (v2.5.2a)[193] to release 88 of the Homo sapiens ENSEMBL version of the genome (GRCh38/hg38 assembly)[194]. A raw count of reads per gene was also obtained with STAR (quantMode GeneCounts option). In order to overcome the high heterogeneity between samples (in terms of the number of raw reads and uniquely mapped reads), first, samples were removed from the analysis if they had less than five million uniquely mapped reads, and then the remaining samples were down-sampled to 30 million mapped reads when needed. The R/Bioconductor package DESeq2 (v1.20.0)[195–197] was used to assess differential expression between experimental groups (Wald statistical test + FDR correction). Prior to processing the differential expression analysis, genes for which the sum of raw counts across all samples was less than two were discarded. Different comparisons were made, which are shown in **Table 8**. A biological significance analysis was conducted using IPA software (Ingenuity® Systems).

Comparisons
1. Mφ1 Endo vs. Mφ1 Control
2. Mφ2 Endo vs. Mφ2 Control
3. Treg Endo vs. Treg Control
4. uNK Endo vs. uNK Control
5. Mφ1 Control vs. Mφ2 Control
6. Mφ1 Endo vs. Mφ2 Endo
7. Mφ Endo vs. Mφ Control

Table 8. Comparisons made between immune populations

3.10. CLASSIFIER DEVELOPMENT

This section corresponds to **Objective 1**. The differentially expressed transcripts found by RNA-Seq were analyzed along with the phenotypic data of the patients in order to determine correlations that determine the presence or absence of endometriosis.

Different analyses were done in order to elucidate if a classifier could be found for each type of endometriosis. There was a total of 37 samples to analyze, and the comparisons made were as follows:

- Adenomyosis (n=6) versus Control (n=19)
- DIE (n=4) versus Control (n=19)
- Ovarian (n=8) versus Control (n=19)
- Endometriosis (Adenomyosis + DIE + Ovarian; n=18) versus Control (n=19)

The phenotypic data analyzed comprised 11 variables: two continuous variables (concentration of RNA and age) and nine discrete variables. The variables included in this study are shown in **Table 9**.

Phenotypic Variables
RNA (ng/µl)
Age
Menstrual Cycle Phase
Endometriosis Clinical History
Hypermenorrhea
Dysmenorrhea
Fertility
Polyps
Myomas
Hormones
Smoker

Table 9. Phenotypic variables of patients included in the biomarker's discovery study.

The steps for the design of the classifier are explained one by one in the following:

1. Detection of batch effects
2. Comparison of statistical models and selection of the best
3. Expression model without RNA normalization
4. Expression model with data normalized by RNA concentration
5. Multi-class model with phenotypic variables (to differentiate between types of endometriosis)
6. Binary model with phenotypic variables (includes the total of endometriosis versus control)

1. Detection of batch effects

The possible batch effect regarding the phenotypic variable RNA was studied. The plots of RNA concentrations based on endometriosis groups and the control group were created.

2. Comparison of statistical models and selection of the best

The objective of this section is to summarize the results of the classificatory models obtained. The models used were:

1. Partial least square regression (PLS)
2. Support vector machine (SVM)
3. K-nearest neighbors (KNN)
4. Quadratic discriminant analysis (QDA)
5. Random Forest (RF)

All these algorithms were used to try to differentiate between control and endometriosis patients as well as to try to classify the different types of the disease. Two datasets were used. One dataset included the raw data and another included the normalized biomarkers (corrected by total RNA concentration). In order to validate the data obtained from the models, the leave-one-out-cross validation (LOOCV) method was used.

For the classificatory models between the control and endometriosis patients, the parameter used to determine which algorithm was better was the area under the curve (AUC). For multiclass models (the ones involving all classes of endometriosis and control patients), the parameter used was accuracy.

- The **AUC** is a parameter to evaluate the goodness of the test that will take values between 1 (perfect test) and 0.5 (useless test). In addition, this area can be interpreted as the probability that the test will correctly classify a pair of individuals, one sick and the other healthy.
- The **accuracy** describes the closeness of a measurement to the true value.
- The **kappa coefficient** is generally thought to be a more robust measure then simple percent agreement calculation, as kappa takes into account the possibility of the agreement occurring by chance.

3. Expression model without RNA normalization

The main objective of this section was to find the most important RNA sequence for differentiating patients with endometriosis. First of all, three principal component analyses (PCAs) were made in order to reduce a large set of variables to a small set that still contains the most of the information in the large set. The first PCA was made with the normalized data. The second was made with the log-transformed data, and the third was made with the same data as the second PCA but the groups were reduced to only two (binary model: endometriosis and control).

As the PLS model was the best of the methods, it was used for the further analysis. Three PLS models were performed because they create a linear regression by projecting the prediction variables and the observable variables into a new space. The first PLS model was made with the normalized data. The second was made with the log-transformed data and the third was made to predict whether a patient has endometriosis or is healthy (binary model). Finally, the most important independent variables (or regressors) affecting each model were obtained.

4. Expression model with data normalized by RNA concentration

As in the previous section, three PCAs were made with the objective of reducing the dimensionality of the data and visualizing it. Next, three PLS models were made with the objective of classifying patients as endometriosis or control and then distinguishing between types of endometriosis. Finally, the variables of importance were visualized for each model. First, three PCAs were made in order to reduce a large set of data to a small dataset. The first one was made with the normalized data. The second was made with the log-transformed data, and the third was also made with the log-transformed data but with only two groups (binary model). After that, a binary PLS model with normalized data was created. It aimed to predict whether a patient was in the control or endometriosis groups. Depending on the principal components used, a different receiving operating curve (ROC) of each model was obtained. In addition, a confusion matrix was also performed in order to visualize how many samples' corresponding groups (control or endometriosis) were correctly predicted by the model. The confusion matrix is a specific table layout that allows visualization of the performance of an algorithm. Each row of the matrix represents the instances in a predicted class, while each column represents the instances in a reference class. In addition, the accuracy-kappa and sensitivity-specificity parameters of the model were calculated. In addition, in order to classify between types of endometriosis, a multi-class PLS model with the normalized data was performed, where the accuracy-kappa parameter was calculated depending on the principal components used. The multi-class PLS model with the log-transformed data was also created, and the parameter-kappa

parameter was calculated depending on the principal components used. Finally, the most important regressors for each model were also observed.

5. Multi-class model with phenotypic variables (to differentiate between types of endometriosis)

The main objective of this section was to find the most important RNA sequences and phenotypic variables to differentiate patients with endometriosis and determine the type of endometriosis. Hence, PCA, PLS analysis using the log-transformed data, and finding the most important independent variables for the model were all carried out, as well as a phenotypic analysis.

6. Binary model with phenotypic variables (includes the total of endometriosis versus control)

The main objective of this section was to find the most important RNA sequences and phenotypic variables to differentiate patients with endometriosis. Again, PCA and PLS analysis using the log-transformed data were carried out, as well as the phenotypic characteristics and the calculation of the most important regressors.

3.11. NANOSTRING TECHNOLOGY

In order to validate the candidate biomarkers found in the classifier (Objective 1), Nanostring® Technology was used. The nCounter® analysis system uses a novel digital barcode technology that offers high levels of sensitivity. This technology uses CodeSets, which are molecular color-coded barcodes specifically designed for each transcript in a single reaction. They contain two 50bp probes (reporter and capture) that hybridize to the RNA target in solution. The reporter probe carries the fluorescent barcode signal and the capture probe immobilizes the hybridized complex for data collection.

The advantages of this technology are that it analyzes the nucleic acid in a single reaction so that samples from different origins (such as FFPE) can be used. To perform the assay, no

enzymes or library preparation are needed. In addition, the system is easy to use and the results are obtained in two days.

nCounter Nanostring® analysis:

1. Quality control prior to analysis
2. Sample processing
3. Post-processing quality control

1. Quality control prior to analysis

Nanostring® recommends the use of 250-300ng of RNA in 4µl, which is the volume that is added to the CodeSet solution containing the probes. Therefore, Nanodrop® analyses of the 285 samples were done before the sample analysis to decide which samples were valid for the study.

2. Sample processing

The sample does not need to be manipulated, as cDNA amplification is not needed. Hence, 4µl of RNA (250-300ng) were added to each well, which should already contain 20µl of master-mix (10µl Reporter CodeSet and 10µl hybridization buffer). The volume is adjusted to 25µl with water and 5µl of Capture probe was added. They were incubated at 65ºC for 12 hours.

Each color-coded barcode was hybridized to a single target-specific probe corresponding to the candidate of interest. After the hybridization process, the CodeSet was washed and the excess of probes was discarded. Then, the purified complexes were inserted into a sample cartridge where they are bind randomly to the surface. An electric current aligned the complexes for immobilization. In this case, 24 cartridges were needed (each one had capacity for 12 samples). Data was collected using a microscope by taking images of the surface where hundreds of thousands of fluorescent barcodes are bound, and the absolute values of the number of copies were obtained **(Figure 13).**

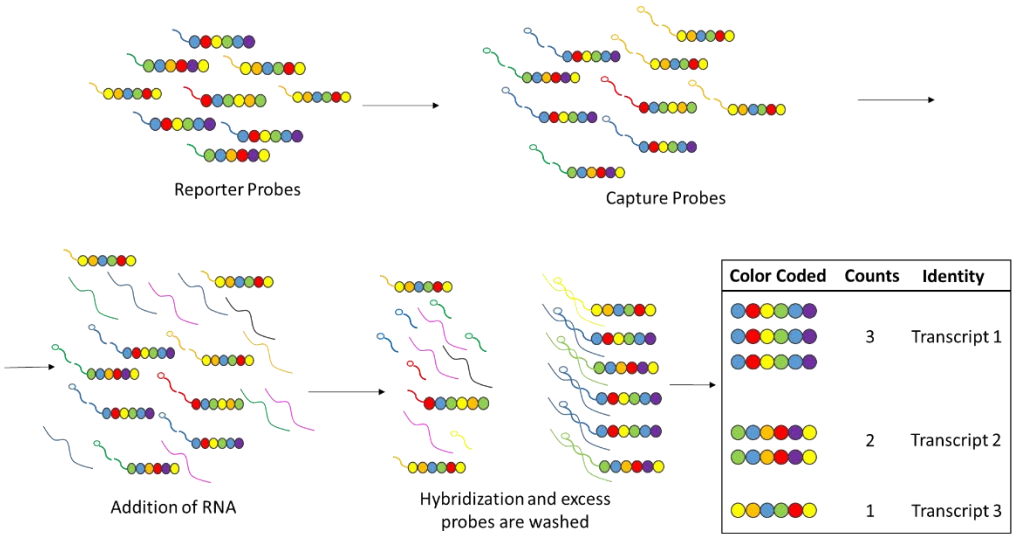


Figure 13. nCounter Technology. Reporter and capture probes are binded to the target. After hybridization, the excess probes are washed and the complexes are counted. Each complex corresponds to one copy of RNA. Then, the color-coded barcodes are counted and identified with their specific target.

3. Post-processing quality control

After sample processing, post-processing quality control was performed using nSolver™ Analysis software, which is an integrated analysis platform for storage, quality controls, and normalization of nCounter data.

3.12. VALIDATION OF THE DEVELOPED CLASSIFIER

In order to validate the model, the sample size was increased from 37 to 285.

First of all, correlation between technologies, RNA-Seq and Nanostring nCounter®, was calculated. After that, two more PLSs were done, one model with only the genetic variables and one with only the phenotypic variables and the objective of observing which kinds of variables were the best predictors. Control or endometriosis class was predicted using the previously developed PLS model. A confusion matrix, statistic parameters (accuracy, kappa, sensitivity, and specificity), and the ROC curve were calculated for each PLS.

Results

4. RESULTS

This section is divided in two sections that correspond to the two specific objectives of this thesis:

1. To discover new biomarkers in the eutopic endometrium of women with endometriosis

2. To evaluate the potential of LGR5 as a biomarker for the disease and the implication of LGR5 and immune cells in the pathophysiology of the disease

In order to give the reader context, a small summary is given at the beginning of the section corresponding to each objective.

4.1. Discovery of new biomarkers in the eutopic endometrium of women with endometriosis

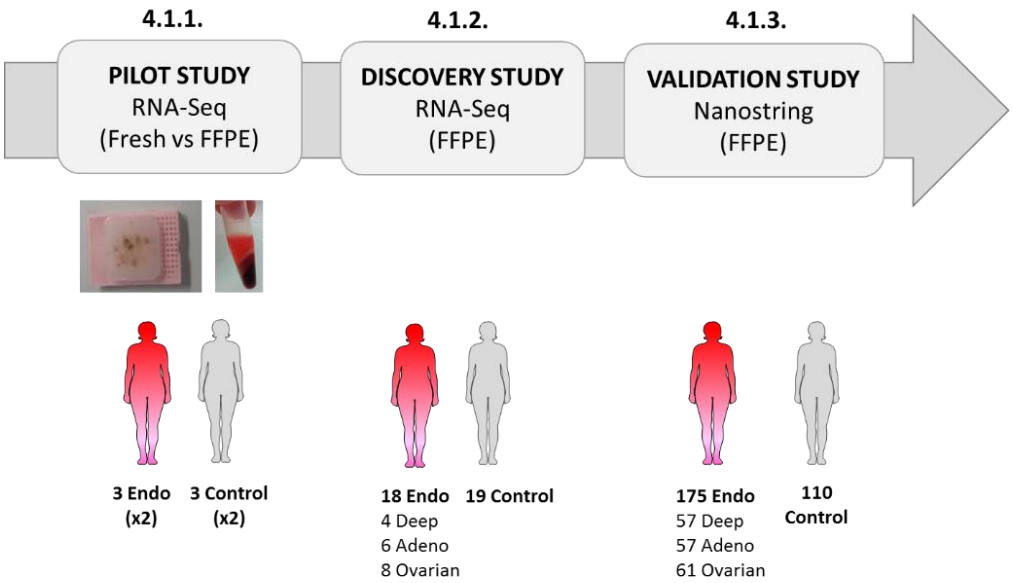


Figure 14. Workflow followed to achieve Objective 1. Each study is numbered as the section where it is explained. Endo: endometriosis; Deep: deep infiltrating endometriosis; Adeno: adenomyosis.

4.1.1. Fresh and FFPE samples are comparable in terms of gene expression

The aim of this study was to evaluate whether fresh eutopic endometrial and FFPE samples were comparable in terms of the gene expression obtained by RNA-Seq. Twelve samples were used: six paired samples from control women (three fresh and three FFPE) and six from women with deep infiltrating endometriosis (three fresh and three FFPE). From this point on, control samples are represented in graphs as N and endometriosis samples as D.

After RNA extraction, between 50-4500ng/µl of RNA were obtained depending on the sample. The samples had to pass the quality controls previous to the library prep. The minimal amount of RNA used for the library prep in this case was 50ng. After the RNA-Seq, quality controls for the fastQ sequences obtained were also performed. The following shows an example of the quality report obtained for one sequence. This report was done for each fastQ sequence obtained in all the RNA-Seq studies.

The first step was to determine the distribution of the per-base quality of the sequences **(Figure 15)**.

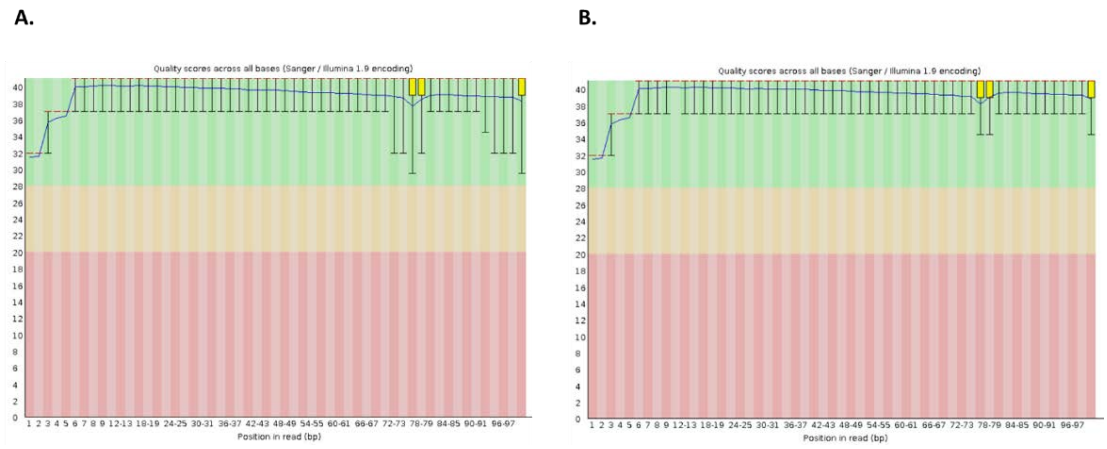


Figure 15. Distribution of per-base quality of the sequences. The plot shows the distribution of base qualities (Y-axis) per position in reads in all sequences (X-axis). The red line in the box represents the median value of quality. The yellow represents the interquartile range (IQR, 25% to 75% of the data in a normal distribution of values). The blue line that spreads out through all boxplots shows the mean quality of the bases. The background colored regions correspond to good (green zone), medium (orange zone), and low (red zone) quality. **A) Per base quality score before preprocessing step. B) Per base quality score after preprocessing step.**

As it can be observed in the previous figure, the sequences were of good quality (represented by the green zone on the graph). After the preprocessing of the data, the distribution of the reads was more uniform. This was observed for all sequences. Then, Per Base Sequence Content plots were performed **(Figure 16)**.

A. **B.**

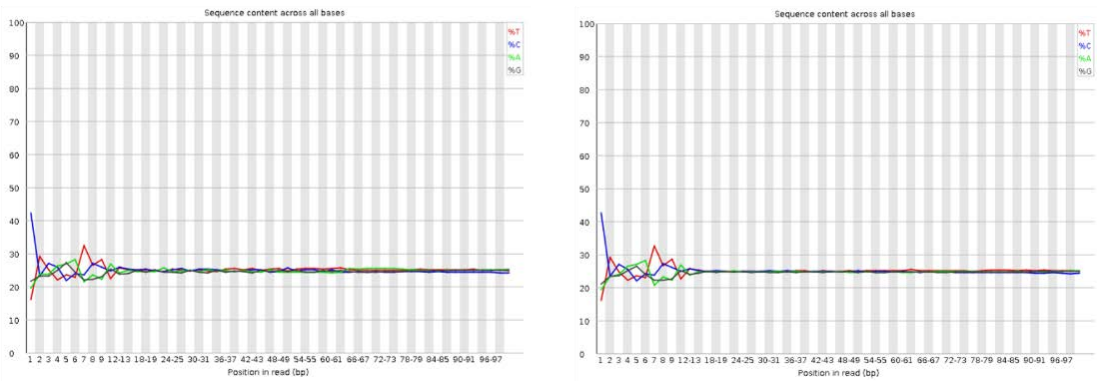


Figure 16. Per-base sequence content plots. These plots show the measurement of the bases content in order to observe the differences between bases of a sequence run. The lines should run parallel to each other in the plot. **A) Per-base sequence content before preprocessing of the data. B) Per-base sequence content after preprocessing of the data.**

After preprocessing the data, the per-base sequence contents were more equal. This can be observed in the previous figure.

In addition, the guanine/cytosine (GC) content was also measured across the whole length of each sequence in a sample and compared to a modeled normal distribution of GC content **(Figure 17)**. In a normal random library, it would be expected to see a normal distribution of GC content where the central peak corresponds to the overall GC content of the underlying genome. A reference distribution curve was used to compare the GC content. If the shape of the distribution is not normal, it would indicate a contaminated library. A normal distribution that is shifted indicates some systematic bias which is independent of base position.

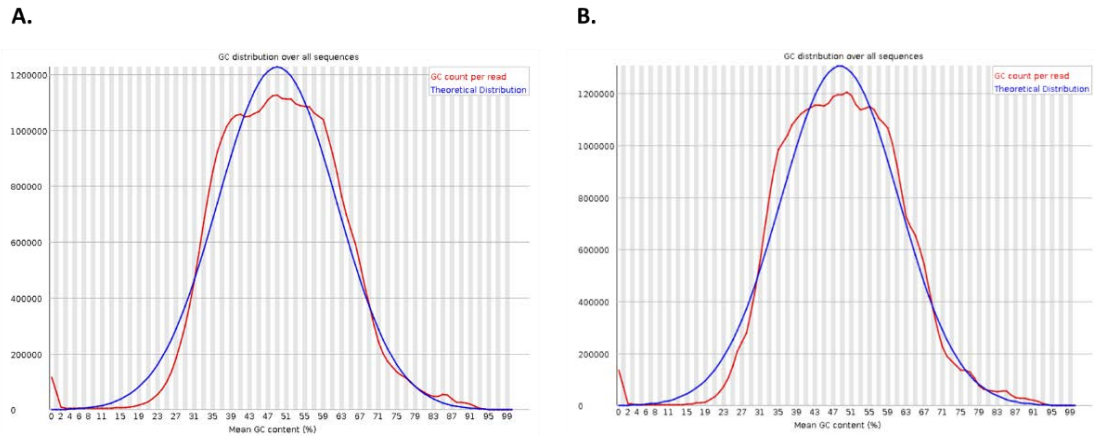


Figure 17. Per-sequence GC content. A) Before preprocessing data. B) After preprocessing data. Blue: theoretical distribution. Red: GC count per read.

As can be observed in the previous figure, the GC content distribution was similar to the reference curve.

After this, per base nitrogen (N) content was measured **(Figure 18)**. If a sequencer is unable to make a base call with sufficient confidence then it will normally substitute an N rather than a conventional base call. **Figure 18** shows the percentage of base calls at each position for which an N was called. It's not unusual to see a very low proportion of Ns appearing in a sequence, especially nearer the end of a sequence. However, if this proportion rises above a few percent, it suggests that the analysis pipeline was unable to interpret the data well enough to make valid base calls.

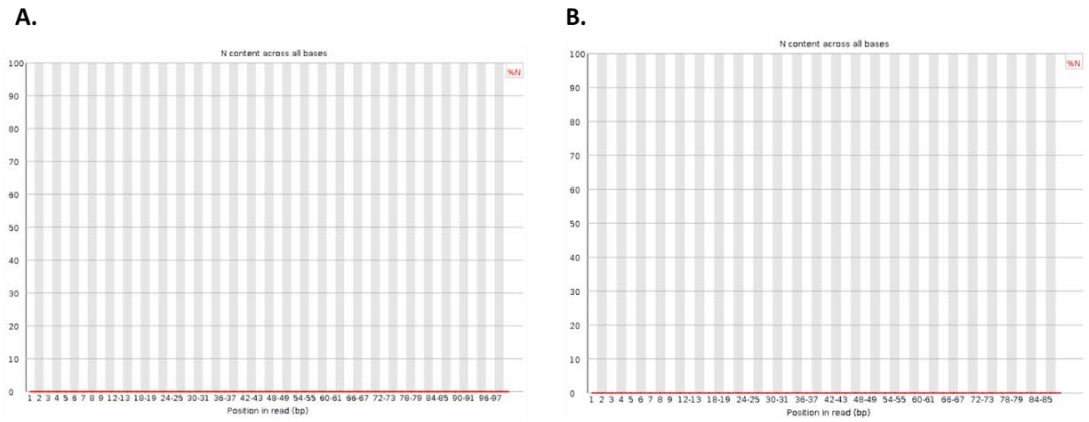


Figure 18. Per-base N content. A) Before preprocessing data. B) After preprocessing data. Red: %N content in the sequence.

As **Figure 18** shows, there was almost 0% N content in the sequences, which suggests that the sequencer was able to interpret the data well.

Sequence length distribution was also analyzed **(Figure 19)**. Some high throughput sequencers generate sequence fragments of uniform length, but others may contain reads of wildly varying lengths. Even within uniform length libraries, some pipelines will trim sequences to remove poor quality base calls from the end. **Figure 19** shows the distribution of fragment sizes. In many cases, this will produce a simple graph showing a peak only at one size, but for variable length FastQ files, this will show the relative amounts of each different size of sequence fragment.

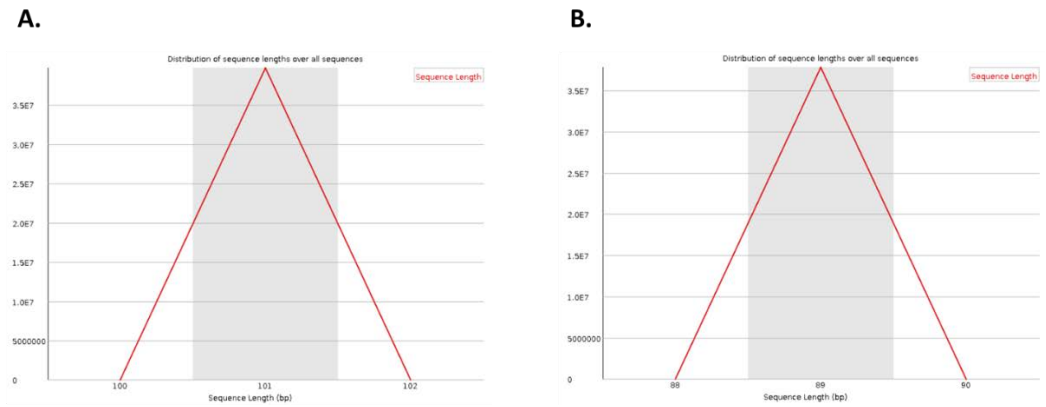


Figure 19. Sequence length distribution per sequence. A) Before preprocessing data. B) After preprocessing data. Red: sequence length.

Finally, the duplication levels were measured **(Figure 20)**. Most sequences would occur only once in the final set. A low level of duplication may indicate a very high level of coverage of the target sequence, but a high level of duplication is more likely to indicate some kind of enrichment bias. This module counts the degree of duplication for every sequence in a library and creates a plot showing the relative number of sequences with different degrees of duplication.

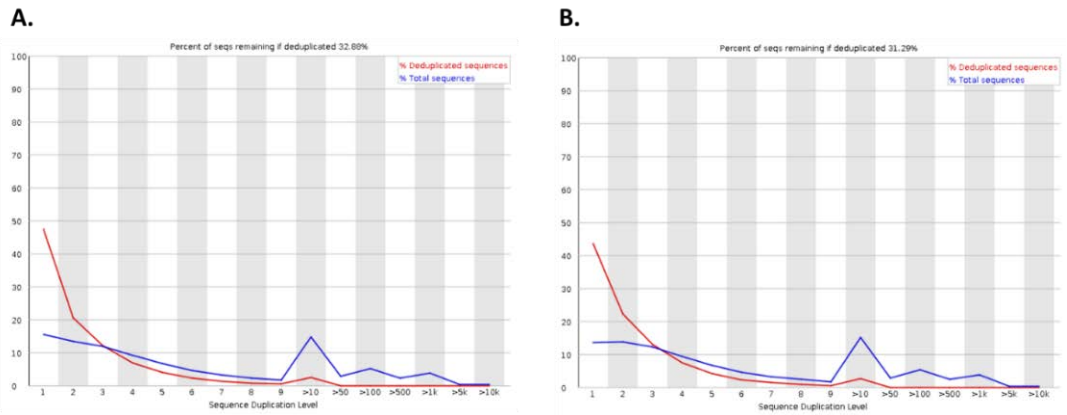


Figure 20. Duplication percentage. A) Before preprocessing data. B) After preprocessing data. The plot shows the proportion of the library that is made up of sequences in each of the different duplication level bins. The blue line shows how the duplication levels the full sequence set are distributed. The red line represents the results when the sequences are de-duplicated. The proportions shown are the proportions of the de-duplicated set, which come from different duplication levels in the original data.

Observing high rates of read duplicates in RNA-seq libraries is common. It may not be an indication of poor library complexity caused by low sample input or over-amplification. It might be caused by such problems, but it is often caused by a very high abundance of a small number of genes (usually ribosomal or mitochondrial housekeeping genes). Sometimes 75% of all reads map to the top 0.1% of expressed genes. The result of such heavy sampling of these genes is a high number of duplicate reads (even when considering read pairs in assessing duplicates)

Once the quality controls for all sequences were completed, different comparisons were made in order to observe the correlation of RNA-Seq data from fresh and FFPE samples from the same patients. First, the raw data was normalized, dividing each sample with its total counts of the library (reads). As can be observed in **Figure 21**, the library size was smaller in the FFPE samples.

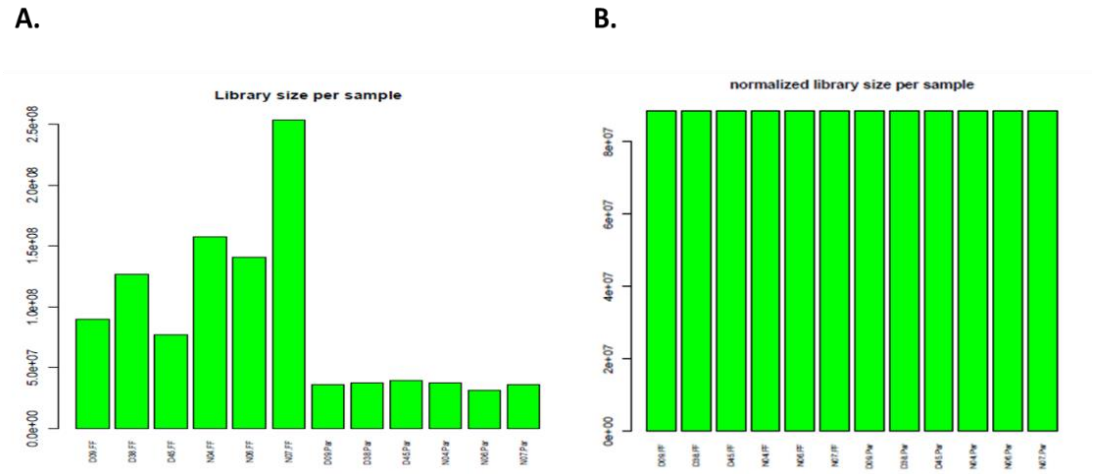


Figure 21. Normalization of the raw data per library size. On the Y-axis, the number read of each library is shown. The samples from the six patients are on the X-axis. The first six correspond to the fresh samples and the last six to the FFPE samples. **Panel A** shows the library size per sample and **Panel B** shows the normalized library sizes per sample.

Then, Pearson correlations of fresh and FFPE expressed genes for each patient and a correlation plot were done. The results can be observed in **Figure 22.A** and in **Figure 22.B** respectively.

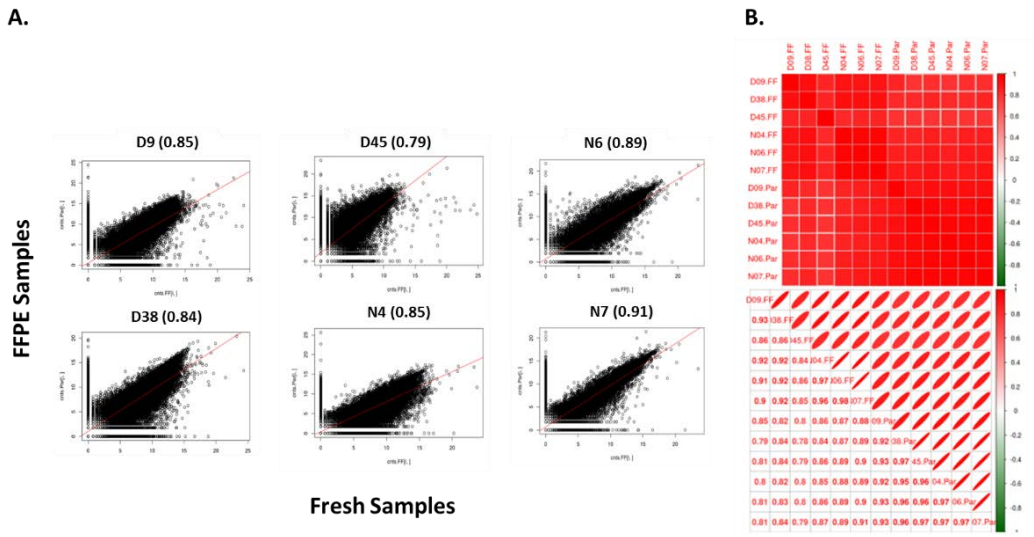


Figure 22. A) Correlations of fresh and FFPE expressed genes per patient. The Y-axis represents the FFPE sample and the X-axis represents the fresh sample. On the top of each patient graph, the name (N for control and D for endometriosis) and the Pearson correlation are shown. **B) Correlation plot** where all patients are shown. Red: correlation close to 1; green: correlation close to -1.

There was a high Pearson correlation between paired samples of different origins **(Figure 22.A)**. In the case of the control patients (N), correlations of 0.89, 0.85, and 0.91 were observed. In the case of the endometriosis patients (D), the correlations observed were 0.85, 0.79, and 0.84. In addition, looking at **Figure 22.B**, where the correlations were done sample by sample, the same conclusion can be deduced.

Finally, correlations for each gene in the samples with different origins were performed. Some examples are shown in **Figure 23**.

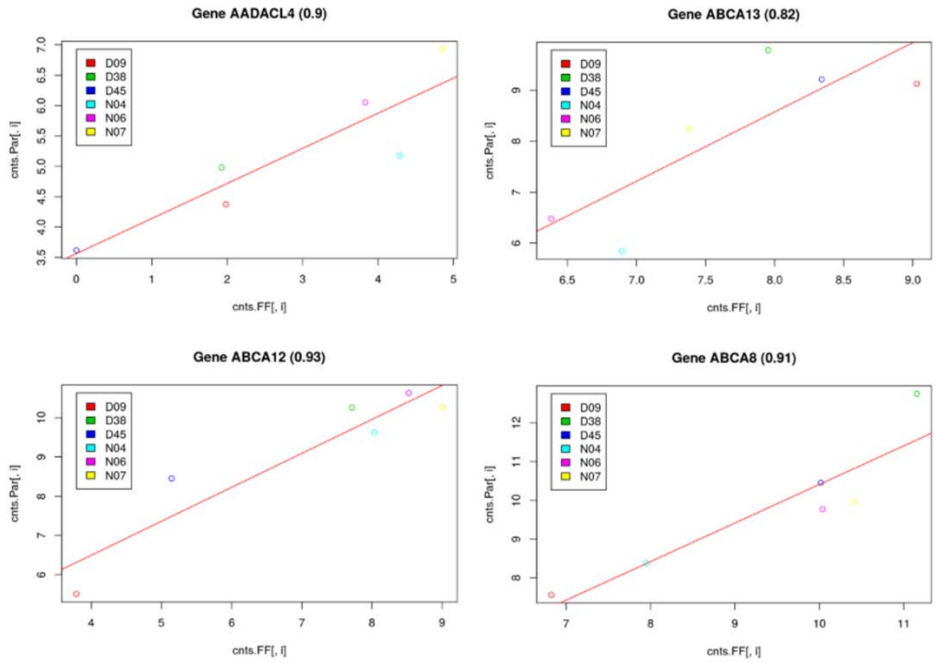


Figure 23. Correlation of samples in each single gene. An example of the correlation of four genes (AADACL4, ABCA13, ABCA12 and ABCA8) in the six patients is shown on the figure. The Y-axis represents the FFPE samples and the X-axis represents the fresh samples. Each patient (endometriosis coded as D and control coded as N) is shown by different color circles.

In general, many genes presented a high correlation. A correlation higher than 0.5 was observed in 7,296 genes, while a correlation higher than 0.75 was found in 2,166 genes. A correlation higher than 0.95 was found in 348 genes.

After the correlation studies, it was concluded that FFPE and fresh samples were comparable in terms of gene expression. Therefore, we decided to conduct the discovery and validation studies using FFPE samples.

4.1.2. Differences between endometriosis and control eutopic endometrium gene expression

In order to discover new biomarkers for the disease from the eutopic endometrium of women with endometriosis, a discovery study using RNA-Seq from 19 control women and 18 endometriosis women was performed.

The concentration of RNA obtained was between 500-1750ng/µl. However, this was from the paraffin embedded samples, which were previously fixed with formaldehyde. This type of sample is usually much degraded. When the quality of the RNA was analyzed by Bioanalyzer (Agilent), the RIN numbers obtained were very low (from 1.2 to 2.6). However, these results were expected and the RNA-Seq was carried out. RNA-Seq analysis and differentially expressed transcripts analysis were completed. In **Figure 24**, an example of the percentage of mapped sequences to the human genome is shown.

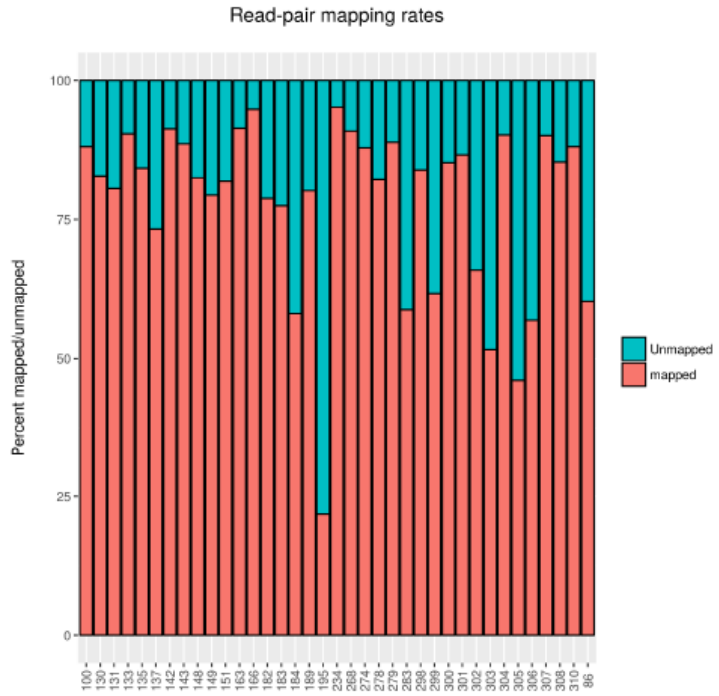


Figure 24. Read-pair mapping rates to the human genome. The Y-axis shows the percentage of mapped/unmapped reads and the X-axis shows the sequenced samples. **Red:** mapped reads. **Blue:** unmapped reads.

For each comparison of study, different analyses were carried out in order to visualize the differences between groups. An example of the graphs obtained is shown below. The comparison used in this example was endometriosis versus control. The endometriosis group includes the three types of endometriosis.

First of all, the levels of gene expression were visualized with: **1)** a per-sample dispersion plot, **2)** a per-sample density plot of the gene expression, and **3)** a per-sample boxplot of the total gene expression **(Figure 25)**.

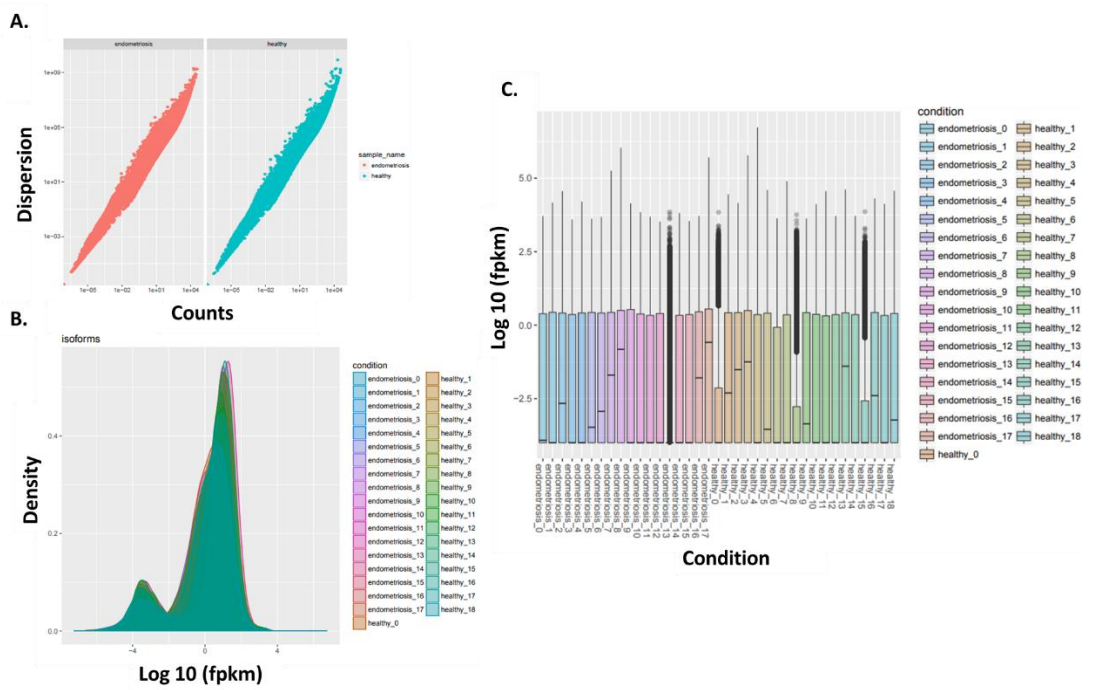


Figure 25. A) Per-sample dispersion plot. The dispersion of the expression of control and endometriosis groups. Red: endometriosis; blue: control. **B) Per-sample density plot of gene expression.** The density of isoforms per sample. Each color represents one sample. **C) Per-sample boxplot of the total gene expression.** This graph shows the number of fragments per kilobase million (fpkm) per sample. Each color represents one patient.

As can be visualized in the previous figure, the dispersion and distribution was similar between the two groups. In addition, the number of reads (log 10 fpkm) was similar in all samples except for four samples, where the number of reads was lower (one endometriosis and three controls).

After the previous step, the correlation between samples was analyzed and different graphs were made for visualization: **1)** a scatter-blot matrix between samples, **2)** a dendogram, **3)** a MA plot of the contrast tested, and **4) a** PCA plot for sample to sample differences **(Figure 26)**.

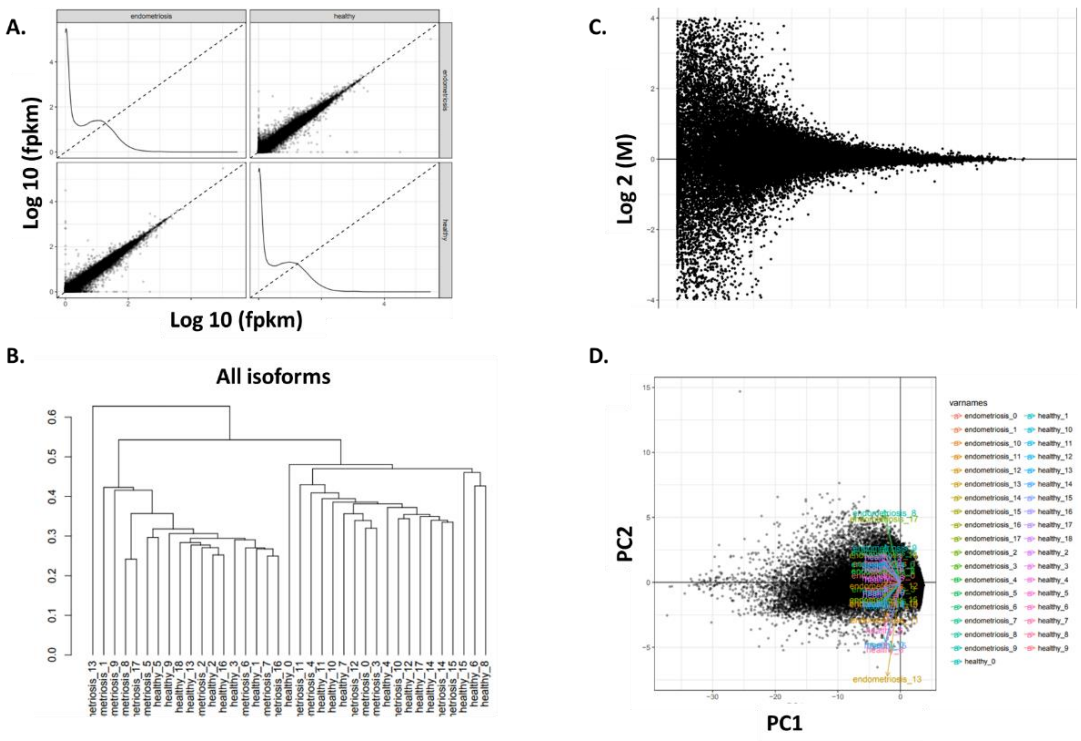


Figure 26. A) Scatter-blot matrix between samples. Correlation of fpkm between endometriosis (left) and control (right) can be observed. **B) Dendogram.** The graph shows the distribution in clusters of the samples (endometriosis and control). **C) MA plot of the contrast tested.** Visualization of the differences between measurements taken in both samples. **D) PCA plot for sample to sample differences.** Principal component analysis for each sample to determine the distribution of both groups.

As can be observed in the graphs of **Figure 26**, the two groups are correlated and did not cluster in different groups **(Figure 26.B)**.

Finally, the identification of differentially expressed genes was performed (q-value<0.05) and visualized as a volcano plot **(Figure 27)**.

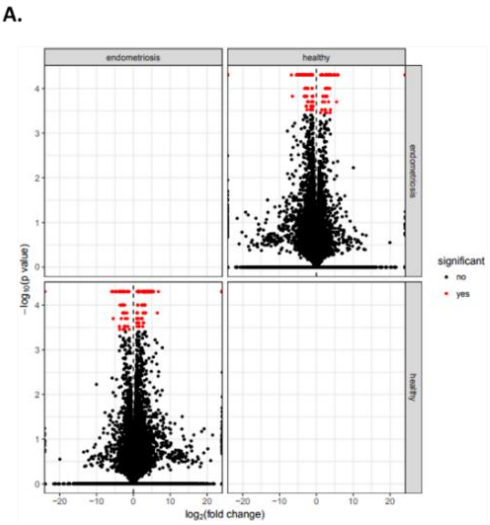


Figure 27. Volcano plot of the contrast tested. The the differentially expressed genes in endometriosis (left) and control (right) groups are shown in red.

After differential expression analysis, 182 differentially expressed transcripts (DETs) were found between endometriosis patients and the control group, with a q-value<0.05 and a FC±1 **(Table 34 in Annex 1)**. Differences between each type of endometriosis versus the control group were also analyzed. The total number of DETs is shown in **Table 10**.

In addition, differences in gene expression between types of endometriosis (ovarian, DIE, and adenomyosis) were investigated and no significant differences were found.

	Comparison			
	Endo vs Ctr	**Ov vs Ctr**	**Adeno vs Ctr**	**Deep vs Ctr**
nº DET (Differentially Expressed Transcripts)	182	35	62	28
nº up	54	23	14	24
nº down	127	12	48	4

Table 10. Comparisons and differentially expressed transcripts (DETs) of endometriosis versus controls. The different comparisons analyzed are listed: Endo vs. Ctr (endometriosis versus control), Ov vs. Ctr (ovarian versus control), Adeno vs. Ctr (adenomyosis versus control) and DIE vs. Ctr (deep infiltrating endometriosis versus control). In the first row, the total DETs are shown. The second row gives the total number of up-regulated DETs in endometriosis and the third gives the total downregulated DETs.

4.1.3. Binary model allows the classification of endometriosis patients

After obtaining the DET lists, the classifier was developed to be able to differentiate between endometriosis and control but also between types of endometriosis. As it was explained in the materials and methods section, six steps were performed:

1. Detection of batch effects
2. Comparison of statistical models and selection of the best
3. Expression model without RNA normalization
4. Expression model with data normalized by RNA concentration
5. Multi-class model with phenotypic variables (to differentiate between types of endometriosis)
6. Binary-class model with phenotypic variables (includes all types of endometriosis versus control)

1. Detection of batch effects:

The possible batch effect regarding the phenotypic variable RNA was studied. A plot of the distribution of concentrations was generated and **Figure 28** was obtained.

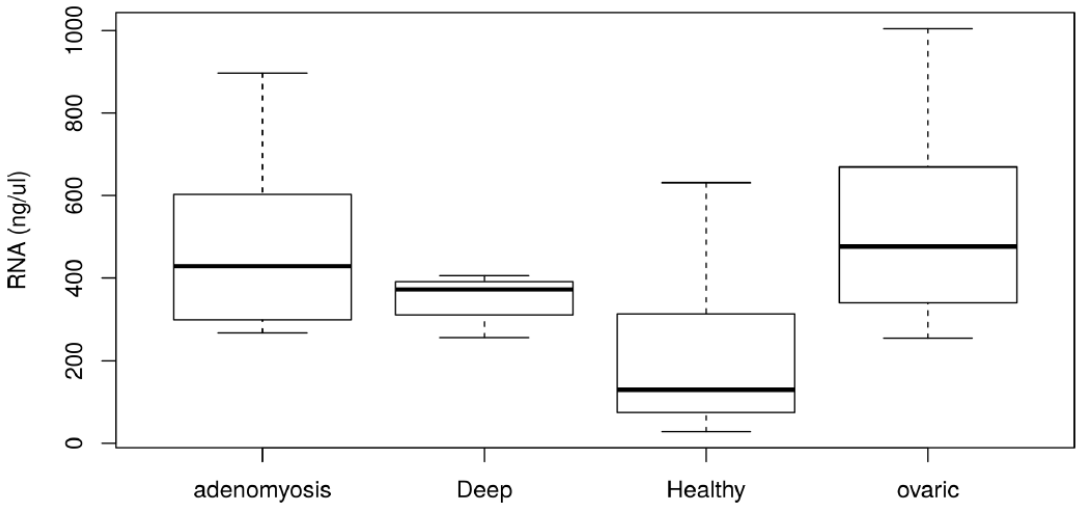


Figure 28. Plot of RNA concentrations of each group. RNA concentration (ng/µl) is shown on the Y-axis and the four groups (adenomyosis, DIE "deep," control "healthy", and ovarian) are shown on the X-axis.

As shown in **Figure 28**, medians of concentrations for the endometriosis classes are higher than the median of the control group. This could suggest that RNA extraction procedures were different or samples were taken at different time or processed differently. However, RNA extraction was performed randomly and samples were taken from the Pathology Department of Vall d'Hebron University Hospital.

2. Comparison of statistical models and selection of the best model

The objective of this section is to summarize the results of the obtained classificatory models. Five models were used and the ROC, accuracy, kappa, sensitivity, and specificity were calculated using the raw and normalized data for the classification of endometriosis and controls (**binary-class model**) **(Table 11 and Table 12)**.

Model	ROC	Accuracy	Kappa	Sensitivity	Specificity
PLS	1.000	0.973	0.946	0.944	1.000
SVM	0.971	0.892	0.784	0.944	0.842
KNN	0.994	0.946	0.891	0.889	1.000
QDA	0.816	0.757	0.515	0.833	0.684
RF	0.982	0.919	0.837	0.889	0.947

Table 11. Five classificatory models results with raw data. The table shows the ROC, accuracy, kappa, sensitivity, and specificity for each classificatory model.

Model	ROC	Accuracy	Kappa	Sensitivity	Specificity
PLS	1.000	0.973	0.946	0.944	1.000
SVM	0.968	0.919	0.838	1.000	0.842
KNN	1.000	0.946	0.891	0.889	1.000
QDA	0.895	0.811	0.622	0.833	0.789
RF	0.985	0.919	0.837	0.889	0.947

Table 12. Five classificatory models results with normalized data. The table shows the ROC, accuracy, kappa, sensitivity, and specificity for each classificatory model.

In addition, the five models were used to determine the accuracy and kappa parameters of the **multiclass model**, which would allow differentiation between types of endometriosis. The results using the raw and the normalized data are shown in **Table 13**.

	Raw data		Normalized data	
Model	**Accuracy**	**Kappa**	**Accuracy**	**Kappa**
PLS	0.757	0.598	0.703	0.494
SVM	0.73	0.58	0.595	0.331
KNN	0.649	0.362	0.676	0.45
QDA	0.486	0.218	0.514	0.271
RF	0.676	0.458	0.73	0.537

Table 13. Five multiclass classificatory models results with raw and the normalized data. The table shows the accuracy and kappa for each model.

As can be observed in the previous tables, the algorithm that had the best classificatory results for both datasets (binary-class and multiclass) was the PLS. In addition, the obtained results indicated better results with the raw dataset than with the normalized dataset, probably due to the influence of the putative RNA concentration batch effect detected.

3. Comparison of the expression model without and with RNA normalization (Steps 3 and 4)

The main objective of this section was to find the RNA sequence that is most important in differentiating patients with endometriosis. Three PCAs were made without RNA normalization and three with RNA normalization. In each case, the first one was made with the normalized data, the second with the log-transformed data, and the third with the log-transformed data but only of the endometriosis and control groups (binary model)**.** Because the PLS model was the best of all methods, it was used for the further analysis. Three PLS models were performed in each case. The first PLS model was made using the normalized data. The second was made using the log-transformed data, and the third was made with the log-transformed data but as a binary model (endometriosis versus control). The following table shows the results of the six PLSs.

	Without RNA normalization						With RNA normalization					
	Normalized data		Log-transform data		Log-transform data (Binary model)		Normalized data		Log-transform data		Log-transform data (Binary model)	
nComp	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	0.51	0.15	0.59	0.29	0.76	0.51	0.51	0.15	0.51	0.15	0.76	0.51
2	0.7	0.49	**0.65**	**0.43**	0.84	0.67	0.7	0.49	0.7	0.49	0.83	0.65
3	0.73	0.53	0.65	0.42	**0.86**	**0.73**	0.73	0.53	0.73	0.53	0.86	0.7
4	**0.73**	**0.54**	0.65	0.42	**0.86**	**0.73**	**0.73**	**0.54**	**0.73**	**0.54**	**0.865**	**0.728**
5	**0.73**	**0.54**	0.62	0.39	0.84	0.68	**0.73**	**0.54**	**0.73**	**0.54**	0.84	0.7
6	0.7	0.5	0.59	0.35	0.84	0.67	0.7	0.5	0.7	0.5	0.84	0.67
7	0.65	0.41	0.59	0.36	0.73	0.46	0.65	0.41	0.65	0.41	0.73	0.46
8	0.65	0.43	0.57	0.32	0.78	0.56	0.65	0.43	0.65	0.43	0.76	0.56
9	0.62	0.37	0.59	0.38	0.76	0.51	0.62	0.37	0.62	0.37	0.78	0.51
10	0.57	0.26	0.49	0.24	0.76	0.51	0.57	0.26	0.57	0.26	0.78	0.51

Table 14. PLS model with the normalized data, log-transformed data, and binary model data. The table shows the parameters' accuracy and kappa depending on the number of principal components used (nComp). The best results in each case are highlighted in grey.

The results indicated that Models 3 and 6 (binary models) could differentiate with 86% accuracy if a patient is a control or has endometriosis independently on the RNA normalization. Models 1, 4, and 5 could differentiate patients with different types of

disease with 73% of accuracy. Finally, the most important variables affecting each model were obtained and are shown in the following figures.

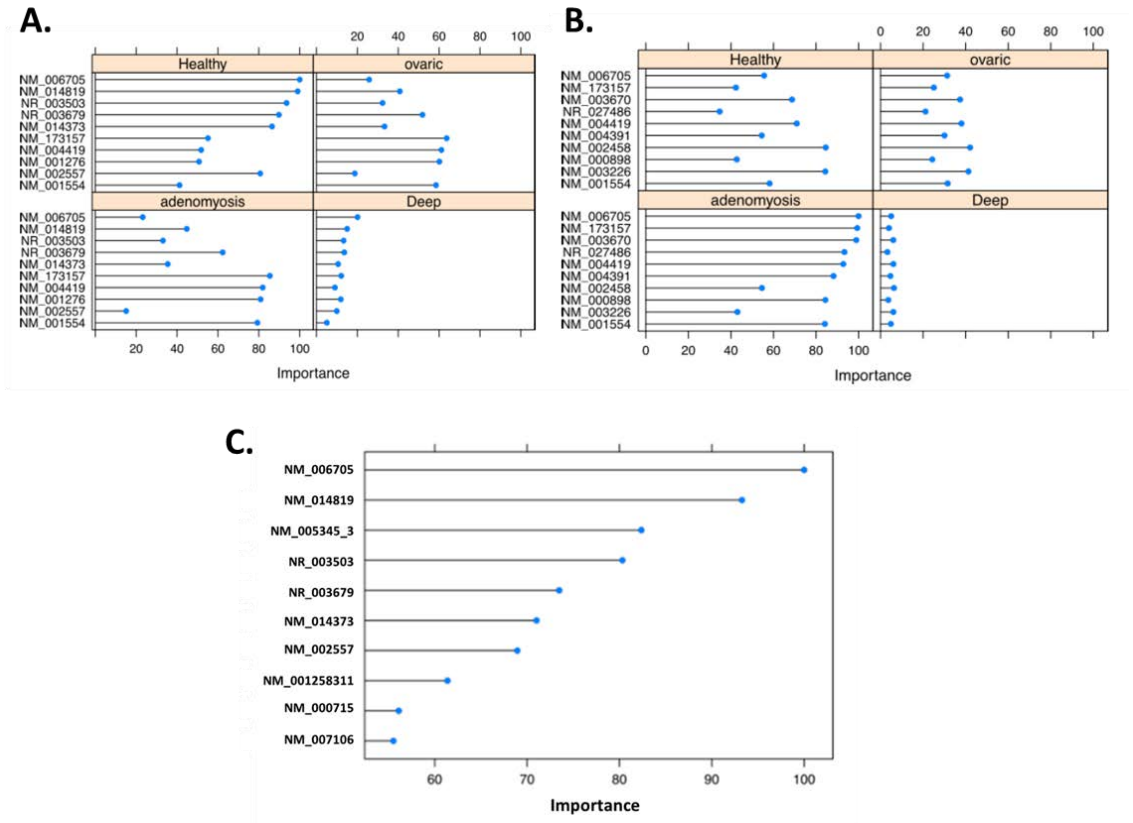


Figure 29. Importance for regression in PLS model without RNA normalization. A) With the normalized data. **B)** With the log-transformed data. **C)** Binary model with the log-transformed data. The most important regressors for each group are shown on the Y-axis and the percentage of importance on the X-axis.

In the previous figure, it can be observed that DIE did not show RNA transcripts as much as the other classes in the PLS model without RNA normalization. The profiles of adenomyosis and control are clearly differentiated from ovarian and DIE. When the contrast endometriosis and control was tested, three important RNAs were found: NM_006705, NM_014819, and NM_005345_3. Thus, those RNAs are candidates for further study to determine their true relationship with the disease.

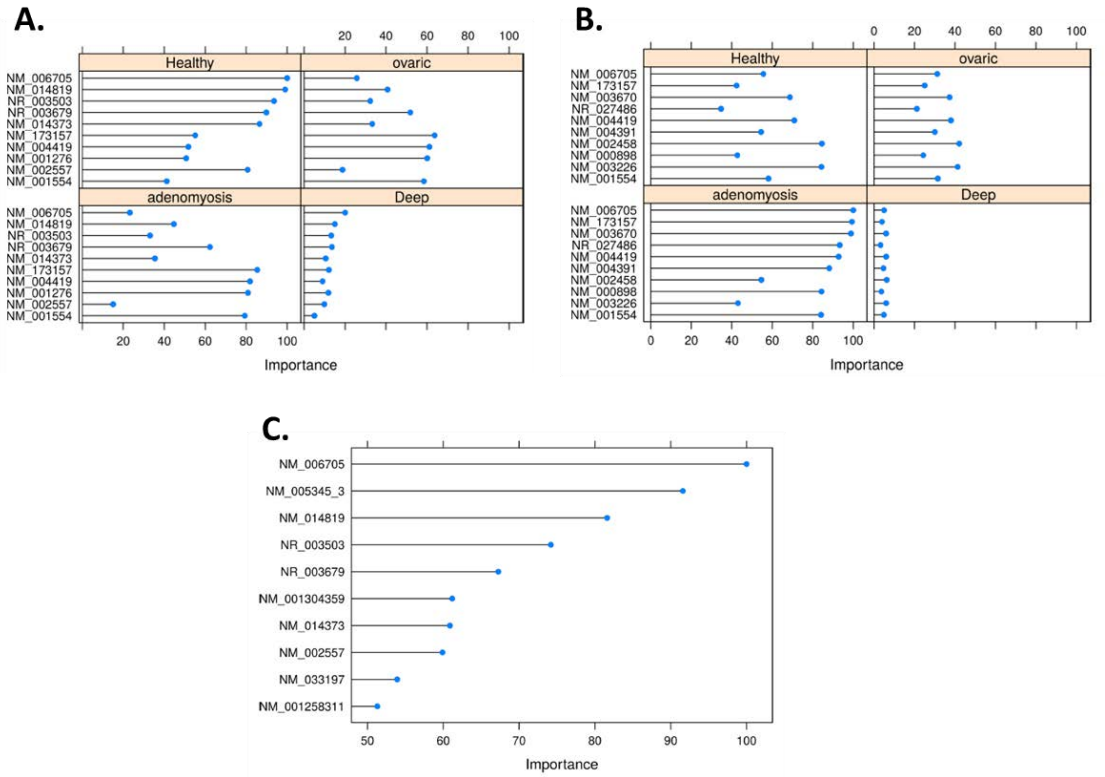


Figure 30. Importance for regression in multiclass PLS model with RNA normalization. A) With the normalized data. **B)** With the log-transformed data. **C)** Binary model with the log-transformed data. The most important regressors for each group are shown on the Y-axis and the percentage of importance on the X-axis.

RNA sequences NM_006705, NM_005345_3, and NM_014819 had more than 80% importance in the classification. This indicates that the RNA normalization does not affect this model, as the previously observed candidate biomarkers without RNA normalization are the same. Therefore, these sequences are candidates to be studied in a more exhaustive way to determine their relationship with endometriosis.

In conclusion, both multi-class models were less useful than the binary classifier model independent of RNA concentration normalization. The binary model was extensively studied. ROC curves, sensitivity, specificity, and a confusion matrix were performed.

nComp	ROC	Sens	Spec
1	0.9	0.67	0.84
2	0.93	0.78	0.89
3	0.94	0.78	0.95
4	0.94	0.78	0.95
5	0.89	0.83	0.84
6	0.85	0.78	0.89
7	0.77	0.67	0.79
8	0.77	0.67	0.89
9	0.76	0.67	0.84
10	0.75	0.67	0.84

Table 15. Results of the binary model with the normalized data. The table shows the ROC results of the model as well as the sensitivity (Sens) and specificity (Spec) depending on the number of principal components used.

	Endometriosis	Control
Endometriosis	14	1
Control	4	18

Table 16. Confusion matrix of the binary model. The table shows the endometriosis and control patients classified as endometriosis or control. Each row of the matrix represents the instances in a predicted class and each column represents the instances in a reference class.

In the binary model, the best model was obtained using either three or with four principal components with a ROC of 0.94, sensitivity of 0.78, and specificity of 0.95. From the confusion matrix (as can be observed in the columns) it can be concluded that of the endometriosis group, 14 samples were predicted correctly (as the endometriosis group) and four were predicted to be in the control group. For the control group, predictions were more accurate, with only one sample predicted as being from the endometriosis group. The ROC curve can be observed in **Figure 31**.

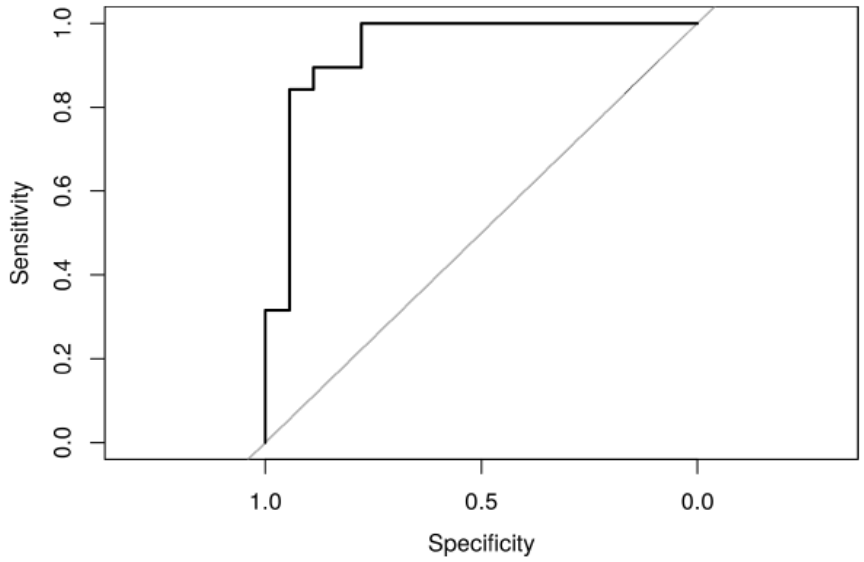


Figure 31. ROC curve of the binary model. The sensitivity is represented on the Y-axis and the specificity on the X-axis. The obtained AUC of the model was 0.9415.

In summary, the results shown above indicated that the model was able to classify whether a patients has endometriosis or not with an accuracy of 86% and an AUC of 0.94.
From here, two PLS models were performed with the objective of discovering if phenotypic variables would increase the sensitivity and specificity of both multi-class and binary models. Both models were performed with the log-transformed data and with RNA normalization (Steps 5 and 6 of the classifier development).

5. Multi-class model with phenotypic variables (to differentiate between types of endometriosis)

The main objective of this section was to find the most important RNA sequences but in this case phenotypic variables were also included to differentiate patients with endometriosis and the type of endometriosis. Hence, PCA and PLS analysis were done.

After the PCA analysis, where the data was obtained from the 182 DET, 74 transcripts of expression were selected. In the following figure, the performed PCA can be observed.

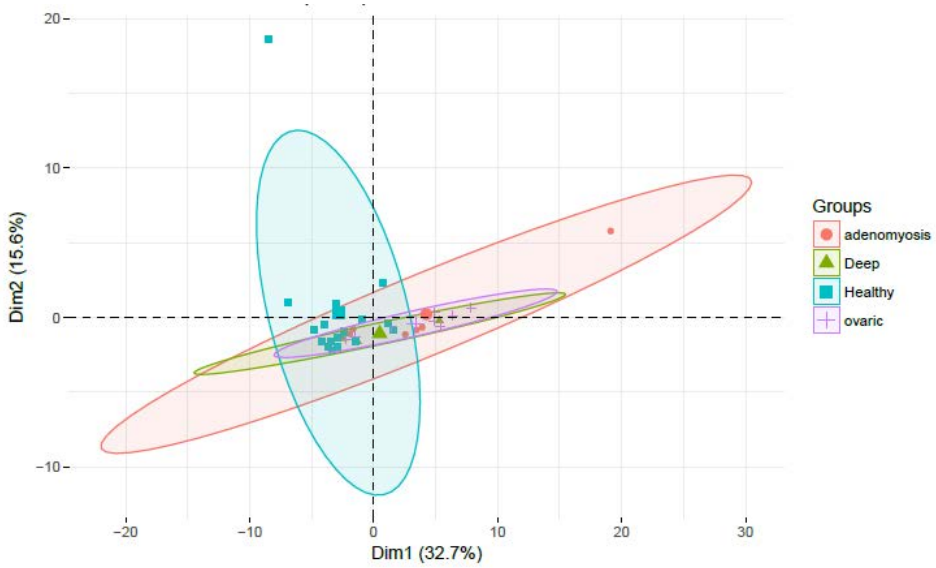


Figure 32. PCA of multi-class analysis. The control group and the three types of endometriosis are represented in the figure. Pink: adenomyosis; green: DIE "Deep"; purple: ovarian; blue: control.

After that, a PLS model using log-transformed data and a confusion matrix were completed **(Table 17 and Table 18)**.

nComp	Accuracy	Kappa
1	0.514	0.153
2	0.676	0.44
3	0.73	0.547
4	0.73	0.552
5	0.703	0.522
6	0.622	0.388
7	0.649	0.447
8	0.676	0.479
9	0.649	0.446
10	0.649	0.439

Table 17. Results of the PLS model. The results are shown depending on the principal components used to perform the model.

As we can see in the previous table, the best results were obtained using four principal components.

	Adenomyosis	DIE	Control	Ovaric
Adenomyosis	3	0	0	1
DIE	0	0	0	0
Control	1	2	19	2
Ovarian	2	2	0	5

Table 18. Confusion matrix of the multi-class PLS model. Each row of the matrix represents the different classes of the study while each column represents the same classes (adenomyosis = 6; DIE = 4; control = 19; ovarian = 8).

Only the control group was perfectly classified. The endometriosis groups were not. Half of the adenomyosis patients were classified well, as were five of the eight ovarian patients. Non of the DIE samples werw classified as such.

Accuracy-kappa were calculated and the obtained values were 0.73 and 0.547 respectively. In addition, sensitivity and specificity **(Table 19)** were also calculated for the multi-class model.

	Sensitivity	Specificity
Adenomyosis	0.5	0.968
DIE	0	1.000
Control	1.000	0.722
Ovarian	0.625	0.862

Table 19. Sensitivity and specificity of the multi-class PLS model. The rows represent each class of the model and the columns represent the sensitivity and specificity of each one.

Despite the obtained results, as can be observed in the confusion matrix, there were groups (DIE and adenomyosis) where the number of patients was too small for the algorithm to be able to correctly classify them. For this reason, the results obtained from the model could not be validated.

Finally, the most important regressors for the model were studied **(Figure 33)**.

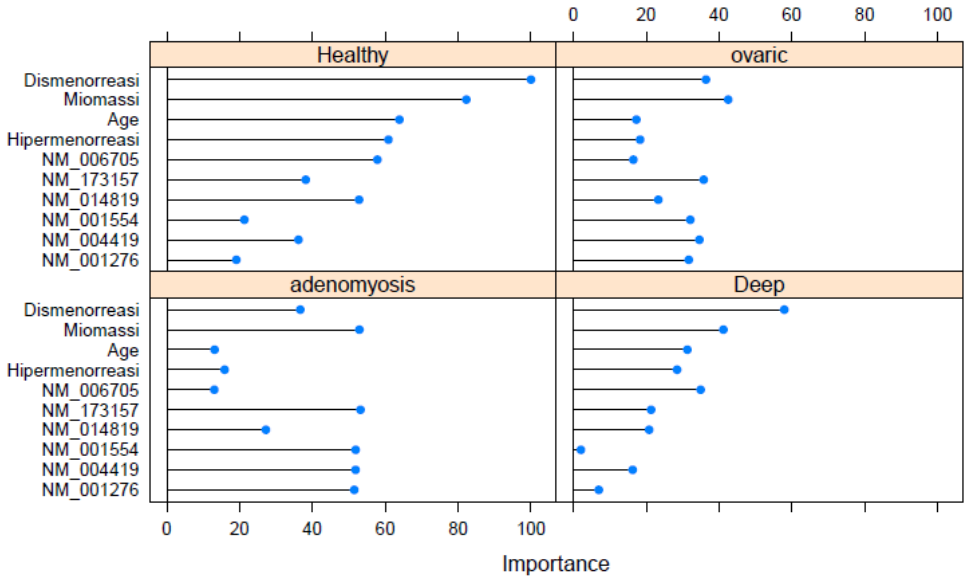


Figure 33. Top10 important regressors for the multi-class PLS model. The top 10 most important regressors for the classification of each studied class are shown in the figure. The phenotypic variables are included in the model. Y-axis: regressors. X-axis: importance of the variables in the model.

The phenotypic variables such as dysmenorrhea, myomas, and hypermenorrhea were very important regressors for the classificatory model. The biomarkers NM_006705, NM_173157, NM_014819, and NM_001554 were also important regressors for the model. Therefore, these RNA sequences (biomarkers) are candidates to be studied in a more exhaustive way to determine their true relationship with endometriosis. With the available data, it could not be confirmed that a classification of endometriosis and controls could be made using this model. Nevertheless, the obtained biomarkers could lead to future research.

6. Binary model with phenotypic variables (includes the total of endometriosis versus control)

The main objective of this section was to find the most important RNA sequences and also the phenotypic variables for differentiating patients with endometriosis. Again, PCA **(Figure**

34) and PLS **(Table 20 and Table 21)** analyses were done as well as the calculation of the most important regressors **(Figure 36)**.

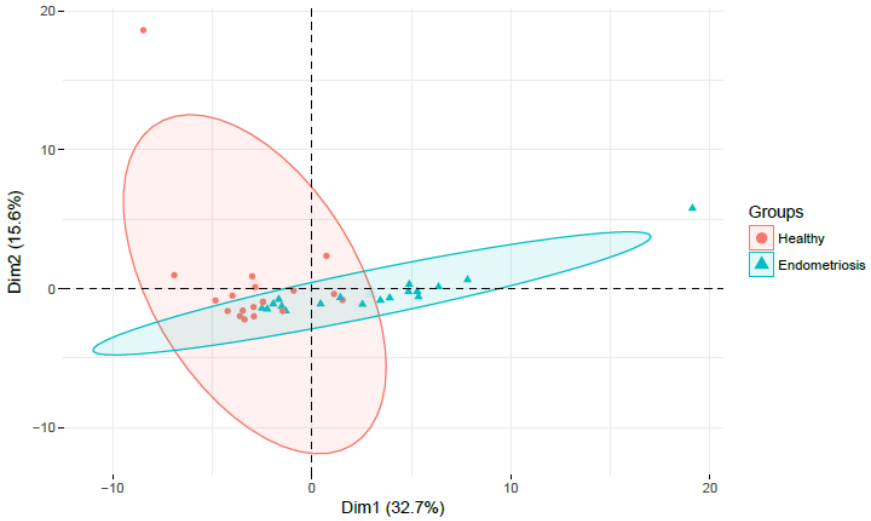


Figure 34. PCA of binary analysis. Control and endometriosis groups are represented in the figure. Pink: control; blue: endometriosis.

nComp	ROC	Sens	Spec
1	0.927	0.947	0.667
2	0.974	1.000	0.889
3	0.977	1.000	0.944
4	0.947	1.000	0.889
5	0.939	1.000	0.833
6	0.942	1.000	0.833
7	0.939	1.000	0.833
8	0.927	1.000	0.833
9	0.927	1.000	0.833
10	0.915	0.947	0.778

Table 20. Results of binary PLS model. The table shows the ROC, sensitivity (Sens), and specificity (Spec) depending on the number of principal components used.

As can be observed in the previous table, the model obtained the best results by using three principal components with a ROC of 0.977, a specificity of 1, and a sensitivity of 0.944.

	Control	Endometriosis
Control	19	1
Endometriosis	0	17

Table 21. Confusion matrix of the binary PLS model. In the rows the two groups of study (control and endometriosis) are represented. The same groups are represented in the columns.

The binary model was able to classify all the control patients in the control group and all but one of the endometriosis patients in the endometriosis group. The accuracy-kappa and the sensitivity-specificity **(Table 22)** were also calculated using three principal components. The values were very close to 1. Therefore, the classifications made by the algorithm were very close to reality. In addition, the specificity was very close to 1 due to the endometriosis patient classified as a control (confusion matrix). The algorithm was able to classify controls correctly, and for this reason the sensitivity was a value of 1.

Accuracy	0.973
Kappa	0.946
Sensitivity	1.000
Specificity	0.944

Table 22. Accuracy-kappa and sensitivity-specificity of the binary PLS model.

Finally, the following figure **(Figure 35)** shows the ROC curve.

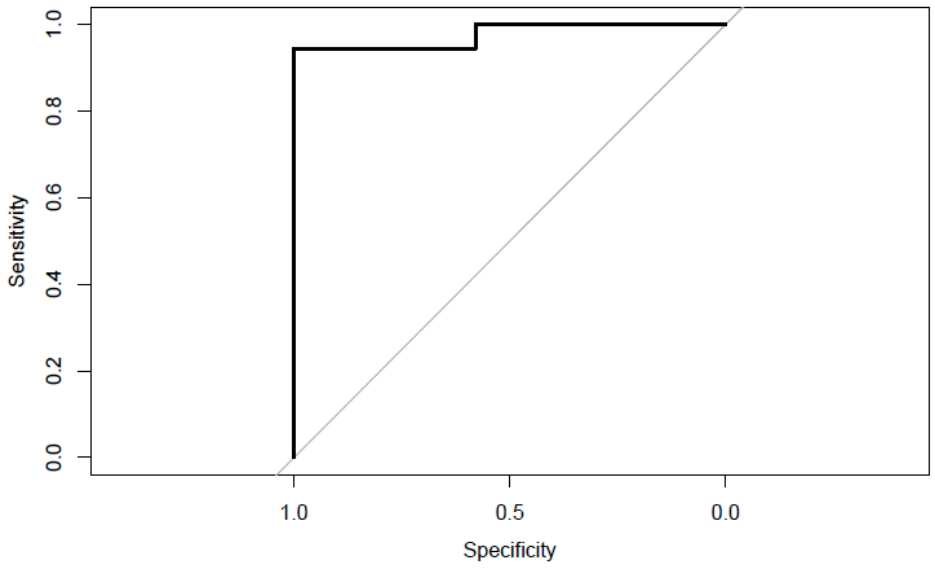


Figure 35. ROC curve of the binary PLS model. The sensitivity is shown on the Y-axis and the specificity on the X-axis. AUC = 0.9766.

As can be seen in **Figure 35**, the AUC had a value of 0.98, which means that the model was capable of differentiating control and endometriosis patients almost perfectly. Finally, the top 10 important variables affecting the model (regressors) were observed **(Figure 36)**.

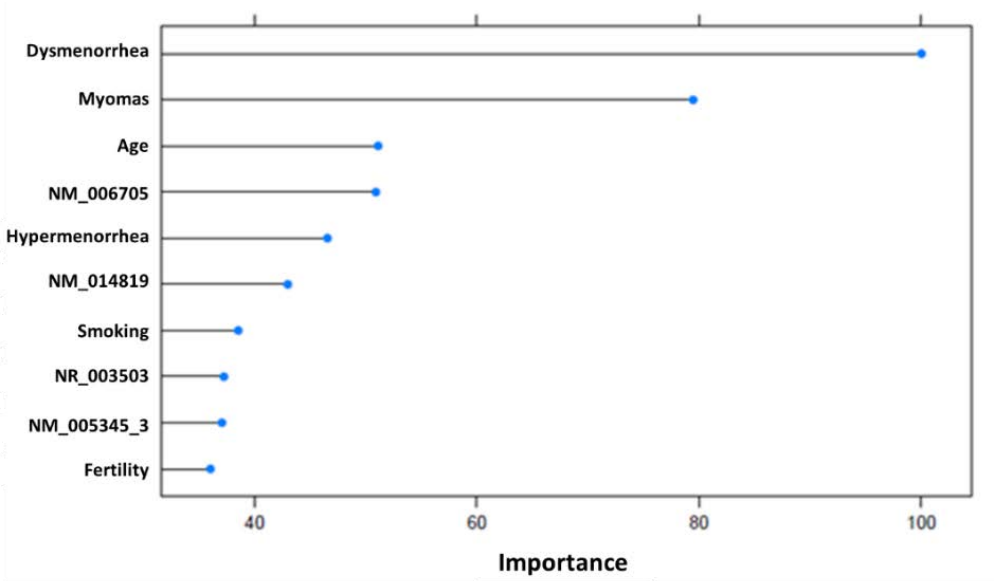


Figure 36. Top10 important regressors for the binary-class PLS model. The top 10 most important regressors for the classification of each studied class are shown in the figure. The phenotypic variables are included in the model. Y-axis: regressors. X-axis: importance of the variables in the model.

Phenotypic variables such as dysmenorrhea, myomas, hypermenorrhea, age, smoking, and fertility are very important regressors for the classificatory model. The biomarkers NM_006705, NM_014819, NR_003503, and NM_005345_3 are also important regressors for the model.

Therefore, these RNA sequences were candidate biomarkers that should be studied and validated in a large set of samples to determine their true relationship with the disease. The genes corresponding to the four candidate transcripts found correspond to Growth Arrest and DNA Damage Inducible Gamma (GADD45G; NM_006705), praja ring finger ubiquitin ligase 2 (PJA2; NM_014819), gamma-glutamyltransferase 8 pseudogene (GGT8P; NR_003503), and Heat Shock Protein Family A Member 1A (HSPA1A; NM_005345_3). GADD45G was up-regulated in endometriosis (FC=1.76; qValue=0.0246), while the remaining three genes were downregulated: PJA (FC=-1; qValue=0.0098), GGT8P (FC=-3.75; qValue=0.0098) and HSPA1A (FC=-1.06; qValue=0.0306).

In conclusion, it was confirmed that the binary model could classify between endometriosis and the control group with the available data. However, the multi-class model was not able to distinguish between types of endometriosis. Therefore, a validation of the biomarkers of the binary classificatory model containing the phenotypic variables was developed.

4.1.4. FFPE samples are feasible for use in Nanostring nCounter

The aim of the validation study was to validate the putative candidates found in the discovery study and determine, with a large sample size, if they would be able to discriminate between endometriosis and control patients. As explained in the materials and methods section, the validation was performed using nCounter® (Nanostring® Technologies) (*see Section 3.11*).

A CodeSet of 20 transcripts was designed containing probes for the four candidate biomarkers that were obtained in the classifier (GADD45G, GGT8P, HSPA1A, and PJA2), five houskeeping genes to normalize the data (ACTB, MRPL19, PSMC4, RPLP0, and SF3A1) and 11 genes that were described in the literature to be de-regulated in eutopic endometrium of women with endometriosis (ADA2, DUSP1, FOS, LINC00882, LOX, MMP11, MMP7, MUC5B, SERPINE1, SLPI, and TFF3)[79,198]. The CodeSet design is shown in **Table 23**.

Gene ID	Accession	Position	Target Sequence
ACTB	NM_001101.2	1011-1110	TGCAGAAGGAGATCACTGCCCTGGCACCCAGCACAATGAAGATCAAGATCATTGCTCCTCCTGAGCGCAAGTACTCCGTGTGGATCGGCGGCTCCATCCT
ADA2	NM_001282229.1	3336-3435	GTCTTGCCCACTCCTGTTTACCCTTCAAGTTCAAGTTCATGTCACTGTCTCAGAGAGGTTTTCCTGTGCTCGCCCTGTTTCTCTCAGGAAGCCTTGCTC
DUSP1	NM_004417.2	988-1087	TCAAGAATGCTGGAGGAAGGGTGTTTGTCCACTGCCAGGCATTTCCCGGTCAGCCACCATCTGCCTTGCTTACCTTATGAGGACTAATCGAGTCAA
FOS	NM_005252.2	1476-1575	ACTCAAGTCCTTACCTCTTCCGGAGATGTAGCAAAACGCATGGAGTGTGTATTGTTCCCAGTGACACTTCAGAGAGCTGGTAGTTAGTAGCATGTTGAGC
GADD45G	NM_006705.3	28-127	CTGAGCTCTGGCTGTCAGTGTGTTCGCCCGCGTCCCCTCCGCGCTCTCGGCTTGTGGATAACTAGCTGCTGGTTGATCGCACTATGACTCTGGAAGAAGT
GGT8P	NR_003503.1	462-561	ACGTGTTCTGGGAAGCGGGAAGGAGACACAGGCCTTGTGTTTCTGAGGCCCAACTTTAGACTGTGCCCTGCTGGGGAGGTGCCAGGGAATGTCTGAGGCT
HSPA1A	NM_005345.5	99-198	GCTTCCCAGAGCGAACCTGTGCGGCTGCAGGCACCGGCGCGTCGAGTTTCCGGCGTCCGGAAGGACCGAGCTCTTCTCGCGGATCCAGTGTTCCGTTTCC
LINC00882	NR_028303.1	127-226	TCGTCCTTAAATATTGCCGATACTTGACCTACGCAAGAGACAATGTCATGTGATTCAGCCTAATATCTCAGAGGATGCAGCATTCAAGGTTCTATCTTGG
LOX	NM_002317.4	1576-1675	CGCTACACAGGACATCATGCGTATGCCTCAGGCTGCACAATTTCACCGTATTAGAAGGCAAAGCAAAACTCCCAATGGATAAATCAGTGCCTGGTGTTCT
MMP11	NM_005940.3	703-802	AGCAGCCAAGGCCCTGATGTCCGCCTTCTACACCTTTCGCTACCCACTGAGTCTCAGCCCAGATGACTGCAGGGGCGTTCAACACCTATATGGCCAGCCC
MMP7	NM_002423.3	312-411	GTGCCAGATGTTGCAGAATACTCACTATTTCCAAATAGCCCAAAATGGACTTCCAAAGTGGTCACCTACAGGATCGTATCATATACTCGAGACTTACCGC
MRPL19	NM_014763.3	365-464	GGAAGTATTCTTCGTGTTACTACAGCTGACCCATATGCCAGTGGAAAAATCAGCCAGTTTCTGGGGATTTGCATTCAGAGATCAGGAAGAGGACTTGGAG
MUC5B	NM_002458.1	16312-16411	GGGACCCGATGGGTTTCCTAAATTTCCCGGGGAGCGGTGGGTCAGCAACTGCCAGTCCTGCGTGTGTGACGAGGGTTCAGTGTCGGTGCAGTGCAAGCCC
PJA2	NM_014819.4	596-695	GAGGAAGGCAGGGATACCTTAGGAAGCAGTACAAATCTTCATAATCACTCTGAGGGAGAGTATATTCCAGGAGCTTGTAGTGCTTCAAGTGTCCAAAATG
PSMC4	NM_006503.2	301-400	CATCGGACAATTTCTGGAGGCTGTGGATCAGAATACAGCCATCGTGGCTCTACCACAGGCTCCAACTATATGTGCGCATCCTGAGCACCATCGATCGG
RPLP0	NM_001002.3	251-350	CGAAATGTTTCATTGTGGGAGCAGACAATGTGGGCTCCAAGCAGATGCAGCAGATCCGCATGTCCCTTCGCGGGAAGGCTGTGGTGCTGATGGGCAAGAA
SERPINE1	NM_000602.4	578-677	CGGAGCACGGTCAAGCAAGTGGACTTTTCAGAGGTGGAGAGACCAGATTCATCATCAATGACTGGGTGAAGACACACAAAAGGTATGATCAGCAACT
SF3A1	NM_001005409.1	236-335	CTTCTAAGCCAGTTGTGGGATTATTTACCCTCCTCCAGAGGTCAGAAATATTGTTGACAAGACTGCCAGCTTTGTGGCCAGAAACGGGCCTGAATTTGA
SLPI	NM_003064.2	331-430	TTTCTGTGAGATGGATGGCCAGTGCAAGCGTGACTTGAAGTGTTGCATGGGCATGTGTGGGAAATCCTGCGTTTCCCCTGTGAAAGCTTGATTCCTGCCA
TFF3	NM_003226.3	586-685	CTGCTGAAAGTTCATATCTGGAGCCTGATGTCTTAACGAATAAAGGTCCCATGCTCCACCCGAGGACAGTTCTTCGTGCCTGAGACTTTCTGAGGTTGTG

Table 23. CodeSet design. The table shows the 20 designed probes and their target sequences for the validation study. The 4 candidate biomarkers are highlighted in grey and the housekeeping in orange. The resting genes were included from literature.

For the nCounter® analysis, three steps were performed, and the results are given below.

1. Quality control prior analysis
2. Sample processing
3. Post-processing quality control

4.1.4.1. Quality control prior analysis

After the FFPE RNA extraction of the 285 samples, the concentration was quantified using Nanodrop®. The obtained RNA concentration was between 8.15ng/µl and 1,876ng/µl. The recommended amount of RNA to use Nanostring® Technology is 300ng. Therefore, the microliters needed to perform the experiment were calculated. Fifty-one samples had less than 300ng, the least containing 40ng. Although these samples had less than the recommended amount, they were also analyzed by the nCounter. In addition, all samples had a ratio of 260/280 and 260/230 close to 2, which indicated that the samples were well purified.

4.1.4.2. Sample processing

Aside from the 51 samples where the total amount of RNA was lower than 300ng, all the samples were processed in the nCounter as instructed. In the case of the 51 samples, a total of 10µl was added to the solution.

4.1.4.3. Post-processing quality control

After the nCounter® processing, the raw data was analyzed using nSolver™ software analysis. All the samples, including the ones that had a smaller amount of RNA, passed the software quality control. Therefore, the number of copies obtained per candidate were normalized with the five housekeeping gene expressions.

4.1.5. FFPE samples gene expression correlate between RNA-Seq and Nanostring

After quality control, the data was normalized to the housekeeping transcripts expression and transformed to log scale before it was used for validation of the model.

The transcripts from RNA-Seq were plotted against the new data provided by nSolver® in order to obtain the correlation between the techniques. This was done for each of the 37 patients used in the discovery phase. The obtained correlation was very high, except for two of the tested transcripts, and few examples of the plots created per transcript is shown in **Figure 37**.

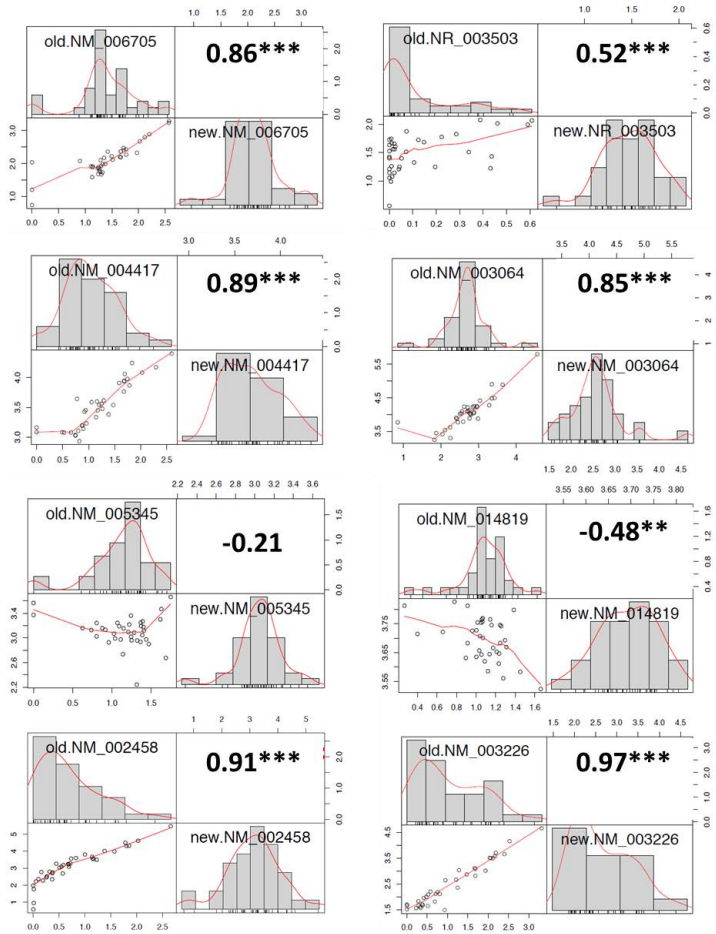


Figure 37. Correlation between techniques. RNA-Seq (X-axis) and Nanostring (Y-axis) expression of the transcripts NM_006705, NR_003503, NM_005345, NM_014819, NM_003226, NM_002458, NM_004417 and NM_003064. On top, the absolute value of the correlation plus the result of the correlation test as stars. On bottom, the bivariate scatterplots, with a fitted line.

4.1.6. Binary-PLS model classifies 60% of patients with endometriosis

The following table shows the comparison of the three PLS models performed: combinatory PLS (with genetic and phenotypic variables), PLS with only genetic variables, and PLS with only phenotypic variables.

	Combinatory model	Genetic variables	Phenotypic variables
Accuracy	0.72	0.5	0.75
Kappa	0.47	0.003	0.52
Sensitivity	0.95	0.49	0.96
Specificity	0.6	0.51	0.64
ROC	0.931	0.49	0.95

Table 24. Statistics on the three PLS models.

The combinatory model has a very high sensitivity (0.95), but the specificity is not very high. When using only genetic variables, the model is much less useful. When using only phenotypic variables, both the sensitivity and specificity increase, indicating that the variables that give importance to the model are the phenotypic variables.

In addition, when looking the confusion matrix of the three models **(Table 25)**, it can be observed that the combinatory model allows the classification of 60% of the patients with endometriosis. When using only the genetic variables, it decreases to 51%. When using the phenotypic variables, it increases to 64%.

	Combinatory Model		Genetic Variables		Phenotypic Variables	
	Control	Endometriosis	Control	Endometriosis	Control	Endometriosis
Control	73	59	38	73	74	53
Endometriosis	4	90	39	76	3	96
Prediction	95%	60%	49%	51%	96%	64%

Table 25. Confusion matrix for the three PLS models: combinatory, genetic, and phenotypic. The table shows how patients are classified following the models. Control=77; endometriosis=149.

4.2. Evaluation of LGR5 as a biomarker for endometriosis

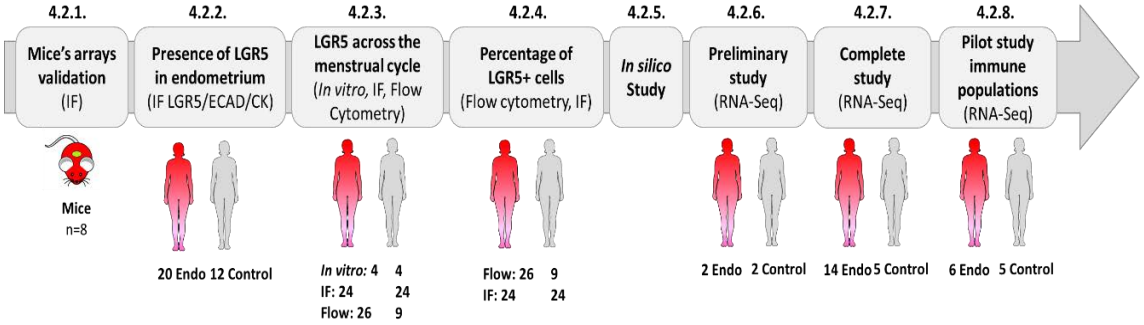


Figure 38. Workflow followed to perform Objective 2. Each study is enumerated in the section where it is explained.

4.2.1. Arrays of the endometriosis mice model were validated

The aim of this section was to validate the arrays performed in the GFP+ cells from the endometriosis mice model, where an overexpression of LGR5 in these cells was observed. Immunofluorescence in the eutopic endometrium of eight mice was performed in order to prove that GFP+ cells located in the stromal compartment overexpressed LGR5. As can be observed in **Figure 39.A**, GFP+ cells co-localized with LGR5 markers in the stromal compartment of the eutopic endometrium of the mice. It was also observed that GFP+ cells co-localized with LGR5 and the epithelial marker CK. Staining of the samples was also performed with the epithelial marker ECAD. As can be observed in **Figure 39.B**, LGR5+ cells located in the stromal compartment co-localized with this epithelial marker as well.

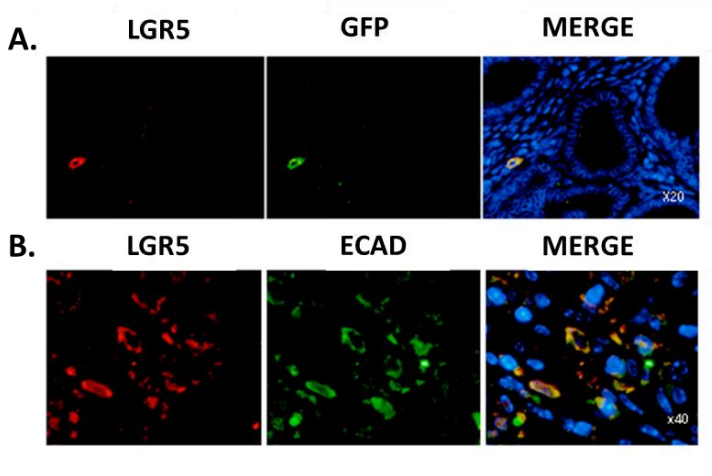


Figure 39. LGR5 staining in eutopic endometrium of the endometriosis mice model. A) Co-localization of LGR5 (red) and GFP (green) in the stromal compartment of the eutopic endometrium of the endometriosis mice model. In the right panel the merged images can be observed, which also show the DAPI probe (blue) at x20 magnification. **B)** Co-localization of LGR5 (red) and ECAD (green) and the merged showing the DAPI probe (blue) at 40x magnification.

4.2.2. LGR5 co-localized with epithelial markers in eutopic endometrium of women with endometriosis

Once it was confirmed that LGR5 co-localized with the migrated cells (GFP+ cells) in the mice model, these results were translated into human eutopic endometrium. A total of 20 eutopic endometrium samples from women with endometriosis and 12 control endometrium samples were stained with LGR5 and the two epithelial markers, CK and ECAD **(Table 26)**.

Type	Co-localization (LGR5/ECAD)	Co-localization (LGR5/CK)
Pelvic (n=3)	2	0
Ovarian (n=9)	7	8
DIE (n=6)	5	3
Adenomyosis (n=2)	2	2
TOTAL endometriosis (n=20)	**16**	**13**
Healthy (n=12)	**0**	**0**

Table 26. **Co-localization of LGR5 with the epithelial markers ECAD and CK.** The table shows the number of patients where LGR5 co-localized with ECAD (left column) and CK (right column).

Interestingly, LGR5+ cells co-localized aberrantly with CK and ECAD in the stromal compartment from 13 and 16 of 20 endometriotic patients, respectively, whereas there was no co-localization in any of 12 donors. An example of this co-localization is shown in **Figure 40**.

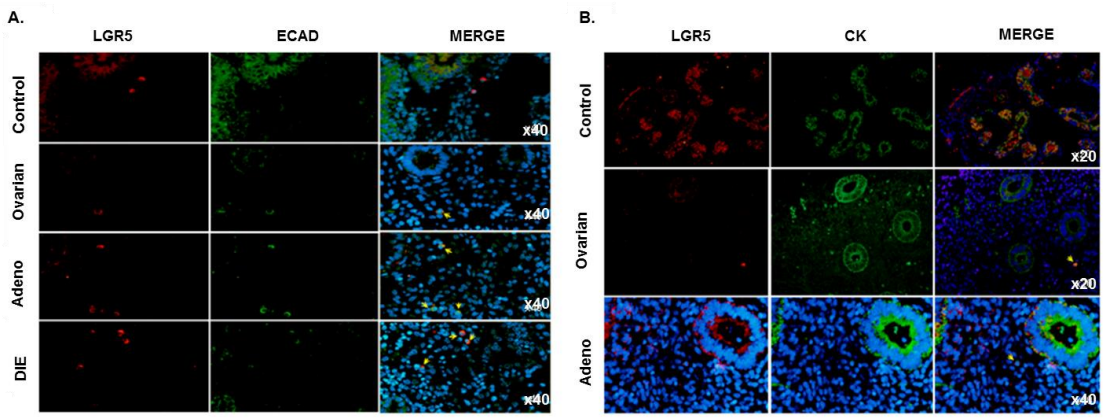


Figure 40. A) Co-localization of LGR5 with ECAD in human eutopic endometrium. B) Co-localization of LGR5 with CK in human eutopic endometrium. In the left panel of each figure, LGR5 is dyed in red. The epithelial markers are dyed in green in the central panel. The right panel shows the merging of both markers. In the first row of figures, the control group is shown, where there is no co-expression of the markers in the stroma. Below, there is a representation of the co-localization of both markers in each type of endometriosis (except pelvic endometriosis) and in the case of B, no co-localization in deep infiltrating endometriosis was found. Cells co-expressing both markers are shown with yellow arrows. Magnification is shown in the right of each row.

4.2.3. LGR5 does not vary throughout the menstrual cycle

The aim of this section was to determine whether LGR5 varies throughout the menstrual cycle. To elucidate that, three approaches were used and are explained below.

4.2.3.1. In vitro studies

In order to elucidate whether LGR5 varies across the menstrual cycle, *in vitro* experiments were performed. Endometrial stromal fibroblasts were treated with estradiol and progesterone over 12 days in order to mimic the proliferative and secretory phases of the menstrual cycle. A total of four controls and three endometriosis (DIE) samples were cultured and treated. No differences were observed in terms of cell morphology **(Figure 41)**.

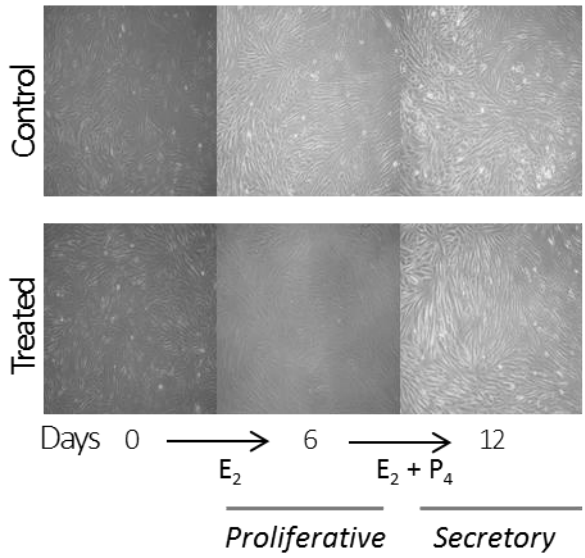


Figure 41. Mimics of the menstrual cycle phases in vitro. In the upper panel is shown the control cells and in the lower panel, the treated cells. On the X-axis, the days of treatment and phases of the menstrual cycle are shown. E_2: estradiol; P_4: progesterone. The figure corresponds to a control group (no endometriosis). N = 4; experiments done per duplicate.

After RNA extraction and RT-qPCR analysis, expression of CYR61 (proliferative phase) and DKK1 (secretory phase) were observed. As can be observed in **Figure 42**, although not significant, CYR61 and DKK1 increased in the proliferative and secretory phases

respectively, confirming that the mimics of the cycle had worked. In addition, a significant variation of LGR5 throughout the menstrual cycle was not observed **(Figure 43)** in control patients or in endometriosis patients, although a non-significant decrease in the marker was observed in the secretory phase in the control cases. Interestingly, the contrary was observed in endometriosis.

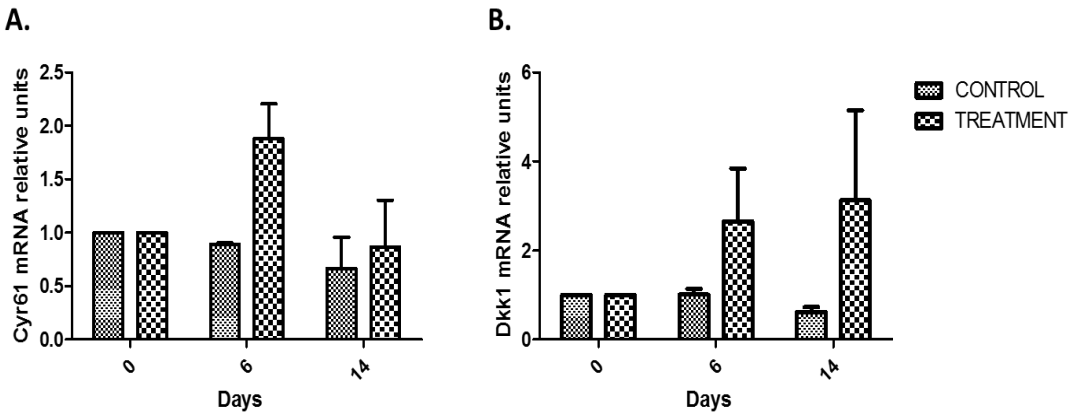


Figure 42. mRNA expression of CYR61 and DKK1 throughout the menstrual cycle in control patients *in vitro.* A) Cyr61 expression. B) Dkk1 expression. On the Y-axis, relative mRNA units of each gene (2^(-ddCt)) are represented. The days of treatment are on the X-axis.

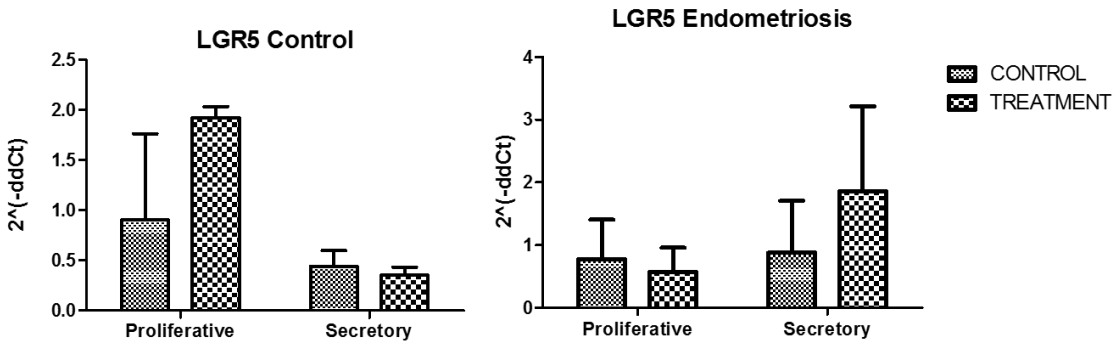


Figure 43. mRNA expression of LGR5 throughout the menstrual cycle. On the Y-axis, relative mRNA units (2^(-ddCt)) are represented. On the X-axis are the phases of the menstrual cycle. **A) Control group** (n=4). **B) Endometriosis group** (n=3).

4.2.3.2. Flow cytometry analysis

Taking advantage of the fact that LGR5+ cells were going to be sorted for the gene expression analysis of LGR5, the concentration of these cells across the menstrual cycle was determined. Around 2% to 7% of LGR5+ cells were obtained from each sample. An example of the LGR5 gating is shown in **Figure 44**.

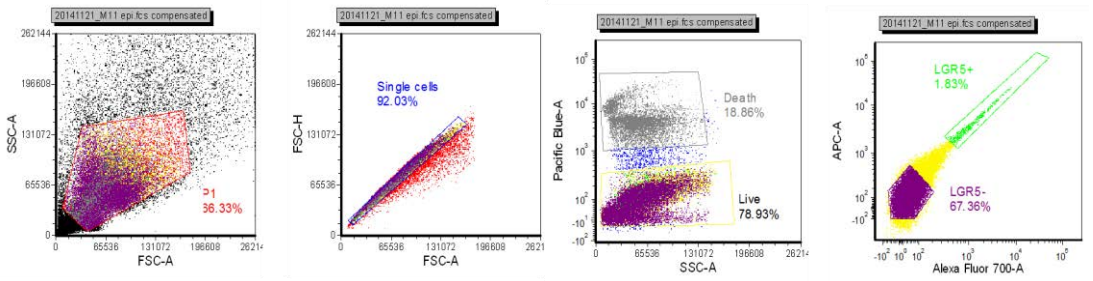


Figure 44. LGR5 gating strategy. First, cells are separated by size (SSC-A/FSC-A). In the second, single cells are obtained and cell debris is excluded (FSC-H/FSC-A). Then, live and dead cells are separated (Pacific blue-A/ SSC-A). Finally, LGR5+/- cells are gated (APC-A/Alexa Fluor 700A), obtaining 1.83% of positive cells, in this example.

First of all, a preliminary study with nine eutopic endometriosis samples was done to elucidate if LGR5 was varying across the menstrual cycle **(Figure 45.A)**. No variation in the percentage of LGR5 was observed. After that, analysis of the 35 sorted samples was performed, and again, no variation of LGR5 across the menstrual cycle was observed **(Figure 45.B)**.

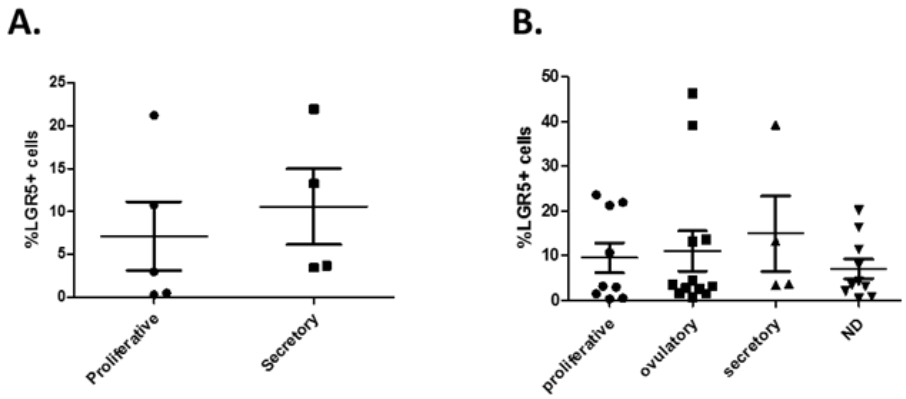


Figure 45. LGR5+ cells across the menstrual cycle. A) Percentage of LGR5+ cells in five proliferative and four secretory eutopic endometrium samples from women with endometriosis. **B)** Percentage of LGR5+ cells in the proliferative phase: 9%; ovulatory phase (egg donors): 12%; secretory phase: 4%, ND (not determined): 10%.

4.2.3.3. Immunofluorescence analysis

Immunofluorescence of 48 eutopic endometrium samples (24 from women with and 24 from women without endometriosis) from different phases of the menstrual cycle was performed and the percentage of the fluorescence intensity mean was calculated for LGR5. The results are shown in **Figure 46**.

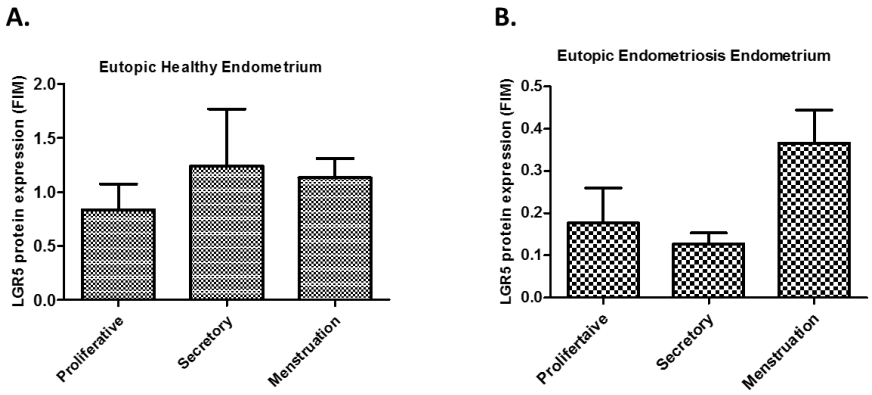


Figure 46. LGR5 protein expression throughout the menstrual cycle. The fluorescence intensity mean (FIM) was calculated for each phase of the menstrual cycle (proliferative, secretory, and menstruation) in control eutopic endometrium **(A)** and eutopic endometrium from women with endometriosis **(B)**.

4.2.4. There are no differences in LGR5 percentage in the epithelium and stroma and there are differences in eutopic endometrium of women with and without endometriosis depending on the technique used

4.2.4.1. Percentage of LGR5+ cells in eutopic and ectopic endometriosis endometrium

Flow Cytometry: Results of the seven comparisons made from flow cytometry analysis of the 35 LGR5+ sorted samples are shown below.

Comparisons
1. Epithelium and Stroma between all groups and Epithelium vs Stoma in each group
2. Total eutopic endometrium (Epithelium + Stroma) between all groups
3. Total eutopic endometrium control vs Total endometriosis
4. Epithelium ectopic vs Epithelium eutopic (patients with endometriosis)
5. Stroma ectopic vs stroma eutopic (patients with endometriosis)
6. Total ectopic vs Total eutopic (4 ovaric and 6 DIE)
7. Ectopic between types of endometriosis (4 ovaric and 6 DIE)

Table 3. Comparisons of LGR5+ cells obtained by flow cytometry.

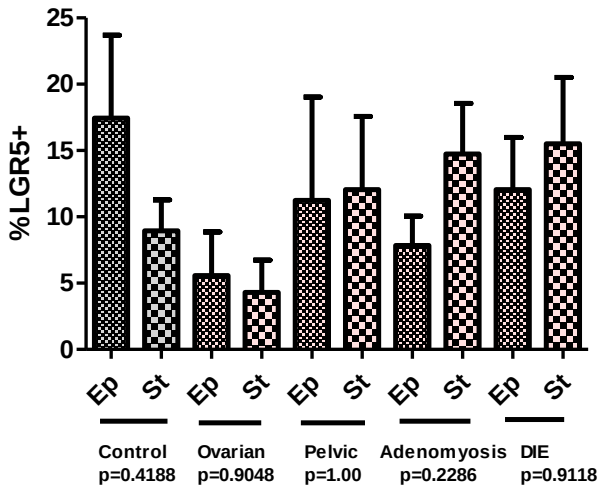


Figure 47. LGR5+ cells in eutopic epithelium and stroma between all groups and in each group. Ep: epithelium; St: stroma. Representation of the LGR5+ cells fraction in the epithelium and the stromal compartments in all groups of study. Grey: control; pink: endometriosis. P value of the t-test performed in each group (epithelium vs. stroma) is shown.

As can be observed in the previous figure, no differences between the epithelium and stromal compartments were observed either in the comparisons of each group (t-test) or in the comparison between groups (one-way ANOVA).

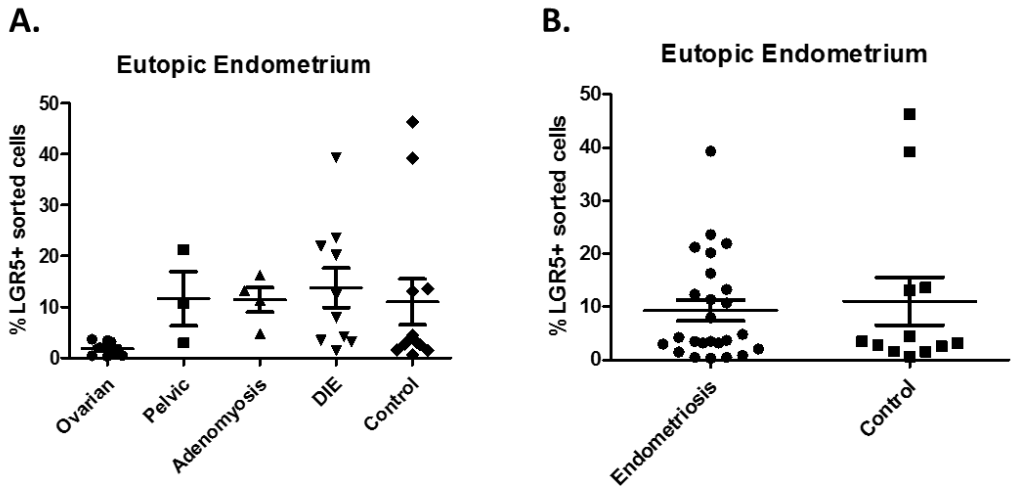


Figure 48. Percentage of LGR5+ cells in total eutopic endometrium measured by flow cytometry. **A)** Percentage of LGR5+ cells in the eutopic endometrium of different types of endometriosis and control samples. One-way ANOVA (P=0.2784). **B) LGR5+ cells in total eutopic endometrium.** The figure shows the percentage of LGR5+ cells in the eutopic endometrium of the total endometriosis versus the control. T-test (P=0.7579).

As the previous figure shows, no significant differences were found between LGR5+ cells in eutopic endometrium of women with endometriosis (differentiating by different types of the disease) and control women. When all of the eutopic endometrium samples were analyzed together (all types of endometriosis) and compared to the control group, no significant differences were found.

In addition, the differences between epithelial and stromal compartment from ectopic lesions against eutopic endometrium in patients with endometriosis were also analyzed. The results are shown in **Figure 49** below. These comparisons were only performed for ovarian endometriosis and DIE because not enough ectopic samples of pelvic endometriosis (n=1) and adenomyosis (n=2) were available for statistical analysis.

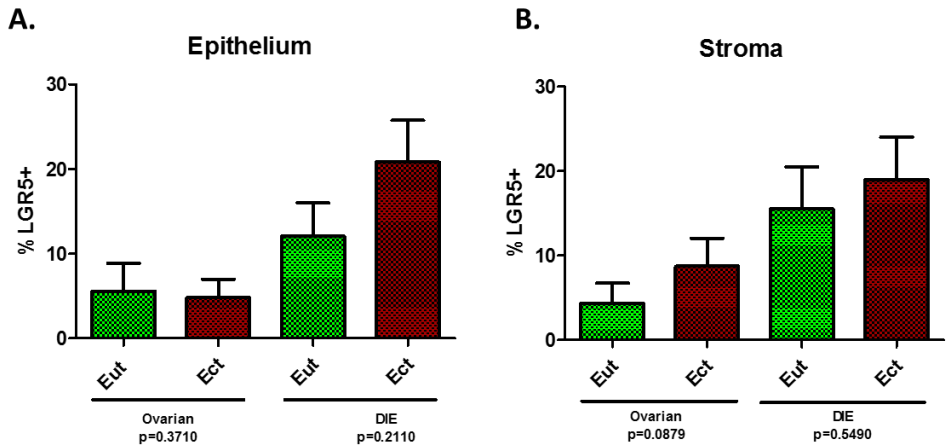


Figure 49. LGR5+ cells in ectopic and eutopic epithelium and stroma in ovarian and DIE endometriosis. The figure shows the percentage of LGR5+ cells in the ectopic (brown) and eutopic (green) endometrium of ovarian endometriosis and DIE. The figure also shows p-values from the t-test. **A)** Epithelial fraction of the endometrium. **B)** Stromal fraction of the endometrium.

As can be observed in **Figure 49**, no significant differences were found between ectopic and eutopic epithelium or in ectopic stroma versus eutopic stroma. However, the total tissue (epithelium and stroma) was analyzed in ovarian endometriosis and DIE and significant differences were found **(Figure 50)**.

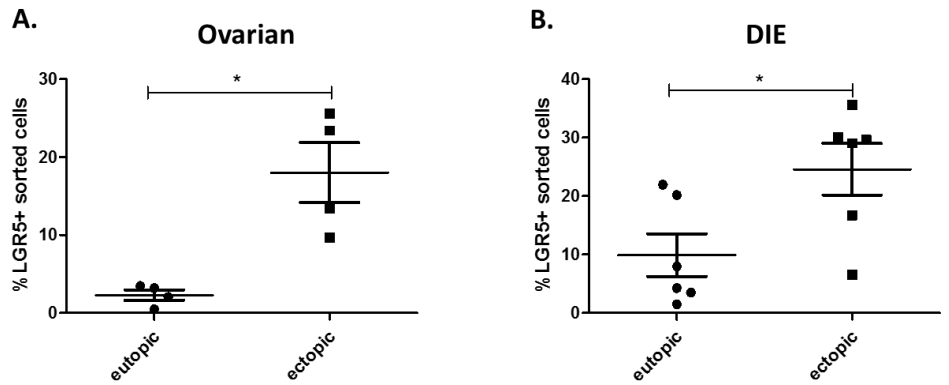


Figure 50. LGR5+ cells in total ectopic and eutopic endometrium in ovarian and DIE endometriosis. The figure shows the percentage of LGR5+ cells in the ectopic and eutopic endometrium of ovarian endometriosis and DIE. **A)** Ovarian endometriosis (p-value=0.0286). **B)** DIE (p-value=0.0411).

In this case, significant differences were observed between ectopic and eutopic endometrium of women with endometriosis. There was a higher LGR5+ cell percentage in ectopic lesions than in the eutopic endometrium.

Finally, in order to determine whether different types of endometriosis contained a higher percentage of LGR5+ cells in ectopic lesions, a t-test was performed between the different types of endometriosis **(Figure 51)**.

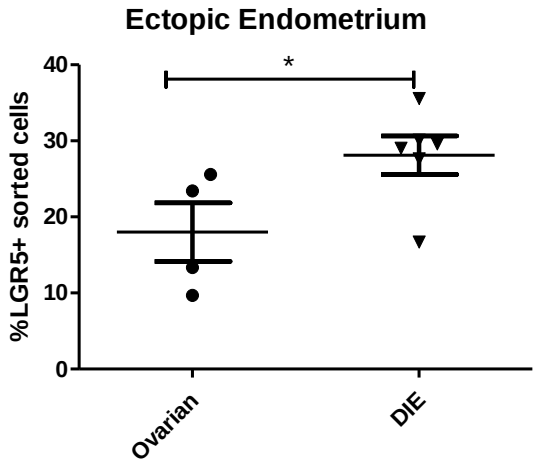


Figure 51. LGR5+ cells in total ectopic endometrium in different types of endometriosis. The figure shows the percentage of LGR5+ cells in the ectopic endometrium of four ovarian endometriosis samples and six DIE samples. T-test (p-value=0.0381).

As it can be observed in the previous figure, a p-value of 0.0381 was found between ovarian and DIE endometriosis, which indicate that more percentage of LGR5+ cells was found in DIE, the most aggressive type of endometriosis.

Immunofluorescence: Immunofluorescence of 48 eutopic endometrium samples (24 from women with and 24 from women without endometriosis) from different phases of the menstrual cycle was performed. The fluorescence intensity mean was calculated for LGR5 and the comparison of eutopic endometrium from control and endometriosis samples was carried out. The results are shown in **Figure 52**.

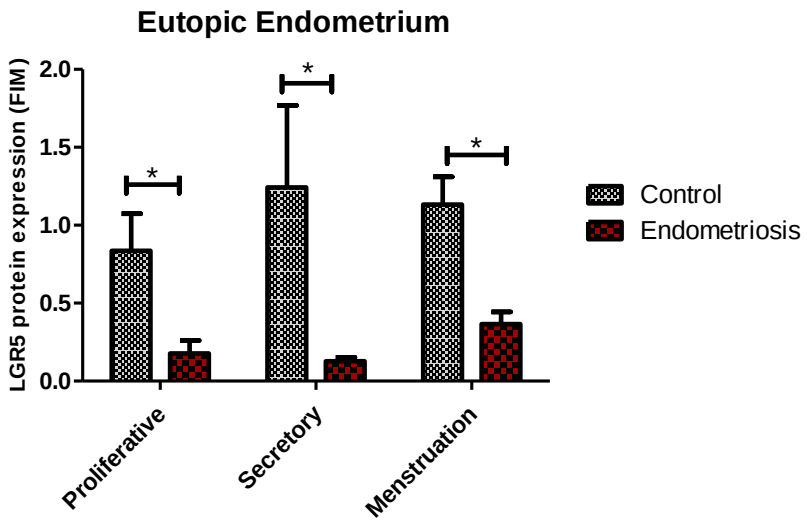


Figure 52. LGR5 protein expression throughout the menstrual cycle in control and endometriosis samples of eutopic endometrium. Fluorescence intensity mean (FIM) is shown on the Y-axis for each phase of the menstrual cycle (proliferative, secretory and menstruation) (t-test for each phase; proliferative: p=0.0242; secretory: p=0.0424; menstruation: p=0.0121).

As can be observed in **Figure 52**, LGR5 protein expression was significantly different between the controls and the endometriosis eutopic endometrium in all the phases of the menstrual cycle.

4.2.5. More than 6,000 genes vary across the menstrual cycle and under FSH stimulation

Pre-analysis of the arrays extracted from GEO (Accession GSE19959) were performed. The results are shown in **Figure 53**.

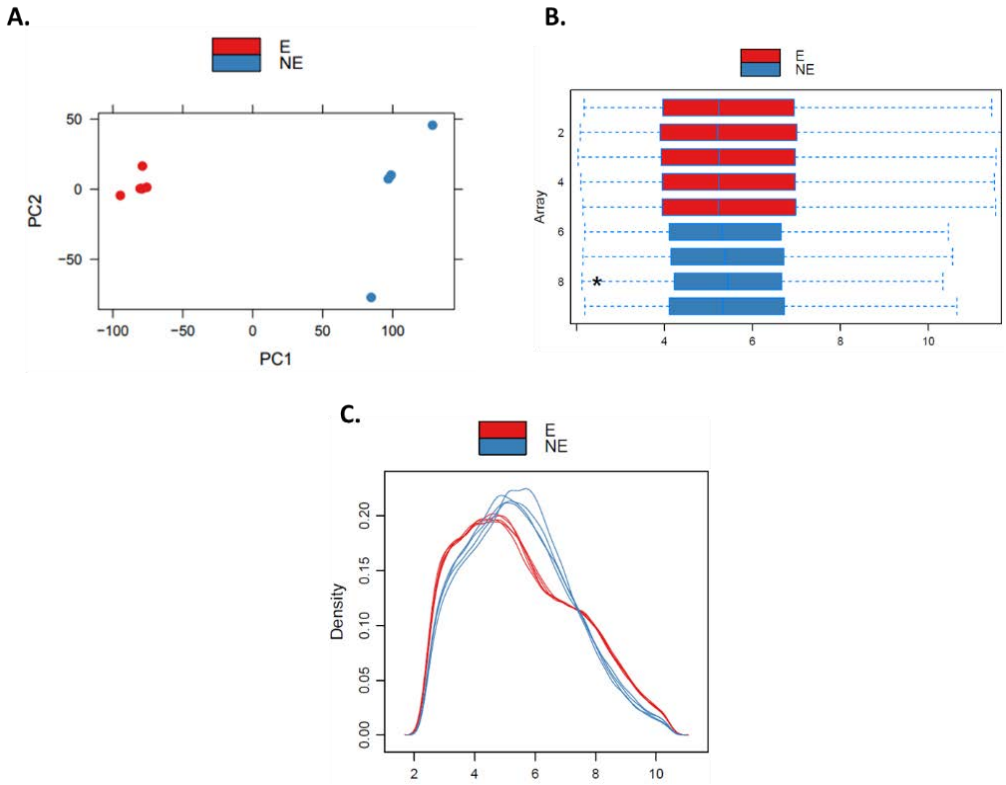


Figure 53. Pre-analysis of arrays from GEO. A) Principal component analysis. The PCA shows that stimulated (E) and non-stimulated (NE) expression arrays were different. **B) Box plot.** Visualization of the data through their quartiles. **C) Density graph.** Visualization of the density of expressions of the different arrays.

After *in silico* analysis, 6,348 genes were excluded from the study. This included 5,315 genes that vary throughout the menstrual cycle and 1,033 DEGs between natural and stimulated cycles. It can be observed in **Figure 54.A** that the proliferative and secretory phases had 3,624 genes in common and 5,315 genes that were unique for each phase. In **Figure 54.B**, it can be observed in the volcano plot that there were significant differences between stimulated and non-stimulated endometrium. In total, 1,033 genes were differentially expressed.

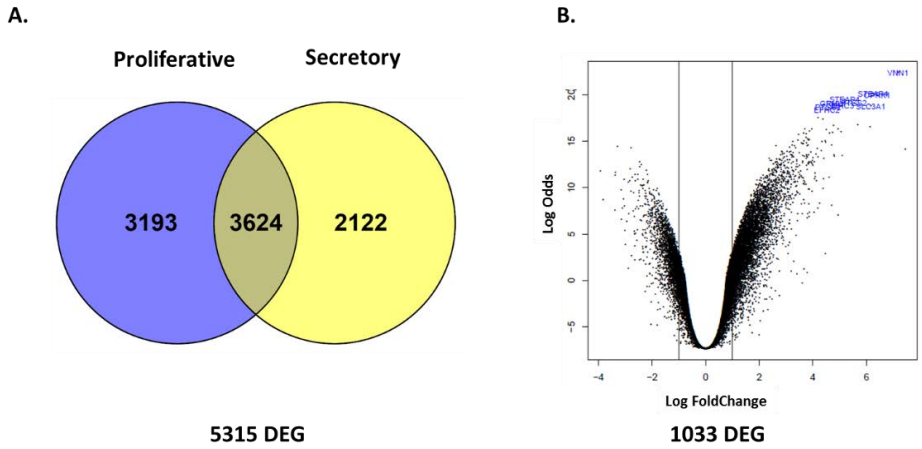


Figure 54. *In silico* studies. A) Venn diagram showing the differentially expressed genes between the proliferative and secretory phases. **B)** Volcano plot showing the differentially expressed genes between a stimulated (by FSH) and non-stimulated endometrium (FC±2 and p-value<0.01).

4.2.6. LGR5+ cells are suitable to perform RNA-Seq and there are no differences between LGR5 in epithelium and stroma

In this study, after the sorting, around 300 to 1,000,000 LGR5+ cells were obtained depending on the sample. After the sorting, RNA extraction was performed and very small amounts of RNA were obtained (7-50ng/µl). Samples were subjected to quality control prior to RNA-Seq by Bioanalyzer®. No intact ribosomal RNA (rRNA) peaks appeared in the Bioanalyzer profile. Their presence would indicate residual cell contamination (eukaryotic: 18S (1869nt), 28S rRNA (5070nt); prokaryotic: 16S (1542nt), 23S rRNA (2906nt)). Samples from the preliminary study (two ovarian and two control) underwent quality controls after the RNA-Seq performed, indicating that despite the low concentration of RNA, it was feasible to use RNA-Seq to study LGR5+ cells gene expression. After statistical analysis and the removal of the genes varying across the menstrual cycle and under FSH stimulation, 40 DEGs were found between ovarian and control (total endometrium: LGR5+ and LGR5- cells; FC±2, FDR<0.05). Thus, it was concluded that the samples were feasible for use in RNA-Seq. In addition, 13 DEGs were found when comparing LGR5+ versus LGR5- cells (FC±2, FDR<0.05). Although 255 DEGs were found between epithelium and stroma (FC±2, p-

value<0.05), when the adjusted p-value (FDR) was applied, no DEGs were observed. Therefore, it was decided to not separate compartments in subsequent samples.

4.2.7. LGR5 is not a biomarker for endometriosis. LGR5+ cells have a macrophage-like phenotype and DIE endometrium presents a special subset of LGR5+ cells

After the sequencing of 38 samples (LGR5+/- cells), 10 comparisons were made and DEGs were found in some groups. The total DEGs found in each comparison are shown in **Table 27**. In this case, the 6,348 DEGs that vary throughout the menstrual cycle and under FSH stimulation were also excluded from the study.

Comparisons	Genes	Significance	LogFold change
1. LGR5+Control vs LGR5+Endo	0	FDR≤0.01	±2
2. LGR5-Control vs LGR5-Endo	502	FDR≤0.01	±2
3. LGR5+Control vs LGR5-Control	0	FDR≤0.01	±2
4. LGR5+Endo vs LGR5-Endo	394	FDR≤0.01	±2
5. LGR5+Ovarian vs LGR5+Pelvic	0	FDR≤0.01	±2
6. LGR5+Ovarian vs LGR5+Adeno	0	FDR≤0.01	±2
7. LGR5+DIE vs LGR5+Ovarian	14023	FDR≤0.001	±2
8. LGR5+Adeno vs LGR5+Pelvic	0	FDR≤0.01	±2
9. LGR5+DIE vs LGR5+Pelvic	14567	FDR≤0.001	±2
10. LGR5+DIE vs LGR5+Adeno	17200	FDR≤0.001	±2

Table 27. Differences between LGR5+/- cells in control women and women with endometriosis. DEGs of different comparisons. Comparisons with significant differences (FDR≤0.01/FDR≤0.001 and FC±2)) are highlighted in grey.

No significant differences (FC±2 and FDR<0.01) were observed when comparing control and endometriosis LGR5+ cells (Comparison 1), which indicated that LGR5 is not a good biomarker for the disease. Furthermore, no differences between LGR5+/- cells in the control group (Comparison 3) were observed. Significant differences were found in Comparison 2 (502 DEGs) and Comparison 4 (394 DEGs).

Interestingly, when comparing types of endometriosis, DEGs were found only in comparisons where DIE was present (Comparisons 7, 9, and 10), indicating that DIE had a special subset of LGR5+ cells.

When analyzing Comparison 4 DEGs (LGR5+Endo vs LGR5-Endo), specific myeloid cell markers were found to be overexpressed, including CD33, CD300E, CD300LF, CD300LB, and CD200R1. In addition, overexpression of monocyte and Mφ markers, such as CD11b (ITGAM), CD163, CD86, CD209, CD14, CD180, CD68, CD84, CD1C, CD1A, CD45, CD53, CD300C, CD1B, CD300A, CD80, CD36, CD74, and CD93 were also found. When analyzing the biological significance using KEGG, 16 upregulated pathways were obtained **(Table 28)**.

Upregulated pathways	Entry
Galactose metabolism	hsa00052
Starch and sucrose metabolism	hsa00500
Butanoate metabolism	hsa00650
Chemokine signaling pathway	hsa04062
Phagosome	hsa04145
Osteoclast differentiation	hsa04380
Cell adhesion molecules	hsa04514
Antigen presenting and presentation	hsa04612
Toll-like receptor signaling pathway	hsa04620
JAK-STAT signaling pathway	hsa04630
Hematopoietic cell lineage	hsa04640
Natural killer cell mediated cytotoxicity	hsa04650
FcgR-mediated phagocytosis	hsa04664
Intestinal immune network for IgA production	hsa04672
Olfactory transduction	hsa04740
Carbohydrate digestion and absorption	hsa04973

Table 28. KEGG upregulated pathways in LGR5+Endo vs LGR5-Endo. The 16 upregulated pathways are shown in the table with their corresponding KEGG entry numbers.

As can be observed in the previous table, some of the upregulated pathways were related to the immune system. Chemokine signaling pathways (hsa04062) had the top enrichment score, followed by Toll-like receptor signaling pathways (hsa04620), Fc epsilon RI signaling pathways (hsa04664), phagosome (hsa04145), natural killer cell mediated cytotoxicity (hsa04650), cell adhesion molecules (CAMs, hsa04514), and antigen processing and presentation (hsa04612). Hematopoietic cell lineage (hsa04640) was also present**.**

As shown in **Figure 55**, LGR5+ cells overexpressed hematopoietic cell lineage markers. Among the myeloid lineage, the CFU-GM node showed the majority of overexpressed markers represented. All these findings strongly suggest that LGR5+ cells may be similar to monocytes and their derivatives (Mφ and dendritic cells). This fact is reinforced by the presence of the overexpressed markers CD45, CD68, CD300C, CD14, and CD163.

A.

Hematopoietic Cell Lineage hsa04640

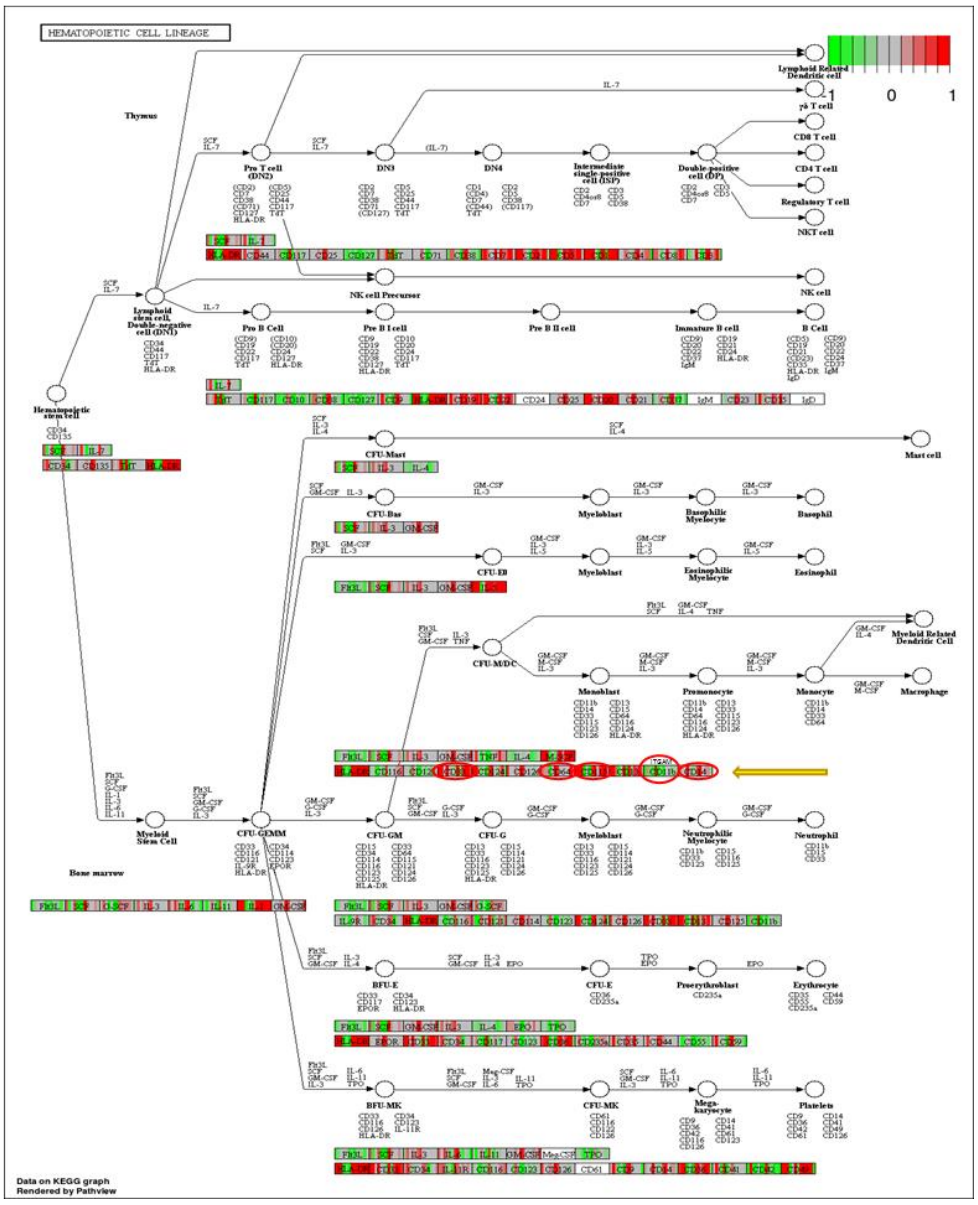
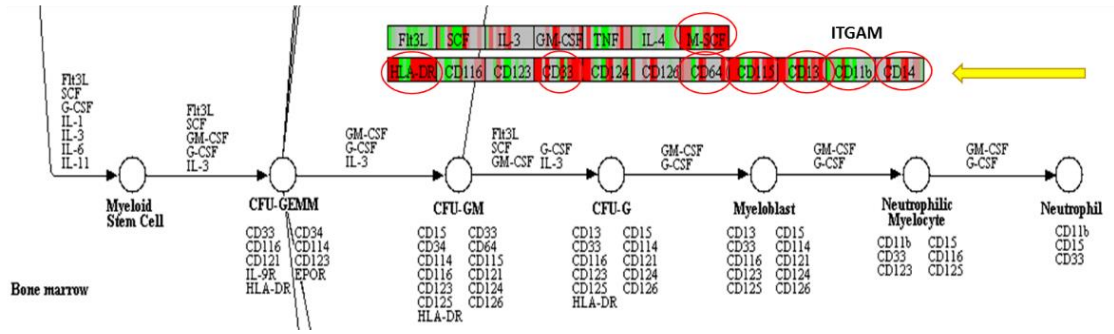


Figure 55. A) Hematopoietic cell linage in LGR5+ cells. Amplification of CFU-GM node in the next page. Red: overexpressed genes. Green: down-expressed genes. Each colored-line in each square represents one patient. The set of all patients determine the color and if the gene is over or down-expressed. **B) Hematopoietic cell linage (amplification of CFU-GM node). Overexpression of myeloid linage.** Overexpressed genes: HLA-DR, CD33, M-SCF, CD64, CD115, CD13, CD11b*, CD14. *CD11b is also called ITGAM. It is shown green (downregulated) in the figure. However, in the DEG list it appeared as ITGAM, which was overexpressed.

Figure 55 (continued).

B.



Finally, significant differences between pelvic endometriosis, ovarian endometriosis, and adenomyosis against DIE were observed (Comparisons 7, 9, and 10).

After biological analysis of these comparisons, common functions and genes were investigated in order to find the unique pathways, molecular functions, and genes overexpressed in DIE. An overexpression of inflammatory responses, immune cell trafficking, and hematological system development and function pathways in LGR5+ cells from DIE compared to other subtypes were found. Molecular and cellular functions, such as the activation of leukocytes and myeloid cells or inflammatory responses were observed, and CCL1, CCL11, DEFB4A, DEFB103A, CRH, PPM1D, and PRKCE were uniquely overexpressed in LGR5+ cells in DIE **(Table 29)**.

LGR5+DIE vs LGR5+ Other endometriosis	
Common functions	**Related DEG**
activation of cells	CCL1
activation of leukocytes	CCL11
activation of myeloid cells	CRH
chemotaxis of T lymphocytes	DEFB103B
inflammatory response	DEFB4A
	PPM1D
	PRKCE

Table 29. Unique molecular functions and genes overexpressed in the LGR5+ cells of DIE.

4.2.8. Macrophages have a different phenotype in eutopic endometrium of women with endometriosis

When working with the immune populations in eutopic endometrium of women with and without endometriosis, very few cells were obtained from the FACS (between one and 61,000 cells). **Table 30** shows the number of cells obtained and used to perform the library-prep and RNA-Seq.

	Mφ1	Mφ2	Treg	uNK
	number sorted cells		number sorted cells	
Control	957	5885	1090	73
	35	47	26	1
	273	2036	615	22
	61	142	9	6
	313	1041	37	32
	number sorted cells		number sorted cells	
Endo	772	9173	684	2420
	11055	60101	535	36590
	492	753	243	22
	183	606	348	317
	4372	34516	259	198
	87	69	59	1

Table 30. Number of immune cells obtained by FACS. Green cells indicate that RNA was extracted previous to the library-prep. In the white cells, library-prep was prepared directly from immune cells in PBS 1x. Endo: endometriosis; Mφ1: macrophages 1; Mφ2: macrophages 2; Treg: Regulatory T cells; uNK: uterine natural killers.

From the 44 sorted immune populations obtained by FACS, 40 samples (with more than 20 cells) were used for the study. Flow cytometry analysis from the 12 sorted samples were performed and results are shown below.

The viability of the samples previous to ICC was around 70%. After the ICC and when they had been sorted, the viability decreased to approximately 30% - 40%.

When CD45+ cells were gated, they only corresponded to 10% of the total sample **(Figure 56)**, and no significant differences were observed between the control and endometriosis groups. After the gating of the immune populations, some comparisons were statistically analyzed and are shown below. No significant differences between groups were observed except for Mφ1 in the control and endometriosis groups ($p = 0.0087$) **(Figure 57.A)**.

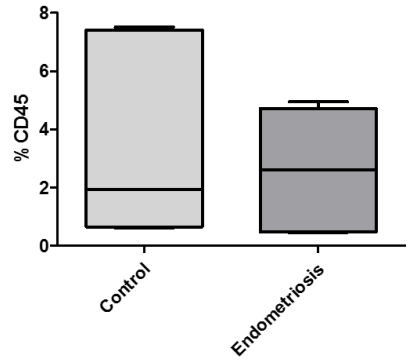


Figure 56. Percentage of CD45 positive cells. Control = 5; endometriosis = 6.

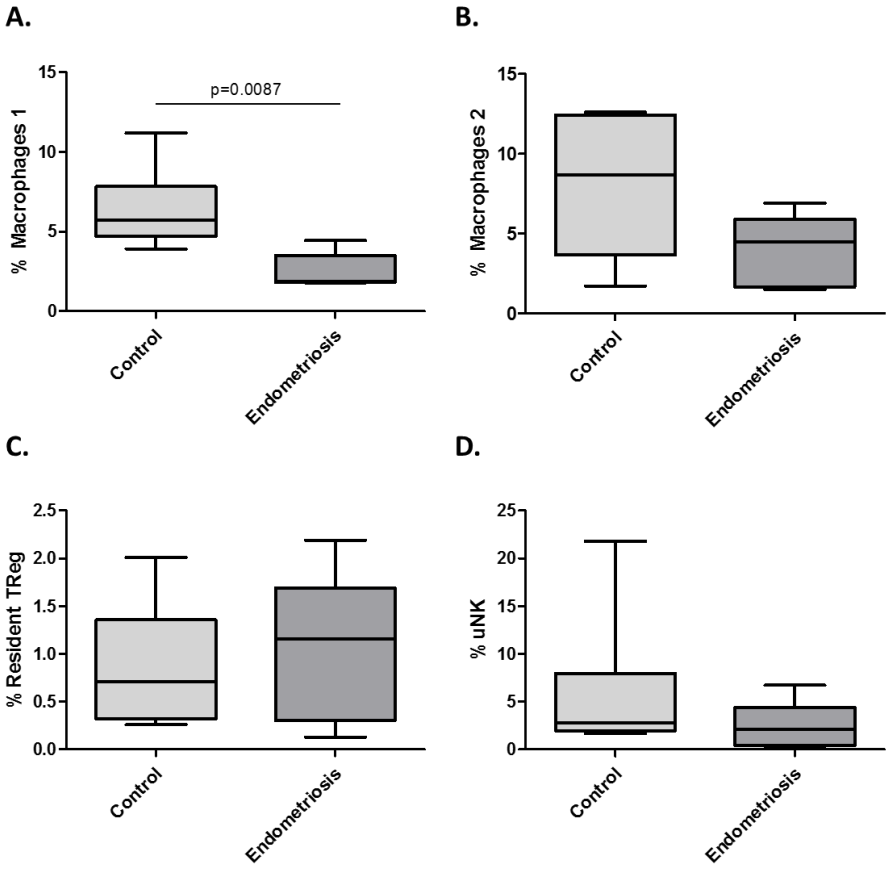


Figure 57. Percentage of immune cells in eutopic endometrium of women with and without endometriosis. Control = 5; endometriosis = 6. **A)** Macrophages 1. **B)** Macrophages 2. **C)** Resident Treg. **D)** Uterine natural killer.

Contamination of immune populations coming from circulation was also calculated, and a significant increase of uNK compared to blood NK was observed in both control and endometriosis groups **(Figure 58)**.

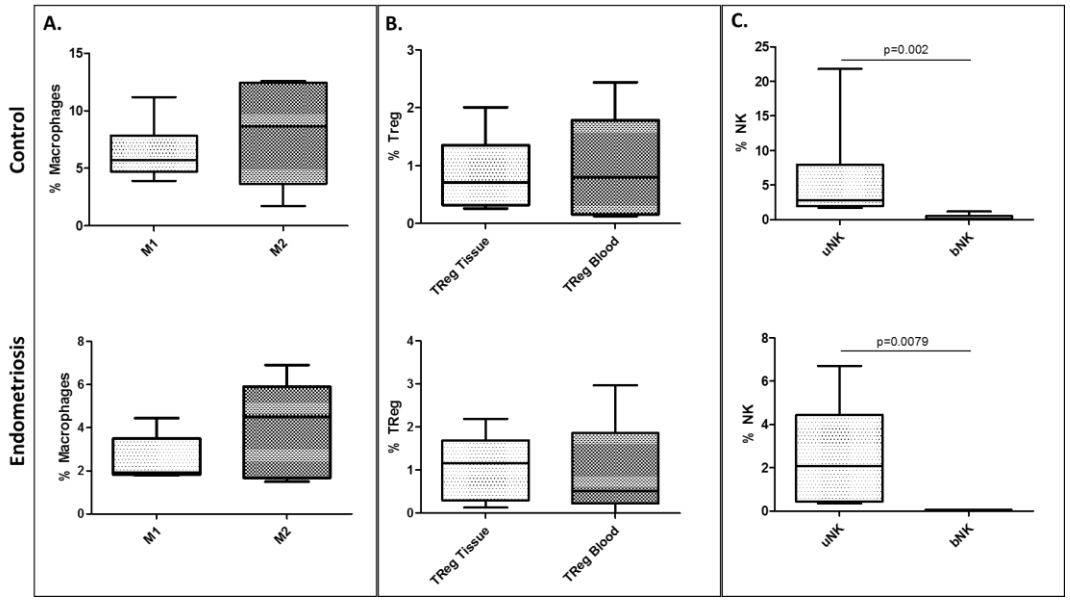


Figure 58. Percentage of immune cells in eutopic endometrium of women with and without endometriosis. In the upper panel, the control comparisons are shown. In the lower panel, the endometriosis comparisons are shown. **A)** Macrophages in control and endometriosis samples. **B)** Resident Treg in control and endometriosis samples. **D)** Uterine natural killer in control and endometriosis samples. Control = 5; endometriosis = 6.

For the gene expression analysis, the samples from which RNA was extracted (12 samples) were subjected to quality controls using a Tapestation system previous to the library-prep preparation. The RINs obtained ranged from 2.5 to 7.5. As the sample size was very small, it was decided that we should proceed with the library prep. After this step, another Tapestation quality control was performed previous to RNA-Seq. Fifteen samples did not undergo the quality controls, but they reached the minimal concentration recommended (10ng) for RNA-Seq. Two samples that did not reach the minimal concentration were included in order to have enough paired samples to compare in subsequent steps.

After the RNA-Seq, quality controls of FastQ sequences were also performed. A range of 5 million to 75 million reads was obtained. Samples with less than 10 million reads that

mapped to the reference genome were eliminated. Thus, 29 of the 40 samples were used for differentially expressed genes analysis. For this reason, the populations were reduced to 5 Mφ1 Endo, 3 Mφ1 Control, 6 Mφ2 Endo, 4 Mφ2 Control, 4 uNK Endo, 2 uNK Control, 3 Treg Endo, and 2 Treg Control. After statistical analysis, DEGs (FDR<0.05 and FC±2) were found in the different comparisons **(Table 31)**.

Comparisons	Sample size	DEG
1. Mφ1 Endo vs. Mφ1 Control	5 vs. 3	1422
2. Mφ2 Endo vs. Mφ2 Control	6 vs. 4	1544
3. Treg Endo vs. Treg Control	3 vs. 2	4146
4. uNK Endo vs. uNK Control	4 vs. 2	793
5. Mφ1 Control vs. Mφ2 Control	3 vs. 4	1260
6. Mφ1 Endo vs. Mφ2 Endo	5 vs. 6	705
7. Mφ Endo vs. Mφ Control	11 vs. 8	1567

Table 31. DEG between comparisons made in immune populations. Mφ1: Macrophages1; Mφ2: Macrophages2; Treg: Regulatory T cells; uNK: uterine Natural Killers.

Due to the fact that comparisons of uNK and Treg comprised less than three samples per group after the exclusion of some samples after the quality control, the statistics were not reliable. Thus, only comparisons comprised of Mφ and total immune populations were studied in more detail by using IPA® software. After the bioinformatics analysis of comparisons comprising Mφ (1, 2, 5, 6, and 7) statistically significant (z-score ± 2) molecular functions, relevant molecules secreted by Mφ1 and Mφ2 and upstream regulators were found and are shown in **Table 32**. Briefly, an increase in cell-cell contact were observed along with the repression of RNA molecular functions in comparisons of Mφ1 Endo versus Mφ1 Control and an increase of calcium concentration, transport of carbohydrates, and internalization of bacteria in the comparison Mφ2 Endo vs Mφ2 Control.

		MφEndo vs MφCtr	Z-Score	Mφ1Endo vs Mφ1Ctr	Z-Score	Mφ2Endo vs Mφ2Ctr	Z-Score	Mφ1 vs Mφ2Endo	Z-Score	Mφ1 vs Mφ2Ctr	Z-Score
Molecular functions		Senescence of fibroblast cell line	-2.09	Cell-cell contact	2.25	Concentration of Ca^{2+}	2.36	Cytotoxicity of leukocytes	-2.18	Growth of *S. Cerevisiae*	-2.00
		Activation of cells	1.95	Repression of RNA	2.19	Transport of carbohydrates	2.21			Engulfment of cells	2.12
						Internalization of bacteria	2.15				
		MφEndo vs MφCtr	**FC**	**Mφ1Endo vs Mφ1Ctr**	**FC**	**Mφ2Endo vs Mφ2Ctr**	**FC**	**Mφ1 vs Mφ2Endo**	**FC**	**Mφ1 vs Mφ2Ctr**	**FC**
Relevant molecules	**Secreted by Mφ1**	CX3CL1	2.6	CX3CL1	1.34	CX3CL1	3.453	CX3CL1	1.44	CX3CL1	2.76
		CCL5	-1.5	CCL5	-3.5	CCL5	1.29	CCL5	-1.43	CCL5	3.31
		CXCL16	1.4	CXCL16	-0.25	CXCL16	2.50	CXCL16	-1.74	CXCL16	1.00
		CXCL9	-0.95	CXCL9	-0.87	CXCL9	-1.12	CXCL9	1.01	CXCL9	0.76
		CXCL10	-1.03	CXCL10	-1.07	CXCL10	-0.99	CXCL10	-0.23	CXCL10	-0.16
		CXCL11	-0.94	CXCL11	-7.47	CXCL11	-1.19	CXCL11	-1.46	CXCL11	4.81
	Polarized Mφ2	Arg1	2.7	Arg1	3.85	Arg1	2.4	Arg1	-1.32	Arg1	-2.77
	Secreted by Mφ2a	CCL24	2.03	CCL24	5.65	CCL24	3.29	CCL24	-1.03	CCL24	-3.39
		CCL22	1.03	CCL22	1.45	CCL22	0.27	CCL22	1.65	CCL22	0.47
		CCL17	0.57	CCL17	1.71	CCL17	-0.42	CCL17	1.16	CCL17	-0.96
		CCL18	-0.2	CCL18	1.09	CCL18	-0.39	CCL18	-1.39	CCL18	-2.88
	Secreted by Mφ2b	CCL1	-1.23	CCL1	-5.6	CCL1	4.44	CCL1	-4.31	CCL1	5.74
	Secreted by Mφ2c	CCL18	-0.2	CCL18	1.09	CCL18	-0.39	CCL18	-1.39	CCL18	-2.88
		CCL16	3.47	CCL16	2.65	CCL16	4.68	CCL16	0.20	CCL16	2.23
		CXCL13	-2.1	CXCL13	-2.42	CXCL13	-1.1	CXCL13	1.44	CXCL13	2.76
		MφEndo vs MφCtr	**Z-Score**	**Mφ1Endo vs Mφ1Ctr**	**Z-Score**	**Mφ2Endo vs Mφ2Ctr**	**Z-Score**	**Mφ1 vs Mφ2Endo**	**Z-Score**	**Mφ1 vs Mφ2Ctr**	**Z-Score**
Upstream regulators		ERG	2,33	TNF	2,58	LH	-2,16	TNF	2,38	miR-483-3p	-2,00
								SATB1	-2,40	TGFB1	2,19
								INHBA	2,44		
								NFkB (complex)	2,78		
								Interferon alpha	2,18		
								IL15	2,22		
								VCAN	-2,00		
								TGFB1	-2,34		

Table 32. Ingenuity Pathways Analysis in macrophages comparisons. Significance: z-score±2. Mφ: macrophages; Ctr: Control; E: Endometriosis; CX3CL1: chemokine (C-X3-C motif) ligand 1; CCL5: C-C Motif Chemokine Ligand 5; CXCL16: C-X-C Motif Chemokine Ligand 16; CXCL9: C-X-C Motif Chemokine Ligand 9; CXCL10: C-X-C Motif Chemokine Ligand 10; CXCL11: C-X-C Motif Chemokine Ligand 11; Arg1: Arginase; CCL24: C-C Motif Chemokine Ligand 24; CCL22: C-C Motif Chemokine Ligand 22; CCL17: C-C Motif Chemokine Ligand 17; CCL18: C-C Motif Chemokine Ligand 18; CCL1: C-C Motif Chemokine Ligand 1; CCL16: C-C Motif Chemokine Ligand 16; CXCL13: chemokine (C-X-C motif) ligand 13; ERG: ETS transcription factor; TGFβ: Tumor Growth Factor beta; TNF: Tumor Necrosis Factor; SATB1: Special AT-Rich Sequence Binding Protein 1; INBHA: Inhibin A, Beta A; NFkB: Nuclear Factor Kappa B Subunit 1; INFα: Interferon alpha; IL15: Interleukin 15; VCAN: Versican; LH: Luteinizing Hormone.

Discussion

5. DISCUSSION

5.1. Discovery of new biomarkers in eutopic endometrium of women with endometriosis

Endometriosis affects around 10% of women in their reproductive years[1,2]. Despite the high prevalence of the disease and its effects on women's daily life (including the economic burden), public and professional awareness of this condition remains poor[199]. Although ovarian endometriosis can be diagnosed by transvaginal or abdominal ultrasonography, detecting peritoneal and some deep lesions is often a considerable diagnostic challenge. In addition, clinical signs and symptoms are commonly not endometriosis-specific, which may slow the diagnosis[199]. To date, no clinically relevant biomarkers or combination of biomarkers is meaningful. Thus, laparoscopy is still considered the gold standard for the diagnosis of the disease. A new approach is needed to avoid costly and invasive surgical intervention associated with morbidity and even mortality.

An ideal classification system for endometriosis should have the following features: able to provide information on the severity and type of endometriosis, has a correlation with the severity and type of symptoms including pain and infertility, accessible, reproducible and easy to perform, and able to provide information regarding prognosis of the disease.
An ideal diagnostic tool should be non-invasive, cheap (or less expensive than surgery), and have a very high sensitivity and specificity. In addition, the sample should be easy to collect, rapid, and easy to process and analyze. In the case of gynecological diseases and when using molecular-based tools (meaning that RNA, proteins, metabolites, etc. are measured), it is also important to take into account the menstrual cycle, because it is known that gene or protein expression vary throughout the menstrual cycle. Thus, a diagnostic tool independent of the menstrual cycle would be the best option for the patients as well as for the physicians.
In this thesis, we tried to develop a new and less-invasive approach for the diagnosis of endometriosis based on a combination-method. The novelty of the developed classifier is that it takes into account endometrial tissue gene expression but also includes phenotypic

variables of the patients. Only analyzing the genetic expression of the tissue in the discovery phase, we already obtained a high sensitivity and specificity (0.78 and 0.95, respectively), an accuracy of 86%, and an AUC of 0.94 by using four principal components, as can be observed in **Table 33**.

Binary PLS Model		
Parameter	**Without phenotypic variables**	**With phenotypic variables**
nComp	4	3
Accuracy	86%	97%
Kappa	0.73	0.95
AUC	0.94	0.98
Sensitivity	0.78	1
Specificity	0.95	0.94
Top10 regressors	**NM_006705**	Dysmenorreha
	NM_005345_3	Myomas
	NM_014819	Age
	NR_003503	**NM_006705**
	NR_003679	Hypermenorrhea
	NM_001304359	**NM_014819**
	NM_014373	Smoking
	NM_002557	**NR_003503**
	NM_033197	**NM_005345_3**
	NM_001258311	Fertility

Table 33. Comparison of binary PLS with and without the phenotypic variables of the patients. The table shows the number of principal components for the model (nComp), the accuracy, the kappa, the AUC, the sensitivity, the specificity and the top 10 important variables (regressors) for each model, with and without the phenotypic variables.

Nevertheless, when we included the phenotypic variables in the algorithm, sensitivity had a value of 1 and specificity a value of 0.94, meaning that all of the healthy patients were diagnosed as healthy while 94% of the patients with endometriosis were classified as such. The accuracy in this case increased to 97%, meaning that there is a 97% possibility that the results are credible in real life. In this case, the AUC was 0.98, indicating that the model almost perfectly predicted the condition of the patients. Interestingly, when phenotypic variables were included in the predictive model, only three principal components were needed to classify the patients. When looking at the top 10 regressors of both classifiers, we observed that the top four differentially expressed transcripts coincided with the four

transcripts of the model with phenotypic variables: NM_006705, which corresponds to GADD45G gene; NM_014819, which corresponds to PJA2; NR_003503, which corresponds to GGT8P; and NM_05345_3, which corresponds to HSPA1A. GADD45G, which was up-regulated in patients with endometriosis, is involved in growth arrest and is inactivated in multiple tumor types such as endometrial cancer[200]. In accordance with our results, Hwang et al. also observed this gene up-regulated in eutopic endometrium of women with endometriosis[78]. In addition, it was described as being present in eutopic endometrium of women with menstrual complaints[201], indicating that it could be involved in the dysmenorrhea present in endometriosis. Interestingly, Arimoto et al. found that GADD45G was downregulated in endometriomas compared to paired eutopic endometrium[202], indicating that its tumor suppressing activity may be involved in the development of the cysts. To our knowledge, PJA2 and GGT8P have not been described in eutopic endometrium of women with endometriosis. PJA2 is responsible for the ubiquitination of cAMP-dependent protein kinase type I and type II-alpha/beta regulatory subunits and for targeting them for proteasomal degradation. It is involved in immune system pathways[203]. GGT8P is a pseudogene and it has not been described in any pathology. In cooperation with other chaperones, HSPA1A stabilize preexistent proteins against aggregation and mediate the folding of newly translated polypeptides in the cytosol as well as within organelles[203]. In an endometriosis mice model it was described as being downregulated in ectopic lesions compared to eutopic endometrium[204].

Regarding the phenotypic variables, the most important regressors were dysmenorrhea, myomas, age, hypermenorrhea, smoking, and fertility. Dysmenorreha and myomas had an importance of more than 80%, while age and hypermenorrhea had an importance of around 50% and smoking and fertility an importance of less than 40%. Dysmenorrhea is one of the symptoms associated with endometriosis[1]. The disease can coexist with myomas (uterine fibroids)[74,75] and is associated with reproductive age and hypermenorrhea[1]. Moreover, myomas could correspond to adenomyomas, which are impossible to distinguish by TVUS. In addition, studies tried to elucidate the effect of tobacco in women with endometriosis and controversial results were obtained[26]. Here, we observed that smoking

may be an indicating factor of suffering endometriosis. Finally, fertility is also present in the classifier. We are aware that it is a difficult factor to take into account because some of the patients are infertile and/or have not tried to conceive yet.

After the validation study, we observed that the phenotypic variables were responsible for giving robustness to the model. When the combinatory model (genetic and phenotypic variables) was used, a sensitivity of 0.95 and a specificity of 0.604 were obtained, while when only genetic variables were present, the values were 0.49 and 0.51 respectively. However, when analyzing phenotypic variables alone, the values were higher – 0.96 and 0.64 respectively – indicating that the phenotypic variables were the variables with the best predictive value. In the discovery phase, the results were very promising using both genetic and phenotypic variables per separate and the statistical parameters increased in the combinatory model. However, in the validation phase, when the sample size was increased, we observed that genetic variables were not robust enough. This variation between phases could be due to the samples' heterogeneity.

Our study has some limitations. For example, we could not include pelvic endometriosis in the study because usually, this type of disease is an accidental finding by doctors and no endometrial biopsy is performed in these patients. The sample size for this type of sample was a limitation. We could only obtain five pelvic endometriosis samples, and therefore, they were not included in the study. Unfortunately, even though we included the other types of the disease, we were not able to discriminate between types of endometriosis, which would have made real progress in the field. As mentioned above, endometriomas can be easily diagnosed without surgery, but pelvic endometriosis is an accidental finding and adenomyosis is very difficult to diagnose previous to hysterectomy. Furthermore, samples used in this study were taken only once from each patient. Even though we demonstrated that our classifier was independent of the menstrual cycle, not all of the complete cycles in the same women express the same genes every month. Therefore, we believe that it should be elucidated whether the diagnostic classifier would have the same expression of the biomarkers in the same patients but in different cycles.

Technically, the candidate biomarkers found in the discovery phase were obtained from FFPE samples, and RNA was degraded. Some important biomarkers could have been missed due to this fact. Nonetheless, we had previously demonstrated a high correlation (>0.79) between fresh and FFPE samples analyzed by RNA-Seq. In addition, Norton et al. studied correlations between fresh and FFPE degraded samples as well as correlations between techniques. They reported a Pearson correlation between samples of >0.94 and >0.80 with Nanosting and RNA-Seq protocols respectively. Moreover, they found a correlation of 0.838 between both technologies using FFPE samples[205], as in our study. We observed a high correlation between both techniques. Illumina also performed some tests comparing the two techniques, and they found a correlation higher than 0.9[206]. The high degree of correlation between Nanostring and RNA-Seq platforms suggests that discovery based in whole transcriptome studies from FFPE material will produce reliable expression data. The fact that we obtained these biomarkers from FFPE samples is also a strength of this project, because the diagnosis of the disease could be made from fixed samples, which would greatly facilitate sample processing and preservation and offers the possibility of testing samples retrospectively. In addition, when a sample is fixed immediately after its collection, the biomarkers should be intact and no modifications in gene expression should occur. In this project, we used an innovative technique to validate the candidate biomarkers. Nanostring nCounter® has been approved by the FDA to test biomarkers of recurrence and subtypes in breast cancer. The developed diagnostic test is called Prosigna®, and it measures the gene expression of 50 genes. A report of the gene expression is generated to help with clinical decisions. Eight biomedical centers are already using this technology in Spain. The reproducibility of nCounter® is very high. Experiments performed by Geiss et al. yielded a replicate correlation coefficient of 0.999[207].

Although our algorithm was not validated, we also included in the CodeSet 11 more genes that were previously described as being de-regulated in eutopic endometrium of women with endometriosis. In the future, we would like to take advantage of this data and try to develop a new classifier, this time with a higher number of patients per group. With the

new classifier, we might be able to increase the specificity and eventually be able to differentiate between types of endometriosis.

Given this research, the prototype design of a new diagnostic kit for endometriosis would be based on Nanostring nCounter® technology and a new algorithm that could include the transcripts analyzed by this technique. A cut-off of the copy number of the expressed genes in the control group and patients with endometriosis should be implemented.

In addition, we believe that some improvements to the diagnostic tool could be performed in the future. For instance, we believe that our predictor could also be tested in other gynecological diseases, such as polyps, myomas, and endometrial and ovarian cancer, to elucidate if the combination of biomarkers is specific to endometriosis, or if it could also indicate other gynecological diseases. This is a possibility because endometriosis shares some symptoms with these gynecological conditions.

It would also be very interesting to test this new combination of biomarkers in samples collected during menses without the need to obtain an endometrial biopsy. Obtaining the same sensitivity and specificity in menstrual fluid would be much less invasive than an endometrial biopsy. In addition, we could also study a new combination of biomarkers in the mucous of the cervix. For this, it would only be necessary to perform a cytology, which is regularly done in the gynecologist's office. Because no endometrial tissue is present in this source, we would have to use a different approach such as proteome or lipidomics. In fact, Dominguez et al. performed a lipidomic profiling study in endometrial fluid in women suffering from ovarian endometriosis and obtained some very promising results. The methods in their study could correctly classify 86% of the samples as endometriosis[208]. It would also be interesting to find an even less invasive method for the diagnosis of endometriosis. In fact, we did a pilot study using peripheral blood where we looked for circulating endometrial cells (CECs). This study is explained in **Annex 2**.

It has been described that between 25% and 50% of infertile women suffer endometriosis and that around 30-50% of women with endometriosis are infertile[4]. Our classifier does not identify women with endometriosis who are also infertile. We believe that it would be a

breakthrough in reproductive medicine if we could include in the study fertile and infertile women suffering from the disease and identify the biomarkers of infertility. Therefore, doctors and patients would have the chance to decide whether they would prefer to perform ART or surgery.

In conclusion, we obtained a predictor model that is less invasive than laparoscopy and allowed us to identify 60% of endometriosis patients as such. However, due to the low specificity, 40% of the women with endometriosis were classified as healthy. Thus, when a women is diagnosed with the disease, we can be sure that she suffers endometriosis with an error of 5%. Therefore, even though almost all of the control women were classified as healthy, when a woman is classified as healthy, we cannot be sure that this is true.

5.2. To evaluate the potential of LGR5 as a biomarker for the disease and the implication of LGR5 and immune cells in the pathophysiology of the disease

Only four studies of LGR5 have been performed in healthy human eutopic endometrium. To our knowledge, this is the first work where LGR5 in eutopic endometrium of women with endometriosis has been explored. Since LGR5 has been well described in some tissues as a stemness marker[209], we postulated that LGR5 could be a universal stem cell marker and could be involved in the pathophysiology of endometriosis. In addition, a previous work done in the endometriosis mice model[108] had shown an abnormal epithelial phenotype in the stromal compartment of the eutopic endometrium of mice with induced endometriosis. Therefore, we hypothesized that it could also be a diagnostic marker for endometriosis. However, as we demonstrated in this thesis, none of our hypotheses were confirmed.

Cervelló et al. demonstrated in healthy endometrium[101] that LGR5+ cells in eutopic endometrium present a monocyte-macrophage-like phenotype, and they do not express any stem cell marker, indicating that these cells are not stem cells in this tissue. However, these cells could be involved in stem cell niche modulation. These results are comparable to our findings in patients with endometriosis.

On the other hand, we demonstrated that 70% and 80% of the patients with endometriosis presented an abnormal co-localization of LGR5 with CK and ECAD respectively in the stromal compartment of the eutopic endometrium. Strikingly, this co-expression was not found in healthy women, showing that LGR5+ cells seem to behave differently in women with endometriosis. In our opinion, this characteristic phenotype could potentially be used as a diagnostic maker of endometriosis. However, when studying RNA-Seq results, we did not observe any significant differences between women with and without endometriosis, indicating that it is not a good biomarker for the disease.

Moreover, previous studies reported the presence of 1% to 2% of LGR5+ cells in healthy endometrium[92], while we observed 2% to 7% of LGR5+ cells in eutopic endometrium of women with endometriosis. Interestingly, in accordance with other groups, our results showed no differences between the percentage of LGR5+ cells present in epithelium and stroma measured by flow cytometry or in gene expression after RNA-Seq analysis[101].

In addition, we also studied the differences between the percentage of the marker in controls and in women with endometriosis. We observed controversial results in this step. When we studied LGR5 percentage using flow cytometry, we did not find differences between the two categories (controls: n=9 and endo: n=26). However, when we measured LGR5 percentage using FIM (fluorescence intensity mean) by IF, we observed a significant increase in LGR5 in healthy endometrium in all phases of the menstrual cycle when compared with endometrium of women with endometriosis (control: n=24 and endo: n=24). This results may be different for several reasons: 1) different methods of sample processing (from fresh tissue or FFPE endometrium), 2) different sample sizes used (in the flow cytometry study the control group represents only one third of the total endometriosis samples), and/or 3) different antibodies used (we used two different antibodies for each technique).

The endometrium is a highly dynamic tissue that changes throughout the menstrual cycle. Thus, we wanted to elucidate if LGR5 was varying across the menstrual cycle in women with endometriosis. It was described that in normal endometrium, LGR5 does not vary

throughout the cycle at the RNA and protein level[92,100]. However, recently, Tempest et al. reported that LGR5 was regulated by progesterone and decreases in the secretory phase of the menstrual cycle[30]. In the discussion of their paper, they claimed that the LGR5 antibodies used in Cervelló et al. study[101], which also correspond to the antibody used in our study (AP2745A), were not specific. They tested the IF antibody and concluded that it was not specific. However, we did not use the same concentration as they did (1:100). We used 1:30 concentration of primary antibody. Furthermore, they used it to perform IHC, and we performed IF. In addition, for the flow cytometry analysis, we did not use the same antibody. In fact, to prove that the LGR5+ cells sorted by FACS were truly positive, we did a cytospin of LGR5+ cells and performed IF with the antibody AP2745A on the slide. As a result, all LGR5+ cells sorted were positive **(Figure 59)**. Therefore, we think that the contradictory results observed by Tempest et al. may have been due to the different techniques and concentrations used.

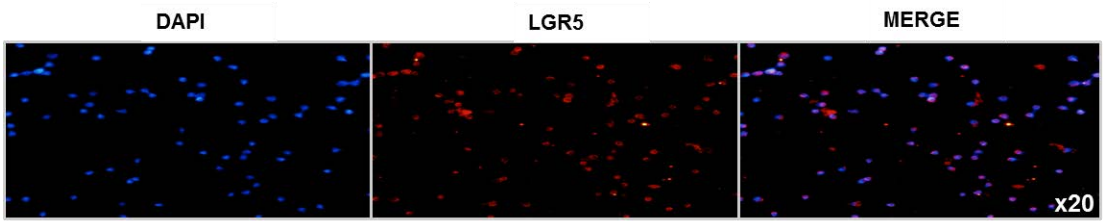


Figure 59. Cytospin LGR5. Cells sorted using NBP1-28904 antibody (for FACS) and stained with AP2745A antibody (for IF). The left panel shows the cells stained by DAPI (blue), the middle panel shows LGR5 in red, and the right panel shows the two images merged. Magnification: x20.

Additionally, in their work it is mentioned that LGR5 varied along the menstrual cycle when *in silico* studies were performed. They amalgamated three lists from publically available microarray datasets of genes expressed during the menstrual cycle in women without hormonal treatment[73,210–212]. They explored the possible regulation of LGR5 by progesterone and suggested that LGR5 could be regulated by this hormone, as they found progesterone binding sites in the promoter of LGR5. Nevertheless, they did not report the direct LGR5 variation throughout the cycle from those lists. On the other hand, in our *in silico* study we studied the genes that vary along the cycle using GEO arrays and observed

that LGR5 was expressed in the proliferative as well as in the secretory phase. In addition, it did not show DEG between phases, reinforcing the idea that LGR5 does not vary along the menstrual cycle in normal endometrium. It is also true that we only used one of the three lists (Talbi et al., 2006) that they used[73]. Interestingly, in accordance with Tempest et al., we observed a slight decrease of LGR5 in the secretory phase when progesterone was added to the medium in the control group in the *in vitro* experiments (*see Figure 43*); reinforcing the idea that LGR5 could be regulated by progesterone. However, the decrease was not significant. Either way, in our study, not only the *in silico* study suggested that LGR5 does not vary across the menstrual cycle in healthy women, but also the *in vitro,* IF, and flow cytometry experiments (*see Section 4.2.3*). Moreover, we also performed these experiments in women with endometriosis and observed that there was no variation throughout the menstrual cycle. However, in this case, we observed a trend towards an increase in the secretory phase. Interestingly, it has been described that stromal fibroblasts have progesterone resistance in women with endometriosis[213]. Although not significant, our results suggest that LGR5 could be regulated by progesterone, as was also suggested by the fact that it has progesterone binding sites in its promoter[30]. Due to the resistance to this hormone in women with endometriosis, the marker is not downregulated in the secretory phase in these women.

In our study, endometrial samples were obtained randomly throughout the menstrual cycle after confirming that the LGR5 percentage did not vary. However, in order to normalize and compare the results, we excluded the genes that significantly changed along the menstrual cycle and in a stimulated cycle (due to the controls used in this study – egg donors). This process has the advantage that our results can be interpreted independently of the phase of the menstrual cycle, but has also limitations, as some relevant genes involved in pathways that change through the menstrual cycle could have been inadvertently excluded. We are aware that we obtained the differentially expressed genes throughout the menstrual cycle and under FSH stimulation from two arrays studies. For this reason, we used IPA software, where we only took into account the DEGs that codify for proteins. Several studies have been done using available datasets from different platforms[214–220], and

it has been demonstrated in other studies that DEGs from arrays and RNA-Seq were similar[163–169,221].

The gene expression profile of LGR5+ human endometrial cells in women with endometriosis showed that members of Wnt pathway were downregulated, in contrast with the migrating cells of the mouse model, where Wnt7a was overexpressed[108]. These results also support the evidence that LGR5+ cells may not behave as stem cells in human eutopic endometrium as previous works have demonstrated[101].

We also observed overexpression of certain hematopoietic markers in LGR5+ cells such as CD33, CD300E, CD300LF, CD300LB, and CD200R1, supporting the fact that LGR5+ cells from human endometrium of women with endometriosis seem to be of myeloid lineage, as has been previously demonstrated to occur in healthy endometrium[101]. In our work, among the myeloid lineage, the colony forming unit-macrophages/dendritic cells (CFU-M/DC) node included the majority of overexpressed markers represented. All these findings strongly suggest that LGR5+ cells may be monocytes and their derivatives (Mφ and dendritic cells). This is reinforced by the presence of the overexpressed markers CD45 (leukocyte marker), CD68 (monocyte-linage marker), CD300C (monocyte marker), CD14 (monocyte and macrophages marker), and CD163 (Mφ2 marker) **(Figure 60)**.

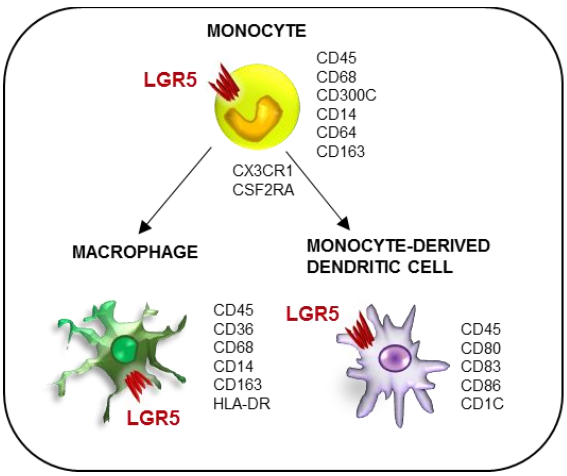


Figure 60. Overexpressed markers of CFU-GM node. Specific markers from the myeloid linage corresponding to monocytes and their derivatives.

Other genes related to this cell lineage were overexpressed in LGR5+ cells. For instance, CX3CR1, a chemokine involved in the adhesion and migration of leukocytes; CSF2RA, a cytokine which controls the production, differentiation, and function of granulocytes; and Mφ; as well as several dendritic cell markers such as CD1C, CD1E, CD83, CD207, and HLA-DR.

Previous works[108] using a mice endometriosis model suggested that a selective migration of cells to the eutopic endometrium could be regulated by an epithelial mesenchymal transition (EMT) process as they expressed aberrantly epithelial markers in the stroma. According to our findings, we suggest that EMT may explain the fact that LGR5+ cells display a monocyte/macrophage gene expression profile, as do cytokeratin in the stroma of the eutopic endometrium. This is also reinforced by overexpression in LGR5+ cells of MMP12, which is a metalloproteinase involved in the degradation of the extracellular matrix and involved in EMT process[222]. A recently published article[101] demonstrated the presence in the endometrium of CD45+ and CD45- LGR5+ cells. Interestingly, there were no significant gene expression profile differences between these groups. In the article is also discussed the existence of two different origins of LGR5+ cells: LGR5+ cells coming from bone marrow and LGR5+ eutopic endometrium resident cells. Therefore, according to these findings, there seems to be a population of LGR5+ macrophage-like cells coming from bone marrow, which are CD45+, and another perivascular population of LGR5+ macrophage-like cells, which are CD45-. In our study, we mostly observed overexpression of the myeloid markers in the LGR5+ population. Taking into account that a significant number of LGR5+ cells co-express CD45+, we tended to conclude that LGR5+ cells were coming from bone marrow. However, we cannot fully exclude the hypothesis that LGR5+ can also come from a trans-differentiation of stromal fibroblasts into an epithelial phenotype by an EMT process that could potentially lead to the migration of these cells and the production of endometrial lesions outside of the uterus.

Mφ are also involved in tissue repair and remodeling[223–225]. In some tumors, Mφ produce factors that foster tumor progression through the production of soluble mediators that support proliferation, angiogenesis, survival, and invasion of malignant cells[226]. It is known

that DIE usually represents an aggressive type of disease with increased invasion, proliferation, and angiogenesis in comparison to other types of endometriosis[1]. Interestingly, when we compared LGR5+ cells from different types of endometriosis, we found seven overexpressed genes in DIE: DEFB103A, DEFB4B, CCL1, CCL11, CRH, PPM1D, and PRKCE, which are all related to inflammatory processes and may have an impact on reproductive outcomes **(Figure 61)**.

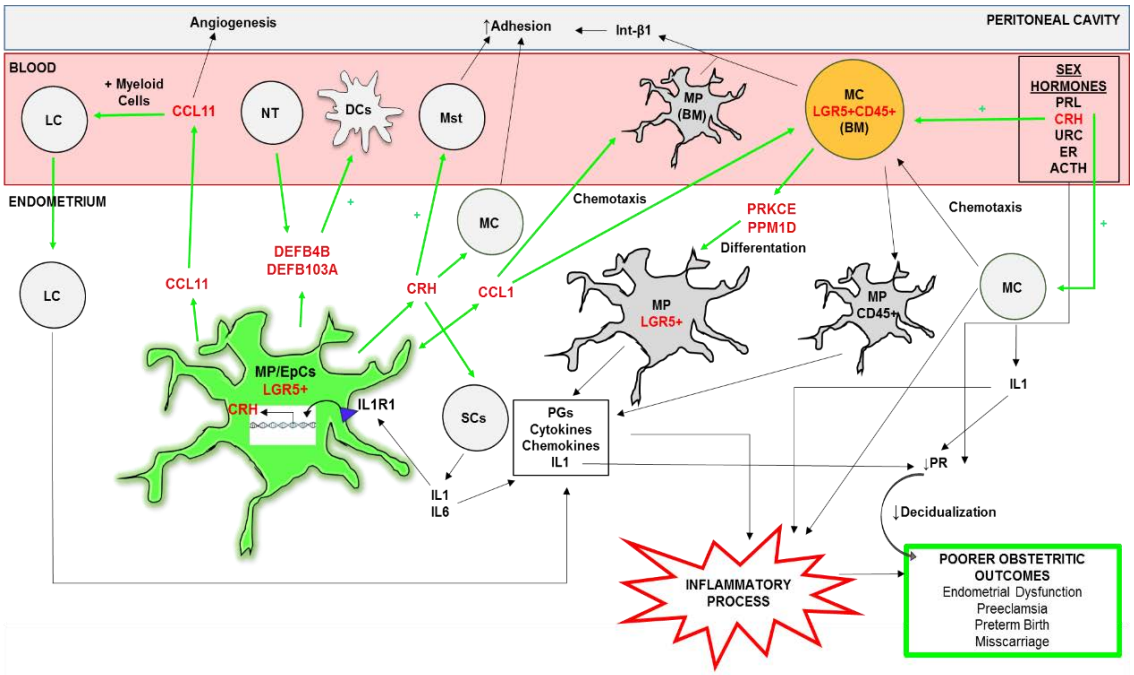


Figure 61. LGR5+ cells from deep infiltrating endometriosis. The figure shows LGR5+ cells (in green) and the seven overexpressed genes in DIE: DEFB103A, DEFB4B, CCL1, CCL11, CRH, PPM1D, and PRKCE (in red). We can observe the effects of the genes in other cell types and the consequences of their expression. They are all related to inflammatory processes that have an impact on reproductive outcomes. MC (BM): monocytes from bone marrow; MP (BM): macrophages from bone marrow; MC: monocytes; MP: macrophages; LC: leukocytes; NT: neutrophils; DCs: dendritic cells; SCs: stromal cells; EpCs: epithelial cells; Mst: mast cells.

Innate defenses of the human endometrium play a critical role in the maintenance of an environment hospitable to fertilization, fetal implantation, and successful pregnancy. Circulating monocytes migrate into tissues and differentiate into Mφ, which play an important role in the initiation, maintenance, and resolution of inflammatory responses. This functions are mediated through the production of innate effectors such as pro-

inflammatory interleukin 1 (IL-1)[227]. IL-1 activates uterine epithelial cells to induce DEFB4B, an innate defensin. CCL1 is a chemokine released by monocytes and Mφ, and acts as a chemoattractant for neutrophils and monocytes into different tissues. It also promotes the expression of integrin-β, which is involved in embryo adhesion processes[228]. Therefore, it seems plausible that LGR5+ cells in DIE could be responsible for recruiting more immune cells in the endometrium by overexpressing CCL1 and potentially affecting embryo implantation.

On the other hand, CCL11 is a chemokine initially identified as a specific chemoattractant protein for eosinophils. However, recent studies indicate that it can have a role in mediating activities of myeloid cells during development and pathological states[229] and that it can have functions in the endometrium other than eosinophil chemoattractant[230]. Further studies showed that the concentration of CCL11 is elevated in the peritoneal fluid of women with severe endometriosis[229,231] and has angiogenic activity[231] that mediates directly angiogenic responses[232], both processes present in DIE. Additionally, Hornung et al. explained that in interacting with other cytokines and immune cells, CCL11 contributes to an inflammatory reproductive tract environment, leading to endometrial or blastocyst dysfunction[229] that could potentially impair implantation.

CRH is found in both epithelial and stromal endometrial compartments[233], although it is mainly produced by epithelial cells[234]. CRH is secreted at inflammatory sites and serves as an autocrine and paracrine modulator[235] with pro-inflammatory properties influencing both innate and acquired immune responses[236], and these properties have been reported in endometriosis[237]. Moreover, CRH has been shown to participate in an immune-regulatory manner in ovulation, luteolysis, decidualization, embryo implantation, and maintenance of human pregnancy[235–239]. An upregulation of CRH was observed in abortions[235,240]. Indeed, it seems plausible that CRH may have a negative reproductive effect in the endometriotic endometrium.

PPM1D controls a number of critical cellular functions such as proliferation, cell cycle arrest, and programmed cell death, and is also implicated in differentiation and regulation

of the activity of hematopoietic stem cells[241]. PPM1D seems to be overexpressed in ovarian clear cell carcinoma (CCC)[242], which is one of the ovarian cancers often associated with endometriosis. Although inconclusive, these initial findings may help to illuminate the link between these two pathological entities.

Finally, PRKCE plays a major role as a critical mediator of several signaling cascades in activated Mφ[243] and is involved in monocyte-derived dendritic cells differentiation[244]. This data supports our proposed theory that LGR5+ cells are monocytes and their derivatives are involved in the innate immune response. Interestingly, PRKCE seems to be overexpressed exclusively in monocytes[245].

We believe that it would be very interesting to study LGR5+ cells in ectopic lesions and compare them with eutopic LGR5+ cells.

Some of the work done on LGR5+ cells was published in Fertility and Sterility last September (2017) **(article cited in Annex 3)**.

Additionally, we were interested in studying Mφ in eutopic endometrium of women with endometriosis after determining that LGR5+ cells were macrophage-like cells that could be involved in the aggressiveness and reproductive outcomes of the disease. Mφ not only maintain organ homeostasis and facilitate host defense and wound healing, but they also underlie the pathogenesis of many chronic diseases[246]. Abnormal distribution of Mφ within the eutopic endometrium could contribute to both the aberrant distribution of immune cells in the pelvic cavity and the abnormal development and gene expression of the eutopic endometrium. In addition, and as was explained in introduction section, Treg and uNK are two populations that are very important in fertility and seem to behave differently in eutopic endometrium in endometriosis. Therefore, we decided to study the gene expression of all these immune populations and observe the possible effects that they could have on the eutopic endometrium of diseased women. This work was performed at Dr. Giudice's lab in UCSF.

We developed a cytometry panel that allowed for separating circulating (blood) immune cells and tissue resident cells and using a gating strategy as well as collecting the activated resident tissue immune cells. Thus, the analyzed immune populations were purely tissue activated resident cells devoid of contamination by circulating immune cells. Of the immune populations isolated, there were insufficient yields of uNK and Treg cells after the quality control post-sequencing for comparing patients and controls (*see results Section 4.2.8*). Therefore, only macrophages were studied in more detail, where sufficient reads in the sequencing were obtained.

As demonstrated in **Table 33**, five comparisons were analyzed using IPA® software. In the comparison of Mφ1 Endometriosis versus Mφ1 Control, there was an increase of the molecular functions cell-cell contact and repression of RNA. In a preliminary study, Khan et al. demonstrated a higher presence of *Escherichia coli (E. coli)* in eutopic endometrium of women with endometriosis[124]. The increased cell-cell contact in Mφ1 in disease is consistent with increased adhesion to bacteria to accomplish bacterial engulfment, although a bacterial pathogenesis of endometrial is not well established. The top deregulated networks in Mφ1 Endometriosis versus Mφ1 Control include cellular development, cellular growth and proliferation, lymphoid tissue structure and development, and infectious diseases, antimicrobial response, and inflammatory response. These data indicate that Mφ1 in endometriosis present a more pro-inflammatory phenotype than Mφ1 in the control group without the disease.

The deregulated molecular functions upregulated in Mφ2 in endometriosis included an accumulation of Ca^{2+} in Mφ2, an increase of transport of carbohydrates, and an internalization of bacteria. These molecular functions are more related to the Mφ1 phenotype than to the Mφ2 phenotype. The question may arise of whether Mφ populations were correctly separated, but when looking at specific markers in the comparison of Mφ1 versus Mφ2 both in controls and endometriosis, we observed that they were correctly sorted. Arginase (Arg1) is the only factor identified so far to detect Mφ2 polarization in human samples[247]. This gene was downregulated in both comparisons,

indicating that it was overexpressed in Mφ2 and thus that the cells were correctly separated. Moreover, Mφ2 were stained and selected by the marker CD163, a specific marker for Mφ2.

A transient increase of Ca^{2+} plays a role in the expression of TNFα by Mφ1. Intracellular Ca^{2+} oscillations are likely to induce permanent changes in Mφ physiology, and a supra-physiologic elevation of Ca^{2+} in mitochondria can be cytotoxic and likely to induce apoptosis in the long term[248]. Our results show that an accumulation of Ca^{2+} prevails in Mφ2 in endometriosis. We also observed a predicted increase of TNFα when comparing Mφ1 of women with versus women without endometriosis, suggesting that Mφ1 are responsible for this Ca^{2+} accumulation in Mφ2 through TNFα secretion. Therefore, Mφ2 in the diseased endometrium may undergo more apoptosis versus Mφ2 in the controls, which would account for why the Mφ1 versus the Mφ2 phenotype is predominant in endometrium in women with versus without the disease[123]. On the other hand, our results show that transport of carbohydrates was increased in Mφ2 in endometriosis. It has been described that glycolysis is increased in Mφ1 and is reduced in Mφ2, and that Mφ polarization may derive from a reprogramming of glucose metabolism[249]. Mφ are phenotypically highly plastic, and their polarization state depends on their microenvironment. It is possible that Mφ2 undergo polarization to Mφ1 due to the altered environment in endometriosis and adopt the Mφ1 pro-inflammatory phenotype[249]. Over the last decade, several studies have suggested that by altering nutrient availability or by blocking specific metabolic pathways, it is possible to skew the Mφ phenotype, altering their effector functions in chronic inflammatory diseases[250]. Therefore, Mφ metabolism modulation opens a new paradigm and a new therapeutic window for treating inflammatory diseases. Finally, the increase of internalization of bacteria in these cells also indicates that they may have an Mφ1 phenotype. Regarding the top deregulated networks in Mφ2 in endometriosis, we observed a deregulation of connective tissue disorders, inflammatory disease and inflammatory response, and drug metabolism, endocrine system development and function and lipid metabolism, which also indicates that Mφ2 in endometriosis have more pro-inflammatory properties than Mφ2 in control women.

In our dataset, we also observed that luteinizing hormone (LH) signaling is predicted to be downregulated in Mφ2 when comparing endometriosis versus controls. One study described that macrophages in the late luteal phase of human endometrium contain LH receptors[251]. This indicates that LH may regulate macrophage functions in gonadal as well as in non-gonadal target tissues. Our results suggest that LH signaling is downregulated in eutopic endometrium of women with endometriosis, leading to a reduction of the implantation rate of the embryo, as it has been shown that endometrial maturation (thickness of the endometrial tissue during the secretory phase) is disturbed in women with low endogenous LH independent of ovarian function[252].

Interestingly, when analyzing upstream regulators of Mφ1 versus Mφ2 in endometriosis, we observed that some molecules that should be downregulated by progesterone were increased, such as NFKB and TNFα. Our samples were obtained in the secretory phase where progesterone levels are high, which means that those molecules should have been downregulated. Instead, they were increased in the IPA. Endometriosis is an estrogen-dependent disorder with a blunted response to progesterone in select cell populations[213], underscoring the plausibility of an endocrine-immune network that might participate in the pathophysiology of the disease. In addition, progesterone resistance might also have an effect on stromal fibroblast decidualization, causing it to not happen correctly. Hence, an implantation failure might occur.

Of note, rather than being distinct macrophage populations, Mφ1 and Mφ2 gene expression signatures often overlap, and the resultant mixed phenotype then depends on the tissue environment. In fact, if we look at the common genes between comparisons of Mφ1 versus Mφ2 in control and endometriosis, we find that 206 of 2,553 genes were expressed by both subpopulations. We observed that Mφ slant to an Mφ1 phenotype in the eutopic endometrium of women with endometriosis[123]. Although we did not observe numerical differences in sorted Mφ subtypes, our results indicate that the Mφ1 phenotype is predominant in this tissue in women with the disease. As mentioned above, the increase of Mφ1 (or the increase of its pro-inflammatory phenotype) in this tissue could be due to

the high presence of LPS, which could be due to the higher presence of *E. coli* in diseased women, although bacterial infection is not a well-accepted paradigm for endometriosis[124]. It has been postulated that the cause of the inflammation in eutopic endometrium in endometriosis may be of bacterial origin[253], and recently, microbiota studies have been performed in women with and without endometriosis to elucidate bacteria's impact on endometrial function[254]. In fact, the endometrial microbiome is important in the acquisition of endometrial receptivity, and pathological modification of this microbiome has been related with poor reproductive outcomes[255]. Therefore, it is plausible that the aberrant expression of Mφ in endometriosis may also be related to the microbial flora in this tissue, and could also impact reproductive outcomes.

Although endometriosis is not a cancer, the development of endometriosis mimics the process of metastasis in cancer. It has been proposed that tumors take advantage of macrophage plasticity to their own benefit. In early phases of cancer, high production of Mφ1 inflammatory mediators activates the adaptive immune response capable of eliminating nascent neoplastic cells, but also appears to support neoplastic transformation[123]. The Mφ1 pro-inflammatory phenotype increases NFKB and downstream events and increases transcription of pro-inflammatory cytokines such as TNFα, IL12, IL23, IL1β, IL6, and ROS. Indeed, it has been described that the NFKB pathway is deregulated in the eutopic endometrium of women with endometriosis,[256] and if we focus on Mφ1 versus Mφ2 in endometriosis, we can observe that the NFKB pathway is activated. This does not occur in Mφ1 versus Mφ2 in control women.

Taking this evidence together, we can conclude that Mφ in eutopic endometrium of women with endometriosis show an aberrant pro-inflammatory phenotype compared to the control group. This increase in inflammation could be related to the pathophysiology of the disease as well as Mφ having an effect in the reproductive outcomes.

These results should be taken with caution, as both studies have several limitations. First, the number of samples assessed was small, which limited the statistical analyses. Additionally, the donors did not undergo laparoscopy prior to endometrial biopsy, and so endometriosis in the controls cannot be accurately excluded. Previous reports[257] have

observed endometriosis in 4% of asymptomatic women undergoing laparoscopic tubal ligation. Therefore, in order to minimize the possible misdiagnosis of endometriosis, we have carefully selected donors of proven fertility and without dysmenorrhea and/or cysts in the ovaries.

Our results show for the first time aberrant LGR5+ cells co-expressing epithelial markers in the stromal compartment of women with endometriosis that have a significantly different expression profile in DIE. In addition, macrophages in the eutopic endometrium of women with disease have a different transcriptome profile than healthy women. Specifically, Mφ1 and Mφ2 have a higher pro-inflammatory phenotype in endometriosis than in the control group, and Mφ2 appear to be predisposed to polarization in women with endometriosis, thus increasing their inflammatory phenotype in disease. All of these findings suggest that eutopic endometrium has different gene signatures depending on the presence or absence of endometriosis. Whether subtypes of the disease affect different subsets of cells has yet to be determined. Moreover, our results may have implications regarding reproductive outcomes, although further research is required to elucidate these issues.

Conclusions

6. CONCLUSIONS

Objective 1:

1. Fresh and FFPE samples are comparable in terms of gene expression, as analyzed using RNA-Seq.

2. With the multiclass PLS predictive model used, the type of endometriosis cannot be elucidated.

3. There is a high correlation between the results obtained with RNA-Seq and Nanostring Technology when using FFPE samples.

4. The predictive model could be validated with a higher number of samples but it could not be translated to clinical use due to the low specificity of the model.

5. Phenotypic variables increase the sensitivity and specificity of the classificatory model.

6. There is a possibility that by using a combination of new biomarkers described as being differentially expressed in women with endometriosis together with the phenotypic variables, a new classifier with a high specificity could be developed.

Objective 2:

7. LGR5+ cells co-expressed epithelial markers in the stromal compartment of women with endometriosis.

8. LGR5 does not vary throughout the menstrual cycle in endometriosis, but it shows a non-significant tendency to increase in the secretory phase while it tends to decrease in control group.

9. LGR5 is not a good diagnostic biomarker for endometriosis.

10. LGR5+ cells may have a macrophage-like phenotype in endometriosis.

11. Deep infiltrating endometriosis showed a special subset of LGR5+ cells compared to other types of endometriosis.

12. LGR5+ in eutopic endometrium may have different gene signatures depending on the type or aggressiveness of the disease and may have implications in reproductive outcomes.

13. Mφ1 and Mφ2 have a higher pro-inflammatory phenotype in eutopic endometrium of endometriosis sufferers than in the control group.

14. Mφ2 could be suffering from polarization in women with endometriosis, which would increase their inflammatory phenotype in cases of the disease.

15. The aberrant expression of Mφ in eutopic endometrium may have implications for reproductive outcomes.

Annexes

7. ANNEXES

ANNEX 1. Differentially expressed transcripts in Endometriosis vs Control (Discovery phase objective 1)

Table 34. DET expressed in Endometriosis vs Control (Discovery Study). Highlighted in grey the 4 candidate biomarkers used in the validation study from the list of 182 DET.

Isoform_ID	Gene_ID	log2_FC	p_value	q_value	Significant
NM_000482	APOA4	-Inf	5.0E-05	0.0098	yes
NM_001001958	OR7G3	-Inf	5.0E-05	0.0098	yes
NM_001004490	OR2AG2	-Inf	5.0E-05	0.0098	yes
NM_001004745	OR5T1	-Inf	5.0E-05	0.0098	yes
NM_001005183	OR6C76	-Inf	5.0E-05	0.0098	yes
NM_001005274	OR4A16	-Inf	5.0E-05	0.0098	yes
NM_001099852	PRAMEF20	-Inf	5.0E-05	0.0098	yes
NM_001282544	CCDC182	-Inf	5.0E-05	0.0098	yes
NM_001289974	LACTBL1	-Inf	5.0E-05	0.0098	yes
NM_004066	CETN1	-Inf	5.0E-05	0.0098	yes
NM_014471	SPINK4	-Inf	5.0E-05	0.0098	yes
NM_030787	CFHR5	-Inf	5.0E-05	0.0098	yes
NR_004402_1	NA	-Inf	5.0E-05	0.0098	yes
NR_027402	FAM223B	-Inf	5.0E-05	0.0098	yes
NR_027402_1	NA	-Inf	5.0E-05	0.0098	yes
NR_033658	NA	-Inf	5.0E-05	0.0098	yes
NR_034134	LINC01249	-Inf	5.0E-05	0.0098	yes
NR_040000	WSCD1	-Inf	5.0E-05	0.0098	yes
NR_110792	LINC01478	-Inf	5.0E-05	0.0098	yes
NR_125874	NA	-Inf	5.0E-05	0.0098	yes
NR_126057	LINC02520	-Inf	5.0E-05	0.0098	yes
NM_001004736	OR5K1	Inf	5.0E-05	0.0098	yes
NM_012351	OR10J1	Inf	5.0E-05	0.0098	yes
NM_013936_3	NA	Inf	5.0E-05	0.0098	yes
NR_037808		Inf	5.0E-05	0.0098	yes
NM_052863	SCGB3A1	5.86	5.0E-05	0.0098	yes
NM_032952	MLXIPL	5.83	5.0E-05	0.0098	yes
NM_003226	TFF3	5.49	5.0E-05	0.0098	yes
NM_016190	CRNN	5.48	2.0E-04	0.0306	yes
NM_138805	FAM3D	4.87	5.0E-05	0.0098	yes

Table 34 ***(Continued)***

NM_002458	MUC5B	4.87	5.0E-05	0.0098	yes
NM_000686	AGTR2	4.81	5.0E-05	0.0098	yes
NM_033197	BPIFB1	4.64	5.0E-05	0.0098	yes
NM_001304359	MUC5AC	4.05	5.0E-05	0.0098	yes
NM_001005181	OR56B4	3.88	5.0E-05	0.0098	yes
NR_040117		3.86	5.0E-05	0.0098	yes
NM_001014450	SPRR2F	3.80	3.0E-04	0.0421	yes
NM_001098514	C16orf89	3.72	3.5E-04	0.0480	yes
NM_019060	CRCT1	3.69	5.0E-05	0.0098	yes
NM_001015038	PAGE2B	3.63	1.0E-04	0.0177	yes
NM_005268	GJB5	3.59	5.0E-05	0.0098	yes
NR_045005	OR10V2P	3.55	5.0E-05	0.0098	yes
NM_003535	HIST1H3J	3.37	5.0E-05	0.0098	yes
NR_109756		3.31	5.0E-05	0.0098	yes
NM_005621	S100A12	3.30	2.0E-04	0.0306	yes
NR_125920	NA	3.25	1.0E-04	0.0177	yes
NM_001271560	NA	2.97	1.0E-04	0.0177	yes
NM_130772	S100Z	2.97	5.0E-05	0.0098	yes
NM_002426	MMP12	2.89	1.5E-04	0.0246	yes
NR_120328	NA	2.87	5.0E-05	0.0098	yes
NM_005430	WNT1	2.86	5.0E-05	0.0098	yes
NM_000519	HBD	2.86	1.0E-04	0.0177	yes
NM_054023	SCGB3A2	2.77	3.5E-04	0.0480	yes
NM_198180	QRFP	2.66	1.0E-04	0.0177	yes
NR_126161	NA	2.65	5.0E-05	0.0098	yes
NM_003064	SLPI	2.57	5.0E-05	0.0098	yes
NR_046200	LINC02138	2.55	5.0E-05	0.0098	yes
NM_001013658	PTX4	2.55	1.5E-04	0.0246	yes
NM_152310	ELOVL3	2.54	5.0E-05	0.0098	yes
NM_001039617	ZDHHC19	2.50	5.0E-05	0.0098	yes
NM_016616	NME8	2.43	3.0E-04	0.0421	yes
NR_110538_1	NA	2.39	1.0E-04	0.0177	yes
NM_001005486	OR4K15	2.37	3.5E-04	0.0480	yes
NR_002942	GAS2L1P2	2.35	3.0E-04	0.0421	yes
NM_001145250	SP9	2.34	1.5E-04	0.0246	yes
NM_145659	IL27	2.25	2.0E-04	0.0306	yes
NM_005564	LCN2	1.87	5.0E-05	0.0098	yes
NM_006705	GADD45G	1.76	1.5E-04	0.0246	yes
NM_032211	LOXL4	1.64	2.0E-04	0.0306	yes
NM_000336	SCNN1B	1.64	5.0E-05	0.0098	yes

Table 34 ***(Continued)***

NM_022370	ROBO3	1.56	5.0E-05	0.0098	yes
NM_004352	CBLN1	1.48	3.0E-04	0.0421	yes
NM_001127608	FAM189A2	1.39	5.0E-05	0.0098	yes
NM_001852	COL9A2	1.14	5.0E-05	0.0098	yes
NM_002557	OVGP1	1.12	3.5E-04	0.0480	yes
NM_014819	PJA2	-1.00	5.0E-05	0.0098	yes
NM_033103	RHPN2	-1.02	2.0E-04	0.0306	yes
NM_001423	EMP1	-1.02	3.0E-04	0.0421	yes
NM_002214	ITGB8	-1.04	2.5E-04	0.0369	yes
NM_005345_3	HSPA1A	-1.06	2.0E-04	0.0306	yes
NM_206876	PPP1CB	-1.09	1.5E-04	0.0246	yes
NM_005746	NAMPT	-1.10	1.5E-04	0.0246	yes
NM_007106	UBL3	-1.11	3.0E-04	0.0421	yes
NM_020755	SERINC1	-1.12	5.0E-05	0.0098	yes
NM_138288	SPTSSA	-1.12	5.0E-05	0.0098	yes
NM_144617	HSPB6	-1.14	1.0E-04	0.0177	yes
NM_001690	ATP6V1A	-1.18	5.0E-05	0.0098	yes
NM_002229	JUNB	-1.18	5.0E-05	0.0098	yes
NM_052886	MAL2	-1.20	2.5E-04	0.0369	yes
NM_002166	ID2	-1.22	1.0E-04	0.0177	yes
NM_006136	CAPZA2	-1.30	5.0E-05	0.0098	yes
NM_003407	ZFP36	-1.39	2.0E-04	0.0306	yes
NM_005443	PAPSS1	-1.40	5.0E-05	0.0098	yes
NM_014373	GPR160	-1.43	5.0E-05	0.0098	yes
NM_006952	UPK1B	-1.48	5.0E-05	0.0098	yes
NM_002228	JUN	-1.50	5.0E-05	0.0098	yes
NM_004419	DUSP5	-1.53	2.5E-04	0.0369	yes
NM_006520	DYNLT3	-1.57	3.0E-04	0.0421	yes
NM_174911	FAM84B	-1.60	3.0E-04	0.0421	yes
NR_003679	HAND2-AS1	-1.66	5.0E-05	0.0098	yes
NM_004666	VNN1	-1.68	5.0E-05	0.0098	yes
NM_000715	C4BPA	-1.69	5.0E-05	0.0098	yes
NM_001964	EGR1	-1.78	5.0E-05	0.0098	yes
NM_001554	CYR61	-1.86	5.0E-05	0.0098	yes
NR_028303	LINC00882	-2.18	2.0E-04	0.0306	yes
NR_051960	FALEC	-2.26	1.0E-04	0.0177	yes
NR_109769	LINC01506	-2.29	5.0E-05	0.0098	yes
NM_003514	HIST1H2AM	-2.31	2.0E-04	0.0306	yes
NM_002551	OR3A2	-2.35	1.0E-04	0.0177	yes
NM_052939	FCRL3	-2.38	2.0E-04	0.0306	yes

Table 34 ***(Continued)***

NM_173080	SPRR4	-2.40	5.0E-05	0.0098	yes
NM_003510	HIST1H2AK	-2.49	5.0E-05	0.0098	yes
NR_110104		-2.54	5.0E-05	0.0098	yes
NM_012113	CA14	-2.55	3.0E-04	0.0421	yes
NM_001017361	KHDC3L	-2.58	5.0E-05	0.0098	yes
NM_001446	FABP7	-2.59	1.0E-04	0.0177	yes
NM_001286811	SPEG	-2.60	1.5E-04	0.0246	yes
NM_001443	FABP1	-2.64	5.0E-05	0.0098	yes
NR_026881	NA	-2.68	2.5E-04	0.0369	yes
NR_034088	LINC00885	-2.68	5.0E-05	0.0098	yes
NR_003525	LRRC37A6P	-2.77	2.5E-04	0.0369	yes
NM_024080	TRPM8	-2.78	5.0E-05	0.0098	yes
NM_020299	AKR1B10	-2.79	5.0E-05	0.0098	yes
NM_207373	C10orf99	-2.85	5.0E-05	0.0098	yes
NR_026935	NA	-2.87	5.0E-05	0.0098	yes
NM_012452	TNFRSF13B	-2.88	5.0E-05	0.0098	yes
NM_000705	ATP4B	-2.90	5.0E-05	0.0098	yes
NM_006274	CCL19	-2.95	5.0E-05	0.0098	yes
NR_103791	NA	-2.96	5.0E-05	0.0098	yes
NR_036534	KCNJ2-AS1	-2.97	5.0E-05	0.0098	yes
NR_121625	NA	-2.98	5.0E-05	0.0098	yes
NM_025243	SLC19A3	-2.99	1.0E-04	0.0177	yes
NM_152997	FDCSP	-3.04	1.5E-04	0.0246	yes
NM_004391	CYP8B1	-3.05	1.0E-04	0.0177	yes
NR_034169	NA	-3.07	5.0E-05	0.0098	yes
NR_026800	FAM30A	-3.08	5.0E-05	0.0098	yes
NR_003268	CTAGE10P	-3.11	5.0E-05	0.0098	yes
NM_001159	AOX1	-3.16	5.0E-05	0.0098	yes
NR_103770	NA	-3.16	1.0E-04	0.0177	yes
NM_152349	KRT222	-3.25	5.0E-05	0.0098	yes
NM_001192	TNFRSF17	-3.25	5.0E-05	0.0098	yes
NM_144646	JCHAIN	-3.27	5.0E-05	0.0098	yes
NM_002411	SCGB2A2	-3.30	1.5E-04	0.0246	yes
NM_000371	TTR	-3.31	5.0E-05	0.0098	yes
NM_173483	CYP4F22	-3.39	5.0E-05	0.0098	yes
NM_012108	STAP1	-3.49	5.0E-05	0.0098	yes
NM_005408	CCL13	-3.54	5.0E-05	0.0098	yes
NM_203347	LCN15	-3.62	5.0E-05	0.0098	yes
NM_000250	MPO	-3.65	5.0E-05	0.0098	yes
NM_000111	SLC26A3	-3.73	5.0E-05	0.0098	yes

Table 34 ***(Continued)***

NR_003503	GGT8P	-3.75	5.0E-05	0.0098	yes
NM_021950	MS4A1	-3.88	5.0E-05	0.0098	yes
NR_110569	NA	-3.90	5.0E-05	0.0098	yes
NM_152338	ZG16	-3.93	5.0E-05	0.0098	yes
NM_001025231	KPRP	-4.07	5.0E-05	0.0098	yes
NR_022006	KIAA0087	-4.07	5.0E-05	0.0098	yes
NR_027486	TBC1D3P2	-4.09	5.0E-05	0.0098	yes
NM_017855	ODAM	-4.17	5.0E-05	0.0098	yes
NM_007227	GPR45	-4.20	5.0E-05	0.0098	yes
NR_034136	LINC02171	-4.25	5.0E-05	0.0098	yes
NR_033376	LINC00222	-4.32	5.0E-05	0.0098	yes
NR_033980	LINC01185	-4.37	5.0E-05	0.0098	yes
NM_001442	FABP4	-4.38	5.0E-05	0.0098	yes
NM_017422	CALML5	-4.44	5.0E-05	0.0098	yes
NR_103835	ANO1-AS2	-4.50	5.0E-05	0.0098	yes
NM_001007534	C3orf56	-4.50	5.0E-05	0.0098	yes
NM_001633	AMBP	-4.55	5.0E-05	0.0098	yes
NR_104606	LINC02212	-4.60	5.0E-05	0.0098	yes
NR_103861	LINC02408	-4.67	5.0E-05	0.0098	yes
NR_110138	PCAT5	-4.77	5.0E-05	0.0098	yes
NR_104171	NA	-4.78	5.0E-05	0.0098	yes
NM_017516	RAB39A	-4.99	5.0E-05	0.0098	yes
NM_000040	APOC3	-5.03	5.0E-05	0.0098	yes
NR_110916	LINC01571	-5.04	5.0E-05	0.0098	yes
NR_026834	LINC00691	-5.06	5.0E-05	0.0098	yes
NR_027127	LINC01959	-5.11	5.0E-05	0.0098	yes
NM_000035	ALDOB	-5.30	5.0E-05	0.0098	yes
NM_001480	GALR1	-5.30	5.0E-05	0.0098	yes
NR_110694	LINC01782	-5.42	5.0E-05	0.0098	yes
NR_027084		-6.51	1.5E-04	0.0246	yes
NM_139248	LIPH	-6.78	5.0E-05	0.0098	yes

ANNEX 2. Presence of circulating endometrial cells in peripheral blood

Recently, the identification of circulating cells from a liquid biopsy has been proposed as a new method for the diagnosis of different cancers such as colon, prostate, renal, pancreatic, and lung cancer[60–64]. A liquid biopsy refers to a test done on a sample of blood to look for circulating tumor cells or pieces of free DNA from tumor cells that are present in the blood. Circulating tumor cells are rare cells that are shed into bloodstream from primary or metastatic tumors and have the potential to initiate metastasis in distant tissues or organs[65,66]. The Food and Drug Administration (FDA) has approved the CellSearch® platform for the detection of circulating tumor cells (CTCs), and it is already being used[67–70]. Normally, CTCs devices such as CellSearch® detect epithelial cells using the epithelial marker Epithelial Cell Adhesion Molecule (EpCAM).

For cancer patients, the enumeration of rare circulating tumor cells (CTCs) in peripheral blood is a strong prognostic indicator of the severity of the cancer. In clinics, it is a very useful tool for cancer screening, early detection, and treatment assessment[258]. Although endometriosis is a benign disease, we believe that circulating endometrial cells (CECs) in peripheral blood are present in patients with endometriosis and could be a useful tool for diagnosis or prognosis like in cancer. In addition, studies with primary endometrial cells from ectopic lesions[259,260] and immortalized cell lines[261] have shown similar invasive properties to tumor cells, thus making their entry into blood circulation possible. Hence, this is a field that has potential and can reach greater importance in the near future. Therefore, the objective of this pilot study was to determine the presence of circulating endometrial cells in blood samples from women with and without endometriosis.

Samples were collected at IVI Barcelona S.L. In the study, nine participants were included: four healthy women who were egg donors and five patients with endometriosis. From those five, two were stimulated with follicular stimulating hormone (FSH) and three were not. All participants were premenopausal women and were obtained from IVI Barcelona S.L. Therefore, in total, nine blood samples of 10 ml each were obtained in K2-EDTA tubes that contained the anticoagulant ethylenediamine tetra-acetic acid (EDTA). To avoid any

possible annoyance during the process in patients with endometriosis, the blood was extracted once the venoclisis was performed, which was necessary to proceed with surgery. To discard the possible cutaneous epithelial cells contamination taken by the needle during sampling, the first milliliter obtained was discarded. After blood collection, tubes were immediately inverted ten times and kept at 4ºC until they were processed. Blood samples collected using this process have to be used within a maximum of four hours after drawing. To obtain CECs, ScreenCell®Cyto Technology (ScreenCell®Cyto CY 4FC) was used following the manufacturer's instructions, which are briefly explained below and shown in **Figure 62**.

First, 3ml of blood were transferred into one 15ml sterile canonical tube and mixed with 4ml of dilution buffer. The tube was incubated for eight minutes at room temperature (RT). During this process, the red blood cells were lysated. After the incubation, the blood was filtered. When all the liquid was filtrated, 1.6ml of PBS 1x was added to wash the filter of debris. The filtration process is very fast; normally it is completed within three to four minutes. After this, a rod that had to be pushed down by holding both sides and letting the circular filter fall on a piece of Whatman® paper. Once the circular filter was dry, it was placed in a p24 well and fixed with formaldehyde 4% for 10 minutes. After this period, the filter was washed three times for five minutes with PBS 1x. Then, the IF of CECs could be performed. IF of CD10 and CK was performed on the isolated CECs for the staining of the stromal and the epithelial compartment respectively.

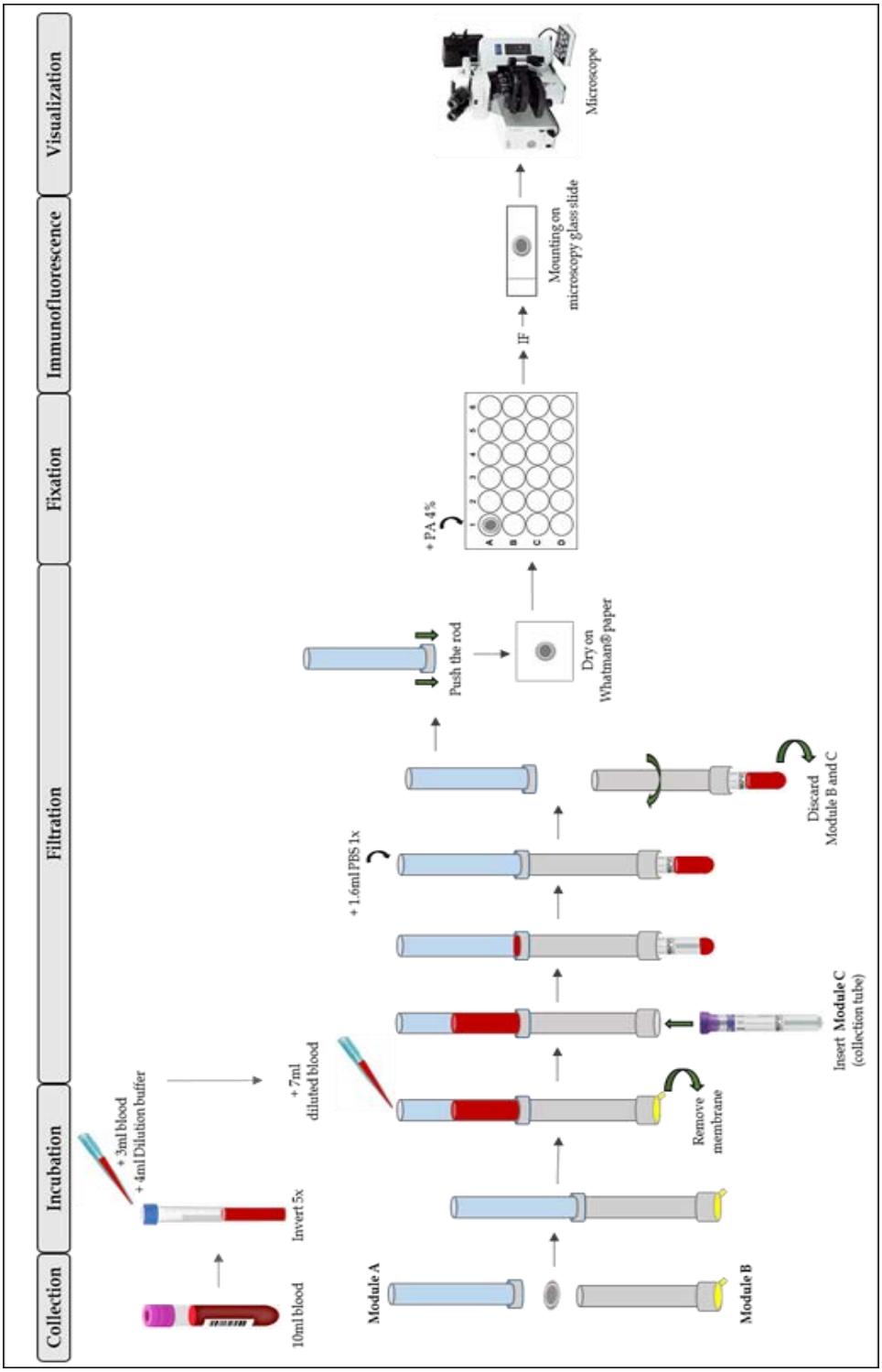


Figure 62. Workflow of CECs isolation. Isolation of CECs by using ScreenCell® Technology.

Once performed the IF and microscope analysis, CECs were found in four of the five studied patients with endometriosis. As expected, they were only observed in patients with endometriosis and not in the four tested healthy donors. The counts of total cells present in the total area of 7mm filter are listed in **Table 35**.

Patient ID	Menstrual Cycle Phase	Type of endometriosis	Treatment	CECs		Hormone Stimulation (FSH)
				CK	CD10	
N37	secretory	Endometrioma	supression	0	17	Non-stimulated
N38	proliferative	DIE	supression	0	0	
N39	menstrual	DIE	surgery	0	3	
N34	ovulatory	Endometrioma	surgery	0	11	Stimulated
N35	proliferative	DIE	surgery	0	5	
N28	ovulatory	Control	-	0	0	Stimulated
N29	ovulatory	Control	-	0	0	
N30	ovulatory	Control	-	0	0	
N31	ovulatory	Control	-	0	0	

Table 35. Patient's characteristics. The table shows the patient ID, the menstrual phase, the type of disease, the treatment followed, hormonal treatment, and the CECs counted in each patient both for the case of CK and CD10.

Surprisingly, only CECs expressing the CD10 marker were observed, and no CK positive cells were found either in the endometriosis group or in the healthy group. No cells expressing CD10 had co-localization with CK markers, indicating that the secondary antibodies did not overlap within themselves. An example of the findings of CD10 in each of the four patients with endometriosis is shown in **Figure 63**.

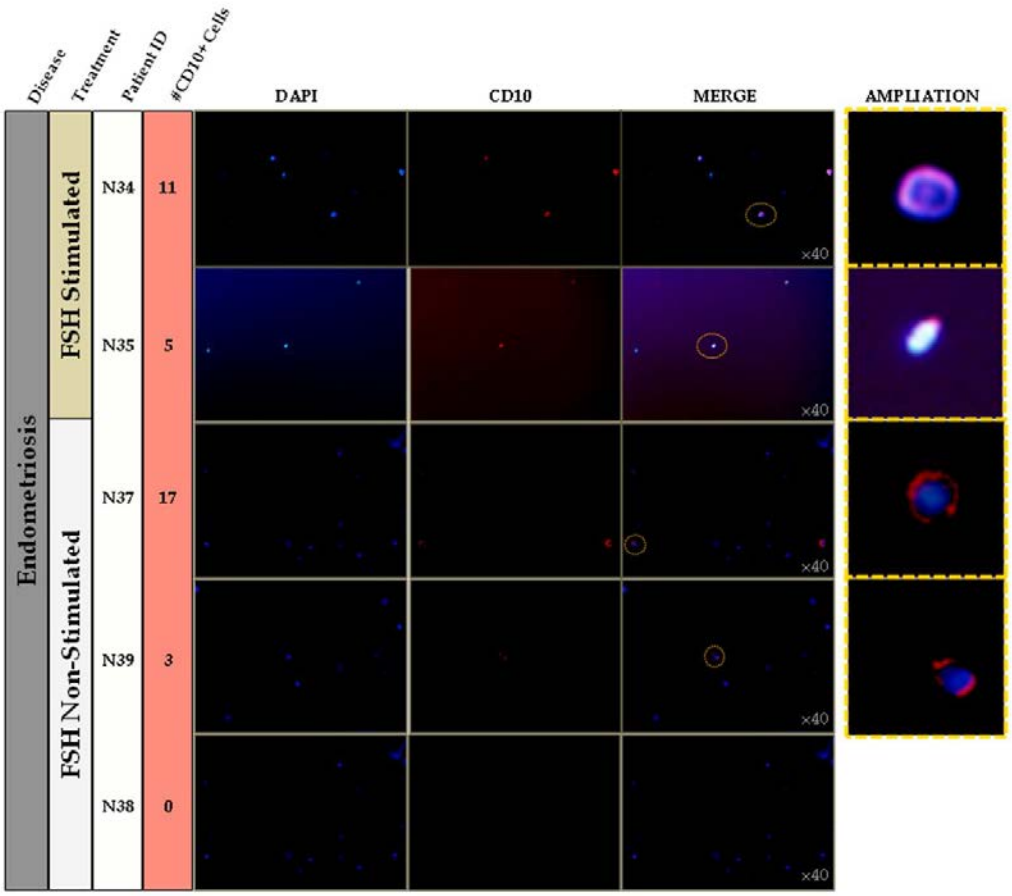


Figure 63. **IF of CECs.** This figure shows the CECs found in endometriosis patients stained with CD10 (red) and DAPI (blue). In the red column, we can observe the number of CECs found in each patient. In yellow circles we can observe the expression of CD10, and in the right panel, there is an amplification of one cell.

To our knowledge, only two groups have looked for circulating cells in peripheral blood in women with endometriosis[262,263]. One of them used MetaCell® Technology, where cells are enriched by filtration. Once they obtained the CECs, they cultured them and performed IHC with CK, vimentin, and CD10 markers to prove the endometrial origin[262]. These groups demonstrated for the first time the presence of CECs in the peripheral blood of women with endometriosis. They found CECs in 4/17 cases of endometriosis. The second group used microfluidic chips. They found that 89.5% of the endometriotic patients presented

CECs while only 15% of the healthy control did[263]. They identified the CECs by the positive expression in the IF of vimentin/cytokeratin and estrogen/progesterone receptor.

In our study, we demonstrated for the first time that the ScreenCell® CYTO is a valid device for isolating CECs from peripheral blood in women with endometriosis, even with the very small amount of CECs that are normally present in a liquid biopsy. The method used is size-based. Cells were trapped in a mesh and then cell culture, after which IF and/or IHC was performed. An advantage of this method is that cells can be stained and their origin can be determined afterwards. Other methods such as CellSearch® require the previous staining of the samples, usually with the epithelial marker EpCAM and CD45, to discard the leukocytes. The previous processing of the sample could modify the phenotypic characteristics of the cells, thus information could be lost. In addition, with these technologies, cells are stained with EpCAM. This could be an issue because not all cancers or diseases produce metastasis with an epithelial origin. Moreover, some cells can suffer from epithelial mesenchymal transition process (EMT), losing the epithelial markers in blood circulation and making their detection impossible using this strategy[264,265]. Hence, CECs capture technologies that are antibody-independent and based solely on the physical properties of cells to detect a wide-range of CECs are increasingly being developed.

Surprisingly, we only could determine the presence of CD10 positive cells (stromal cells) in the peripheral blood of women with endometriosis, but we did not find epithelial cells. Supporting our results, Bobek et al. found the presence of stromal cells but not epithelial cells in peripheral blood using MetaCell® filters[262]. Furthermore, Chen et al. isolated CECs using microfluidics chips. They found that patients with endometriosis presented CECs independently of menstrual cycle phases[263], which is very useful in clinics as patients can be tested at any time of the cycle. However, they did not mention whether the cells were epithelial or stromal cells. They used primary antibodies for both epithelial and stromal markers together and the same secondary antibody, making it impossible to discriminate between the two markers. We hypothesized that the invasiveness of the endometriosis can be attributed to endometrial stromal cells, which create support for endometrial glandular

cells' growth. Alternately, it could be that epithelial cells suffer EMT when they spread to the blood circulation and that the antigen may be lost during this process[258].

Interestingly, one patient who had deep infiltrating endometriosis (DIE) and was under suppression treatment (GnRH) for two months did not show the presence of any CEC. This could be an indicator of the disease activity, but this is only a speculation as we did not have enough samples to test this theory. Moreover, no differences in the CECs count during the different phases of the menstrual cycle were observed, in accordance with Chen et al's study[263].

A very interesting test that we would like to perform is to determine the CECs content in the same women tested before and after the surgery or before and after the suppression treatment. This information would be very valuable in determining if this user friendly method could be useful as a diagnostic and even as a prognostic tool. Moreover, if a decrease of CECs in women receiving suppression treatment or after surgery were observed, it could also be a perfect tool for determining the activity of the disease and the effect of certain pharmacological therapies. Indeed, we also want to study the presence of CECs in different types of endometriosis and see if we could identify if the CECs counts could be related to the aggressiveness of the disease as well. In a future perspective, we would like to increase the sample size and to test more patients, and establish a cut-off of the number of cells present in blood from healthy women.

Although CECs are promising biomarkers for endometriosis with great potential for a noninvasive diagnostic assay, further research is needed, and we believe that size enrichment is a very rudimentary technique. Single-cell sequencing of the captured cells may also be helpful in defining more specific biomarkers of CECs.

ANNEX 3. Main achievements of this thesis

❖ Publications arising from this thesis

1. "Aberrant expression of epithelial leucine-rich repeat containing G protein–coupled receptor 5–positive cells in the eutopic endometrium in endometriosis and implications in deep-infiltrating endometriosis". Vallvé-Juanico J, Suárez-Salvador E, Castellví J, Ballesteros A, Taylor HS, Gil-Moreno A, Santamaria X. ***Fertil. Steril.*** **108, 858–867.e2 (2017).**

2. "The immune environment in endometrium of women with endometriosis". Vallvé-Juanico J, Houshdaran S and Giudice L. **Review in preparation to be submitted in Nature Immunology reviews.**

3. "LGR5 does not vary throughout the menstrual cycle in endometriosis eutopic endometrium". Vallvé-Juanico J, Barón C, Suárez-Salvador E, Castellví J, Ballesteros A, Gil-Moreno A, Santamaria X. **Article to be submitted in Fertility and Sterility.**

❖ Publications in collaboration

1. "Progesterone and testosterone-derived progestins down-regulate CCL2 in endometrial stromal fibroblasts in vitro". Houshdaran S, Chen JC, Vallvé-Juanico J, Balayan S, Irwin JC, Giudice LC. **Article to be submitted in Endocrinology.**

2. "Advances in endometrial cancer protein biomarkers for use in the clinic". Elena Martinez-Garcia, Carlos Lopez-Gil, Irene Campoy, Julia Vallve, Eva Coll, Silvia Cabrera, Santiago Ramon y Cajal, Xavier Matias-Guiu, Jan van Oostrum, Jaume Reventos, Antonio Gil-Moreno, Eva Colas. **Expert Review of Proteomics, 2017.**

3. "Leucine-rich repeat–containing G-protein–coupled receptor 5–positive cells in the endometrial stem cell niche". Cervelló I, Gil-Sanchis C, Santamaría X, Faus A, Vallvé-Juanico J, Díaz-Gimeno P, et al. **Fertil Steril, 2017.**

❖ Conference presentations arising from this thesis

1. **"Predictive Model for Endometriosis Diagnosis Based in Uterine Aspirates".** Vallvé-Juanico J, Suárez-Salvador E., Castellví J, Hugh S Taylor, Gil-Moreno A and Santamaria X. **Poster presentation**. 64th Annual Scientific Meeting SRI, Orlando, Florida, USA. 14-18th March, 2017. (International)

2. **"Endometrial genetic expression profile for the diagnosis of endometriosis".** Vallvé Juanico J, Suárez-Salvador E., Castellví J, Christian Barón, Gil-Moreno A and Santamaria X. **Poster presentation**. 10th Scientific Conference VHIR Vall d'Hebron Research Institute (VHIR), Barcelona, Spain. 13-16th December, 2016.

3. **"Understanding the role of LGR5 in endometriosis".** Vallvé-Juanico J., Suárez-Fernández E., Castellví J., Barón C., Gil-Moreno A. and Santamaria X. **Poster presentation.** VI Congreso de Jóvenes Investigadores de la RTICC. Salamanca University, Salamanca, Spain. 23rd September, 2016.

4. "**New approach for the diagnosis of endometriosis".** Vallvé-Juanico J. Suárez-Fernández E., Castellví J, Gil-Moreno A and Santamaria X. **Oral Communication.** 2nd SEUD Congress. Barcelona, Spain. 12-14th May, 2016. (International)

5. "**Understanding the role of LGR5 in endometriosis".** Vallvé-Juanico J., Suárez-Fernández E., Castellví J., Barón C., Gil-Moreno A. and Santamaria X. **Poster presentation.** 9th Scientific Conference VHIR. Fundació Hospital Universitari Vall d'Hebron Institut de Recerca (VHIR), Barcelona, Spain. 10-11th September, 2015.

6. **"Understanding LGR5 as a prognostic and diagnostic marker for endometriosis".** Vallvé-Juanico J. Suárez Fernández E., Castellví J, Gil-Moreno A and Santamaria X. **Oral Communication.** 8th Scientific Conference VHIR. Vall d'Hebron Research Institute (VHIR), Barcelona, Spain. 11-12th December, 2014.

❖ Grants

1. Industrial Doctorate AGAUR (Agency for Administration of University and Research Grants, Catalonia, Spain)

2. SRI International Grant (Society of Reproductive Investigation, USA)

❖ Internships

During a total of 8 months, I worked in in the laboratory of Dr. Linda Giudice located in the Center of Reproductive Sciences (CRS) in the Department of Obstetrics, Gynecology and Reproductive Sciences at University of California San Francisco (UCSF). Here I performed the experiments with immune populations in eutopic endometrium of women with and without endometriosis and is where I will follow my next steps of my scientific career as a postdoctoral researcher.

References

8. REFERENCES

1. Giudice, L. C. Clinical Practice: Endometriosis. *N. Engl. J. Med.* **362,** 2389–2398 (2010).
2. Becker, C. M. *et al.* World Endometriosis Research Foundation Endometriosis Phenome and Biobanking Harmonisation Project: I. Surgical phenotype data collection in endometriosis research. *Fertil. Steril.* **102,** 1213–1222 (2014).
3. Hawkins, S. & Matzuk, M. Author Manuscript Menstrual Cycle : Basic Biology. *Ann. N. Y. Acad. Sci.* **1135,** 10–18 (2008).
4. Nisolle, M. & Donnez, J. Peritoneal endometriosis, ovarian endometriosis, and adenomyotic nodules of the rectovaginal septum are three different entities. *Fertil. Steril.* **68,** 585–596 (1997).
5. Evaluación, A. De & Sanitarias, D. T. Revisión de la evidencia científica. (2007).
6. Bergqvist, A., Ljungberg, O. & Myhre, E. Human endometrium and endometriotic tissue obtained simultaneously: A comparative histological study. *Int. J. Gynecol. Pathol.* **3,** 135–145 (1984).
7. Sourial, S., Tempest, N. & Hapangama, D. K. Theories on pathogenesis of endometriosis. **2014,** (2014).
8. Sampson, J. A. Metastatic or embolic endometriosis, due to the menstrual dissemination of endometrial tissue into the venous circulation. **III,** 81–82 (1927).
9. Society, A. G. Peritoneal Endometriosis, Due to the Menstrual Dissemination of Endometrial Tissue into the Peritoneal Cavity. *Am. J. Obstet. Gynecol.* **15,** 101–110 (1928).
10. Bianco, B. *et al.* The possible role of genetic variants in autoimmune-related genes in the development of endometriosis. *Hum. Immunol.* **73,** 306–315 (2012).
11. Ahn, S. H. *et al.* Pathophysiology and Immune Dysfungtion in Endometriosis. *Biomed Res. Int.* **2015,** 1–12 (2014).
12. Ferguson, B. R., Bennington, J. L. & Haber, S. L. Histochemistry of mucosubstances and histology of mixed müllerian pelvic lymph node glandular inclusions. Evidence for histogenesis by müllerian metaplasia of coelomic epithelium. *Obstet. Gynecol.* **33,** 617–25 (1969).
13. Sasson, I. & Taylor, H. Stem cells and the pathogenesis of endometriosis. *Ann. N. Y. Acad. Sci.* **1127,** 106–115 (2008).
14. Sasson, I. E. & Taylor, H. S. Stem cells and the pathogenesis of endometriosis. in *Annals of the New York Academy of Sciences* **1127,** 106–115 (2008).
15. Burney, R. O. & Giudice, L. C. Pathogenesis and pathophysiology of endometriosis. *Fertility and Sterility* **98,** 511–519 (2012).
16. Berbic, M. *et al.* Macrophage expression in endometrium of women with and without endometriosis. *Hum. Reprod.* **24,** 325–332 (2009).
17. Jones, R. K., Searle, R. F. & Bulmer, J. N. Apoptosis and bcl-2 expression in normal human

endometrium, endometriosis and adenomyosis. *Hum. Reprod.* **13,** 3496–3502 (1998).

18. Taylor, R. N., Lebovic, D. I. & Mueller, M. D. Angiogenic factors in endometriosis. *Ann. N. Y. Acad. Sci.* **955,** 89-100, 396–406 (2002).

19. Jeziorska, M., Salamonsen, L. A. & Woolley, D. E. Mast cell and eosinophil distribution and activation in human endometrium throughout the menstrual cycle. *Biol. Reprod.* **53,** 312–20 (1995).

20. Huhtinen, K. *et al.* Endometrial and endometriotic concentrations of estrone and estradiol are determined by local metabolism rather than circulating levels. *J. Clin. Endocrinol. Metab.* **97,** 4228–4235 (2012).

21. Viganò, P., Parazzini, F., Somigliana, E. & Vercellini, P. Endometriosis: Epidemiology and aetiological factors. *Best Practice and Research: Clinical Obstetrics and Gynaecology* **18,** 177–200 (2004).

22. Sapkota, Y. *et al.* Meta-analysis identifies five novel loci associated with endometriosis highlighting key genes involved in hormone metabolism. *Nat. Commun.* **8,** (2017).

23. Mason, A. Incidence of endometriosis by study population and diagnostic method: the ENDO Study. **19,** 389–399 (2009).

24. Vessey, M. P., Villard-Mackintosh, L. & Painter, R. Epidemiology of endometriosis in women attending family planning clinics. *Bmj* **306,** 182–184 (1993).

25. Barbieri, R. L. Reproductive history and endometriosis among premenopausal women. *Obstet. Gynecol.* **104,** 965–974 (2004).

26. McLeod, B. S. & Retzloff, M. G. Epidemiology of endometriosis: An assessment of risk factors. *Clinical Obstetrics and Gynecology* **53,** 389–396 (2010).

27. Liu, Y. & Zhang, W. Association between body mass index and endometriosis risk: a meta-analysis. *Oncotarget* **8,** 46928–46936 (2017).

28. Nnoaham, K. E. *et al.* Impact of endometriosis on quality of life and work productivity : a multicenter study across ten countries. *Fertil. Steril.* **96,** 366–373 (2011).

29. De Felip, E. *et al.* Dioxin-like compounds and endometriosis: A study on Italian and Belgian women of reproductive age. *Toxicol. Lett.* **150,** 203–209 (2004).

30. Tempest, N., Baker, A. M., Wright, N. A. & Hapangama, D. K. Does human endometrial LGR5 gene expression suggest the existence of another hormonally regulated epithelial stem cell niche? *Hum. Reprod.* **33,** 1–11 (2018).

31. Vila-Vives, J. M., Martínez-Conejero, J. A. & Pellicer, A. Effect of adenomyosis on implantation. *Reproductive BioMedicine Online* **24,** 584 (2012).

32. Petraglia, F., Serour, G. I. & Chapron, C. The changing prevalence of infertility. in *International Journal of Gynecology and Obstetrics* **123,** S4–S8 (2013).

33. Tanbo, T. & Fedorcsak, P. Endometriosis-associated infertility: aspects of pathophysiological mechanisms and treatment options. *Acta Obstet. Gynecol. Scand.* **96,** 659–667 (2017).

34. Khan, K. N. *et al.* Differential macrophage infiltration in early and advanced endometriosis and adjacent peritoneum. *Fertil. Steril.* **81,** 652–661 (2004).

35. Borrelli, G. M. *et al.* Chemokines in the pathogenesis of endometriosis and infertility. *J. Reprod. Immunol.* **98,** 1–9 (2013).

36. Pellicer, A., Oliveira, N., Ruiz, A., Remohí, J. & Simón, C. Exploring the mechanism(s) of endometriosis-related infertility: an analysis of embryo development and implantation in assisted reproduction. *Hum. Reprod.* **10 Suppl 2,** 91–7 (1995).

37. Bulun, S. E. *et al.* Progesterone resistance in endometriosis: link to failure to metabolize estradiol. *Mol. Cell. Endocrinol.* **248,** 94–103 (2006).

38. Leyendecker, G., Kunz, G., Wildt, L., Beil, D. & Deininger, H. Uterine hyperperistalsis and dysperistalsis as dysfunctions of the mechanism of rapid sperm transport in patients with endometriosis and infertility. *Hum. Reprod.* **11,** 1542–51 (1996).

39. Society, E. & Reproduction, H. Pain and endometriosis. *Pain* **132,** 22–25 (2012).

40. Stovner, L. J., Aegidius, K. & Linde, M. Endometriosis and headache. *Curr. Pain Headache Rep.* **15,** 415–419 (2011).

41. Wykes, C. B., Clark, T. J. & Khan, K. S. Accuracy of laparoscopy in the diagnosis of endometriosis: A systematic quantitative review. *BJOG An Int. J. Obstet. Gynaecol.* **111,** 1204–1212 (2004).

42. Ballweg, M. Lou. Impact of endometriosis on women's health: Comparative historical data show that the earlier the onset, the more severe the disease. *Best Pract. Res. Clin. Obstet. Gynaecol.* **18,** 201–218 (2004).

43. Moore, J. *et al.* A systematic review of the accuracy of ultrasound in the diagnosis of endometriosis. *Ultrasound Obstet. Gynecol.* **20,** 630–634 (2002).

44. Eskenazi, B. *et al.* Validation study of nonsurgical diagnosis of endometriosis. *Fertil. Steril.* **76,** 929–935 (2001).

45. Chapron, C. *et al.* Anatomical distribution of deeply infiltrating endometriosis: Surgical implications and proposition for a classification. *Hum. Reprod.* **18,** 157–161 (2003).

46. Duffy, J. *et al.* Laparoscopic surgery for endometriosis (Review). *Cochrane Database Syst. Rev.* 79 (2014). doi:10.1002/14651858.CD011031.pub2.www.cochranelibrary.com

47. Theron, E., Shaw, G. B. & Action, S. Revised American Society for Reproductive Medicine classification of endometriosis: 1996. *Soc. Work (South Africa)* **43,** 283–290 (2007).

48. Johnson, N. P. *et al.* World endometriosis society consensus on the classification of endometriosis. *Hum. Reprod.* **32,** 315–324 (2017).

49. Agarwal, S. K. Impact of six months of GnRH agonist therapy for endometriosis. Is there an age-related effect on bone mineral density? *J Reprod.Med.* **47,** 530–534 (2002).

50. Strimbu, K. & Tavel, J. a. What are Biomarkers? *Curr Opin HIV AIDS* **5,** 463–466 (2011).

51. Tobergte, D. R. & Curtis, S. *The GALE Encyclopedia of Science*. *Journal of Chemical Information and Modeling* **53,** (Gale Research, 2013).

52. Sinaii, N., Cleary, S. D., Ballweg, M. L., Nieman, L. K. & Stratton, P. High rates of

autoimmune and endocrine disorders, fibromyalgia, chronic fatigue syndrome and atopic diseases among women with endometriosis: a survey analysis. *Hum. Reprod.* **17,** 2715–24 (2002).

53. Ek, M. *et al.* Gastrointestinal symptoms among endometriosis patients-A case-cohort study. *BMC Womens. Health* **15,** 1–10 (2015).

54. Parazzini, F., Esposito, G., Tozzi, L., Noli, S. & Bianchi, S. Epidemiology of endometriosis and its comorbidities. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **209,** 3–7 (2017).

55. May, K. E. *et al.* Peripheral biomarkers of endometriosis: A systematic review. *Hum. Reprod. Update* **16,** 651–674 (2010).

56. Gueye, N. A. *et al.* in *Biomarkers for endometriosis in saliva, urine, and peritoneal fluid* 141–163 (Springer International Publishing, 2017). doi:10.1007/978-3-319-59856-7_8

57. Vitonis, A. F. *et al.* Title : WERF Endometriosis Phenome and Biobanking Harmonization Project (EPHect): II . Clinical and covariate phenotype data collection in endometriosis research . *Fertil. Steril.* **102,** 1223–32 (2014).

58. Liu, E. *et al.* Urinary biomarkers for the non-invasive diagnosis of endometriosis. *Cochrane database Syst. Rev.* **12,** CD012019 (2015).

59. Nisenblat, V. *et al.* Blood biomarkers for the non-invasive diagnosis of endometriosis (Review). (2016). doi:10.1002/14651858.CD012179.www.cochranelibrary.com

60. Desitter, I. *et al.* A new device for rapid isolation by size and characterization of rare circulating tumor cells. *Anticancer Res.* **31,** 427–441 (2011).

61. El-Heliebi, A. *et al.* Are morphological criteria sufficient for the identification of circulating tumor cells in renal cancer? *J. Transl. Med.* **11,** 1–17 (2013).

62. Kulemann, B. *et al.* Circulating tumor cells found in patients with localized and advanced pancreatic cancer. *Pancreas* **44,** 547–550 (2015).

63. Chen, C. *et al.* HHS Public Access. **73,** 813–826 (2016).

64. Mascalchi, M. *et al.* Prevalence and number of circulating tumour cells and microemboli at diagnosis of advanced NSCLC. *J. Cancer Res. Clin. Oncol.* **142,** 195–200 (2016).

65. Allard, W. J. *et al.* Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients With Nonmalignant Diseases Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients With Non. *Clin. Cancer Res.* **10,** 6897–6904 (2005).

66. Riethdorf, S., Wikman, H. & Pantel, K. Review: Biological relevance of disseminated tumor cells in cancer patients. *Int. J. Cancer* **123,** 1991–2006 (2008).

67. Cristofanilli, M. *et al.* Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *… Engl. J. …* 781–791 (2004).

68. Bidard, F.-C. *et al.* Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data. *Lancet Oncol.* **15,** 406–414 (2014).

69. Budd, G. T. *et al.* Circulating tumor cells versus imaging - Predicting overall survival in metastatic breast cancer. *Clin. Cancer Res.* **12,** 6403–6409 (2006).

70. Hayes, D. F. *et al.* Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival. *Clin. Cancer Res.* **12,** 4218–4224 (2006).

71. Garcia-Velasco, J. A. *et al.* Is endometrial receptivity transcriptomics affected in women with endometriosis? A pilot study. *Reprod. Biomed. Online* **31,** 647–654 (2015).

72. Ahn, S. H., Singh, V. & Tayade, C. Biomarkers in endometriosis: challenges and opportunities. *Fertility and Sterility* **107,** 523–532 (2017).

73. Talbi, S. *et al.* Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. *Endocrinology* **147,** 1097–121 (2006).

74. Maclaran, K., Agarwal, N. & Odejinmi, F. Co-Existence of Uterine Myomas and Endometriosis in Women Undergoing Laparoscopic Myomectomy: Risk Factors and Surgical Implications. *J. Minim. Invasive Gynecol.* **21,** 1086–1090 (2014).

75. Uimari, O., Järvelä, I. & Ryynänen, M. Do symptomatic endometriosis and uterine fibroids appear together? *J. Hum. Reprod. Sci.* **4,** 34 (2011).

76. Shen, L. *et al.* High prevalence of endometrial polyps in endometriosis-associated infertility. *Fertil. Steril.* **95,** 2722–2724.e1 (2011).

77. Zheng, Q.-M. *et al.* Risk of endometrial polyps in women with endometriosis: a meta-analysis. *Reprod. Biol. Endocrinol.* **13,** 103 (2015).

78. Hwang, K. R., Choi, Y. M., Kim, J. J., Jeon, H. W. & Hong, M. A. DNA Microarray Analysis of Gene Expression in Eutopic Endometrium from Patients with Endometriosis. *Adv. Reprod. Sci.* **5,** 75–96 (2017).

79. Zhao, L. *et al.* Identification of global transcriptome abnormalities and potential biomarkers in eutopic endometria of women with endometriosis: A preliminary study. *Biomed. Reports* **6,** 654–662 (2017).

80. Vodolazkaia, A. *et al.* Evaluation of a panel of 28 biomarkers for the non-invasive diagnosis of endometriosis. *Hum. Reprod.* **27,** 2698–2711 (2012).

81. Manuscript, A. & Structures, T. Proteomic identification of neurotrophins in the eutopic endometrium of women with endometriosis. **6,** 247–253 (2009).

82. Fowler, P. A. *et al.* An investigation of the effects of endometriosis on the proteome of human eutopic endometrium: A heterogeneous tissue with a complex disease. *Proteomics* **7,** 130–142 (2007).

83. Stephens, A. N. *et al.* Post-translational modifications and protein-specific isoforms in endometriosis revealed by 2D DIGE. *J. Proteome Res.* **9,** 2438–2449 (2010).

84. Houshdaran, S. *et al.* Aberrant Endometrial DNA Methylome and Associated Gene Expression in Endometriosis. *Biol. Reprod.* (2016). doi:10.1095/biolreprod.116.140434

85. Kyama, C. M. *et al.* Evaluation of endometrial biomarkers for semi-invasive diagnosis of endometriosis. *Fertil. Steril.* **95,** 1338–1343.e3 (2011).

86. Kyama, C. M. *et al.* ProteinChip** ProteinChip (Ciphergen Biosystems, Inc., Fremont, CA). technology is a useful method in the pathogenesis and diagnosis of endometriosis: a preliminary study. *Fertil. Steril.* **86,** 203–209 (2006).

87. Fassbender, A. *et al.* TRIzol treatment of secretory phase endometrium allows combined proteomic and mRNA microarray analysis of the same sample in women with and without endometriosis. *Reprod. Biol. Endocrinol.* **8,** 123 (2010).

88. Gupta, D. *et al.* Endometrial biomarkers for the non-invasive diagnosis of endometriosis (Review) Endometrial biomarkers for the non-invasive diagnosis of endometriosis. (2016). doi:10.1002/14651858.CD012165.Copyright

89. Nisenblat, V. *et al.* Combination of the non-invasive tests for the diagnosis of endometriosis (Review). *Cochrane Database Syst. Rev.* (2016). doi:10.1002/14651858.CD012281.www.cochranelibrary.com

90. Li, T., He, H., Liu, R., Wang, S.-X. & Pu, D.-M. Isolation and identification of epithelial and stromal stem cells from eutopic endometrium of women with endometriosis. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **178,** 89–94 (2014).

91. Figueira, P. G. M., Abr??o, M. S., Krikun, G. & Taylor, H. Stem cells in endometrium and their role in the pathogenesis of endometriosis. *Ann. N. Y. Acad. Sci.* **1221,** 10–17 (2011).

92. Gil-Sanchis, C. *et al.* Leucine-rich repeat-containing G-protein-coupled receptor 5 (Lgr5) as a putative human endometrial stem cell marker. *Mol. Hum. Reprod.* **19,** 407–414 (2013).

93. Hufnagel, D., Li, F., Cosar, E., Krikun, G. & Taylor, H. S. The Role of Stem Cells in the Etiology and Pathophysiology of Endometriosis. doi:10.1055/s-0035-1564609

94. Barker, N., Es, J. H. Van & Jaks, V. Very Long-term Self-renewal of Small Intestine , Colon , and Hair Follicles from Cycling Lgr5 + ve Stem Cells Very Long-term Self-renewal of Small Intestine , Colon , and Hair Follicles from Cycling Lgr5 + ve Stem Cells. **LXXIII,** 351–356 (2008).

95. Haegebarth, A. & Clevers, H. Wnt signaling, lgr5, and stem cells in the intestine and skin. *Am. J. Pathol.* **174,** 715–721 (2009).

96. Barker, N. *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* **449,** 1003–1007 (2007).

97. Kumar, K. K., Burgess, A. W. & Gulbis, J. M. Structure and function of LGR5: An enigmatic G-protein coupled receptor marking stem cells. *Protein Sci.* **23,** 551–565 (2014).

98. Carmon, K. S., Gong, X., Lin, Q., Thomas, A. & Liu, Q. R-spondins function as ligands of the orphan receptors LGR4 and LGR5 to regulate Wnt/ -catenin signaling. *Proc. Natl. Acad. Sci.* **108,** 11452–11457 (2011).

99. Sun, X., Jackson, L., Dey, S. K. & Daikoku, T. In pursuit of leucine-rich repeat-containing G protein-coupled receptor-5 regulation and function in the uterus. *Endocrinology* **150,** 5065–5073 (2009).

100. Krusche, C. A., Kroll, T., Beier, H. M. & Classen-Linke, I. Expression of leucine-rich repeat-containing G-protein-coupled receptors in the human cyclic endometrium. *Fertil. Steril.* **87,** 1428–1437 (2007).

101. Cervelló, I. *et al.* Leucine-rich repeat–containing G-protein–coupled receptor 5–positive cells in the endometrial stem cell niche. *Fertil. Steril.* (2016). doi:10.1016/j.fertnstert.2016.10.021

102. Vallvé-Juanico, J. *et al.* Aberrant expression of epithelial leucine-rich repeat containing G protein–coupled receptor 5–positive cells in the eutopic endometrium in endometriosis and implications in deep-infiltrating endometriosis. *Fertil. Steril.* **108,** 858–867.e2 (2017).

103. Cervelló, I. *et al.* Human endometrial side population cells exhibit genotypic, phenotypic and functional features of somatic stem cells. *PLoS One* **5,** (2010).

104. Mantovani, A. *et al.* The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* **25,** 677–686 (2004).

105. Kim, M. Y. *et al.* Tumor Self-Seeding by Circulating Cancer Cells. *Cell* **139,** 1315–1326 (2009).

106. Melo, F. de S. e *et al.* A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. *Nature* **543,** 676–680 (2017).

107. Shimokawa, M. *et al.* Visualization and targeting of LGR5+ human colon cancer stem cells. *Nature* (2017). doi:10.1038/nature22081

108. Santamaria, X., Massasa, E. E. & Taylor, H. S. Migration of cells from experimental endometriosis to the uterine endometrium. *Endocrinology* **153,** 5566–5574 (2012).

109. Jensen, A. L. *et al.* A Subset of Human Uterine Endometrial Macrophages is Alternatively Activated. *Am. J. Reprod. Immunol.* **68,** 374–386 (2012).

110. Givan, A. L. *et al.* Flow cytometric analysis of leukocytes in the human female reproductive tract: comparison of fallopian tube, uterus, cervix, and vagina. *Am. J. Reprod. Immunol.* **38,** 350–359 (1997).

111. Wira, C. R., Fahey, J. V., Sentman, C. L., Pioli, P. A. & Shen, L. Innate and adaptive immunity in female genital tract: Cellular responses and interactions. *Immunol. Rev.* **206,** 306–335 (2005).

112. Starkey, P. M., Clover, L. M. & Rees, M. C. Variation during the menstrual cycle of immune cell populations in human endometrium. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **39,** 203–7 (1991).

113. Salamonsen, L. A., Zhang, J. & Brasted, M. Leukocyte networks and human endometrial remodelling. *J. Reprod. Immunol.* **57,** 95–108 (2002).

114. Ning, F., Liu, H. & Lash, G. E. The Role of Decidual Macrophages During Normal and Pathological Pregnancy. *Am. J. Reprod. Immunol.* **75,** 298–309 (2016).

115. Deloia, J. A., Ph, D., Stewart-akers, A. M. & Brekosky, J. Effects of exogenous estrogen on uterine leukocytes recruitment. **77,** (2002).

116. Wood, G. W. macrophage. (1989).

117. Bonatz, G. *et al.* Macrophage- and lymfocyte-subtypes in the endometrium during different phases of the ovarian cycle. *Int. J. Gynaecol. Obstet.* **37,** 29–36 (1992).

118. King, A. Uterine leukocytes and decidualization. *Human Reproduction Update* **6,** 28–36 (2000).

119. Gordon, S. & Taylor, P. R. Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* **5,** 953–964 (2005).

120. Mantovani, A., Sica, A. & Locati, M. Macrophage polarization comes of age. *Immunity* **23,** 344–346 (2005).

121. Cominelli, A. *et al.* Matrix metalloproteinase-27 is expressed in CD163+/CD206+M2 macrophages in the cycling human endometrium and in superficial endometriotic lesions. *Mol. Hum. Reprod.* **20,** 767–775 (2014).

122. Ding, J., Shen, J., Rana, N., Fernandez, B. B. & Dmowski, W. P. Relationship between apoptosis and the number of macrophages in eutopic endometrium from women with and without endometriosis. *Fertil. Steril.* **78,** 830–835 (2002).

123. Takebayashi, A. *et al.* Subpopulations of Macrophages within Eutopic Endometrium of Endometriosis Patients. *Am. J. Reprod. Immunol.* **73,** 221–231 (2015).

124. Khan, K. N. *et al.* Escherichia coli contamination of menstrual blood and effect of bacterial endotoxin on endometriosis. *Fertil. Steril.* **94,** 2860–2863 (2010).

125. Gebel M, H. *et al.* Spontaneous apoptosis of endometrial tissue is impaired in women with endometriosis. *Fertil. Steril.* **69,** 1042–1047 (1998).

126. Liu, G., Wu, C., Wu, Y. & Zhao, Y. Phagocytosis of apoptotic cells and immune regulation. *Scand. J. Immunol.* **64,** 1–9 (2006).

127. Matarese, G., De Placido, G., Nikas, Y. & Alviggi, C. Pathogenesis of endometriosis: Natural immunity dysfunction or autoimmune disease? *Trends Mol. Med.* **9,** 223–228 (2003).

128. Al-Jefout, M., Andreadis, N., Tokushige, N., Markham, R. & Fraser, I. A pilot study to evaluate the relative efficacy of endometrial biopsy and full curettage in making a diagnosis of endometriosis by the detection of endometrial nerve fibers. *Am. J. Obstet. Gynecol.* **197,** (2007).

129. Tokushige, N., Markham, R., Russell, P. & Fraser, I. S. High density of small nerve fibres in the functional layer of the endometrium in women with endometriosis. *Hum. Reprod.* **21,** 782–787 (2006).

130. Liuzzi, F. J. & Tedeschi, B. Peripheral nerve regeneration. *Neurosurg Clin N Am* **2,** 31–42 (1991).

131. Wu, J., Xie, H., Yao, S. & Liang, Y. Macrophage and nerve interaction in endometriosis. 1–11 (2017). doi:10.1186/s12974-017-0828-3

132. Vercellini, P. *et al.* Site of origin of epithelial ovarian cancer: The endometriosis connection. *Br. J. Obstet. Gynaecol.* **107,** 1155–1157 (2000).

133. Takamura, M. *et al.* Simultaneous Detection and Evaluation of Four Subsets of CD4 + T Lymphocyte in Lesions and Peripheral Blood in Endometriosis. (2015). doi:10.1111/aji.12426

134. Evans, J. & Salamonsen, L. A. Inflammation, leukocytes and menstruation. *Rev. Endocr. Metab. Disord.* **13,** 277–288 (2012).

135. Berbic, M., Ng, C. H. M. & Fraser, I. S. Inflammation and endometrial bleeding. *Climacteric* **17 Suppl 2,** 47–53 (2014).

136. Arruvito, L., Sanz, M., Banham, A. H. & Fainboim, L. Expansion of CD4+CD25+and FOXP3+ Regulatory T Cells during the Follicular Phase of the Menstrual Cycle: Implications for Human Reproduction. *J. Immunol.* **178,** 2572–2578 (2007).

137. Sakaguchi, S., Miyara, M., Costantino, C. M. & Hafler, D. A. FOXP3 + regulatory T cells in the human immune system. *Nat. Rev. Immunol.* **10,** 490–500 (2010).

138. Guerin, L. R., Prins, J. R. & Robertson, S. A. Regulatory T-cells and immune tolerance in pregnancy : a new target for infertility treatment ? **15,** 517–535 (2009).

139. Chen, S. *et al.* Expression of the T regulatory cell transcription factor FoxP3 in peri-implantation phase endometrium in infertile women with endometriosis. *Reprod. Biol. Endocrinol.* **10,** 34 (2012).

140. Berbic, M. *et al.* The role of Foxp3+ regulatory T-cells in endometriosis: A potential controlling mechanism for a complex, chronic immunological condition. *Hum. Reprod.* **25,** 900–907 (2010).

141. Liu, H. & Lang, J. H. Is abnormal eutopic endometrium the cause of endometriosis? The role of eutopic endometrium in pathogenesis of endometriosis. *Med. Sci. Monit.* **17,** RA92-9 (2011).

142. de Barros, I. B. L. *et al.* 'What do we know about regulatory T cells and endometriosis? A systematic review'. *Journal of Reproductive Immunology* **120,** 48–55 (2017).

143. Basta, P. *et al.* The frequency of CD25+CD4+ and FOXP3+ regulatory T cells in ectopic endometrium and ectopic decidua. *Reprod. Biol. Endocrinol.* **8,** 116 (2010).

144. Podgaec, S. *et al.* Foxp3 expression in deep rectosigmoid endometriosis lesions and its association with chronic pelvic pain. *J. Reprod. Immunol.* **104–105,** 96–99 (2014).

145. Podgaec, S. *et al.* CD4 + CD25 high Foxp3 + Cells Increased in the Peritoneal Fluid of Patients with Endometriosis. **3,** 2–9 (2012).

146. Thiruchelvam, U., Wingfield, M. & O'Farrelly, C. Increased uNK Progenitor Cells in Women With Endometriosis and Infertility are Associated With Low Levels of Endometrial Stem Cell Factor. *Am. J. Reprod. Immunol.* **75,** 493–502 (2016).

147. Parkin, K. L. & Fazleabas, A. T. Uterine Leukocyte Function and Dysfunction: A Hypothesis on the Impact of Endometriosis. *Am. J. Reprod. Immunol.* 1–7 (2016). doi:10.1111/aji.12487

148. Manaster, I. *et al.* Endometrial NK Cells Are Special Immature Cells That Await Pregnancy. *J. Immunol.* **181,** 1869–1876 (2008).

149. Flynn, L. *et al.* Menstrual Cycle Dependent Fluctuations in NK and T-Lymphocyte Subsets from Non-Pregnant Human Endometrium. *Am. J. Reprod. Immunol.* **43,** 209–217 (2000).

150. Lee, J. Y., Lee, M. & Lee, S. K. Role of endometrial immune cells in implantation. *Clinical and Experimental Reproductive Medicine* **38,** 119–125 (2011).

151. King, A. E. & Critchley, H. O. D. Oestrogen and progesterone regulation of inflammatory processes in the human endometrium. *J. Steroid Biochem. Mol. Biol.* **120,** 116–126 (2010).

152. Mselle, T. F. *et al.* Unique characteristics of NK cells throughout the human female reproductive tract. *Clin. Immunol.* **124,** 69–76 (2007).

153. Salamonsen, L. A. & Woolley, D. E. Menstruation: Induction by matrix metalloproteinases and inflammatory cells. *Journal of Reproductive Immunology* **44,** 1–27 (1999).

154. Tang, A. W., Alfirevic, Z. & Quenby, S. Natural killer cells and pregnancy outcomes in women with recurrent miscarriage and infertility: A systematic review. *Hum. Reprod.* **26,** 1971–1980 (2011).

155. Bulmer, J. N., Morrison, L., Longfellow, M., Ritson, A. & Pace, D. Granulated lymphocytes in human endometrium: histochemical and immunohistochemical studies. *Hum. Reprod.* **6,** 791–8 (1991).

156. Lysakova-Devine, T. & O'Farrelly, C. Tissue-specific NK cell populations and their origin. *J. Leukoc. Biol.* **96,** 981–990 (2014).

157. Li, X. F. *et al.* Angiogenic Growth Factor Messenger Ribonucleic Acids in Uterine Natural Killer Cells. **86,** (2015).

158. Thiruchelvam, U., Wingfield, M. & O'Farrelly, C. Natural Killer Cells: Key Players in Endometriosis. *Am. J. Reprod. Immunol.* **74,** 291–301 (2015).

159. Jones, R. K., Bulmer, J. N. & Searle, R. F. Immunohistochemical characterization of stromal leukocytes in ovarian endometriosis: comparison of eutopic and ectopic endometrium with normal endometrium. *Fertil. Steril.* **66,** 81–89 (1996).

160. Herington, J. L., Bruner-Tran, K. L., Lucas, J. A. & Osteen, K. G. Immune interactions in endometriosis. *Expert Rev. Clin. Immunol.* **7,** 611–626 (2011).

161. Petta, C. A. *et al.* Natural killer cells and telomerase in the endometrium of patients with endometriosis. *J. Endometr.* **2,** 182–188 (2010).

162. Van Vaerenbergh, I. *et al.* Progesterone rise on HCG day in GnRH antagonist/rFSH stimulated cycles affects endometrial gene expression. *Reprod. Biomed. Online* **22,** 263–271 (2011).

163. Black, M. B. *et al.* Comparison of Microarrays and RNA-Seq for Gene Expression Analyses of Dose-Response Experiments. **137,** 385–403 (2014).

164. Nookaew, I. *et al.* A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays : a case study in Saccharomyces cerevisiae. **40,** 10084–10097 (2012).

165. Izadi, F., Zarrini, H. N., Kiani, G. & Jelodar, N. B. Open access Hypothesis A comparative analytical assay of gene regulatory networks inferred using microarray and RNA-seq datasets Open access. **12,** 340–346 (2016).

166. Zhang, W. *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. 1–12 (2015). doi:10.1186/s13059-015-0694-1

167. Fumagalli, D. *et al.* Transfer of clinically relevant gene expression signatures in breast cancer : from Affymetrix microarray to Illumina RNA-Sequencing technology. 1–12 (2014).

168. Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M. & Gilad, Y. RNA-seq : An assessment of technical reproducibility and comparison with gene expression arrays. 1509–1517 (2008). doi:10.1101/gr.079558.108.

169. Bradford, J. R. *et al.* A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. (2010).

170. Sultan, M. *et al.* A Global View of Gene Activity and Alternative Splicing by Deep Sequencing of the Human Transcriptome. *Science (80-.).* **321,** 956–960 (2008).

171. Andrews, S. FastQC - A quality control tool for high throughput sequence data. (2010). at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

172. Hannon. FASTX Toolkit http://hannonlab.cshl.edu/fastx_toolkit/index.html by Hannon Lab.

173. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).

174. Qiagen's CLCBio CLC Genomics Workbench. http://www.clcbio.com/products/clc-genomics-workbench/.

175. CLCBio Manuals - RNA-seq.

http://www.clcsupport.com/clcgenomicsworkbench/754/index.php?manual=RNA_Seq_analysis.html.

176. EBI Ensembl http://ftp.ensembl.org/pub/release-75/gtf/homo_sapiens/Homo_sapiens.GRCh37.75.gtf.gz.

177. Aanes, H. *et al.* Normalization of RNA-sequencing data from samples with varying mRNA levels. *PLoS One* **9,** e89158 (2014).

178. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27,** 863–864 (2011).

179. Hannon Lab. FASTX-Toolkit. at <http://hannonlab.cshl.edu/fastx_toolkit/index.html>

180. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).

181. Langmead, B. Alignment with Bowtie. 1–24 (2011). doi:10.1002/0471250953.bi1107s32.Aligning

182. (https://ccb.jhu.edu/software/tophat/igenomes.shtml).

183. Trapnell, C. *et al.* Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nat. Biotechnol.* **28,** 511–515 (2011).

184. www.bioconductor.org.

185. Extraction of Differential Gene Expression Version 2.5.2.

186. Benjamini, Y. & Hochberg, Y. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).

187. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **40,** (2012).

188. Ashburner, M. *et al.* Gene Ontology: Tool for The Unification of Biology. *Nat. Genet.* **25,** 25–29 (2000).

189. Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D. & Woolf, P. J. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10,** 161 (2009).

190. Luo, W. *et al.* shinyjs: Perform Common JavaScript Operations in Shiny Apps using Plain R Code. *{PLoS} Comput. Biol.* **9,** 1830–1831 (2013).

191. Okonechnikov, K., Conesa, A. & García-Alcalde, F. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* **32,** 292–294 (2015).

192. Schmieder, R., Lim, Y. W. & Edwards, R. Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28,** 433–435 (2012).

193. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

194. Ftp://ftp.ensembl.org/pub/release-88/fasta/homo_sapiens. ENSEMBL: Homo sapiens, release 88.

195. Wolfgang Huber, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* **12,** 115–121 (2015).

196. Foundation for Statistical Computing, Vienna, A. R Core Team (2014). R: A language and environment for statistical computing. *http://www.R-project.org/.*

197. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 1–21 (2014).

198. Herndon, C. N. *et al.* Global Transcriptome Abnormalities of the Eutopic Endometrium from Women with Adenomyosis. *Reprod. Sci.* **23,** 1289–1303 (2016).

199. Zondervan, K. T. *et al.* Endometriosis. *Nat. Rev. Dis. Prim.* **4,** 9 (2018).

200. Wong, Y. F. *et al.* Identification of molecular markers and signaling pathway in endometrial cancer in Hong Kong Chinese women by genome-wide gene expression profiling. *Oncogene* **26,** 1971–82 (2007).

201. Critchley, H. O. D. *et al.* Gene expression profiling of mid to late secretory phase endometrial biopsies from women with menstrual complaint. *Am. J. Obstet. Gynecol.* **195,** (2006).

202. Arimoto, T. *et al.* Genome-wide cDNA microarray analysis of gene-expression profiles involved in ovarian endometriosis. *Int. J. Oncol.* **22,** 551–560 (2003).

203. STRING interaction network - WEE1. *ELIXIR Core Data Resource* (2017). at <https://string-db.org/cgi/network.pl?taskId=JsHk7tPMxF6h>

204. Kusakabe, K. T. *et al.* DNA microarray analysis in a mouse model for endometriosis and validation of candidate factors with human adenomyosis. *J. Reprod. Immunol.* **85,** 149–160 (2010).

205. Norton, N. *et al.* Gene expression, single nucleotide variant and fusion transcript discovery in archival material from breast tumors. *PLoS One* **8,** 1–11 (2013).

206. Illumina. RNA-Seq Offers Significant Advantages in Immuno-Oncology Research. 1–4 (2017). at <https://www.illumina.com/content/dam/illumina-marketing/documents/products/whitepapers/truseq-rna-acess-vs-nanostring-white-paper-1170-2017-002.pdf>

207. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26,** 317–325 (2008).

208. Domínguez, F. *et al.* Lipidomic profiling of endometrial fluid in women with ovarian endometriosis. *Biol. Reprod.* **96,** 772–779 (2017).

209. Barker, N. & Clevers, H. Leucine-Rich Repeat-Containing G-Protein-Coupled Receptors as Markers of Adult Stem Cells. *Gastroenterology* **138,** 1681–1696 (2010).

210. Burney, R. O. *et al.* Gene expression analysis of endometrium reveals progesterone resistance and candidate susceptibility genes in women with endometriosis. *Endocrinology* **148,** 3814–3826 (2007).

211. Nguyen, H. P. T., Sprung, C. N. & Gargett, C. E. Differential expression of Wnt signaling molecules between pre- and postmenopausal endometrial epithelial cells suggests a population of putative epithelial stem/progenitor cells reside in the basalis layer. *Endocrinology* **153,** 2870–2883 (2012).

212. Jemt, A. *et al.* Comprehensive RNA sequencing of healthy human endometrium at two time points of the menstrual cycle^{<xref ref-type='fn' rid='afn2'>†</xref>} *Biol. Reprod.* **96,** 24–33 (2016).

213. Patel, B. G., Rudnicki, M., Yu, J., Shu, Y. & Taylor, R. N. Progesterone resistance in endometriosis: origins, consequences and interventions. *Acta Obstet. Gynecol. Scand.* **96,** 623–632 (2017).

214. Ben-porath, I. *et al.* An embryonic stem cell – like gene expression signature in poorly differentiated aggressive human tumors. **40,** 499–507 (2008).

215. Venet, D., Dumont, J. E. & Detours, V. Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome. **7,** (2011).

216. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. **36,** 1090–1098 (2004).

217. Luo, H. *et al.* Understanding and predicting binding between human leukocyte antigens (HLAs) and peptides by network analysis. *BMC Bioinformatics* **16,** S9 (2015).

218. Vang, Y. S. & Xie, X. HLA class I binding prediction via convolutional neural networks. (2016).

219. Wirapati, P. *et al.* Meta-analysis of gene expression profiles in breast cancer : toward a unified understanding of breast cancer subtyping and prognosis signatures. **10,** 1–11

220. Varn, F. S., Ung, M. H., Lou, S. K. & Cheng, C. Integrative analysis of survival-associated gene sets in breast cancer. 1–16 (2015). doi:10.1186/s12920-015-0086-0

221. Sultan, M. *et al.* No Title. **956,** (2012).

222. Madala, S. K. *et al.* Matrix Metalloproteinase 12-Deficiency Augments Extracellular Matrix Degrading Metalloproteinases and Attenuates IL-13– Dependent Fibrosis. doi:10.4049/jimmunol.0903008

223. Wynn, T. A. & Vannella, K. M. Macrophages in Tissue Repair, Regeneration, and Fibrosis. *Immunity* **44,** 450–462 (2016).

224. Lech, M. & Anders, H. J. Macrophages and fibrosis: How resident and infiltrating mononuclear phagocytes orchestrate all phases of tissue injury and repair. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1832,** 989–997 (2013).

225. Wattananit, S. *et al.* Monocyte-Derived Macrophages Contribute to Spontaneous Long-Term Functional Recovery after Stroke in Mice. *J. Neurosci.* **36,** 4182–4195 (2016).

226. Low, H. B., Png, C. W., Li, C., Wang, D. Y. & Justin, S. B. Monocyte-derived factors including PLA2G7 induced by macrophage-nasopharyngeal carcinoma cell interaction promote tumor cell invasiveness. (2016).

227. Pioli, P. A. *et al.* Lipopolysaccharide-induced IL-1β production by human uterine macrophages up-regulates uterine epithelial cell expression of human β-defensin 2. *J. Immunol.* **176,** 6647–6655 (2006).

228. Shi, Y.-L., Luo, X.-Z., Zhu, X.-Y. & Li, D.-J. Combination of 17$^{N}_{L}$-estradiol with the environmental pollutant TCDD is involved in pathogenesis of endometriosis via up-regulating the chemokine I-309 –CCR8. doi:10.1016/j.fertnstert.2006.11.129

229. Hornung, D. *et al.* Localization in tissues and secretion of eotaxin by cells from normal endometrium and endometriosis. *J. Clin. Endocrinol. Metab.* **85,** 2604–2608 (2000).

230. Zhang, J., Lathbury, L. J. & Salamonsen, L. A. Expression of the chemokine eotaxin and its receptor, CCR3, in human endometrium. *Biol. Reprod.* **62,** 404–411 (2000).

231. OuYang, Z. *et al.* Interleukin-4 induces expression of eotaxin in endometriotic stromal cells. *Fertil. Steril.* **94,** 58–62 (2010).

232. Salcedo, R. *et al.* Eotaxin (CCL11) induces in vivo angiogenic responses by human CCR3+ endothelial cells. *J. Immunol.* **166,** 7571–8 (2001).

233. Gravanis, A., Makrigiannakis, A., Zoumakis, E. & Margioris, A. N. Endometrial and myometrial corticotropin-releasing hormone (CRH): Its regulation and possible roles. *Peptides* **22,** 785–793 (2001).

234. Zoumakis, E. *et al.* Cycle and age-related changes in corticotropin-releasing hormone levels in human endometrium and ovaries. *Gynecol Endocrinol* **15,** 98–102 (2001).

235. Kalantaridou, S. N. *et al.* Corticotropin-releasing hormone, stress and human reproduction: An update. *Journal of Reproductive Immunology* **85,** 33–39 (2010).

236. Makrigiannakis, A. *et al.* Corticotropin-releasing hormone (CRH) and immunotolerance of the fetus.

237. Vergetaki, A. *et al.* Differential Expression of CRH, UCN, CRHR1 and CRHR2 in Eutopic and Ectopic Endometrium of Women with Endometriosis. *PLoS One* **8,** (2013).

238. Mastorakos, G., Karoutsou, E. I. & Mizamtsidi, M. Corticotropin releasing hormone and the immune/inflammatory response. *Eur. J. Endocrinol.* **155,** S77–S84 (2006).

239. Kalantaridou, S. N., Makrigiannakis, A., Zoumakis, E. & Chrousos, G. P. Stress and the female reproductive system. in *Journal of Reproductive Immunology* **62,** 61–68 (2004).

240. Minas, V. *et al.* Abortion is associated with increased expression of FasL in decidual leukocytes and apoptosis of extravillous trophoblasts: A role for CRH and urocortin. *Mol. Hum. Reprod.* **13,** 663–673 (2007).

241. Uyanik, B., Grigorash, B. B., Goloudina, A. R. & Demidov, O. N. REVIEW ARTICLE DNA damage-induced phosphatase Wip1 in regulation of hematopoiesis , immune system and in fl ammation. *Nat. Publ. Gr.* 1–5 (2017). doi:10.1038/cddiscovery.2017.18

242. Tan, D. S. P. *et al.* PPM1D is a potential therapeutic target in ovarian clear cell carcinomas. *Clin. Cancer Res.* **15,** 2269–2280 (2009).

243. Castrillo, A. *et al.* Protein Kinase C Is Required for Macrophage Activation and Defense Against Bacterial Infection. *J Exp Med* **194,** 1231–1242 (2001).

244. Aksoy, E., Amraoui, Z., Goriely, S., Goldman, M. & Willems, F. Critical role of protein kinase C epsilon for lipopolysaccharide-induced IL-12 synthesis in monocyte-derived dendritic cells. *Eur J Immunol* **32,** 3040–3049 (2002).

245. Oshevski, S. *et al.* Differential expression of protein kinase C isoform transcripts in human hematopoietic progenitors undergoing differentiation. *Biochem Biophys Res Commun* **263,** 603–609 (1999).

246. Parisi, L. *et al.* Review Article Macrophage Polarization in Chronic Inflammatory Diseases : Killers or Builders ? *J. Immunol. Res.* **2018,** (2018).

247. Galván-Peña, S. & O'Neill, L. A. J. Metabolic Reprograming in Macrophage Polarization. *Front. Immunol.* **5,** 1–6 (2014).

248. Clapham, D. E. Calcium signaling. *Cell* **80,** 259–268 (1995).

249. Blagih, J. & Jones, R. G. Polarizing macrophages through reprogramming of glucose metabolism. *Cell Metab.* **15,** 793–795 (2012).

250. Di Conza, G. & Ho, P. C. Metabolic adaptation of macrophages in chronic diseases. *Cancer Lett.* **414,** 250–256 (2018).

251. Zhang, Y. M., Rao, C. V. & Lei, Z. M. Macrophages in human reproductive tissues contain luteinizing hormone/chorionic gonadotropin receptors. *Am. J. Reprod. Immunol.* **49,** 93–100 (2003).

252. Min, J. Y., Park, M. H., Lee, J. K., Kim, H. J. & Park, Y. K. Gonadotropin-releasing hormone modulates immune system function via the nuclear factor-κB pathway in murine Raw264.7 macrophages. *Neuroimmunomodulation* **16,** 177–184 (2009).

253. Cicinelli, E. *et al.* Chronic endometritis: correlation among hysteroscopic, histologic, and bacteriologic findings in a prospective trial with 2190 consecutive office hysteroscopies. *Fertil. Steril.* **89,** 677–684 (2008).

254. Puca, J., Hoyne, G. F. & Hoyne, G. Microbial dysbiosis and disease pathogenesis of endometriosis, could there be a link? *Allied J. Med. Res. Allied J Med Res* **11,** 1–9 (2017).

255. Moreno, I. *et al.* Evidence that the endometrial microbiota has an effect on implantation success or failure. *Am. J. Obstet. Gynecol.* **215,** 684–703 (2016).

256. Zhou, B. *et al.* A Functional Promoter Polymorphism in *NFKB1* Increases Susceptibility to Endometriosis. *DNA Cell Biol.* **29,** 235–239 (2010).

257. Bloski, T. & Pierson, R. Endometriosis and Chronic Pelvic Pain: Unraveling the Mystery Behind this Complex Condition. *Nursing for women's health* **12,** 382–395 (2008).

258. Lv, P., Tang, Z., Liang, X., Guo, M. & Han, R. P. S. Spatially gradated segregation and recovery of circulating tumor cells from peripheral blood of cancer patients. *Biomicrofluidics* **7,** 1–9 (2013).

259. Renner, S. P. *et al.* Inhibition of adhesion, proliferation, and invasion of primary endometriosis and endometrial stromal and ovarian carcinoma cells by a nonhyaluronan adhesion barrier gel. *Biomed Res. Int.* **2015,** (2015).

260. Guan, Y. T. *et al.* Overexpression of chloride channel-3 is associated with the increased migration and invasion ability of ectopic endometrial cells from patients with endometriosis. *Hum. Reprod.* **31,** 986–998 (2016).

261. Zeitvogel, A., Baumann, R. & Starzinski-Powitz, A. Identification of an Invasive, N-Cadherin-Expressing Epithelial Cell Type in Endometriosis Using a New Cell Culture Model. *Am. J. Pathol.* **159,** 1839–1852 (2001).

262. Bobek, V., Kolostova, K. & Kucera, E. Circulating endometrial cells in peripheral blood. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **181,** 267–274 (2014).

263. Chen, Y. *et al.* Evaluation of circulating endometrial cells as a biomarker for endometriosis. *Chin. Med. J. (Engl).* **130,** 2339–2345 (2017).

264. Aktas, B. *et al.* Stem cell and epithelial-mesenchymal transition markers are frequently overexpressed in circulating tumor cells of metastatic breast cancer patients. *Breast Cancer Res.* **11,** 1–9 (2009).

265. Papadaki, M. A. Co-expression of putative stemness and epithelial-to- mesenchymal transition markers on single circulating tumour cells from patients with early and metastatic breast cancer. 1–10 (2014).