RNA BASED BIOMARKERS FOR PREDICTION OF THE ENDOMETRIAL WINDOW OF IMPLANTATION

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

MOLECULAR BIOLOGY AND GENETICS

By

Ege Dedeoğlu

December 2020

RNA BASED BIOMARKERS FOR PREDICTION OF THE ENDOMETRIAL WINDOW OF IMPLANTATION

By Ege Dedeoğlu

We certify that we have read this dissertation and that in our opinion it is fully adequate in scope and in quality, as a thesis for the degree of Master of Science.

Ali Osmay Güre
(Advisor)

Özlen Konu Karakayalı

Evrim Ünsal

Approved for Graduate School of Engineering and Science

Ezhan Karaşan
Director of the Graduate School of Engineering and Science

ABSTRACT

RNA BASED BIOMARKERS FOR PREDICTION OF THE ENDOMETRIAL WINDOW OF IMPLANTATION

Ege Dedeoğlu

M.Sc. in Molecular Biology and Genetics

Advisor: Ali Osmay Güre

December 2020

Early reproductive failure is the most common issue related to successful pregnancies, as around 30% of all conceptions reach live birth. The path to a successful pregnancy is reliant on the successful implantation of the embryo to the endometrium. This event requires three major components; a viable embryo ready for implantation, a receptive endometrium in which the implantation will occur, and healthy crosstalk between the embryo and receptive endometrium. It is estimated that two out of all three implantation failures are related to endometrial origin. This has led many researchers to attempt to elucidate the mechanism behind endometrial receptivity and generate a prediction of successful implantation of endometrial origin. Although there have been plenty of articles on this subject, there is still no consensus regarding standard endometrial receptivity biomarkers.

Additionally, most of these articles' findings cannot find their way into clinics. This is highlighted by the fact that the success rate of embryo implantation in ART applied in clinics is only around 10%. This study aimed to identify novel methods and biomarkers to predict the endometrium's receptivity, which could be applied in clinics

easier and faster than the current kits in the market. We took several different approaches to achieve this aim. The first was to identify and validate particular miRNAs that showed a change in expression levels of the different days of the endometrial cycle in a healthy women's serum. Our bioinformatical analysis has yielded ten miRNAs that show statistical differences in the human endometrium and being expressed in the serum. Downstream RNA-Seq and qPCR experiments have validated specific miRNAs previously predicted and identified novel miRNAs used for this purpose. The second was using in silico methods, identifying novel genes present in the endometrium that can predict the optimal point of receptivity. If considered and validated in vitro, this novel gene-list will be a cheaper but still as powerful alternative to the current endometrial test kit used in clinics today. Further validation RNA-Seq experiments on healthy and infertile females will elucidate our novel biomarkers' strength, designed to be used in ART clinics worldwide. Furthermore, upon building on these findings, it is possible to uncover previously overlooked mechanisms leading to women's implantation success.

Keywords: Endometrial receptivity, miRNA, serum, implantation, embryo, in silico, RNA-Seq

ÖZET

ENDOMETRİYAL İMPLANTASYON PENCERESİNİN RNA TEMELLİ BİYOBELİRTEÇLER İLE TAHMİN EDİLMESİ

Ege Dedeoğlu

Moleküler Biyoloji ve Genetik, Yüksek Lisans

Tez Danışmanı: Ali Osmay Güre

Aralık 2020

Günümüzde başarısız hamileliğin en yaygın kaynağı, erken gebelik kaybıdır, başarılı gebeliklerin yalnızca %30 canlı doğum ile sonuçlanır. Başarılı bir hamileliğin gerçekleşmesi, embriyonun endometriyum'a başarılı implantasyonuna dayalıdır. Başarılı bir implantasyon gerçekleşmesi için üç temel unsura ihtiyacı vardır; implantasyona hazır sağlıklı bir embriyo, implantasyonun gerçekleşeceği reseptif bir endometriyum ve bu ikisi arasında sağlıklı karşılıklı etkileşim hali. Başarısız implantasyonların üçte ikisinin endometriyal kaynaklı olduğu düşünülmektedir. Bu, birçok araştırmacının endometriyal reseptivite mekanizmalarını araştırmalarına ve dolayısıyla endometriyal kökenli başarılı implantasyon tahmini gerçekleştirmeye itmiştir. Bu konu üzerinde birçok araştırma yayınlanmasına karşın, endometriyal biyobelireçleri hakkında hala standardize bir fikir birliği bulunmamaktadır. Ek olarak, bu araştırmaların sonuçlarının neredeyse hiçbiri, klinik uygulamalarda yerlerini bulamıyorlar. Bu gerçek, yardımcı üreme teknikleri (ART) uygulayan kliniklerdeki, embriyo implantasyon başarısının %10 civarında olmasıyla daha da belirginleşmekte. Bu araştırmadaki amacımız, kliniklere uygulanması mevcut kitlere nazaran daha

kolay ve hızlı uygulanabilir, implantasyon penceresinin tahminde kullanılabilecek yeni metodolojiler ve biyobelirteçler tespit etmektir. Bu amaca ulaşmak için birkaç farklı yol denedik. Bunlardan birincisi, sağlıklı kadınların kan serumlarından, endometriyal siklus sırasında değişim gösteren miRNA tespit etmekti. Biyoenformatik analizimiz, 10 adet miRNA'nın bu doğrultuda kullanılabilineceğini gösterdi. Devamında yapılan kPZR ve RNA-Sekanslama deneyleri, bu 10 miRNA arasından, doğrulanabilir olanları tespit etmemizi sağladı. Ek olarak, endometriyal reseptivite tahmini yapılabilecek yeni miRNA'ların tespit edilmesini sağladı. İkincil olarak in silico yöntemler kullanarak, temel üç endometriyal fazın tahmininde kullanılabilecek yeni genlerin tespiti ve validasyonuydu. İsmi NERS-17 olan bu yeni gen listesi, uygun görüldüğü ve in vitro validasyonu tamamlandığı takdirde, şu anda kliniklerde kullanılan alternatiflerinden daha ucuz olacak ama tahmin gücünden bir şey eksilmeyecektir. Ileride gerçekleşecek RNA-Sekanslama deneyleri, dünya çapındaki ART klinkiernde kullanılmak üzere, hem sağlıklı kontrollerde hem in vitro fertilizasyon (IVF) tedavisi gören infertil kadınlarda, tespit edilen yeni biyobelirteçlerimizin gücünü daha iyi anlamamızı sağlayacaktır. Son olarak, bu bulguların üzerinde durarak daha önce dikkat edilmeyen ve kadınlarda implantasyon başarısın belirleyen mekanizmalar daha iyi anlaşılabilir.

Anahtar kelimeler: Endometriyal reseptivite, miRNA, serum, implantasyon, embriyo, in silico, RNA-Sekanslama

ACKNOWLEDGEMENTS

First of all, I would like to express my deep gratitude to my supervisor, Assoc. Prof. Dr. Ali Osmay Güre for his constant encouragement, guidance, and his patience with me through this 3.5-year adventure I have embarked on. I am fortunate and grateful for him challenging me with ideas that previously would seem impossible to me, and through his support, showing me, I have the will to overcome such obstacles.

Secondly, I would like to thank our collaborators, a St Mikrogen Genetic Diagnostic Center and Gen-Art In Vitro Fertilization Women's Health and Biotechnology Center. Without their professional advice and support, this thesis would not be possible. I would like to thank Prof. Dr. Volkan Baltacı, Dr. Aysun Baltacı, Assoc. Prof. Dr. Evrim Ünsal, Asst. Prof. Dr. Süleyman Aktuna, Spc. Dr. Leyla Özer, Opr. Dr. Sertaç Şen, Ahmet Atik, and Spc. Cemile Engüzel, for their active contributions and constant understanding during the project.

I would like to offer my appreciation to the members of my thesis jury: Assoc. Prof. Özlen Konu Karakayalı, for her understanding and assistance, during this thesis, and my academic journey overall and Prof. Dr. Volkan Baltacı, as without his active supervision and passion for this project, I would not be where I am today.

I sincerely would like to offer my most sincere thanks to both Barış Küçükkaraduman and Muhammad Waqas Akbar. For keeping me grounded and motivated in times of distress and always being there to offer their helping hand, without their brotherly friendship and guidance, these long years would have gone by much harder. I would

like to thank previous and current AOG Lab members: Alper Poyraz, Şükrü Atakan, Hamza Yusuf Altun, Seçil Demirkol Canlı, Can Türk, Kerem Mert Şenses, Murat İşbilen, Abbas Güven Akçay, Farid Ahadli, Marzana Ishraq, Farid Sadati, Shila Azizolli, İsli Cela, Noor Niaz & Ronak Naeemaee for their assistance during my studies in any shape or form. I would also like to thank my students that I had the chance to supervise during my Master's, especially Osama Jawad and Hilal Üstün, whose hunger for discovery motivated me to be a better researcher.

I want to thank my dearest and closest friends for the constant support they provided, with laughter and conversation, and the fond memories we all share; I know that I cherish them with all my heart throughout my life. May we share our happiness in the future.

I want to thank Naz Bozbeyoğlu, Irem Evcili, and Ibrahim Oğuzhan Tarman on a separate note for their invaluable friendship through our 3.5 adventure that we embarked on together. The constant laughter they provided and assistance in any technical issues during my experiments were terrific. I would also like to thank Ünal Metin Tokat, Pelin Gülizar Ersan, Tuğçe Bildik, Aslı Bartan & Berfu Saraydar, for their friendship.

Finally, I owe my greatest thanks to my family; my father Emin Dedeoğlu, my beautiful mother Tülin Dedeoğlu & my older brother Boran Dedeoğlu. Throughout my life, I have felt their love and support in everything I wanted to do. I am profoundly grateful for them providing me with endless opportunities to improve myself as a person and in a loving environment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADH1A	Alcohol Dehydrogenase 1A (Class I), Alpha Polypeptide
ALDH2	Aldehyde Dehydrogenase 2 Family Member
ART	Artificial Reproductive Technologies
BH	Benjamini-Hochberg
C2	Control 2
C5	Control 5
CATSPERB	Cation Channel Sperm Associated Auxiliary Subunit Beta
CPM	Counts Per Million
CV	Coefficient of Variation
DEMiR	Differentially Expressed microRNA
DGCR8	DGCR8 Microprocessor Complex Subunit
DKK1	Dickkopf WNT Signaling Pathway Inhibitor 1
DNA	Deoxyribonucleic Acid
DPP4	Dipeptidyl Peptidase 4
EBI	European Bioinformatics Institute
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial to Mesenchymal Transition
ENA	European Nucleotide Archive
EP	Early Proliferative
ERA	Endometrial Receptivity Assay

ES	Early Secretory Phase
FET	Frozen Embryo Transfer
FJX1	Four-Jointed Box Kinase 1
GEO	Gene Expression Omnibus
GPX3	Glutathione Peroxidase 3
HDAC1	Histone Deacetylase 1
HRP	Horse Radish Peroxidase
HSD17B2	Hydroxysteroid 17-Beta Dehydrogenase 2
IUD	Intrauterine Device
IVF	In Vitro Fertilization
KEGG	Kyoto Encyclopedia of Genes and Genomes
L2FC	Log 2 transformed Fold Change
LDLR	Low-Density Lipoprotein Receptor
LH	Luteinizing Hormone
LNA	Locked Nucleic Acid
LP	Late Proliferative
LS	Late Secretory
MCD	Menstrual Cycle Day
miRNA	micro Ribonucleic Acid
MMP	Matrix Metalloproteases
MMP11	Matrix Metallopeptidase 11
MMP7	Matrix Metallopeptidase 7
MYOCD	Myocardin
NERS-17	Novel Endometrial Receptivity Signature - 17

NGS	Next-Generation Sequencing
OTUD4	OTU Deubiquitinase 4
PAEP	Progestagen Associated Endometrial Protein
PART1	Prostate Androgen-Regulated Transcript 1
PCA	Principal Component Analysis
PCOS	Polycystic Ovary Syndrome
PCR	Polymerase Chain Reaction
PKHD1L1	Polycystic Kidney And Hepatic Disease 1 -Like 1
PRO	Proliferative Phase
RAB10	RAB10, Member RAS Oncogene Family
RIF	Repeated Implantation Failure
RMA	Robust Multichip Average
RNA	Ribonucleic Acid
RT	Reverse Transcription
RTKN2	Rhotekin 2
SCARA5	Scavenger Receptor Class A Member 5
SEMA3A	Semaphorin 3A
SIAH3	Siah E3 Ubiquitin Protein Ligase Family Member 3
SLC15A2	Solute Carrier Family 15 Member 2
SLC30A2	Solute Carrier Family 30 Member 2
STXBP6	Syntaxin Binding Protein 6
SSM	Single Sample Matrix
TSPAN8	Tetraspanin 8

UMI	Unique Molecular Identifier
WOI	Window of Implantation

INTRODUCTION

1.1 Statistics of reproductive success

It is estimated that over 75% of all conceptions result in early reproductive failures [1]. It was also reported that more than 50% of ongoing pregnancies are lost at pre-clinical stages [2, 3]. Although this conception rate is higher in young couples, it is below 30% [4]. With the dawn of the omics-era after the human genome project, many researchers have tried to understand the mechanisms that govern successful pregnancies to increase the numbers stated above [5].

The omics studies regarding reproduction suffer from low reproducibility and low success in increasing ART applications' success rates. This is highlighted by the fact that transferred embryos' worldwide success rate is estimated to be lower than 10% [6]. A study done by Boomsa and colleagues highlight a more optimistic outlook. Their study, done with 210 women undergoing IVF, showed that 51.4% of transferred embryos were successfully implanted. However, the rate of successful pregnancies was lower than 25% [7]. Illustrated in **Figure 1.1** [8].

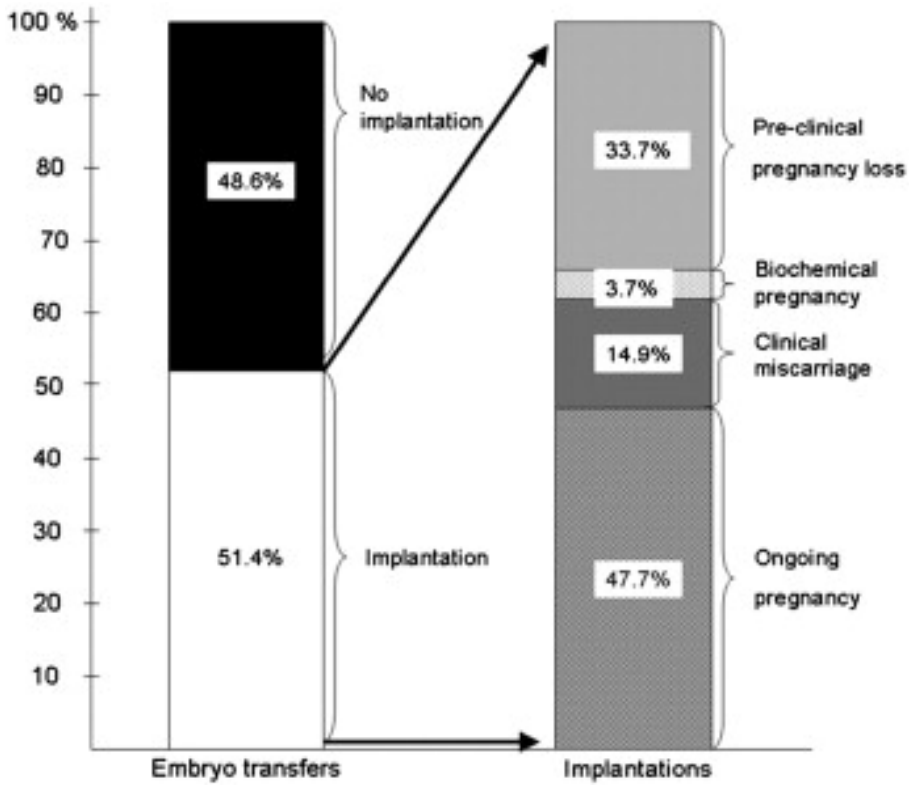


Figure 1.1. Graphical representation of successful ongoing pregnancies of 210 women undergoing IVF. [8] (See Appendix for Copyright Permissions)

Endometrial receptivity and embryonic implantation

1.2.1 Embryo Implantation to the Receptive Endometrium.

The endometrial tissue is a dynamic tissue that undergoes significant cyclic physiological changes under the influence of hormones, primarily estrogen (E), progesterone (P), during the menstrual cycle of each female [9].

The process in which invasion of the embryo inside the tissue is called implantation. The endometrium's importance is that it allows the implanted embryo to develop into a healthy conceptus, resulting in a live birth. However, the endometrium is one of the most problematic issues for the embryo to implant, as the embryo can implant only certain days to the endometrial tissue; this phenomenon is called receptivity. The time in which the endometrium is receptive to the endometrium is termed as the window

of implantation (WOI) [10]. The WOI occurs around days 19-23 of a 28-day menstrual cycle. It lasts around 3-to-5 days and generally occurs seven days after the LH Surge happens. The implantation process can be summarized in three steps: 1) Apposition, where an unstable contact of the embryo with the endometrium is present, 2) Adhesion, when this contact between the endometrium and embryo is much more stable, 3) Invasion, where the trophoblast invades and migrates inside the uterine tissue [6]. Summarized in **Figure 1.2**

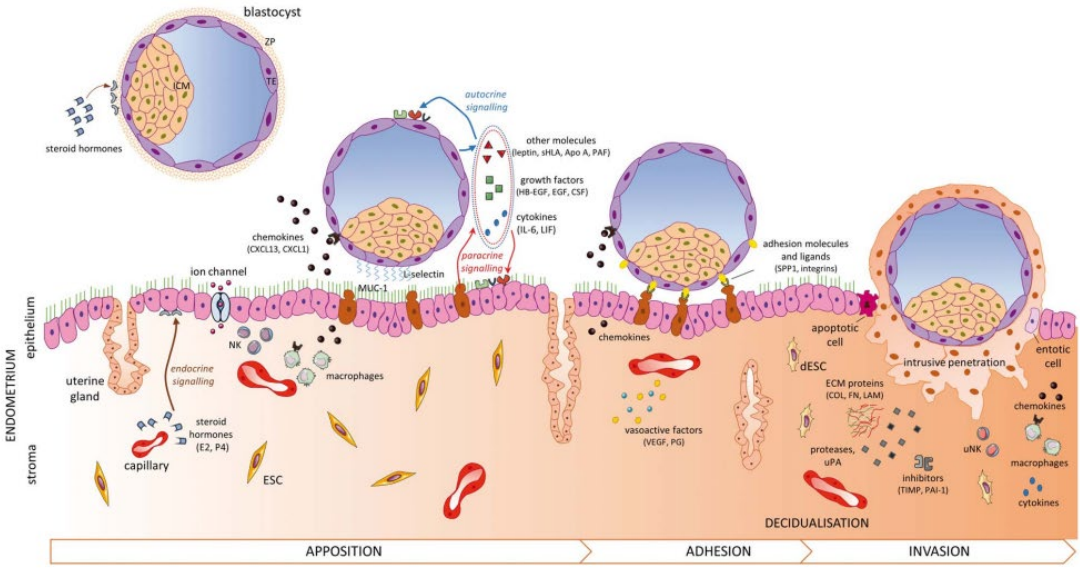


Figure 1.2. Schematic of the implantation process. [6] (See Appendix for copyright permission.)**The Window of Implantation (WOI)**

1.2.2.1 Different phases of the endometrial cycle.

The endometrial cycle consists of three major phases, menstruation (M), the proliferative phase (P), and the secretory phase (S). The latter two phases can further be compartmentalized to early proliferative (EP), late proliferative (LP), early secretory (ES), mid secretory (MS/WOI), and late secretory (LS). Each phase has its transcriptomic signature and biological processes.

The menstrual phase generally occurs from the first days to day 5 of the cycle. The main processes occurring during menstruation are tissue breakdown, apoptosis, tissue breakdown, and DNA repair [11, 12]. However, tissue desquamation is the primary biological process during this phase, facilitated by matrix metalloproteinases (MMP) [13].

The proliferative phase occurs after the menstrual phase until ovulation, which occurs around the 12^{th} to 14^{th} day of the menstrual cycle. An increase in E levels activates this phase. During this phase, biological processes are primarily tissue remodeling, vasculogenesis, angiogenesis, cellular differentiation, and proliferation [12, 14, 15]. MMPs are repressed at this phase of the endometrial cycle. Hox family genes are increased in expression during this phase, facilitating glandular differentiation.

The ES phase occurs directly after ovulation and up until the WOI (day 19-23). It is marked by a dramatic inhibition of cell division, which is inferred by the stark repression of growth factor genes [16]. Additionally, as this phase is priming the endometrial tissue for implantation, it is metabolically very active compared to the P phase [12].

As stated in the previous section, the MS phase, or WOI, occurs during menstrual cycle days 19-23. During this phase, critical biological processes are adhesion, innate immune response, and response to wounding. Genes coding for the complement system are regulated during this phase, compared to the ES phase. These responses are similar to those of tissue injury [17]. This, in turn, might indicate that the embryo

can implant to the endometrium when it is “injured,” in other words, when it is weaker compared to other stages of the cycle [18].

The last phase of the endometrial cycle is the LS phase. It is characterized by an increase in an inflammatory response, primarily with apoptosis and matrix degradation, to prepare the endometrial tissue for shedding. The processes stated here destroy the endometrium's receptive status, closing the WOI until the next cycle [14].

1.2.2.2 Complications of the Window of Implantation (WOI).

Complications in WOI are identified, most accurately in women with repeated implantation failure (RIF). RIF has been defined as the failure to achieve a clinical pregnancy after transferring a minimum of four good quality embryos in at least three fresh or frozen embryos in a woman with an age less than 40 [19, 20]. This means these women suffer from a complication of endometrial origin.

It has been shown that two significant complications impair optimal embryo implantation by affecting the WOI. The first being that the WOI is displaced or the expected duration is shortened [21]. This was highlighted that Valdes and colleagues identified that more than 25% of RIF patients have a displaced WOI. Displacement of the WOI can occur in three distinct ways; 1) delayed, 2) advanced, and 3) shorter than expected [22], summarized in **Figure 1.3.** The other being that in some patients with RIF, the WOI is compromised completely, treating RIF as a pathology [23]. Sebastian-Leon and colleagues have recently generated two signatures, a “disrupted” model and a “displaced” model. Combining these models has classified more than half of RIF patients as having both a displaced and pathological WOI [24].

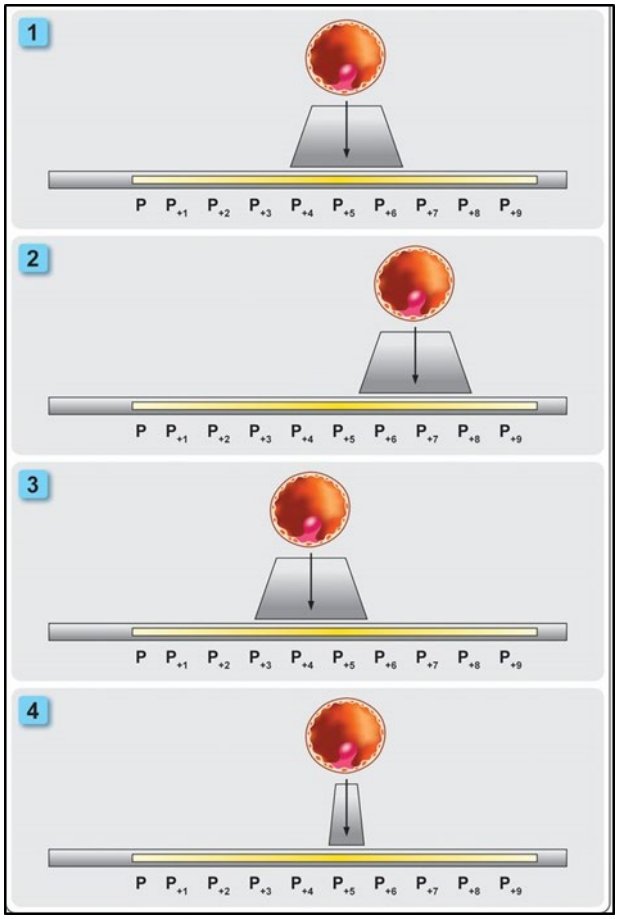


Figure 1.3. Schematic of displacement of WOI. [25] (See Appendix for copyright permission)

1.2.2.3 Detection of the Window of Implantation (WOI).

Conventional markers used in endometrial receptivity detection are outdated and unable to predict endometrial receptivity accurately. Two of the most common conventional markers are endometrial thickness & Noyes Criteria. Previously endometrial thickness > 7 mm via ultrasound was considered a receptive endometrium. However, analysis done with more than 39,000 women identified the sensitivity as 99% while the specificity at 3%, which is very poor [26]; another study concurred that endometrial thickness should be ignored when deciding for endometrial receptivity [27]. Likewise, the Noyes Criteria was once considered the

gold standard of endometrial dating [28], was shown in separate studies not to be linked with successful pregnancy outcomes [29, 30]. Such results could be linked to the fact that, due to the absence of a standardized measure, the observation is done on different machines and different people's decision-making process. This will inevitably introduce variability that would not be present in current biomarkers.

1.2.2.3.1 Molecular markers of receptivity

The most popular molecular marker is the LH Surge of the individual. As stated previously, the WOI is synonymous with LH+7 & LH+8 in the reproductive biology field. This is far from a safe prediction, as it has been shown that this method is also not reliable, as there is patient to patient variability [31].

Therefore, the focus shifted towards molecular biomarkers. Besides transcriptomic biomarkers, protein-based [32] and cytokine-based [7, 33] biomarkers were also reported. These studies did not yield promising results, which is evident in that they are not currently used in IVF clinics.

Out of the molecular biomarkers, those that use a transcriptomic signature are most often applied in ART clinics[34-36]. Among these, the endometrial receptivity array (ERA) remains the most popular worldwide due to its high success in predicting the WOI [37], although there have been some issues raised against it as well [38]. The endometrial receptivity array (ERA) is an in vitro diagnostic kit that determines the personalized WOI. Its major application is in ART clinics worldwide and was developed by igenomix®. ERA was developed as a 238 gene panel that can accurately predict four different stages of the endometrial cycle, the P, ES, WOI, and LS, with 99.8% sensitivity and 88.6% specificity [39]. Briefly, an endometrial biopsy is

collected at either LH+7 or P+5 of a natural IVF cycle. The result will indicate whether that biopsy is receptive, pre-receptive, or post-receptive. In the receptive result, the patients' embryos are frozen and transferred (FET) in the patient's next cycle. If the result yields a non-receptive response, the healthcare specialist applying the test, will require a second endometrial biopsy the following cycle on a different day, until the WOI is determined. Application of ERA on 6000 patients showed that in ~30% of patients are classified as non-receptive. Out of those 1700 patients, 85% were classified as pre-receptive, 12.6% as post receptive and 2.4% as proliferative[40]. This finding supports the claim, that women do have personalized WOI (pWOI), independent of LH Surge, which reinforces the idea, of a prediction tool, independent of LH Surge, to detect the pWOI.

However, there are some drawbacks to the ERA, and these are as follows. 1) The usage of an endometrial biopsy introduces an unwanted risk to the patient. 2) It cannot detect a disrupted WOI and cannot provide any recommendation to the healthcare provider in that regard. 3) It can be used only in natural cycles and mild hormonal therapy treatment, not conventional IVF patients. This reduces its application area tremendously, especially in countries like Turkey, where IVF treatments are much more popular than natural cycle IVF. 4) The predicted WOI and the WOI that the patients get embryo transferred are not the same. 5) A 238 gene panel causes the ERA's price to be high, ranging between $800-$1000, depending on your country, according to igenomix. This makes it mostly unaffordable, especially in developing countries. 6) Its requirement of repetition can cause unwanted monetary and psychological detriment to the patient. 7) Finally, it was shown by the developers of

the ERA test that fresh embryo transfers yield higher rates of successful pregnancies compared to FET [41].

1.3 miRNAs in Endometrial Receptivity and Embryo Implantation

Mature miRNAs are endogenous ncRNAs with sizes ~22 nucleotides, their primary and most studied function being epitranscriptomic regulation of gene expression by targeting mRNAs, silencing them [42, 43]. MiRNAs are involved in many biological events, such as differentiation, epithelial-to-mesenchymal transition (EMT), apoptosis, morphogenesis, and cellular metabolism [44-46]. They have also been implicated in both endometrial receptivity and embryo quality assessment. Especially in the implantation process, being a key player in the crosstalk between the two requires successful implantation [47]. Their importance as potential biomarkers in many phenotypes has been highlighted in recent years [48, 49]. This is since their expression changes in circulating fluids like serum are very stable, and can withstand extensive long-terms storage, provided appropriate storage conditions [50].

1.3.1 miRNAs in embryonic development and determining embryo viability

MiRNAs are shown to have an essential role in embryonic development, as shown with specific knockout experiments. Knocking out genes that play critical roles in miRNA biogenesis' canonical pathway showed halting of embryonic development at different days [51]; however, the development halting occurs after the pre-implantation stage the embryo [52]. Studies done on human IVF embryos have identified individual miRNAs that might be markers of embryo viability. Researchers have compared spent embryonic culture medium of implanted and non-implanted embryos [53, 54]. For instance, miR-661 was reported to be upregulated in blastocyst

culture media in non-implanted embryos [55]. However, these studies lack consistency amongst themselves, making it difficult for clinic application.

1.3.2 miRNAs in endometrial receptivity.

It has been shown by a multitude of studies that individual miRNAs exhibit changes in expression levels during the WOI. miRNAs that silence proliferation was up-regulated in the WOI, compared to pre-receptive state [56]. Individual miRNAs' expression was also found to be down-regulated between the ES and WOI [57]. MiRNAs also govern the mesenchymal state in the endometrium; as a member of the miR-200 family, miR-429 was downregulated during WOI, thus facilitating an EMT-like phenotype [58, 59]. MiRNAs play an integral role in forming the endometrium's receptive state, which is not surprising, as their role in other significant biological processes is known.

1.3.3 Circulating miRNAs at different stages of the endometrial cycle

Two studies have attempted to identify a change during the different endometrial cycle stages in circulating biofluids [60, 61]. Rekker et al. have used plasma of normal cycling women. Their cohort consists of 9 women, plasma collected at only four-time points of the endometrial cycle, and the miRNA expressions were calculated using qPCR, with an Exiqon Human Panel I assay. They could not identify a miRNA that exhibited expression change during the menstrual cycle [60]. However, the study has significant issues: 1) The study was done in 2013, meaning the panel used during the experiments did not contain all expressed miRNAs. 2) Using qPCR instead of RNA-Seq reduces the sensitivity of your detection. 3) The number of sample days collected and patients may not be enough to gain statistical significance.

Kresowik et al. did a similar experiment but checked the expression difference of 8 miRNAs, previously known to show differences during the WOI, at the endometrial tissue level. Additionally, they only compared the P phase with the WOI. The advantage of their study is that they also checked the miRNA expressions in endometrial biopsies. This allowed them to discover that most of the miRNAs observed showed parallelity in their expression changes in the serum compared to the endometrium. They identified miR-31 as significantly changing in both the endometrium and serum of healthy cycling women [61]. Even though their experimental setup had issues similar to the previous study, this is a promising result for miRNAs' idea as potential non-invasive biomarkers of the WOI.

1.4 Rationale and Aims of the Study

A key component of successful implantation is that the endometrium should be receptive to the viable embryo. Although many studies identify potential biomarkers for detecting the WOI, almost none can be applied in clinics due to inconsistencies. The most popular kit that makes endometrial receptivity prediction is the ERA. ERA achieved what previous markers of receptivity could not; it had proved: 1) The WOI is personalized for every female, 2) that with transcriptomic approaches, the WOI was quantifiable, and 3) it is superior to prior approaches. However, it has two significant drawbacks.

1) The requirement of an endometrial biopsy forces the patient to undergo a FET. It also inhibits its usage for IVF patients undergoing standard ovarian stimulation protocols. These patients are the actual people the ERA should be used for.

2) With an average price of 900$, it is not affordable for most of the world's population. One of the main reasons for this price tag is that the ERA is a 238 gene panel and checked with RNA-Seq technology.

In that line, our study had the following aims:

- •Identify and potentially validate serum miRNAs that can be used in WOI prediction for a precursor finding for eventual miRNA-serum based endometrial receptivity detection
  - Identification of differentially expressed miRNAs during the WOI that is also expressed in the serum, in silico. Validation of these miRNA expressions with miRNA-Seq and qPCR.
  - Identify novel miRNAs that show expression change exclusively in the serum throughout the menstrual cycle.
- Generation of a predictive gene-list to be used in endometrial biopsies, as a cheaper alternative to ERA, that can predict the different phases of the endometrial cycle

CHAPTER 2

MATERIALS AND METHODS

2.1 Detection of candidate serum microRNAs for the Window of Implantation (WOI) prediction

2.1.1 GEO Dataset selection for candidate microRNA analyses for in silico analyses.

As our study's focus was to generate a signature based on healthy, fertile females, we focused on datasets that contained samples obtained from healthy females at this stage of the research. Another focus was that any study would contain no data available from serum samples of cycling women. We obtained datasets obtained from endometrial biopsies or tissues. We used the keywords: “endometrium,” “window of implantation,” “microRNA,” “menstruation,” and “endometrial receptivity.” We wanted to exclude the terms that were related to any diseases of endometrial origin and that might impair optimal receptivity of participants, henceforth datasets that included the terms; “cancer,” “endometriosis,” “repeated implantation failure (RIF),” & “polycystic ovary syndrome (PCOS)” were excluded. The resulting search yielded five different studies. One of these studies contained only 1 sample per endometrial stage. Therefore it was discarded from the study. The remaining four datasets were used in downstream analyses. The following table (**Table 2.1**) summarizes the characteristics of each dataset.

Table 2.1 Summary of the datasets that have been selected for analyses.

GEO Accession Number	Sample Number	Sample Type	Endometrial Stages	Technology	Year
GSE34435 [57]	7	Endometrial Biopsy	ES & WOI	Agilent miRNA Microarray	2011
GSE44558 [62]	20	Endometrial Fluid Aspirate	E.PRO, L.PRO, ES, WOI, LS	Agilent miRNA Microarray	2013
GSE86491 [63]	14	Endometrial Biopsy	PRO & WOI	Illumina miRNA-Seq	2016
GSE109053 [64]	12	Endometrial Biopsy	ES & WOI	Illumina miRNA-Seq	2018

2.1.2 Finding differentially expressed serum microRNAs as candidate WOI predictors.

2.1.2.1 Finding differentially expressed miRNAs for microarray datasets.

The raw expression data for both datasets were downloaded from the GEO database. The *norRMA* method was utilized to normalize the Agilent microarray's raw values, which is an alternative to quantile normalization, the method generally recommended for Agilent microarrays [65]. Briefly, it was shown that using the RMA method [66], without background correction to signal intensity, the signal precision between biological replicates had the least variability. Therefore, this method was used to get a summarized miRNA expression data. Upon the obtained data, the differential gene

expression was calculated via the “RankProd 2.0” R statistical software package [67]. The reason for using this package was that it was much superior to a Student’s t-test when comparing a low amount of samples (biological replicate number < 10) [68] that are also unpaired. A gene was considered as significantly differentially expressed when p.adj<0.05 & |L2FC|>1.5. The WOI was taken as a reference point in differential expression analysis.

2.1.2.2 Finding differentially expressed miRNAs for RNA-Seq datasets.

Both the datasets are paired RNA-Seq data, meaning that there are two fastq files for each sample. This increases the accuracy of the analysis. The fastq files for each dataset were downloaded from EBI-ENA. Pre-processing and obtaining read counts for each miRNA-Seq sample was done on the Ubuntu 16.04 LTS terminal. The fastq pairs for each sample were quality controlled via “fastqc.” Adapter trimming and quality trimming steps were done with the “TrimGalore” tool. TrimGalore reads the first million reads and infers the adapter sequence itself; in this case, it has not been provided by the user. Both of our datasets have used the Illumina Small RNA Sequencing adapter, which sequence is “TGGAATTCTCGG.” Upon adapter trimming, reads with three bases lower than the quality score (Phred score) 30 were discarded. The paired reads are matched to one another, and the trimmed reads are written to a separate folder. Alignment and obtaining read counts were done utilizing an automated tool called miARma-Seq [69]. This tool allows the user to do the standard RNA-Seq procedures in a fully customizable and automated way. The tools support multiple alignment tools. However, the miRNA reads were aligned with the “bowtie2” tool [70]. The parameters for bowtie2 were as follows: --local -p 8 -q --phred33 -D 20 -R 3 -N 0 -L 8 -i S,1,0.50 [71]. The alignment was done on the GRCh38.p13 genome downloaded from the GenCode website. Upon alignment

"featureCounts" tool was used to obtain the raw read counts. The gtf file required for "featureCounts" was obtained from miRbase release 22 [72] and filtered for homo sapiens miRNAs. This process yielded the raw read counts for both datasets. The raw read counts were filtered using the "edgeR" R package's "filterByExpr" function [73]. This function filtered out certain miRNAs with a CPM < 1 in 70% of all the samples., this yielded the final miRNAs to be normalized. Normalization and differential gene expression was calculated in a paired manner using the "DESeq2" R package [74]. miRNAs that had BH adjusted p-value < 0.05 were considered to be differentially expressed. Normalized read counts were log-transformed in base 2. The WOI was taken as a reference point in differential expression analysis.

2.1.2.3 Cross-referencing differentially expressed miRNAs in serum samples.

After identifying the differentially expressed miRNAs in different endometrial samples, our focus was to identify those highly expressed in the serum. For this purpose, we searched GEO Database and found a miRNA-Seq experiment that contained both serum and whole blood samples, with the GEO Accession code GSE100467 [75]. We merged the raw data uploaded to GEO by following the authors' direction using the "isomiRs" R package [76]. This yielded the raw read counts of the 39 serum samples and 77 whole blood samples. The raw read counts were filtered using the filterByExpr function of the edgeR function and normalized using DESeq2, and normalized read counts were log-transformed in base 2. The filtration stringency was higher than expression analysis, wanted to reduce false-positive results as much as possible. Henceforth, miRNAs that had CPM < 5 in 80% of samples were discarded.

The differentially expressed miRNAs from the four datasets were cross-referenced with the serum miRNAs using the web-tool Venny 2.1 [77]. This yielded the miRNAs

that are differentially expressed in endometrial tissue and are expressed highly in serum samples. The remaining miRNAs from the four distinct datasets were compared with one another. miRNAs that were common amongst at least two datasets were considered for further analysis. The resulting miRNAs were all checked for serum expression levels and coefficient of variation values, and whole blood expression. MiRNAs that were not annotated properly were only common amongst the microarray datasets, showed conflicting fold change between datasets were filtered out. MiRNAs that also showed high serum variance and expression (%CV >= 8%, mean > 5) while showing low blood variance and expression (%CV > 8%, mean <=5). Graphs necessary for drawing expression values were done on GraphPad Prism 8 (GraphPad Software Inc., La Jolla, CA, USA).

2.2 In vitro validation & discovery of candidate serum miRNAs

2.2.1 Serum collection from healthy controls.

For this study, a total of 10 patients were recruited. From each patient, blood samples were collected in varying amounts in 7 days during their menstrual cycle. These seven days were chosen to represent the different stages of the endometrial cycle. The days for each patient are 9,11,13,16,19,21 & 24. For the patients to be considered for the study, they needed to suffice certain criteria. They needed to have at least one child, born through natural conception, age<40, have regular and consistent menstrual cycles each month, not be using oral contraceptives or IUDs, in the case they must have used it at least three months before serum collection. The LH surge of each control would be determined with an ELISA experiment. The serum collection protocol for controls' ethics committee approval was obtained from Gazi University Hospital, Ankara,

Turkey. The controls were recruited into the study by Bilkent University and Mikrogen Genetic Diagnostics Lab, Ankara, Turkey. Upon whole blood, collection serum was extracted using standard operating procedures [78]. The serum samples were aliquoted by 250 µl and were stored in -80℃ for future use.

2.2.2 ELISA Measurements for LH Surge Detection.

Determination of LH surge was important for potential normalization of expression profiles according to the LH Surge. To determine the LH Surge, an ELISA was performed on frozen serum samples of the control participants. A readily available ELISA Kit (Thermo Fisher Scientific, USA, Cat No: EHLH) was used following the manufacturer's directions to achieve this. Briefly, 100µl of serum sample was plated too, that each column would be a separate patient and each row would be a particular day of the menstrual cycle. The remaining two columns were used for standards, which were diluted six times. After plating the serum and standards, incubate at room temperature by gently shaking for 2.5 hours. Wash each well with 300µl of 1X Wash Buffer 4 times after the last wash blot on a clean paper towel. Add 100µl of biotinylated antibody to each well and incubate at room temperature for 1 hour by gentle shaking. Rewash 4 times again. Add 100µl Streptavidin-HRP solution, incubate for 45 minutes, by gentle shaking at room temperature. Rewash, four times. Addition of 100µl TMB substrate to each well followed by 30-minute incubation by gentle shaking at room temperature and in the dark. Addition of 50µl Stop Solution to each well. The manufacturer provided all of the reagents required, including the ELISA plate. The serum samples were plated in singlets. ELISA readings were measured by an ELISA reader (Molecular Devices Vmax, USA) within 30 minutes of stop solution addition and read on 450nm wavelength.

2.2.3 MiRNA Isolation, Quantification, and Quality Control Analysis.

2.2.3.1 MiRNA Isolation.

Before isolation, the 250µl aliquoted frozen serum samples are thawed on ice for around 1-2 hours. After thawing, the serum samples are centrifuged at 13000 rpm (Biofuge Pico, Heraeus Instruments, Germany) at 4℃ for 10 minutes. Before extraction, regardless of the kit being used, this process has yielded more consistent results in miRNA expression-based studies [79]. The 200µl serum is used as the starting material for extraction. The miRNeasy Serum/Plasma Advanced Kit (Qiagen Inc., USA, Cat. No: 217204) was utilized. The isolation protocol was carried out by following the manufacturer's instructions. For RNA isolation control downstream, 3.5µl of 1:125 diluted cel-miR-39a template spike-in (Qiagen Inc., USA, Cat No: 339390) is added to the isolation step. Isolated RNA is eluted with 20µl Nuclease Free Water (provided with the kit). The cDNA reaction is set-up immediately. After elution, the RNA samples are stored in -80℃.

2.2.3.2 cDNA Generation.

According to the manufacturer's directions, cDNA synthesis reactions were carried out using the miRCURY LNA miRNA RT Kit (Qiagen Inc., USA, Cat. No: 339340). As a control for the cDNA synthesis, the UniSp6 template (provided with the kit) was added as 0.5µl for each 10µl cDNA reaction. The cDNA reaction setup is as follows: 1) 60 min at 42℃, 2) 5 min at 95℃, 3) Infinite storage at 4℃.

2.2.3.3 Quantification of candidate miRNAs with qPCR.

Locked nucleic acid (LNA) enhanced custom primer assays (Qiagen Inc., USA, Cat. No: 339317) were used for the qPCR Reaction. The primers for the target miRNAs &

controls (UniSp6, cel-miR-39a-3p, hsa-miR-451a, hsa-miR-23a-3p, hsa-miR-103a-3p & hsa-miR-2861) were resuspended with 220µl Nuclease-Free Water and stored in -20℃ until further use.

The qPCR reaction was prepared with the miRCURY LNA SYBR Green PCR Kit (Qiagen Inc, USA, Cat. No: 339347). Each reaction setup per well is as follows: 5µl of miRCURY LNA SYBR GREEN Master Mix, 3 µl of 1:30 diluted cDNA template, 1 µl of LNA miRNA primer assay & 1 µl of Nuclease-Free Water (hsa-miR-451a's reactions contained 2 µl of primer since it is concentration was shown to behalf of the other solutions). 96 well, low profile white 0.1 ml plates (BIOplastics, Netherlands, Cat. No: B17489) were used for the reaction, while the LightCycler 480 II (Roche, USA) was utilized measuring the Ct values. Each plate contained a single well without a cDNA template for each sample plated as a duplicate.

The reaction setup of the LightCycler is as follows: 1) 10-minute activation at 95℃, 2) 45 cycles of 15 seconds at 95℃ followed by 1 minute at 56℃ & 3) Melting Curve analysis was done between 56℃ to 90℃ with 1℃ increments.

A combination of cel-miR-39a-3p & internal control miRNA (hsa-miR-23a-3p) was used as a reference for ΔΔCt method. The geometric mean of their Ct values was treated as an internal control for normalization.

2.2.3.4 Quality Control.

2.2.3.4.1 Quality Control of isolated miRNA solutions.

For the quality assessment of miRNAs, the Agilent Bioanalyzer 2100 (Agilent Technologies, USA) was utilized. Specifically, the Agilent Small RNA Kit was utilized (Agilent Technologies, USA, Cat. No. 5067-1548). The procedure was executed according to the manufacturer's directions. Alongside quality control, the Agilent Bioanalyzer can determine miRNA concentration much better than a NanoDrop Spectrophotometer. A total of 11 samples can be analyzed simultaneously since the analysis' performance varies when all 11 samples are loaded; in some cases, samples were analyzed in duplicates.

2.2.3.4.2 Hemolysis, RNA Isolation Quality & Reverse Transcription Quality Control.

The hemolysis status of the serum samples of individuals was assessed via qPCR. It has previously been shown that the Ct difference between hsa-miR-23a and hsa-miR-451a can detect hemolysis with greater precision compared to other methods [80, 81]. According to both the manufacturer & the literature ΔCt (miR23a Ct - miR451a Ct) < 5, the sample is considered not to have undergone hemolysis. If ΔCt > 5 & ΔCt < 7, low hemolysis levels might have occurred, downstream analysis can continue with caution. If ΔCt > 7, hemolysis has occurred, the sample should be discarded from the study.

In the case of both the RNA Isolation Control (cel-miR-39a-3p) & Reverse Transcription Control (UniSp6), according to the manufacturer's recommendations, the variation of the Ct values in samples for both oligonucleotides should be less than

2%. If it is more significant than for any of them, that process was not executed properly, and the data re-analysis should occur.

2.3 miRNA Sequencing Analysis

2.3.1 Library Preparation & miRNA Sequencing.

Twelve samples were chosen to be sequenced from two controls (6 samples per control), these being Control 2 (C2) & Control 5 (C5). After miRNA isolation & Quality Control, samples were quantified using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific, USA), with the Qubit miRNA Assay (Thermo Fisher Scientific, USA, Cat. No: Q32881). The Qubit Fluorometer is the most reliable source for miRNA quantification in serum and plasma samples [82, 83]. Although this step is unnecessary, it was essential for us to see that the starting material was sufficient, which is greater than 100 ng for each sample.

The library preparation was done using the QIASeq miRNA Library Kit (Qiagen Inc., USA, Cat. No: 331502) & indexed with QIASeq miRNA NGS 12 Index IL (Qiagen Inc., USA, Cat. No: 331592). As recommended by the manufacturer, 5 µl of RNA from each of the 12 samples was the library preparation's starting material. The procedure was carried out diligently according to the manufacturer's directions, with the appropriate dilutions of reagents, again recommended in the kit's handbook. The magnetic stand was the MagJet Separation Rack (Thermo Fisher Scientific, USA, Cat No: MR02). Upon finishing, a product of a 15 µl sequencing library was obtained.

Library concentrations were measured using the Qubit Fluorometer with the hsDNA Assay Kit (Thermo Fisher Scientific, USA, Cat No: Q32854). The molarity of each library was calculated by $\frac{\left(X\frac{ng}{\mu l}\right)(10^6)}{112450} = Y\,nm$. Upon calculation, each library was diluted to 4 nM by adding (Y-4) µl Nuclease-Free Water to the 4 µl sequencing library. The library's quality control was assessed using an Agilent Bioanalyzer 2100 with the HS DNA Kit (Agilent Technologies, USA, Cat No: 5067-4626).

The sequencing reaction was to be done on a NextSeq 500 machine. Since we have 12 samples to sequence, the mid-output kit for 300 cycles (Illumina, USA, Cat. No: 20024905) was chosen, which yields a maximum of 130M reads (>10M reads allocated per sample). The final library concentrations were diluted to 1.2 nM from 4 nM according to the manufacturer's directions and multiplexed into a single tube. (NextSeq System User Guide, Illumina, USA). While sequencing the BaseSpace sequencing hub was not used, sequencing data was directly written on a sequencer's hard drive. The fastq files for each lane were written separately, resulting in 48 (4 lanes, 12 samples) fastq files. This is done to infer if a lane-specific issue might have occurred during sequencing. The sequencing reaction is arranged so that the reads are 75 bp and are single-end reads; the sequencing process takes around ~22 hours for our specific setup.

2.3.2 Processing of the sequencing reads and raw read counts.

2.3.2.1 Using CLC Main Workbench 20.

The fastq files were processed using the CLC Main Workbench 20 (Qiagen Inc., USA) and the Biomedical Genomics Analysis (Qiagen Inc., USA). The Biomedical

Genomics Analysis package contains a specific and ready-to-use workflow for fastq files, sequenced with the QIASeq miRNA Library Kit with the "QIASeq miRNA Analysis" tool. The fastq files were imported using the "Illumina read importer" tool, samples' lane reads were merged to create a single fastq file per sample while importing. The QIASeq miRNA Analysis tool processes the raw reads with all necessary steps and outputs a raw read count report. Briefly, these steps are: 1) Adapter & Quality trimming of reads. 2) Align trimmed reads to miRbase v22 fasta file, 3) Create UMI reads and eliminate ligation artifacts that might introduce bias to data. Adequate sequencing depth was conferred by calculating the UMI counts per miRNA. The total amount of UMI's a particular miRNA had is divided by the total read count of that miRNA to obtain this value. It has been previously reported that a UMI count per gene greater than 1 is adequate sequencing depth [84]. Therefore, we discarded miRNAs that have UMI counts per miRNA<1 for further analyses. The average UMI counts per miRNA per sample were calculated as a measure of overall sequencing depth. This value was expected to be greater than 3.3 [85].

2.3.2.2 Normalization of raw read counts and determination of novel miRNAs for potential WOI prediction.

Raw read counts were normalized and filtered using the DESeq2 and edgeR, R packages described in **Section 2.1.2.2**. Two separate approaches were used to identify certain miRNAs predicting the WOI regardless of when the LH surge occurs. The first was to identify differentially expressed miRNAs using the DESeq2 algorithm. Since the number of samples per endometrial phase is relatively low, the samples were grouped as Pre-Receptive (PRO+ES) & WOI. Differential expression was determined as adjusted (BH corrected) p-value < 0.05 & |L2FC| > 1. The second method was to identify statistically significantly correlated miRNAs between Control 2 and Control

5 (R>0.56) miRNAs that also showed 1.5 FC (~0.7 |L2FC|). The DESeq2 normalized miRNA expression values of the six samples of each control were correlated using Microsoft Excel. The %CV was calculated for each miRNA per patient with the formula: $\mathrm{CV} = \frac{\bar{X}}{\mathrm{SD}}$. miRNAs correlated and had high CV (>15%) were considered for the analysis. Potential novel endogenous miRNA controls were also inferred using this strategy.

2.4 Generation of mRNA-based transcriptomic signature for WOI prediction.

2.4.1 Determination of the discovery and validation datasets.

We wanted to generate our signature from RNA-Seq experiments exclusively, as it can screen the full transcriptome, not just ncRNA, at the same time compared and with higher confidence compared to microarrays. Datasets that compared the proliferative (PRO) and early secretory (ES) phases of the endometrial cycle to the WOI were considered. Below is a table (**Table 2.2**) summarizing the datasets that were used in this study. The validation dataset was selected based on the criteria that should contain all the different stages of the endometrial cycle. Only specific microarray experiments matched this criterion. The one with the most significant number of samples was considered as the validation cohort.

Table 2.2 Summary of the discovery and validation datasets**.**

GEO Accession Number	Sample Number	Endometrial Stages	Technology	Year
GSE86491 [63]	14	ES & WOI	Illumina RNA-Seq	2011
GSE98386 [86]	40	ES & WOI	Illumina RNA-Seq	2013
GSE132711 [87]	13	PRO & WOI	Illumina RNA-Seq	2016
GSE111976 [88]	6	ES & WOI	Illumina RNA-Seq	2018
GSE29981	20	E**.**PRO, L**.**PRO, ES, WOI	Affymetrix microarray (hg133plus2.0)	2011

2.4.2 Processing of expression data.

The raw read counts of GSE86491, GSE98386 & GSE111976 were downloaded from the GEO webpage. GSE132711 fastq files were downloaded and processed accordingly. Principal Component Analysis (PCA) was done to assess any present batch effects that the studies may have. PCA was done with R package "ggbiplot." For each dataset, DESeq2 was used to calculate differential gene expression. Genes that had BH correct p-value < 0.05 & |L2FC|>1.5 was considered as differentially expressed. Pre-filtration was done on each dataset as described in **Section 2.1.2.2.** GSE86491 was the only dataset that had no requirement for pre-processing. The downloaded raw read counts were analyzed directly, as stated above.

GSE111976 sequencing data was obtained from raw single-cell sequencing, which contains sequencing information from up to 6 different cell-types. To get an

expression profile of the endometrium and the recommendations from the authors of the relevant study, the raw read counts were merged by summing each sequencing result by sample. This yielded three samples in the early secretory (ES) phase, while three samples were in the WOI. These files were used in downstream analysis.

In GSE98386, the researchers reported two distinct batches in the study's series matrix file. After DESeq2 normalization and read count transformation, this expression data was used for batch correction [89]. Briefly, after data filtration and DESeq2 normalization, the batches were merged, resulting in 20,027 genes. The batches were corrected using the "ComBat" function of the "sva" R package [90]. Re-visualization was done on the batch corrected data. Differential gene expression on the batch corrected data was done via limma [91] package in R following the protocol described by Law and colleagues [92]. DESeq2 could not be used as the data was not raw read counts.

Finally, for the GSE132711 dataset, the raw sequencing counts were not readily available. Therefore, the fastq files were downloaded from EBI-ENA. Reads were quality controlled with "fastqc," & quality trimming was done with TrimGalore. Briefly, the adapter sequence was inferred by the program automatically. This was identified as Illumina Universal 3' Adapter, the sequence being: "AGATCGGAAGAGC." The average read quality was determined as Phred score => 30. After adapter and quality trimming, reads that were shorter than 20 basepairs were discarded. The reads were aligned to GRCh38.p13 via the HISAT2 read aligner [93]. The default parameters were used for the alignment. Duplicate reads were marked and removed using the "MarkDuplicates" function of PicardTools on the Ubuntu terminal.

The validation stringency was set to “LENIENT” instead of the default value. The raw read counts were obtained using featureCounts [94], with default parameters. After obtaining the raw read counts, differential gene expression was obtained, as stated above.

For the validation dataset, GSE29981, an end to end workflow for differential gene expression for Affymetrix microarray data was used [95]. This workflow is available as a Bioconductor package titled “maEndToEnd.” The raw data were downloaded from the ArrayExpress database via the accession “E-GEOD-29981”. After downloaded, the data is normalized using the RMA method, using the “oligo” package in R. This is a one-step process, as the function, background corrects the data, normalizes it, and log2 transforms it giving it an expression value. After normalization, a histogram was drawn to determine where an appropriate expression cut-off could be made at. Assays that did not have signal intensities larger than 2 in 18 samples out of 20 were discarded. Out of 54,675 probes, 36,951 remained. Annotation of the probe-ID’s were done using the “AnnotationDbi” & “pd.hg.u133.plus.2” packages. Probes with multiple gene names assigned to them were discarded, resulting in 1758 probes being discarded, leaving 35,223 probes as the final expression set. Upon visualization of the PCA data, a single sample was an outlier; this sample was discarded from the analysis, leaving 19 samples. It was decided to do a better clustering rather than menstrual cycle day of different patients. Therefore, the k-means method was used to generate three distinct clusters on the PCA data to generate a more reliable validation dataset. These three clusters would be the proliferative phase (PRO) with n=11, the early secretory phase (ES) with n= 4 & the window of implantation (WOI) with n=4. In the case of multiple probes assigned to a

single gene, the probe with the highest %CV was considered the representative probe for that gene.

2.4.3 Determining the gene signature for WOI prediction.

After differential gene expression values were determined, the differential genes were grouped according to their fold-change direction, creating eight distinct groups (4-dataset number * 2 – Fold Change direction). According to their adjusted p-value (smallest p-value is 1) and |L2FC| values (most significant fold change is 1) separately, each gene was ranked in ascending order within their group. The datasets that compare the same phases with the WOI were cross-referenced with one another (GSE86491 & GSE132711 – PRO v WOI) (GSE98386 & GSE111976 – ES v WOI). Genes that had conflicting directions in differential gene expression were identified. The gene with the lowest rank in either its significance or L2FC rank was identified for both upregulation and down-regulation in each dataset. These were considered thresholds for that data sets' specific group. This means there would be eight distinct thresholds for our analysis if all datasets had conflicting fold changes in every direction. This was not the case as there were no conflicting genes between GSE98386 down-regulated genes and GSE111976, resulting in 6 different thresholds. Shared genes between the two datasets that did not have conflicting fold changes were considered for further analysis. However, those with lower ranks among these genes than the thresholds were discarded for each dataset group. These were the final candidate genes for the prediction algorithm.

The groups were merged according to the phase difference and direction. Each group contains information from two distinct datasets. This information is the significance

rank and L2FC rank of that specific gene, according to their rank-sum values (sum of dataset one significance rank, dataset 1 L2FC rank, dataset two significance rank & dataset 2 L2FC rank). The top 5 genes were selected from each group (PRO-UP, PRO-DOWN, ES-UP & ES-DOWN), resulting in a 20-gene signature.

2.4.4 Testing the predictive value of the 20-gene signature.

2.4.4.1 Creation of the discovery dataset.

None of the discovery datasets used had sufficient samples or contained all the endometrial cycle stages. Therefore, they merged using Z-score standardization to merge the gene expression info from all four datasets [96]. This resulted in a combined dataset with 71 samples: 13 samples in PRO, 23 samples in ES, and 35 samples in the WOI. This dataset had 8728 genes. Genes from the 20-gene signature that was not amongst the 8728 were replaced by the next gene for that group, thus recreating the 20-gene signature.

2.4.4.2 Determination of the prediction of the gene signature.

A single sample matrix (SSM) approach was taken for the prediction method. Briefly, the average expression value for each gene of the signature is calculated for each phase of the endometrial cycle (PRO-ES-WOI) from the discovery cohort. The resulting 3 x 20 matrix is the SSM, which will predict the validation class. The validation dataset did not contain three genes of the 20-gene signature (DKK1, MYOCD & TSPAN8); this resulted in a final 17-gene signature, titled Novel Endometrial Receptivity Signature – 17 **(NERS17)** from now on. The SSM values of the different phases NERS17 signature were correlated with the validation cohort's expression values separately. The prediction of an individual sample to a particular phase was

determined by the phase which yielded the highest correlation coefficient out of the three phases. Relevant heatmaps regarding the analyses were drawn with CLUSTER 3.0 & Java TreeView program.

CHAPTER 3

RESULTS

3.1 In silico determination of candidate miRNAs.

The work of Kresowik et al. [61] has shown that specific miRNAs show parallelity, in terms of their expression profiles, between the endometrial tissue and serum, between the P phase and WOI. This finding reinforces our hypothesis that miRNAs could be behaving this way, not discovered. Therefore, our first goal was to identify miRNAs that show altered expression during the WOI compared to the different stages of the endometrial cycle.

After identifying the four distinct datasets to be used in our analyses (**Section 2.1.1), t**o identify potential miRNAs for WOI prediction, we had to determine miRNAs that showed significant differential expression (DEmiRs) between the phases of the endometrial cycle. Differential expression analysis yielded 29, 95, 128 & 89 DEmiRs for the four datasets analyzed (**Figure 3.1)**.

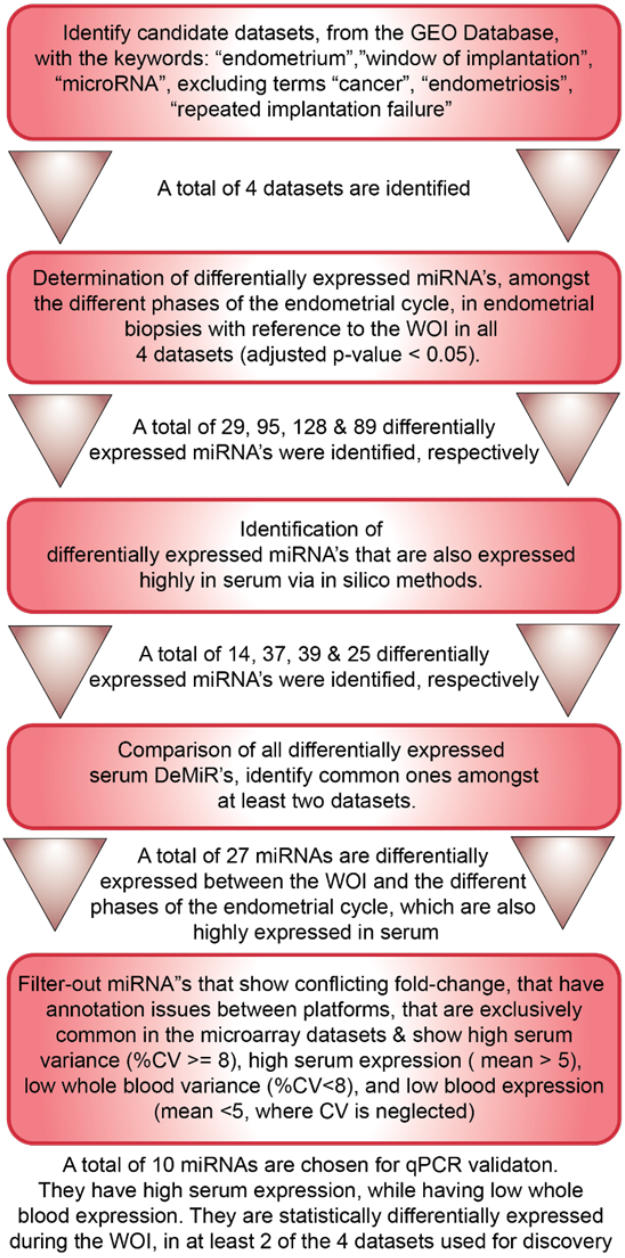


Figure 3.1 Identification of 10 candidate miRNAs as candidates of WOI prediction from serum in healthy cycling women. The bioinformatical workflow of identifying miRNAs with a distinct expression profile in the endometrium is also highly expressed in serum samples.

Before moving, we wanted to identify in GSE34435 [57], specific miRNAs that showed different expression patterns that might have been missed during differential

expression analyses. Out of the five different sample collection days, probes that (4 from pre-receptive endometrium and one from the WOI), probes that had |L2FC|>2 were graphed. This analysis yielded six distinct miRNA profiles (**Figure 3.2**), further reinforcing our claim that distinct miRNA patterns occur in the endometrium. A combination of these miRNA profiles could predict the WOI if these miRNAs also behave accordingly in the serum.

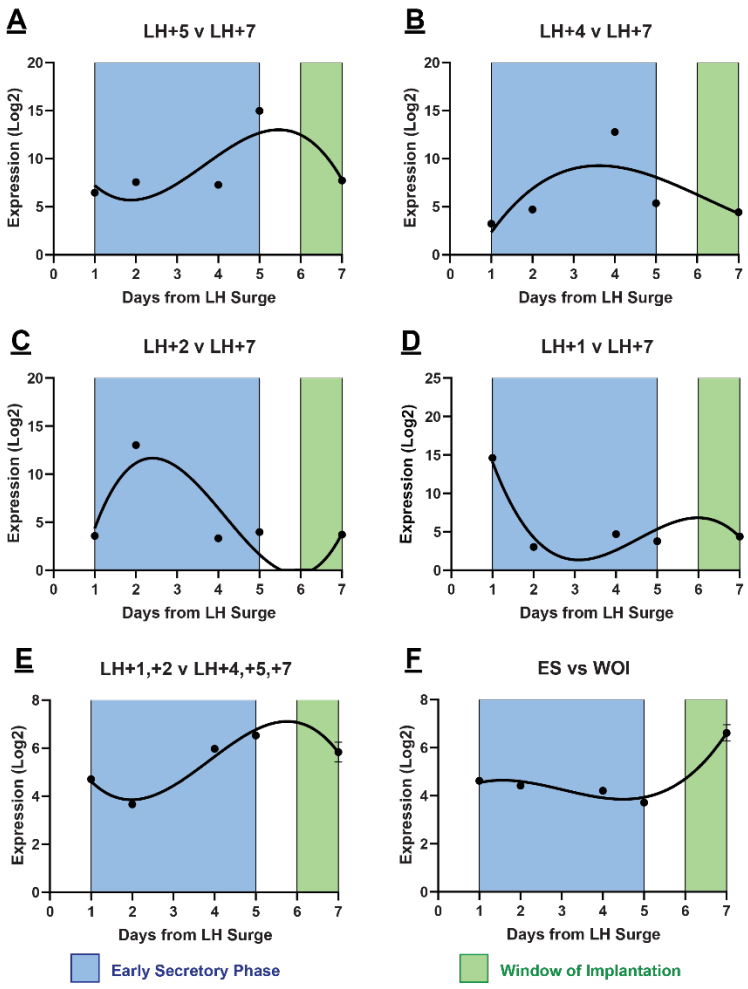


Figure 3.2. Graphs of probes of individual miRNAs in GSE34435. A) 7 probes were determined to behave similarly. B) 16 probes behave similarly. C) 22 probes behave similarly. D) 23 probes behave similarly. E & F) 1 probe was found to behave these ways.

After, determination of DEmiRs, it was checked how much of these were expressed in the serum. An independent study was found, GSE100467, that contained RNA-expression data from 39 patients. One hundred two miRNAs were identified, with high confidence, expressed in the serum [75]. These miRNAs were cross-referenced with endometrial DEmiRs, resulting in 14,37,39 & 25 serum DEMiRs (**Table 3.1)**

Table 3.1 Summary table of the datasets and their results.

GEO Accession	Tissue Type	Endometrial Cycle Phase				# DEmiRs	# DEmiRs in serum
		P	ES	WOI	LS		
GSE34435	EB					29	14
GSE44558	EFA					95	37
GSE86491	EB					128	39
GSE109053	EB					89	25

miRNAs were then cross-referenced with one another to see if there was one that was common in all four datasets. This was not the case, however, as no miRNA was shared between all four datasets. Then we focused on miRNAs that were common in at least two separate datasets. As a result, 27 miRNAs remained **(Figure 3.3).**

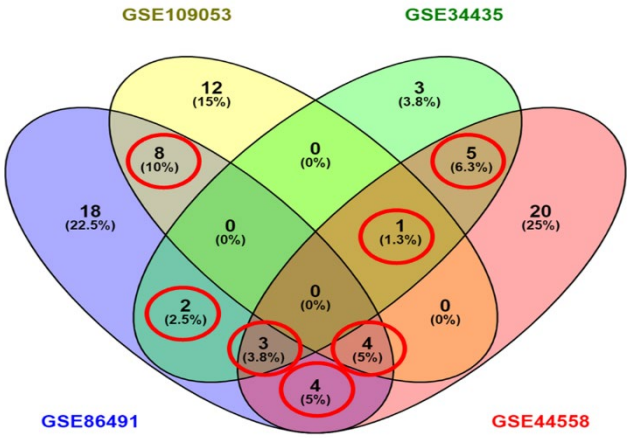


Figure 3.3. Venn Diagram representation of a comparison of serum DEmiRs between datasets.

Finally, the 27 miRNAs were filtered out depending on specific criteria. miRNAs that showed annotation problems (1 miRNA) were common in only microarray studies (5 miRNA) had different fold change direction in the same phase comparison (2 miRNAs). The remaining 19 miRNAs were filtered according to whole blood expression. Our focus was to identify miRNAs that are variable in the serum and have high expression in the serum while having low blood variance. This was prioritized as we wanted to miRNAs affected by hemolysis less than others on the list. Hence, miRNAs had $\geq 8\%$ CV in serum and CV > 8% in blood and had high serum expression values. Out of 19, only 5 met these criteria.

We then decided to lift the serum expression filter; this introduced three more miRNAs. Finally, we decided if blood expression is low then, blood CV can be neglected, as its effect on expression would be minimal; this introduced two new miRNAs, finalizing the list with ten miRNAs. These miRNAs were: hsa-miR-375-3p (**Figure 3.4.A**), hsa-miR-148a-3p (**Figure 3.4.B.**), hsa-miR-30b-5p, hsa-miR-30a-5p, hsa-miR-199a-3p, hsa-miR-146b-5p, hsa-miR-150-5p, hsa-miR-1307-3p, hsa-miR-130b-3p & hsa-miR-29a-3p **(Supplementary Figure 1)**.

To get a better insight into these ten miRNAs. They were, investigated using network analysis tools called miRNet [97]. This tool checks the input miRNA list's common targets, using several miRNA-mRNA interaction databases (**Figure 3.5**). Results from our analyses yielded that the top 3 biological processes done with KEGG pathway analyses [98] involved in the targets of these 10 miRNAs were; 1) Focal Adhesion (adj.p.value = 10^{-5}), 2) Progesterone Mediated Oocyte Maturation (adj.p.value = $5.7 * 10^{-4}$) and 3) Endometrial Cancer (adj.p.value = 0.0114). All of them are related to

endometrial function. This finding further validates the potential strength of our selected miRNAs.

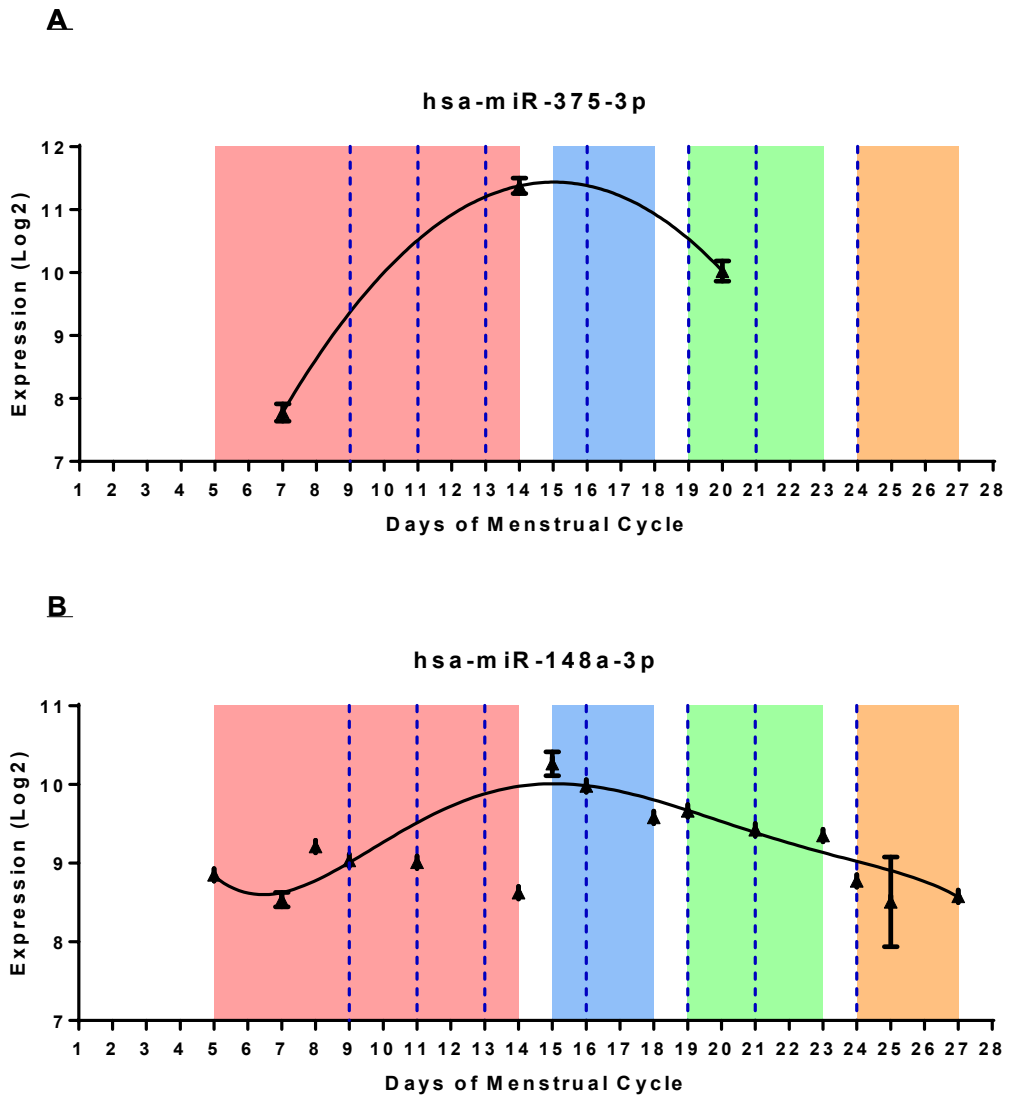


Figure 3.4. in Silico profile of miRNAs are dynamic in endometrial tissue. A) hsa-miR-375-3p is increasing between the P phase and WOI but decreasing between ES and WOI. B) hsa-miR-148a-3p shows a direct and dramatic P to ES transition and a consistent decrease until the M phase.

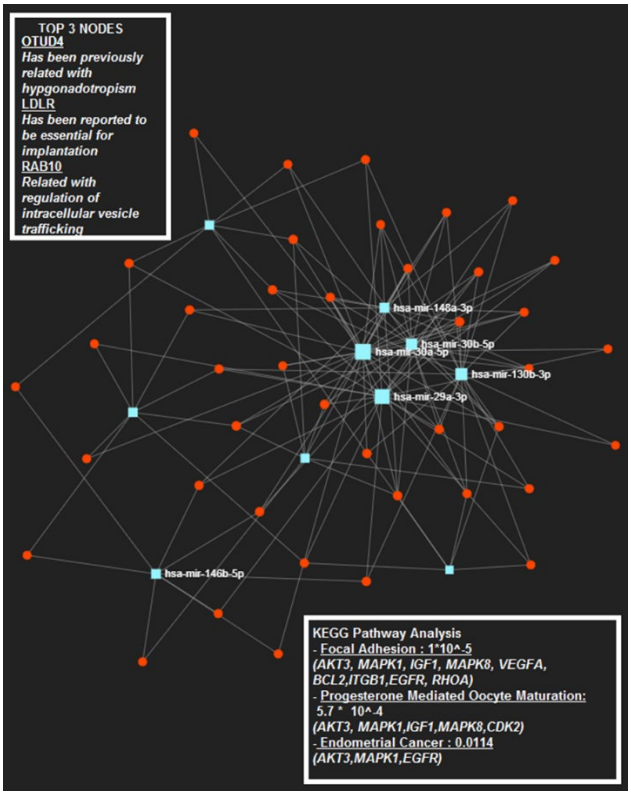


Figure 3.5. Network Analysis of the candidate ten miRNAs. The top three genes that are commonly targeted are OTUD4, LDLR, and RAB10. The latter two have been previously implicated in implantation. The significantly enriched pathways are also related to the endometrial tissue and implantation.

3.2 Validation and identification of miRNAs for potential WOI prediction.

We wanted to test one of our central hypotheses: miRNAs that behave in serum samples as they do in endometrial tissue. We did not have paired endometrial tissue samples with serum samples; instead, our *in-silico* candidates were chosen for future validation with RNA-Seq and qPCR.

Additionally, the study's primary goal was to identify a serum miRNA signature, irrespective of LH Surge day. As mRNA transcripts predict the WOI with better

accuracy than LH Surge, it is well within reason for miRNAs to assume that miRNAs also could act similarly. Another advantage of this would be in patients with impaired or displaced WOI; the LH Surge is useless in predicting the WOI. As we have only ten miRNAs of candidacy out of more than 2500, we decided to do a miRNA sequencing experiment. Unfortunately, we were only able to sequence two patients' serum during this time. Even in that case, we have gotten some promising results.

To achieve an overall profile of all miRNAs, we wanted to collect more serum samples than other studies. Therefore, we collected samples from different days of the menstrual cycle, reflecting the changes occurring during the endometrial cycle. Days 9 and 11 were for the P/LP phase. Days 13,16 were for the ES phase; Days 19 & 21 were for the WOI, and day 24 was for the LS phase of the endometrial cycle.

3.2.1 Detection of the LH Surge and Hemolysis rates in serum samples

3.2.1.1 LH Surge detection of serum samples.

Determining the LH Surge of patients was critical to us, as it allowed us to potentially determine miRNAs that acted according to the LH Surge while also identifying those that behave independently.

Serum baseline LH levels were previously shown to be variable tremendously amongst different people. This phenomenon was observed in our results as well (**Figure 3.6).** Differences almost of more than 28-fold in baseline LH levels were observed (min = 0.119 pg/ml, max= 3.438 pg/ml). Additionally, Control 3, Control 8 & Control 10 **(Figure 3.6C, H, J)** of our control samples did not show the LH Surge. The most likely reason for this is that the LH Surge occurred when serum was not

collected from the participant. However, 7/10 participants showed a clear LH Peak (**Figure 3.6).** It was also observed that the baseline of overall LH concentration varied amongst the control samples. In conjunction, the LH peak concentrations of high baseline LH individuals is higher **(Figure 3.8.A)**.

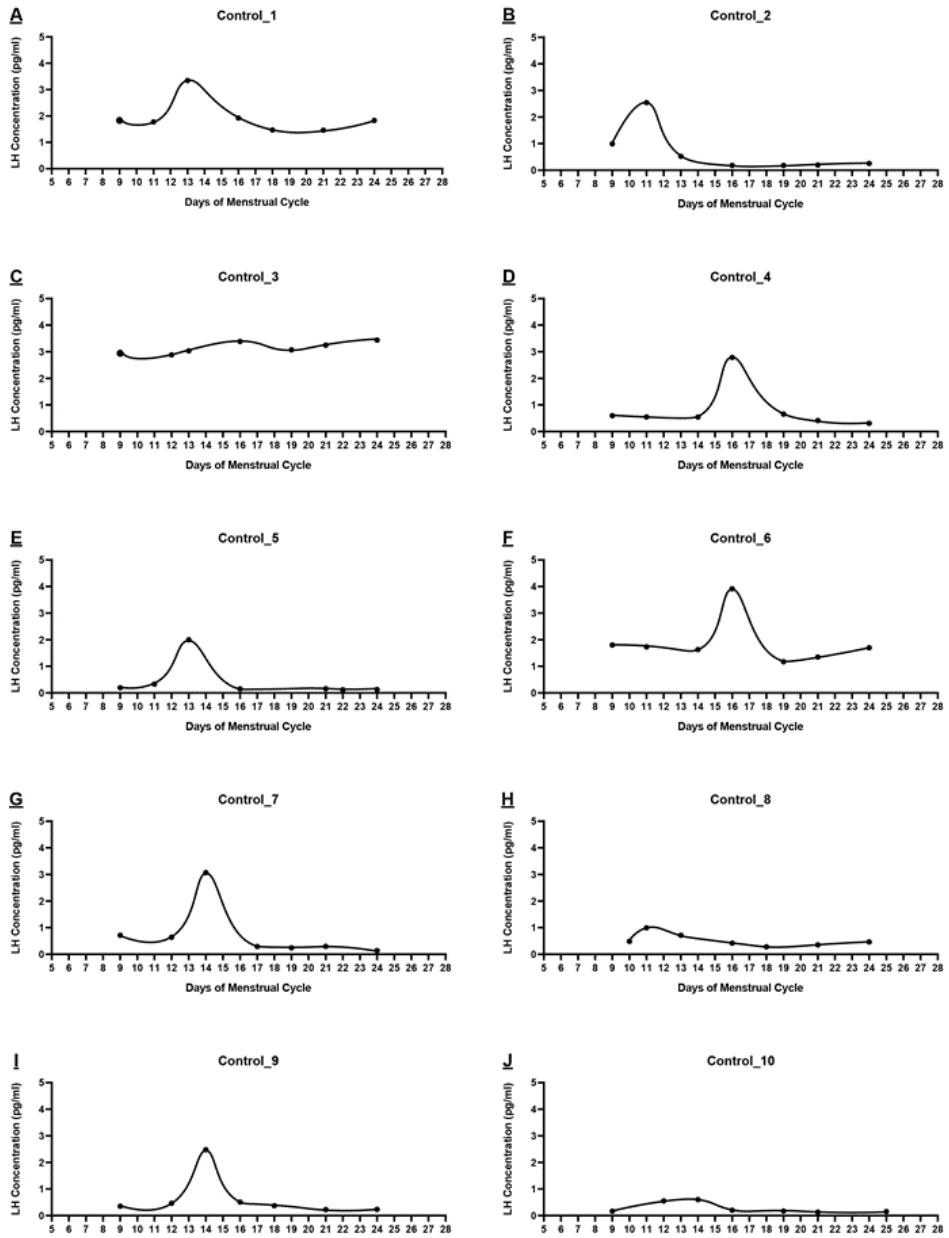


Figure 3.6. LH concentrations of controls' days of serum collection. Control 3, 8 & 10's LH peak could not be adequately detected.

3.2.1.2 Determination of hemolysis in samples.

Hemolysis detection was crucial for understanding which patients' sample can be used in further downstream analysis. C2 and C5 samples were chosen for downstream sequencing analyses. Therefore, hemolysis of these serum samples needed to be evaluated. Although C5 seemed to have more hemolysis, both samples could be considered safe for usage as no Ct difference was more significant than 7 **(Table 3.2)**.

Table 3.2. Hemolysis values of C2 and C5 serum samples.

Sample ID	Hemolysis Control
C2.16	4.12
C2.19	4.06
C2.21	4.88
C2.24	4.58
C5.9	5.26
C5.13	5.45
C5.16	5.68
C5.21	6.47
C5.22	5.62
C5.24	5.57

3.2.2 miRNA-Sequencing Analyses.

3.2.2.1 MiRNA Sequencing Quality Control

MiRNA sequencing procedure was done as described in **Section 2.3.** The libraries that were going to be sequenced showed two distinct peaks in their Agilent Bioanalyzer tracks (**Figure 3.7)**. This is important as in the library preparation kit manual, the peak observed at 157 bp is stated to be an adapter-dimer, while the 180 bp peak is the miRNA peak. In this case, it is required to do a gel extraction on the miRNA Library; however, we could not do this due to scheduling conflicts of the sequencer.

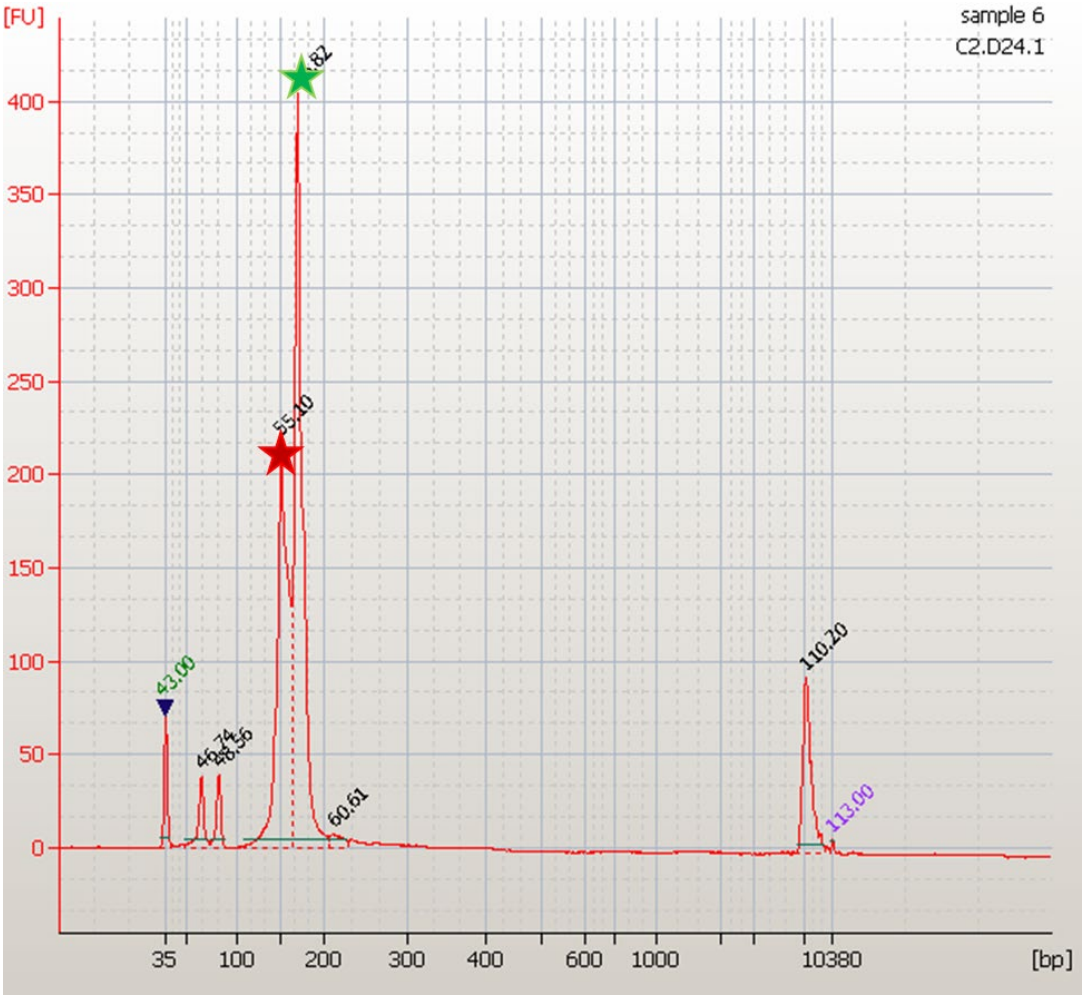


Figure 3.7. Agilent HS DNA Electrogram image of C2.D24 sample. Although this electrogram is for one sample, each sample's electrogram had the same adapter-dimer peak (red-star) at 157 bp while a strong miRNA peak at 180 bp (green-star).

This resulted in small library sizes after processing was done of the reads **(Table 3.3)**. The low number of library sizes suggests that most sequencing reactions sequenced the adapter-dimers instead of the miRNAs. However, the average UMI counts per miRNA (UCpM) was calculated to infer each sample's sequencing depth. This showed that almost all the samples sequenced have adequate sequencing depth, as each sample, except C2.D24, has UCpm>3.50. This concludes that even though the retained library size is relatively small per sample, the data obtained is viable for

further analysis. However, in the future, this fundamental aspect will be focused on in greater detail.

Table 3.3 Final library sizes after analyses.

Sample_ID	Initial Library Size	Final Library Size	% of retained library	Average UMI counts per miRNA
C2-D10_S1	8,984,137	1,567,419	17.4%	4.38
C2-D13_S2	8,319,955	1,536,643	18.5%	3.72
C2-D16_S3	7,371,767	1,839,125	24.9%	4.06
C2-D19_S4	8,436,434	2,814,513	33.4%	4.11
C2-D21_S5	8,821,565	2,720,757	30.8%	3.85
C2-D24_S6	8,123,170	3,196,961	39.4%	3.46
C5-D9_S7	6,606,418	1,943,002	29.4%	4.32
C5-D13_S8	7,227,596	1,845,784	25.5%	4.38
C5-D16_S9	8,109,182	2,494,944	30.8%	3.79
C5-D21_S10	10,615,795	5,016,331	47.3%	4.44
C5-D22_S11	6,987,707	2,918,124	41.8%	4.16
C5-D24_S12	8,312,580	2,306,947	27.8%	4.20

3.2.2.2 miRNA expression in serum of cycling healthy women.

Out of our candidate miRNAs, only two of them showed considerable variation in the serum samples. The first one being, hsa-miR-375-3p. Although this miRNA showed a change in its RNA-Seq expression profile **(Figure 3.8A)**, it had inconsistent results with its *in-silico* expression profile **(Figure 3.4A).** The *in-silico* prediction showed a dramatic increase in expression from the PRO phase to the ES phase, then a slight decrease during the WOI. The RNA-Seq experiments showed a gradual decrease from the PRO phase through the ES phase and the WOI. Even though there is parallelism between the expression levels of the prediction and the RNA-Seq experiment between

the ES and WOI, as they both decrease during WOI, this decrease is not observed in the RNA-Seq results.

Additionally, it is observed that the 22nd day of Control 5 (C5.D22) shows a dramatic increase in expression value, which is not observed in Control 2. This may occur because, on the 22nd day of the menstrual cycle, a phenomenon may occur that we cannot report in Control 2. After all, no serum sample is collected that day. Regardless of the inconsistency between samples and the *in-silico* prediction, it is crucial to understand that a particular miRNA's serum profile may not be the same as the endometrial tissue. hsa-miR-375-3p still shows a consistent and steady decrease during the PRO and ES phases in healthy cycling women's serum samples. This is an important finding as this miRNA still holds potential in being used as a biomarker for endometrial receptivity in serum.

The second miRNA that showed variation was hsa-miR-30b-5p. Interestingly between the two patients, it was shown to have opposite behaviors **(Figure 3.8B)**. However, Control 2 showed consistent results with the *in-silico* prediction of this miRNA **(Supplementary Figure 1A)**. A gradual increase in the expression profile from the start of the endometrial cycle, peaking at the WOI. Regardless of the inconsistency between Control 2 and Control 5, the variation observed, and Control 2 is consistent with *in-silico* predictions, suggests that specific miRNAs can behave as they do in the endometrial tissue in healthy females.

Out of the ten candidate miRNAs, other than these two, no other showed any significant difference during the menstrual cycle. To hsa-miR-130b-3p was chosen **(Figure 3.8C)**; however, the remaining seven behave similarly (data not shown). One

thing of importance is that, as seen in all three miRNAs, C5.D22 shows a dramatic difference either with high expression or low expression. As stated above, this might be due to a biological phenomenon, but possibly due to sample degradation. This does not change the fact that few candidates (20%) showed variance during the menstrual cycle. So few miRNAs were captured because our *in-silico* identification of miRNAs, although significant, did not have a lot of fold change in the tissue, to begin with **(Supplementary Figure 1)**. Therefore it is not illogical to see these expression values in serum.

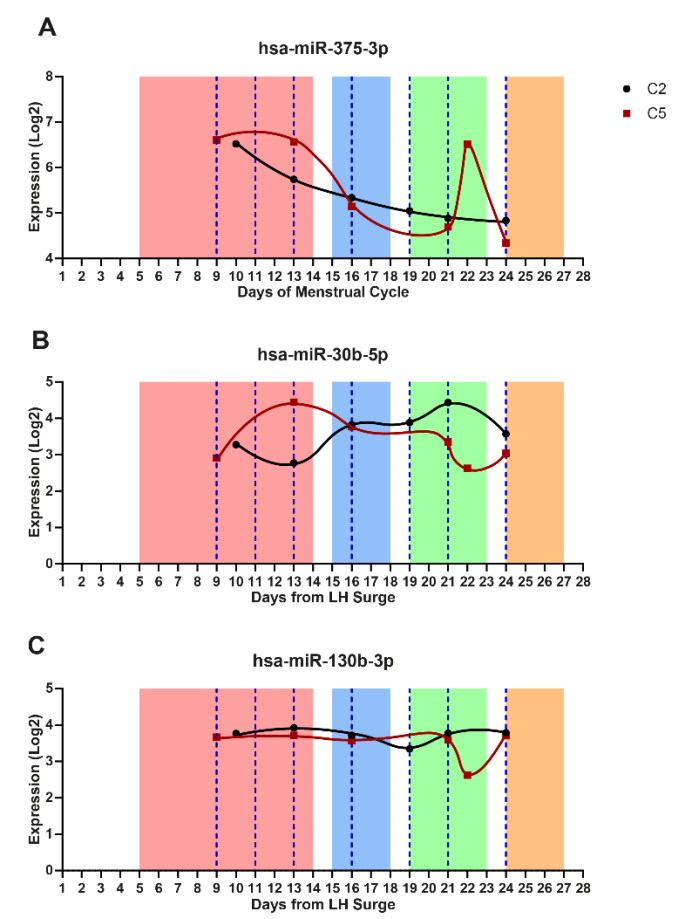


Figure 3.8. *In-silico* methods determined the serum profile of individual candidate miRNAs. A) Out of all ten miRNAs, this miRNA was the one that behaved most consistently. A decrease is observed after the ES phase towards the WOI. B) hsa-

miR-30b-5p was another miRNA that exhibited variation in serum. However, C2 showed a profile close to *in-silico* prediction, C5 exhibited opposite behavior. C) The remainder of the candidate miRNAs behaved like has-miR-130b-3p, essentially non-varying throughout the menstrual cycle.

We have identified seven novel miRNAs that show similarity in their behavior between the two controls and show fold change between the different phases of the menstrual cycle **(Figure 3.9)**. These seven miRNAs are separated into four groups: **1)** miRNAs that are statistically differentially expressed between Pre-Receptive (PRO & ES) and Receptive (WOI) phases of the endometrial cycle **(Figure 3.9A)**: hsa-miR-183-5p & hsa-miR-106-5p, **2)** miRNAs that can potentially distinguish between PRO-ES **(Figure 3.9B)**: hsa-miR-2277-3p & hsa-miR-148b-5p, **3).** miRNAs that can potentially distinguish between ES-WOI **(Figure3.9C)**: hsa-miR-7108-3p, **4)** miRNAs that can distinguish potentially between both PRO-ES & ES-WOI **(Figure 3.9D)**: hsa-miR-4730 & hsa-miR-615-3p. Out of these seven, the last group members, hsa-miR-4730 & hsa-miR-615-3p, show the most promise. However, it is essential to note that the remaining five might increase these miRNA's predictive power when used in conjunction with one another. These miRNAs exhibit a unique and very dynamic expression pattern during the endometrial cycle of serum samples. The functional analysis of these miRNAs and their targets has not been performed yet.

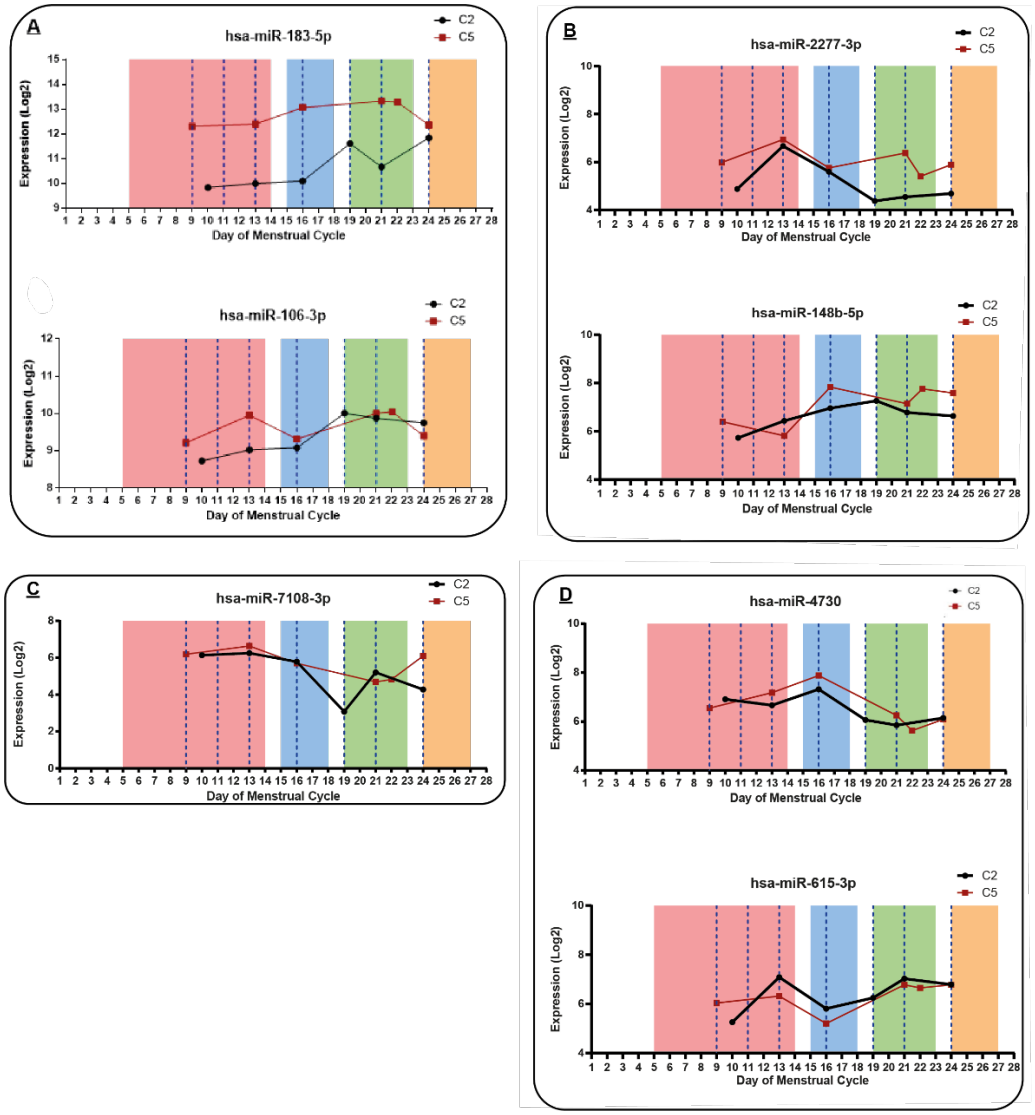


Figure 3.9. Seven novel miRNAs that show distinct profiles of miRNA expression. These miRNAs can be used as potential biomarkers of the WOI, with the possible combination of hsa-miR-375-3p and hsa-miR-30b-5p.

3.2.3 Quantitative Polymerase Chain Reaction (qPCR) for Validation of RNA-Seq experiments.

As to double-check our findings obtained via RNA-Seq, qPCR was performed on serum samples of C2 and C5. Only hsa-miR-375-3p and hsa-miR-30b-5p expressions were measured for the two controls (C2 & C5).

Upon Ct value retrieval, the ΔΔCt method was used to calculate the two miRNAs' relative expression. The RNA-Seq and qPCR measurements of Control 2 **(Figure 3.10A)** and Control 5 **(Figure 3.10B)** were compared with each other with correlation for each miRNA. The results show a significant correlation (R> 0.56) between the RNA-Seq measurements and the qPCR results for both miRNAs in the separate controls. This confirms that technical issues do not compromise the RNA-Seq readings and that they can be trusted. It also shows that C5 behaves differently in terms of its miRNA expression to C2 and the *in-silico* predictions. Finally, and most importantly, it also shows that, even though miRNA qPCR is challenging to optimize, it is a valid option for miRNA-Seq validation. **(Figure 3.10)**.

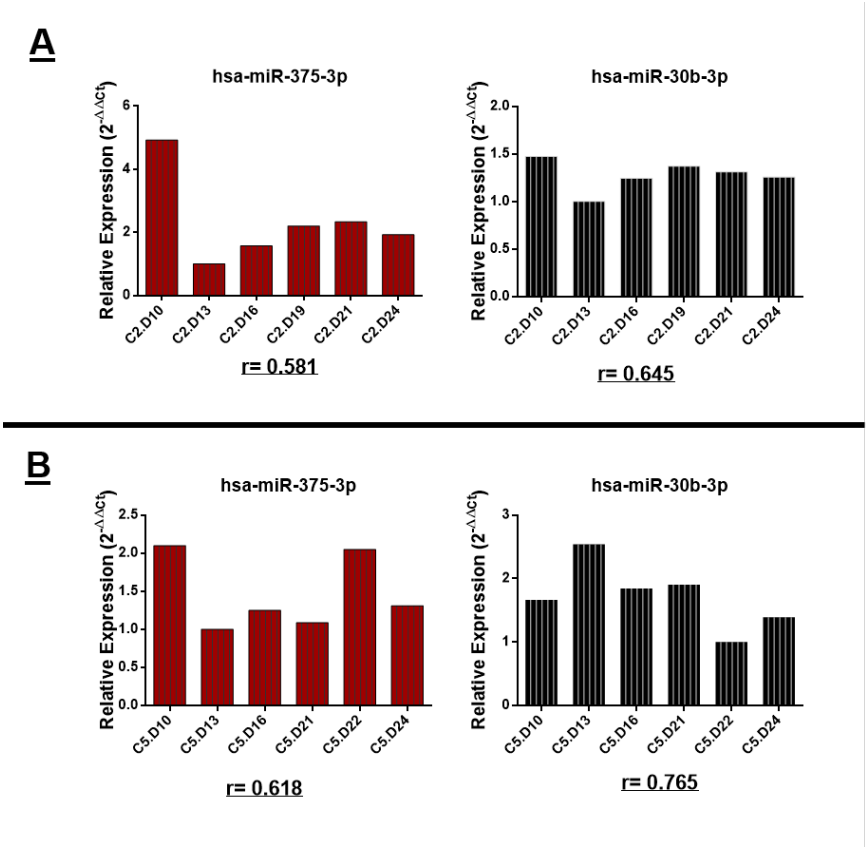


Figure 3.10. The qPCR method was able to validate the expression profile of has-miR-375-3p and hsa-miR-30b-5p. . **A)** qPCR expression values of both hsa-miR-375-3p & hsa-miR-30b-5p for Control 2 has shown significant correlation (R>0.576) with these miRNAs' RNA-Seq expression values. **B)** qPCR expression values of both hsa-miR-375-3p & hsa-miR-30b-5p for Control 5 has shown significant correlation (R>0.576) with these miRNAs' RNA-Seq expression values.

Overall, our RNA-Seq experiment was able to identify novel miRNAs that could potentially be used in WOI prediction. Two of the miRNAs identified via *in-silico* methods were validated and showed a dynamic profile in the endometrium to be considered in the future. It must be kept in mind that experiments were done with two controls; the results obtained show promise and bear the potential to be used in WOI prediction in the future.

3.3 Generation and Validation of an *in-silico* mRNA based novel endometrial receptivity signature (NERS-17)

3.3.1 Identification of genes for NERS-17

3.3.1.1 Principal Component Analysis of Discovery Datasets

The discovery and characteristics of the datasets were described in full detail in **Section 2.4.1**. To see if these datasets (GSE86491, GSE98386, GSE132711 & GSE111976) do not contain any technical batches or see if our batch correction methodology had successfully removed the technical batches. GSE86491, GSE132711 did not show any batches and therefore were used in the downstream analysis **(Figure 3.11A, C)**. GSE98386 showed two batches **(Figure 3.11B-above).**

This batch effect was reported by the authors as well. However, after batch correction using the ComBat function from the sva Bioconductor package, the batch effects were corrected **(Figure 3.11.B-below).** Finally, GSE111976 contained different phases of the cycle; this was seen in the PCA graph. As we intended only to use the six ES and WOI samples, their separation was sought out and observed clearly **(Figure 3.11D).**

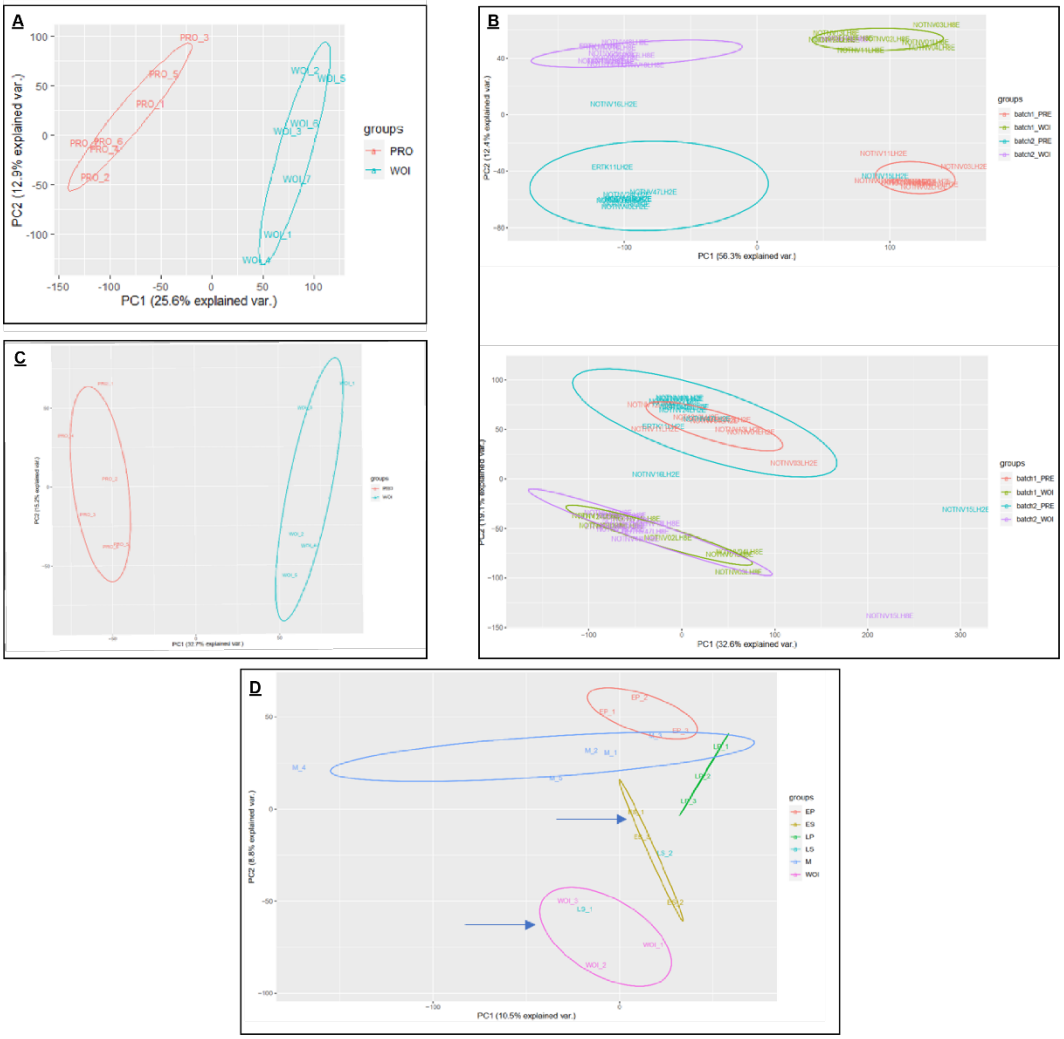


Figure 3.11 The PCA graphs of the discovery datasets. A) PCA graph of GSE86491, a clear separation between the PRO stage (red) and the WOI (blue) is observed. **B)** PCA graph of GSE98386. There were two batches separated from PC1 in the original data (above). After batch correction via ComBat showed no batch effect, separation of the endometrial stages is clear (below). **C)** PCA graph of

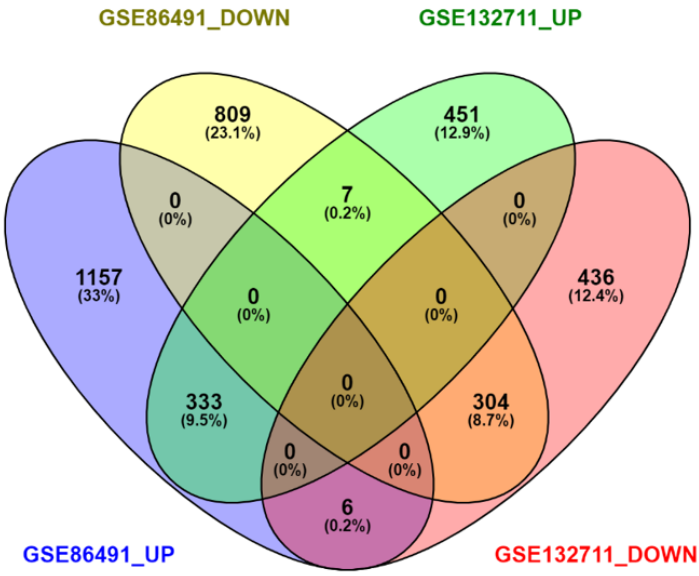
GSE132711, a clear separation between the PRO stage (red) and the WOI (blue) is observed. **D)** PCA graph of GSE111976. This dataset contains six separate endometrial phases, all of which are separated. Since this dataset will be used for ES v WOI, those data are separated clearly (arrows).

3.3.1.2 Identification of genes to be used for signature generation.

Upon visualization, differential gene expression analysis was done via DESeq2 on all four discovery datasets. This yielded a total of 1496 up- and 1120 downregulated genes for GSE86491, 805 up-, and 678 down-regulated genes for GSE98386, 791 up-, and 746 downregulated genes for GSE132711 and 342 up- and 306 downregulated genes for GSE111976. We later ranked the differentially expressed genes regardless of their direction based on their significance and |L2FC| values, separately. This was done as to determine the genes that had the most statistical and practical importance.

To get a final set of genes to choose the gene-list from, we decided to further the analysis with genes common between the two datasets that compare the same endometrial stage with the WOI. Therefore, to understand the genes showing the difference between the PRO and WOI (PRO genes), we compared the differentially expressed genes between GSE86491 & GSE132711. Similarly, to identify genes that show the difference between ES & WOI (ES genes), GSE98386 & GSE111976 differentially expressed genes were compared. The comparison to identify PRO genes yielded 333 genes that are upregulated (PRO_UP) and 304 genes that are downregulated (PRO_DOWN). 13 (7 in GSE86491_UP v GSE132711_DOWN) genes showed conflicting L2FC directions **(Figure 3.12).** The highest rank of either the significance rank or the |L2FC| rank was taken as cutoff **(Table 3.4).** The cutoffs for GSE86491_UP, GSE86491_DOWN, GSE132711_UP & GSE132711_DOWN is

respectively 637,289,482,613. The common genes with a lower rank than these are filtered accordingly to get a final list of PRO_genes.



PROLIFERATIVE PHASE

Figure 3.12 The comparison of GSE86491 and GSE132711 to obtain the PRO genes. A total of 637 genes (333 up- and 304 downregulated were considered for further analysis). The conflicting genes were used to determine a cutoff for further filtering of the 637 genes.

Table 3.4 The ranks and threshold values of the PRO genes

SIGNIFICANCE RANK

GSE86491_UP vs GSE132711_DOWN	86_UP	132_DOWN
KIAA1324	**845**	**263**
SEMA3G	1128	309
LCN12	1173	522
AC008763.1	1367	612
PKHD1L1	1398	403
GPR132	1438	710

L2FC RANK

GSE86491_UP vs GSE132711_DOWN	86_UP	132_DOWN
KIAA1324	1382	**694**
SEMA3G	973	836
LCN12	925	1464
AC008763.1	987	1906
PKHD1L1	**637**	1115
GPR132	1364	2610

SIGNIFICANCE RANK

GSE86491_DOWN vs GSE132711_UP	86_DOWN	132_UP
LHFPL3-AS2	**613**	717
AMZ1	888	730
KCNK2	891	528
CSTA	951	779
SEC14L6	1036	483
CYS1	1080	**482**
MMP10	1082	791

L2FC RANK

GSE86491_DOWN vs GSE132711_UP	86_DOWN	132_UP
LHFPL3-AS2	**189**	1388
AMZ1	876	1419
KCNK2	631	932
CSTA	756	1520
SEC14L6	1091	842
CYS1	682	**840**
MMP10	375	1538

The same process was repeated for ES genes. However, because GSE111976 has few differentially expressed genes, the number of shared genes and conflicting genes was smaller than PRO. ES_UP had 104 genes, while ES_DOWN had 50 genes. Conflicting genes were only 2, and both were between GSE98386_UP and GSE111976_DOWN **(Figure 3.13)**. The cutoff threshold was determined the same way described above. There were only two cutoffs in ES genes, GSE98386_UP and GSE111976_DOWN, the cutoffs being 194 & 153, respectively. **(Table 3.5)**.

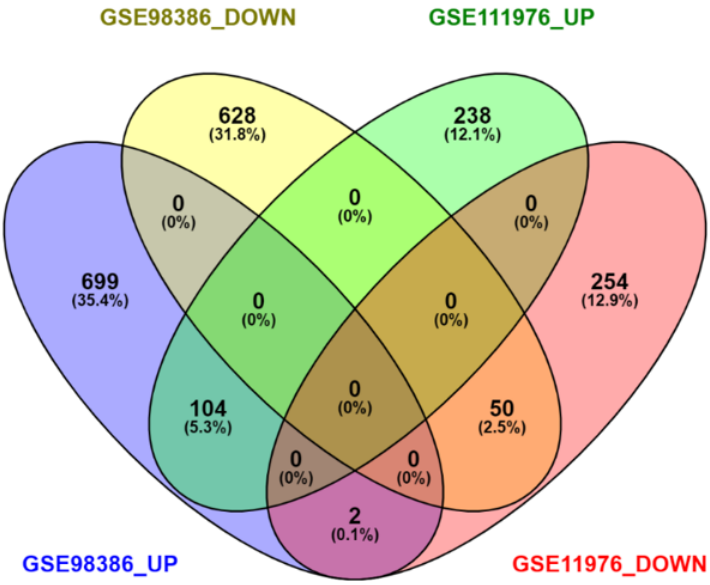


Figure 3.13 The comparison of GSE98386 and GSE111976 to obtain the ES genes. A total of 154 genes (104 up- and 50 downregulated were considered for further analysis). The conflicting two genes were used to determine a cutoff for further filtering of the 154 genes.

Table 3.5 The ranks and threshold values of the ES genes

SIGNIFICANCE RANK

GSE98386_UP vs GSE111976_DOWN	98_UP	119_DOWN
B3GNT7	**194**	**244**
FUT2	689	304

L2FC RANK

GSE98386_UP vs GSE111976_DOWN	98_UP	119_DOWN
B3GNT7	**208**	262
FUT2	433	**153**

Upon filtering the specific datasets with their respective cutoff values. The final gene list that would be used to determine the gene signature was obtained. This final gene

list had 197 unique genes, with 79 in PRO_UP, 42 in PRO_DOWN, 46 in ES_UP, and 39 in ES_DOWN, with 13 genes being common between PRO_UP and ES_UP **(Figure 3.14).** Out of these 197 genes, five were selected from each group based on their rank-sum values **(Materials and Methods),** resulting in 20 genes.

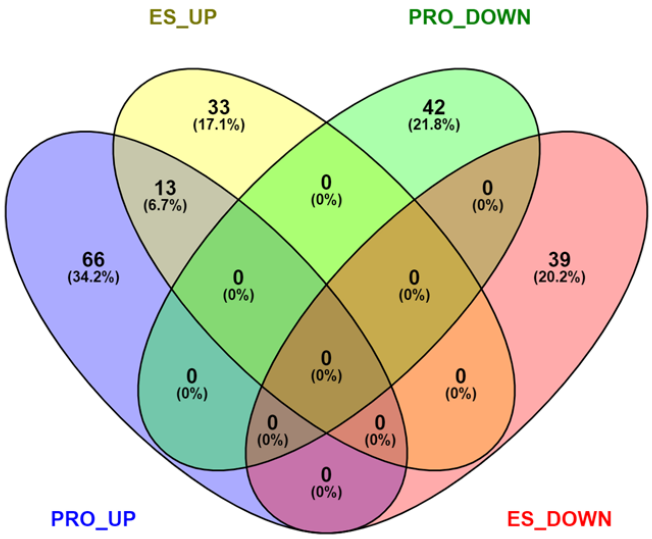


Figure 3.14 The final comparison of genes that will be used for WOI prediction signature generation. A total of 197 genes passed the filter that we created. These genes are high confidence, differentially expressed genes with WOI as a reference.

3.3.1.3 Creation of a Discovery Dataset and SSM value generation.

A discovery dataset was created to generate the SSM scores of the genes. This was done by Z-score standardization and merging the four discovery datasets. The genes that were commonly expressed amongst all four were considered for further analysis. This dataset contains 8728 genes with 71 samples (13 Proliferative, 23 Early secretory, and 35 WOI). If a from the top 20 is not contained in the discovery dataset, we replaced it with the gene(s) that had the next highest rank that is contained in the

discovery dataset (e.g., if the 2nd and fourth gene was not contained in the dataset of the PRO_DOWN group, they were replaced by the 6th and seventh genes). This yielded us the final 20 genes of the signature. The average expression of each phase (PRO, ES & WOI) of each of the 20 genes were obtained; this was our SSM values; thus, our final signature was created.

3.3.2 Generation and validation of NERS-17

3.3.2.1 The validation dataset for the gene signature

Upon generating a 20 gene signature from the discovery dataset, we wanted to validate our signature's strength in predicting the three endometrial phases (PRO, ES & WOI). To achieve this, we used an independent validation dataset. To add strength to the validation, a microarray dataset was used. As our signature was generated from RNA-Seq data, it would show that it is validated and can be utilized cross-platform, increasing its applicability in the case of successful prediction from microarray data.

The validation dataset was obtained from GSE29981. The dataset did not explicitly state the endometrial phases of the samples contained within it. Instead, it provided a menstrual cycle day (MCD). A PCA of the data showed that the MCD of these patients did not reflect the endometrial stages, as particular ES and WOI samples showed to be more closely related to the PRO phase **(Figure 3.15A)**. To generate a valid prediction, the samples should be classified as correctly as possible. Henceforth, we have re-classified the samples to have 3 clusters based on the k-means algorithm **(Figure 3.15B).** There was an outlier in the data that introduced bias to the algorithm; therefore, it was discarded, leaving 19 samples in the validation dataset. The final validation dataset contained 19 samples (11 in PRO, 4 in ES, and 4 in WOI).

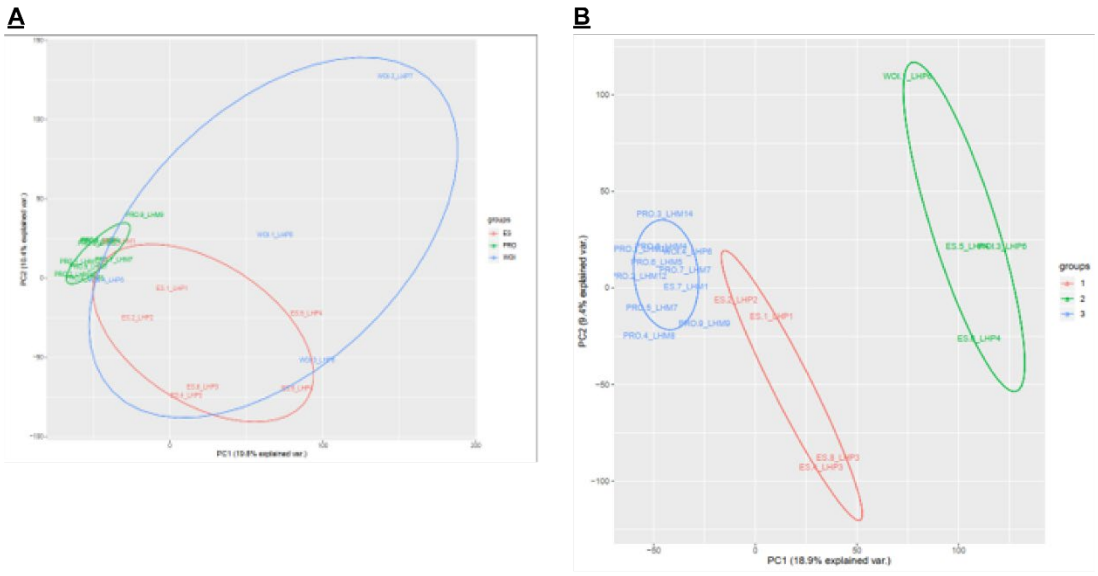


Figure 3.15 PCA graphs of the validation dataset. A) PCA graph before k-means classification. Green-PRO phase, Red- ES phase, Blue – WOI. **B)** PCA graph after k-means classification. Three clear clusters are formed. Blue-PRO, Red – ES, Green – WOI.

3.3.2.2 Generation of NERS-17

The 20 gene list was set to be applied to the validation dataset; however, three genes were not contained in the validation dataset. Therefore, they were discarded from the final signature, leaving the signature with 17 genes that are novel in predicting endometrial receptivity, hence the name, novel endometrial receptivity signature -17 (NERS-17). The genes contained in the signature are as follows; PRO_UP genes: SLC30A2, HSD17B2, SCARA5 **(Figure 3.16A)**, PRO_DOWN genes: FJX1, MMP7, MMP11, SEMA3A, RTKN2 **(Figure 3.16B)**, ES_UP genes: CATSPERB, GPX3, PAEP, DPP4 **(Figure 3.16C)** and ES_DOWN genes: PKHD1L1, STXBP6, PART1, SLC15A2, SIAH3 **(Figure 3.16D)**.

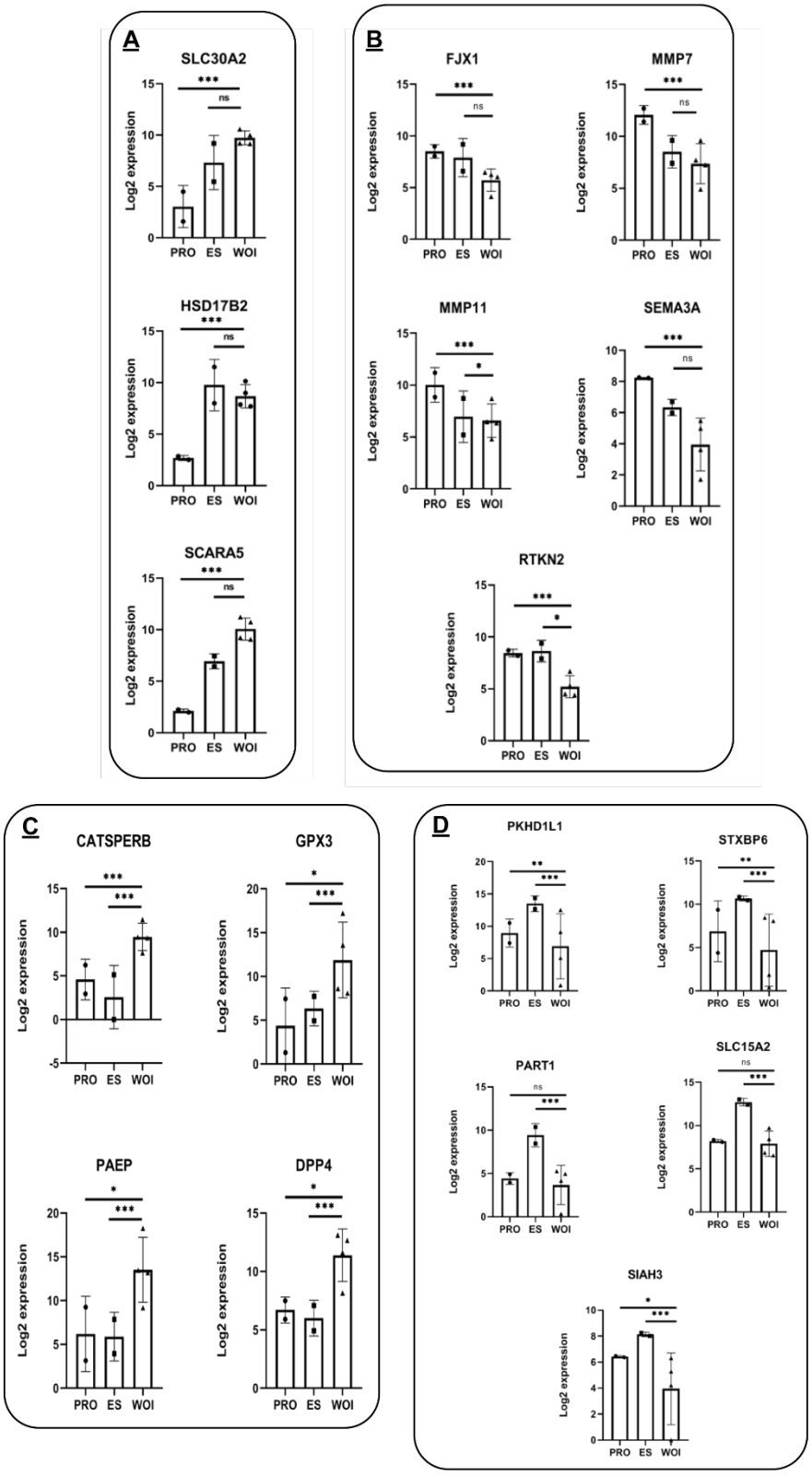


Figure 3.16 Genes that constitute the NERS-17 gene signature. A) Genes that are upregulated in PRO compared to WOI. **B)** Genes that are downregulated in PRO

compared to WOI. **C)** Genes that are upregulated in ES compared to WOI. **D)** Genes that are downregulated in ES compared to WOI.

3.3.3 Validation of NERS-17

Upon generation of NERS-17, we applied the signature to the validation dataset. Our results showed that our signature could predict the correct endometrial stage with 100% accuracy **(Figure 3.17).** The gene signature is potentially able to separate the proliferative phase into two separate groups as well. This could indicate further usage if it can distinguish between the early proliferative and late proliferative stages of the endometrial cycle.

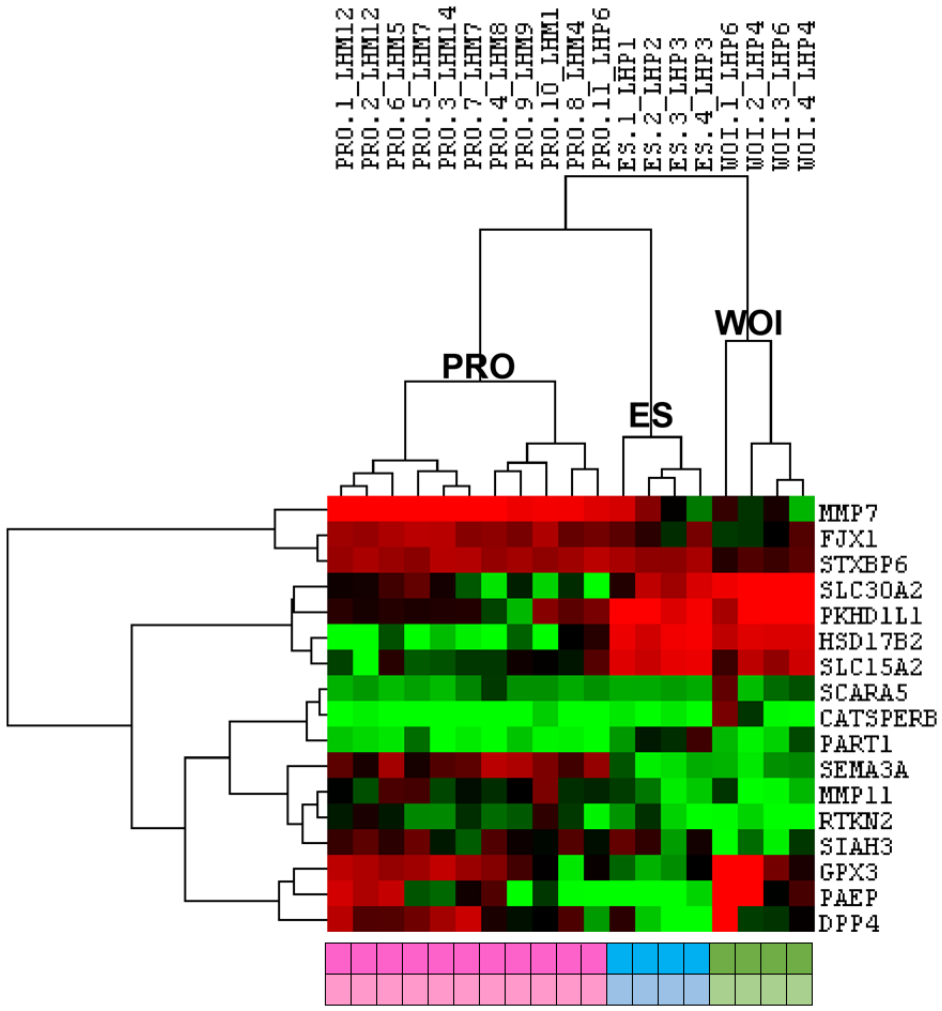


Figure 3.17 The heatmap of the validation dataset is classified accurately by NERS-17. The samples' actual assignment (below darker shade colors) and the samples' prediction (below lighter shaded of colors) match in 100% of the samples. The proliferative phase is seen to be separated into two distinct groups, implicating a future application of NERS-17 in separating the two separate phases of the proliferative phase, being early proliferative and late proliferative.

After successfully validating our gene signature in an independent microarray dataset, we wanted to see our gene signature's novelty concerning the ERA. This showed that out of the 17 genes, four are shared amongst the 258 genes in the ERA **(Figure 3.18)**. This shows that the NERS-17 is novel in terms of most of the genes it contains. However, it contains similarity in predicting the endometrial receptivity with the most powerful tool currently used. The common genes with NERS-17 and ERA are PAEP, DPP4, SLC15A2, and GPX3.

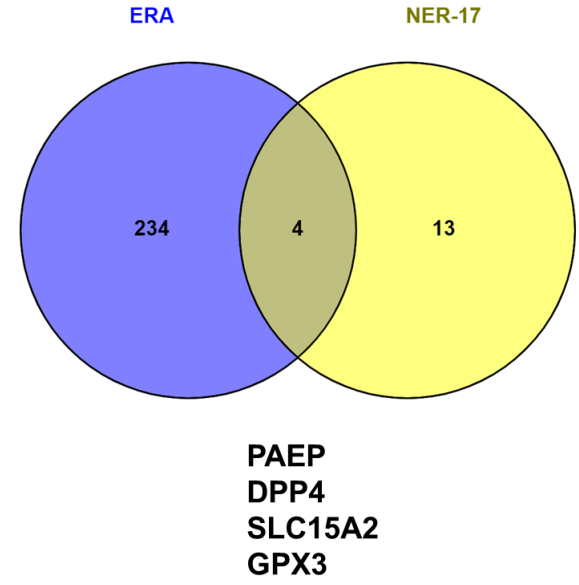


Figure 3.18 The comparison between ERA and NERS-17. The novelty of NERS-17 is highlighted upon comparison with the ERA, having more than ~78% that are novel in predicting the endometrial receptivity.

Overall, we successfully generated a novel gene signature that can predict the WOI and other phases of the endometrial cycle with the PRO phase and the ES phase. With genes that were not linked to endometrial receptivity before. This opens the possibility of this gene-list being used as an alternative to ERA being much cheaper as it contains fewer genes. However, the NERS-17 gene signature still requires rigorous validation experiments for further validation and, more importantly, highlights the potential of its predictive capabilities.

CHAPTER 4

DISCUSSION

Implantation is the first process of many necessary to achieve a successful pregnancy. It is a complex molecular event that shows significant variation between different mammals [99, 100]; hence the mechanisms governing it are still not fully understood. Understanding the implantation process would help clinicians determine specific markers that implicate a high possibility of implantation, which would help treat patients suffering from pregnancy-related complications. Researchers have attempted to identify such markers previously [101], but these markers were not robust enough to be applied in clinics in place of the gold standard of endometrial dating at the time [28]. However, with the dawn of the "genomic era" after the human genome project, genomic, metabolomic, and especially transcriptomic studies on both the embryo and the endometrium increased dramatically, thus creating a new field, "reproductomics. However, answering the challenging question of implantation is not easy, as clearly, from many studies that go under the radar and do not get the attention they deserve. This is because the field of "reproductomics" is saturated with almost identical articles that get to the same conclusions but have no reproducibility [102-104]. Out of these studies, one has been able to identify the WOI with high sensitivity and specificity, and consistently, called the Endometrial Receptivity Assay (ERA) [34]. This diagnostic tool is still being used today around the globe. Other alternatives have been developed by researchers [36] but are not widely used as they do not provide a clear advantage over the ERA. The ERA's drawbacks exist as it utilizes an endometrial biopsy, limiting its applicability in clinics and conventional IVF patients. It also

utilizes 238 genes to predict the WOI, making it unaffordable, reducing its usage amongst mid to low-income patients. Therefore, it was vital for us to address these shortcomings, build on the ideas previously developed in the field, and with novel approaches, find biomarkers for WOI prediction.

Our main goal was to focus on the prediction of WOI that is more applicable to IVF patients. The first idea to achieve this was to focus on a less invasive procedure. Therefore, we wanted to focus on WOI prediction from serum. It is well studied that miRNAs, compared to mRNAs, are more stable in serum samples, even in long term storage. This property makes them prime candidates as biomarkers. Henceforth our goal was to identify miRNAs from serum samples that could be utilized in this regard. As the field of "reproductomics" has scarcely focused on this aspect [60, 61], we came up with the methodology and application to discover and validate individual miRNAs ourselves. The second idea was to decrease the number of genes used to predict the WOI by a significant amount and achieve this without losing predictive power. This would allow for a decrease in expenses compared to the ERA and make it more applicable with more readily available technologies worldwide.

4.1 Identification and validation of miRNA in serum samples for potential WOI prediction.

In our study, we initially aimed to discover miRNAs for a pipeline used in WOI prediction and predict the other stages of the endometrial cycle. Our approach's strategy was to initially identify those miRNAs that are statistically expressed differentially expressed in the WOI concerning other phases of the endometrial cycle.

This was done via using *in-silico* methods. We then wanted to identify those highly expressed in serum samples, assuming they would behave in serum as they did in the endometrium, as previous literature suggested [61]. Thus, we identified ten miRNAs highly expressed in serum and differentially expressed during the WOI, in both the PRO and the ES phases. This filter removed most genes that had the highest L2FC values between the phases and the WOI. To increase the number of candidate miRNAs predicted by the WOI, we did not set an L2FC filter. This allowed most of the miRNAs in the 10-miRNA list to have minimal expression changes between the endometrial stages. Another issue is that the datasets that the expression prediction graphs were obtained from did not contain a measurement between PRO and ES, except for one (GSE44558). Therefore, in most cases, the miRNA's actual expression profile could not be obtained via *in-silico* methods. This pushed us to explore the serum expression during the menstrual cycle ourselves, as this data was not available and not been comprehensively studied using NGS.

Identifying the 10-miRNA list prompted us to validate and explore if these miRNAs would; behave as they do in the endometrial tissue and whether our methodology successfully predicted such miRNAs. We wanted to validate their expression values in serum using RNA-Seq technology and obtain our data. The RNA-Seq experiments showed mixed results. The most apparent result was that eight of the ten of our candidate miRNAs did not have expression changes that could be detected during the menstrual cycle. This most probably is a result of, as stated above, the L2FC values being small in the endometrial tissue. Therefore, even if the identified miRNA influences the WOI post-transcriptionally, in the endometrial tissue. Its regulation is local rather than distant by these miRNAs. We obtained another result because two of

our miRNAs showed variation in regard between the different phases of the endometrial cycle. These miRNAs were hsa-miR-375-3p and hsa-miR-30b-5p.

The miRNA hsa-miR-375-3p was previously implicated to play an important in the endometrium but in a different context of that which we implicated it. Initially, it was upregulated in the endometrial tissue of patients undergoing controlled ovarian stimulation compared to healthy controls [103]. Hsa-miR-375-3p has also been shown to be a vital miRNA in endometriosis [105, 106]. Endometriosis is a fibrotic condition defined by the presence and growth of endometrium-like tissue outside the uterine cavity (ectopically) [107]. The functional role of hsa-miR-375-3p in endometriosis development is not entirely understood. The miRNA was shown to increase expression in the endometriotic lesion as a whole [108].

Conversely, when investigated at the cellular level, it shows a decrease in stromal cells obtained from the lesions [109]. However, to date, no study has linked miR-375 with endometrial receptivity or implantation success. The *in-silico* predictions we had regarding hsa-miR-375-3p are consistent with what the literature states. In IVF patients, progesterone is much higher than that of a healthy female during their WOIs. This suggests that hsa-miR-375-3p expression in the endometrium is correlated with progesterone, which increases the miRNA expression as reaching the WOI to be expected. However, our RNA-Seq findings did not show the same result, highlighting that serum expression profile might not reflect that of the endometrium for specific miRNAs. Regardless, as this miRNA shows a decrease between the PRO and ES phases, it retains its importance as a potential serum biomarker for endometrial stages. An interesting finding regarding this miRNA expression was that in C5, Day 22, it had a spike in expression levels, which could be validated via qPCR afterward. This

might be because we did not collect the serum sample from C2 on Day 22; we could not detect such an increase, which is an indicator of a biological phenomenon.

The other miRNA that showed variation that was among our 10-miRNA list was hsa-miR-30b-5p. This miRNA is a member of the miR-30 family, which has been widely reported to play a crucial role in endometrial receptivity and implantation in murine models and humans [62]. Hsa-miR-30b-5p has been shown to increase during the WOI compared to the Pre-Receptive endometrium[57]. It has also been shown to be localized in extracellular vesicles of the endometrium. It shows a consistent increase in these vesicles' expression as the endometrium reaches a receptive state [110]. This event suggests that miR-30d affects the implanting embryo as it is well known that the embryo takes up the miR-30 family in such cases. This might indicate the consistency between our *in-silico* prediction and RNA-Seq result for Control 2. As this miRNA is known to be secreted, it might behave in such a manner in the serum sample. However, Control 5 showed an opposite trend to both *in-silico* results and Control 5. This might indicate that Control 5's miRNAs might have been affected by hemolysis and necessitates a more comprehensive RNA-Seq experiment with more individuals.

Our validation experiments of both miR-375-3p and miR-30b-5p via qPCR showed significant values of correlation. This shows that our sequencing experiments' sequencing depth was adequate to detect biological phenomena, as stated previously. More importantly, it shows that our methodology for miRNA detection via qPCR works and can identify serum miRNA profiles' differences. With the qPCR re-validation, the first aspect of our research was completed; however, we still wanted to

see if there were miRNAs that were not in our 10-miRNA list that could hold potential in predicting the WOI.

In that line, we have discovered seven distinct miRNAs that could potentially predict the WOI and the different stages of the endometrial cycle. These miRNAs were separated into four distinct groups. The first group was statistically differentially expressed miRNAs in the Receptive endometrium, and it contained two miRNAs. The first of these was hsa-miR-183-5p. MiR-183-5p has been implicated with endometrial function in previous studies [111, 112]. It was shown to enhance endometriosis's invasive properties by being over-expressed in endometrial stromal cells [113, 114]. Additionally, this miRNA was implicated with endometrial receptivity and embryo implantation. MiR-183-5p downregulation showed a decrease in implantation rates in murine samples, as well as its expression was shown to be modulated by estrogen [115]. The other miRNA that is in this group was hsa-miR-106-3p. It has not been investigated in the context of the endometrium. However, it has been identified as a serum biomarker for different biological conditions such as colorectal cancer, hepatocellular carcinoma, and brain-related disorders [116-119]. It has been shown to promote EMT to increase the metastatic properties in cancers [120, 121]. This is in line with what is known regarding implantation, as we observe an increase of hsa-miR-106b-3p during the receptive endometrium, although during the implantation process, the endometrium goes under MET, endometrial receptivity is characterized by EMT.

Two other groups of miRNAs were the ones that could potentially predict precisely between the PRO-ES and those that could do so between ES-WOI. The miRNAs in

the first group were hsa-miR-2277-3p and hsa-miR-148b-3p. Hsa-miR-2277-3p has not been shown to be involved in the endometrium previously. However, it was shown to be transported through exosomes and have paracrine effects on human colon cancer cells. It was shown to increase proliferation, migration, and invasion of colon cancer cell lines [122]. Hsa-miR-148b-3p has been shown to modulate the adaptive immune response. Upon inhibition, the HLA receptors' effects were abolished, and the adaptive immune response was inhibited [123]. The effect of hsa-miR-148b-3p on the adaptive immune system in a similar manner has been shown to affect the endometrium as well, albeit in bovines [124]. The miRNA, the sole member of the ES-WOI group, hsa-miR-7108-3p, had no prior literature regarding its role in any biological phenomenon. This shows that the serum could hold potential in discovering miRNAs that were previously overlooked.

The final group of miRNAs was those that could be used to predict both PRO-ES and ES-WOI potentially. The first miRNA in this group is has-miR-4730. Just as miR-7108-3p, it is a novel miRNA and has not been studied extensively. However, it was shown to have a high predictive power of survival in pancreatic cancer patients [125]. The final miRNA that we have shown to have potential in predicted endometrial stages was hsa-miR-615-3p. It has been shown that this miRNA plays a role in EMT and promotes tumor growth in breast cancer; however, the functional mechanisms have not been elucidated yet [126].

These seven miRNAs are important candidates for a potential diagnostic tool for WOI prediction from healthy individuals' serum. They have been identified by not normalizing the MCDs of the Control participants to the LH surge, which gives them

a unique property of being able to predict the WOI without relying on LH Surge information. However, it might be important to consider normalizing the LH days to see if new miRNAs are identified in the process. Regarding our pipeline, it might be important to see the expression profiles of these seven miRNAs in the endometrial tissue, basically reverse engineering our approach in identifying the 10-miRNA list. All in all, our pipeline showed that: 1) particular miRNAs could behave similarly between the endometrial tissue and serum samples with respect to the endometrial stage, 2) There are novel miRNAs that have not yet been implicated with the implantation process that could be important for the process and 3) Certain miRNAs involved in the implantation process, show variation in serum samples, and could be used for endometrial stage prediction.

4.2 Generation of a novel gene list for endometrial stage prediction

We were able to generate a gene signature that contains seventeen genes titled Novel Endometrial Receptivity Signature – 17 (NERS-17). It could potentially predict not just the WOI but also other endometrial stages. The novel gene-list's motivation was to counter the problems in ERA's applicability due to the high number of genes (238 genes) it contains. This number increases the test's overall price, as it cannot be run as much in the next generational sequencer. Decreasing the number of genes in the test would cheapen the costs per patient immensely if done on a targeted RNA-Seq sequencer. It also opens the possibility of making the test a qPCR-based test, making it more accessible worldwide. Even though we had 100% accuracy predicting the three endometrial stages in a validation set, it is still essential to test our gene list's accuracy in different cohorts. Unfortunately, the number of studies that contain all three of the endometrial stages is scarce online.

Additionally, these studies have low reliability since they have used unpaired samples and are done on microarray platforms. Hence it is possible to do a meta-analysis, merging different RNA-Seq and microarray data to create one massive validation dataset. Additionally, alongside *in-silico* validation, *in-vitro* validation is also required. As it would be challenging to obtain endometrial tissue throughout the menstrual cycle of a female, the validation could initially be done on endometrial cancer cell lines. One option could be to compare our gene signature's strength to ERA and other gene signatures. Unfortunately, this approach's issue is that replicating the ERA is impossible, as researchers have no access to their prediction algorithm. Therefore, an SSM that contains the 238 genes might be trained in the same manner NERS-17 was trained, and a comparison could be made that way. However, this would not reflect the actual prediction of ERA. As stated previously in the **Results**, our gene signature only has four genes out of the 17 commons with the ERA. To further understand the novelty of our gene signature, we must compare them with other published signatures. Our gene signature consists of 17 genes, 3 in PRO_UP, 5 in PRO_DOWN, 4 in ES_UP, and 5 in ES_DOWN.

The three genes that are in PRO_UP are SLC30A2, HSD17B2, and SCARA5. SLC30A2 is a zinc transporter. It has previously been overexpressed in WOI compared to ES and PRO [63, 127]. Its functional implications in implantation have not been elucidated in humans; however, it has been shown to play a role in nutrient transport in marsupial implantation [128]. This is probably the case in the endometrium, as it is well known that sugar metabolism is highly active during the WOI. HSD17B2 is a gene known to cause endometriosis of the ovary and progesterone resistance [129, 130]. It is known to modulate steroid hormones,

primarily estrogen and progesterone; for instance, one of its functions is to catalyze 17-beta-estradiol to estrone, deactivating it. It is also shown that it is modulated by progesterone expression, and expression increase during the secretory phase [131]. SCARA5 was shown to have a vital role in the decidualized stromal cell during the mid-secretory phase of the endometrial cycle [132]. The gene has been shown to play a very intricate role in decidualization, and the loss of its expression during the mid-secretory phase causes early pregnancy losses [133].

FJX1 is the first among the five downregulated genes in the PRO phase compared to the WOI. It was shown to increase the eutopic endometrial tissue in patients who have endometriosis [134]. This is in line with what our results suggest, as, in healthy females, you would expect lower levels of FJX1 in the endometrial tissue. MMP7 and MMP11 are the two matrix metalloproteinases that show decreased WOI expression [135, 136]. This makes sense as it is well documented that these genes play a role in menstruation rather than embryo implantation or endometrial receptivity. SEMA3A, the fourth gene on the list, is a diffusible axonal repellent usually involved in brain and bone metabolism [137, 138]. However, it has been linked to the endometrium as well. Consistent findings were reported with our results, as SEMA3A decreased the secretory phase of the endometrial cycle as it is also a pro-angiogenesis gene [139]. The last gene in this category is RTKN2. This gene has not yet been implicated with embryonic implantation. However, it is known to promote cell proliferation in various cancers [140, 141]. This is in line with the previous genes that constitute this group.

The third group of NERS-17 is upregulated in WOI compared to the ES (ES_UP). The first of these genes is CATSPERB. This gene normally controls sperm motility and is

a calcium channel [142]. It has not been implicated with endometrial function previously. The second gene, GPX3, showed an increase in expression values during the WOI compared to the ES in different studies [143, 144]. Its main function is to protect cells from oxidative damage. PAEP and DPP4 alongside GPX3 have been consistently upregulated in the WOI compared to the ES phase of endometrial tissue. [145, 146].

The final group of the signature is the ES_DOWN group. PKHD1L1, the first gene of this group, has not been linked to either embryo implantation or endometrial receptivity. Its exact function is unknown; however, it showed specific upregulation in T lymphocytes, suggesting a cellular immunity role [147]. STXBP6 is again a novel gene in the context of endometrial function. It has been implicated in breast cancer and lung adenocarcinoma [148, 149]. It has been shown to inhibit the activation of T-cells as well [150]. This makes with the profile we see, as the WOI is an immunologically active phase of the endometrial cycle. Another transcript that shows the decrease in the WOI is PART1. PART1 is not a protein-coding gene but rather a long non-coding RNA (lncRNA). It has been linked to being a diagnostic predictor of ovarian cancer [151] and modulating proliferation in prostate cancer cells [152]. SLC15A2 is the fourth gene in this group. It is a solute carrier that is responsible for peptide absorption. It has been shown that its overexpression causes endometrial dysfunction in murine models [153]. The final gene, SIAH3, is an E3 ubiquitin ligase. It is not involved in the endometrial function and is more involved in localizing specific genes to the mitochondria, especially those linked to Parkinson's disease [154].

In conclusion, we have generated a novel gene-list titled NERS-17 that contains only of17 genes and can predict different phases of the endometrial cycle. The genes that constitute NERS-17 are mostly genes that have not been previously linked to endometrial function before. This makes our gene signature novel as it has highlighted potential functions of genes not previously explored.

4.3 Conclusion and Future Perspectives

Overall, our research has laid the groundwork for a complete pipeline to determine and validate serum miRNAs that could potentially be used in the future for WOI prediction. Additionally, we have shown that a 17 gene list could be used instead of 238 to predict the endometrial stages. We will have to do miRNA Sequencing experiments with more controls to validate our findings and potentially find new miRNAs for this purpose. Our main goal was to convert serum miRNAs to a kit to be used by both healthy and IVF patients. In this thesis, we have generated a prototype and a good idea of moving forward concerning that aim. After sequencing controls, IVF patients will be screened similarly. In conclusion, our research shows high promise for the future of in vitro prediction of the WOI.

BIBLIOGRAPHY

1. Roberts, A.D.L., C.R. , *Where have all the conceptions gone?* Lancet, 1975. **1**(7907): p. 636-7.
2. Wilcox, A.J., et al., *Incidence of early loss of pregnancy.* N Engl J Med, 1988. **319**(4): p. 189-94.
3. Ellish, N.J., et al., *A prospective study of early pregnancy loss.* Hum Reprod, 1996. **11**(2): p. 406-12.
4. Bellver, J. and C. Simon, *Implantation failure of endometrial origin: what is new?* Curr Opin Obstet Gynecol, 2018. **30**(4): p. 229-236.
5. Altmae, S., et al., *Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium.* Hum Reprod Update, 2014. **20**(1): p. 12-28.
6. Hernandez-Vargas, P., M. Munoz, and F. Dominguez, *Identifying biomarkers for predicting successful embryo implantation: applying single to multi-OMICs to improve reproductive outcomes.* Hum Reprod Update, 2020. **26**(2): p. 264-301.
7. Boomsma, C.M., et al., *Endometrial secretion analysis identifies a cytokine profile predictive of pregnancy in IVF.* Hum Reprod, 2009. **24**(6): p. 1427-35.
8. Koot, Y.E., et al., *Molecular aspects of implantation failure.* Biochim Biophys Acta, 2012. **1822**(12): p. 1943-50.
9. Gellersen, B. and J.J. Brosens, *Cyclic decidualization of the human endometrium in reproductive health and failure.* Endocr Rev, 2014. **35**(6): p. 851-905.
10. Achache, H. and A. Revel, *Endometrial receptivity markers, the journey to successful embryo implantation.* Hum Reprod Update, 2006. **12**(6): p. 731-46.
11. Ponnampalam, A.P., et al., *Molecular classification of human endometrial cycle stages by transcriptional profiling.* Mol Hum Reprod, 2004. **10**(12): p. 879-93.
12. Ruiz-Alonso, M., D. Blesa, and C. Simon, *The genomics of the human endometrium.* Biochim Biophys Acta, 2012. **1822**(12): p. 1931-42.
13. Salamonsen, L.A. and D.E. Woolley, *Matrix metalloproteinases in normal menstruation.* Hum Reprod, 1996. **11 Suppl 2**: p. 124-33.
14. Diaz-Gimeno, P., et al., *Transcriptomics of the human endometrium.* Int J Dev Biol, 2014. **58**(2-4): p. 127-37.
15. Maruyama, T. and Y. Yoshimura, *Molecular and cellular mechanisms for differentiation and regeneration of the uterine endometrium.* Endocr J, 2008. **55**(5): p. 795-810.
16. Talbi, S., et al., *Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women.* Endocrinology, 2006. **147**(3): p. 1097-121.
17. Huang, J., et al., *A comparison of transcriptomic profiles in endometrium during window of implantation between women with unexplained recurrent implantation failure and recurrent miscarriage.* Reproduction, 2017. **153**(6): p. 749-758.

18. Ashary, N., A. Tiwari, and D. Modi, *Embryo Implantation: War in Times of Love.* Endocrinology, 2018. **159**(2): p. 1188-1198.
19. Coughlan, C., et al., *Recurrent implantation failure: definition and management.* Reprod Biomed Online, 2014. **28**(1): p. 14-38.
20. Yang, W., et al., *Coagulation parameters predictive of repeated implantation failure in Chinese women: A retrospective study.* Medicine (Baltimore), 2020. **99**(48): p. e23320.
21. Valdes, C.T., A. Schutt, and C. Simon, *Implantation failure of endometrial origin: it is not pathology, but our failure to synchronize the developing embryo with a receptive endometrium.* Fertil Steril, 2017. **108**(1): p. 15-18.
22. Gomez, E., et al., *Human Endometrial Transcriptomics: Implications for Embryonic Implantation.* Cold Spring Harb Perspect Med, 2015. **5**(7): p. a022996.
23. Macklon, N., *Recurrent implantation failure is a pathology with a specific transcriptomic signature.* Fertil Steril, 2017. **108**(1): p. 9-14.
24. Sebastian-Leon, P., et al., *Asynchronous and pathological windows of implantation: two causes of recurrent implantation failure.* Hum Reprod, 2018. **33**(4): p. 626-635.
25. Galliano, D., et al., *ART and uterine pathology: how relevant is the maternal side for implantation?* Hum Reprod Update, 2015. **21**(1): p. 13-38.
26. Craciunas, L., et al., *Conventional and modern markers of endometrial receptivity: a systematic review and meta-analysis.* Hum Reprod Update, 2019. **25**(2): p. 202-223.
27. Griesinger, G., S. Trevisan, and B. Cometti, *Endometrial thickness on the day of embryo transfer is a poor predictor of IVF treatment outcome.* Hum Reprod Open, 2018. **2018**(1): p. hox031.
28. Noyes, R.W., A.T. Hertig, and J. Rock, *Dating the endometrial biopsy.* Am J Obstet Gynecol, 1975. **122**(2): p. 262-3.
29. Driessen, F., et al., *The significance of dating an endometrial biopsy for the prognosis of the infertile couple.* Int J Fertil, 1980. **25**(2): p. 112-6.
30. Balasch, J., et al., *The usefulness of endometrial biopsy for luteal phase evaluation in infertility.* Hum Reprod, 1992. **7**(7): p. 973-7.
31. Direito, A., et al., *Relationships between the luteinizing hormone surge and other characteristics of the menstrual cycle in normally ovulating women.* Fertil Steril, 2013. **99**(1): p. 279-285 e3.
32. Hannan, N.J., et al., *2D-DiGE analysis of the human endometrial secretome reveals differences between receptive and nonreceptive states in fertile and infertile women.* J Proteome Res, 2010. **9**(12): p. 6256-64.
33. Boomsma, C.M., et al., *Cytokine profiling in endometrial secretions: a non-invasive window on endometrial receptivity.* Reprod Biomed Online, 2009. **18**(1): p. 85-94.
34. Diaz-Gimeno, P., et al., *A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature.* Fertil Steril, 2011. **95**(1): p. 50-60, 60 e1-15.
35. Diaz-Gimeno, P., et al., *The accuracy and reproducibility of the endometrial receptivity array is superior to histology as a diagnostic method for endometrial receptivity.* Fertil Steril, 2013. **99**(2): p. 508-17.
36. Enciso, M., et al., *Development of a new comprehensive and reliable endometrial receptivity map (ER Map/ER Grade) based on RT-qPCR gene expression analysis.* Hum Reprod, 2018. **33**(2): p. 220-228.

37. Mahajan, N., *Endometrial receptivity array: Clinical application.* J Hum Reprod Sci, 2015. **8**(3): p. 121-9.
38. Bassil, R., et al., *Does the endometrial receptivity array really provide personalized embryo transfer?* J Assist Reprod Genet, 2018. **35**(7): p. 1301-1305.
39. Robert, C.A., *Endometrial receptivity array for individualized determination of endometrial receptivity.* Hum Reprod Open, 2020. **2020**(2): p. hoaa014.
40. Katzorke, N., et al., *Diagnosis of Endometrial-Factor Infertility: Current Approaches and New Avenues for Research.* Geburtshilfe Frauenheilkd, 2016. **76**(6): p. 699-703.
41. Simon, C., et al., *A 5-year multicentre randomized controlled trial comparing personalized, frozen and fresh blastocyst transfer in IVF.* Reprod Biomed Online, 2020. **41**(3): p. 402-415.
42. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
43. Shivdasani, R.A., *MicroRNAs: regulators of gene expression and cell differentiation.* Blood, 2006. **108**(12): p. 3646-53.
44. Abba, M.L., et al., *MicroRNA Regulation of Epithelial to Mesenchymal Transition.* J Clin Med, 2016. **5**(1).
45. Song, L. and R.S. Tuan, *MicroRNAs and cell differentiation in mammalian development.* Birth Defects Res C Embryo Today, 2006. **78**(2): p. 140-9.
46. Ouellet, D.L., et al., *MicroRNAs in gene regulation: when the smallest governs it all.* J Biomed Biotechnol, 2006. **2006**(4): p. 69616.
47. Galliano, D. and A. Pellicer, *MicroRNA and implantation.* Fertil Steril, 2014. **101**(6): p. 1531-44.
48. Chen, X., et al., *Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases.* Cell Res, 2008. **18**(10): p. 997-1006.
49. Montani, F., et al., *miR-Test: a blood test for lung cancer early detection.* J Natl Cancer Inst, 2015. **107**(6): p. djv063.
50. Glinge, C., et al., *Stability of Circulating Blood-Based MicroRNAs - Pre-Analytic Methodological Considerations.* PLoS One, 2017. **12**(2): p. e0167969.
51. Wang, Y., et al., *DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal.* Nat Genet, 2007. **39**(3): p. 380-5.
52. Liu, W., et al., *MicroRNA and Embryo Implantation.* Am J Reprod Immunol, 2016. **75**(3): p. 263-71.
53. Capalbo, A., et al., *MicroRNAs in spent blastocyst culture medium are derived from trophectoderm cells and can be explored for human embryo reproductive competence assessment.* Fertil Steril, 2016. **105**(1): p. 225-35 e1-3.
54. Paul, A.B.M., S.T. Sadek, and A.M. Mahesan, *The role of microRNAs in human embryo implantation: a review.* J Assist Reprod Genet, 2019. **36**(2): p. 179-187.
55. Cuman, C., et al., *Human Blastocyst Secreted microRNA Regulate Endometrial Epithelial Cell Adhesion.* EBioMedicine, 2015. **2**(10): p. 1528-35.
56. Kuokkanen, S., et al., *Genomic profiling of microRNAs and messenger RNAs reveals hormonal regulation in microRNA expression in human endometrium.* Biol Reprod, 2010. **82**(4): p. 791-801.

57. Altmae, S., et al., *MicroRNAs miR-30b, miR-30d, and miR-494 regulate human endometrial receptivity.* Reprod Sci, 2013. **20**(3): p. 308-17.
58. Li, Z., et al., *MicroRNA-429 functions as a regulator of epithelial-mesenchymal transition by targeting Pcdh8 during murine embryo implantation.* Hum Reprod, 2015. **30**(3): p. 507-18.
59. Liang, J., S. Wang, and Z. Wang, *Role of microRNAs in embryo implantation.* Reprod Biol Endocrinol, 2017. **15**(1): p. 90.
60. Rekker, K., et al., *Circulating microRNA Profile throughout the menstrual cycle.* PLoS One, 2013. **8**(11): p. e81166.
61. Kresowik, J.D., et al., *MicroRNA-31 is significantly elevated in both human endometrium and serum during the window of implantation: a potential biomarker for optimum receptivity.* Biol Reprod, 2014. **91**(1): p. 17.
62. Vilella, F., et al., *Hsa-miR-30d, secreted by the human endometrium, is taken up by the pre-implantation embryo and might modify its transcriptome.* Development, 2015. **142**(18): p. 3210-21.
63. Sigurgeirsson, B., et al., *Comprehensive RNA sequencing of healthy human endometrium at two time points of the menstrual cycle.* Biol Reprod, 2017. **96**(1): p. 24-33.
64. Teder, H., et al., *TAC-seq: targeted DNA and RNA sequencing for precise biomarker molecule counting.* NPJ Genom Med, 2018. **3**: p. 34.
65. Lopez-Romero, P., et al., *Processing of Agilent microRNA array data.* BMC Res Notes, 2010. **3**: p. 18.
66. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-64.
67. Del Carratore, F., et al., *RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets.* Bioinformatics, 2017. **33**(17): p. 2774-2775.
68. Breitling, R. and P. Herzyk, *Rank-based methods as a non-parametric alternative of the T-statistic for the analysis of biological microarray data.* J Bioinform Comput Biol, 2005. **3**(5): p. 1171-89.
69. Andres-Leon, E., R. Nunez-Torres, and A.M. Rojas, *miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis.* Sci Rep, 2016. **6**: p. 25749.
70. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nat Methods, 2012. **9**(4): p. 357-9.
71. Tam, S., M.S. Tsao, and J.D. McPherson, *Optimization of miRNA-seq data preprocessing.* Brief Bioinform, 2015. **16**(6): p. 950-63.
72. Kozomara, A. and S. Griffiths-Jones, *miRBase: annotating high confidence microRNAs using deep sequencing data.* Nucleic Acids Res, 2014. **42**(Database issue): p. D68-73.
73. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
74. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
75. Juzenas, S., et al., *A comprehensive, cell specific microRNA catalogue of human peripheral blood.* Nucleic Acids Res, 2017. **45**(16): p. 9290-9301.

76. Pantano, L., X. Estivill, and E. Marti, *SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells.* Nucleic Acids Res, 2010. **38**(5): p. e34.
77. Oliveros, J.C. *Venny. An interactive tool for comparing lists with Venn's diagrams.* (2007-2015); Available from: https://bioinfogp.cnb.csic.es/tools/venny/index.html.
78. Tuck, M.K., et al., *Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group.* J Proteome Res, 2009. **8**(1): p. 113-7.
79. Li, Y. and K.V. Kowdley, *Method for microRNA isolation from clinical serum samples.* Anal Biochem, 2012. **431**(1): p. 69-75.
80. Blondal, T., et al., *Assessing sample and miRNA profile quality in serum and plasma or other biofluids.* Methods, 2013. **59**(1): p. S1-6.
81. Shah, J.S., P.S. Soon, and D.J. Marsh, *Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs.* PLoS One, 2016. **11**(4): p. e0153200.
82. Wright, K., et al., *Comparison of methods for miRNA isolation and quantification from ovine plasma.* Sci Rep, 2020. **10**(1): p. 825.
83. Garcia-Elias, A., et al., *Defining quantification methods and optimizing protocols for microarray hybridization of circulating microRNAs.* Sci Rep, 2017. **7**(1): p. 7725.
84. Vieth, B., et al., *A systematic evaluation of single cell RNA-seq analysis pipelines.* Nat Commun, 2019. **10**(1): p. 4667.
85. Sena, J.A., et al., *Unique Molecular Identifiers reveal a novel sequencing artefact with implications for RNA-Seq based gene expression analysis.* Sci Rep, 2018. **8**(1): p. 13121.
86. Altmae, S., et al., *Meta-signature of human endometrial receptivity: a meta-analysis and validation study of transcriptomic biomarkers.* Sci Rep, 2017. **7**(1): p. 10077.
87. Chi, R.A., et al., *Human Endometrial Transcriptome and Progesterone Receptor Cistrome Reveal Important Pathways and Epithelial Regulators.* J Clin Endocrinol Metab, 2020. **105**(4).
88. Wang, W., et al., *Single-cell transcriptomic atlas of the human endometrium during the menstrual cycle.* Nat Med, 2020. **26**(10): p. 1644-1653.
89. Birnbaum, D.J., et al., *A 25-gene classifier predicts overall survival in resectable pancreatic cancer.* BMC Med, 2017. **15**(1): p. 170.
90. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-3.
91. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.
92. Law, C.W., et al., *RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR.* F1000Res, 2016. **5**.
93. Kim, D., et al., *Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype.* Nat Biotechnol, 2019. **37**(8): p. 907-915.
94. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.

95. Klaus, B. and S. Reisenauer, *An end to end workflow for differential gene expression using Affymetrix microarrays.* F1000Res, 2016. **5**: p. 1384.
96. Zahid, K.R., et al., *mTOR/HDAC1 Crosstalk Mediated Suppression of ADH1A and ALDH2 Links Alcohol Metabolism to Hepatocellular Carcinoma Onset and Progression in silico.* Front Oncol, 2019. **9**: p. 1000.
97. Chang, L., et al., *miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology.* Nucleic Acids Res, 2020. **48**(W1): p. W244-W251.
98. Kanehisa, M., *The KEGG database.* Novartis Found Symp, 2002. **247**: p. 91-101; discussion 101-3, 119-28, 244-52.
99. Menezo, Y.J. and F. Herubel, *Mouse and bovine models for human IVF.* Reprod Biomed Online, 2002. **4**(2): p. 170-5.
100. Lee, K.Y. and F.J. DeMayo, *Animal models of implantation.* Reproduction, 2004. **128**(6): p. 679-95.
101. Nikas, G., *Pinopodes as markers of endometrial receptivity in clinical practice.* Hum Reprod, 1999. **14 Suppl 2**: p. 99-106.
102. Haouzi, D., et al., *Identification of new biomarkers of human endometrial receptivity in the natural cycle.* Hum Reprod, 2009. **24**(1): p. 198-205.
103. Sha, A.G., et al., *Genome-wide identification of micro-ribonucleic acids associated with human endometrial receptivity in natural and stimulated cycles by deep sequencing.* Fertil Steril, 2011. **96**(1): p. 150-155 e5.
104. Hu, S., et al., *Transcriptomic changes during the pre-receptive to receptive transition in human endometrium detected by RNA-Seq.* J Clin Endocrinol Metab, 2014. **99**(12): p. E2744-53.
105. Braza-Boils, A., et al., *MicroRNA expression profile in endometriosis: its relation to angiogenesis and fibrinolytic factors.* Hum Reprod, 2014. **29**(5): p. 978-88.
106. Logan, P.C., P. Yango, and N.D. Tran, *Endometrial Stromal and Epithelial Cells Exhibit Unique Aberrant Molecular Defects in Patients With Endometriosis.* Reprod Sci, 2018. **25**(1): p. 140-159.
107. Vigano, P., et al., *Time to redefine endometriosis including its pro-fibrotic nature.* Hum Reprod, 2018. **33**(3): p. 347-352.
108. Saare, M., et al., *Challenges in endometriosis miRNA studies - From tissue heterogeneity to disease specific miRNAs.* Biochim Biophys Acta Mol Basis Dis, 2017. **1863**(9): p. 2282-2292.
109. Rekker, K., et al., *Differentially-Expressed miRNAs in Ectopic Stromal Cells Contribute to Endometriosis Development: The Plausible Role of miR-139-5p and miR-375.* Int J Mol Sci, 2018. **19**(12).
110. Tan, Q., et al., *MicroRNAs in Small Extracellular Vesicles Indicate Successful Embryo Implantation during Early Pregnancy.* Cells, 2020. **9**(3).
111. Estella, C., et al., *miRNA signature and Dicer requirement during human endometrial stromal decidualization in vitro.* PLoS One, 2012. **7**(7): p. e41080.
112. Tang, S. and Y. Dai, *RNA sequencing reveals significant miRNAs in Atypical endometrial hyperplasia.* Eur J Obstet Gynecol Reprod Biol, 2018. **225**: p. 129-135.
113. Shi, X.Y., et al., *Downregulation of miR-183 inhibits apoptosis and enhances the invasive potential of endometrial stromal cells in endometriosis.* Int J Mol Med, 2014. **33**(1): p. 59-67.

114. Chen, J., et al., *MiR-183 Regulates ITGB1P Expression and Promotes Invasion of Endometrial Stromal Cells.* Biomed Res Int, 2015. **2015**: p. 340218.
115. Akbar, R., et al., *miR-183-5p regulates uterine receptivity and enhances embryo implantation.* J Mol Endocrinol, 2020. **64**(1): p. 43-52.
116. Liu, H., et al., *Colorectal cancer-derived exosomal miR-106b-3p promotes metastasis by down-regulating DLC-1 expression.* Clin Sci (Lond), 2020. **134**(4): p. 419-434.
117. Moshiri, F., et al., *Circulating miR-106b-3p, miR-101-3p and miR-1246 as diagnostic biomarkers of hepatocellular carcinoma.* Oncotarget, 2018. **9**(20): p. 15350-15364.
118. Guo, R., et al., *A 9-microRNA Signature in Serum Serves as a Noninvasive Biomarker in Early Diagnosis of Alzheimer's Disease.* J Alzheimers Dis, 2017. **60**(4): p. 1365-1377.
119. Norsworthy, P.J., et al., *A blood miRNA signature associates with sporadic Creutzfeldt-Jakob disease diagnosis.* Nat Commun, 2020. **11**(1): p. 3960.
120. Qiao, G., et al., *Effects of miR106b3p on cell proliferation and epithelialmesenchymal transition, and targeting of ZNRF3 in esophageal squamous cell carcinoma.* Int J Mol Med, 2019. **43**(4): p. 1817-1829.
121. Mannavola, F., G. Pezzicoli, and M. Tucci, *DLC-1 down-regulation via exosomal miR-106b-3p exchange promotes CRC metastasis by the epithelial-to-mesenchymal transition.* Clin Sci (Lond), 2020. **134**(8): p. 955-959.
122. Gao, Q., et al., *Functional Passenger-Strand miRNAs in Exosomes Derived from Human Colon Cancer Cells and Their Heterogeneous Paracrine Effects.* Int J Biol Sci, 2020. **16**(6): p. 1044-1058.
123. Conti, B.J., et al., *Propolis modulates miRNAs involved in TLR-4 pathway, NF-kappaB activation, cytokine production and in the bactericidal activity of human dendritic cells.* J Pharm Pharmacol, 2016. **68**(12): p. 1604-1612.
124. Wu, H., et al., *IFN-tau Mediated Control of Bovine Major Histocompatibility Complex Class I Expression and Function via the Regulation of bta-miR-148b/152 in Bovine Endometrial Epithelial Cells.* Front Immunol, 2018. **9**: p. 167.
125. Yang, W., H.Y. Ju, and X.F. Tian, *Hsa-miR-4730 as a new and potential diagnostic and prognostic indicators for pancreatic cancer.* Eur Rev Med Pharmacol Sci, 2020. **24**(17): p. 8801-8811.
126. Lei, B., et al., *miR-615-3p promotes the epithelial-mesenchymal transition and metastasis of breast cancer by targeting PICK1/TGFBRI axis.* J Exp Clin Cancer Res, 2020. **39**(1): p. 71.
127. Chan, C., et al., *Discovery of biomarkers of endometrial receptivity through a minimally invasive approach: a validation study with implications for assisted reproduction.* Fertil Steril, 2013. **100**(3): p. 810-7.
128. Whittington, C.M., et al., *Transcriptomic changes in the pre-implantation uterus highlight histotrophic nutrition of the developing marsupial embryo.* Sci Rep, 2018. **8**(1): p. 2412.
129. Brosens, I., J.J. Brosens, and G. Benagiano, *The eutopic endometrium in endometriosis: are the changes of clinical significance?* Reprod Biomed Online, 2012. **24**(5): p. 496-502.
130. Osinski, M., et al., *HSD3B2, HSD17B1, HSD17B2, ESR1, ESR2 and AR expression in infertile women with endometriosis.* Ginekol Pol, 2018. **89**(3): p. 125-134.

131. Kitawaki, J., et al., *Progesterone induction of 17beta-hydroxysteroid dehydrogenase type 2 during the secretory phase occurs in the endometrium of estrogen-dependent benign diseases but not in normal endometrium.* J Clin Endocrinol Metab, 2000. **85**(9): p. 3292-6.
132. Duncan, W.C., et al., *Ectopic pregnancy as a model to identify endometrial genes and signaling pathways important in decidualization and regulated by local trophoblast.* PLoS One, 2011. **6**(8): p. e23595.
133. Lucas, E.S., et al., *Recurrent pregnancy loss is associated with a pro-senescent decidual response during the peri-implantation window.* Commun Biol, 2020. **3**(1): p. 37.
134. Chang, H.J., et al., *Overexpression of Four Joint Box-1 Protein (FJX1) in Eutopic Endometrium From Women With Endometriosis.* Reprod Sci, 2018. **25**(2): p. 207-213.
135. Kaitu'u, T.J., et al., *Matrix metalloproteinases in endometrial breakdown and repair: functional significance in a mouse model.* Biol Reprod, 2005. **73**(4): p. 672-80.
136. Zhang, J. and L.A. Salamonsen, *In vivo evidence for active matrix metalloproteinases in human endometrium supports their role in tissue breakdown at menstruation.* J Clin Endocrinol Metab, 2002. **87**(5): p. 2346-51.
137. Fukuda, T., et al., *Sema3A regulates bone-mass accrual through sensory innervations.* Nature, 2013. **497**(7450): p. 490-3.
138. Hayashi, M., et al., *Autoregulation of Osteocyte Sema3A Orchestrates Estrogen Action and Counteracts Bone Aging.* Cell Metab, 2019. **29**(3): p. 627-637 e5.
139. Ferreira, G.D., et al., *Expression of semaphorin class 3 is higher in the proliferative phase on the human endometrium.* Arch Gynecol Obstet, 2018. **297**(5): p. 1175-1179.
140. Liao, Y.X., et al., *Silencing of RTKN2 by siRNA suppresses proliferation, and induces G1 arrest and apoptosis in human bladder cancer cells.* Mol Med Rep, 2016. **13**(6): p. 4872-8.
141. Ji, L., et al., *RTKN2 is Associated with Unfavorable Prognosis and Promotes Progression in Non-Small-Cell Lung Cancer.* Onco Targets Ther, 2020. **13**: p. 10729-10738.
142. Liu, J., et al., *CatSperbeta, a novel transmembrane protein in the CatSper channel complex.* J Biol Chem, 2007. **282**(26): p. 18945-52.
143. Riesewijk, A., et al., *Gene expression profiling of human endometrial receptivity on days LH+2 versus LH+7 by microarray technology.* Mol Hum Reprod, 2003. **9**(5): p. 253-64.
144. Farimani Sanoee, M., et al., *Effect of myomectomy on endometrial glutathione peroxidase 3 (GPx3) and glycodelin mRNA expression at the time of the implantation window.* Iran Biomed J, 2014. **18**(2): p. 60-6.
145. Horcajadas, J.A., et al., *Controlled ovarian stimulation induces a functional genomic delay of the endometrium with potential clinical implications.* J Clin Endocrinol Metab, 2008. **93**(11): p. 4500-10.
146. Burmenskaya, O.V., et al., *Transcription profile analysis of the endometrium revealed molecular markers of the personalized 'window of implantation' during in vitro fertilization.* Gynecol Endocrinol, 2017. **33**(sup1): p. 22-27.

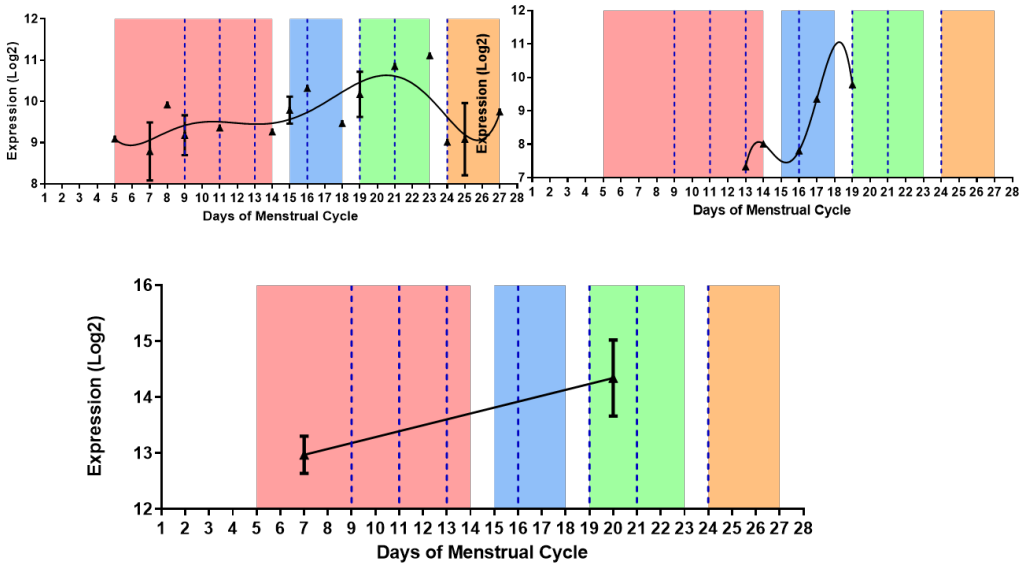
147. Hogan, M.C., et al., *PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression.* Hum Mol Genet, 2003. **12**(6): p. 685-98.
148. Fernandez, S.V. and J. Russo, *Estrogen and xenoestrogens in breast cancer.* Toxicol Pathol, 2010. **38**(1): p. 110-22.
149. Lenka, G., et al., *Identification of Methylation-Driven, Differentially Expressed STXBP6 as a Novel Biomarker in Lung Adenocarcinoma.* Sci Rep, 2017. **7**: p. 42573.
150. Liu, Y., et al., *Identification of STXBP6-IRF1 positive feedback loop in regulation of PD-L1 in cancer.* Cancer Immunol Immunother, 2020.
151. Zhao, Q. and C. Fan, *A novel risk score system for assessment of ovarian cancer based on co-expression network analysis and expression level of five lncRNAs.* BMC Med Genet, 2019. **20**(1): p. 103.
152. Sun, M., et al., *LncRNA PART1 modulates toll-like receptor pathways to influence cell proliferation and apoptosis in prostate cancer cells.* Biol Chem, 2018. **399**(4): p. 387-395.
153. Chen, J., et al., *Transcriptomic changes and potential regulatory mechanism of intrauterine human chorionic gonadotropin co-cultured with peripheral blood mononuclear cells infusion in mice with embryonic implantation dysfunction.* Ann Transl Med, 2020. **8**(4): p. 99.
154. Hasson, S.A., et al., *High-content genome-wide RNAi screens identify regulators of parkin upstream of mitophagy.* Nature, 2013. **504**(7479): p. 291-5.

APPENDIX

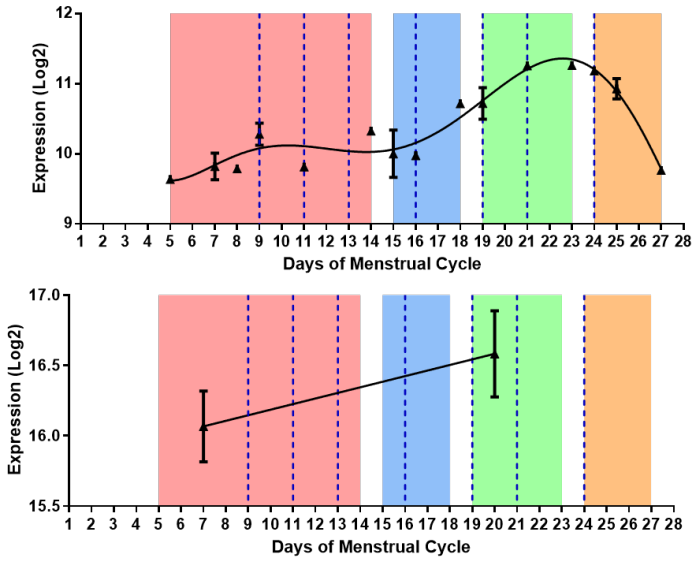
SUPPLEMENTARY FIGURES

Supplementary Figure 1: The remaining miRNA candidates obtained from *in-silico* prediction.

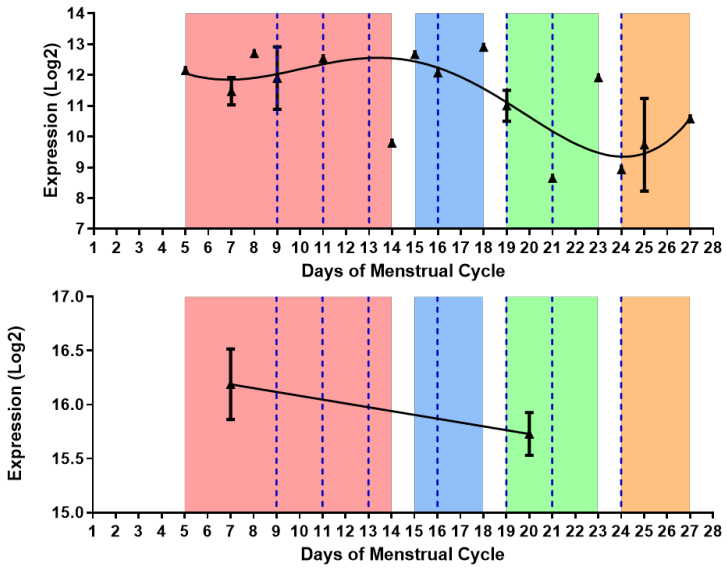
A. hsa-miR-30b-5p



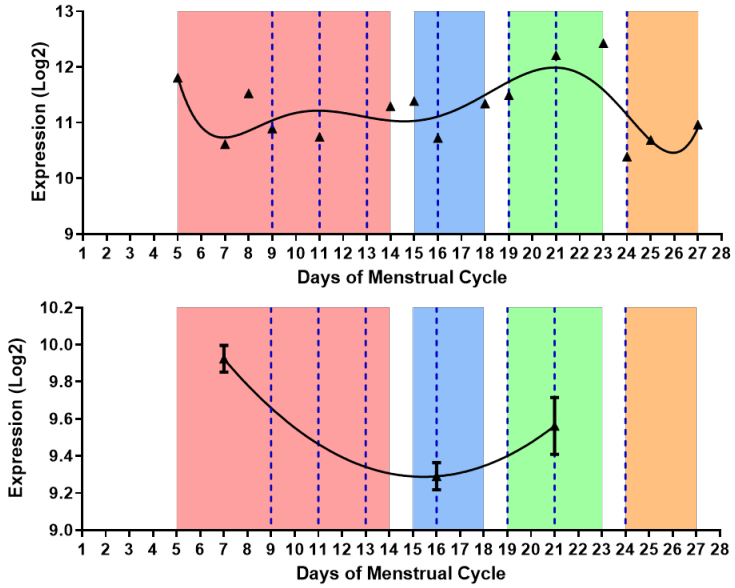
B. hsa-miR-30a-5p



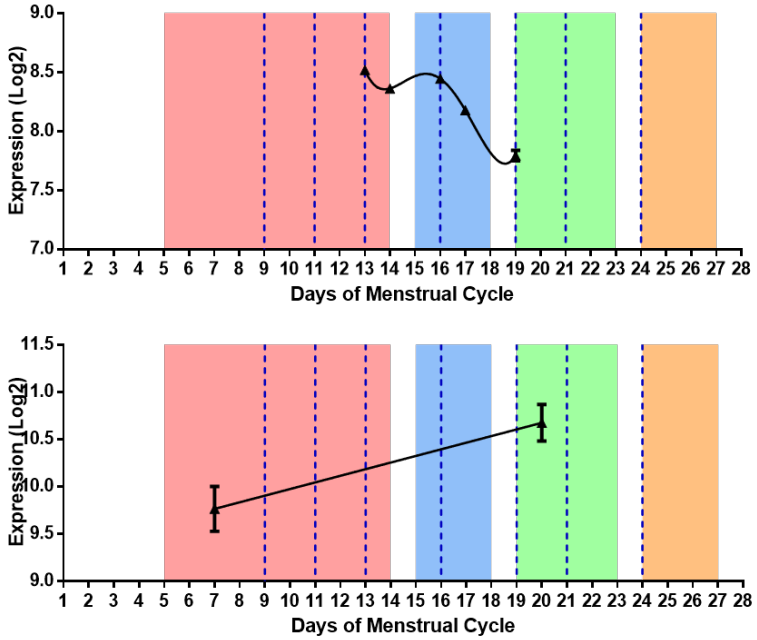
C. hsa-miR-199a-3p



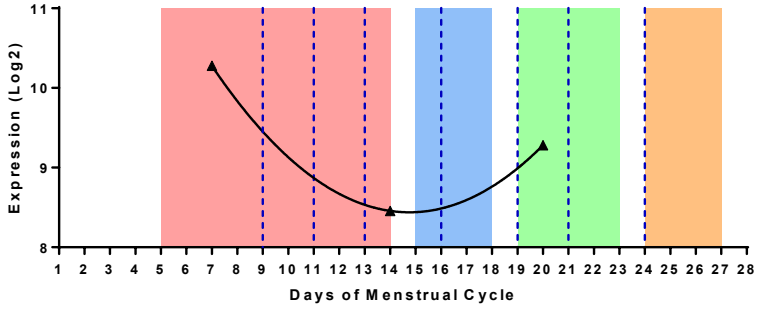
D. hsa-miR-29a-3p



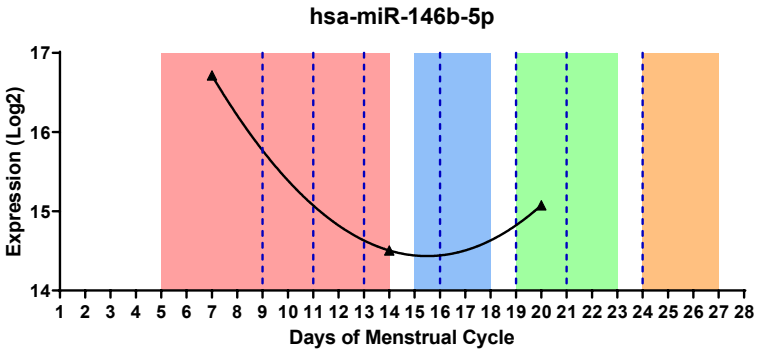
E. hsa-miR-130b-3p



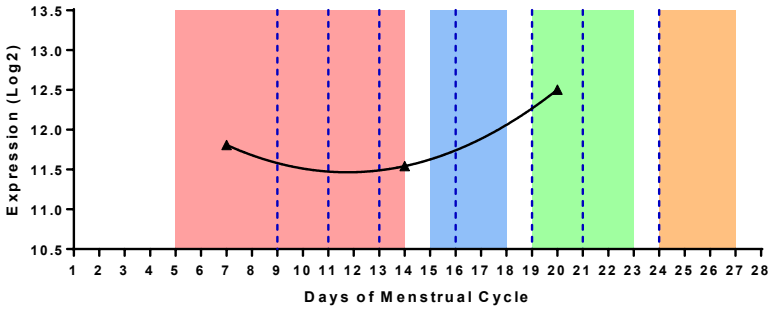
F. hsa-miR-150-5p



G. hsa-miR-146b-5p



H. hsa-miR-1307-3p



COPYRIGHT PERMISSIONS

Copyright Permission for Figure 1.1

ELSEVIER LICENSE

TERMS AND CONDITIONS

This Agreement between Mr. Ege Dedeoğlu ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4962510714678
License date	Dec 05, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease
Licensed Content Title	Molecular aspects of implantation failure
Licensed Content Author	Y.E.M. Koot,G. Teklenburg,M.S. Salker,J.J. Brosens,N.S. Macklon
Licensed Content Date	Dec 1, 2012
Licensed Content Volume	1822
Licensed Content Issue	12
Licensed Content Pages	8
Start Page	1943
End Page	1950
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1

Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	RNA BASED BIOMARKERS FOR PREDICTION OF THE ENDOMETRIAL WINDOW OF IMPLANTATION
Institution name	Bilkent University
Expected presentation date	Dec 2020
Portions	Fig 1
Requestor Location	Mr. Ege Dedeoğlu Ankara, Çankaya 06810 Turkey Attn: Bilkent University
Publisher Tax ID	GB 494 6272 12
Total	**0.00 USD**

Copyright Permission for Figure 1.2

OXFORD UNIVERSITY PRESS LICENSE

TERMS AND CONDITIONS

This Agreement between Mr. Ege Dedeoğlu ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4962570609269
License date	Dec 05, 2020
Licensed Content Publisher	Oxford University Press
Licensed Content Publication	Human Reproduction Update
Licensed Content Title	Identifying biomarkers for predicting successful embryo implantation: applying single to multi-OMICs to improve reproductive outcomes
Licensed Content Author	Hernández-Vargas, Purificación; Muñoz, Manuel
Licensed Content Date	Feb 25, 2020
Licensed Content Volume	26
Licensed Content Issue	2
Type of Use	Thesis/Dissertation
Requestor type	Educational Institution/Non-commercial/ Not for-profit
Format	Print and electronic

Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	RNA BASED BIOMARKERS FOR PREDICTION OF THE ENDOMETRIAL WINDOW OF IMPLANTATION
Institution name	Bilkent University
Expected presentation date	Dec 2020
Portions	Figure 1
Requestor Location	Mr. Ege Dedeoğlu Ankara, Çankaya 06810 Turkey Attn: Bilkent University
Publisher Tax ID	GB125506730
Total	**0.00 USD**

Copyright Permission for Figure 1.3

OXFORD UNIVERSITY PRESS LICENSE

TERMS AND CONDITIONS

This Agreement between Mr. Ege Dedeoğlu ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4962760126583
License date	Dec 05, 2020
Licensed Content Publisher	Oxford University Press
Licensed Content Publication	Human Reproduction Update
Licensed Content Title	ART and uterine pathology: how relevant is the maternal side for implantation?
Licensed Content Author	Galliano, Daniela; Bellver, José
Licensed Content Date	Aug 25, 2014
Licensed Content Volume	21
Licensed Content Issue	1
Type of Use	Thesis/Dissertation
Requestor type	Educational Institution/Non-commercial/ Not for-profit
Format	Print and electronic

Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	RNA BASED BIOMARKERS FOR PREDICTION OF THE ENDOMETRIAL WINDOW OF IMPLANTATION
Institution name	Bilkent University
Expected presentation date	Dec 2020
Portions	Figure 4
Requestor Location	Mr. Ege Dedeoğlu Ankara, Çankaya 06810 Turkey Attn: Bilkent University
Publisher Tax ID	GB125506730
Total	**0.00 USD**