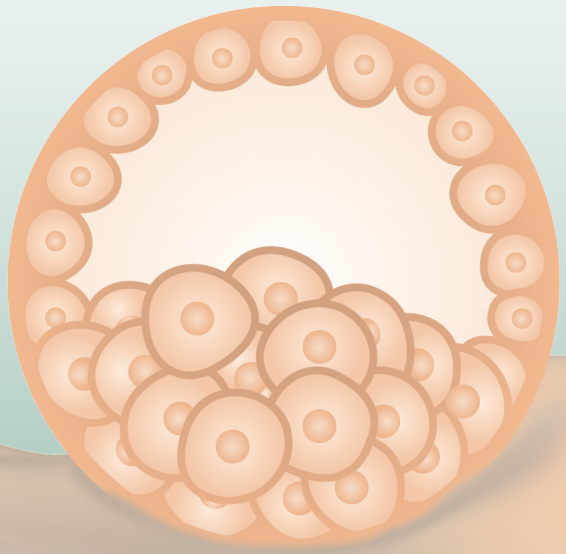
VNIVERSITAT DE VALÈNCIA

PhD Program in Biomedicine and Biotechnology



Pathological endometrial function characterization in the mid-secretory phase in infertile patients

Author:

Josefa María Sánchez Reyes

Bachelor´s degree in Basic and Experimental Biomedicine

Master´s degree in Genetics and Evolution

Directors:

Dr. Patricia Díaz Gimeno

Prof. José Alejandro Remohí Giménez

Valencia, February 2023

VNIVERSITAT DE VALÈNCIA

PhD Program in Biomedicine and Biotechnology

International PhD thesis

Pathological endometrial function characterization in the mid-secretory phase in infertile patients

Author:

Josefa María Sánchez Reyes

Bachelor's degree in Basic and Experimental Biomedicine

Master's degree in Genetics and Evolution

Directors:

Dr. Patricia Díaz Gimeno

Prof. José Alejandro Remohí Giménez

Valencia, February 2023

VNIVERSITAT DE VALÈNCIA

Dra. Patricia Díaz Gimeno, Doctora en Ginecología y Obstetricia, Investigadora Principal Miguel Servet del Instituto de investigación sanitaria La Fe y líder de grupo en Medicina Reproductiva Genómica y de Sistemas en Fundación IVI.

CERTIFICA:

Que el trabajo de investigación titulado: ***"Pathological endometrial function characterization in the mid-secretory phase in infertile patients"*** ha sido realizado íntegramente por Josefa María Sánchez Reyes bajo mi dirección. Dicha memoria está concluida y reúne todos los requisitos para su presentación y defensa como TESIS DOCTORAL ante un tribunal.

Y para que así conste a los efectos oportunos, firmo la presente certificación en Valencia a 30 de enero del 2023.

Fdo. Dra. Patricia Díaz Gimeno

VNIVERSITAT DE VALÈNCIA

Prof. José Alejandro Remohí Giménez, profesor en Ginecología y Obstetricia, fundador y co-presidente del Instituto Valenciano de Infertilidad (IVI).

CERTIFICA:

Que el trabajo de investigación titulado: "***Pathological endometrial function characterization in the mid-secretory phase in infertile patients***" ha sido realizado íntegramente por Josefa María Sánchez Reyes bajo mi dirección. Dicha memoria está concluida y reúne todos los requisitos para su presentación y defensa como TESIS DOCTORAL ante un tribunal.

Y para que así conste a los efectos oportunos, firmo la presente certificación en Valencia a 30 de enero del 2023.

Fdo. Prof. José Alejandro Remohí Gimenez

To the love of my life.

ACKNOWLEDGMENT

Desde que tengo memoria, siempre me he preguntado “el porqué de las cosas”, dicho por mi madre y de otro modo “no me callaba ni debajo del agua”. Pronto empecé a destacar en el ámbito académico y siempre consideré las asignaturas de ciencias mis favoritas en el colegio. Mi pasión por los animales hizo que en un principio quisiera ser veterinaria, pero mi tía Mari me hizo cambiar de opinión advirtiéndome que tendría que vivir situaciones muy desagradables.

Al inicio de la adolescencia, ya en el instituto, me empecé a interesar más por la alimentación y el deporte, queriendo ser nutricionista. Fue en bachillerato, movida por mis inquietudes en ciencias de la salud, cuando decidí estudiar el, en aquel entonces, novedoso Grado en Biomedicina Básica y Experimental de la Universidad de Sevilla.

Nada más iniciar la carrera, ya empecé a buscar mi pasión dentro del mundo de la ciencia, desde luego, la genética molecular era protagonista, pero también ganaba terreno la investigación aplicada a la clínica. Gracias a Migue, un amigo de mis tíos Manoli y Carlos (a los que siempre estaré agradecida), pude conocer de cerca el funcionamiento de un laboratorio de reproducción asistida y fue amor a primera vista. Esta disciplina aunaba técnicas moleculares y translación clínica, convirtiéndose en la opción perfecta para mí.

Sin embargo, por diversos motivos, no pude realizar el Máster en Reproducción Humana Asistida de la Universidad de Valencia, así que me decanté por mi otra pasión y estudié el Máster en Genética y Evolución de la Universidad de Granada. Tras finalizar, yo seguía queriendo trabajar en el campo de reproducción asistida y no pensaba rendirme hasta conseguirlo. Tras dos años de incesante búsqueda y de pluriempleo, por fin, surgió la oportunidad que tanto deseaba, y empecé a trabajar como estudiante predoctoral en

Fundación IVI (IIS La Fe) con una beca ACIF de la Universidad de Valencia. El día que me dieron la noticia, fue de los más felices de mi vida.

Tengo que decir, que todo este recorrido no fue nada fácil. Ya de niña, mi paso por el colegio fue muy duro debido a mi sobrepeso y mis logros académicos. Desde el inicio de la adolescencia, debía aprender a compaginar estudios y trabajo, ya que siempre que podía iba a echar una mano en el negocio familiar, valorando desde muy joven lo que era el esfuerzo y el sacrificio. Además, vengo de una familia chipionera de trabajadores, que, por circunstancias de la vida, no han podido cursar más allá de los estudios primarios o secundarios, por lo que, yo he sido pionera en este sentido.

Ahora bien ¿qué puedo decir de mi periodo predoctoral? pues que ha sido la etapa más dura de mi vida, pero a la vez la que más me ha enriquecido a nivel personal y profesional. Por suerte o por desgracia me ha tocado enfrentarme, sin mucha tregua, a multitud de circunstancias adversas a nivel laboral y personal, y considero que eso me ha hecho ser una mejor versión de mí misma. Durante este tiempo, he aprendido a caer y levantarme más fuerte, a valorarme más y a priorizar lo que realmente importa en la vida. Decidí que nunca más, nada ni nadie volvería a apagar mi alegría y borrar mi sonrisa. Actualmente, yo no sería quien soy si no hubiera pasado por todo esto y si no me hubiera topado con las personas que lo han hecho posible.

En primer lugar, dar las gracias a mis directores de tesis, en especial, a la Dra. Patricia Díaz Gimeno, por su paciencia, tiempo, empatía, profesionalidad y ser la mejor directora que se puede tener. Gracias por haberme dado la oportunidad de trabajar en este campo y aprender tantísimo sobre transcriptómica endometrial. Gracias por haberme apoyado por encima de todas las cosas y por haber intentado sacar siempre lo mejor de mí. Te estaré eternamente agradecida porque si no hubiera sido por ti, yo no habría sacado esta tesis adelante, ni habría vivido esta experiencia tan completa. Agradecer también al Prof. José

Remohí el haberme acogido como su doctoranda, aportando la dirección e infraestructura necesaria en esta etapa.

Al Prof. Antonio Pellicer, por sus aportaciones y permitirme desarrollar mi etapa predoctoral en Fundación IVI. Al Dr. Nicolás Garrido por su comprensión e implicación, así como haber sido el primero en brindarme la oportunidad de trabajar como estudiante predoctoral en Fundación IVI. También al Dr. Emre Seli y al Dr. Juan Antonio García Velasco. Thank you, Dr. Emre, for your suggestions during our meetings.

Por supuesto, gracias a todo el personal de las clínicas IVI Valencia, Bilbao, Barcelona y Madrid que ha contribuido al reclutamiento de pacientes y la obtención de muestras, en especial a Antonio Pellicer, Carmina Vidal, Imma Sánchez Ribas, Elena Labarta, Juan Antonio García Velasco, Juan Giles, Ernesto Bosch, Agustín Ballesteros, Gemma Castillón, Marcos Ferrando, Graciela Kohls, Francesca Gelosi, Laura Caracena, Isabel Llorens, Cristina Gaya, Marga Esbert, Mónica Toribio y Fernando Quintana. A todos los profesionales de los servicios de Biobanco y Genómica del IIS La Fe, incluidos los que siguieron otras trayectorias: Raquel Amigo, Carolina, Manu, Ibon, Cristina Cardona, Lola y Jose, por su profesionalidad y por tratarme siempre como una más. Gracias también a Lourdes Fernández y a Ester Castillo de Illumina®, por su asesoramiento técnico, así como a José Guillem de Lab Courier por haber estado al frente de los envíos.

A mis compañeros de equipo, del departamento de Medicina Reproductiva Genómica y de Sistemas. Empezando por la Dra. Patricia Sebastián, por su gran implicación y soporte a todos los niveles, mil gracias por haberme hecho aprender tanto sobre bioinformática y sobre la vida misma. A la Dra. Almudena Devesa, mi primera mentora y amiga, mil gracias por tu profesionalidad y tu ayuda. A Pablo García, por su gran corazón y estar ahí siempre que lo he necesitado. A Ismael y Antonio, porque además de ser grandes compañeros de trabajo, son los mejores amigos que se puede tener, me faltaría espacio en

estas páginas para agradeceros todas las risas, horas de esfuerzo compartidas, conversaciones intensas, apoyo incondicional, en definitiva, todos los momentos vividos y lo muchísimo que me habéis aportado. Agradecer a Diana Martí, por ser tan especial y haberse convertido en una de mis mejores amigas en tan poco tiempo, vales oro y te quiero mil. Gracias también a Asunta y Fran, que, aunque hayamos coincidido poco, ha sido un placer. Por supuesto agradecer a Alejandro Alemán que dejó huella por su trabajo y apoyo en el proyecto y a Manu, por su implicación y humor tan característico. Gracias también a mi querido David, por ser tan bueno y hacerme la mentora más afortunada, así como a Héctor y Miguel, por su positividad y apoyo.

A los investigadores principales Irene Cervelló, Paco Domínguez, Sonia Herraiz y Hortensia Ferrero por acogerme tan bien y ofrecerme su ayuda siempre que lo he necesitado. Por supuesto, gracias a mis compañeros de laboratorio, incluidos los que han seguido otros caminos: Ana B., Ana C., Hannes, Nuria, Silvia, María M., Andrea, Zaira, Gaby, Indra, Emilio, Elena, Pedro, Alba y María. En especial, dar las gracias a Amparo, Ali y Jessica, por su apoyo técnico siempre con una sonrisa. Gracias también a Sara (amabilidad personificada), Lucía (mi consejera top), Robert (mi loquillo preferido), Majo (mi cachorra), Luismi (super guerrero indiscutible), María G. (la anfitriona más increíble), Irene (la guía turística con más estilo), Anita (mi alcaldesa con gracia), Adolfo (el orador con más arte), Noe y Marina (mis risitas contagiosas), por haber estado siempre ahí, ante todo. A María Cristina, por haber sido una cura para mi autoestima y haberme regalado tanta sabiduría, “no te escondas nunca” porque eres increíble. A Yassmin, mi florecilla, por su cariño y ternura incondicionales, además de haber sido la huésped más maravillosa. Sois mis personas favoritas y os quiero muchísimo amigas. Gracias a todos por ser los mejores compañeros que se puede tener y por haberme hecho sentir como en casa. Mil gracias por esas quedadas, comidas, consejos, risas y sobre todo por hacer que

el periodo predoctoral parezca más bonito. Por supuesto, gracias a mi alma gemela canadiense-italiana, mi mejor amiga en fundación, Rosalba. Por tus correcciones en esta tesis, por ser tan maravillosa y regalarme tantos buenos momentos. Gracias a la vida por haberme hecho coincidir en esta etapa contigo. Gracias a tus preciosas hijas Serena y Sophie, haber sido una luz en el camino, fruto de la historia de amor entre tú y mi querido Ismael, que tuve la suerte de vivir desde el principio.

A todos los profesionales de soporte, administración, UAGI e IVI *education* con los que he tenido la suerte de coincidir: Inma, Carmen, Davinia, Regina, Guille, Dago, Alba B., Irene U, Pablo G, Juanma, Héctor, Silvia, Lucía C., Mª Luisa, María, Helena y Koke por regalarme tan buenos ratos. En especial, a Marcos, por ser el primero en ayudarme con los trámites administrativos; Leo, por su carisma y solventarme los problemas con Windows; Alfredo, por su esfuerzo con las exportaciones y sus "buenos días" matutinos; Lorena, por su fortaleza y positividad; Elena, por su bondad; Alba, por ser tan detallista; y Loreto, por su apoyo a pesar de la distancia. Por supuesto, gracias a mi amigo Víctor, por tener siempre un hueco para mí en su mini amarillo y ser tan espectacular a todos los niveles; y a mi querido Alejandro, por ser tan bueno, comprensivo, creer tanto en mí y estar ahí por encima de todo. Gracias por darle verdadero significado a la palabra amistad, no cambies jamás. Gracias también a mi querida Eva, Paula, Josep y Nando, por hacer real que las parejas de mis amigos, sean mis amigos.

To staff of Juno Genetics: Tishy, Dhruti, Araz, Elpida, Louisa, Nahomi, Georgina, Columba, Hadis, Megan, Kishlay, Ayman, Millie, Suleman, Zeynep, Alexander, Nada etc. Thank you for kindly welcoming me into your laboratory and for making my stay in Oxford an amazing experience. I would like to give special thanks to Prof. Dagan Wells and Dr. Katharina Spath for their kindness and support. Por supuesto, dar las gracias a mi maravillosa compañera de estancia y amiga, Nuria Soler, gracias por tu apoyo y todos los

momentos vividos. Sin ti la estancia no habría sido lo mismo. También agradecer a Nuria y Tudor, su hospitalidad y amabilidad durante la estancia. Gracias también a mis nuevos compañeros de trabajo y amigos de Juno Genetics España: Nacho, Carlos, Elena, Marta, Michelle, Blai, Óscar, Sara, Roberto, Celia, Javi, Marta y Marcos, con especial mención a Claudia, Macarena y Carmen. Gracias por regalarme vuestro apoyo cada día y hacer que todo merezca la pena, sois los mejores y me siento muy afortunada.

Gracias también a todos los profesionales del Colegio Público Principe Felipe, el IES Caepionis, el IES Salmedina, la Universidad de Sevilla, la Universidad de Granada y la Universidad de Valencia por haber hecho posible mi formación. Especial mención a mis maestros del colegio María José, Ana María, Esther, Juan Manuel, Luis y Pepe. A mis profesores del instituto Javi, Fedriani, María, Juan Emilio, José Antonio, Alfonso, Joaquín, Luis, Pablo, Abraham, Antonio, Moisés, Mercedes, Salvador, Pepe Villagrán, Pepe Mellado y Pepe Castro. A mis profesores y/o tutores del grado y máster Rafael Fernández Chacón, José López Barneo, Patricia Ortega Sáenz, José Casadesús, Javier Vitorica, Carmen Garnacho, Antonio Núñez, Francisco Romero Campero, Pablo Huertas, Sebastián Chávez, Alberto Pascual, María Luz Montesinos, Carolina Sousa, Miguel Burgos, Marta Pérez, Juan Pedro Martínez, Rafael Jiménez y Josefa Cabrera. Dar las gracias también a Lázaro, George, Samantha y Tomi por ser los mejores profes de inglés.

A mis maravillosos amigos de Chipiona: Andrés, Jairo, Bea, Toni, Joaquín, Patri, Loly, Jesús, Ramón, Inma, Ángel, María José, Fran, Camilo, Vane M., Juan, Silvia, Christian, Diego, Nazaret, Rocío, Valeria y Murga, con especial mención a mi Vane, Conchita y Regli, por ser los mejores amigos que se puede tener, regalarme tan buenos momentos y haberme apoyado siempre a pesar de la distancia. A mi familia biomédica, mi queridísimo "Bar Coyote Hundo": Ismael, Jesús, Manolo, Álvaro, Laura, Juanlu, Zarza y Andrea, con especial mención a las integrantes del "*Sex and the City*", Marina, Carmen y Susana.

Gracias por haber sido mi máximo pilar a lo largo de estos años, por todas las anécdotas, conversaciones y alegrías que me habéis dado. Gracias por ser únicos y eternos. Os adoro.

A mi familia: tita Manoli, tito Carlos, tita Mari, primo Carlos, mi hermano Antonio, mi cuñada Noelia, mi madrina Cande, mis primas Carmen y Helena, mi cuñado Adri, mi concuñada Carmen, mis suegros Pili y Paco, a todos mis titos y primos lejanos o postizos, por creer en mí por encima de todas las cosas, llorar con mis penas y reír con mis alegrías. Gracias por apoyarme durante toda mi vida hasta con distancia de por medio. Abuelitos míos, a vosotros también os lo agradezco, que sé que desde algún lugar me seguís cuidando. A mis perritos Laika, Linda, Iker, Khaleesi y Kora por ser mis compañeros más leales y cariñosos durante esos largos días de estudio y trabajo. En especial agradecer a mis queridos sobrinos Antoñito y Ale, por llenar mi vida de luz y ternura, estoy orgullosa de ser vuestra tata. Por supuesto a mis padres, Figenia y Antonio, gracias a vosotros soy la mujer que soy hoy y todo lo debo a vuestro apoyo, sacrificio, paciencia, comprensión, amor incondicional y los valores que me habéis transmitido. Os quiero y no sé qué haría sin todos y cada uno de ustedes.

Finalmente, quería dedicar esta tesis a ti, al amor de mi vida, mi Fran. Gracias por ser mi compañero de vida y darme tu apoyo día tras día. Gracias por quererme tanto y hacerme la mujer más afortunada del mundo. Gracias por ser como eres y por seguirme donde quiera que vaya. Gracias por ser mi pasado, mi presente y mi futuro. Sin ti nada de lo que soy hoy y de lo que he conseguido habría sido posible. Te quiero.

GRACIAS A TODOS

"Si una persona es perseverante, aunque sea dura de entendimiento, se hará inteligente; y aunque sea débil, se transformará en fuerte."
Leonardo da Vinci

El presente trabajo de tesis doctoral ha sido realizado en el grupo de Medicina Reproductiva Genómica y de Sistemas de Fundación IVI, que forma parte del grupo de Biomarcadores, Medicina Genómica, Estadística y Análisis masivo de datos en Reproducción Humana Asistida del Instituto de Investigación Sanitaria (IIS) La Fe, en colaboración con el Dpto. de Pediatría, Obstetricia y Ginecología de la Facultad de Medicina de la Universidad de Valencia, en España. Así como en las clínicas IVI de Valencia, Madrid, Barcelona y Bilbao, y en los laboratorios de Juno Genetics asociados al *Nuffield Department of Women´s & Reproductive Health* de la Universidad de Oxford, en Reino Unido.

El desarrollo de esta tesis ha sido posible gracias al proyecto PI19/00537 y el contrato Miguel Servet (CP20/00118), financiados por el Instituto de Salud Carlos III (Ministerio de ciencia e innovación) y cofinanciados por la Unión Europea (FEDER) "Una manera de hacer Europa", además del proyecto 1706-FIVI-048-PD (Fundación IVI), todos ellos asignados a la Dra. Patricia Díaz Gimeno. Así como a las subvenciones del programa ACIF de la Generalitat Valenciana para personal investigador en formación de carácter predoctoral (ACIF/2018/072) y del programa BEFPI para estancias de contratos predoctorales en centros de investigación fuera de la Comunidad Valenciana (BEFPI/2020/028) asignadas a la doctoranda Josefa María Sánchez Reyes.

RESUMEN

INTRODUCCIÓN

El endometrio humano. Anatomía y función.

El útero es un órgano interno localizado en la zona pélvica que forma parte del aparato reproductor femenino en mamíferos. En humanos, el útero tiene forma simple, de pera invertida y una cavidad triangular. Este órgano puede dividirse en tres zonas: fundus (región superior con forma de cúpula), cuerpo (región que conecta con los ovarios a través de las trompas de falopio) y el istmo (región estrecha entre el cuerpo y el cuello del útero). Por su parte, la pared uterina está dividida en tres capas: perimetrio (la más externa y delgada), miometrio (capa media de músculo liso) y endometrio (capa mucosa interna que comunica con la cavidad uterina), siendo esta última la más compleja y objeto de estudio en esta tesis doctoral. El endometrio puede dividirse en cuatro compartimentos: epitelial (que se divide en epitelio luminal y glándulas epiteliales), estromal (compuesto por matriz extracelular y células estromales endometriales), vascular (formado por la red de capilares que se extienden desde el miometrio) e inmune (compuesta principalmente por células T, células B, macrófagos y células *natural killer* uterinas). Además, puede ser dividido en dos regiones: funcional (parte de los compartimentos endometriales que se someten a cambios durante la menstruación) y basal (parte de los compartimentos endometriales invariable y que funciona como suplemento para la regeneración de una nueva capa funcional).

Por otro lado, el endometrio es un tejido multicelular dinámico que cambia cíclicamente en términos de función y apariencia durante el ciclo menstrual. El ciclo menstrual humano dura unos 21-35 días y está orquestado por niveles cíclicos hormonales inducidos por el eje hipotálamo-hipofisario. La hormona liberadora de gonadotropina (GnRH; del inglés, *gonadotropin releasing hormone*), liberada por el hipotálamo, induce la liberación cíclica

de la hormona estimulante de folículos (FSH; del inglés *follicle-stimulating hormone*) y de la hormona luteinizante (LH; del inglés *luteinizing hormone*) en la hipófisis. FSH y LH se unen a sus receptores ováricos correspondientes e inducen la secreción de estrógenos y progesterona respectivamente, que actuarán a nivel endometrial. El ciclo menstrual (idealmente de 28 días) puede dividirse en tres fases: menstruación (0-5 días), proliferativa (5-14 días) y secretora (14-28 días). La menstruación consiste en la renovación del compartimento funcional del endometrio en ausencia de implantación embrionaria, y se caracteriza por un decrecimiento de los niveles hormonales. La fase proliferativa consiste en un engrosamiento del tejido endometrial frente a los altos niveles de estrógenos, mientras que la fase secretora consiste en la diferenciación del tejido endometrial para su preparación para la implantación embrionaria en respuesta a los altos niveles de progesterona. La principal función del endometrio es precisamente, desarrollar su capacidad receptiva para hacer posible la implantación embrionaria y mantener el embarazo, situándose el periodo de máxima receptividad endometrial en la fase secretora media. Se trata de un periodo corto situado entre los días 19-24 del ciclo menstrual, denominado ventana de implantación, (WOI; del inglés *window of implantation*), que es complejo, multifactorial y variable en tiempos entre las mujeres implicando cambios a nivel molecular, celular y tisular. El estudio de la WOI resulta clave para la comprensión del potencial reproductivo de la mujer y posibles problemas de fertilidad.

La implantación embrionaria requiere de un diálogo sincronizado entre el embrión con buena calidad y el endometrio receptivo durante la WOI. La WOI está regulada por una amplia variedad de citoquinas, factores de crecimiento, prostaglandinas, enzimas y moléculas de adhesión entre otras, para coordinar que las distintas fases de la implantación tengan lugar: aposición (primer contacto entre blastocisto y el endometrio receptivo), adhesión (las células trofoblásticas del blastocisto se unen al epitelio

endometrial) e invasión (las células trofoblásticas cruzan el epitelio endometrial y se expanden hacia el estroma para alcanzar los vasos sanguíneos maternos). La implantación va seguida del embarazo temprano (hasta 12 semanas), siendo la tolerancia inmunológica materno-fetal un mecanismo molecular clave para el establecimiento de este periodo. El estudio de los eventos moleculares coordinados que tienen lugar durante la implantación y el embarazo temprano es importante también para comprender posibles problemas de fertilidad.

Infertilidad relacionada con el factor endometrial. Fallo de implantación recurrente.

La infertilidad afecta a un 8-12% de parejas en edad reproductiva y se define como "una enfermedad del aparato reproductor asociada a fallos para lograr embarazo clínico tras 12 meses o más de vida sexual regular sin protección". Este hecho ha aumentado la importancia y la popularidad de los tratamientos de reproducción asistida (TRAs). La infertilidad de origen femenino supone un 20-35% de los casos y puede ser debida a la edad, problemas endocrinos, alteraciones inmunológicas, infecciones del tracto genital, estilo de vida y desórdenes uterinos. Entre estos desórdenes uterinos destaca el fallo de implantación recurrente (FIR), en el que se centra esta tesis doctoral. Si bien es cierto que el FIR no presenta una definición universal, suele estar asociado a la ausencia de implantación tras al menos tres transferencias con embriones de buena calidad en los TRAs. El término "fallo de implantación" implica que el embrión no implanta en el endometrio en las fases de adhesión o invasión (ausencia de gonadotropina coriónica beta (β-hCG) en orina o sangre), o bien en fases muy tempranas tras la invasión (detección de β-hCG pero no la formación de saco gestacional, es decir, aborto bioquímico). Mientras que el término fallo de implantación puede aplicarse tanto a gestaciones espontáneas

como a pacientes de TRA, el FIR sólo puede aplicarse en el contexto de TRA. En cuanto a la prevalencia de FIR, tampoco está clara, pero ha sido estimada entre 5-66,7%.

El FIR es una patología muy compleja de etiología múltiple. Las causas del FIR pueden proceder del embrión o de la madre. En el factor embrionario quedan englobadas las causas gamética, genética y las condiciones del TRA. Posibles medidas diagnósticas pueden ser: buenos sistemas de testeo de la integridad del ácido desoxirribonucleico (ADN) espermático, empleo de inyección intracitoplasmática de espermatozoides seleccionados morfológicamente, evaluación morfológica adecuada de ovocitos y espermatozoides, revisiones periódicas de los protocolos de TRA, uso del sistema *time lapse* para la observación morfológica de los embriones o diagnóstico genético preimplantacional. En cuanto a medidas terapéuticas, tendríamos el establecimiento de protocolos de estimulación ovárica adecuados, un buen sistema de selección de gametos y embriones, programas de donación de ovocitos y esperma, y optimización de los protocolos de TRA. Por otro lado, en el factor materno quedan englobados el estilo de vida, alteraciones uterinas (hiprosálpinx, miomas, pólipos, grosor endometrial etc.), problemas endocrinos (tiroides, diabetes y niveles de prolactina), las condiciones del TRA y otras posibles comorbilidades como trombofilias, síndrome antifosfolípido, alteraciones inmunológicas o endometritis. Algunas opciones diagnósticas son las evaluaciones personalizadas de las pacientes, ultrasonidos, histeroscopia, histerosalpingograma y análisis de sangre. En cuanto a opciones terapéuticas tendríamos las recomendaciones saludables, disección quirúrgica, salpingectomía, tratamiento hormonal y optimización de los protocolos de TRA. En conclusión, el FIR no sólo presenta falta de consenso en cuanto a definición y prevalencia, sino también en el establecimiento de medidas diagnósticas y terapéuticas estandarizadas considerando todas las posibles causas. La investigación en esta línea podría ayudar al adecuado diagnóstico y tratamiento de las

pacientes y por tanto a disminuir la carga psicológica, económica y de consumo de tiempo tras repetidos fallos de implantación.

Sin embargo, cuando el fallo de implantación no se puede atribuir a ninguna de estas causas, en la literatura se habla normalmente de FIR de origen endometrial. El FIR endometrial es el más difícil de diagnosticar y tratar, ya que puede ocurrir en un útero que en apariencia es normal según técnicas de imagen, o que es anormal, pero fue tratado farmacológica o quirúrgicamente, tras descartar el factor embrionario y otras causas ambientales. Las dos posibles causas recientemente descritas de FIR endometrial son la presencia de una WOI desplazada o asíncrona con el desarrollo del embrión y/o una WOI alterada o patológica, siendo esta última en la que nos focalizamos en esta tesis doctoral. El diagnóstico del FIR endometrial requiere del uso de tecnologías avanzadas que permitan estudiar la causa subyacente molecular y desarrollar estrategias terapéuticas personalizadas, resultando interesante seguir investigando en esta línea.

Medicina de precisión y estratificación transcriptómica.

Los avances tecnológicos de las dos últimas décadas han conducido a un cambio en el estudio de los procesos biológicos, pasando de una aproximación de "molécula a molécula" a la evaluación simultánea de todos los genes, transcritos o proteínas presentes en una célula o tejido en cualquier condición y tiempo; resultando en la aparición de las denominadas "ciencias ómicas" tales como genómica, transcriptómica o proteómica, respectivamente, entre otras. Las llamadas "ciencias ómicas" están jugando un papel clave para los avances en la medicina de precisión en la práctica clínica pues permiten caracterizar molecularmente enfermedades complejas. La medicina personalizada o de precisión se describe como el conjunto de estrategias de prevención, diagnóstico y tratamiento tomadas a medida en función de las características individuales de los

pacientes. También se denomina como medicina estratificada, ya que suele centrarse en la estratificación de la población de pacientes (nueva taxonomía) empleando datos a gran escala de origen clínico, estilo de vida, comportamiento, características genéticas y otros biomarcadores, yendo más allá de la típica aproximación basada en "signos y síntomas". Establecer nuevas taxonomías o grupos dentro de la enfermedad permite mejorar su tratamiento e identificación, debido a que proporciona un mayor conocimiento sobre las causas críticas que la originan. Entre las ciencias ómicas, la transcriptómica es una herramienta muy útil en medicina de precisión porque permite el estudio global de la expresión génica relacionada con las enfermedades, y ha resultado clave para el estudio de la función endometrial. Concretamente, un conjunto de genes cuya expresión génica es característica de un fenotipo concreto se conoce como firma génica y puede ser empleado como biomarcador de una enfermedad.

Actualmente, una de las principales tecnologías de alto rendimiento para la obtención de datos transcriptómicos es el *RNA-Sequencing* (RNA-Seq), que queda englobado dentro de la secuenciación de última generación (NGS¸ del inglés *next-generation sequencing*). La NGS incluye una serie de técnicas y aproximaciones que permiten la rápida determinación de las secuencias de millones de fragmentos de ADN de forma paralelizada, masiva y automatizada en un único proceso de secuenciación. Una de las principales plataformas de RNA-Seq empleada en la actualidad es Illumina®, debido a su alto rendimiento y los múltiples protocolos adaptados a todo tipo de aplicaciones que dispone con metodologías bien establecidas para evitar sesgos y perfiles erróneos. Todos los protocolos Illumina® tienen una serie de fases comunes: (1) extracción de ácido ribonucleico (ARN) de la muestra de tejido, (2) retrotranscripción del mismo a ADN complementario (ADNc), (3) ligación de los fragmentos de ADNc con los adaptadores

(permiten la identificación y secuenciación), (4) amplificación mediante la reacción en cadena de la polimerasa (PCR; del inglés *polymerase chain reaction*), (5) selección de los fragmentos (purificación por tamaño) y (6) secuenciación por síntesis (se obtendrán las lecturas de cada transcrito, directamente proporcional al nivel de expresión).

Una vez que los datos transcriptómicos son generados, estos deben ser analizados computacionalmente empleando una serie de aproximaciones bioinformáticas, que comúnmente son: (1) pre-procesamiento (asignación de las lecturas a cada muestra según los adaptadores, controles de calidad, alineamiento de las lecturas con respecto a un genoma de referencia para obtener los conteos, filtros de calidad y normalización), (2) análisis exploratorio (detección y eliminación de muestras que no siguen el patrón de comportamiento y corrección de efectos tanda), (3) análisis de expresión diferencial (obtención de genes expresados diferencialmente en las distintas condiciones de estudio, como patología vs. control), (4) análisis funcional (estudio de los procesos biológicos o las rutas en las que están implicados los genes diferencialmente expresados). Además, la inteligencia artificial es una de las aproximaciones bioinformáticas más útiles para llevar a cabo estratificación transcriptómica en el contexto de una medicina de precisión. Esta engloba al aprendizaje automático (ML; del inglés *machine learning*), que se define como el estudio de algoritmos computacionales que pueden mejorar automáticamente, aprendiendo de los datos e identificando patrones, con intervención humana limitada. En transcriptómica, estos algoritmos son aplicados a los datos de expresión génica normalizados y corregidos, usando aprendizaje supervisado o no supervisado. El aprendizaje no supervisado ocurre usando muestras no etiquetadas con el objetivo de estructurar los datos y definir perfiles transcriptómicos. Sin embargo, el aprendizaje supervisado se aplica a muestras etiquetadas (ej. patológica o control) para descubrir

nuevas firmas de genes, realizar predicciones de muestras desconocidas y clasificarlas según su perfil transcriptómico. Los principales algoritmos de predicción recomendados en clínica son *support vector machine* (SVM), *k-nearest neighbors* (kNN) y *random forest* (RF). Este tipo de modelos son denominados predictores transcriptómicos y se desarrollan y validan internamente a partir de un conjunto de muestras de referencia denominado "conjunto de entrenamiento", mientras que se validan externamente en un conjunto independiente de muestras denominado "conjunto de prueba".

En conclusión, enfermedades complejas tales como el FIR endometrial debido a una WOI patológica pueden ser abordadas desde el punto de vista de la medicina de precisión, siendo la transcriptómica, así como las tecnologías para su obtención y las herramientas bioinformáticas asociadas, claves para la estratificación de estas enfermedades y su estudio molecular.

Aplicaciones de la transcriptómica endometrial y contexto del estudio.

El endometrio es un tejido dinámico a nivel molecular y morfológico que implica fenotipos complejos y multifactoriales relacionados con la función endometrial, tales como la receptividad endometrial y el FIR de origen endometrial. El estudio de la WOI y sus alteraciones puede ayudar a determinar el estatus endometrial óptimo en clínica. Clásicamente, la WOI era estimada determinando los días del ciclo menstrual, el tiempo de ovulación, el nivel hormonal o los métodos histológicos. Otras aproximaciones han incluido los ultrasonidos, el aspirado del fluido endometrial, la histeroscopia o la identificación de biomarcadores moleculares únicos relacionados con el proceso de implantación. Sin embargo, con el advenimiento de las "ciencias ómicas", la tendencia

ha ido cambiando hacia el estudio de biomarcadores multi-ómicos de la receptividad endometrial.

Concretamente, la transcriptómica ha demostrado ser un método preciso y reproducible para el estudio de la WOI y ofrecer más información molecular del endometrio que permita distinguir nuevos patrones en los procesos y enfermedades en dirección a una medicina personalizada. De hecho, se han realizado numerosos estudios transcriptómicos para definir la WOI sana, a menudo comparando el endometrio en diferentes fases del ciclo menstrual o comparando pacientes FIR con controles para caracterizar alteraciones de la WOI. Estos estudios propusieron diferentes firmas de genes como biomarcadores endometriales, y aunque la mayoría de ellas fueron publicadas para una mayor comprensión molecular de la función endometrial, otras firmas han sido empleadas para desarrollar predictores transcriptómicos para la estratificación de pacientes y/o procedimientos diagnósticos.

Los predictores transcriptómicos endometriales para la evaluación de la WOI y el FIR endometrial han sido aplicados en dos estudios principales. El primer estudio, de la Dra. Diaz-Gimeno y colaboradores, pionero en el uso de predictores con datos de endometrio, determinó una firma transcriptómica que fue patentada y empleada para el desarrollo del análisis de receptividad endometrial (ERA test®; del inglés *endometrial receptivity analysis*), una herramienta diagnóstica para identificar la WOI. Esta firma transcriptómica permitió también describir que el FIR endometrial puede ser originado a partir de la asincronía entre el embrión y el endometrio debido a un desplazamiento del periodo de la WOI y se ha empleado en clínica para la personalización de transferencias embrionarias en ciclo sustituido (HRT; del inglés *hormone replacement therapy*), aunque los beneficios

son muy controvertidos y hay que seguir investigando. En contraste, el otro estudio, de la Dra. Koot y colaboradores, propone una firma transcriptómica considerando el FIR endometrial resultado de una WOI patológica, con independencia del desplazamiento, que muestra un carácter muy heterogéneo. Si bien es cierto que este estudio fue pionero en la corrección del efecto de la progresión endometrial previa al estudio de la patología para evitar posibles sesgos, estuvo focalizado en el ciclo natural (siendo HRT el más empleado para la preparación endometrial en clínica) en una población desbalanceada (más pacientes controles que FIR) y los resultados fueron insuficientes para desarrollar un procedimiento diagnóstico estandarizado para las pacientes.

Con el objetivo de clarificar estos controvertidos resultados, otro estudio también basado en predictores transcriptómicos, fue llevado a cabo por el grupo de la Dra. Diaz-Gimeno y demostró que el FIR endometrial puede estar originado por ambas, una WOI desplazada y/o una WOI patológica. Esto no sólo permitió la descripción de una nueva taxonomía de FIR, sino también la introducción de una metodología que permitiera distinguir clínicamente entre una paciente que pudiera beneficiarse de una transferencia embrionaria personalizada por tener una WOI desplazada y una paciente con una WOI patológica que debe ser caracterizada para su traslación a la clínica como procedimiento diagnóstico. En este estudio, las pacientes estaban en ciclo natural y no fue posible realizar un seguimiento clínico debido a la naturaleza *in-silico* de los datos procedentes de la Dra. Koot y colaboradores, así pues, son necesarios más estudios prospectivos clínicos para corroborar la existencia de la WOI patológica. Además, el FIR endometrial debido a una WOI patológica parece englobar un perfil heterogéneo poco conocido molecularmente y se requieren más estudios transcriptómicos para su caracterización y explorar nuevas estrategias para evaluar el factor endometrial en pacientes infértiles.

OBJETIVOS

El **objetivo principal** es identificar y caracterizar la ventana de implantación patológica en una cohorte de pacientes de fecundación *in vitro* con preparación endometrial mediante ciclo sustituido.

Los **objetivos específicos** son:

1. Desarrollar un modelo de predicción capaz de identificar la ventana de implantación patológica independientemente de los desplazamientos en las pacientes en ciclo sustituido.
2. Estratificar la población de pacientes para definir los diferentes perfiles transcriptómicos relacionados con la ventana de implantación patológica.
3. Identificar las asociaciones clínicamente relevantes de los grupos definidos transcriptómicamente con los resultados reproductivos.
4. Estudiar los mecanismos moleculares y las alteraciones funcionales entre los perfiles transcriptómicos definidos.

METODOLOGÍA

Diseño del estudio y obtención de muestras endometriales de las pacientes.

Las pacientes de este estudio prospectivo multicéntrico promovido por Fundación IVI fueron reclutadas entre enero de 2019 y agosto de 2020 en las clínicas IVI de Valencia, Madrid, Barcelona y Bilbao. Los criterios de inclusión fueron: pacientes de fecundación *in vitro* (FIV) recomendadas para evaluación endometrial con preparación endometrial mediante HRT con estrógenos y progesterona sin análogos de GnRH; calidad embrionaria garantizada por diagnóstico genético preimplantacional o por pertenecer a un programa

de donación (< 35 años); tener indicación de transferencia embrionaria personalizada en el contexto de una microinyección intracitoplásmica de espermatozoide (ICSI; del inglés, *intracytoplasmic sperm injection*); edad = 18-50 años; índice de masa corporal (IMC) = 19-30 kg/m^2; y grosor endometrial > 6,5 mm con aspecto trilaminar en el décimo día del ciclo menstrual. Los criterios de exclusión fueron: tener como única indicación de tratamiento factor masculino con semen propio; patologías uterinas serias no tratadas; síntomas pre-menopáusicos severos; enfermedades metabólicas y/o sistémicas graves no controladas; y tratamiento concomitante que pudiera interferir con el tratamiento reproductivo. El estudio fue aprobado por el comité ético del Instituto Valenciano de Infertilidad (1706-FIVI-048-PD) y todas las participantes reclutadas firmaron el consentimiento informado. El seguimiento clínico fue llevado a cabo durante el periodo del estudio y la información clínica de interés fue exportada de la base de datos SIVIS siguiendo la ley de protección de datos. Las pacientes fueron clasificadas según la definición clínica de FIR como FIR (≥ 3 fallos de implantación) y controles (< 3 fallos de implantación), entendiendo como fallo de implantación un resultado de β-hCG negativa o aborto bioquímico tras la transferencia con un embrión de buena calidad. Las poblaciones fueron comparadas considerando las variables de confusión más importantes y su comportamiento transcriptómico mediante un análisis de componentes principales (ACP) en R (versión 4.0.5, 2021-03-31).

Tal como está establecido en el protocolo habitual de la evaluación endometrial en las clínicas IVI, la biopsia endometrial fue obtenida del fondo uterino mediante una cánula Pipelle de Cornier® (CCD Laboratories, Paris, Francia) en condiciones estériles durante la ventana de implantación, normalmente 5 días después del inicio de la toma de progesterona en el ciclo HRT. La biopsia sobrante, destinada para este estudio, fue

introducida en un 1,8 mL Nunc cryotube® (Thermo scientific, Madrid, España) con RNAlater® (Sigma-Aldrich, Madrid, España). Los criotubos fueron mantenidos a 4ºC como máximo 1 mes hasta ser enviados a Fundación IVI (IIS La Fe) donde fueron limpiadas con Dulbecos PBS® (Capricorn Scientific, Labclinics, Barcelona, España), secadas mediante pases a través de una placa 90x20 mm Cell Culture Dish® (SPL Life Sciences, Gyeonggi-do, Korea) con ayuda de dos Surgical Blades (No.24; Braun, Hessen, Alemania) y almacenadas a -80ºC en un nuevo criotubo etiquetado con código anonimizado. Los aspectos inusuales de las muestras fueron evaluados.

Implementación de un protocolo de *RNA-Sequencing* óptimo y análisis transcriptómico.

Las biopsias endometriales fueron disgregadas manualmente con cuchillas sobre una placa o empleando el Tissulyser II (Qiagen, Hilden, Alemania) y el ARN total fue extraído empleando el miRNeasy mini kit® (Qiagen, Hilden, Alemania) en tandas de 5-10 muestras bien en Fundación IVI o en el servicio de biobanco (IIS La Fe). La calidad del ARN fue evaluada empleando el NanoDrop ONE® (AF-00342; Thermo Fisher Scientific, Valencia, España) y la 4200 TapeStation System® (Agilent, Valencia, España) bien en fundación IVI o en el servicio de genómica (IIS La Fe). Las muestras que no cumplieron con los criterios de calidad requeridos para este estudio fueron excluidas, siendo estos: ratio 260/280 ~2; ratio 260/230 = 1,8-2,2; *RNA integrity number* (RIN) ≥ 3 y porcentaje de fragmentos de ARN superiores a 200 nucleótidos (DV200) $\geq 70\%$. El protocolo seleccionado para la generación de las librerías fue el *AmpliSeq for Illumina Transcriptome Human Gene Expression Panel*, incluyendo un ARN universal humano (Agilent, Valencia, España) como control positivo, agua libre de ARNasas (Qiagen,

Hilden, Alemania) como control negativo y dos muestras como réplicas técnicas en cada tanda de secuenciación. La secuenciación fue llevada a cabo con el sistema NextSeq500/550 con un diseño pareado de 150 ciclos y 10 millones (M) de lecturas por muestra según las recomendaciones de Illumina®. Primero se llevó a cabo un estudio piloto con 40 muestras en Fundación IVI (IIS La Fe) y una vez comprobada la validez del protocolo, el resto de las muestras fueron secuenciadas en Juno Genetics (Oxford) o en el servicio de genómica (IIS La Fe). Las muestras con demasiadas (> 12 M) o muy pocas (< 5 M) lecturas fueron secuenciadas de nuevo permaneciendo como duplicadas. Las variables técnicas de interés fueron consideradas en los análisis posteriores. Los principales parámetros de secuenciación fueron considerados para evaluar el proceso, que en este caso debían ser: densidad de clusters = 170-220 K/mm2; clusters que pasan el filtro (PF) ≥ 90%; rendimiento = 50-60 Gb; y > 80% de datos > Q30.

En cuanto al pre-procesamiento y el análisis exploratorio, los datos crudos procedentes del secuenciador fueron demultiplexados mediante bcl2fastq y evaluados usando FastQC. STAR fue empleado para el mapeo de las lecturas con respecto a un genoma de referencia integrado por los 20.802 amplicones incluidos en el panel AmpliSeq de transcriptoma completo empleado. Los conteos crudos fueron obtenidos empleando featureCounts, una herramienta que permite calcular parámetros como el *Phred quality score* (Q) y los conteos de baja calidad (Q < 30) fueron filtrados. Todos estos pasos se llevaron a cabo usando Snakemake (versión 7.3.4) de Python (versión 3.8). El comportamiento de las muestras se estudió mediante un ACP. Los controles, así como las réplicas y las muestras duplicadas que no cumplían con el rango óptimo de lecturas (5-12 M), las muestras aisladas (*outliers*) y las muestras con datos clínicos insuficientes para disponer de clasificación clínica, fueron eliminadas. Los genes con 0 conteos fueron descartados, así como los genes con baja expresión, filtrando por bajos conteos por millón (CPM < 1)

usando EdgeR. Los conteos restantes fueron normalizados usando Voom y cuantiles. Los posibles efectos tanda técnicos (tanda de extracción de ARN, concentración de ARN, ratio 260/280, ratio 260/230, RIN, DV200 y tanda de secuenciación), así como los efectos tanda demográficos (clínica de reclutamiento, edad, IMC, etnia, alergia, medicación, tabaco y alcohol) fueron evaluados mediante ACP y corregidos en caso necesario mediante modelos lineales con el paquete lima. A continuación, las muestras fueron clasificadas según un predictor transcriptómico basado en 73 genes de tiempo previamente desarrollado por nuestro grupo de investigación, para así corregir el efecto de la progresión endometrial y centrar el estudio en la patología. Todos estos pasos fueron llevados a cabo en R y los gráficos fueron generados con el paquete ggplot2.

Desarrollo de un modelo de predicción, estratificación de las pacientes y seguimiento clínico.

Las muestras normalizadas y corregidas fueron divididas en conjuntos de entrenamiento (80%) y de prueba (20%) manteniendo una proporción de 1 paciente FIR por cada 3 controles (1:3). Ambas poblaciones fueron comparadas considerando las principales variables de confusión y estudiando su comportamiento transcriptómico mediante ACP. El conjunto de entrenamiento fue empleado para la selección de la potencial firma biomarcadora de la WOI patológica y el desarrollo y validación interna de un modelo de predicción, mientras que el conjunto de prueba fue empleado para la validación externa del modelo y su optimización. Concretamente, el algoritmo CorrelationAttributeEval fue empleado para ordenar los genes de forma decreciente en función de su poder predictivo para FIR en el conjunto de entrenamiento. A continuación, los algoritmos SVM, kNN y RF (parámetros por defecto; validación cruzada 5-fold 80:20; 100 veces) fueron

implementados para estudiar la capacidad predictiva de diferentes conjuntos de los genes ordenados incrementando el tamaño de 1 en 1 (0-300 primeros genes), 10 en 10 (301-500 primeros genes), 100 en 100 (501-1000 primeros genes), 200 en 200 (1001-2000 primeros genes) y 500 en 500 (hasta la totalidad de genes). Para cada algoritmo se seleccionó la firma de genes con mayor precisión y entre ellas, se escogió la firma con mayor número de genes para identificar la WOI patológica en este estudio. Se llevó a cabo un método balanceado probabilístico usando la firma seleccionada y testando todas las combinaciones singulares y por parejas entre los algoritmos SVM, kNN y RF. Para ello, el conjunto de entrenamiento fue dividido 100 veces (iteraciones) considerando las mismas muestras de la condición minoritaria (FIR) y seleccionando aleatoriamente el número de muestras equivalente de la condición mayoritaria (control). Las muestras del conjunto de prueba fueron testadas en cada iteración ofreciendo como resultado una probabilidad media de patología por muestra. Las muestras fueron clasificadas según esta probabilidad como FIR ($\geq$ 0,5) o controles ($<$ 0,5). Parámetros predictivos como la precisión, sensibilidad y especificidad fueron calculados para el conjunto de prueba de forma independiente para la validación externa de los modelos de predicción. Se seleccionó como modelo óptimo para identificar la WOI patológica aquel con mejores parámetros. Para la validación interna del modelo seleccionado, se empleó un proceso de validación cruzada usando el conjunto de entrenamiento (parámetros por defecto; validación cruzada 5-fold 80:20; 10 veces) y se calcularon los parámetros predictivos evitando sobreentrenamiento. Todos estos pasos fueron llevados a cabo en R (versión 4.0.5, 2021-03-31) y en Weka. Todos los gráficos fueron generados usando ggplot2.

Las pacientes de nuestra población fueron clasificadas como controles (c) o FIR debido a una WOI patológica (p) según el modelo balanceado óptimo. La estratificación se

estableció según la probabilidad de patología en los perfiles c1 (≤ 0.2), c2 (0.2-0.5), p2 [0.5-0.8) y p1 (≥ 0.8). Estos grupos fueron comparados considerando las principales variables de confusión. Para estudiar la relevancia clínica de estos perfiles se realizó un seguimiento y se compararon sus resultados reproductivos. Concretamente, se calcularon (considerando el resultado de la primera transferencia embrionaria tras la obtención de la biopsia) y compararon en R, las tasas de embarazo (TE), de embarazo evolutivo (TEE), de aborto bioquímico (TAB) y de aborto clínico (TAC). Adicionalmente, la TE acumulada fue calculada teniendo en cuenta todas las transferencias embrionarias.

Estudio molecular de los perfiles transcriptómicos y validación experimental.

Para llevar a cabo el estudio molecular se aplicó un análisis de expresión diferencial (DEA; del inglés *differential expression analysis*) y un análisis funcional (GSEA; del inglés *gene set enrichment analysis*) comparando los diferentes perfiles usando el paquete ClusterProfiler en R. Las dos bases de datos consultadas fueron *Kyoto Encyclopedia of Genes and Genomes* (KEGG) y *Gene Ontology* (GO) y sólo las funciones enriquecidas fueron seleccionadas (tasa de falsos positivos (FDR; del inglés *false discovery rate)* < 0,05). Para el enriquecimiento con KEGG (versión Sept-2021), las rutas anotadas fueron filtradas según el número de genes asociados (0-1000) y las relacionadas con enfermedades y fármacos fueron eliminadas. Para el enriquecimiento de GO (versión Dic-2021), se emplearon las tres ontologías (procesos biológicos, funciones moleculares y componentes celulares), sólo las asociaciones con evidencia experimental fueron incluidas y las anotaciones fueron filtradas según el número de genes asociados (5-300). La concordancia a nivel funcional entre los diferentes perfiles fue evaluada usando Cohen´s Kappa index.

Finalmente, la expresión de los seis genes diferencialmente expresados (DEGs; del inglés *differentially expressed genes*) más interesantes entre los perfiles fue evaluada con la PCR cuantitativa (qPCR) usando cebadores específicos (Invitrogen, Thermo Fisher Scientific MA, EE. UU) diseñados con Primer-BLAST, en veinte muestras de ARN. El ADNc fue sintetizado usando PrimeScript Reagent Kit (Perfect Real Time, Takara, Shiga, Japón) en un termociclador T3000 (Biometra, Dublín, Irlanda). La qPCR fue llevada a cabo en un StepOnePlus Real-Time PCR System (Applied Biosystems, CA, EE. UU) usando Power-Up SYBR Green (Thermo Fisher Scientific, MA, EE. UU) y el gen de la actina beta (*ACTB*) como *housekeeping*. Los cebadores fueron testados usando un ARN universal humano (Agilent, Valencia, España) como control positivo y agua libre de ARNasas (Qiagen, Hilden, Alemania) como control negativo, analizando cada muestra por duplicado. La expresión relativa fue calculada usando el método ΔΔCt y fue comparada entre los distintos perfiles de interés. Los potenciales efectos tanda relacionados con el procedimiento experimental y el tiempo endometrial fueron evaluados. Las tendencias de expresión obtenidas con qPCR y RNA-Seq fueron comparadas. Los pasos fueron realizados en R y los gráficos generados con ggplot2.

RESULTADOS Y DISCUSIÓN

Datos transcriptómicos y caracterización clínica de las pacientes.

De las 291 biopsias endometriales que se obtuvieron, 276 disponían de tejido suficiente y fueron extraídas. De estas, 41 fueron excluidas por no cumplir los criterios de calidad de ARN y otras 40 por no disponer de resultados reproductivos, quedando 195 muestras que fueron secuenciadas en seis tandas. No fue hallada ninguna relación entre la calidad del ARN y los aspectos inusuales anotados sobre la apariencia de las muestras. Las

librerías generadas mostraron la longitud óptima (202-267 bp), los parámetros de secuenciación indicaron que el proceso fue adecuado y los datos obtenidos, aunque un poco bajos en densidad de clusters, tenían calidad suficiente. Aunque hubo heterogeneidad en el número de lecturas, la mayoría se situaron en el rango óptimo establecido (5-12 M). Tras verificar su esperado comportamiento en el ACP, los controles fueron eliminados, así como las réplicas y las 11 muestras fueron filtradas según el número de lecturas óptimo. De las 195 muestras únicas restantes, 2 muestras fueron descartadas por tener un comportamiento aislado y 62 por tener datos clínicos incompletos o transferencias insuficientes para establecer la clasificación clínica. Finalmente, 131 muestras fueron incluidas en los análisis. Por otro lado, 656 de los 20.802 genes medidos con el panel de RNA-Seq obtuvieron 0 conteos y fueron eliminados; mientras que 5.472 genes fueron filtrados según los conteos por millón (CPM $\geq$ 1) permaneciendo 14.674 genes. Tras el proceso de normalización, el único efecto tanda que fue identificado en el ACP fue el correspondiente a la tanda de secuenciación, el cuál fue corregido. El efecto del tiempo endometrial también fue identificado y corregido adecuadamente.

Las 131 pacientes fueron clasificadas clínicamente como FIR (n = 32) y controles (n = 99). Con respecto a la caracterización clínica, la población fue homogénea y por tanto comparable (p-valor > 0,05) según la clínica, años de infertilidad y tipo, edad, IMC y clasificación del tiempo endometrial, con ligeras diferencias en el ciclo empleado para las transferencias. Sin embargo, el número de transferencias (p-valor = 2,20e-16) y de fallos de implantación (p-valor = 2,20e-16) mostraron diferencias significativas como era de esperar según los criterios empleados para la clasificación clínica. El ACP comparando las muestras controles y FIR mostró la misma homogeneidad.

Estos resultados muestran el apropiado diseño experimental de nuestro estudio para continuar con los objetivos de esta tesis. En comparación con el único modelo de predicción para el FIR con independencia del tiempo, desarrollado previamente por Koot y colaboradores, en este estudio hemos empleado la tecnología de Illumina® RNA-Seq en lugar de Agilent® microarrays, mejorando la reproducibilidad y la detección de la expresión génica, obteniendo mayor cantidad de datos de calidad (131 muestras y 14.674 genes, frente a las 115 muestras y 12.198 genes de Koot). Además, hemos aplicado una metodología más fiable para la corrección del tiempo endometrial para focalizar el estudio en la función patológica, clasificando a las pacientes según un predictor transcriptómico previamente diseñado por nuestro grupo de investigación, mientras que Koot y colaboradores emplearon una clasificación basada en los días de la LH medida en orina, un método menos fiable que no permite distinguir la variabilidad endometrial molecular. Finalmente, la clasificación clínica empleada ha sido basada en los datos clínicos de las pacientes antes y después de la obtención de la biopsia y no sólo en los previos como ocurre en el estudio de Koot, por lo que los resultados predictivos serán más completos y realistas.

Modelo de predicción, nueva taxonomía y relevancia clínica.

Una vez que la población de pacientes fue dividida en conjunto de entrenamiento (n = 105) y prueba (n = 26), se determinó que ambos grupos eran comparables (p-valor > 0,05) según todas las posibles variables de confusión evaluadas (clínica, años y tipo de infertilidad, IMC, clasificación del tiempo endometrial, número de transferencias y de fallos de implantación, así como el tipo de ciclo empleado en las transferencias embrionarias). La única variable clínica que mostró ligeras diferencias fue la edad (p-

valor = 0,04), sin embargo, esto fue debido probablemente al efecto del pequeño tamaño del conjunto de prueba en comparación con el de entrenamiento. El análisis transcriptómico mostró la misma homogeneidad comparando ambos grupos. En cuanto a la firma característica de la WOI patológica, un conjunto de 236 genes fue seleccionado como firma óptima con el algoritmo kNN (Precisión = 79,20%). Con respecto a la selección del modelo balanceado probabilístico óptimo, la combinación entre SVM y kNN ofreció los mejores parámetros predictivos tras la validación externa con el conjunto de prueba (Precisión = 77%; Sensibilidad = 67%; Especificidad = 80%). Por otro lado, en el proceso de validación cruzada de este modelo empleando el conjunto de entrenamiento, la distribución de los parámetros de predicción también fue apropiada, con valores ligeramente superiores al conjunto de prueba como cabía esperar: precisión media = 83% (mínimo = 78%; máximo = 86%), sensibilidad media = 76% (mínimo = 69%; máximo = 85%) y especificidad media = 85% (mínimo = 80%; máximo = 91%).

Considerando la clasificación dada por el modelo y la probabilidad de patología obtenida para cada muestra, las pacientes predichas como FIR fueron estratificadas en p1 [n = 24 (18,32%)] y p2 [n = 14 (10,69%)], mientras que las predichas como control, en c2 [n = 32 (24,43%)] y c1 [n = 61 (46,56%)]. En cuanto a la caracterización clínica, todos estos perfiles fueron comparables considerando todas las variables de confusión evaluadas mencionadas previamente, con las excepciones (p-valor < 0,05) del número de transferencias embrionarias y de fallos de implantación (comparaciones p1 vs. c1, p1 vs. c2, p1 vs. p2, p2 vs. c1 y p2 vs. c2) que es lógico considerando los criterios que hemos empleado para la clasificación de las pacientes. Finalmente, el tipo de ciclo empleado en las transferencias embrionarias (comparación p1 vs. c1), la edad (comparación p1 vs. c2) y la clínica (comparación c2 vs. c1) mostraron ligeras diferencias. Estos resultados

muestran que los perfiles transcriptómicos son comparables pues están poco influenciados por las variables clínicas.

Considerando la relevancia clínica, los perfiles patológicos tuvieron tasas de embarazo inferiores a los perfiles control (29-57% vs. 71-78% TE and 50-57% vs. 76-91% TEE, respectivamente) mientras que las tasas de aborto fueron superiores (12-43% vs. 0-8% TAB and 0-43% vs. 9-17% TAC) destacando el perfil p1 con la mayor TAB (43%) y el p2 con la mayor TAC (43%). Las diferencias significativas en TE fueron obtenidas comparando p1 con c1 (p-valor = 1,15e-03) y con c2 (p-valor = 3,62e-04) verificando que p1 es el perfil patológico relacionado con la menor TE. En el caso de la TEE, las diferencias significativas fueron halladas comparando p1 con c1 (p-valor = 0,05) y p2 con c1 (p-valor = 0,01), lo que señala a c1 como el perfil control asociado a la mayor TEE. Con respecto a TAB, diferencias significativas fueron obtenidas comparando p1 y c1 (p-valor = 1,79e-03) lo que indica que p1 es un perfil patológico más asociado a aborto bioquímico, mientras que, en el caso de la TAC, comparando p2 y c1 (p-valor = 0,05), lo que indica que p2 es un perfil patológico más relacionado con aborto clínico. En cuanto a la TE acumulada, fue menor en los perfiles patológicos, mostrando una tendencia creciente de p1 a c1 (p1 = 38%; p2 = 76%; c2 = 81%; c1 = 93%).

Retomando el estudio de Koot y colaboradores, priorizaron una firma de 303 genes característicos de la patología (con sólo 6 coincidencias con nuestros 236 genes) empleando una metodología de selección basada en la ratio de ruido, muy empleada con datos de microarrays. En este estudio, basado en RNA-Seq, hemos empleando una metodología más adecuada en innovadora acorde a nuestros datos transcriptómicos basada en un algoritmo de selección de genes y en la evaluación de su poder predictivo. Además, proponemos el primer modelo balanceado para el FIR debido a una WOI patológica evitando sesgos de los resultados, a diferencia de Koot y colaboradores (que

emplea una población desbalanceada con más pacientes control que FIR), ofreciendo buenos parámetros predictivos, mejorando incluso la sensibilidad de 58,3% a 67%. Sin embargo, la precisión y la especificidad son moderadas, probablemente debido a la naturaleza binaria del modelo de predicción FIR vs. control, que no es capaz de captar completamente la heterogeneidad transcriptómica de la función endometrial patológica. Por tanto, aunque los resultados sean prometedores, se debe seguir investigando en esta línea para desarrollar herramientas innovadoras de translación clínica.

Finalmente, este es el primer estudio que estratifica la función endometrial con independencia de las variaciones del ciclo menstrual según la probabilidad de patología dada por un modelo probabilístico balanceado. Esta metodología nos ha mostrado que las pacientes subfértiles en HRT tienen un gradiente de al menos cuatro perfiles endometriales transcriptómicos con diferente significado clínico y pronóstico. Este hecho permite dividir la heterogeneidad molecular de la patología endometrial ofreciendo las claves de un buen diseño basado en este gradiente y no en una clasificación dicotómica. Además, nuestros resultados reproducen previos hallazgos en ciclo natural (pero en este caso con seguimiento clínico y en HRT), de los más empleados protocolos para preparación endometrial en transferencias embrionarias que permite controlar los niveles hormonales sin afectar a la funcionalidad del endometrio. Este hallazgo nos permite verificar que este tipo de ciclo no trata la patología endometrial pues sigue siendo detectable en nuestro estudio.

Por tanto, la nueva taxonomía propuesta en este trabajo podría ayudar a mejorar las aproximaciones diagnósticas y terapéuticas de las pacientes con FIR debido a una WOI patológica en dirección a una medicina de precisión.

Caracterización molecular de los grupos estratificados y validación experimental.

En cuanto a la caracterización molecular, los grupos más diferentes fueron p2 y c1 con 47 DEGs, seguidos de p1 y c1 con 3 DEGs. Adicionalmente, 14 DEGs fueron hallados en la comparación de p2 con ambos perfiles control (c2 y c1 conjuntamente) coincidiendo 13 de ellos con los 47 DEGs entre p2 y c1.

El mayor número de diferencias funcionales fue también hallado comparando p2 y c1 (54) y p1 con c1 (38). La mayoría de estas funciones en p2 vs. c1 fueron reguladas a la baja en p2 y mayormente relacionadas con la respuesta inmune, así como con el metabolismo y la producción de energía. Sin embargo, las funciones halladas en p1 vs. c1 fueron mayormente reguladas al alza en p1 y relacionadas con la respuesta inmune. Además, los dos perfiles patológicos (p1 y p2) fueron similares con ningún DEG y solo un DEG fue encontrado entre los perfiles control (c2 vs. c1) que fue común entre los hallados en p1 vs. c1. La comparación entre cada perfil patológico con c2 no mostró ningún DEG, señalando a c2 como perfil más cercano a la patología. Funcionalmente, p1 estuvo más asociado a funciones reguladas al alza relacionadas con la respuesta inmune, así como a proliferación y diferenciación en comparación con p2. La misma tendencia se observó comparando p1 y c2. En cuanto a p2 vs. c2, también con más funciones reguladas al alza, además de relacionadas con la respuesta inmune, también se hallaron funciones relacionadas con la expresión génica y la degradación de proteínas. En cuanto a los perfiles control, c2 estuvo más asociado a funciones reguladas al laza relacionadas con el sistema nervioso y la percepción sensorial, así como la respuesta hormonal.

Estos resultados muestran el sentido molecular y fisiológico de la nueva taxonomía definida. Los perfiles patológicos tienen mal pronóstico y están más relacionados con una respuesta inmune alterada. Concretamente, p1 con la menor TE y la mayor TAB, está más

asociado a una respuesta inmune excesiva contra el embrión en estadios tempranos como ha sido previamente descrito en otros trabajos sobre fallo de implantación. Sin embargo, p2, con la mayor TAC, es un perfil más inmuno tolerante y relacionado con aborto en estadios tardíos de embarazo probablemente debido a la falta de respuesta metabólica, siendo la primera vez que se relaciona aborto y deficiencia metabólica en el endometrio. En cuanto a los perfiles control, de buen pronóstico, c1 mostró la mayor TEE por encima de c2, lo que muestra el carácter molecular de c2 más cercano hacia la patología, siendo más inmuno tolerante, pero con una percepción sensorial excesiva e influencia de la respuesta hormonal. Con respecto al Cohen´s kappa index, no se halló apenas concordancia funcional entre las funciones enriquecidas obtenidas en las diferentes comparaciones entre los grupos de pronóstico, mostrando la heterogeneidad molecular que hay tras la patología endometrial.

En cuanto a la validación experimental de los resultados moleculares, los 3 DEGs entre p1 vs. c1 (coincidiendo 1 con el DEG hallado entre c2 y c1), así como los 3 genes con mayor *fold change* (FC) de los 13 DEGs que coincidieron entre p2 vs. los perfiles control y p2 vs. c1, fueron evaluados mediante qPCR. Todos los genes validados mostraron la misma tendencia de expresión que en RNA-Seq excepto *solute carrier familiy 17 member 8* (*SLC17A8*). En la comparación entre p1 vs. c1, *DNA microRNA-mediated repression inhibitor 1 (DND1) y LOC644172*, sinónimo de *mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2* (*MAPK8IP1P2*); en p2 vs. c1, *CF transmembrane conductance regulator* (*CFTR*) y *V-set domain containing T cell activation inhibitor 1* (*VTCN1*); y en c2 vs. c1, *LOC644172*. La expresión de actina beta *(ACTB)* como gene *housekeeping,* mostró un |FC| = 1 propio de un gen constitutivo. Por tanto, *DND1, SYT10, LOC644172, CFTR* y *VTCN1* fueron consistentes entre las distintas tecnologías reforzando su papel como potenciales biomarcadores de la patología endometrial.

En comparación con el estudio de Koot y colaboradores, este proyecto se focaliza en la caracterización molecular de los grupos de pronóstico de la nueva taxonomía y no en la clasificación clínica inicial, ofreciendo resultados más completos y rigurosos según la información clínica y transcriptómica. Además, en este estudio, los resultados moleculares de interés han sido evaluados usando también qPCR. Si bien es cierto que ya la tecnología de RNA-Seq es suficientemente precisa, aplicar otra tecnología para medir la expresión génica permite reforzar los resultados. Las inconsistencias halladas entre las distintas tecnologías podrían deberse a un FC < 2 como ocurre con el *SLC17A8*. Concretamente, la relación de estos genes con la función endometrial patológica ha sido descrita por primera vez en este estudio, con la excepción del gen *CFTR*. Todas las diferencias moleculares y funcionales encontradas entre los diferentes perfiles podrían ser clave para el descubrimiento de nuevos marcadores y/o nuevas dianas terapéuticas en investigaciones futuras, favoreciendo al diagnóstico y tratamiento de las pacientes infértiles.

Limitaciones y perspectiva futura.

El haber realizado un modelo de predicción dicotómico (patológico/control) hace que los resultados de predicción no sean óptimos, ya que se trata de perfiles heterogéneos desde el punto de vista molecular. Así mismo, necesitaríamos tamaños muestrales mayores para definir mejor los grupos y entrenar un modelo con más perfiles. Además, este estudio está focalizado en pacientes que se someten a ciclo sustituido y el tiempo fue limitado para profundizar en el estudio molecular mediante farmacología de sistemas.

Por tanto, el desarrollo de un modelo de predicción basado en los cuatro perfiles transcriptómicos (rompiendo con el modelo dicotómico y clasificando de manera más

precisa), el aumento del tamaño muestral de la población de estudio, el testeo de otros ciclos de preparación endometrial, así como el desarrollo de un modelo de farmacología de sistemas basado en los hallazgos moleculares de este estudio, podrían ser los siguientes pasos de esta tesis para la futura implementación de medidas diagnósticas y terapéuticas estandarizadas para pacientes infértiles.

CONCLUSIONES

1. La ventana de implantación patológica con independencia de la progresión endometrial tiene un perfil transcriptómico heterogéneo en las pacientes sometidas a ciclo sustituido (como fue previamente reportado en pacientes en ciclo natural) que puede ser identificado moderadamente con un modelo de clasificación binario basado en una firma biomarcadora de 236 genes. Por tanto, el ciclo sustituido no es útil para tratar la función endometrial patológica, ya que después de este tratamiento se sigue detectando la patología en nuestra población de estudio.

2. Las mujeres sujetas a una función endometrial patológica pueden ser estratificadas en cuatro grupos transcriptómicos según la probabilidad de patología dada por nuestro modelo de predicción. De hecho, los perfiles siguen un gradiente de prognosis, habiendo dos controles con buen pronóstico y dos patológicos con mal pronóstico.

3. Los perfiles control están asociados con moderados cambios en la función endometrial e implicaciones clínicas, mientras que los patológicos están

relacionados con alteraciones evidentes en la función endometrial y capacidad reproductiva comprometida. Concretamente, las pacientes estratificadas del grupo control con mejor pronóstico, tienen la mayor tasa de embarazo evolutivo, mientras que las del otro grupo control, más cercano a la patología, presentan una excesiva percepción sensorial y respuesta hormonal. En contraste, las pacientes estratificadas en el grupo patológico con el peor pronóstico tienen mayores tasas de aborto bioquímico y bajas tasas de embarazo debido a la leve tolerancia inmunológica materno-fetal, mientras que el otro grupo patológico presenta mayores tasas de aborto clínico siendo un perfil más inmuno tolerante asociado al déficit de nutrientes y energía en estadios tardíos del embarazo.

4. *DND1, SYT10, LOC644172, CFTR* y *VTCN1*, los genes más diferencialmente expresados entre los perfiles transcriptómicos mostraron tendencias similares en los análisis de RNA-Seq y qPCR reforzando su evidencia como potenciales biomarcadores del fallo de implantación recurrente endometrial.

5. Este estudio sienta las bases para el diseño de herramientas de última generación basadas en esta nueva estratificación transcriptómica de pacientes con infertilidad de origen endometrial, para proporcionar pronósticos precisos y facilitar el descubrimiento de procedimientos diagnósticos y terapéuticos alternativos y más personalizados en el contexto de una medicina reproductiva de precisión.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL TABLES

ABBREVIATIONS

A	
Acc	accuracy
ACTB	beta-actin/ actina beta
ACYP1	acylphosphatase
AI	artificial intelligence
AKT	serine/threonine protein kinase
APS	antiphospholipid syndrome
ART	assisted reproduction treatment(s)/ tratamiento de reproducción asistida (TRA)
Ave.Expr	average expression
B	
B	billion
BCL	binary base call
β-hCG	(beta) human chorionic gonadotropin
BMI	body mass index/ índice de masa corporal (IMC)
BMR	biochemical miscarriage rate/ tasa de aborto bioquímico (TAB)
bp	base pair(s)
BP	biological processes
C	
c1	control group 1
c2	control group 2
C1	pre-receptive
C2	receptive 1
C3	receptive 2
C4	post-secretory
ºC	degree(s) Celsius
CC	cellular components
CD	combinatorial dual
cDNA	complementary DNA/ ADN complementario (ADNc)
CFTR	CF transmembrane conductance regulator
cm	centimeter
CMR	clinical miscarriage rate/ tasa de aborto clínico (TAC)
COX2	cyclooxygenase 2
CPM	counts per million
CS	correlation score
CV	cross-validation
D	
DC	dendritic cell(s)
DEA	differential expression analysis
DEG	differentially expressed gene(s)
DNA	deoxyribonucleic acid/ ácido desoxirribonucleico (ADN)
DND1	microRNA-mediated repression inhibitor 1
dNK	decidual natural killer
DV200	RNA fragments with more than 200 nucleotides
E	
E2	17β-estradiol
ECM	extracellular matrix
E-cadherin	endothelial cadherin
EEC	endometrial epithelial cell(s)
EGF	epidermal growth factor
EGFL-7	EGF-like domain 7
EPC	epithelial progenitor cell(s)

ERA	endometrial receptivity analysis
ERK1/2	extracellular signal-regulated protein kinases 1/2
ER Map	Endometrial Receptivity Map
ESC	endometrial stromal cell(s)
F	
FC	fold change
FDA	Food and Drugs Administration
FDR	false discovery rate
FEA	functional enrichment analysis
FFPE	formalin-fixed paraffin-embedded
FOXO1	forkhead box protein O1
FSH	follicle-stimulating hormone
G	
g	gram(s)
Gb	gigabase(s)
GnRH	gonadotropin-releasing hormone
GO	Gene Ontology
GSEA	gene set enrichment analysis
H	
h	hour(s)
HB-EGF	heparin binding-EGF
HK	housekeeping gene
HRT	hormone replacement therapy
I	
i	index

ICM	inner cell mass
ICSI	intracytoplasmic sperm injection
ID	identifier
IIS	Instituto de Investigación Sanitaria/ medical research institute
IL	interleukin
IMSI	intracytoplasmic morphologically-selected sperm injection
IU	international units
IVF	*in vitro* fertilization/ fecundación *in vitro* (FIV)
IVI	Instituto Valenciano de Infertilidad
J	
JAK	Janus kinases
K	
K	kilocluster(s)
kb	kilobase(s)
KEGG	kyoto encyclopedia of genes and genomes
kg	kilogram(s)
kNN	k-nearest neighbors
L	
LH	luteinizing hormone
LIF	leukemia-inhibitory factor
LOC644172	synonym of mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2 (*MAPK8IP1P2*)
M	
M	million(s)
m (2)	(square) meter(s)
MAPK	mitogen-activated protein kinase

MAPK8IP1P2	mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2
Max	maximum
MF	molecular functions
mg	milligram(s)
Min	minimum
miRNA	microRNA(s)
mL	milliliter(s)
ML	machine learning
mm (2)	(square) millimeter(s)
MMP	matrix metalloproteinase(s)
mRNA	messenger RNA
MSC	mesenchymal stem cell(s)
MUC-1	mucin 1
MV(P)	micronized vaginal progesterone
N	
n	number
NA	not available
NC	negative control
ng	nanogram(s)
NGS	next-generation sequencing
NLRP1	NLR family pyrin domain containing 1
NLRP5	NLR family pyrin domain containing 5
nM	nanomolar
No.	number of
O	
ONT	Oxford Nanopore Technologies
OPN	osteopontin
OPR	ongoing pregnancy rate/ tasa de embarazo evolutivo (TEE)
ORA	over-representation analysis
P	
p1	pathological group 1
p2	pathological group 2
P4	pregn-4-ene-3,20-dione
P5	sequence 5 that binds to flow cell
P7	sequence 7 that binds to flow cell
PacBio	Pacific Biosciences
PAI-1	plasminogen activator inhibitor 1
PBS	phosphate buffered saline
PC	positive control
PC1	first principal component
PC2	second principal component
PCA	principal component analysis/ análisis de componentes principales (ACP)
PF	passing filter
pg	picogram(s)
PGD	preimplantation genetic diagnosis
PGE2	prostaglandin E2
PGT	preimplantation genetic testing
pM	picomolar
PR	pregnancy rate/tasa de embarazo (TE)
preRNA	precursor RNA
Prob	prediction-probability of pathology
Q	
Q (30)	Phred quality score
QC	quality control
(q)PCR	(quantitative) polymerase chain reaction

R	
Rd1 SP	sequencing primer binding site read 1
Rd2 SP	sequencing primer binding site read 2
RF	random forest
RIF	recurrent implantation failure/ fallo de implantación recurrente (FIR)
RIN	RNA integrity number
RNA	ribonucleic acid/ ácido ribonucleico (ARN)
RNA-Seq	RNA-Sequencing
RPL	recurrent pregnancy loss
RT	reverse transcription
S	
S	sensitivity
SBS	sequencing by synthesis
siRNA	small interfering RNA(s)
SLC17A8	solute carrier familiy 17 member 8
SNR	signal-to-noise ratio
Sp	specificity
SSC	somatic stem cell(s)
STAR	Spliced Transcripts Alignment to Reference
STAT	signal transducer and activator of transcription protein
SVM	support vector machine
SYT10	synaptotagmin

T	
TE	Tris-EDTA
TED	transcriptomic endometrial dating
TGF-β	transforming growth factor beta
TIMP	tissue inhibitor(s) of MMPs
Tr	training set
TRAF3IP1	TRAF3 interacting protein 1
Ts	test set
U	
ul	microliter
uNK	uterine natural killer
uPA	urokinase-type plasminogen activator
V	
VTCN1	V-set domain containing T cell activation inhibitor 1
VWC2	Von Willebrand factor C domain containing 2
W	
WIN	Window Implantation
WOI	window of implantation
Z	
ZNF738	zinc finger protein 738

I. INTRODUCTION

"Nothing in life is to be feared, it is only to be understood."

Marie Curie

INTRODUCTION

1. The human endometrium

1.1. Anatomy of the uterus and composition of the endometrium

The **female reproductive system** consists of the uterus, ovaries, fallopian tubes, cervix, vagina and external genitalia (perineum and vulva) in most mammals. The **uterus** is the largest multifunctional organ in the female reproductive system, which is located internally, within the pelvic area, between the bladder and the rectum. Its functions include sperm transport and storage, endocrine communication with the ovary, embryo implantation, pregnancy recognition, placenta formation, gestation, and eventual expulsion of offspring at birth (Hayssen & Orr, 2017).

In humans, the uterus has a simplex form, with an inverted pear silhouette, and a triangular shaped cavity. It is on average 7.5 cm long, 5 cm wide, 1.7 cm thick, and weighs approximately 60 g in a non-pregnant state. This organ can be divided into three zones: the uterine **fundus** (or upper domed-shaped region), the **body** of the uterus (that connects to the ovaries through the fallopian tubes), and the uterine **isthmus** (the narrow region between the body and the cervix). The broad, uterosacral, and round ligaments support the uterus within the pelvic cavity. Blood supply to the uterus is principally sustained by the uterine artery, which branches off directly from the internal iliac artery, while venous drainage occurs through the tributaries that lead to the internal iliac vein (Jones & Lopez, 2013). The **uterine wall** consists of three principal layers: the **perimetrium**, which is the thin outermost layer covering the external surface; the **myometrium**, which is the thick middle layer of smooth muscle; and finally, the **endometrium**, the innermost, mucus-

producing and most complex layer of the uterus (Simon et al., 2009), that this PhD dissertation will focus on.

Notably, the **endometrium** can be subdivided into four compartments: epithelial, stromal, vasculature and immune (Bergmann et al., 2021; Simon et al., 2009). The **endometrial epithelia** includes luminal epithelium and endometrial glands, which are both composed of columnar endometrial epithelial cells (EECs) that can have apical motile cilia (also known as microvilli). A reservoir of endogenous endometrial somatic stem cells (SSCs), called epithelial progenitor cells (EPCs), are found in the base of the glands. The luminal epithelium establishes the boundary to the uterine cavity, while uterine glands consist of a connected glandular epithelium that extends towards the myometrial border. The **endometrial stroma** is a connective tissue consisting of extracellular matrix (ECM) and endometrial stromal cells (ESCs), that are mainly fibroblasts. Mesenchymal stem cells (MSCs) are also found in this compartment as a reservoir of SSCs. On the other hand, the **endometrial vasculature** is mainly a capillary bed arising from the myometrium. Specifically, the uterine radial arteries branch into basal arteries, which in turn, branch into spiral arteries. Notably, the endothelial cells that line the microvessels and are adjacent to MSCs, are another reservoir of SSCs. Finally, the **immune component of the endometrium**, which accounts for an estimated 10-15% of the stromal compartment, protects against pathogens and balances the commensal microbiome. It mainly consists of T-cells, B-cells, macrophages and uterine natural killer (uNK) cells. Further, the endometrium can also be divided into the functional and basal layers. The **functional layer** is the portion of the endometrial compartments that changes dynamically throughout each menstrual cycle and is shed during menstruation, while the **basal layer** remains in place to stimulate the regeneration of a new functional layer ***(Figure 1)***.

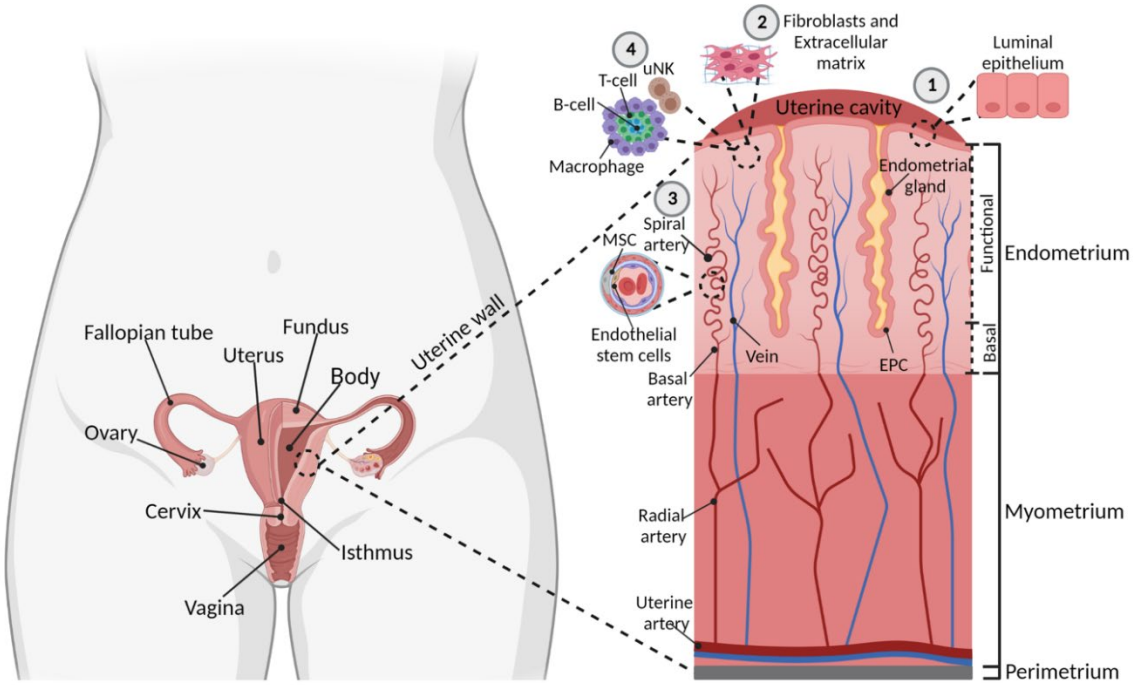


Figure 1. Anatomy of the uterus and composition of the endometrium.
The main organs of human female reproductive system are shown on the left side including the uterus, the largest one. The different elements of the uterine wall appear on the right side, focusing on the composition of the human endometrium: epithelial compartment (1); stromal compartment (2); vascular compartment (3) and immune compartment (4). EPC, epithelial progenitor cell; MSC, mesenchymal stem cell; uNK, uterine natural killer cells. Created with BioRender.com (2021).

1.2. The menstrual cycle and the window of implantation (WOI)

The **endometrium** is a highly dynamic multicellular tissue, that undergoes cyclical changes in function and appearance throughout the menstrual cycle. Similar to humans, apes, old world monkeys and some new world monkeys also undergo a menstrual cycle characterized by external bleeding due to the shedding of the endometrium (Catalini & Fedder, 2020). Most other mammalian species have estrous cycles, where the uterus is remodelled throughout the cycle but the endometrial lining is not shed (Billhaq et al., 2020).

In humans, the age at menarche (the beginning of the menstrual cycle) is on average 12 years old, while the age at natural menopause (marked by the depletion of the ovarian reserves, and thereby, cessation of menstruation) is on average 51 years old. The **human menstrual cycle** lasts on average 28 days (although it often varies between 26-35 days, depending on the woman) and it is orchestrated by the cyclical levels of steroid hormones produced by the hypothalamic-pituitary-ovarian axis. Briefly, the pulsatile release of gonadotropin-releasing hormone (GnRH) in the hypothalamus regulates the frequency of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) release by the anterior pituitary. FSH and LH then bind to their specific ovarian receptors to regulate folliculogenesis and ultimately, the secretion of oestrogen (17β-estradiol; E2) and progesterone (pregn-4-ene-3,20-dione; P4), respectively. In turn, E2 and P4 bind their cognate endometrial receptors to regulate the menstrual cycle (Canelon & Boland, 2020; Critchley et al., 2020; Mihm et al., 2011). Overall, the human menstrual cycle involves changes at the endometrial and ovarian level (Bergmann et al., 2021; Simon et al., 2009) ***(Figure 2).*** At the endometrial level, three main phases can be distinguished: menstruation, proliferative and secretory. In parallel, there are two ovarian phases: follicular and luteal. Concretely, menstruation and the proliferative endometrial phases coincide with the ovarian follicular phase, while the secretory endometrial phase coincides with the luteal phase. Characteristics of these phases of the human menstrual cycle are detailed below:

- **Menstruation/follicular phase** (day 0-5): the degradation of the corpus luteum (a progesterone-secreting structure generated at the ovulation site of the dominant follicle) and decreasing hormone levels induce endometrial shedding and initiate another menstrual cycle if no embryo implants into the endometrium. This phase is

characterized by the shedding of the functional layer of the endometrium, which, as previously described, includes the luminal epithelium and spiral arteries.

- **Proliferative/follicular phase** (day 5-14): in response to an increase in FSH, the oestrogen produced by the developing follicle in the ovaries stimulates the growth and proliferation of the endometrial tissue, thickening the endometrial lining from approximately 2 to 12 mm. This phase is characterized by the regeneration of the luminal epithelium, mediated by the EPCs migrating from glands and differentiating into enlarged columnar cells; the predominance of fibroblast-like cells in the stroma; the presence of immune aggregates (i.e., B cells, T cells and macrophages); and the angiogenesis, where spiral arteries sprout from the basal arteries. The end of this phase is marked by ovulation, which occurs 24-36 h after the LH spike.
- **Secretory/luteal phase** (day 14-28): ovulation (defined by the release of a mature oocyte from the ovary) leads to the development of a corpus luteum, that increases progesterone levels, which in turn, induces endometrial tissue differentiation to prepare for potential embryo implantation. The secretory phase can be subdivided into early-, mid-, and late-secretory. During this phase, the endometrial glands increase secretions while coiling, and the luminal epithelium grows apical cellular protrusions (pinopodes) to facilitate implantation. Further, the stromal MSCs surrounding the spiral arteries differentiate into a secretory phenotype, while the fibroblasts convert to decidualized, secretory fibroblasts. The abundant spiral arteries increase blood supply and the presence of immune cells is maximal. In the case of implantation, these immune cells establish maternal tolerance to the embryo, while if no successful implantation occurs, the uNK cells and macrophages are recruited to remove the decidualized cells.

Proper functioning of the endometrium is key for women's reproductive potential and fertility. The **main functions of the endometrium** are to reach its receptive state, thereby

enabling embryo implantation, and maintain early pregnancy. The period of maximal **endometrial receptivity** in the menstrual cycle is known as **window of implantation (WOI)** (Harper, 1992; Wilcox et al., 1999). This short period normally occurs in the mid-secretory phase, between days 19-24 of the menstrual cycle (Dominguez et al., 2003; Navot et al., 1991), and represents a complex and multifactorial process, involving molecular, cellular and tissue changes, which may vary between women (Lessey & Young, 2019; Talbi et al., 2006). Considering both endometrial receptivity and the WOI, the secretory phase of the endometrium are further subdivided into pre-receptive, receptive and post-receptive (Diaz-Gimeno et al., 2011). **Characterization of the WOI is essential** for optimizing women's reproductive potential and determining possible causes of endometrial-factor infertility.

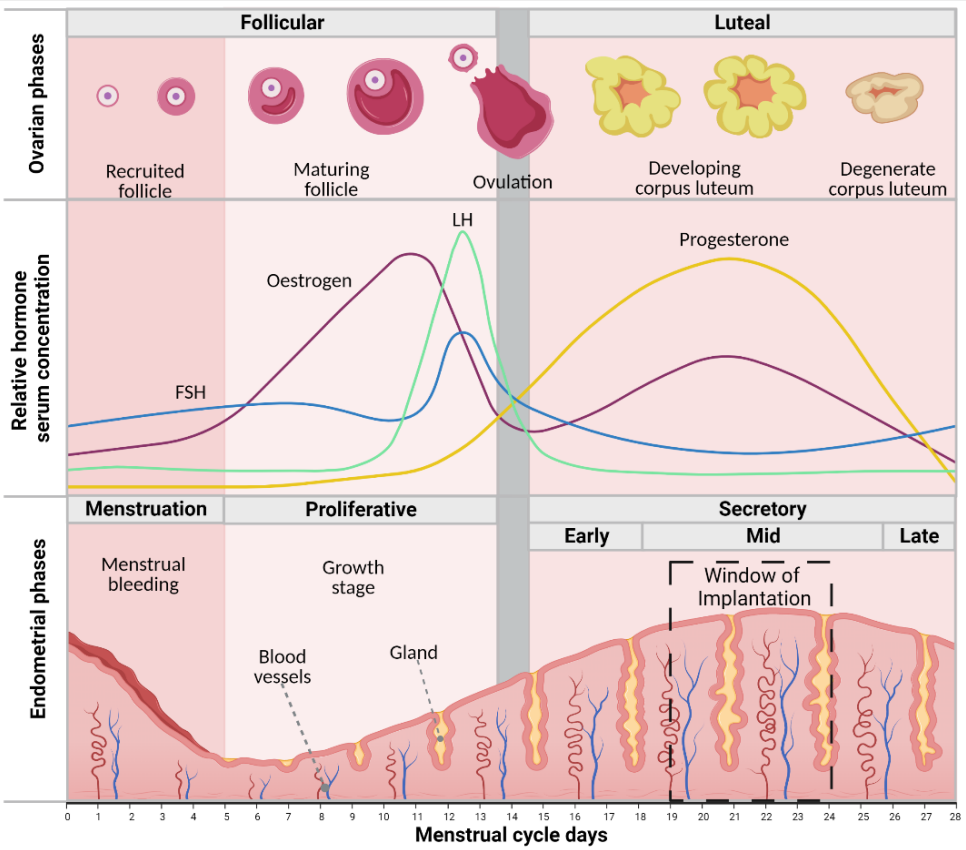


Figure 2. The human menstrual cycle.
Diagram showing the relative concentrations of serum follicle-stimulating hormone (FSH), luteinizing hormone (LH), oestrogen, and progesterone, corresponding with stages of follicle growth and changes in endometrial morphology. Created with BioRender.com (2021).

1.3. Molecular mechanisms regulating implantation and early fetal development

Successful **embryo implantation** requires a **synchronized feto-maternal crosstalk** (between a good-quality embryo and a receptive endometrium) **during the WOI**. The WOI is regulated by a wide variety of cytokines, growth factors, prostaglandins, enzymes, and adhesion molecules, amongst other molecules. Accordingly, the process of implantation involves a coordinated sequence of molecular events, which is attracting researchers interest to understand possible fertility problems (Ochoa-Bernal & Fazleabas, 2020).

Following successful fertilization in the infundibulum, the zygote (i.e., fertilized oocyte or one-cell embryo) undergoes a sequence of mitotic cell divisions as it migrates through the fallopian tube, towards the uterine cavity. By day 5 or 6, the embryo becomes a **blastocyst**, a complex multicellular structure of 32-256 blastocoels, with an inner cell mass (ICM) that develops into the embryo, and an outer layer of cells (termed trophoblasts) that develop into the placenta. Prior to implantation, the blastocyst moves freely within uterine cavity until it reaches the upper and posterior wall in the midsagittal plane, in the fundus of the uterus (Kim & Kim, 2017). As detailed below, embryo implantation involves a **tightly-regulated sequence of events**, mediated by several molecular mechanisms (Massimiani et al., 2020) ***(Figure 3)***:

1. **Apposition**. The first contact between a good-quality embryo and the receptive endometrium. In order to implant, blastocysts need to "hatch" out of their zona pellucida (i.e., protective glycoprotein coating), align their ICM towards the endometrium, and bind the leukemia-inhibitory factor (LIF) secreted by the receptive endometrium (through the LIF receptors expressed by the blastocoels). In response to

the microvilli on the trophoblasts interdigitating with the pinopodes of the endometrial epithelium, LIF mediates the shift from proliferative to differentiated luminal epithelium through the Janus kinases (JAK)-signal transducer and activator of transcription protein (STAT) signaling pathway. Notably, LIF also drives stromal proliferation, through regulation of the epidermal growth factor (EGF) pathway, where heparin binding-EGF (HB-EGF) is one of the most important EGF family members. Meanwhile, trophoblast cells begin to secrete molecules like human chorionic gonadotropin (hCG; a biochemical indicator of the establishment of early pregnancy), and interleukin 1 beta (IL-1β), which initiates the extracellular signal-regulated protein kinases 1/2 (Erk1/2) pathway in endometrial epithelial cells. The consequent expression of cyclooxygenase 2 (COX2) leads to the production of prostaglandin E2 (PGE2) in endometrial stromal cells, promoting decidualization and vascular permeability of the endometrium.

2. **Adhesion**. Trophoblasts attach to the endometrial epithelium mediated by L-selectin, integrin αvβ3 and endothelial cadherin (E-cadherin) adhesion molecules. Specifically, integrin αvβ3 is dually expressed in pinopodes and trophoblasts, and binds its glycoprotein ligand osteopontin (OPN), that is expressed in the endometrial epithelium. Alternatively, if the implanting embryo encounters a glycocalyx associated with the luminal epithelium, the glycoprotein mucin 1 (MUC-1) creates a protective barrier that prevents the blastocyst from binding to an area with poor chances of implantation. Other molecules associated with adhesion are CD98 (a receptor expressed in the surface of epithelial cells), and the NOTCH1 receptors and ligands (expressed in the trophoblast and the endometrial epithelium, respectively). Indeed, activation of the NOTCH1 signaling pathway by hCG up-regulates expression of forkhead box protein O1 (FOXO1), that is essential for the decidualization process.

3. **Invasion**. Trophoblasts cross throughout the epithelium and invade the decidualized endometrial stroma. Then trophoblasts differentiate into inner cytotrophoblasts and outer syncytiotrophoblasts. Trophoblast cells mediate the reconstruction of maternal spiral arteries in order to maintain a high blood flow between the fetus and the mother. In fact, invasion is associated with remodelling of the ECM, through tissue degradation mediated by multiple proteinases, including urokinase-type plasminogen activator (uPA; produced by trophoblasts) and matrix metalloproteinases (MMPs; expressed by endometrial cells at the feto-maternal interface). Specifically, MMP-2 and MMP-9 are produced in response to IL-1β, transforming growth factor beta (TGF-β), or hCG. Notably, trophoblast invasion is a tightly-regulated process, subject to negative feedback from tissue inhibitors of MMPs (TIMPs) and plasminogen activator inhibitor 1 (PAI-1; the main inhibitor of uPA) downstream of TGF-β signaling. Finally, EGF-like domain 7 (EGFL-7), also mediates the invasion of trophoblast cells, through activation of NOTCH, mitogen-activated protein kinase (MAPK), and serine/threonine protein kinase (AKT) signaling pathways.

The implantation process is followed by **early fetal development**, the period normally assigned to first trimester of pregnancy (12 weeks) (Farquharson et al., 2005). One of the main biological processes in the maintenance of early pregnancy is **feto-maternal immune tolerance** (Norwitz et al., 2001), a phenomenon that refers to several mechanisms in the decidua that protect the allogeneic embryo from the maternal immune system, while maintaining defences against possible pathogens (Yang et al., 2019). This feto-maternal interface consists of decidual stromal and immune cells that maintain bidirectional communication with the trophoblast cells to regulate the molecular functions that maintain a successful pregnancy. Decidual NK (dNK) cells predominate in the human decidua in the first trimester of pregnancy (representing approximately 90% of all cells),

followed by macrophages (15-20%), T cells (5-15%) and dendritic cells (1-2%) (Wang & Li, 2020). Elucidating the complexities of the molecular mechanisms underlying feto-maternal immune tolerance may help reveal etiologies of reproductive disorders.

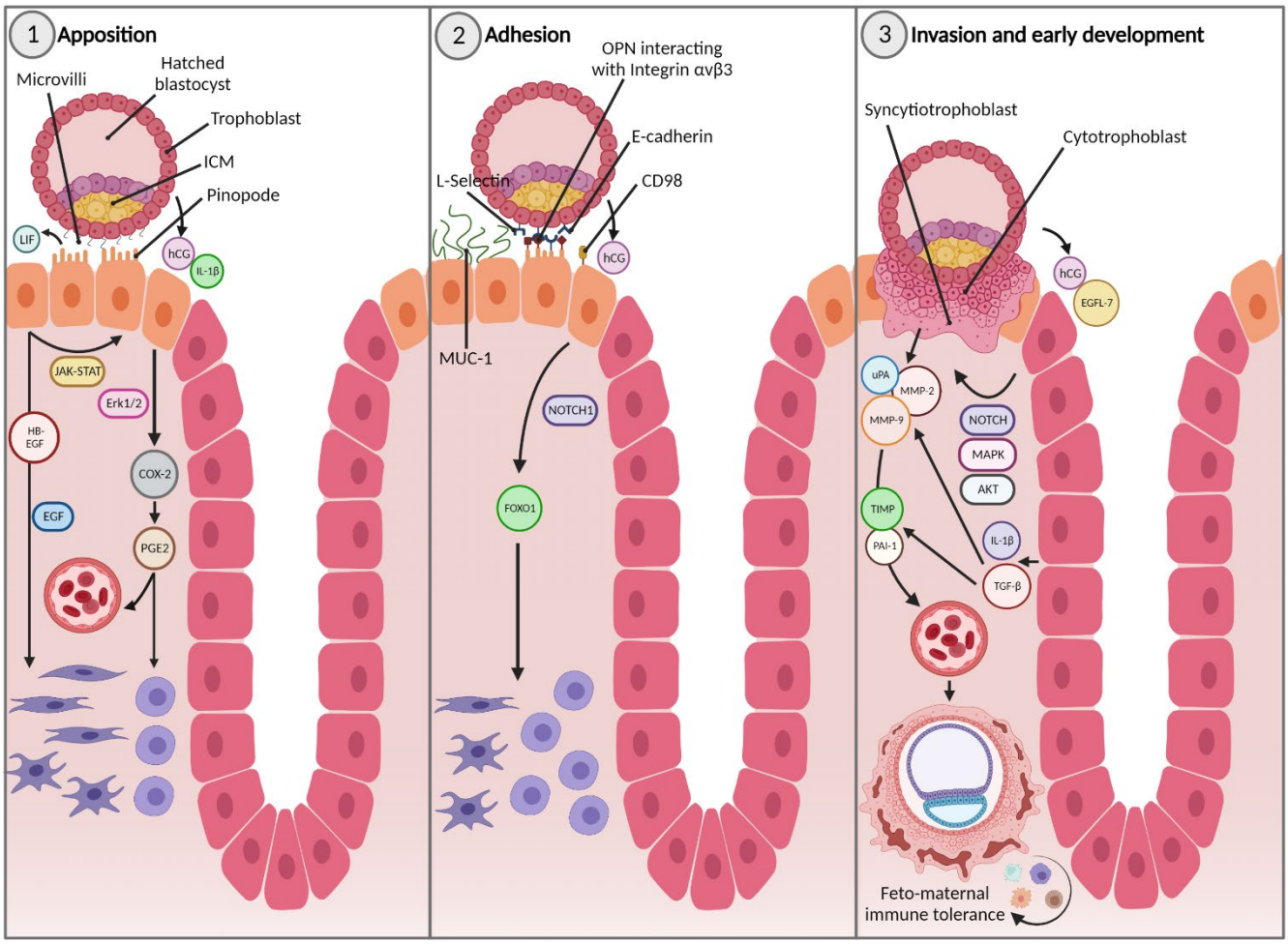


Figure 3. Molecular mechanisms in embryo implantation and early development

The main molecular mechanisms implicated in the different phases of human embryo implantation: apposition (1), adhesion (2), invasion and early development (3). AKT, serine/threonine protein kinase; COX-2, cyclooxygenase 2; E-cadherin, endothelial cadherin; EGFL-7, epidermal growth factor-like domain 7; ERK1/2, extracellular signal-regulated protein kinases; FOXO1, forkhead box protein O1; HB-EGF, heparin binding epidermal growth factor; hCG, human chorionic gonadotropin; ICM, inner cell mass; IL-1β, interleukin 1 beta; JAK-STAT, Janus kinases-signal transducer and activator of transcription protein; LIF, leukemia-inhibitory factor; MAPK, mitogen-activated protein kinase; MMP-2, matrix metalloproteinases 2; MMP-9, matrix metalloproteinases 9; MUC-1, mucin 1; OPN, osteopontin; PAI-1, plasminogen activator inhibitor 1; PGE2, prostaglandin E2; TGF-β, transforming growth factor β; TIMP, tissue inhibitor of MMPs; uPA, urokinase-type plasminogen activator. Created with BioRender.com (2022).

2. Recurrent implantation failure (RIF)

2.1. Endometrial-factor infertility: RIF

Infertility currently affects 8-12% of reproductive-aged couples worldwide (Vander Borght & Wyns, 2018), and is defined by the World Health Organization as "a disease of the reproductive system defined by failure to achieve clinical pregnancy after 12 months or more of regular unprotected sexual intercourse" (Gurunath et al., 2011). This fact has increased the importance and popularity of assisted reproduction treatments (ARTs). While this definition of infertility is based on a given period of time, sterility is a permanent state of infertility. On the other hand, **subfertility**, which in some cases may be used interchangeably with infertility, is defined as any type of reduced fertility in couples unsuccessfully trying to conceive. Further, infertility can be categorized as secondary or primary, if the person has respectively conceived a baby in the past or not (Vander Borght & Wyns, 2018).

Human reproduction is highly inefficient - only 30-40% of all spontaneous conceptions progress to clinical pregnancy (Macklon et al., 2002) due to a plethora of contributing male and/or female factors. This PhD dissertation will focus on **female-factor infertility**, which accounts for 20-35% of infertility cases (Massimiani et al., 2020), and may derive from a wide variety of causes, such as age, endocrine disorders, immunological alterations, genital tract infections, lifestyle-related factors and several uterine disorders (Bala et al., 2021; Devesa-Peiro et al., 2020; Khizroeva et al., 2019; Liu et al., 2011; Ravel et al., 2021; Unuane et al., 2011). Specifically, **uterine disorders** include well-established uterine pathologies, such as endometriosis, adenomyosis, uterine myomas, endometrial hyperplasia, endometrial cancer, endometrial polyps, or uterine malformations, whose effect on endometrial receptivity is usually poorly understood and

sometimes controversial (Alansari & Wardle, 2012; Da Broi et al., 2019; Gambadauro & Gudmundsson, 2017; Harada et al., 2016; Horne & Critchley, 2007; Lessey & Kim, 2017; Miravet-Valenciano et al., 2017; Munro, 2019; Turocy & Rackow, 2019; Zepiridis et al., 2016). On the other hand, it is widely accepted that **recurrent pregnancy loss (RPL)** and **recurrent implantation failure (RIF)** are disorders that contribute to uterine-factor infertility, with their negative repercussions on endometrial function and embryo implantation and development, however, the complexities of their underlying molecular mechanisms merit further investigation (Bashiri et al., 2018; Diaz-Gimeno et al., 2017; Ford & Schust, 2009; Jauniaux et al., 2006; Koot et al., 2016; Moustafa & Young, 2020; Sebastian-Leon et al., 2018).

Despite RPL and RIF having similar etiologies and treatments, they are considered distinct pathological conditions. **RPL** is a distressing pregnancy disorder experienced by ~2.5% of women trying to conceive, that is defined as the failure of two or more clinical pregnancies during the first 20–24 weeks of gestation, and includes embryonic and fetal losses (Dimitriadis et al., 2020). Similarly, a clinical miscarriage denotes the loss of an intrauterine pregnancy, previously confirmed by ultrasonography or histology, prior to 20–24 weeks of gestation. Of note, ectopic pregnancies and molar pregnancies are not considered in this definition of RPL. Further, the inclusion of non-consecutive losses (interspersed with a successful pregnancy) or early pregnancy losses also called biochemical miscarriage (before the first 10 weeks of gestation) is contentious (Dimitriadis et al., 2020; Kolte et al., 2015).

Attempting to improve the implantation rate in humans, from approximately 30% per natural cycle, has become a real challenge in ARTs (Kim & Kim, 2017). **Implantation failure** refers to an embryo that does not implant in the maternal endometrium. Implantation can fail during initial attachment or invasion stages, resulting in an

undetectable pregnancy [i.e., absence of urine or serum beta hCG (β-hCG)], or following successful invasion of the embryo through the luminal surface of the endometrium (when the β-hCG produced by the embryo may be detected in urine or blood), but the process becomes disrupted prior to the formation of an intrauterine gestational sac (which is clinically referred to as a biochemical miscarriage) (Bashiri et al., 2018; Coughlan et al., 2014). Further, implantation failure is diagnosed in both people who try to conceive spontaneously and with the help of ARTs, while RIF only applies to patients undergoing ARTs.

RIF is an imprecisely defined reproductive disorder, characterized by the absence of implantation (and subsequent pregnancy) after at least three transfers with good-quality embryos (Bashiri et al., 2018; Coughlan et al., 2014; Koot et al., 2016; Macklon, 2017). It is estimated that embryos account for a third of implantation failures, while suboptimal endometrial receptivity and/or altered communication at the feto-maternal interface are responsible for the remaining two thirds (66.7%) (Craciunas et al., 2019). In addition, the prevalence of RIF was recently estimated to be 10% (Cimadomo et al., 2021; Somigliana et al., 2018), however, another study found that 95.2% of patients with RIF achieved pregnancy after three consecutive quality embryo transfers, arguing that RIF prevalence was realistically only 5% in their cohort (Pirtea et al., 2021) ***(Figure 4)*****.** Taken together, the **discrepancies in clinically defining RIF and its prevalence impede accurate diagnoses and prevent patients undergoing ARTs from receiving appropriate treatments** that support their endometrial function, embryo implantation, and ultimately, the establishment of a successful pregnancy, becoming the **major object of study in this dissertation.**

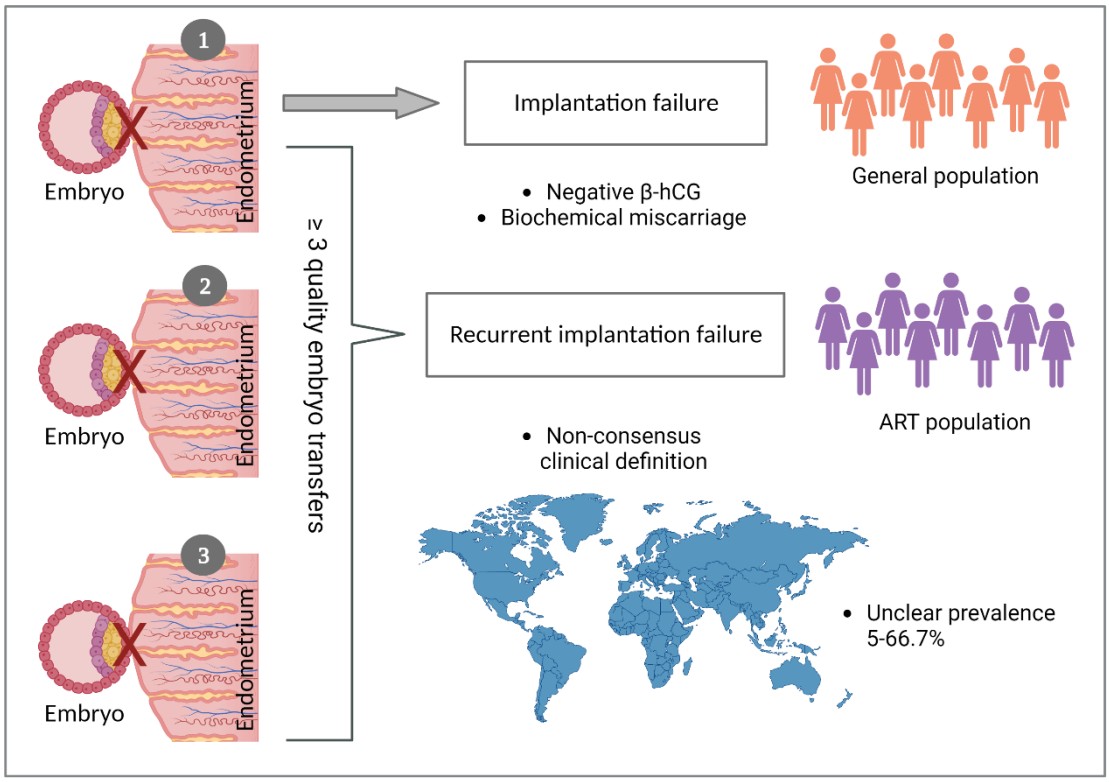


Figure 4. Implantation failure vs. recurrent implantation failure.
This diagram compares implantation failure and recurrent implantation failure. ART, assisted reproduction treatment; β-hCG, beta human chorionic gonadotropin. Created with BioRender.com (2022).

2.2. Etiologies of RIF

RIF is a complex pathology with multiple possible etiologies. Although most clinical definitions of RIF disregard embryonic causes (because most embryonic causes are easily identified and controlled in the clinical setting), RIF can be caused by either embryonic or maternal factors ***(Figure 5)*** (Bashiri et al., 2018; Coughlan et al., 2014; Simon & Laufer, 2012).

Embryonic factors that contribute to RIF include **gametic and/or genetic causes and ART conditions**. Embryonic quality reflects the quality of the progenitor oocyte and sperm. Compromised oocyte quality is often suspected as a cause of RIF when there was

a poor response to ovarian stimulation (Ferraretti et al., 2011) or aggressive ovarian stimulation (Verberg et al., 2009). Alternatively, sperm deoxyribonucleic acid (DNA) damage and abnormal spermatozoa morphology also reduce embryo quality and implantation rates (Künzle et al., 2003; Zini et al., 2008). Further, chromosomal aneuploidies (often associated with advanced maternal age) in embryos often lead to pregnancy loss and implantation failure (Nagaoka et al., 2012) as a natural selection mechanism. Finally, embryo manipulation and the transfer procedure itself can directly or indirectly affect embryo quality (Das & Holzer, 2012; Guerif et al., 2004).

The **maternal factors** contributing to RIF include **lifestyle factors** (e.g., smoking, alcohol intake or diet), **uterine alterations, ART conditions,** along with **endocrine disorders or other comorbidities**. Obese patients [with a body mass index (BMI) > 30 kg/m^2] have the highest chance of implantation failure when compared with patients of normal weight (BMI = 18.50-24.99 kg/m^2) (Moragianni et al., 2012; Orvieto et al., 2009). In addition, smoking and alcohol consumption have been associated with a significantly increased risk of implantation failure (Rossi et al., 2011; Waylen et al., 2009). In terms of uterine alterations, a thin endometrium (< 8 mm in the secretory phase) (Miwa et al., 2009; Vartanyan et al., 2020) and hydrosalpinxes (Kodaman et al., 2004; Palagiano et al., 2021) have decreased implantation rates, while the implications of other pathologies, that impede endometrial function but have not yet been well linked to RIF (e.g., severe uterine adhesions, large myomas, endometrial polyps, or uterine malformations), are still being explored (Devesa-Peiro et al., 2020). Implantation success can also be hindered by unsterile and/or traumatic transfer procedures (Schoolcraft et al., 2001). The kind of ART cycle has recently been postulated to affect endometrial function, like the excessive peak of progesterone in some stimulated cycles (Lawrenz & Fatemi, 2017), but requires further investigation. Endocrine disorders [e.g., untreated thyroid disease, uncontrolled diabetes,

or variations in the prolactin level (Timeva et al., 2014)] have been associated to RIF, as well as other comorbidities like hereditary thrombophilias (Azem et al., 2004), and antiphospholipid syndrome (APS) (Blank & Shoenfeld, 2010). In addition, chronic immune responses with excessive presence of NK (Santillan et al., 2015; Yamada et al., 2003) and T helper cells (Nakagawa et al., 2015), or chronic endometritis due to infection (Johnston-MacAnanny et al., 2010), could be also related to implantation failure, but further research is required.

Furthermore, when any of the factors above are attributed to implantation failure, the term of **RIF of endometrial origin or endometrial RIF** is employed in the literature. Several groups have recently argued that endometrial RIF may be caused by a displaced/**asynchronous WOI** and/or a **pathological/altered WOI** (Diaz-Gimeno et al., 2011; Koot et al., 2016; Sebastian-Leon et al., 2018) but more research is required. In this regard, endometrial progression to the WOI must be synchronized with early fetal development, and the endometrium must function adequately during the WOI, to achieve a suitable environment for successful embryo implantation and establishment of pregnancy.

Due to its multifaceted and complex etiology, RIF is currently one of the most difficult fertility conditions to manage clinically. **This dissertation will particularly focus on endometrial RIF that results from a pathological WOI (independent of displacement)**, to address the knowledge gap regarding the combination of molecular alterations that lead to this disorder and set the foundation for further clinical studies that aim to improve the diagnosis, treatment, and ultimately, reproductive capacity of affected patients.

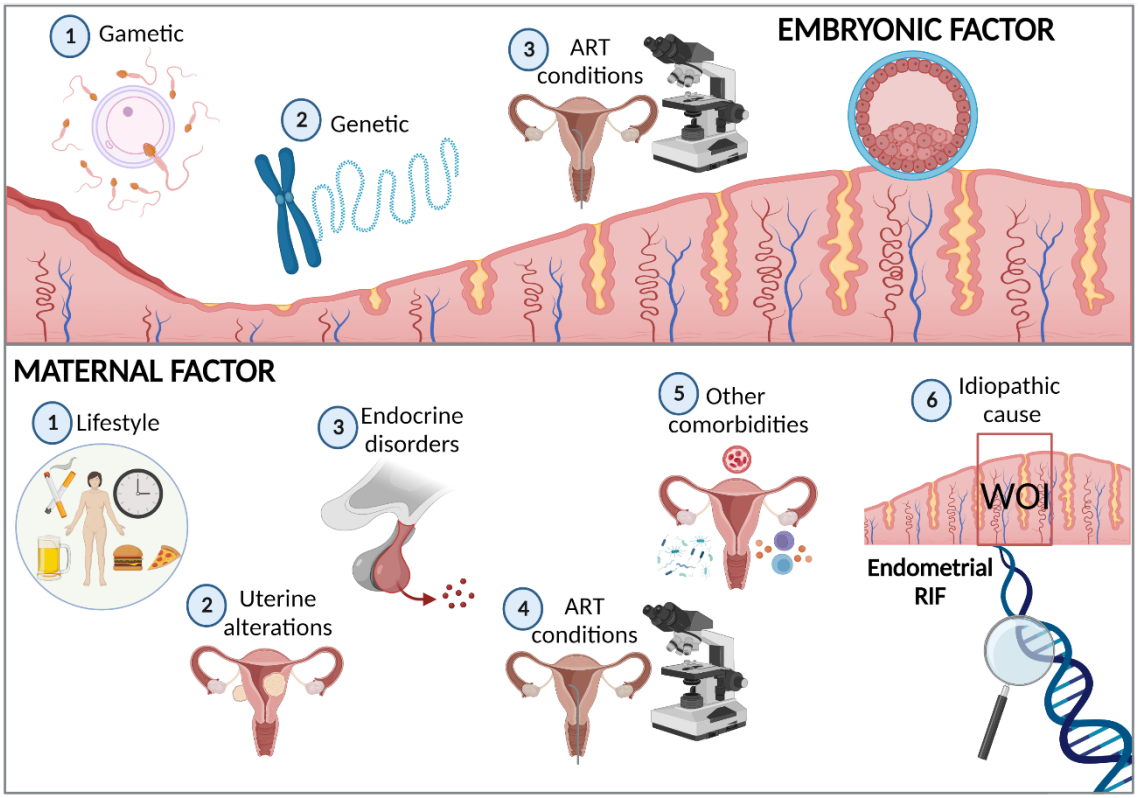


Figure 5. Embryonic and maternal factors contributing to recurrent implantation failure. *Multiple embryonic (above) and maternal (below) factors contribute to recurrent implantation failure (RIF). Concretely, endometrial RIF can be due to a displaced and/or a pathological window of implantation (WOI), the latter of which being the focus of this thesis. ART, assisted reproduction treatment. Created with BioRender.com (2022).*

2.3. RIF diagnosis and therapeutic strategies

Even after several decades of clinicians offering ARTs, the **diagnostic criteria for RIF remain controversial.** Reaching a consensus on how to define RIF would not only provide more accurate calculations of its prevalence, but also aid in establishing rigorous and standardized diagnostic and therapeutic strategies that consider all embryonic and maternal factors ***(Table 1).***

Table 1. Summary of the main evaluation and solution measures in the clinical management of recurrent implantation failure.

Origin	Factor	Evaluation	Potential solution
Embryonic	Gametic	Sperm DNA integrity test, IMSI, morphological evaluation (oocytes and sperm)	Safe ovarian stimulation protocols, Selection of competent gametes or referral to donation programs
	Genetic	Time-lapse incubation, PGT	Selection of the highest quality embryos
	ART conditions	Periodic revisions	Optimization of ARTs
Maternal	Lifestyle	Detailed history	Healthy recommendations
	Uterine alterations	Ultrasonography, hysteroscopy, hysterosalpingography	Surgical resection or salpingectomy
	Endocrine disorders	Ultrasonography, blood test	Hormonal therapy
	ART conditions	Periodic revisions	Optimization of ARTs
	Thrombophilia, APS, immunological alterations, endometritis	Not standardized / experimental	Mitigate confounding effects of comorbidities and/or their treatment
	Endometrial RIF	Molecular technologies	Personalized strategies*

*APS, antiphospholipid syndrome; ARTs, assisted reproduction treatments; DNA, deoxyribonucleic acid; IMSI, intracytoplasmic morphologically-selected sperm injection; PGT, preimplantation genetic testing; RIF, recurrent implantation failure. *This thesis dissertation could contribute to develop personalized strategies for patients with endometrial RIF in the future.*

- **Embryonic factors**

Safe ovarian stimulation protocols that ensure the retrieval of good-quality oocytes (Takahashi et al., 2004), and a good system for testing sperm DNA integrity (Das & Holzer, 2012), or rigorous sperm selection via intracytoplasmic morphologically-selected

sperm injection [(IMSI); that is still under research] (Shalom-Paz et al., 2015), are some of the measures that clinicians use to obtain **quality gametes** for patients with RIF undergoing ARTs. However, in cases where these strategies are not feasible, semen and oocyte donation programs are recommended. On the other hand, the use of a time-lapse incubation system for rigorous morphologic observation could help to select **good-quality embryos** (Cruz et al., 2011). Although the identification of aneuploid embryos through preimplantation genetic diagnosis [PGD; commonly known as preimplantation genetic testing (PGT)] could potentially improve implantation rates, only the cohort of RIF patients with advanced maternal age (≥ 35 years old) may benefit (Coughlan, 2018; Harper et al., 2006). The optimization of **ART conditions** such as an adequate embryo culture, a suitable assisted hatching and the use of ultrasound-guided embryo transfer with a catheter adapted to remove cervical mucus, could all contribute to ensure the production of good-quality embryos that reduce the risk of implantation failure (Coughlan, 2018; Das & Holzer, 2012).

- **Maternal factors**

Once the embryonic factors are evaluated and quality embryo is ensured, the next step is to assess the potential maternal factors. First, taking detailed patient histories allows clinicians to evaluate each patient's **lifestyle habits** and provide personalized recommendations to implement healthy habits. Routine ultrasounds, hysteroscopy and hysterosalpingography may help identify some **uterine alterations**, and surgical dissection or salpingectomy are some of the measures to treat these alterations. On the other hand, **endocrine disorders** are diagnosed by ultrasound and/or blood tests and are usually controlled with hormonal therapies prior to initiating ARTs. Periodically revising and maintaining high standards for **ART conditions** also helps improve reproductive

outcomes. In addition, diagnostic and therapeutic options for **other comorbidities** such as thrombophilia, APS, immunological alterations or endometritis are less standardized in the clinical practice or still experimental (Bashiri et al., 2018; Coughlan et al., 2014, 2018; Moustafa & Young, 2020; Simon & Laufer, 2012; Timeva et al., 2014).

Endometrial RIF is one of the most difficult maternal factors to evaluate and treat because its molecular mechanisms are poorly understood. Even when embryonic and potential environmental risk factors have been addressed, endometrial RIF may still occur in an uterus that appears normal on imaging techniques, or in an abnormal uterus that was previously treated (pharmacologically or surgically). In this regard, studies, such as **this thesis dissertation**, that **employ advanced technologies** (such as those detailed in the next section) **to reveal which molecular mechanisms are dysregulated in RIF at the endometrial level**, will set the foundation for the development of novel personalized therapeutic strategies (Diaz-Gimeno et al., 2011; Koot et al., 2016; Sebastian-Leon et al., 2018). Improving clinical management of RIF could help decrease the psychological, physical, and socioeconomic challenges that patients face with repeated implantation failures, as well as the frustration of the professionals that try to manage these cases and cost on the public health care system (Bashiri et al., 2018; Coughlan, 2018).

3. Transcriptomics for the study of endometrial factor

3.1. Omic sciences and precision medicine: Transcriptomics

The exponential technological advances of the last two decades have revolutionized the ability to study biological processes, from single molecules to simultaneous evaluation of the whole genome. The emergence of the **omic sciences** (i.e., genomics, transcriptomics, or proteomics) initially entailed the identification of genes, transcripts or proteins, present

in a single cell or tissue, at any given time and condition (Manzoni et al., 2018). However, the field of omics is ever-expanding with more specific and/or improved functional approaches, such as functional genomics (the analysis of how genetic sequences affect gene expression profiles), epigenomics (the assessment of epigenetic modifications), exomics (the analysis of exons), metabolomics (the analysis of metabolites), secretomics (the analysis of secreted products), lipidomics (the large-scale analysis of lipid species), interactomics (the study of biomolecule interactions) and pharmacogenomics (the study of how a person´s genes affect their drug responses), to name a few (Hernandez-Vargas et al., 2020; Olivier et al., 2019). The fact that the omic sciences made it possible to identify multiple types of potential biomarkers, or therapeutic targets, in different biological processes, fostered the leap from the reductionist to global/holistic analytical approaches emerging in research (Manzoni et al., 2018; Olivier et al., 2019). The multidisciplinary analysis and interpretation of omic data, as a whole, through effective and integrative methodologies, with fundamental support of bioinformatics, brought about "The Big Data Era" (Yu & Zeng, 2018).

A major goal of **biomedical research** is to characterize the molecular changes that underlie the development and progression of **complex human diseases**. As a result, omic sciences have been postulated to be key for advancing precision medicine in clinical practice (Manzoni et al., 2018; Olivier et al., 2019). **Precision or personalized medicine** is understood as tailored preventative, diagnostic and therapeutic strategies that take into account the individual characteristics of each patient to improve their clinical outcomes. Notably, precision medicine is also referred to as **stratified medicine** because it uses large-scale data including clinical, lifestyle, behavioural, genetic, and other biomarker information, to stratify patient populations (i.e., developing disease taxonomies), which goes beyond the conventional "signs-and-symptoms" approach. Establishing novel

taxonomies or disease groups improves treatment precision and phenotypic recognition due to better understanding of critical causes of pathology (Koenig et al., 2017).

Transcriptomics is a recognized tool in precision medicine, approved by the Food and Drugs Administration (FDA) (Shi et al., 2010; Su et al., 2014), that reveals the disease-related changes in global gene expression patterns. Transcriptomics differs from genomics, which focuses on alterations in gene sequences rather than their expression and employs more reproducible measurement techniques than proteomics ***(Figure 6)***. The **transcriptome** refers to the sum of all ribonucleic acid (RNA) transcripts in a cell, including coding [1-4%; messenger RNA (mRNA)] and non-coding (> 95%; ribosomal, transfer, small nuclear, small interfering, micro- and long non-coding RNAs) transcripts. Although analyzing mRNA provides direct insight into cell- and tissue- specific gene expression features (Manzoni et al., 2018), other non-coding RNA molecules are indirectly involved in gene expression regulation, and can thus provide additional insight (Wei et al., 2017).

In 2001, the completion of the Human Genome Project (Craig Venter et al., 2001) transformed the understanding of molecular genetics and catapulted the potential of precision medicine. Indeed, the Human Genome Project showed that the human genome only has about 20,000-25,000 protein-coding genes (Carrasco-Ramiro et al., 2017). Originally, a gene was simply defined as a molecular unit of heritable information, however, once the central dogma of molecular biology was established, genes were redefined as a segment of DNA, encoding a protein with a specific function. Accordingly, **gene expression** was characterized by transcription (the conversion of the DNA sequence to an RNA copy) followed by translation (decoding of the RNA copy to produce a chain of amino acids, which ultimately folds into active proteins).

However, the concept of gene expression, as we know it today, is much more complex, as it accounts for regulation processes at the transcriptional, post-transcriptional, and post-translational levels (Feero et al., 2010). Transcriptional regulation includes epigenetic changes, such as DNA methylation and histone modifications, changes in the three-dimensional structure of DNA, as well as the RNA-DNA and protein-DNA interactions. Regarding post-transcriptional regulation, alternative splicing is a process where single-precursor RNA (preRNA) molecules can yield multiple RNA products (depending on the cellular environment) including small non-coding RNA molecules [e.g., microRNAs (miRNA) and small interfering RNAs (siRNA)] that are involved in RNA silencing. Interestingly, non-coding RNA molecules are implicated with epigenetic changes at the transcriptional level (Wei et al., 2017). The post-transcriptional modifications of immature proteins also contribute to the diverse phenotypes produced by the human genome, by yielding multiple proteins implicated in different functions.

It is very important to understand how gene expression is regulated to interpret transcriptomic findings. Concretely, **gene signatures,** or the set of genes whose expression is characteristic of a specific phenotype, may be used as biomarker of a disease, and have become very useful in precision medicine (Magic et al., 2007).

Omic-based approaches for precision medicine have been implemented in multiple medical specialties, including **reproductive medicine**. Understanding the intricacies of infertility, and factors influencing the success rates of ARTs, is complicated; hence, every step of patient care, from identifying the cause(s) of infertility to the transfer of healthy embryos, needs to be precise (Zhang & Yu, 2020).

Endometrial RIF due to a pathological WOI (independent of displacement) supposes a great challenge because it is a complex and multifactorial disorder, however, **transcriptomics** has already proven to be useful to study **endometrial function** in the context of precision medicine (Diaz-Gimeno et al., 2011, 2014, 2017; Koot et al., 2016; Sebastian-Leon et al., 2018), and therefore **may aid in the molecular characterization of the pathological WOI.**

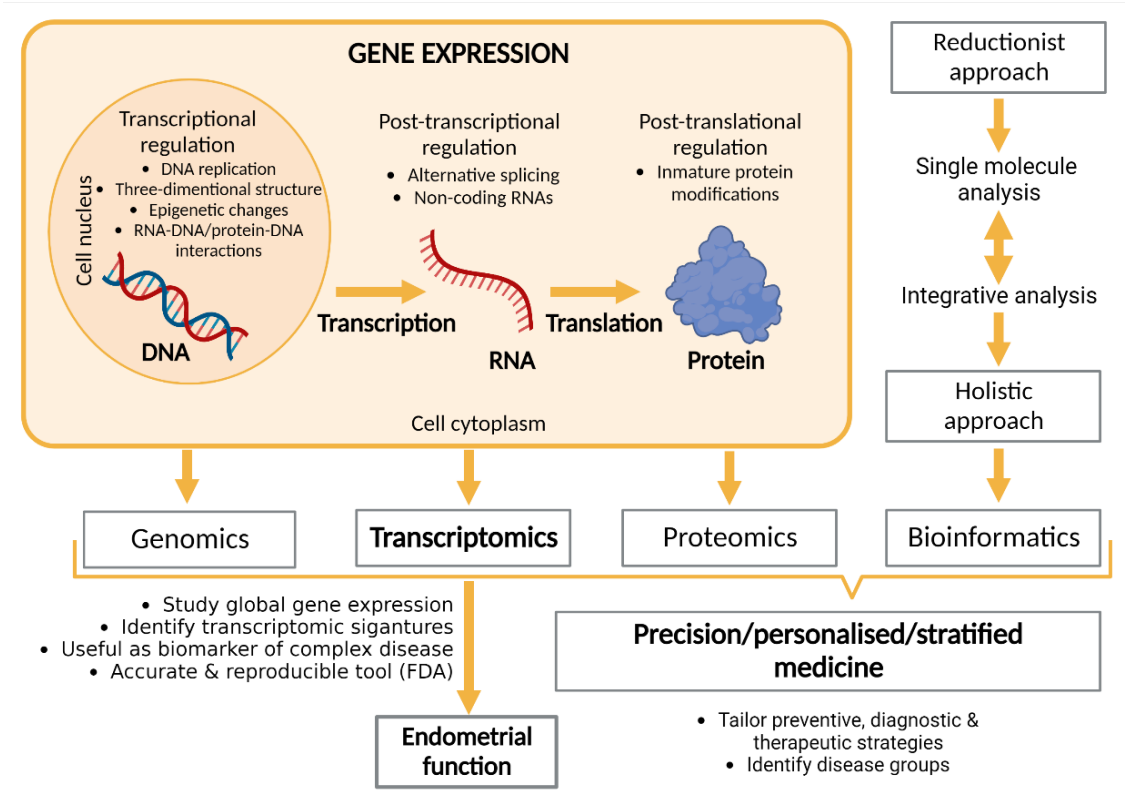


Figure 6. Transcriptomics and precision medicine to study endometrial function.
This diagram depicts the complexity of gene expression and its relationship with omic sciences, highlighting the application of transcriptomics in precision medicine to study endometrial function. DNA, deoxyribonucleic acid; FDA, Food and Drugs Administration; RNA, ribonucleic acid. Created with BioRender.com (2022).

3.2. Technologies to obtain transcriptomic data: RNA-Sequencing

The development of the Northern Blot in 1977 facilitated RNA-based gene expression studies, allowing researchers to study a single or a small group of genes. However, this kind of analysis was deprecated with techniques that provided more precise information,

like **quantitative polymerase chain reaction (qPCR)** developed in 1992 (Segundo-Val & Sanz-Lozano, 2016). Notably, qPCR is still employed to study specific genes, but does not have the capacity to analyse the whole transcriptome at once. The **two main high-throughput technologies** to measure global gene expression are **microarrays** (also known as DNA chips), and next-generation sequencing (NGS) technologies like **RNA-Sequencing (RNA-Seq)**.

Microarrays were originally developed around 1995, as the first hybridization-based approach to evaluate the transcriptome (Schena et al., 1995). A microarray is a solid support of glass, plastic, or nylon, that contains a collection of DNA probes complementary to a large number of genes of interest. Each gene is represented by a single or multiple probes, in defined positions of the support, called probe cells. Microarrays can be designed to measure expression from a specific group of genes, or from all genes of the transcriptome. Currently, microarray platforms are mainly commercially supplied by Affymetrix®, Agilent®, and Illumina®. Regarding the methodology, RNA is extracted from experimental and reference biological samples, reverse transcribed to complementary DNA (cDNA), and fluorescently labelled with different colours to compare gene expression in the distinct study conditions (e.g., pathology vs. control) (Schena et al., 1995; Segundo-Val & Sanz-Lozano, 2016). The cDNA is then deposited on the microarray to hybridize with the corresponding probes, and the microarray plate is scanned to read the fluorescent hybridization signal from each array spot. Finally, the expression level of each gene represented in the microarray is visualized with the fluorescence intensity.

The first-generation sequencing method was proposed by Sanger in 1997 and was employed for the Human Genome Project in 2001. With the subsequent technological

advances, NGS emerged in 2006, allowing parallel, massive, automated, and rapid sequencing of millions of DNA fragments in a single run (Manzoni et al., 2018).

RNA-Seq was developed in 2008, as a method for determining the order of nucleotides from cDNA fragments, in order to quantify global gene expression (Shendure et al., 2017). In fact, RNA-Seq proposed a revolutionary approach to transcriptomic profiling (Wang et al., 2009), that has been gaining popularity over microarrays due to its ample range of benefits ***(Table 2).***

Currently, the main **RNA-Seq platforms** are Illumina®, Ion Torrent® from Thermo Fisher Scientific, Pacific Biosciences (PacBio®) and Oxford Nanopore Technologies (ONT®). However, **Illumina® is the most employed platform** due to its high-throughput capacity, variety of protocols for different applications, and well-established methodologies that avoid biases and error profiles (Stark et al., 2019) ***(Table 3).***

Table 2. Comparison between microarrays and RNA-Sequencing technologies.

Microarrays	RNA-Sequencing
Known transcripts only	Novel transcripts sequences
Does not provide alternative splicing information	Provides structural variations & alternative splicing information
Lower sensitivity and accuracy (no single-base resolution)	Higher sensitivity and accuracy
Lower reproducibility (complicated experimental comparisons)	Higher reproducibility
Lower dynamic range (cross-hybridization & saturation signals)	Higher dynamic range
Limited sample comparisons	Unlimited sample comparisons
Lower cost	Higher cost (but is decreasing)
qPCR validation with some genes is necessary	qPCR validation with some genes is recommended

Advantages of RNA-Sequencing technique (on the right) with respect to Microarrays (on the left) are summarised in this table. qPCR, quantitative polymerase chain reaction.

Table 3. Comparison of the main RNA-Sequencing platforms.

Platform	Illumina	Ion Torrent	PacBio	ONT
NGS type	Second generation	Second generation	Third generation	Third generation
Sequencing type	Synthesis	Semiconductor	Single molecule at real time	Single molecule at real time
Reads type	Short-read (50-500 bp)	Short-read (200-600 bp)	Long-read (> 50 kb)	Long-read (1-10 kb)
Throughput	100-1,000 times more than long-reads	100-1,000 times more than long-reads	500,000-1 M reads/run	500,000-1 M reads/run
Input molecule	cDNA from RNA (even degraded)	cDNA from RNA (even degraded)	cDNA from RNA (not degraded)	RNA or cDNA from RNA (not degraded)
Sample preparation	More steps	More steps	Less steps	Less steps
Protocols and computational workflows	Largest catalogue	Large catalogue	More suited to *de novo* transcriptome analysis or similar complex analysis	More suited to *de novo* transcriptome analysis or similar complex analysis. Ribonucleotide modifications can be detected
Biases and error profiles	Very well understood	Somewhat understood	Not well understood	Not well understood

The main characteristics of Illumina®, Ion Torrent®, Pacific Biosciences (PacBio®) and Oxford Nanopore Technologies (ONT®) platforms are presented. bp, base pair; cDNA, complementary deoxyribonucleic acid; kb, kilobase; M, million; NGS, next-generation sequencing; RNA, ribonucleic acid.

The **core steps of Illumina® protocols** are as follows (Stark et al., 2019) ***(Figure 7)***:

1. **RNA extraction.** Total RNA or a specific type of RNA is extracted from biological samples, and is fragmented, or not, depending on the protocol. Several protocols exist to extract RNA manually or using commercial kits, and the necessary RNA input will depend on the Illumina® protocol.
2. **Reverse transcription.** Extracted RNA is converted to cDNA through the action of reverse transcriptases.
3. **Adaptor ligation.** Some adaptors are ligated to the 3′ and 5′ ends of cDNA fragments. These adaptors include sequences employed to identify the original

sample of the fragments called barcodes or indexes [e.g., dual indexes like i5 and i7], link with the sequencing support (flow cell) and clustering [e.g., P5 and P7], as well as provide sequencing primer binding sites for read 1 and read 2 [e.g., Rd1 SP and Rd2 SP].

4. **Amplification.** The cDNA fragments with adaptors are amplified by polymerase chain reaction (PCR) according to the Illumina® protocol.
5. **Size selection.** Amplicons are selected by size (150-200 bp) using bead-based library purification.
6. **Sequencing.** After generating the library, a sequencing step is required. Concretely, Illumina® systems employ sequencing by synthesis (SBS).

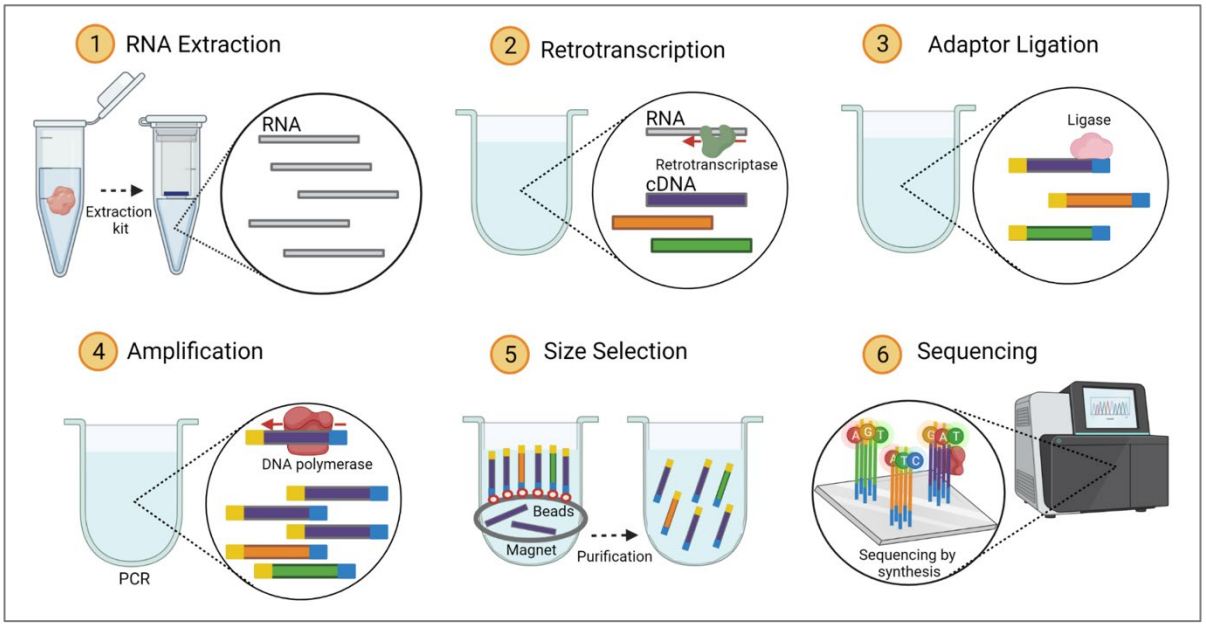


Figure 7. General workflow of Illumina® RNA-Sequencing protocols.
Different phases are numbered and ordered from 1 to 6. In sequencing by synthesis, individual purified DNA molecules are first clustered on a flow cell using a bridge amplification. DNA polymerase then synthesizes the cDNA of these fragments using 3′ blocked fluorescently-labelled nucleotides. In each cycle of sequencing, the growing DNA stand is imaged, to detect which of the four fluorophores has been incorporated and generate the reads (order of nucleotides of each DNA fragment). cDNA, complementary deoxyribonucleic acid; DNA, deoxyribonucleic acid; PCR, polymerase chain reaction; RNA, ribonucleic acid. Created with BioRender.com (2021).

To design a good Illumina® RNA-Seq experiment, several factors need to be taken into **particular consideration** (Stark et al., 2019):

- **RNA integrity.** Most Illumina® protocols require good-quality RNA samples. To avoid RNA degradation, samples should be cleaned on ice, in RNase-free conditions, and stored at -20ºC or -80ºC (ideally with RNA-Later, which maintains RNA integrity). Principal indicative parameters of RNA quality include purity ratios of absorbance against proteins (where A260/A280 values of 1.8-2.0 indicate pure samples) or other contaminants (where A260/230 values of ~2.2 indicate pure samples), an RNA integrity number (RIN) > 6, and/or > 70% of RNA fragments with more than 200 nucleotides (DV200).
- **Level of replication.** Sufficient biological replicates that represent the innate variabilities among samples in the study condition should be included in the experimental design. Ideally, a study should include a minimum of 4 to 6 biological replicates in each condition. Notably, a pilot study with a few samples is recommended to develop before the complete RNA-Seq study. On the other hand, technical replicates ensure the consistency of the technology or laboratory procedure employed. Although RNA-Seq is considered very accurate, some positive and negative controls may also be sequenced.
- **Sequencing read depth.** Read depth refers to the target number of sequence reads to be obtained for each sample and is calculated by dividing the total number of reads employed by the number of samples. For gene expression studies in eukaryotic genomes, 10-30 million (M) read depths per sample are generally accepted. However, when only relatively large changes in the expression of main genes are of interest, and there are adequate replicates, less sequencing may be sufficient.

- **Length sequencing reads.** Length reads are proportional to the number of sequencing cycles, and longer reads cover the sequenced DNA better. While this may be less relevant for gene expression studies that aim to determine where in the transcriptome each read came from, longer reads benefit more qualitative RNA-Seq assays, particularly those aimed at identifying specific isoforms.
- **Single- or paired-end sequencing reads.** Single-end sequencing uses only one end (either 3´or 5´) of each cDNA fragment to generate the read, while paired-end generates two reads for each fragment (one for the 3´direction and another for the 5´direction). Notably, complex assays that require more coverage often use paired-end sequencing, while single-end sequencing can be used for gene expression studies.

In addition, depending on the research objective and starting samples, multiple **Illumina® RNA-Seq protocols** are available for different applications (Illumina, 2021c):

- **mRNA sequencing.** It is used to quantify gene expression, identify known and novel isoforms in the coding transcriptome, detect gene fusions, and measure allele-specific expression.
- **Targeted RNA-Sequencing.** It is employed to analyse gene expression in a specific set of genes of interest, or all genes, via enrichment or amplicon-based approaches. Since these protocols are often used to sequence just a representative region of the gene (known as an assay), they are compatible with low and/or degraded RNA inputs from difficult samples [e.g., formalin-fixed paraffin-embedded (FFPE) ones].
- **Ultra-low-input and single-cell RNA-Seq.** It uses deep RNA-Seq to examine the signals and behaviour of a given cell in the context of its microenvironment, which has been very useful for studying specific processes, such as differentiation, proliferation, and tumorigenesis.

- **RNA exome capture sequencing.** It uses sequence-specific captures of the coding regions of the transcriptome to analyze RNA exomes. These protocols are also compatible with low-quality or limited starting material.
- **Total RNA sequencing.** It measures gene and transcript abundance and detects both known and novel features of coding or noncoding RNAs.
- **Small RNA sequencing.** It sequences small RNA species, such as miRNAs, to elucidate the role of noncoding RNAs in gene silencing and post-transcriptional regulation of gene expression.
- **Ribosome profiling.** It sequences ribosome-protected mRNA fragments, to gain a complete view of the active ribosomes in a given cell, at a specific point in time, and predicts protein abundance.

Likewise, **Illumina® sequencing systems** should be selected according to the chosen protocol and the study objective (Illumina, 2021d) ***(Table 4)***.

Performance parameters to take into account when evaluating the sequencing process in a specific system include the cluster density (quantity of clusters that attach to the flow cell surface), passing filter (PF; percentage reflecting the purity of the signal from each cluster), yield (output of sequencing process), and Phred quality score (Q30; a probability measure that the nucleobases were detected correctly during sequencing), among others.

Table 4. Comparison of the characteristics of the main Illumina® RNA-Sequencing systems.

System	Max reads/run	Max length/reads (bp)	Max output (Gb)	Run time (h)	Main protocols
iSeq100	4 M	2x150	1.2	9.5-19	Targeted
MiniSeq	25 M	2x150	7.5	4-24	Targeted, small RNA
MiSeq	25 M	2x300	15	4-55	Targeted, small RNA
NextSeq500/550	400 M	2x150	120	12-30	Targeted, Single-cell, small RNA
NextSeq1000/2000	1.1 B	2x150	330	11-48	mRNA, Exome, Total RNA, small RNA, Ribosome
NovaSeq6000	20 B	2x250	3000	13-44	Single-cell, Exome, Total RNA

The main characteristics of iSeq100, Miniseq, MiSeq, NextSeq500/550, NextSeq1000/2000 and NovaSeq6000 Illumina® systems are summarized. B, billion; bp, base pair; Gb, gigabase; h, hours; M, million; Max, maximum; mRNA, messenger RNA; RNA, ribonucleic acid.

Once the sequencing process is finished and evaluated, generated data should be preprocessed at computational level and the reads are mapped against a reference genome to obtain the number of reads per transcript, that are proportional to its expression level. Thanks to the use of indexes or barcodes, the gene expression from several samples can be simultaneously studied to compare different conditions (e.g., pathology vs. control) (Stark et al., 2019).

Overall, we considered the **Illumina® RNA-Seq technology to be the best option for the molecular characterization of the pathological WOI in the endometrium of patients with RIF.**

3.3. Bioinformatic approaches for transcriptomic analyses

Bioinformatics is a relatively new multidisciplinary science, that combines knowledge from computational science, statistics, and biology, among other disciplines, to carry out *in silico* analyses of large datasets generated by high-throughput technologies (Yu & Zeng, 2018). Bioinformatic tools are often used to compare transcriptomic profiles of two sample sets (i.e., pathology vs. control).

Once the sequencing is completed, the raw RNA-Seq data undergoes **transcriptomic analysis**, that typically consists of preprocessing and exploratory analyses, followed by differential expression analysis (DEA) and functional analysis ***(Figure 8).***

A) Preprocessing and exploratory analysis

The raw data files generated by the sequencers [usually in binary base call (BCL) sequence format] contain base-called reads and integrated index or barcode sequences that help assign reads to their sample of origin using bcl2fastq. This process, which is called demultiplexing, generates FASTQ files that require a quality control (QC) using FastQC (Andrews, 2020; Illumina, 2021b). To obtain the counts (reads per transcript; which are proportional to the gene expression level), the reads are then mapped to a known transcriptome (or annotated reference genome) using alignment tools like TopHat (Kim et al., 2013), Spliced Transcripts Alignment Reference [STAR; (Dobin et al., 2013)], or HISAT (Kim et al., 2015), and quantified with tools such as RSEM (Li & Dewey, 2011), CuffLinks (Trapnell et al., 2012), and HTSeq (Anders et al., 2015), or its equivalent, featureCounts (Liao et al., 2014). The quantified read counts are usually combined into an expression matrix that has a row for each expression feature (transcript or gene) and a column for each sample. Quantified counts are then filtered, and normalized [usually by quartile or median normalization (Dillies et al., 2013; Stark et al.,

2019)], to account for differences in read depth, expression patterns and technical biases. Following preprocessing, exploratory analyses detect and remove outliers, and batch effects, prior to downstream analysis. The most popular method to achieve this is principal component analysis (PCA), which identifies features that produce the highest variations in gene expression between samples (Jollife & Cadima, 2016), facilitating removal of unwanted effects using linear models.

B) Differential expression analysis

Once data are normalized and corrected, a DEA can be applied to compare the mean expression values of each gene and provides a list of significantly up-regulated/down-regulated genes between samples from different studied conditions [e.g., pathological and control]. The statistical tests employed during this analysis will depend on the experimental design (e.g., number of groups, use of paired samples from the same individual, or independent samples etc.). Since DEA is performed independently for thousands of genes, the obtained p-values must be corrected for multiple testing (especially in case vs. control studies), usually by establishing the false discovery rate (FDR; ratio of false positives to the total number of positives) (Benjamini & Hochberg, 1995). Some of the most employed tools for filtering, normalization, and DEA are edgeR (Robinson et al., 2010), DESeq2 (Love et al., 2014), and limma+voom (Law et al., 2014; Stark et al., 2019).

C) Functional analysis

Functional enrichment analysis (FEA) is one of the main strategies that combines the differential expression levels obtained from DEA, with information from other biological databases, to help identify which biological processes or signaling pathways are most affected in the studied conditions (e.g., pathology vs. control). ClusterProfiler (Yu et al.,

2012) is commonly employed to apply this strategy, that includes over-representation analysis (ORA) and gene set enrichment analysis (GSEA) methods. Specifically, ORA compares a gene list of interest [e.g., differentially expressed genes or (DEGs)] with a background gene list (e.g., a list of all genes evaluated in the study) to distinguish cell functions that are significantly associated to the DEGs. Alternatively, GSEA considers the expression of all evaluated genes (not only the significant ones) ordered according to a feature of interest (such as the p-value or the fold change (FC) obtained from the DEA) and can be used to determine whether a defined set of genes involved in a particular annotated cell function (e.g., immune response) shows overall up-regulated/down-regulated expression in the studied conditions. Since this analysis is performed independently for thousands of functions, p-values must also be corrected for multiple testing, as in DEAs (Khatri et al., 2012; Subramanian et al., 2005).

Due to their high curation level and reliability, Gene Ontology (GO) (Ashburner et al., 2000) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) are biological databases that are widely used to retrieve functional information associated to DEGs. For instance, GO annotates genes to three ontologies: biological processes (BP), the set of molecular events used by cells, tissues, organs, or the whole organism for their functions; molecular functions (MF), the basic activities of a gene product at the molecular level; and cellular components (CC), the parts of the cell and extracellular environment in which a gene product performs a function. Notably, this gene ontology information is mainly inferred from experimental, phylogenetic, or computational evidence. On the other hand, KEGG PATHWAY annotates genes to pathway maps, that illustrate the molecular networks related to metabolism, genetic or environmental information processing, cellular processes, organismal systems and drug development.

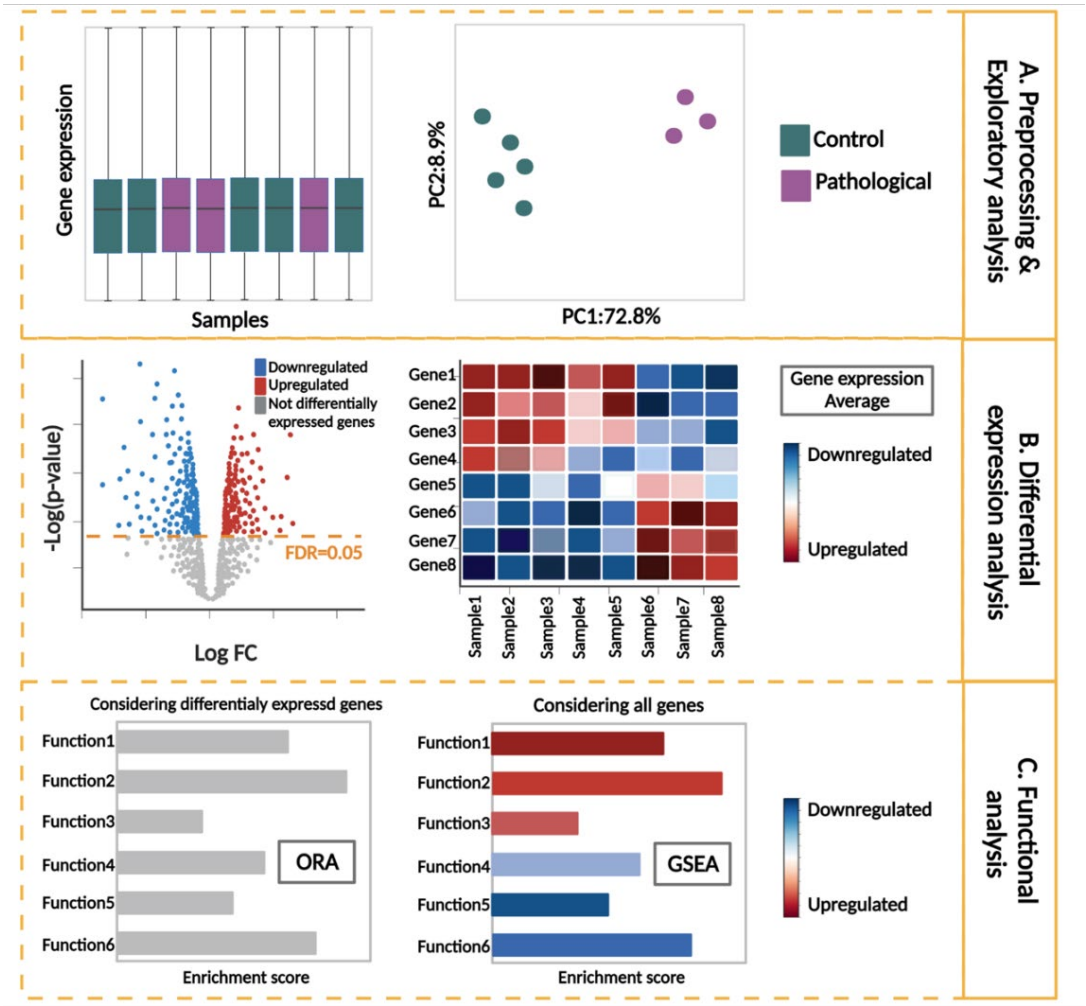


Figure 8. Main steps of transcriptomic data analysis.
Transcriptomic data analysis includes (A) preprocessing and exploratory analysis [e.g., through a boxplot and a principal component analysis (PCA) plot], (B) differential expression analysis (often visualized with a volcano plot and heatmap) and (C) functional analysis (typically represented by bar plots). FC, fold change; GSEA, gene set enrichment analysis; ORA, over-representation analysis; PC1, first component; PC2, second component. Created with BioRender.com (2022).

3.4. Artificial intelligence in precision medicine and transcriptomics

Machine learning (ML) is a branch of artificial intelligence (AI) based on the premise that computerized algorithms can gradually improve decision-making, by learning from data and identifying patterns, with limited human intervention (Maceachern & Forkert, 2021; Mitchell, 1997). Due to the versatile and successful use of ML approaches in various fields, ranging from astronomy (to analyse galaxy images) (Kuminski et al.,

2014), banking (to detect fraudulent transactions) (Perols, 2011), to information technology (to filter spam in email servers) (Guzella & Caminhas, 2009), they have recently been attracting interest in the context of **precision medicine**. In this regard, ML can quickly identify and/or predict patterns in complex clinical and biological data, substantially **improving the efficiency of prognostic and diagnostic healthcare systems** (Gui & Chan, 2017).

A prognosis is an approximation of outcomes (e.g., patient´s susceptibility to disease, likelihood of disease recurrence, life expectancy, or response to treatment), whereas a diagnosis is an identification of disease by examination. ML approaches have been extensively tested in oncology (Van't Veer et al., 2002), and were reported to improve predictions of cancer prognoses by 15-25% (Kourou et al., 2015). In terms of diagnoses, ML algorithms are capable of screening patients and stratifying them by risk, which not only assists physicians in clinical decision making, but also improves the precision of treatment. Thus far, ML models have been built to characterize diseases such as congenital cataracts (Long et al., 2017), skin cancer (Esteva et al., 2017), autism (Wall et al., 2012) and **endometrial receptivity or RIF** (Diaz-Gimeno et al., 2011, 2013, 2017, 2021; Koot et al., 2016; Sebastian-Leon et al., 2018), among others.

In **transcriptomics**, normalized and corrected gene expression data can be input into unsupervised or supervised ML algorithms. **Unsupervised learning** is useful for structuring data and defining profiles (e.g., via clustering methods like k-means) in unlabelled samples, while **supervised learning** algorithms are employed to discover new gene signatures using labelled samples (e.g., pathology and control), so they can predict specific patterns in subsequent unlabelled samples with a certain accuracy [(Acc), quality of being precise], sensitivity [(S), ability to correctly identify patients with a given pathology] and specificity [(Sp), ability to correctly identify people without a given

pathology]. Notably, supervised ML methodology includes an internal validation through an extensive cross-validation process that gives an error estimation avoiding overfitting (Diaz-Gimeno et al., 2014, 2017).

Best practices for applying ML algorithms to transcriptomic data for clinical applications include random forest (RF), support vector machine (SVM) and k-nearest neighbors (kNN) algorithms (Shi et al., 2010; Su et al., 2014). Models developed with these algorithms also called **transcriptomic predictors,** can be modified through the interface Weka (Witten et al., 2016), to be able to combine clinical information and transcriptomic patterns in a reference set of samples (training set), and subsequently externally-validated for robustness in an independent set of samples (test set) (Diaz-Gimeno et al., 2017). In addition, **balancing the training set** according to the studied conditions (e.g., the same number of pathological and control samples) is key to building a proper prediction model without bias. In clinical studies, resampling (over-sampling or under-sampling), is one of the most employed approaches (Alahmari, 2020).

Therefore, ML can be used to predict classifications based on transcriptomic profiles, supporting its use for **disease stratification** ***(Figure 9)***, **especially in the context of complex diseases like endometrial RIF due to a pathological WOI.**

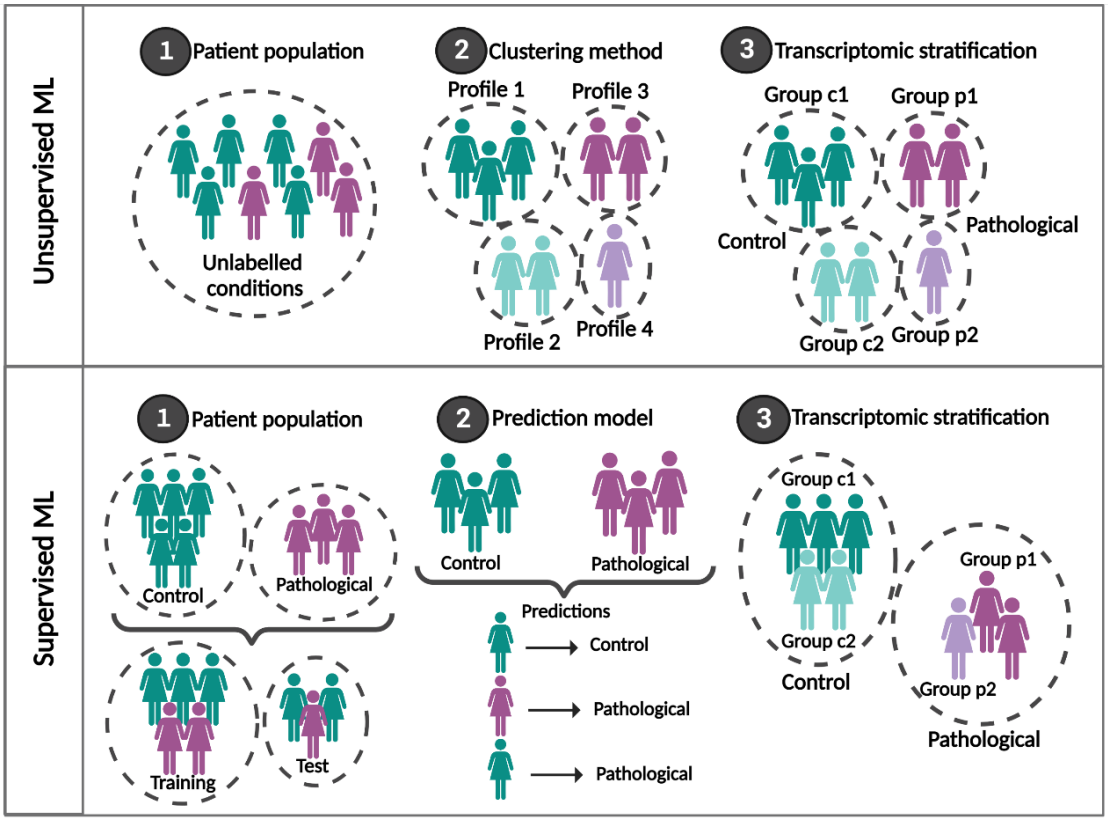


Figure 9. Machine learning for transcriptomic stratification.
Unsupervised (above) and supervised (below) machine learning (ML) methods for transcriptomic stratification are represented. Note, although patients may be clinically classified as control, the supervised ML model may predict that these patients present a pathological transcriptomic profile. Created with BioRender.com (2022).

3.5. Main applications of endometrial transcriptomics and current research context

The endometrium is a dynamic tissue at the morphological and molecular level that implies **complex and multifactorial phenotypes related to the endometrial function**, like **endometrial receptivity** or **RIF of endometrial origin** (Diaz-Gimeno et al., 2011; Koot et al., 2016). **Studying the WOI, and its alterations,** can help clinicians ascertain the optimal period of endometrial receptivity, promote implantation, and ultimately help patients achieve pregnancy. Classically, the WOI was estimated by determining menstrual cycle days, timing ovulation, measuring hormone levels, or using histological

assessments (Noyes et al., 1975). However, recent approaches have included ultrasound, endometrial fluid aspirates, and hysteroscopy (Craciunas et al., 2019), along with single molecular biomarkers of implantation (e.g., adhesion molecules, prostaglandins, growth factors, cytokines, or matrix metalloproteinases) (Fox & Lessey, 2018). Meanwhile, emerging strategies involve studying multi-omic biomarkers of endometrial receptivity (Hernandez-Vargas et al., 2020). In this regard, precise and reproducible **transcriptomic analyses of the WOI** offer in-depth molecular information about endometrial dynamics, and highlight the dysregulated endometrial functions that lead to **complex and multifactorial disease phenotypes**, which can be used to advance precision medicine (Diaz-Gimeno et al., 2014, 2017; Sebastian-Leon et al., 2018).

Transcriptomic studies that have sought to **define the healthy WOI**, often compare the endometrium in different phases to characterize the endometrial receptivity status (Altmäe et al., 2017; Bhagwat et al., 2013; Borthwick et al., 2003; Carson et al., 2002; Diaz-Gimeno et al., 2011, 2017, 2021; Haouzi et al., 2009; Kao et al., 2002; Mirkin et al., 2005; Ponnampalam et al., 2004; Punyadeera et al., 2005; Riesewijk et al., 2003; Sigurgeirsson et al., 2017; Talbi et al., 2006), or RIF patients to controls, to **characterize alterations of the WOI** (Altmäe et al., 2010; Bastu et al., 2019; Bersinger et al., 2008; Koler et al., 2009; Koot et al., 2016; Ledee et al., 2011; Sebastian-Leon et al., 2018; Shi et al., 2018; Tapia et al., 2008). These transcriptomic studies evaluated different endometrial gene signatures, using distinct sample cohorts, technologies, and methodologies. While most signatures improved **understanding of endometrial function at the molecular level**, other signatures have been employed to develop endometrial transcriptomic predictors for **patient stratification and/or diagnostic procedures** (Diaz-Gimeno et al., 2011, 2017, 2021; Koot et al., 2016; Sebastian-Leon et al., 2018).

Endometrial transcriptomic predictors to evaluate the human WOI and endometrial RIF have been applied in two independent studies (Diaz-Gimeno et al., 2011; Koot et al., 2016). The first study (Diaz-Gimeno et al., 2011), that innovatively used predictors with endometrial data, established a transcriptomic signature that was patented and used to develop the endometrial receptivity analysis [ERA test®; (Igenomix, 2021)], which was widely used for identifying the WOI, in clinical practice (Ruiz-Alonso et al., 2013). This transcriptomic signature highlighted that **endometrial RIF** originates from an asynchrony between the embryo and the endometrium, that is due to a **displacement of WOI timing**. Although Diaz-Gimeno's signature helped personalize embryo transfers by tailoring the hours of progesterone treatment required during hormone replacement therapy (HRT) (Ruiz-Alonso et al., 2013), the clinical benefit of personalized embryo transfer remains controversial (Bassil et al., 2018; Cozzolino et al., 2020, 2022; Doyle, Combs, et al., 2022; Doyle, Jahandideh, et al., 2022; Simon et al., 2020). Nevertheless, over the last decade, the ERA test® has served as a precursor for other transcriptomic-based diagnostic tools for endometrial receptivity, such as the Window Implantation test [WIN test® (Hamamah, 2013)], Endometrial Receptivity Map [ER Map® (Enciso et al., 2018)] and transcriptomic endometrial dating [TED® (Diaz-Gimeno et al., 2021)], among others.

Currently, the **transcriptomic evaluation of endometrial-factor infertility** requires an endometrial biopsy collection from the patient, who is usually undergoing **HRT**. Although HRT is mainly indicated for alleviating the effects of menopause (Mumusoglu et al., 2021), it can also effectively prepare the endometrium of *in vitro* fertilization (IVF) patients for embryo transfer (Cagnacci & Venier, 2019). Unlike natural and ovarian stimulation cycles, HRT involves the administration of exogenous oestrogen and progesterone, to balance hormonal levels and control menstrual cycle progression,

without altering endometrial function, respectively. Ideally, endometrial biopsies are collected approximately five days after the beginning of progesterone treatment [P+5; to coincide with the theoretical window of implantation (Ruiz-Alonso et al., 2013)], and promptly subjected to RNA-Seq and subsequent transcriptomic analysis for endometrial diagnosis. Alternatively, Macklon´s group (Koot et al., 2016) proposed a transcriptomic signature based on the premise that endometrial RIF results from a **pathological or disrupted WOI**, independent of displacements with a heterogeneous profile. This group was the first to **correct endometrial progression effects (**considering the LH day when the endometrial biopsy was collected) before studying the pathology, to avoid possible biases that mask potential endometrial pathology biomarkers. However, findings from this study were focus on an unbalanced population (much more control than pathological samples) in natural cycles, and the results were not robust enough to be translated into a diagnostic tool.

These controversial results were recently addressed by Diaz-Gimeno's group, who demonstrated that endometrial RIF could originate from either **a displaced WOI, or a pathological WOI, which may present independently, or simultaneously, in a given patient with endometrial RIF** (Sebastian-Leon et al., 2018). This study not only revealed a new RIF taxonomy, but also established a methodology for clinically distinguishing a patient who only had a displaced WOI and could benefit from a personalized embryo transfer from a patient with a disrupted WOI that requires further characterization/diagnosis. Limitations of this study include that patients were in natural cycles, and clinical follow-up was not feasible due to the *in silico* nature of the analysis [with microarray data from Macklon´s group (Koot et al., 2016)].

In these contexts, the work presented in the herein **doctoral thesis dissertation** was designed to **corroborate the existence of the pathological WOI**, using a prospective clinical study, and **molecularly characterize the heterogeneous character** of the disrupted WOI, using transcriptomic analyses, to develop **alternative strategies for evaluating endometrial-factor infertility**, particularly in patients with endometrial RIF ***(Figure 10).***

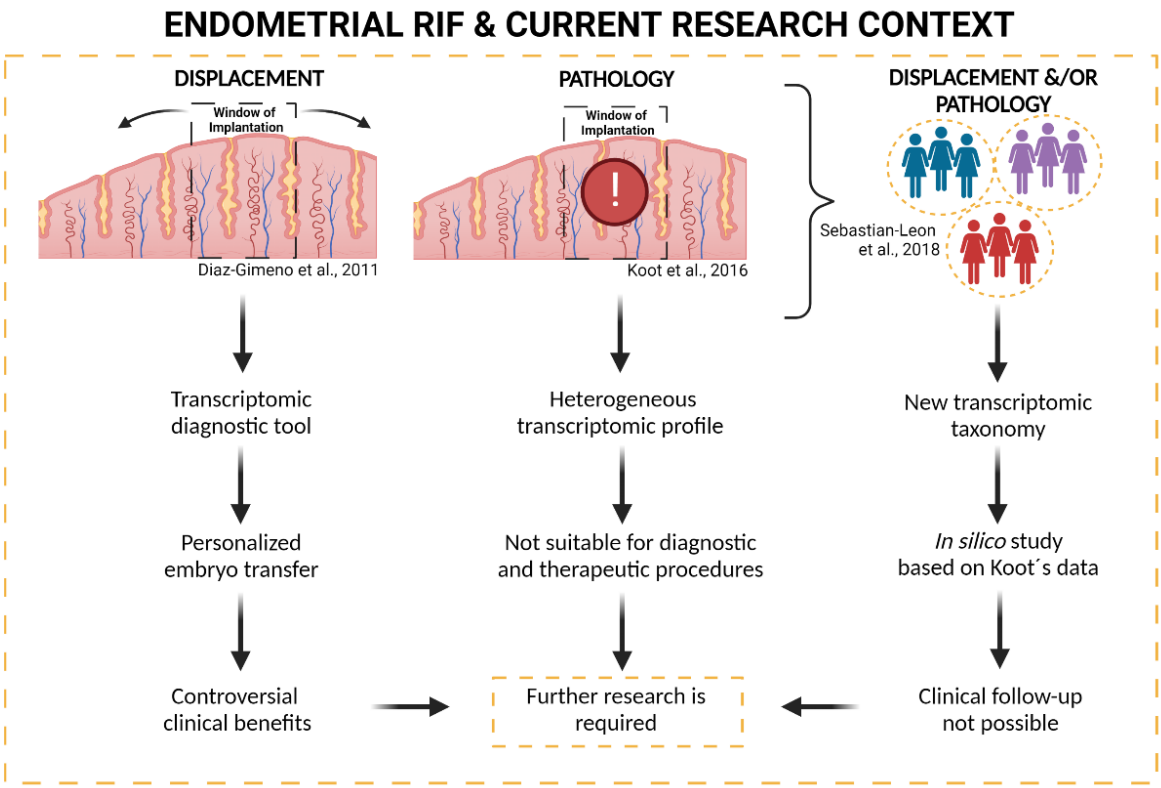


Figure 10. Endometrial recurrent implantation failure and current research context.
The main findings from studies using machine learning algorithms to analyse the endometrial transcriptome of patients with recurrent implantation failure (RIF) are summarized, along with the motivation for the development of this doctoral thesis, which is based on the molecular characterization of the pathological window of implantation, independent of displacement. Created with BioRender.com (2022).

II. HYPOTHESIS

"No great discovery was ever made without a bold guess."

Isaac Newton

HYPOTHESIS

Based on the evidence of a pathological window of implantation independent of endometrial progression (with a heterogenous transcriptomic profile) existing in patients with recurrent implantation failure, undergoing natural cycles for *in vitro* fertilization/embryo transfer, we reasoned that a pathological window of implantation also exists and can be characterized in a cohort of patients with recurrent implantation failure, undergoing hormone replacement therapy for *in vitro* fertilization/embryo transfer.

Further, assessing the transcriptomic profiles of the endometrial biopsies of these patients using RNA-Sequencing and artificial intelligence can shed light on their heterogenous nature, facilitate patient stratification, and ultimately, improve the precision of diagnoses and treatment of endometrial-factor infertility.

III. OBJECTIVES

"Perseverance is a virtue of the less brilliant."

Santiago Ramón y Cajal

OBJECTIVES

The **main objective** of this thesis dissertation was to identify and characterize the pathological window of implantation in a cohort of *in vitro* fertilization patients undergoing endometrial preparation with hormone replacement therapy.

Accordingly, the **specific objectives** were to:

1. Develop a machine learning prediction model capable of identifying the pathological window of implantation, independent of displacements, in the patients undergoing hormone replacement therapy.
2. Stratify the patient population in order to define the different transcriptomic profiles related to the pathological window of implantation.
3. Identify the clinically-relevant association(s) of the transcriptomically-defined groups with reproductive outcomes.
4. Study the underlying molecular mechanisms and functional alterations between the different transcriptomic profiles.

IV. EXPERIMENTAL DESIGN

"Science is organized knowledge. Wisdom is organized life."

Immanuel Kant

EXPERIMENTAL DESIGN

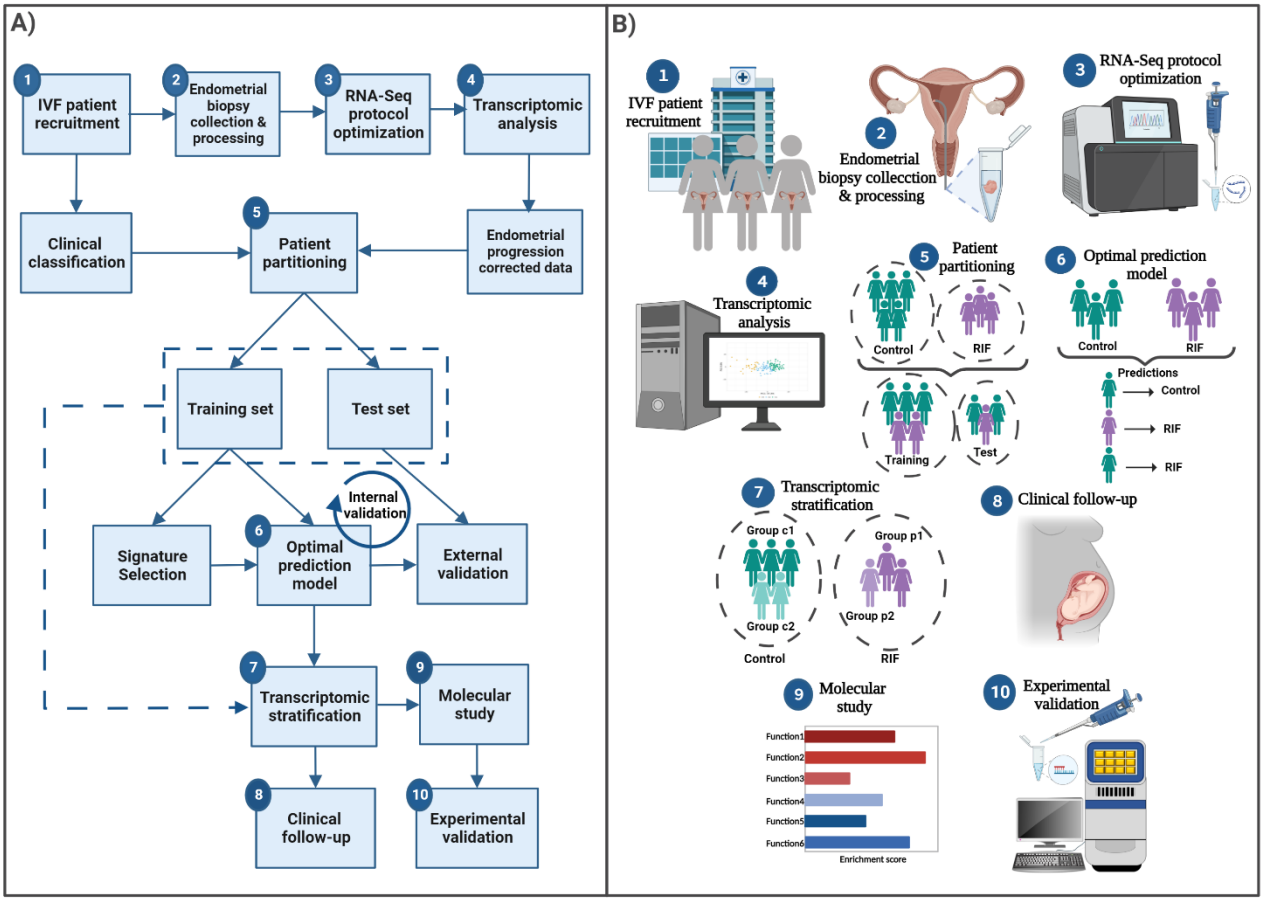


Figure 11. Experimental design.

The ordered sequence of steps that have been performed along this thesis are presented as a flowchart (A) and with illustrations (B). *(1) In vitro fertilization (IVF) patients recommended for endometrial evaluation, in the context of a personalized embryo transfer, inside the routine of clinical practice, were recruited from different IVI clinics. Baseline clinical data and reproductive outcomes were collected from patients to determine their broad clinical classification [control or recurrent implantations failure (RIF)]. (2) Endometrial biopsies were collected in the mid-secretory phase, for the patients' clinical endometrial evaluation during hormone replacement therapy (HRT), and the excess tissue from these biopsies was processed for this transcriptomic study. (3) A whole transcriptome RNA-Sequencing (RNA-Seq) protocol was optimized to process the endometrial biopsies. (4) Transcriptomic data from endometrial biopsies were analysed including a correction of menstrual cycle variations to focus the study on the endometrial pathology. (5) Patient were partitioned into training and test sets considering the clinical classification. (6) The optimal prediction model based on a gene signature of the pathological window of implantation, was selected and internally-validated using the training set. This model was also externally-validated using the test set. (7) The patient population was stratified into different transcriptomically-defined groups according to the optimal machine learning model. (8) A clinical follow-up of the stratified patients was carried out to study the clinical relevance of the new taxonomy. (9) The molecular mechanisms underlying the different transcriptomic profiles were studied using a functional analysis. (10) The most interesting potential biomarkers were validated using quantitative chain reaction (qPCR). Created with BioRender.com (2022).*

V. MATERIAL & METHODS

"You never fail until you stop trying."

Albert Einstein

MATERIAL & METHODS

1. Sample collection from a subfertile patient population

1.1. Study design and participants

Patients were recruited for this multicentric, prospective study between January 2019 and August 2020, in the IVI clinics located in Valencia, Madrid, Barcelona and Bilbao, under the direction of the IVI Foundation (IIS La Fe, Valencia, Spain). A clinical follow-up was carried out during the course of this thesis to record the reproductive outcomes of the participants.

Patients were considered eligible to participate in the study if they:

- were undergoing routine endometrial evaluation prior to *in vitro* fertilization (IVF);
- were prescribed hormone replacement therapy (HRT), with external administration of estradiol valerate (6 mg/day) and micronized vaginal progesterone (MVP; 400 mg/12h) without gonadotropin-releasing hormone (GnRH) agonist for endometrial preparation;
- had good-quality embryos, guaranteed by preimplantation genetic testing (PGT) or oocyte donation (from women < 35 years old);
- were undergoing intracytoplasmic sperm injection (ICSI) for personalized embryo transfer;
- were aged 18-50 years;
- had a body mass index (BMI) of 19-30 kg/m^2;
- presented an endometrial thickness of > 6.5 mm, with a trilaminar aspect on the tenth day of their menstrual cycle (indicative of endometrial proliferation during the proliferative phase).

Patients were excluded if they presented male-factor infertility (with autologous sperm) as the only fertility treatment indication, untreated reproductive pathologies that may compromise endometrial function (e.g., myomas, polyps, adhesions, adenomyosis, uterine malformations and hydrosalpinx), severe pre-menopausal symptoms, uncontrolled systemic or metabolic disorders, or were co-administered medication that can interfere with their reproductive treatments.

1.2. Ethical approval and data protection

This study was approved by the Ethics Committee of the Instituto Valenciano de Infertilidad (IVI), Valencia, Spain (1706-FIVI-048-PD). Written informed consent was obtained from all participants. Relevant clinical information of each participant (e.g., recruiting clinic, type of infertility, type of transfer cycle, years of infertility, treatment indication, patient age, BMI, ethnicity, allergies, comorbidities, drug or alcohol use, smoking habits, stage of endometrial progression, number of transfers, embryonic conditions and reproductive outcomes) was exported from an internal database of medical records (SIVIS) in accordance with the data protection law.

1.3. Clinical classification of the patients

Participating patients were clinically classified as RIF or control, based on their history of implantation failure before and after endometrial biopsy collection, and according to the most common clinical definition of RIF (Bashiri et al., 2018; Coughlan et al., 2014; Koot et al., 2016; Macklon, 2017) ***(Figure 12).***

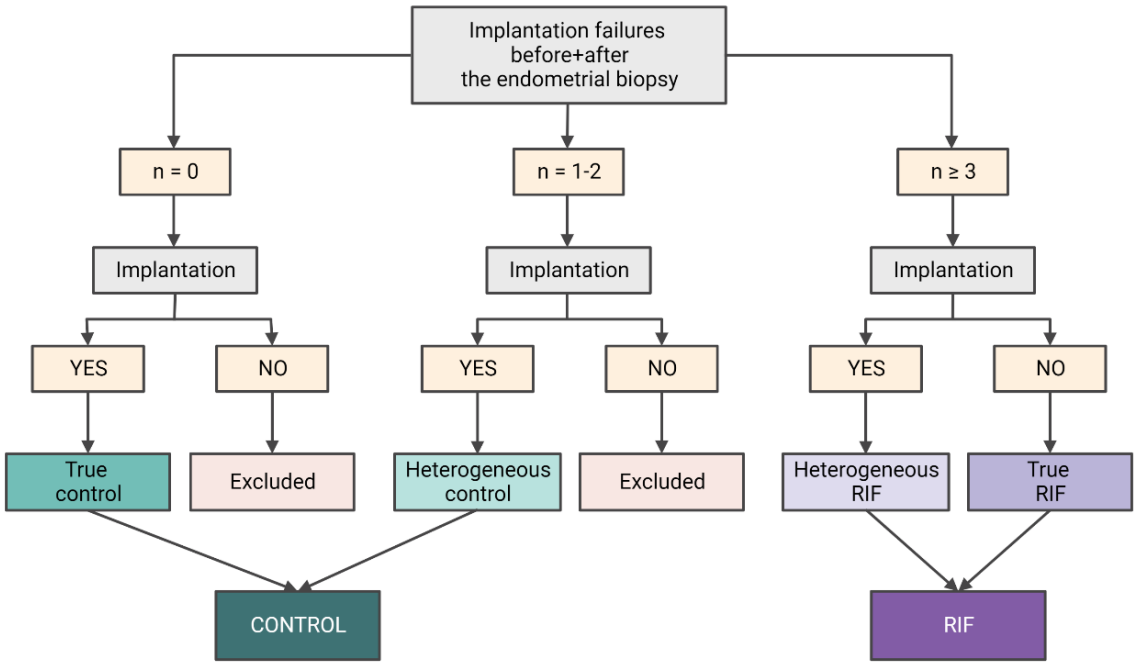


Figure 12. Clinical classification of study participants.
Implantation was considered to have failed when patients presented a negative beta chorionic gonadotropin (β-hCG) value (≤ 10 IU/L) 14-16 days after in vitro fertilization or biochemical miscarriage (defined by a positive serum β-hCG value of ≥ 10 IU/L, but absence of pregnancy within the first 10 weeks of gestation) following a good-quality embryo transfer. Implantation was considered successful in patients who had a biochemical pregnancy (defined by a positive serum β-hCG), clinical pregnancy (defined by a positive serum β-hCG along with the presence of an intrauterine gestational sac and/or fetal heartbeat detected by transvaginal ultrasound within the first 12 weeks of gestation), clinical miscarriage (loss of an ultrasound detected pregnancy within the first 20-24 weeks of gestation), ongoing pregnancy (or ongoing viable intrauterine pregnancy with a detectable fetal heartbeat after 12 weeks of gestation), or live birth (which indicated the end of a successful pregnancy, with the birth of a healthy baby that survives at least the first week of life). For the purposes of this study, ectopic pregnancy was not regarded as a successful implantation or an implantation failure, since the gestational sac is detected outside the uterus. n, number; RIF, recurrent implantation failure. Created with BioRender.com (2022).

The main confounding variables of both groups (i.e., recruiting clinic, years and type of infertility, age, BMI, stage of endometrial progression, number of transfers and implantation failures as well as type of transfer cycle) were compared, and their transcriptomic behaviour was analysed using principal component analysis (PCA) (Jollife & Cadima, 2016). Statistical differences were calculated employing Fisher´s test for qualitative variables (i.e., clinic, type of infertility, stage of endometrial progression and type of transfer cycle), or student´s t-test for normally distributed (Shapiro's test > 0.05)

or Wilcoxon´s test for non-normally distributed (Shapiro´s test < 0.05) quantitative variables (i.e., years of infertility, age, BMI, number of transfers and implantation failures). In all cases, p-values < 0.05 were considered statistically significant. All these analyses were carried out in R environment (version 4.0.5, 2021-03-31) (R Core Team, 2020), and all plots were generated using the ggplot2 package (Wickham, 2011).

1.4. Endometrial biopsy collection

In accordance with standard clinical protocol for endometrial evaluation in the IVI clinics, HRT was prescribed to study participants for endometrial preparation. Estradiol valerate (6 mg/day; Meriestra, Novartis, Barcelona, Spain) was administered from the first or second day of menstruation. After ten days on oestrogen therapy, if the endometrial thickness was ≥ 6.5 mm with a trilaminar pattern, serum progesterone was < 1.0 ng/mL and serum oestrogen > 100 pg/mL, MVP [400 mg/12h; Utrogestan (SEID, Barcelona, Spain) or Progeffik (Effik, Madrid, Spain)] was administered for secretory phase support. An endometrial biopsy was collected from the uterine fundus of each participant, with a cannula Pipelle de Cornier® (CCD Laboratories, Paris, France) in sterile conditions during the expected WOI period, which is usually five days after initiating progesterone treatment (P+5). Notably, HRT cycles are shorter than natural cycles, thus, the start of the WOI is moved forward from approximately the 19th to the 16th day of the cycle ***(Figure 13A)***.

The tissue remaining from the biopsy, after a portion was retained for the patient's clinical use, was immediately transferred into 1.8 mL Nunc cryotubes® (Thermo Scientific, Madrid, Spain) with 900 uL of RNAlater® (Sigma-Aldrich, Madrid, Spain) to maintain RNA integrity ***(Figure 13B)***. The samples were stored at 4ºC (for up to 1 month) until they were transported to the IVI Foundation (IIS La Fe, Valencia, Spain) for

transcriptomic analysis, under recommended RNAlater® conditions. Each cryotube was anonymized, with a code reflecting the recruiting clinic, biopsy collection date, and patient identification.

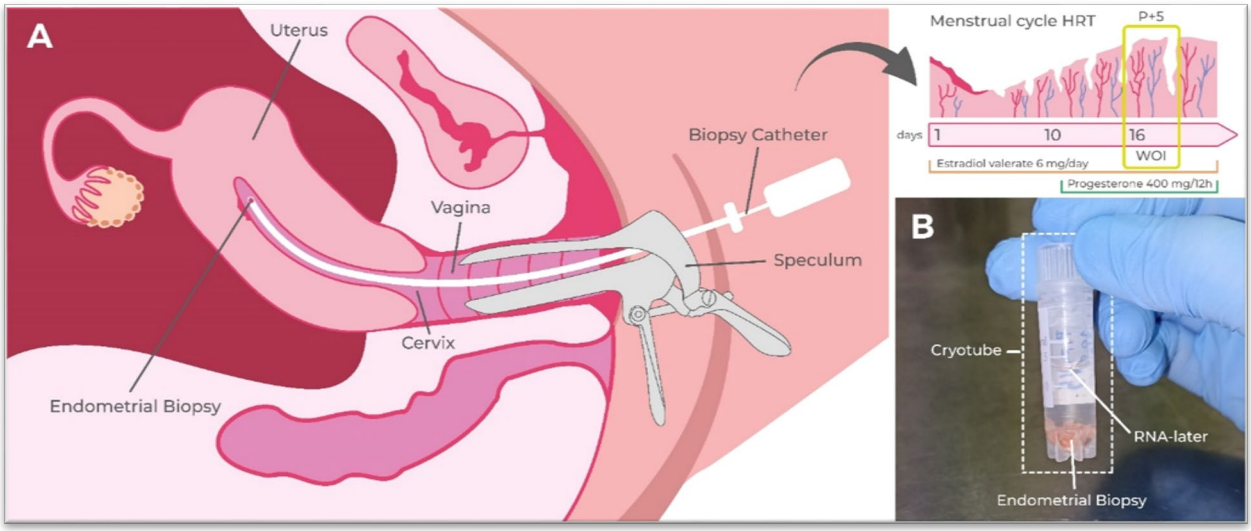


Figure 13. Biopsy collection procedure.
(A) *Schematic depiction of endometrial sampling during the expected window of implantation (WOI) beginning on the 16th day of hormone replacement therapy (HRT) cycles.* ***(B)*** *Macroscopic appearance of endometrial biopsy in a cryotube with RNAlater®. h, hour; mg, milligram; P, progesterone.*

1.5. Endometrial sample processing

Upon reception, the unusual aspects of sample appearance (e.g., low quantity of tissue, very fragmented biopsies, presence of blood or other biological fluid, insufficient volume of RNAlater® or a very cloudy supernatant in the cryotube) were noted and considered in the subsequent analysis of RNA quality. Endometrial samples were cleaned on ice, in RNase-free conditions, with Dulbeco's phosphate buffered saline (PBS®; Capricorn Scientific, Labclinics, Barcelona, Spain) ***(Figure 14A)***, rolled on a 90x20 mm Cell Culture Dish® (SPL Life Sciences, Gyeonggi-do, Korea) using two sterile surgical blades (No.24; Braun, Hessen, Germany) to remove residual fluid ***(Figure 14B)***, transferred to a new cryotube with the same anonymized code, and stored at -80ºC until RNA extraction.

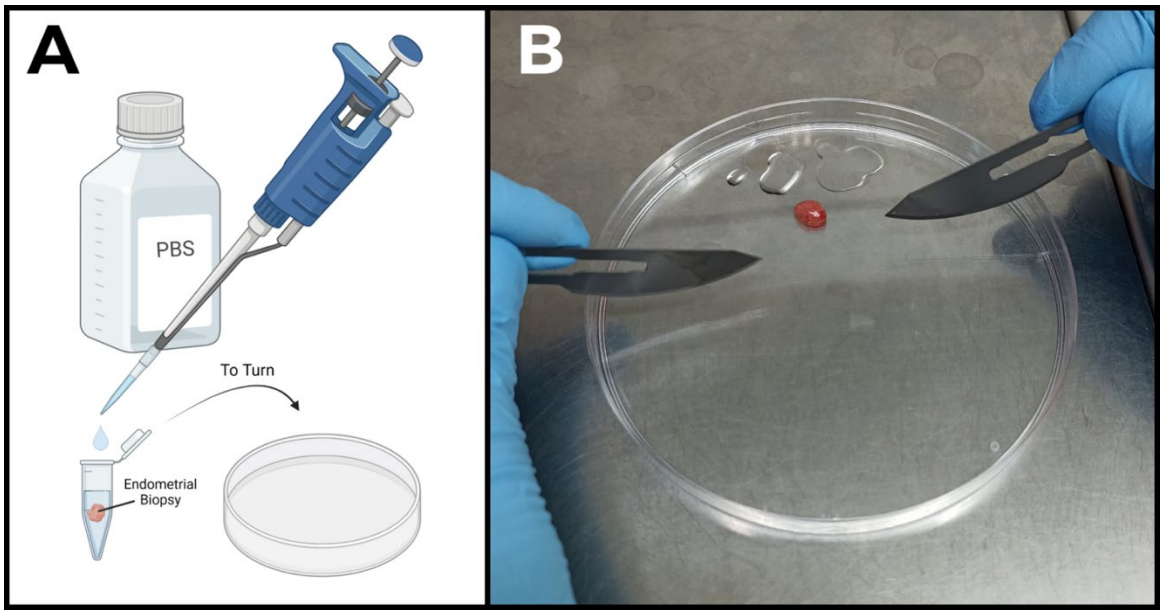


Figure 14. Endometrial sample processing procedure.
(A) *Upon reception, endometrial biopsies were rinsed using phosphate buffered saline (PBS). Created with BioRender.com (2022).* ***(B)*** *Macroscopic appearance of a human endometrial sample after removing the excess fluid (that can be seen at the top of the cell culture plate).*

2. Implementation of an optimal RNA-Sequencing protocol and transcriptomic analysis

2.1. RNA extraction and quality assessment

Endometrial samples were disrupted manually ***(Figure 15A-B)***, or using the Tissulyser II (Qiagen, Hilden, Germany), in RNase-free conditions. Total RNA was extracted using the miRNeasy Mini Kit® (Qiagen, Hilden, Germany) ***(Figure 15C)*** in the laboratories of the IVI Foundation or biobank facilities (IIS La Fe). RNA extractions were programmed in batches of 5-10 samples, and any technical issues were noted to study their effect on the transcriptomic behaviour of the samples. RNA quality was assessed with the NanoDrop ONE® (AF-00342; Thermo Fisher Scientific, Valencia, Spain) in the laboratories of the IVI Foundation, or 4200 TapeStation System® (Agilent, Valencia,

Spain) by the Genomics department of the IIS La Fe. RNA samples were stored at -80ºC and excluded from subsequent transcriptomic analysis if they did not meet the required quality criteria (i.e., a 260/280 ratio of ~2.0, a 260/230 ratio between 1.8-2.2, RIN ≥ 3 and DV200 ≥ 70%).

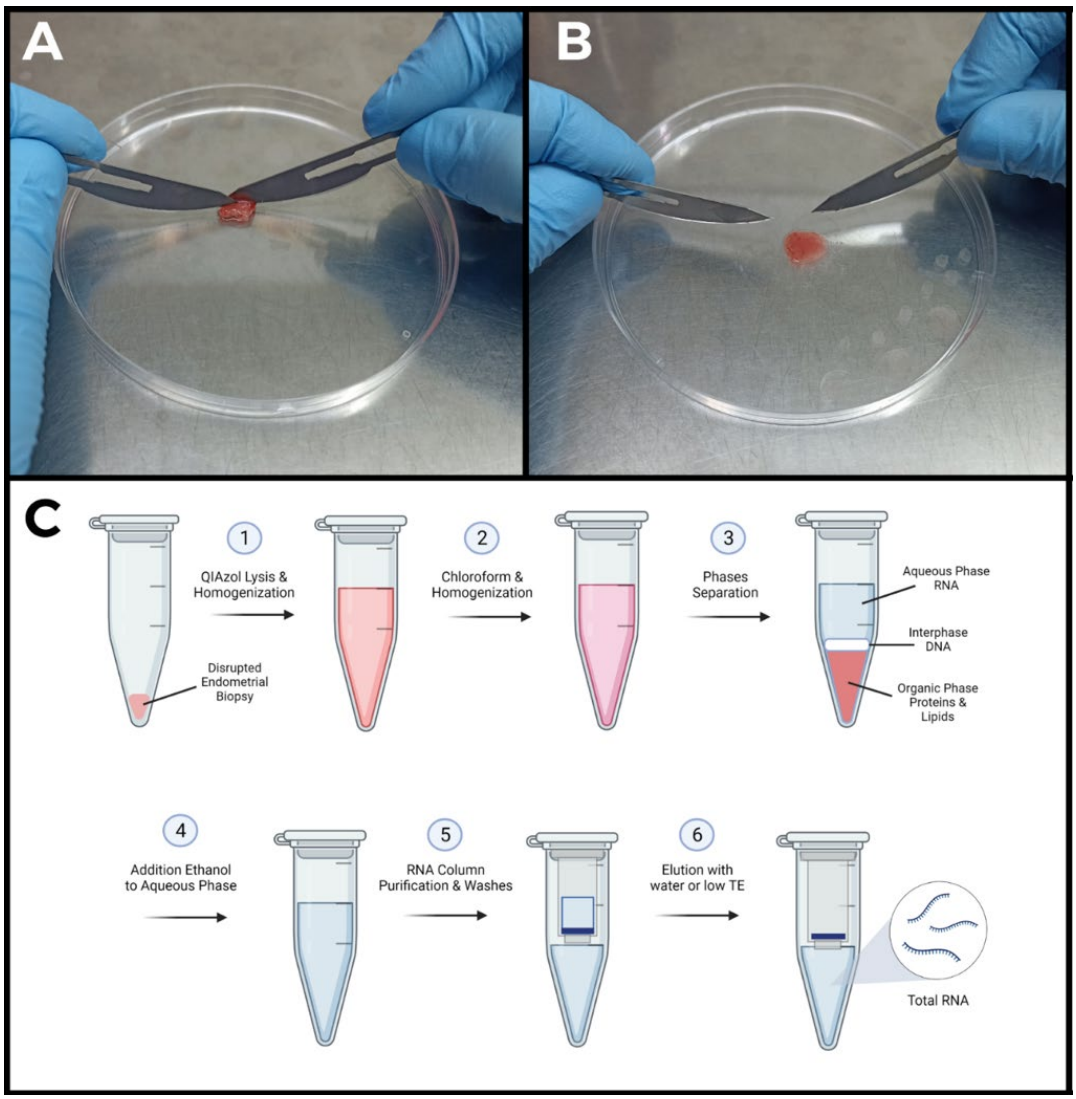


Figure 15. Sample disruption and RNA extraction procedure.
The macroscopic appearance of endometrial biopsies before ***(A)*** *and after* ***(B)*** *manual sample disruption is shown.* ***(C)*** *Workflow of RNA extraction using the miRNeasy Mini Kit® (Qiagen, Hilden, Germany). DNA, deoxyribonucleic acid; RNA, ribonucleic acid; TE, Tris-EDTA. Created with BioRender.com (2022).*

2.2. Library generation and sequencing

The endometrial RNA libraries were generated using the AmpliSeq for Illumina® Transcriptome Human Gene Expression Panel (Illumina, 2021a, 2022a) according to manufacturer's instructions ***(Figure 16)***. This panel targets amplicons in the whole transcriptome (20,802 amplicons of 104 bp), and was selected due to its compatibility with high- and low-quality RNA samples from human tissue. Each sequencing batch included a universal human RNA (Agilent, Valencia, Spain) as a positive control, RNase-free water (Qiagen, Hilden, Germany) as a negative control, and two biological samples as technical replicates. The sequencing process was carried out on a NextSeq500/550 system, using a paired-end design, with 150 cycles and 10 M reads per sample, according to Illumina® recommendations.

A pilot study with 40 RNA samples was carried out in IVI Foundation to validate the protocol, and the remaining RNA samples were subsequently sequenced in Juno Genetics laboratories (Oxford, UK) during research stay (between September-December 2020), and the Genomic Department of the IIS La Fe. Samples with excessive (> 12 M) or insufficient (< 5 M) reads were resequenced, remaining as duplicated, to optimize the number of reads. The samples in each sequencing batch were recorded to evaluate the possible technical effects on the transcriptomic behaviour of the samples.

To validate the performance of our sequencing process, we verified that our clusters had an adequate density (170-220 K/mm^2) and high PF (≥ 90%), and that we yielded an estimated 50-60 Gb output with high read quality (> 80% data > Q30) (Illumina, 2022c, 2022b).

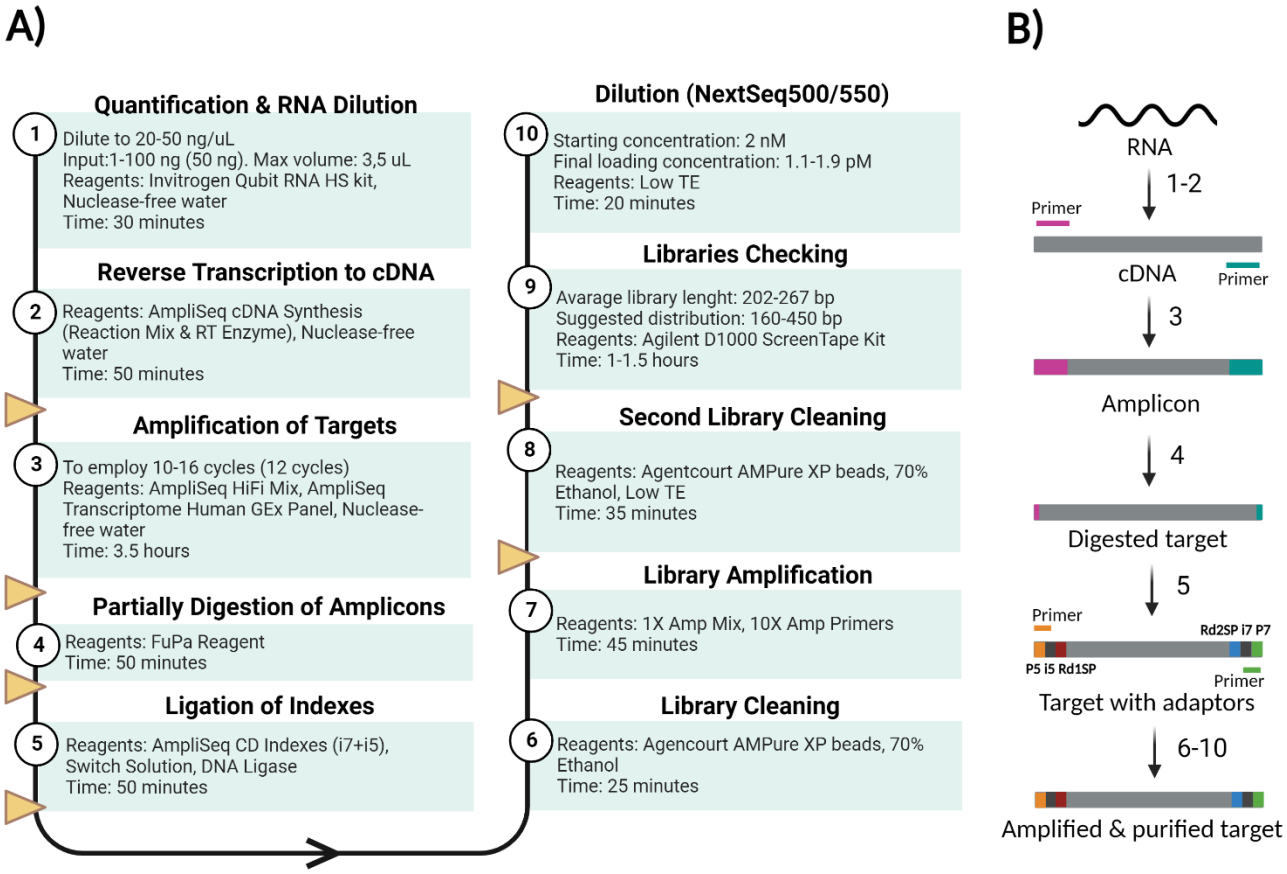


Figure 16. Optimized Illumina® AmpliSeq library generation procedure used in the study. ***(A)*** *The main steps of the Illumina® AmpliSeq protocol are listed, along with the necessary reagents, optimized parameters in brackets, and required time for each step. Orange triangles indicate safe stopping points in the protocol.* ***(B)*** *Molecular events corresponding with the different steps of the library generation. Bp, base pair; CD, combinatorial dual; cDNA, complementary DNA; i, index; i5, index sequence 5 or barcode 5; i7, index sequence 7 or barcode 7; Max, maximum; ng, nanogram; nM, nanomolar; P5, sequence 5 that binds to flow cell; P7, sequence 7 that binds to flow cell; pM, picomolar; Rd1 SP, sequencing primer binding site read 1; Rd2 SP, sequencing primer binding site read 2; RNA, ribonucleic acid; RT, reverse transcription; TE, Tris-EDTA; uL, microliter. Figure adapted from* (Illumina, 2021a, 2022a). *Created with BioRender.com. (2022).*

2.3. Transcriptomic data preprocessing and exploratory analysis

Raw sequencing data was demultiplexed using bcl2fastq, and its quality was evaluated using FastQC (Andrews, 2020; Illumina, 2021b). Reads from the whole transcriptome RNA-Seq library were aligned against a reference sequence set containing the 20,802 transcripts targeted by the AmpliSeq panel, using STAR (Dobin et al., 2013), and quantified using featureCounts (Liao et al., 2014), to exclude low quality counts ($Q < 30$).

This pipeline was written in Python (version 3.8) (Van Rossum & Drake, 2009) and implemented in Snakemake (version 7.3.4) (Mölder et al., 2021).

After studying the behaviour of the samples using a PCA (Jollife & Cadima, 2016), controls were excluded, and replicates and duplicates were filtered according to the optimal number of reads (5-12 M). Then, we removed outliers and samples that could not be classified (due to incomplete clinical data), as well as genes with zero counts. Low expressed genes were filtered by low counts per million (CPM < 1) using the EdgeR package (Chen et al., 2016).

The remaining raw counts were normalized using Voom (Law et al., 2014) and quantile normalization (Hansen et al., 2012). Then, PCA was used to assess possible technical batch effects (e.g., from RNA extraction, NanoDrop concentration, 260/280 and 260/230 ratios, RIN, DV200, Qubit concentration, and sequencing) and demographic batch effects (i.e., recruiting clinic, patient age, BMI, ethnicity, allergies, drug or alcohol use and smoking habits) that could bias gene expression behaviour.

Finally, the limma package (Ritchie et al., 2015) was employed to remove the unwanted effects using linear models ***(Figure 17).*** This pipeline was implemented in R (version 4.0.5, 2021-03-31) (R Core Team, 2020) and plots were generated using the ggplot2 package (Wickham, 2011).

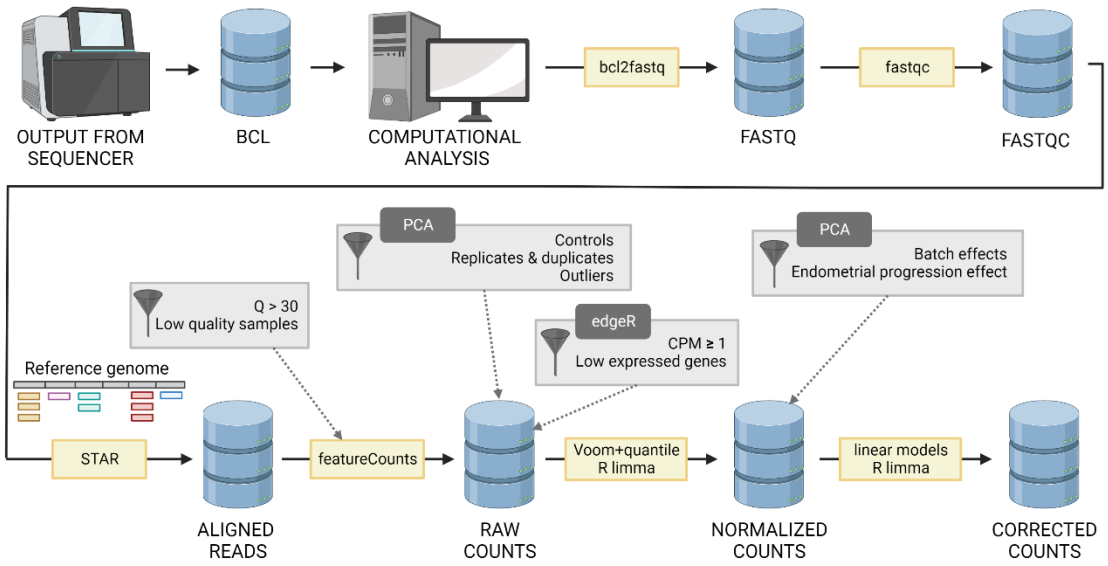


Figure 17. Preprocessing and exploratory analysis of transcriptomic data.
Bioinformatic workflow depicting the transformation of raw sequencing data into final transcript counts. The bioinformatic tools (yellow text boxes), filters (light grey text boxes) applied with the bioinformatics packages/analyses (dark grey text boxes), and type of data file obtained from each step (blue cylinders) of our transcriptomic analysis are shown. BCL, binary base call; CPM, counts per million; PCA, principal components analysis; Q, Phred quality score; QC, quality control; STAR, Spliced Transcripts Alignment to Reference. Created with BioRender.com (2022).

2.4. Removing the effects of endometrial progression

To classify normalized and corrected data according to the stage of the patient's endometrial progression, we used a transcriptomic predictor based on a signature of 73 endometrial timing genes, previously developed by our group (Diaz-Gimeno et al., 2021). Classified samples were represented using PCA (Jollife & Cadima, 2016), and the timing batch effect was removed using the linear models of the limma package (Ritchie et al., 2015). This strategy eliminates possible biases of menstrual cycle variations, to ensure that the observed transcriptomic behaviour of the samples reflects the endometrial pathology (Devesa-Peiro et al., 2021) ***(Figure 17).*** In addition, this endometrial classification was also employed to evaluate the menstrual cycle variations as possible

confounding variable amongst different groups of patients in the following analysis. This pipeline was implemented in R (version 4.0.5, 2021-03-31) (R Core Team, 2020) and plots were generated using the ggplot2 package (Wickham, 2011).

3. Patient stratification considering the endometrial pathological function

3.1. Developing a prediction model that distinguishes the pathological WOI

Normalized and corrected data were assigned to the training (80%) or test set (20%), with one RIF patient for every three controls (proportion 1:3). The main confounding variables of the training and test sets were compared, and their transcriptomic behaviour was analysed with PCA (Jollife & Cadima, 2016). Statistical differences were calculated employing Fisher´s test for qualitative variables (i.e., clinic, type of infertility, stage of endometrial progression and type of transfer cycle); student´s t-test for normally distributed (Shapiro's test > 0.05) or Wilcoxon´s test for non-normally distributed (Shapiro´s test < 0.05) quantitative variables (i.e., years of infertility, age, BMI, number of transfers and implantation failures). In all cases, p-values < 0.05 were considered statistically significant.

The training set was employed to identify the pathological WOI signature and develop a balanced probabilistic model that was internally-validated, while the test set was used to validate the robustness of this model externally. Briefly, the gene selection algorithm CorrelationAttributeEval (Witten et al., 2016) was used to evaluate the worth of each measured gene from the training set, by analysing the Pearson´s correlation coefficient

between each gene and the pathological condition. The resulting correlation scores (CS; ranging from 0 to 1) were then used to arrange all these genes in decreasing order. To determine the maximum number of genes that can be input into the model to provide the most accurate prediction, the SVM (Noble, 2006), kNN (Zhang, 2016) and RF (Breiman, 2001) algorithms were independently implemented using default parameters in the training set. In each case, stratified five-fold cross-validation (80:20; 100 times) was conducted using different sets of the ordered genes, increasing in size [i.e., 1-by-1 (0-300 first genes), 10-by-10 (301-500 first genes), 100-by-100 (501-1000 first genes), 200-by-200 (1001-2000 first genes), or 500-by-500 (for the total number of genes)]. The largest signature obtained from the three algorithms was considered as the potential biomarker signature of the pathological WOI for this study ***(Figure 18)***.

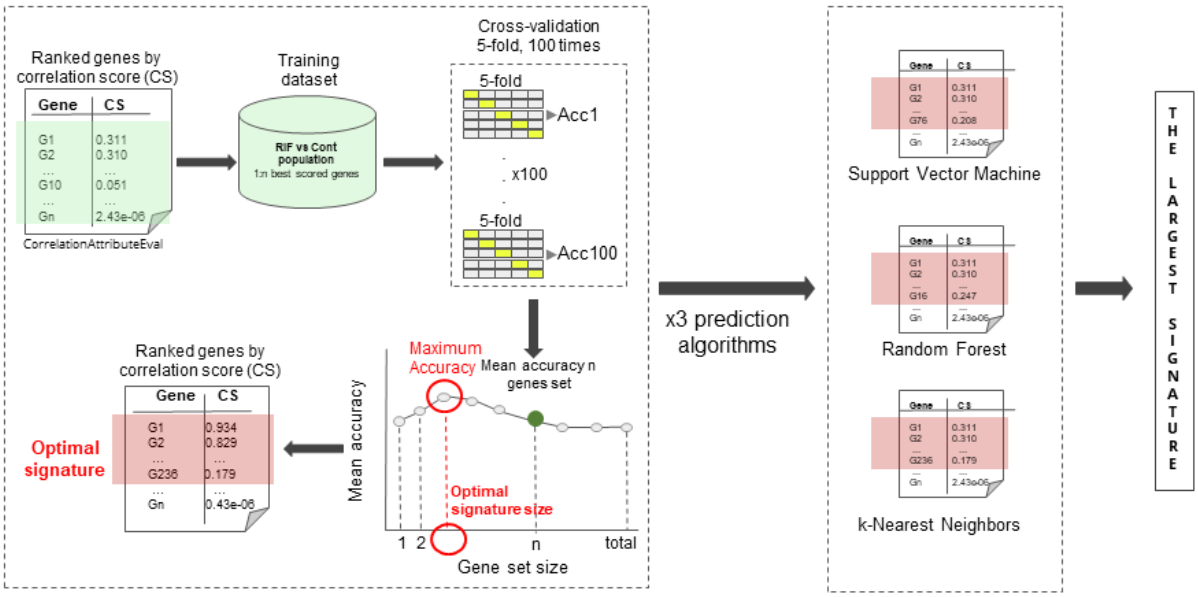


Figure 18. Workflow for establishing the pathological window of implantation gene signature. *The total genes from training set were ranked in decreasing order, according to their correlation with the pathological window of implantation (regardless endometrial progression). Their predictive power was evaluated using five-fold cross-validation (80:20, 100 times). The maximum accuracy value (Acc) was used to select the optimal signature size from each of the three different machine learning algorithms (i.e., support vector machine, random forest and k-nearest neighbors). The largest signature of all was selected for further analysis. Cont, control; n, unspecific number of; RIF, recurrent implantation failure.*

Due to the imbalanced class of our cohort (much more control than RIF samples), and the immensity and complexity of the entire endometrial transcriptomic gene set, it was important to undersample the majority class (controls) to avoid biases that would cause the algorithm to perform poorly (i.e., incorrectly identify the minor class of RIF patients with pathological WOI). To this end, we split our training set into 100 balanced subgroups, containing an equal number of RIF and random control samples. The three base algorithms (i.e., SVM, kNN and RF) were applied individually, or in pairs (i.e., SVM+kNN, SVM+RF and kNN+RF), using the pathological WOI signature established in the previous step. This balanced probabilistic model was applied to the test set for 100 iterations, to generate a prediction-probability of pathology for each sample. A mean prediction-probability threshold of 0.5 was then used to classify samples as RIF (≥ 0.5) or control (< 0.5).

To validate the prediction model and the selected signature externally, the accuracy, sensitivity and specificity were calculated independently for the test set in R (version 4.0.5, 2021-03-31) (R Core Team, 2020) ***(Figure 19)***.

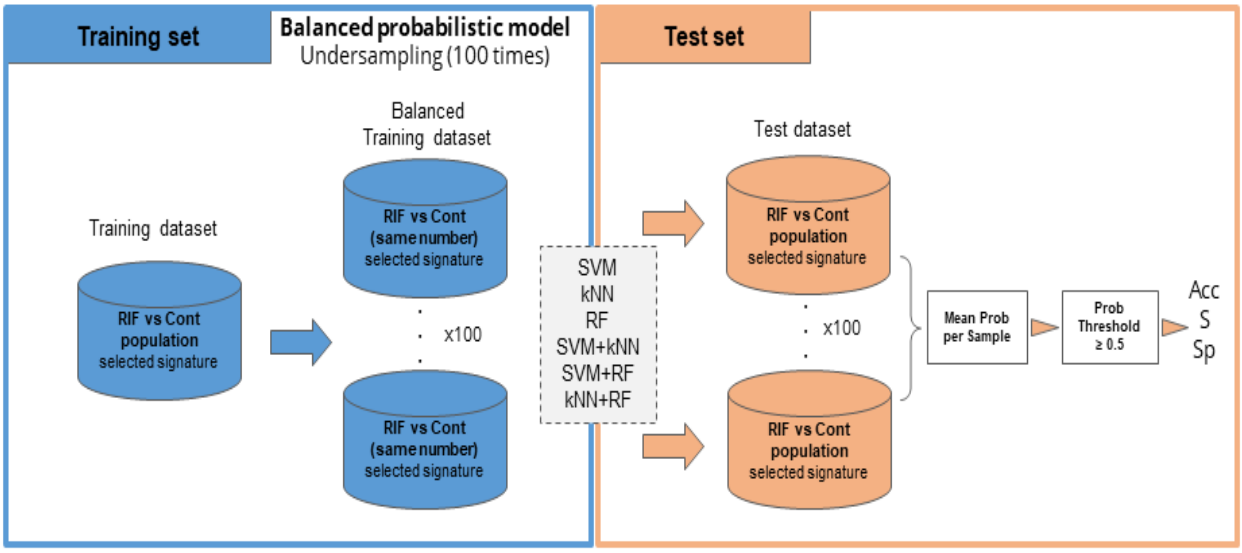


Figure 19. Development of a balanced probabilistic model for classifying endometrial pathology.
Selected gene signature and training set were employed for creating a total of 100 balanced datasets using different algorithms (grey square). In the balanced process based on undersampling, the same samples from the minority condition (RIF) were considered and the same number of samples from control was randomly selected. The robustness of the model was assessed externally with test set. Performance parameters were calculated considering the classification predicted according to the probability of pathology (Threshold ≥ 0.5). Acc, accuracy; Cont, control; kNN, k-nearest neighbors; n, number of; Prob, prediction-probability of pathology; RF, random forest; RIF, recurrent implantation failure; S, sensitivity; Sp, specificity; SVM, support vector machine.

Finally, the algorithm with the best overall performance was selected as the optimal prediction model of the pathological WOI. Additionally, the best prediction model was internally-evaluated through a five-fold cross-validation with a stratified ten-fold split using the training set in order to calculate performance parameters avoiding overfitting ***(Figure 20).***

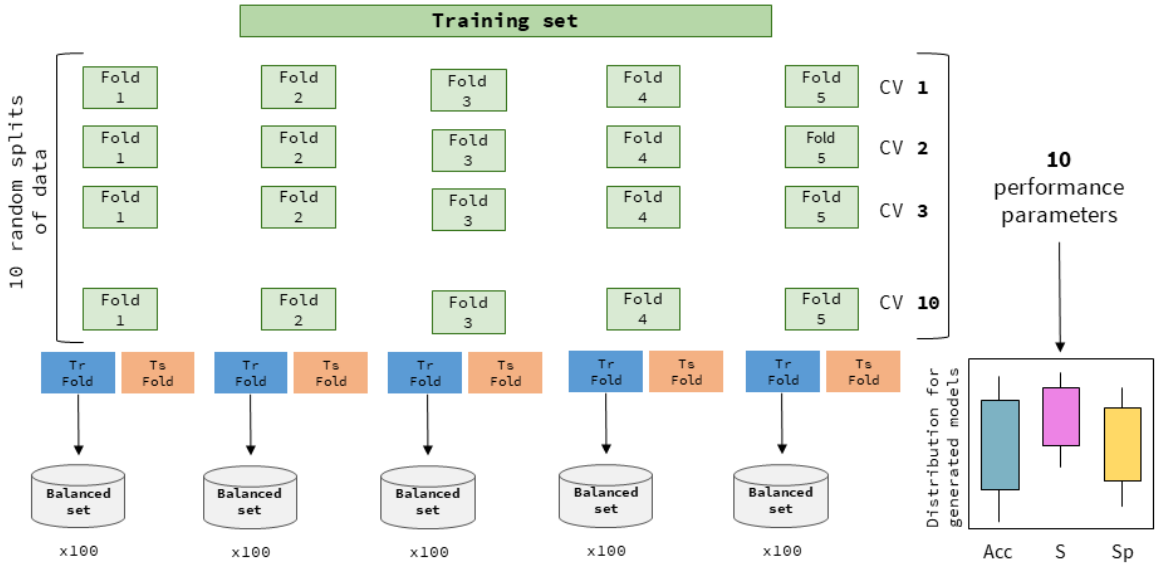


Figure 20. Five-fold cross-validation process applied to training set for the internal validation.
Each of the five folds was subdivided into training (Tr Fold) and test (Ts fold) in a cross-validation (CV) process using the training set. Imbalanced effect observed in pathological condition with respect to control condition was corrected creating 100 balanced datasets randomly from Tr Fold. Distribution of different performance parameters (Acc, accuracy; S, sensitivity; Sp, specificity) were calculated considering the different models generated in the 10 splits.

Therefore, the external validation is necessary to optimize the prediction model and the selected signature in a set of samples independent from the set employed for their selection. Meanwhile, the internal validation is important to assess the prediction performance of the model and the signature in the set of reference without bias. The modifications to the base algorithms were all implemented through Weka (Witten et al., 2016) and RWeka (Hornik et al., 2008). All plots were generated using the ggplot2 package (Wickham, 2011).

3.2. Transcriptomic stratification and clinical follow-up

Samples from the training and test sets were classified as control (c) or RIF due to a pathological WOI (p) according to the optimal balanced probabilistic model. The mean prediction-probabilities of pathology generated by our model allowed the further stratification of our samples into c1 (≤ 0.2), c2 (0.2-0.5), p2 [0.5-0.8), or p1 (≥ 0.8) groups ***(Figure 21).***

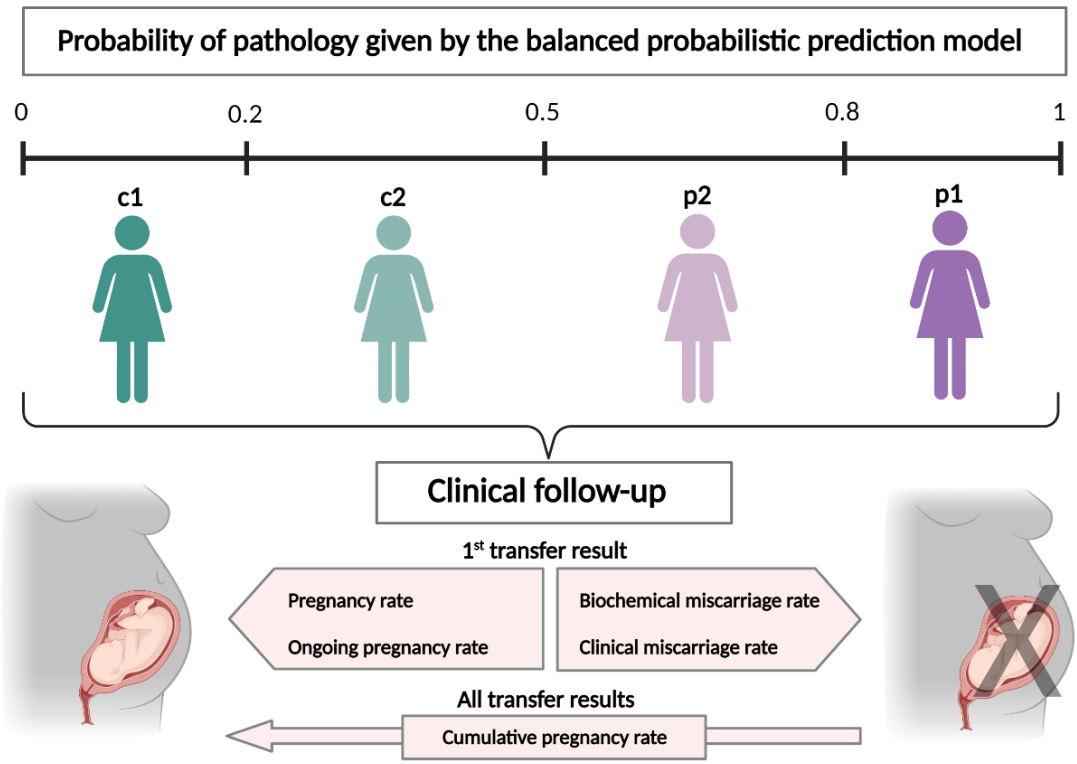


Figure 21. Transcriptomic stratification and clinical follow-up.
(Top) Patients were stratified into four transcriptomically-defined groups (c1 or c2 for controls; p2 or p1 for pathological RIF) according to the mean prediction-probabilities generated by our balanced probabilistic prediction model, that used a threshold of 0.5 to distinguish between control and pathological samples. (Bottom) A follow-up of the patients' reproductive outcomes was conducted to identify the trends associations with each transcriptomic profile. Created with BioRender.com (2022).

The homogeneity across these groups was evaluated, taking into account confounding variables, to ensure the clinical differences observed were indeed associated with the transcriptomic profile and not biased towards the patients' baseline characteristics. Statistical differences were calculated employing Fisher´s test for qualitative variables (i.e., clinic, type of infertility, stage of endometrial progression and type of transfer cycle); student´s t-test for normally distributed (Shapiro's test > 0.05) or Wilcoxon´s test for non-normally distributed (Shapiro´s test < 0.05) quantitative variables (i.e., years of infertility, age, BMI, number of transfers and implantation failures). In all cases, p-values < 0.05 were considered statistically significant.

A clinical follow-up of the reproductive outcomes (following the first transfer after the biopsy collection) of the patients was conducted to identify the trends associated with the four transcriptomic profiles. The pregnancy rate (PR) was defined as the proportion of clinically-confirmed pregnancies. The ongoing pregnancy rate (OPR) was defined as the number of ongoing pregnancies divided by the total number of clinically-confirmed pregnancies. Additionally, the cumulative pregnancy rate (cumulative PR) was defined as the total number of transfers the patient needed to successfully achieve pregnancy and was calculated considering all the embryo transfers. The clinical miscarriage rate (CMR) reflected the losses of ultrasound-detected pregnancies prior to 20-24 weeks of gestation, while the biochemical miscarriage rate (BMR) was calculated with the number of biochemical pregnancies (detected by positive serum β-hCG values, without visualization

of the gestational sac in the first 10 weeks of gestation), over the total number of pregnancies. Statistical differences between these reproductive outcomes across the groups (i.e., c1, c2, p1 and p2) were calculated using Fisher´s tests in R (version 4.0.5, 2021-03-31) (R Core Team, 2020) and plots were generated using the ggplot2 package (Wickham, 2011).

3.3. Molecular study of the stratified patients

To identify the DEGs in each of our transcriptomically-defined groups, we applied a DEA, calculating statistical differences using the student´s t-test. We then performed a GSEA (considering the FC from DEA), using ClusterProfiler (Yu et al., 2012), which queried KEGG (Kanehisa & Goto, 2000) and GO (Ashburner et al., 2000) to identify affected functions ***(Figure 22)***. The adjusted p-value for each gene/function was corrected by the FDR (Benjamini & Hochberg, 1995), and only significant genes/functions (FDR < 0.05) were selected for further analysis.

Notably, for KEGG enrichment (version Sept-2021), annotated pathways were filtered by those with less than 1000 associated genes, and unrelated to diseases or drugs. Regarding GO enrichment (version Dec-2021), the three ontologies were used (i.e., biological processes, molecular functions and cellular components), however, only experimentally-evidenced GO-gene associations were included, and annotated GO terms were filtered by stablishing associated genes (5-300). Finally, functional concordance between the groups was calculated using Cohen´s Kappa index (Cohen, 1960). All these steps were carried out in R (version 4.0.5, 2021-03-31) (R Core Team, 2020) and plots were generated using the ggplot2 package (Wickham, 2011).

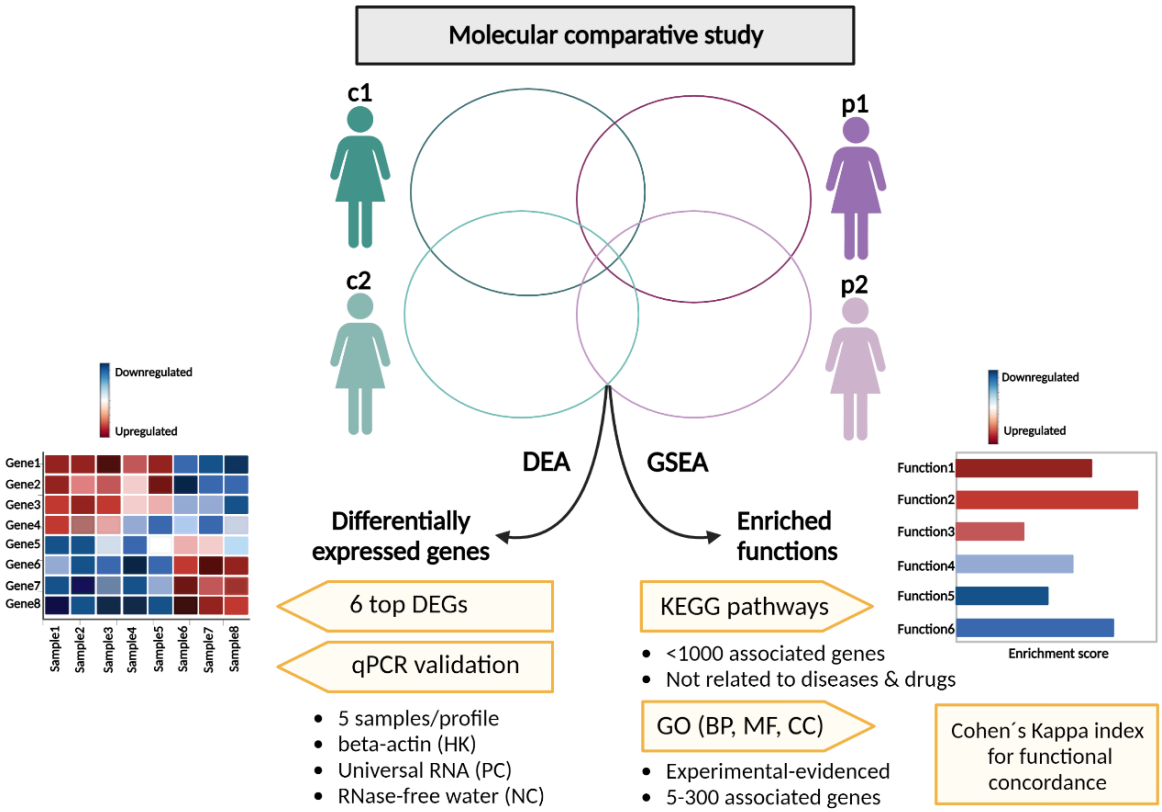


Figure 22. Molecular study of the transcriptomic profiles and validation of potential biomarkers.
BP, biological processes; c1, control group 1; c2, control group 2; CC, cellular components; DEA, differential expression analysis; DEGs, differentially expressed genes; GO, Gene Ontology; GSEA, gene set enrichment analysis; HK, housekeeping gene; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular functions; NC, negative control; p1, pathological group 1; p2, pathological group 2; PC, positive control; qPCR, quantitative polymerase chain reaction. Created with BioRender.com (2022).

3.4. Experimental validation of potential biomarkers

The expression of six potential biomarkers (selected among the DEGs between the transcriptomic groups) was evaluated with quantitative PCR (qPCR) ***(Figure 22).*** Specific primers (Invitrogen, Thermo Fisher Scientific MA, USA) designed with Primer-BLAST (Ye et al., 2012) were employed to measure gene expression in twenty of the RNA samples (five from each of the four transcriptomic profiles). Specific primer sequences are detailed in ***Supplemental table 1***.

Complementary DNA was synthesized using the PrimeScript Reagent Kit (Perfect Real Time, Takara, Shiga, Japan) in a Thermocycler T3000 (Biometra, Dublin, Ireland). Then, qPCR was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems, CA, USA) using Power-Up SYBR Green (Thermo Fisher Scientific, MA, USA) and beta-actin (*ACTB*) as a housekeeping gene. Primers were tested using a universal human RNA (Agilent, Valencia, Spain) as a positive control, and RNase-free water (Qiagen, Hilden, Germany) as a negative control, with each sample analysed in duplicate.

Finally, relative gene expression was calculated using the ΔΔCt method (Schmittgen & Livak, 2008). Possible batch effects related to the experimental procedure and endometrial timing classification were evaluated using an ANOVA test. Gene expression tendencies from the qPCR were compared with those we previously obtained with RNA-Seq. Calculations and comparisons were performed in R (version 4.0.5, 2021-03-31) (R Core Team, 2020) and plots were generated using the ggplot2 package (Wickham, 2011).

VI. RESULTS & DISCUSSION

"Science is always worth it because its discoveries, sooner or later, are always applied."

Severo Ochoa

RESULTS & DISCUSSION

1. Transcriptomic data from the subfertile patient population

1.1. Quality of RNA samples

Of the 291 endometrial biopsies collected (one per patient), 276 samples were received with sufficient tissue for total RNA extraction with the miRNeasy Mini Kit® (> 30 mg); 15 were received with insufficient tissue (< 30 mg) and thus could not be included in the study. Rigorous assessment of RNA quality then excluded 41/276 (14.85%) samples, and 17% of the remaining samples (40/235) had incomplete clinical follow-up for their classification. In the context of this study, the high sample loss (32.99%; 96/291) was partially due to the limited tissue remaining for research after the patient's clinical endometrial evaluation, and the requirement to have available clinical data of failed implantation(s) following a minimum of embryo transfers (hindering the initial clinical classification as RIF or control).

Most of the 195 samples that were sequenced had adequate quality (260/280 ratio ~2, 260/230 ratio between 1.8-2.2, RIN ≥ 3 and DV200 ≥ 70%), with the exception of two samples with a DV200 ~30% (that were included for having a RIN ≥ 3), and 29 samples with a 260/230 ratio below 1.8 (that were included for having the remaining parameters suitable). Specifically, the median of their 260/280 ratio was 2.09 ***(Figure 23A)***; 260/230 ratio was 1.98 ***(Figure 23B);*** RIN was 5.3 ***(Figure 23C)***; and DV200 was 83.68 ***(Figure 23D).*** Despite these moderate RIN values, the samples met the remaining quality parameters, and thus we selected the AmpliSeq for Illumina® Transcriptome Human Gene Expression Panel, which targets the whole transcriptome, and is compatible with low-

and high-quality RNA samples from human tissue, to optimize RNA sequencing. Notably, unusual macroscopic appearance of the samples did not appear to be related to their RNA quality.

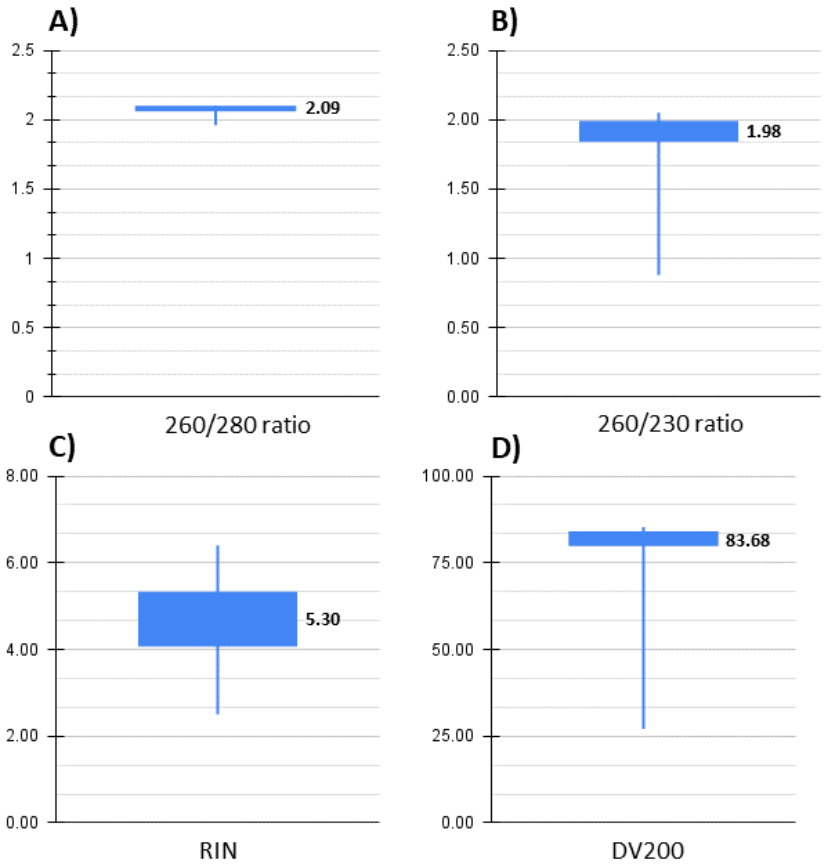


Figure 23. Distribution of quality parameters from RNA samples selected for sequencing. *Boxplots for the distribution of the 260/280 ratios **(A)**, 260/230 ratios **(B)**, RIN values **(C)** and DV200 values **(D)** from the 195 RNA samples selected for sequencing process. The median of each distribution is indicated in bold, to the right of the corresponding boxplot. DV200, RNA fragments with more than 200 nucleotides; RIN, ribonucleic acid integrity number.*

1.2. Validation of RNA-Sequencing performance

The RNA-Seq libraries were generated for the 195 eligible RNA samples (with 202-267 bp fragments, as we expected) in six batches ***(Table 5)***, with an overall favourable performance ***(Table 6)***. While the pilot study and batch 1 exhibited optimal performance

in all parameters, the estimated yield of run 5 was slightly lower than the minimum optimal value (45.9 vs. 50 Gb, respectively). The cluster density was low (≤ 160 K/mm^2) for batches 2-5, however the elevated cluster PFs we obtained in all batches indicated adequate signal of cluster occupancy. Further, underclustering is preferable to overclustering, since underclustering maintains high data quality despite the slight reduction in transcriptomic data output, and overclustering can hinder run performance, reduce the Q30 and total data output, as well as possibly introduce artefacts (Illumina, 2022d). Based on these considerations and premises, we reasoned that the transcriptomic data we generated was of high enough quality to continue with the study.

Table 5. RNA sequencing batches.

Batch	Location	Unique samples	Replicates	Controls	Duplicates	Justification for duplication	Samples sequenced
Pilot (0)	Spain	34	B11 + V29	RNase-free water + Universal RNA	0	-	40
1	Oxford	29	B11 + V29		6	> 12 M or < 5 M reads probably due to prolonged exposure to ethanol in the pilot study	40
2		35	B11 + V29		0	-	40
3		34	B11 + V37		0	-	40
4		27	B11 + V37		5	< 5 M reads in former batches probably derived from RNA degradation	37
5	Spain	36	B11		0	-	40
Total	**-**	**195**	**19**	**12**	**11**	**-**	**237**

RNA sequencing of the eligible 195 samples was conducted in six sequencing batches (0-5). The sequencing location, number of unique samples included, anonymized codes (i.e., B11, V29 and V37) of the two technical replicates employed, negative and positive controls, number of samples whose sequencing was duplicated (along with the justification), and final number of samples sequenced. M, million; *RNA, ribonucleic acid.*

Table 6. Performance parameters of sequencing systems.

Batch	Loading concentration (pM)	Cluster density (K/mm^2)	Clusters PF (%)	Estimated Yield (Gb)	Data > Q30 (%)
Optimal values	1.1-1.9	170-220	≥ 90	50-60	> 80
Pilot (0)	1.5	186	92.3	71.3	94.5
1	1.6	178	93.4	71	95.2
2	1.8	140	95.7	56.9	96.3
3	1.9	160	94.8	64	95.8
4	1.9	140	95.8	55.8	96.8
5	1.6	114	94.9	45.9	95.9

The optimal performance parameters of the RNA-Sequencing protocol we employed are indicated as a reference for the values we obtained with each sequencing batch (0-5). Gb, gigabase; K/mm^2, kilocluster per square millimeter; PF, passing filter; pM, pico molar; Q30, Phred quality score.

1.3. Behaviour of transcriptomic data

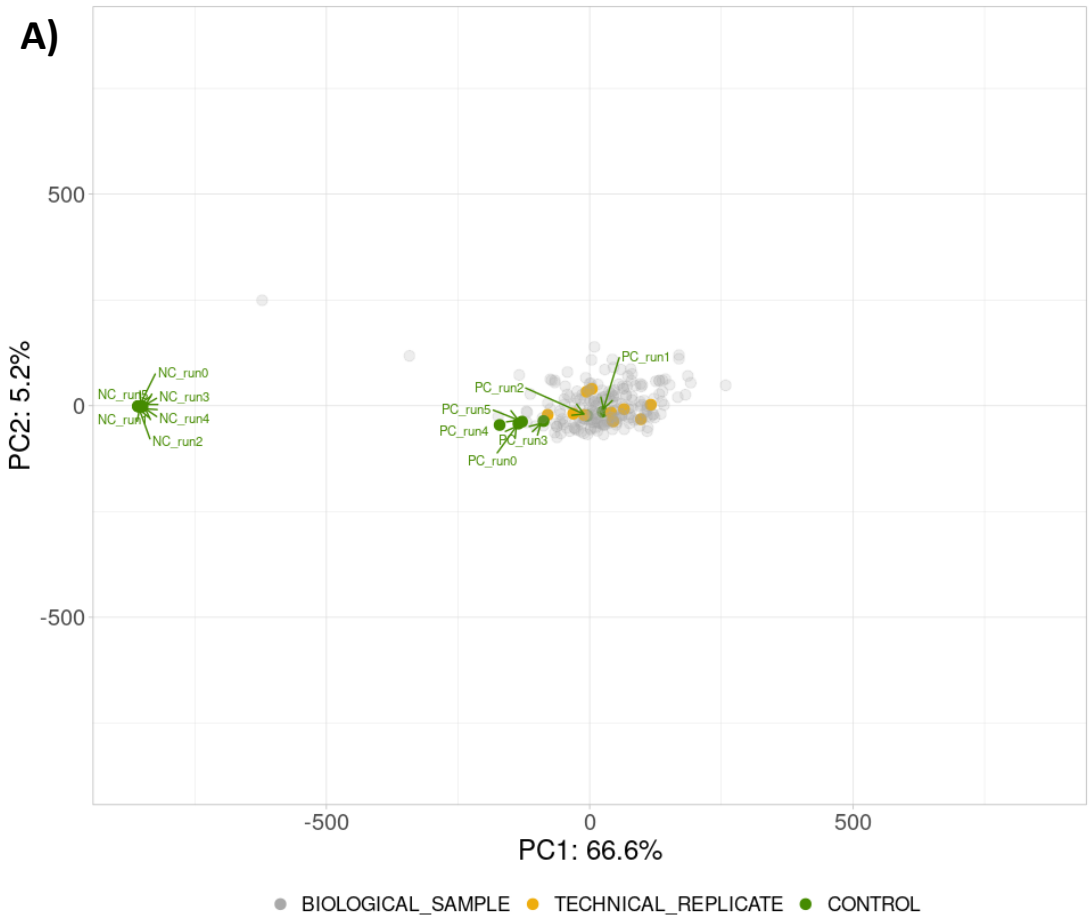
Most of our samples (72.56%) produced 5-12 M reads, as expected. Specifically, 14.36% produced < 5 M, 40.08% produced [5-8) M, 22.36% produced [8-10) M, 10.12% produced [10-12) M and 13.08% produced ≥ 12 M. The Illumnina® RNA-Seq protocol we employed was designed to obtain 10 M reads/sample, however variation was anticipated due to the heterogeneous transcriptomic character of the samples. By filtering out the controls as well as replicate and 11 duplicate samples without 5-12 M reads, the range of reads became 8.40% with < 5 M, 47.33% with [5-8) M, 19.84% with [8-10) M, 13.74% with [10-12) M and 10.69% with ≥ 12 M. Notably, this process did not affect the downstream transcriptomic analysis, as we normalized and corrected the data to make them comparable and look for their differences at the transcriptomic level.

Regarding transcriptomic behaviour in principal component analysis (PCA), negative and positive controls exhibited distinguishable patterns, with positive controls close to biological samples, as we expected ***(Figure 24A)***. Notably, technical replicates displayed

similar transcriptomic behaviour, demonstrating the reproducibility of our techniques ***(Figure 24B)***. Among the 195 unique samples (after removing controls, replicates and duplicated), two were considered outliers because they behaved differently than the rest of the samples ***(Figure 25),*** and were removed. Notably, these samples had a DV200 ~30% but were included for having a RIN ≥ 3. This fact verifies that DV200 is a key parameter to measure RNA quality considering these kinds of RNA-Seq protocols. Finally, 62 samples were excluded for the patients dropping out of the study, having incomplete clinical data or insufficient embryo transfers for clinical classification, leaving 131 samples for subsequent analysis. On the other hand, out of the 20,802 genes measured by our RNA-Seq panel, 656 were removed for having zero counts, and 5,472 genes were excluded for having CPM < 1. Finally, 14,674 genes were included in subsequent analysis, which supports the estimates of there being 14,000 genes expressed in the human endometrium (The Human Protein Atlas, 2022).

After the normalization process, no batch effects from the RNA extraction batch, NanoDrop concentration, 260/280 or 260/230 ratios, RIN, DV200 or Qubit concentration, were visually detected after testing the experimental variables in our patients ***(Figure 26A-G)***. The samples from run 1 and run 5 unexpectedly clustered ***(Figure 26H),*** and consequently, this sequencing run batch effect was corrected prior to subsequent analysis.

We found no demographic-related batch effects (i.e., recruiting clinic; patient age at biopsy collection, BMI, ethnicity, allergies; use of tobacco, alcohol, or other drugs) among our samples ***(Figure 27)***. Finally, an effect of endometrial progression was also identified with PCA, using the transcriptomic endometrial dating (TED) classification, based on 72 out of 73 timing genes on which this predictor is based on ***(Figure 28A),*** and it was removed with success ***(Figure 28B***).


A)
500
0
-500
PC2: 5.2%
NC_run0
NC_run5
NC_run3
NC_run1
NC_run4
NC_run2
PC_run1
PC_run2
PC_run5
PC_run4
PC_run3
PC_run0
-500
0
500
PC1: 66.6%
BIOLOGICAL_SAMPLE
TECHNICAL_REPLICATE
CONTROL

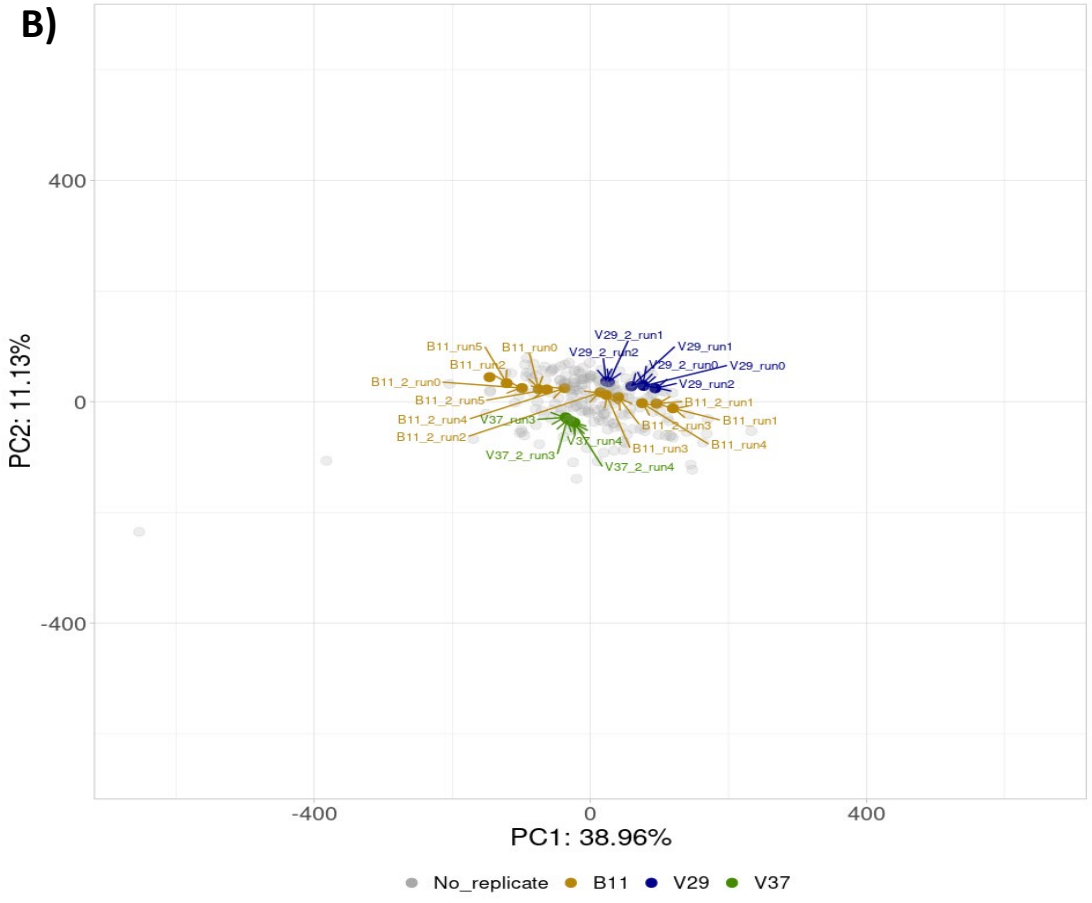

B)
400
0
-400
PC2: 11.13%
B11_run5
B11_run0
B11_run1
B11_2_run0
B11_2_run5
B11_2_run4
B11_2_run2
V37_run3
V37_run4
V37_2_run3
V37_2_run4
B11_run3
B11_2_run3
V29_2_run1
V29_2_run2
V29_run1
V29_2_run0
V29_run0
V29_run2
B11_2_run1
B11_run1
B11_run4
-400
0
400
PC1: 38.96%
No_replicate
B11
V29
V37

Figure 24. Transcriptomic behaviour of control and technical replicates employed in the RNA-Sequencing protocol.

(A) *Principal component analysis (PCA) plot including all 20,802 genes detected by the panel and all 237 samples (including duplicated, replicates and controls). Negative controls (NC) and positive controls (PC) are indicated by green dots, technical replicates are indicated by yellow dots, and remaining biological samples are represented by the grey dots in the background.* ***(B)*** *PCA plot after removing controls. The technical replicates are distinguished by coloured dots (B11 in yellow, V29 in blue and V37 in green) while the remaining biological samples were indicated by the grey dots in the background. In both plots, the X axis represents the variability (%) of the data in the first component (PC1) and the Y axis represents the variability of the second component (PC2).*

Figure 25. Identification of outliers considering sample behaviour.

Principal component analysis (PCA) plot including the 195 unique samples. Samples arerepresented as coloured points according to the sequencing batches (labelled run 0-5). Two outliers (V10 and BI16; outlined in the red circle) were removed prior to subsequent analysis. The X axis represents the variability (%) of the data in the first component (PC1) and the Y axis represents the variability of the second component (PC2).

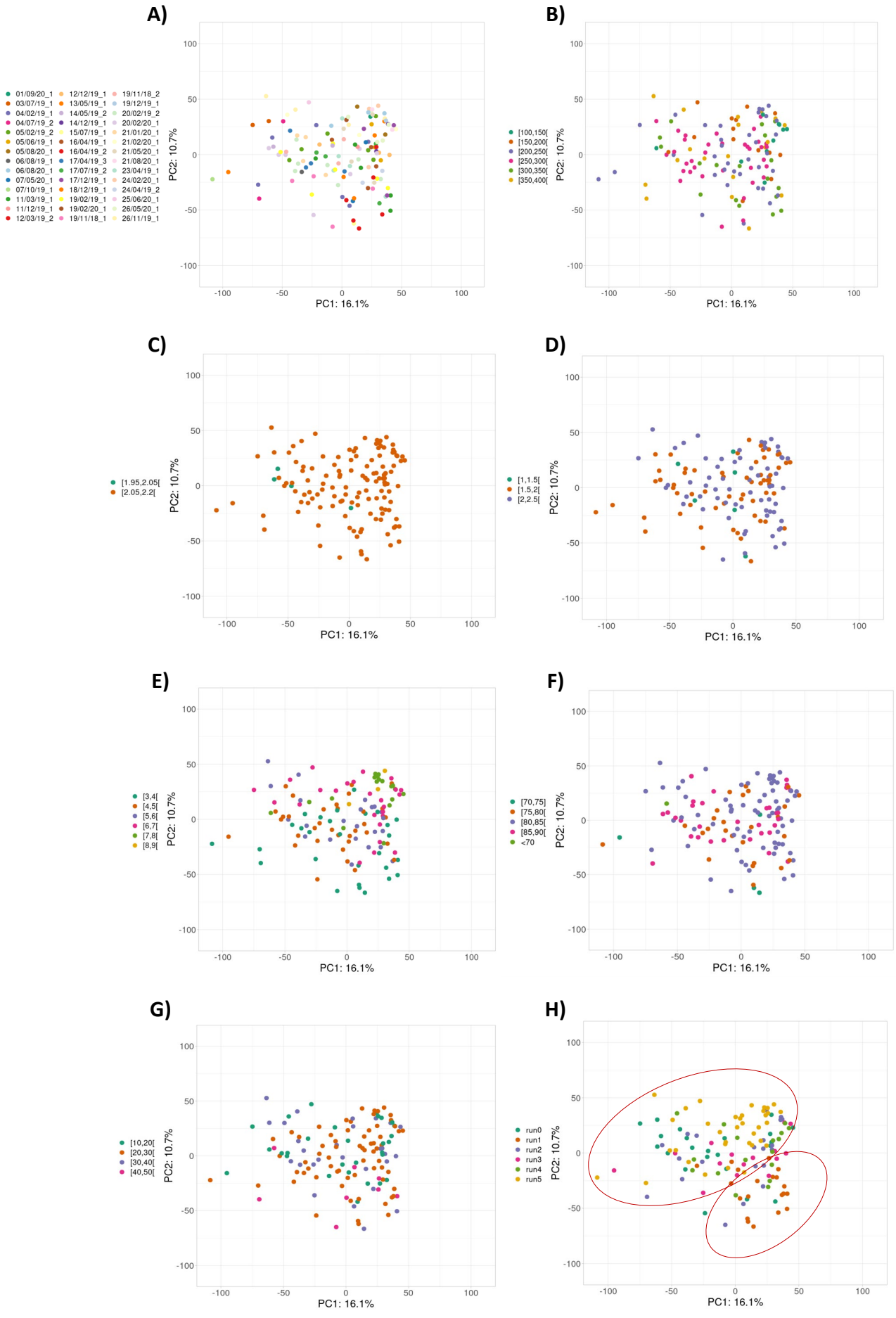

A)
01/09/20_1
03/07/19_1
04/02/19_1
04/07/19_2
05/02/19_2
05/06/19_1
05/08/20_1
06/08/19_1
06/08/20_1
07/05/20_1
07/10/19_1
11/03/19_1
11/12/19_1
12/03/19_2
12/12/19_1
13/05/19_1
14/05/19_2
14/12/19_1
15/07/19_1
16/04/19_1
16/04/19_2
17/04/19_3
17/07/19_2
17/12/19_1
18/12/19_1
19/02/19_1
19/02/20_1
19/11/18_1
19/11/18_2
19/12/19_1
20/02/19_2
20/02/20_1
21/01/20_1
21/02/20_1
21/05/20_1
21/08/20_1
23/04/19_1
24/02/20_1
24/04/19_2
25/06/20_1
26/05/20_1
26/11/19_1
PC2: 10.7%
PC1: 16.1%
B)
[100,150[
[150,200[
[200,250[
[250,300[
[300,350[
[350,400[
PC2: 10.7%
PC1: 16.1%
C)
[1.95,2.05[
[2.05,2.2[
PC2: 10.7%
PC1: 16.1%
D)
[1,1.5[
[1.5,2[
[2,2.5[
PC2: 10.7%
PC1: 16.1%
E)
[3,4[
[4,5[
[5,6[
[6,7[
[7,8[
[8,9[
PC2: 10.7%
PC1: 16.1%
F)
[70,75]
[75,80[
[80,85[
[85,90[
<70
PC2: 10.7%
PC1: 16.1%
G)
[10,20[
[20,30[
[30,40[
[40,50[
PC2: 10.7%
PC1: 16.1%
H)
run0
run1
run2
run3
run4
run5
PC2: 10.7%
PC1: 16.1%

Figure 26. Evaluation of experimental batch effects with regards to sample behaviour.
Principal component analysis (PCA) plots evaluating eight potential experimental batch effects using the transcriptomic behaviour of the final 131 samples selected for the study. ***(A)*** *Ribonucleic acid (RNA) extraction batch. The colour of the dots represents the 48 dates when the RNA was extracted from the samples.* ***(B)*** *NanoDrop concentration. The colour of the dots represents a range of RNA concentration (ng/uL), where blue indicates [100-150); orange, [150-200); purple, [200-250); pink, [250-300); green, [300-350); yellow, [350-400).* ***(C)*** *260/280 ratio. Each colour represents a range of values, where blue indicates [1.95-2.05) and orange indicates [2.05-2.2).* ***(D)*** *260/230 ratio. Each colour represents a range of values, where blue indicates [1-1.5); orange, [1.5-2); purple [2-2.5).* ***(E)*** *RNA integrity number (RIN). Each colour represents of a range of values, where blue indicates [3-4); orange, [4-5); purple, [5-6); pink, [6-7); green, [7-8); yellow, [8-9).* ***(F)*** *RNA fragments larger than 200 nucleotides (DV200). Each colour represents a range of values (%), where blue indicates [70-75); orange, [75-80); purple, [80-85); pink, [85-90); green ~70.* ***(G)*** *Qubit concentration. Each colour represents a range of concentration (ng/uL), where blue indicates [10-20); orange, [20-30); purple, [30-40); pink, [40-50).* ***(H)*** *Sequencing run. The colour of the dots reflects the six sequencing runs, where blue indicates Run0; orange, Run1; purple, Run2; pink, Run3; green, Run4; yellow, Run5. Notably, the batch effect found with sequencing run1 and run5 (outlined with a red circle) was corrected prior to subsequent analysis. In all plots, the X axis represents the variability (%) of data in the first component (PC1) and the Y axis represents the variability of the second component (PC2).*

Taken together, these findings confirmed that our methodology produced sufficient quality data, with a non-biased transcriptomic behaviour, to continue with the objectives of this thesis. Building on the findings of Koot et al. (Koot et al., 2016), who proposed the only RIF prediction model (independent of menstrual cycle effect) that exists to date, we employed Illumina® RNA-Seq technology instead of two-channel Agilent® microarrays, improving reproducibility of gene expression detection (Wang et al., 2009) and increasing the abundance of high-quality transcriptomic data. Concretely, we analysed 131 samples for 14,674 filtered genes, while Koot et al. included 115 patients and 12,198 genes. We additionally applied a more reliable methodology for removing the endometrial progression effect when classifying patients into transcriptomically-defined groups according to menstrual cycle variations, using an algorithm previously designed by our research group (Diaz-Gimeno et al., 2021). Alternatively, Koot et al. employed a classification based on the quantity of luteinizing hormone (LH) measured in urine (to determine the days since ovulation), which is imprecise (Cano & Aliaga, 1995; Direito et al., 2013) as it cannot distinguish the molecular variability in the endometrium of women.

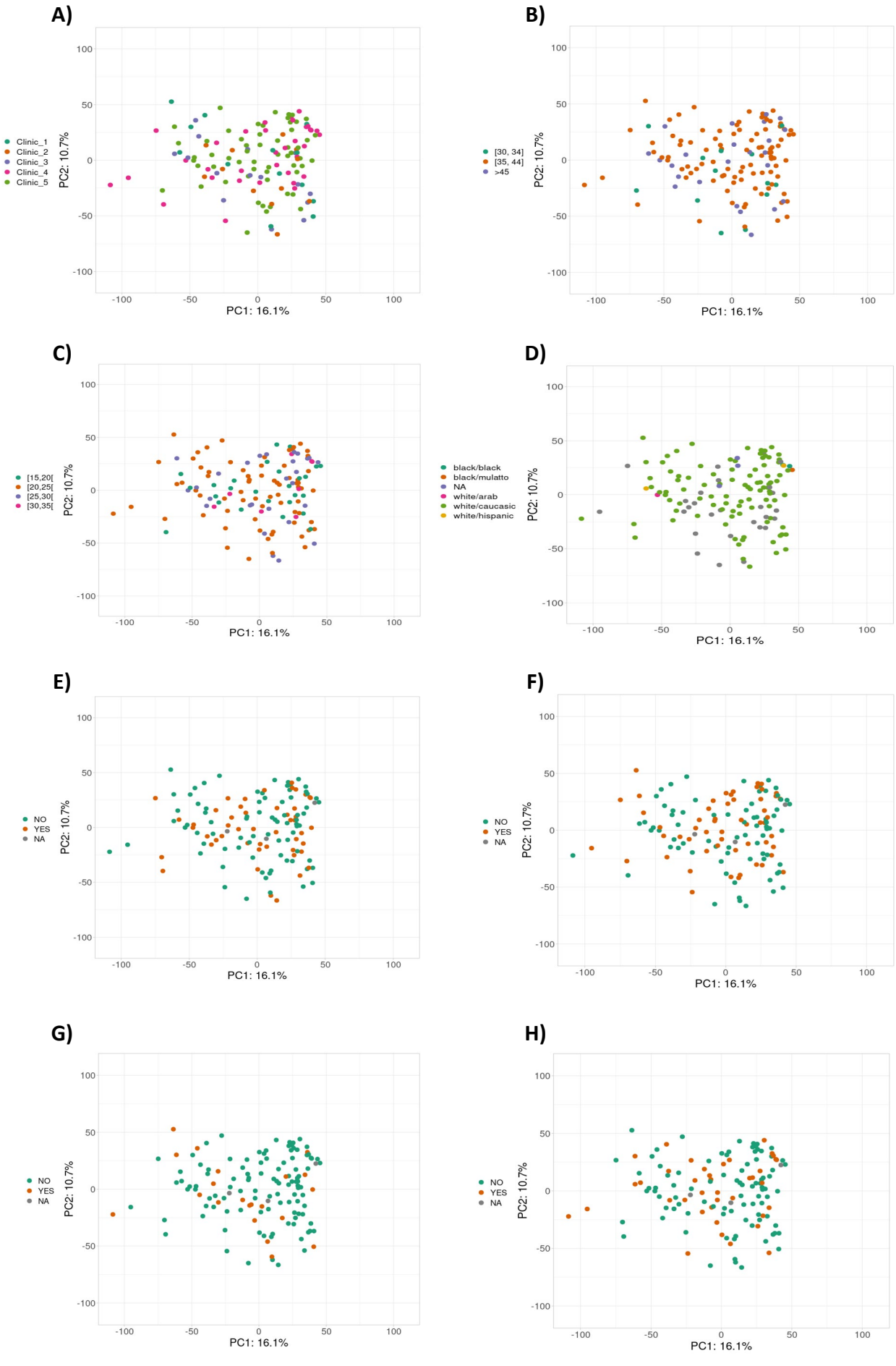

A)
Clinic_1
Clinic_2
Clinic_3
Clinic_4
Clinic_5
B)
[30, 34]
[35, 44]
>45
C)
[15,20[
[20,25[
[25,30[
[30,35[
D)
black/black
black/mulatto
NA
white/arab
white/caucasic
white/hispanic
E)
NO
YES
NA
F)
NO
YES
NA
G)
NO
YES
NA
H)
NO
YES
NA
PC1: 16.1%
PC2: 10.7%

Figure 27. Evaluation of demographic batch effects based on sample behaviour.
Principal component analysis (PCA) plots of the 131 selected samples evaluating potential demographic batch effects. ***(A)*** *Recruiting clinic. Clinics 1-5 are represented by the different coloured dots (blue, orange, purple, pink and green, respectively).* ***(B)*** *Patient´s age when the biopsy was collected. Considering that patients ≥ 35 years old have an advanced maternal age, each colour represents an age range, where blue indicates [30-34]; orange [35-44]; purple ≥ 45.* ***(C)*** *Patient´s body mass index (BMI). The colour of the dots represents ranges of BMI, where blue indicates [15-20); orange, [20-25); purple, [25-30); and pink reflects some obese patients (7) with BMI [30-35) that were not removed due to the lack of batch effects and the limited sample size used in this study.* ***(D)*** *Patient ethnicity. The colour of the dots represents different ethnicities, where blue indicates patients of African descent; orange, Mulatto; pink, Arab; green, Caucasian; yellow, Hispanic; and purple indicates those that were not available (NA). For patient allergies* ***(E)****, use of other drugs* ***(F)****, tobacco* ***(G)*** *or alcohol* ***(H)*** *consumption, blue dots represent no, orange indicate yes and purple dots indicate the data was not available. In all plots, the X axis represents the variability (%) of data in the first component (PC1) and the Y axis represents the variability of the second component (PC2).*

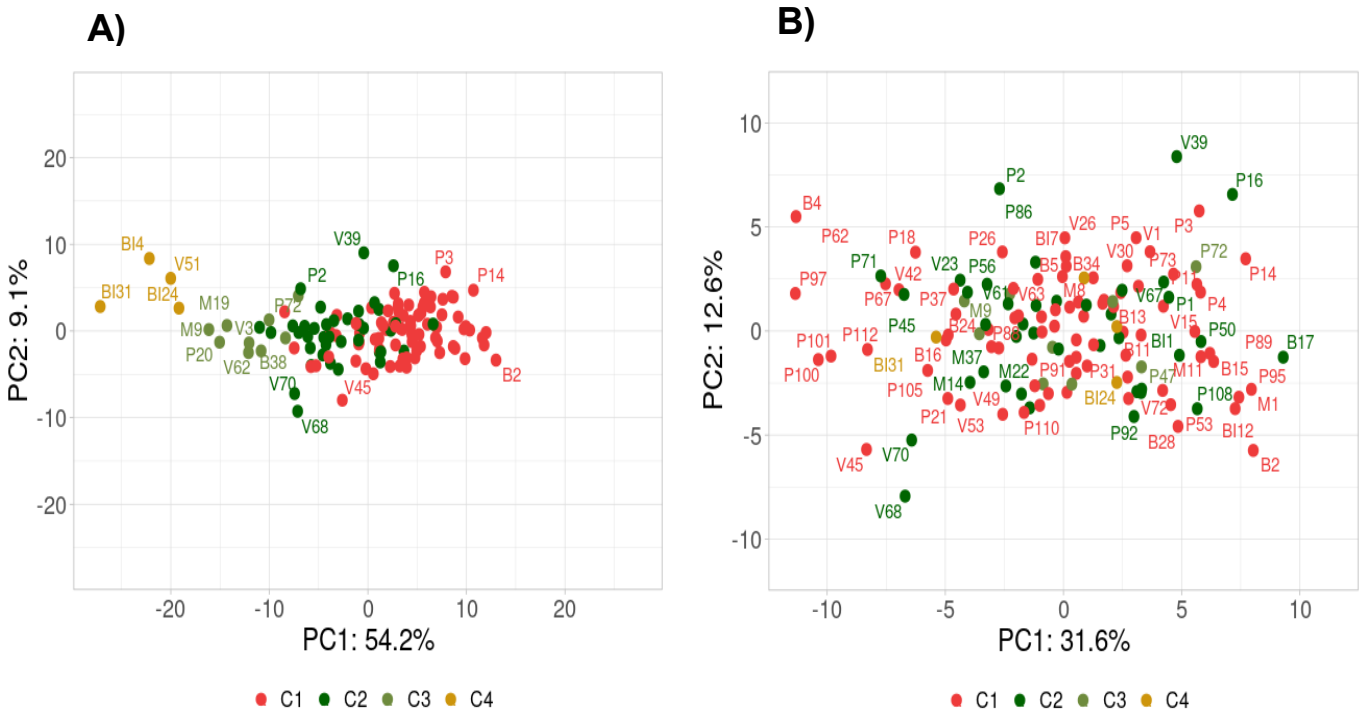


Figure 28. Correction of transcriptomic variations due to endometrial progression effect to focus the study on the pathology.
Principal component analysis (PCA) plots of the 131 selected samples sequenced with RNA-Seq, prior to ***(A)*** *and after* ***(B)*** *removing the endometrial timing effect using the linear models after classifying samples with the transcriptomic endometrial dating (TED) tool recently developed by our group (Diaz-Gimeno et al., 2021). Notably, 72/73 genes included in TED were found in our dataset. Samples are represented as coloured dots, based on their endometrial progression classification, where red indicates pre-receptive (C1); dark green, receptive 1 (C2); light green, receptive 2 (C3); yellow, post-secretory (C4). In both plots, the X axis represents the variability (%) of the data in the first component (PC1) and the Y axis represents the variability of the second component (PC2).*

2. Clinical characterization of the subfertile patient population

2.1. Clinical classification

The 131 samples selected for analysis were clinically classified as RIF (n = 32) or control (n = 99), depending on if the patients respectively had ≥ 3 or < 3 implantation failures following transfers with good-quality embryos ***(Table 7)***. Due to the uneven number of control and RIF patients (which highlighted RIF as a minority condition), we balanced the sample size of the two groups to avoid AI predictions biasing the controls over the pathology (Blagus & Lusa, 2015). Further, since our classification was based on the clinical history of patients before and after biopsy collection [unlike Koot´s study (Koot et al., 2016), that only considered the history prior to biopsy collection], our balanced prediction model was founded on more complete clinical profiles, providing realistic results.

Table 7. Clinical classification of study population.

Clinical Classification		Total	
CONTROL	**True control**	50	**99**
	Heterogeneous control	49	
RIF	**True RIF**	19	**32**
	Heterogeneous RIF	13	
Total		**131**	

Participating patients were broadly classified as control and recurrent implantation failure (RIF), based on their clinical history of implantation failures. The control group was subdivided into true control (patients that achieved implantation after the first transfer) or heterogeneous control (patients that achieved implantation after the second or third transfer). Alternatively, the RIF group was subdivided into true RIF (patients that did not achieve implantation after at least three transfers) or heterogeneous RIF (patients that achieved implantation after the fourth transfer or more). We focused the study in the comparison between RIF and control groups (rather than the subgroups) due to the discrete sample size of our patient population.

2.2. Comparison of RIF and control patients

RIF and control conditions in patient population were homogeneous and statistically comparable (p-value > 0.05) in terms of potential cofounding variables, such as the recruiting clinic, years and type of infertility, age, BMI and endometrial progression classification. All the patients were undergoing HRT before the biopsy collection, following by a frozen and single embryo transfer. However, the pre- and post-biopsy transfers were performed in HRT or natural cycles with fresh or frozen embryos, as well as single (the majority) or double (only 12 transfers in 10/131 patients). All the transferred embryos were day 5/6 blastocysts with good quality. Although the majority of patients were undergoing HRT considering all cycles before and after the biopsy collection, the remaining patients (12/131) were distributed into minority subgroups with natural and HRT cycles which may be the reason for the slight differences (p-value = 0.01) found in the comparison RIF vs. control ***(Figure 29A)***.

In addition, the number of transfers (p-value = 2.20e-16) and implantation failures (p-value = 2.20e-16) were significantly different between the two groups, as was expected from the clinical classification criteria employed in this study ***(Figure 29A)***. Accordingly, the transcriptomic behaviour of both groups was homogenous in the PCA ***(Figure 29B)***.

However, the principal components PC1 and PC2, which are the most commonly used to visualize the variability of data in PCA, represent 28.8% of the variability between groups, indicating the necessity for employing additional strategies (i.e., artificial intelligence algorithms) that consider the larger spectrum of underlying variables in pathological endometrial function, to more accurately stratify patients.

A)

Clinical variables	RIF	Control	p-value (statistical test)
No. patients	32	99	—
Clinic	Clinic1 = 5 Clinic2 = 6 Clinic3 = 1 Clinic4 = 9 Clinic5 = 11	Clinic1 = 8 Clinic2 = 6 Clinic3 = 14 Clinic4 = 25 Clinic5 = 46	0.06 (Fisher´s)
Infertility years	2.92 (1.68) NA = 5	3.12 (2.88) NA = 8	0.62 (Wilcoxon's)
Infertility type	Primary = 24 Secondary = 5 NA = 3	Primary = 77 Secondary = 15 NA = 7	1 (Fisher's)
Age (years)	41.94 (4.20)	40.42 (4.50)	0.09 (t-test)
BMI	22.30 (3.84)	23.20 (3.71)	0.19 (Wilcoxon´s)
Endometrial progression	C1 = 20 C2 = 8 C3 = 1 C4 = 3	C1 = 59 C2 = 31 C3 = 8 C4 = 1	0.11 (Fisher´s)
No. Transfers	4.16 (1.19)	1.72 (0.81)	**2.20e-16** (Wilcoxon´s)
No. Implantation failures	3.75 (1.08)	0.72 (0.81)	**2.20e-16** (Wilcoxon´s)
Transfer cycle type	HRT = 27 HRT/Natural = 0 Natural/HRT = 1 Natural = 0 HRT/Natural/HRT = 3 HRT/Natural/HRT/Natural = 1	HRT = 92 HRT/Natural = 4 Natural/HRT = 2 Natural = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	**0.01** (Fisher's)

B)

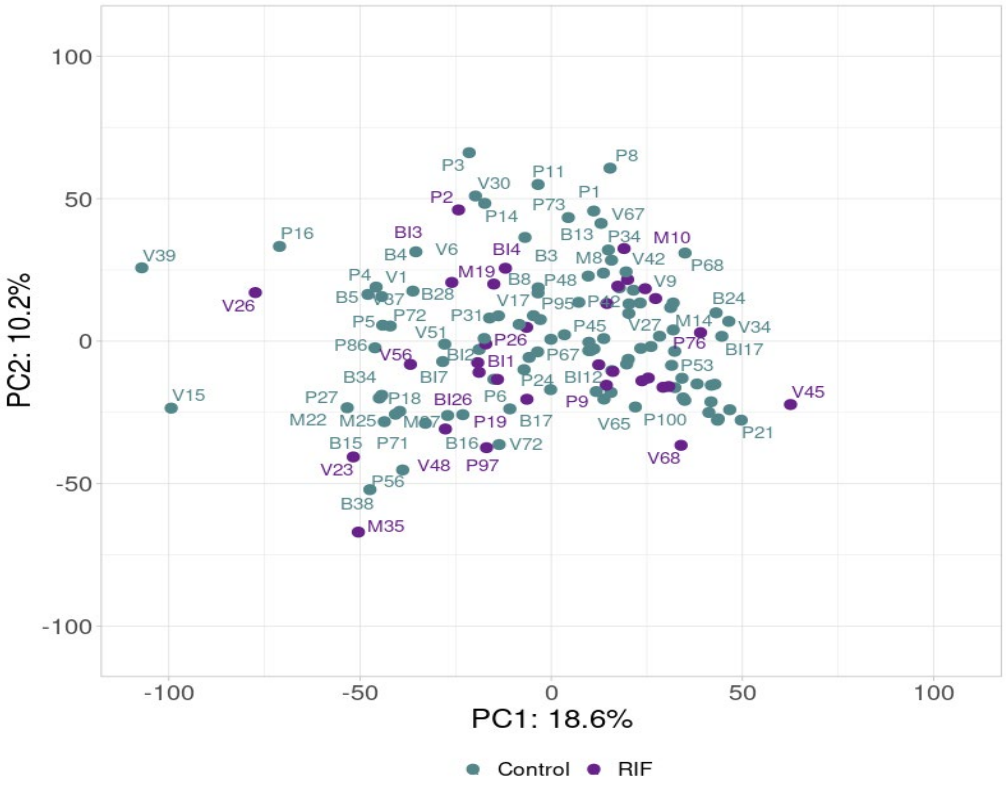


Figure 29. Comparison of the baseline characteristics of the recurrent implantation failure and control groups.
(A) *Baseline characteristics of the study population. Qualitative variables (i.e., recruiting clinic, type of infertility, endometrial progression classification and transfer cycle type) were compared by Fisher´s test. Normally (Shapiro's test > 0.05) and non-normally distributed (Shapiro´s test < 0.05) quantitative variables (years of infertility, age, body mass index (BMI), number of transfers and implantation failures) were compared using student´s t-test or Wilcoxon´s test, respectively. Data are presented as a mean and standard deviation (in brackets). In all cases, p-values < 0.05 were considered statistically significant (in bold). C1, pre-receptive; C2, receptive 1; C3, receptive 2; C4, post-secretory; HRT, hormone replacement therapy; NA, not available; No., number of.* ***(B)*** *Principal component analysis (PCA) plot. Samples (n = 131) are represented as coloured points according to their clinical classification [green, control; purple, recurrent implantation failure (RIF)]. The X axis represents the data variability (%) in the first component (PC1) and the Y axis, the variability of the second component (PC2).*

3. A transcriptomic predictor for the pathological WOI

3.1. Comparison of the training and test sets

The samples were assigned to training (n = 105) and test (n = 26) sets, both with a RIF:control proportion of 1:3. Training and test sets were homogeneous and statistically comparable (p-value > 0.05) in terms of the potential cofounding variables, such as study condition, recruiting clinic, years and type of infertility, BMI, endometrial progression classification, number of transfers and implantation failures as well as type of cycle for embryo transfers. Patients in the test set were significantly older (p-value = 0.04; ***Figure 30A***), however this was probably an effect of the discrete size of the test set in comparison to the training set. In terms of transcriptomic data, both sets appeared homogenous by PCA and the percentage of data variability explained was 30% approximately **(*Figure 30 B)***. Altogether, these results indicated the training and test sets were comparable and suitable for the development of a robust prediction model, avoiding bias in the prediction tasks.

A)

Clinical variables	Training set	Test set	p-value (statistical test)
No. patients	105	26	—
No. patients per condition	RIF = 26 Control = 79	RIF = 6 Control = 20	1 (Fisher´s)
Clinic	Clinic1 = 12 Clinic2 = 9 Clinic3 = 13 Clinic4 = 30 Clinic5 = 41	Clinic1 = 1 Clinic2 = 3 Clinic3 = 2 Clinic4 = 4 Clinic5 =16	0.26 (Fisher´s)
Infertility years	3.13 (2.49) NA = 10	2.86 (3.28) NA = 3	0.38 (Wilcoxon´s)
Infertility type	Primary = 83 Secondary = 14 NA = 8	Primary = 18 Secondary = 6 NA = 2	0.23 (Fisher´s)
Age (years)	40.37 (4.37)	42.50 (4.48)	**0.04** (t-test)
BMI	22.95 (3.84)	23.11 (3.41)	0.75 (Wilcoxon´s)
Endometrial progression	C1 = 62 C2 = 31 C3 = 9 C4 = 3	C1 = 17 C2 = 8 C3 = 0 C4 = 1	0.50 (Fisher´s)
No. Transfers	2.29 (1.46)	2.42 (1.10)	0.31 (Wilcoxon´s)
No. Implantation failures	1.44 (1.65)	1.54 (1.27)	0.40 (Wilcoxon´s)
Transfer cycle type	HRT = 93 HRT/Natural = 4 Natural/HRT = 3 Natural = 1 HRT/Natural/HRT = 3 HRT/Natural/HRT/Natural = 1	HRT = 26 HRT/Natural = 0 Natural/HRT = 0 Natural = 0 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	0.93 (Fisher´s)

B)

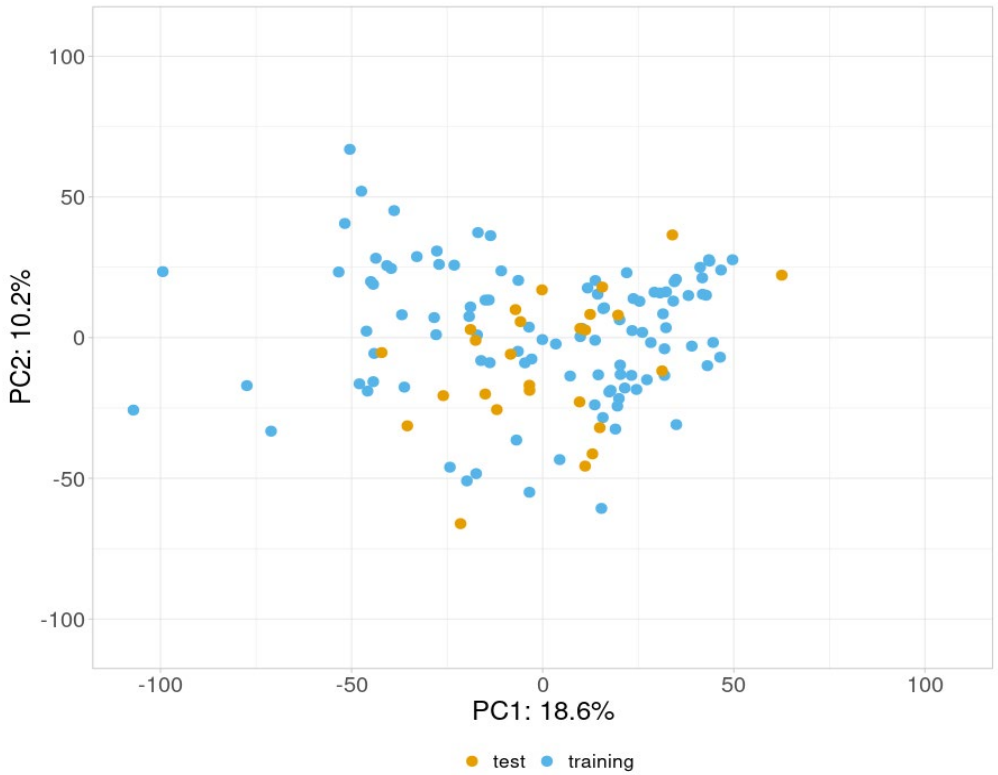


Figure 30. Comparison between training and test sets in patient population.
(A) *Baseline characteristics of the patients in the training and test sets. Fisher´s test was employed to compare qualitative variables (i.e., study condition, recruiting clinic, type of infertility, endometrial progression classification and type of transfer cycle). Normally (Shapiro's test > 0.05) and non-normally distributed (Shapiro´s test < 0.05) quantitative variables (i.e., years of infertility, age, body mass index (BMI), number of transfers and implantation failures), were compared using student´s t-test or Wilcoxon´s test, respectively. Data are presented as a mean and standard deviations (in brackets). In all cases, p-values < 0.05 were considered statistically significant (in bold). C1, pre-receptive; C2, receptive 1; C3, receptive 2; C4, post-secretory; HRT, hormone replacement therapy; NA, not available; No, number of; RIF, recurrent implantation failure.* ***(B)*** *Principal component analysis (PCA) plot. Samples (n = 131) are represented as coloured points according to their partitioning condition (blue, training set; orange, test set). The X axis represents the data variability explained (%) in the first component (PC1) and the Y axis, the variability explained of the second component (PC2).*

3.2. A gene signature that predicts the pathological WOI

The kNN algorithm predicted the largest accurate signature (Acc = 79.20%), with respect to the SVM and RF algorithms, including 236 genes related to the pathological window of implantation ***(Supplemental table 2).***

Interestingly, only 6 of the 303 genes selected by Koot et al. (Koot et al., 2016) coincided with our signature of 236 genes, resulting in a 97.46% difference. The six genes shared between both signatures were zinc finger protein 738 (*ZNF738*), NLR family pyrin domain containing 1 (*NLRP1*), acylphosphatase 1 (*ACYP1*), TRAF3 interacting protein 1 (*TRAF3IP1*), Von Willebrand factor C domain containing 2 (*VWC2*) and NLR family pyrin domain containing 5 (*NLRP5*), and were considered in the molecular study.

Koot *et al.* established their signature based on the signal-to-noise ratio (SNR), following standard methodology for the interpretation of microarray gene expression data (Koot et al., 2016). However, by employing RNA-Seq (which obtains more precise gene expression data than microarrays) followed by the CorrelationAttributeEval gene selection algorithm (to order the genes according to their potential predictive power for pathological endometrial function), and SVM, kNN and RF algorithms (to assess the

prediction performance of the selected genes), we provide a more robust methodology and refined potential biomarker signature for the detection of the pathological RIF.

3.3. The balanced probabilistic prediction model: external and internal validations

The optimal balanced probabilistic model was obtained by combining SVM and kNN algorithms. Using selected gene signature, the test set was predicted with 77% accuracy, 67% sensitivity and 80% specificity, which altogether provided the best performance in the external validation ***(Table 8)***.

Table 8. Comparison of prediction model outputs in the external validation.

Predictive parameter	External validation of the prediction model with test set					
	SVM	kNN	RF	**SVM+kNN**	SVM+RF	kNN+RF
Acc (%)	81	69	65	**77**	81	69
S (%)	33	50	50	**67**	33	50
Sp (%)	95	75	70	**80**	95	75

Base algorithms [e.g., support vector machine (SVM), k-nearest neighbors (kNN) and random forest (RF)] were employed alone or in combination to optimize the balanced prediction model. The test set was used to validate each model externally, calculating the accuracy (Acc), sensitivity (S) and specificity (Sp) of the classifications. The one with the best overall predictive parameters (in bold) was selected as the prediction model for endometrial pathology.

The five-fold cross-validation of the SVM+kNN model, using the stratified ten-fold split of the balanced training set, also provided adequate distributions of predictive parameter values in the internal validation [mean Acc = 83% (min = 78%; max = 86%), mean S = 76% (min = 69%; max = 85%) and mean Sp = 85% (min = 80%; max = 91%)] that were slightly higher than in the external validation as we expected, because the training set was

employed to select the gene signature and develop the prediction model. Therefore, we propose the first balanced prediction model (based on the combination of SVM and kNN algorithms) for endometrial RIF that avoids bias in prediction tasks. Compared to Koot *et al.*, who employ an unbalanced population, we improved the sensitivity from 58.3% to 67% (Koot et al., 2016). However, our accuracy and specificity are still moderate, probably due to the binary nature of the prediction model (RIF vs. control), which does not reflect the full transcriptomic heterogeneity of pathological endometrial function due to the complexity and multifactorial character of RIF. Therefore, although these results are very promising, further research is needed to strengthen these tools for clinical translation, especially in the context of improving precision medicine for patients with endometrial-factor infertility.

4. A new transcriptomic taxonomy for the pathological endometrial function

4.1. Clinical relevance of the transcriptomically-defined groups

Given the predicted classification and the probability of pathology output by the model, samples predicted as RIF due to a pathological WOI were stratified into p1 [n = 24 (18.32%)] and p2 [n = 14 (10.69%)], while samples predicted as control were stratified into c2 [n = 32 (24.43%)] and c1 [n = 61 (46.56%)]. All four transcriptomically-defined groups were comparable in terms of potential confounding variables, such as years and type of infertility, BMI and endometrial progression classification. The significant differences (p-value < 0.05) between the number of transfers and implantation failures (between p1 vs. c1; p1 vs. c2; p1 vs. p2; p2 vs. c1; p2 vs. c2) were expected, due to the

clinical criteria for diagnosing RIF. On the other hand, the slight differences in the type of transfer cycle (between p1 and c1), age (between p1 and c2) and recruiting clinic (between c1 and c2) were probably due to the dissimilarity in the sample size among the different transcriptomic profiles ***(Table 9)***.

Overall, the comparability of our four transcriptomic profiles confirmed that they were only poorly influenced by a few baseline characteristics, indicating that the methodology we employed to stratify patients, based on the probability of pathology, is robust enough to account for clinical differences in reproductive outcomes, reliably identify potential biases and highlight the transcriptomic heterogeneity of pathological endometrial function. Additionally, our findings supported the necessity of correcting for the endometrial progression effect. By having this variable balanced among our groups, we effectively removed the transcriptomic variations induced by the menstrual cycle that can mask the pathological functions of the endometrium [which were postulated in a previous study by our group (Devesa-Peiro et al., 2021)], corroborating the differences between the groups.

Table 9. Baseline characteristics of the stratified patients.

Clinical variable	p1 vs. c1		p-value (statistical test)
	p1	c1	
No. Patients	24	61	-
Clinic	Clinic 1 = 4 Clinic 2 = 4 Clinic 3 = 1 Clinic 4 = 7 Clinic 5 = 8	Clinic 1 = 1 Clinic 2 = 6 Clinic 3 = 9 Clinic 4 = 17 Clinic 5 = 28	0.06 (Fisher´s)
Infertility years	2.8 (1.79) NA = 4	2.94 (2.86) NA = 4	0.55 (Wilcoxon´s)
Infertility type	Primary = 18 Secondary = 3 NA = 3	Primary = 48 Secondary = 10 NA = 3	1 (Fisher´s)
Age (years)	42.5 (3.56)	40.77 (4.41)	0.07 (t-test)

BMI	22.22 (3.68)	23.62 (3.97)	0.12 (Wilcoxon`s)
Endometrial progression	C1 = 16 C2 = 6 C3 = 1 C4 = 1	C1 = 40 C2 = 17 C3 = 3 C4 = 1	0.91 (Fisher´s)
No. Transfers	4.21 (1.32)	1.75 (0.85)	**9.55e-12** (Wilcoxon`s)
No. Implantation failures	3.79 (1.14)	0.75 (0.85)	**2.77e-13** (Wilcoxon`s)
Transfer cycle type	HRT =20 Natural = 0 HRT/Natural = 0 Natural/HRT = 1 HRT/Natural/HRT = 2 HRT/Natural/HRT/Natural = 1	HRT = 56 Natural = 0 HRT/Natural = 4 Natural/HRT = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	**0.03** (Fisher´s)
Clinical variable	**p1 vs. c2**		**p-value (statistical test)**
	p1	**c2**	
No. Patients	24	32	-
Clinic	Clinic1 = 4 Clinic2 = 4 Clinic3 = 1 Clinic4 = 7 Clinic5 = 8	Clinic1 = 7 Clinic2 = 1 Clinic3 = 4 Clinic4 = 7 Clinic5 = 13	0.38 (Fisher´s)
Infertility years	2.8 (1.70) NA = 4	3.31 (3.10) NA = 4	0.68 (Wilcoxon´s)
Infertility type	Primary = 18 Secondary = 3 NA = 3	Primary = 23 Secondary = 6 NA = 3	0.72 (Fisher´s)
Age (years)	42.5 (3.56)	39.84 (4.97)	**0.02** (t-test)
BMI	22.22 (3.68)	22.53 (3.47)	0.72 (Wilcoxon´s)
Endometrial progression	C1 = 16 C2 = 6 C3 = 1 C4 = 1	C1 = 17 C2 = 9 C3 = 5 C4 = 1	0.58 (Fisher´s)
No. Transfers	4.21 (1.32)	1.78 (0.91)	**3.77e-09** (Wilcoxon´s)
No. Implantation failures	3.79 (1.14)	0.81 (1)	**5.08e-10** (Wilcoxon´s)
Transfer cycle type	HRT = 20 Natural = 0 HRT/Natural = 0 Natural/HRT = 1 HRT/Natural/HRT = 2 HRT/Natural/HRT/Natural = 1	HRT = 30 Natural = 1 HRT/Natural = 0 Natural/HRT = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	0.29 (Fisher´s)

Clinical variable	p1 vs. p2		p-value (statistical test)
	p1	p2	
No. Patients	24	14	-
Clinic	Clinic1 = 4 Clinic2 = 4 Clinic3 = 1 Clinic4 = 7 Clinic5 = 8	Clinic1 = 1 Clinic2 = 1 Clinic3 = 1 Clinic4 = 3 Clinic5 = 8	0.63 (Fisher´s)
Infertility years	2.8 (1.70) NA = 4	3.58 (1.87) NA =1	0.20 (Wilcoxon´s)
Infertility type	Primary = 18 Secondary = 3 NA = 3	Primary = 12 Secondary = 1 NA = 1	1 (Fisher´s)
Age (years)	42.5 (3.56)	40.14 (4.43)	0.10 (t-test)
BMI	22.22 (3.68)	22.53 (3.34)	0.58 (Wilcoxon´s)
Endometrial progression	C1 = 16 C2 = 6 C3 = 1 C4 = 1	C1 = 6 C2 = 7 C3 = 0 C4 =1	0.31 (Fisher´s)
No. Transfers	4.21 (1.32)	2.71 (1.44)	**4.67e-03** (Wilcoxon´s)
No. Implantation failures	3.79 (1.14)	2 (1.75)	**1.46e-03** (Wilcoxon´s)
Transfer cycle type	HRT = 20 Natural = 0 HRT/Natural = 0 Natural/HRT = 1 HRT/Natural/HRT = 2 HRT/Natural/HRT/Natural = 1	HRT = 13 Natural = 0 HRT/Natural = 0 Natural/HRT = 0 HRT/Natural/HRT = 1 HRT/Natural/HRT/Natural = 0	1 (Fisher´s)
Clinical variable	**p2 vs. c1**		**p-value (statistical test)**
	p2	**c1**	
No. Patients	14	61	-
Clinic	Clinic1 = 1 Clinic2 = 1 Clinic3 = 1 Clinic4 = 3 Clinic5 = 8	Clinic1 = 1 Clinic2 = 6 Clinic3 = 9 Clinic4 = 17 Clinic5 = 28	0.67 (Fisher´s)
Infertility years	3.58 (1.87) NA = 1	2.94 (2.86) NA = 4	0.08 (Wilcoxon´s)
Infertility type	Primary = 12 Secondary = 1 NA = 1	Primary = 48 Secondary = 10 NA = 3	0.68 (Fisher´s)
Age (years)	40.14 (4.43)	40.77 (4.41)	0.64 (t-test)
BMI	22.53 (3.34)	23.62 (3.97)	0.48 (Wilcoxon´s)

Endometrial progression	C1 = 6 C2 = 7 C3 = 0 C4 = 1	C1 = 40 C2 = 17 C3 = 3 C4 = 1	0.13 (Fisher´s)
No. Transfers	2.71 (1.44)	1.75 (0.85)	**0.02** (Wilcoxon´s)
No. Implantation failures	2 (1.75)	0.75 (0.85)	**0.01** (Wilcoxon´s)
Transfer cycle type	HRT = 13 Natural = 0 HRT/Natural = 0 Natural/HRT = 0 HRT/Natural/HRT = 1 HRT/Natural/HRT/Natural = 0	HRT = 56 Natural = 0 HRT/Natural = 4 Natural/HRT = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	0.35 (Fisher´s)
Clinical variable	**p2 vs. c2**		**p-value (statistical test)**
	p2	**c2**	
No. Patients	14	32	-
Clinic	Clinic1 = 1 Clinic2 = 1 Clinic3 = 1 Clinic4 = 3 Clinic5 = 8	Clinic1 = 7 Clinic2 = 1 Clinic3 = 4 Clinic4 = 7 Clinic5 = 13	0.70 (Fisher´s)
Infertility years	3.58 (1.87) NA = 1	3.31 (3.10) NA = 4	0.26 (Wilcoxon´s)
Infertility type	Primary = 12 Secondary = 1 NA = 1	Primary = 23 Secondary = 6 NA = 3	0.40 (Fisher´s)
Age (years)	40.14 (4.43)	39.84 (4.97)	0.84 (t-test)
BMI	22.53 (3.34)	22.53 (3.47)	0.98 (Wilcoxon´s)
Endometrial progression	C1 = 6 C2 = 7 C3 = 0 C4 = 1	C1 = 17 C2 = 9 C3 = 5 C4 = 1	0.24 (Fisher´s)
No. Transfers	2.71 (1.44)	1.78 (0.91)	**0.03** (Wilcoxon´s)
No. Implantation failures	2 (1.75)	0.81 (1)	**0.02** (Wilcoxon´s)
Transfer cycle type	HRT = 13 Natural = 0 HRT/Natural = 0 Natural/HRT = 0 HRT/Natural/HRT = 1 HRT/Natural/HRT/Natural = 0	HRT = 30 Natural = 1 HRT/Natural = 0 Natural/HRT = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	0.67 (Fisher´s)
Clinical variable	**c2 vs. c1**		**p-value (statistical test)**
	c2	**c1**	
No. Patients	32	61	-

Clinic	Clinic1 = 7 Clinic2 = 1 Clinic3 = 4 Clinic4 = 7 Clinic5 = 13	Clinic1 = 1 Clinic2 = 6 Clinic3 = 9 Clinic4 = 17 Clinic5 = 28	**0.02** (Fisher´s)
Infertility years	3.31 (3.10) NA = 4	2.95 (2.86) NA = 4	0.28 (Wilcoxon´s)
Infertility type	Primary = 23 Secondary = 6 NA = 3	Primary = 48 Secondary = 10 NA = 3	0.77 (Fisher´s)
Age (years)	39.84 (4.97)	40.77 (4.41)	0.38 (t-test)
BMI	22.53 (3.47)	23.62 (3.97)	0.23 (Wilcoxon´s)
Endometrial progression	C1 = 17 C2 = 9 C3 = 5 C4 = 1	C1 = 40 C2 = 17 C3 = 3 C4 = 1	0.27 (Fisher´s)
No. Transfers	1.78 (0.91)	1.75 (0.85)	0.95 (Wilcoxon´s)
No. Implantation failures	0.81 (1)	0.75 (0.85)	0.95 (Wilcoxon´s)
Transfer cycle type	HRT = 30 Natural = 1 HRT/Natural = 0 Natural/HRT = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	HRT = 56 Natural = 0 HRT/Natural = 4 Natural/HRT = 1 HRT/Natural/HRT = 0 HRT/Natural/HRT/Natural = 0	0.18 (Fisher´s)

Fisher´s test was employed for the comparison of qualitative variables (i.e., recruiting clinic, endometrial progression classification, type of infertility and transfer cycle). Normally distributed (Shapiro's test > 0.05) and non-normally distributed (Shapiro´s test < 0.05) quantitative variables (i.e., age, body mass index (BMI), number of transfers and implantation failures and years of infertility) were compared using student´s t-test or Wilcoxon´s test, respectively. Data are presents as a mean and standard deviations (in brackets). In all cases, p-values < 0.05 were considered statistically significant (in bold). C1, pre-receptive; C2, receptive 1; C3, receptive 2; C4, post-secretory. HRT, hormone replacement therapy; NA, not available; No., number of.

Regarding the clinical relevance of this new taxonomy, reproductive outcomes were analysed considering the first embryo transfer after biopsy collection. In comparison to patients within the control profiles, patients with the pathological profiles had lower pregnancy rates (71-78% vs. 29-57% PR and 76-91% vs. 50-57% OPR, respectively), and accordingly, higher miscarriage rates (0-8% vs. 12-43% BMR and 9-17% vs. 0-43% CMR). Notably, the p1 and p2 transcriptomic profiles were respectively associated with

the highest BMR (43%) and CMR (43%). We found significant differences between the PRs of p1 and c1 (p-value = 1.15e-03) and c2 (p-value = 3.62e-04), distinguishing p1 as the pathological group with the lowest PR. Meanwhile, the significant differences in the OPRs between p1 and c1 (p-value = 0.05), and p2 and c1 (p-value = 0.01), highlighted c1 as the control group with the highest OPR. We also found significant differences in BMR between p1 and c1 (p-value = 1.79e-03), marking p1 as the pathological group with the highest BMR. Further, we observed significant differences between the CMRs of p2 with c1 (p-value = 0.05), identifying p2 as the pathological group with the highest CMR ***(Figure 31A)***. As for the cumulative PR, which was lower in pathological profiles, there was an increasing trend from p1 to c1 profiles (with 38% for p1, 76% for p2, 81% for c2 and 93% for c1; ***Figure 31B***).

This study innovatively stratified subfertile patients undergoing HRT, according to their endometrial transcriptome (corrected for endometrial progression) and probability of RIF computed by a balanced probabilistic model. Indeed, these patients fell into four transcriptomically-defined groups, with distinct clinical implications and prognosis. This methodology not only highlighted the molecular heterogeneity underlying endometrial RIF, but also improved the reliability of predictions with respect to those based on a dichotomic classification (Koot et al., 2016), which do not sufficiently encompass the complexity and multifactorial nature of endometrial-factor infertility.

A)

Prognosis Groups		No. samples (%)	Stratification Criteria	Clinical Class (No.)	Calculated rates (%)	Fisher´s test PR (p-value)	Fisher´s test OPR (p-value)	Fisher´s test BMR (p-value)	Fisher´s test CMR (p-value)
RIF	p1	24 (18.32%)	Prob ≥ 0.8 in the prediction model	RIF (24)	PR = 29 OPR = 57 BMR = 43 CMR = 0	p1 vs. p2 = 0.18	p1 vs. p2 = 1	p1 vs. p2 = 0.28	p1 vs. p2 = 0.24
	p2	14 (10.69%)	Prob = (0.8-0.5] in the prediction model	control (8) RIF (6)	PR = 57 OPR = 50 BMR = 12 CMR = 43	p1 vs. c2 = **3.62e-04** p1 vs. c1 = **1.15e-03**	p1 vs. c2 = 0.37 p1 vs. c1 = **0.05**	p1 vs. c2 = 0.06 p1 vs. c1 = **1.79e-03**	p1 vs. c2 = 1 p1 vs. c1 = 1
Control	c2	32 (24.43%)	Prob = (0.5-0.2) in the prediction model	control (31) RIF (1)	PR = 78 OPR = 76 BMR = 8 CMR = 17	p2 vs. c2 = 0.17 p2 vs. c1 = 0.35	p2 vs. c2 = 0.20 p2 vs. c1 = **0.01**	p2 vs. c2 = 1 p2 vs. c1 = 0.16	p2 vs. c2 = 0.31 p2 vs. c1 = **0.05**
	c1	61 (46.56%)	Prob ≤ 0.2 in the prediction model	control (60) RIF (1)	PR = 71 OPR = 91 BMR = 0 CMR = 9	c2 vs. c1 = 0.47	c2 vs. c1 = 0.15	c2 vs. c1 = 0.13	c2 vs. c1 = 0.43

B)

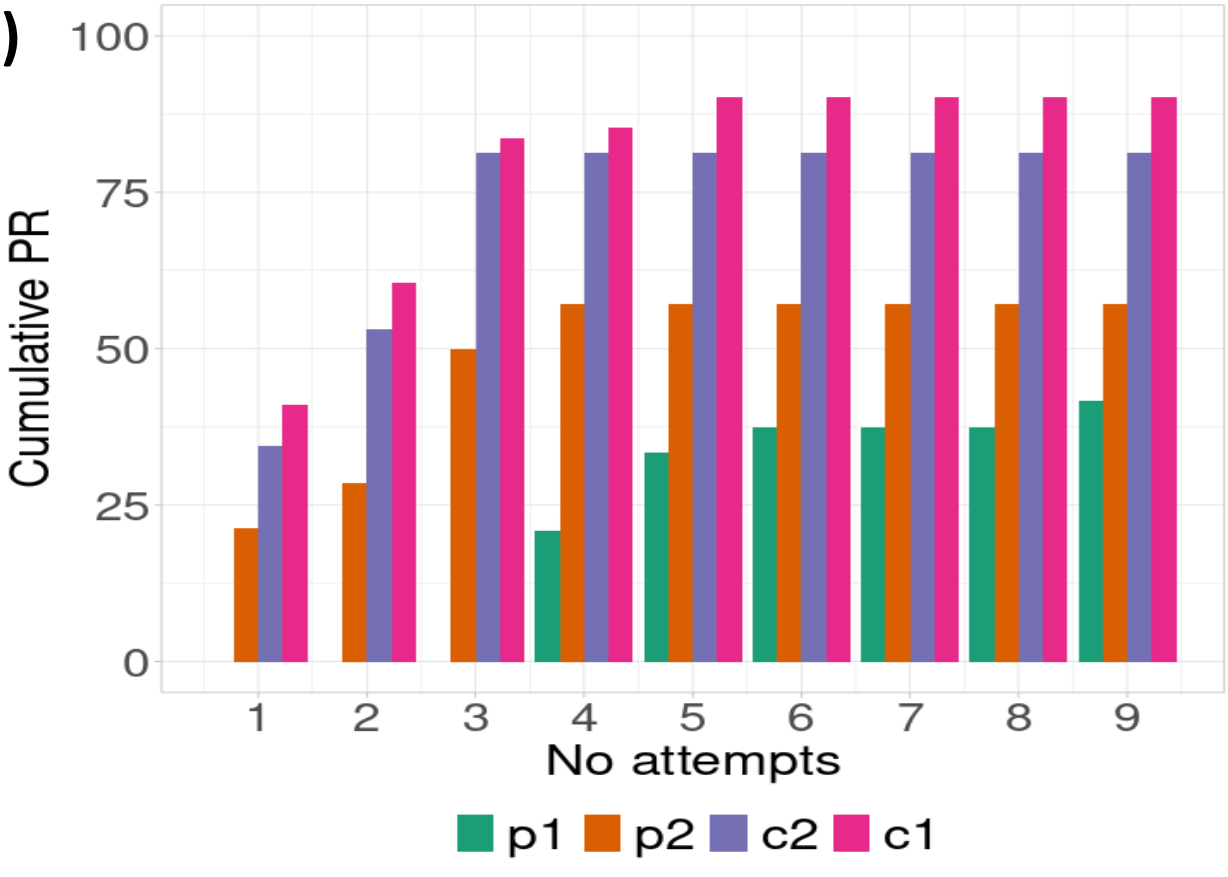


Figure 31. Clinical relevance of the new taxonomy for endometrial-factor infertility.
(A) *Reproductive outcomes of the stratified patients following the first transfer after biopsy collection. Results were compared using Fisher's test. In all cases, p-values ≤ 0.05 were considered statistically significant (in bold). BMR, biochemical miscarriage rate. CMR, clinical miscarriage rate; No., number of; OPR, ongoing pregnancy rate; PR, pregnancy rate; Prob, probability of pathology given by the balanced probabilistic prediction model based on the combination of support vector machine (SVM) and k-nearest neighbors (kNN) algorithms.* ***(B)*** *Cumulative pregnancy rates (cumulative PR) based on the total number of embryo transfers required (No. attempts) to achieving a successful pregnancy.*

Our findings complement those of Koot *et al.*, who described how pathological transcriptomic profiles (independent of endometrial progression) change in natural cycles (Koot et al., 2016). However, we study these changes in HRT, that is indicated for endometrial preparation prior to frozen embryo transfer in ART patients, to control hormonal levels without affecting the endometrial function (Kalem et al., 2018; Mumusoglu et al., 2021). Since we were able to identify the pathological WOI transcriptomically in our study, it indicated that HRT was not treating the pathology.

This is the first prospective study to propose a transcriptomically-based taxonomy for patients with endometrial RIF, and include clinical follow-up of their reproductive outcomes, unlike previous studies (Koot et al., 2016; Sebastian-Leon et al., 2018). This taxonomy could help improve the precision of diagnosis and treatment of IVF patients, as the algorithms employed in this study were able to distinguish between the profiles with good prognosis (c1 and c2 profiles) and poor prognosis (p1 and p2 profiles). Analysing the reproductive outcomes of the patients included in the study, we confirmed that at one extreme, the c1 profile had the best prognosis, while at the other extreme, the p1 prognosis group had the worst prognosis.

4.2. Molecular characterization of the transcriptomic profiles

Molecularly, the most significant distinctions were found between p2 and c1 profiles, which had 47 DEGs among them (FDR < 0.05), followed by p1 and c1, with 3 DEGs ***(Table 10, Figure 32)***. Interestingly, these 47 DEGs intersected with 13 of the 14 DEGs found between p2 and both control profiles (c2 and c1 together). The highest number of functional differences was also found comparing p2 with c1 (54 functions), and p1 with c1 (38 functions) ***(Table 11, Supplemental table 3)***. Among the 54 significantly affected functions in p2, the 30 down-regulated ones were mainly related to immune response, as well as metabolism and energy production, whereas the remaining 24 up-regulated functions were mainly related to gene expression regulation and protein degradation, as wells as nervous system and sensory perception. On the other hand, 27/38 functions from the p1 vs. c1 comparison were up-regulated in p1, and mainly related to immune response, while the remaining 11 functions were down-regulated in p1, and mainly related to cellular movement and ciliary processes.

Furthermore, the two pathological profiles (p1 and p2) were similar (with no DEGs) and the only DEG found between control profiles (*LOC644172*) was common between p1 and c1 ***(Table 10, Figure 32).*** No DEGs were found between c2 and either pathological profile, indicating c2 has more transcriptomic similarity to the pathological profiles. In comparison with p2, p1 had 7/13 up-regulated functions related to immune response, as well as proliferation and differentiation, and 4/12 down-regulated functions associated with gene expression and protein degradation ***(Table 11, Supplemental table 3).***

Table 10. Significant differentially expressed genes between different transcriptomic profiles.

Comparison	Gene ID	Ave.Expr	FC	FDR
p1 vs. c1	DND1	0.754	10.517	2.79E-03
	SYT10	0.269	-3.342	1.61E-02
	LOC644172	0.099	6.156	1.61E-02
p2 vs. c1	LYPD6	1.433	-2.116	6.40E-05
	C8orf47	3.286	-2.166	1.36E-02
	OCA2	1.128	-1.934	1.36E-02
	PTGS2	3.652	-2.140	1.36E-02
	DACT2	1.298	-2.073	1.36E-02
	CFTR	1.675	-3.244	1.36E-02
	PPIL6	2.144	-1.589	1.36E-02
	DNAI2	1.562	-2.066	1.38E-02
	VTCN1	0.383	-2.729	1.40E-02
	SLC17A8	0.652	-2.272	1.47E-02
	LGR5	4.659	-1.927	1.75E-02
	RIIAD1	1.877	-1.812	2.21E-02
	LINC00617	4.381	-1.815	2.21E-02
	C12orf75	5.253	-1.510	2.34E-02
	MAB21L1	1.332	1.605	2.74E-02
	UNC5C	2.588	1.689	2.91E-02
	LOC100287718	2.481	-1.731	2.94E-02
	AKAP14	2.497	-1.807	3.30E-02
	SPAG17	1.923	-2.145	3.30E-02
	TEKT1	1.467	-2.150	3.30E-02
	LOC643037	0.317	-1.883	3.53E-02
	ESPN	1.131	-1.856	3.53E-02
	FAM81A	3.126	-1.645	3.57E-02
	CXorf30	-0.069	-2.149	3.57E-02
	DYDC1	0.131	-2.251	3.57E-02
	CCDC65	2.274	-1.679	3.57E-02
	LAD1	5.052	-1.391	3.57E-02
	C1orf87	0.578	-1.885	3.59E-02
	DNAH10	1.665	-1.805	3.69E-02
	BDKRB2	3.133	1.745	3.85E-02
	SYT17	0.560	-1.901	3.95E-02
	SCNN1G	2.835	-2.115	3.95E-02
	FGL1	1.414	-1.764	4.00E-02
	SLC4A4	0.862	-2.153	4.14E-02
	RASL10B	1.495	-1.456	4.14E-02
	FRMPD2	1.536	-2.027	4.14E-02
	PCDH19	2.753	1.616	4.14E-02
	MORN5	1.746	-1.652	4.14E-02
	CHST6	0.615	-1.983	4.14E-02
	C1orf194	3.731	-1.638	4.52E-02
	ADGB	0.071	-2.382	4.52E-02
	KCNE3	2.666	-1.544	4.52E-02
	STOML3	1.057	-1.953	4.52E-02
	AK8	1.838	-1.430	4.52E-02
	C15orf26	0.224	-2.071	4.52E-02
	CCDC39	1.887	-1.684	4.52E-02
	C6orf165	2.557	-1.610	4.89E-02
c2 vs. c1	**LOC644172**	0.208	5.529	1.84E-02

Genes with significant values (FDR < 0.05) in the differential expression analysis (DEA) comparing the different prognosis groups are presented. The differentially expressed genes (DEGs) shared between the different comparisons are in bold. Ave.Expr; average expression of each gene in the dataset; FC, fold change; FDR, p-values adjusted by the false discovery rate; Genes ID, Hugo gene nomenclature committee symbol identifier.

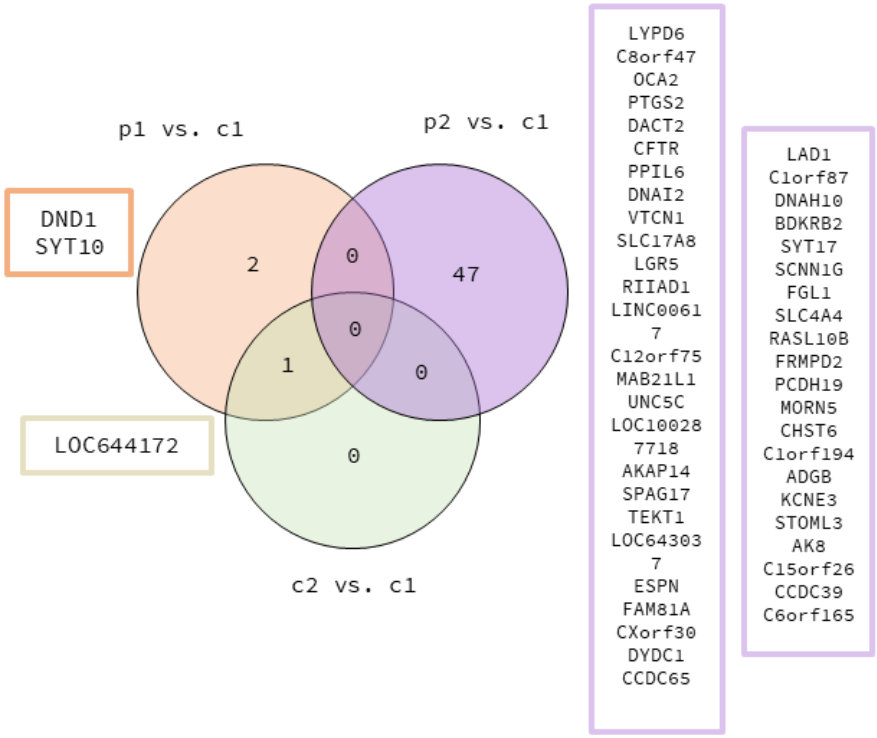


Figure 32. Venn diagram of the differentially expressed genes among the different transcriptomic profiles.
Venn diagram highlighting the differentially expressed genes (DEGs) identified by comparing different prognosis groups and their intersections. The comparisons without any DEGs were not included in the diagram (p1 vs. c2; p1 vs. p2; p2 vs. c2). Hugo gene nomenclature committee symbols of the DEGs are indicated in the coloured text boxes beside the corresponding comparison.

Similarly, when comparing p1 and c2, 9/17 of the up-regulated functions were linked to immune response, as well as proliferation and differentiation, however, 2/6 down-regulated functions were related to signal transduction and potassium activity. In comparison to c2, the up-regulated functions in p2 were associated with immune response (6/22) or gene expression and protein degradation (7/22), while most (5/14) of down-regulated functions were linked to metabolism and energy production. Between control profiles, c2 had 6/10 up-regulated functions related to nervous system and sensory

perception or hormonal response, and 9/13 down-regulated functions linked to immune response, as well as proliferation and differentiation**.**

Table 11. Summary of the molecular and functional differences between different prognosis groups.

Comparison	No. DEGs	No. enriched functions	No. up-regulated functions	No. down-regulated functions	Up-regulated functional groups	Down-regulated functional groups
p1 vs. c1	**3**	**38**	**27**	**11**	**immune response (9)** nervous system & sensory perception (5) hormonal response / nervous system & sensory perception (1) signal transduction (3) metabolism & energy production (3) cellular movement & ciliary processes (2) proliferation & differentiation (1) cellular adhesion & membranes (1) gene expression & protein degradation (1) longevity & senescence (1)	**cellular movement & ciliary processes (6)** signal transduction / potassium transport (2) metabolism & energy production (1) absorption processes (1) proliferation & differentiation (1)
p1 vs. c2	0	23	17	6	**immune response (8)** **immune response / proliferation & differentiation (1)** metabolism & energy production (3) signal transduction (2) proliferation & differentiation (1) nervous system & sensory perception (1) gene expression & protein degradation (1)	**signal transduction (1)** **signal transduction / potassium activity (1)** hormonal response / nervous system & sensory perception (1) metabolism & energy production (1) absorption processes (1) insulin secretion (1)
p1 vs. p2	0	25	13	12	**immune response (6)** **immune response / proliferation & differentiation (1)** cellular adhesion & membranes (3) gene expression regulation & protein degradation (2) nervous system & sensory perception (1)	**gene expression & protein degradation (4)** metabolism & energy production (2) cellular adhesion & membranes (2) angiogenesis, coagulation & blood pressure (2) hormonal response / nervous system & sensory perception (1) immune response (1)
p2 vs. c1	**47**	**54**	**24**	**30**	**gene expression & protein degradation (6)** **nervous system & sensory perception (5)** **hormonal response / nervous system & sensory perception (1**) immune response (3) cellular adhesion & membranes (3) signal transduction (2) angiogenesis, coagulation & blood pressure (2) longevity & senescence (1) metabolism & energy production (1)	**immune response (7)** **immune response / proliferation & differentiation (3)** **metabolism & energy production (9)** **hormonal response / metabolism & energy production (1)** cellular adhesion & membranes (3) cellular movement & ciliary processes (3) gene expression & protein degradation (2) insulin secretion (1) absorption processes (1)

p2 vs. c2	0	36	22	14	**gene expression & protein degradation (7)** **immune response (6)** cellular adhesion & membranes (3) metabolism & energy production (3) nervous system & sensory perception (1) hormonal response / nervous system & sensory perception (1) angiogenesis, coagulation & blood pressure (1)	**metabolism & energy production (5)** cellular movement & ciliary processes (2) gene expression & protein degradation (2) cellular adhesion & membranes (2) immune response (1) immune response / proliferation & differentiation (1) absorption processes (1)
c2 vs. c1	1	23	10	13	**nervous system & sensory perception (5)** **hormonal response / nervous system & sensory perception (1)** cellular adhesion & membranes (3) signal transduction (1)	**immune response (5)** **immune response / proliferation & differentiation (4)** gene expression & protein degradation (3) cellular movement & ciliary processes (1)

To aid interpretation of the functional results, biological functions (the number of which was indicated in brackets) were organized in functional groups. Comparisons with the highest numbers of enriched functions are in bold. DEGs, differentially expressed genes; No, Number of.

Furthermore, none of the six genes shared between Koot´s signature and our signature were differentially expressed in our transcriptomic profiles, which is likely due to the differences in the experimental designs between both studies. Regarding the intersections between the 236 genes from our signature and the genes with enriched functions among the prognosis groups, we found the highest number of shared traits (27 intersections) in p1 vs. c1 and p2 vs. c1 comparisons (which was expected, as they should be the genes most implied in endometrial pathology), followed by p2 vs. c2 (20 intersections), p1 vs. c2 (18 intersections), c2 vs. c1 (18 intersections), and p1 vs. p2 (12 intersections). These differences between the profiles complement our molecular findings, by highlighting the heterogeneity of endometrial RIF and supporting the need to be able to distinguish patients with different transcriptomic profiles to improve their care.

According to the Cohen´s kappa index values we computed, there was almost no concordance (values ≤ 0.34) between the enriched functions found between the different comparisons among the prognosis groups, further demonstrating the molecular heterogeneity underlying endometrial RIF ***(Figure 33).***

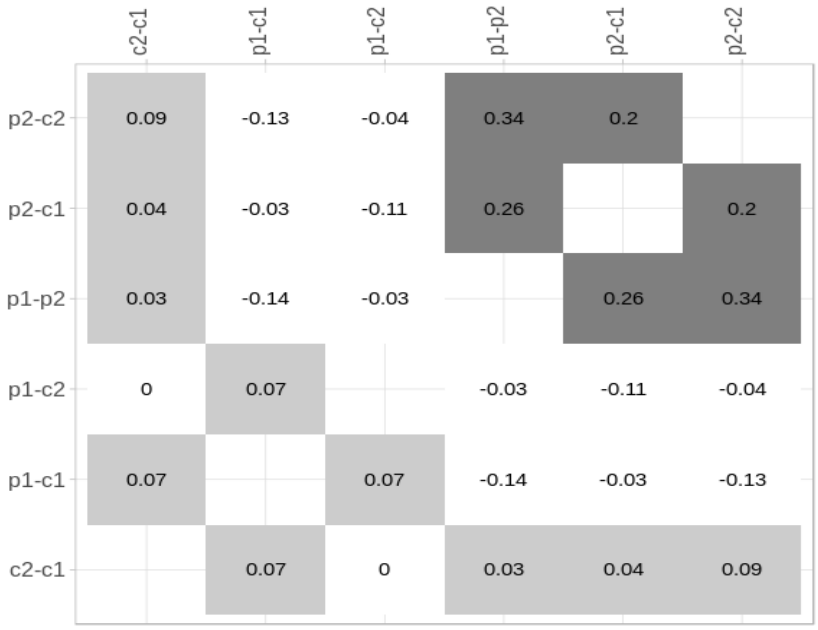


Figure 33. Cohen´s Kappa values from comparisons of enriched functions between different transcriptomic profiles.
Cohen´s Kappa coefficient (k) is a statistic we employed to assess the concordance of enriched functions from different comparisons between transcriptomic groups. Values can be interpreted as follows: k < 0 indicates no agreement, [0.0-0.20] slight agreement, [0.21-0.40] fair agreement, [0.41-0.60] moderate agreement, [0.61-0.80] substantial agreement, and [0.81-1.00] almost perfect agreement. The white-to-black colour gradient of the squares reflects the concordance (from least to most, respectively). Functional comparisons of prognosis groups are indicated on the top and left side of the figure.

Altogether, our results reveal the molecular complexity and altered physiological functions in the endometrium of women in the distinct prognosis groups. Pathological transcriptomic profiles with poor prognosis were found to be related to heightened immune responses. Concretely, p1, with the lowest PR and the highest BMR, was associated to an excessive immune response against the developing embryo in early pregnancy stages, supporting previous postulates by Sargent *et al.* who reported an

association between cytotoxic T cells and recurrent early abortions (Sargent et al., 1988). On the other hand, we found that p2, with the highest CMR, was initially immunotolerant, but led to miscarriage in late pregnancy due to the lack of metabolic responses which provides a novel association between miscarriage and metabolic deficiency. Regarding the control profiles with good prognosis, c1 was associated with a higher OPR than c2, which showed some molecular homogeny with the pathological profiles, and was more immunotolerant, but with an excessive sensory perception and hormonal response.

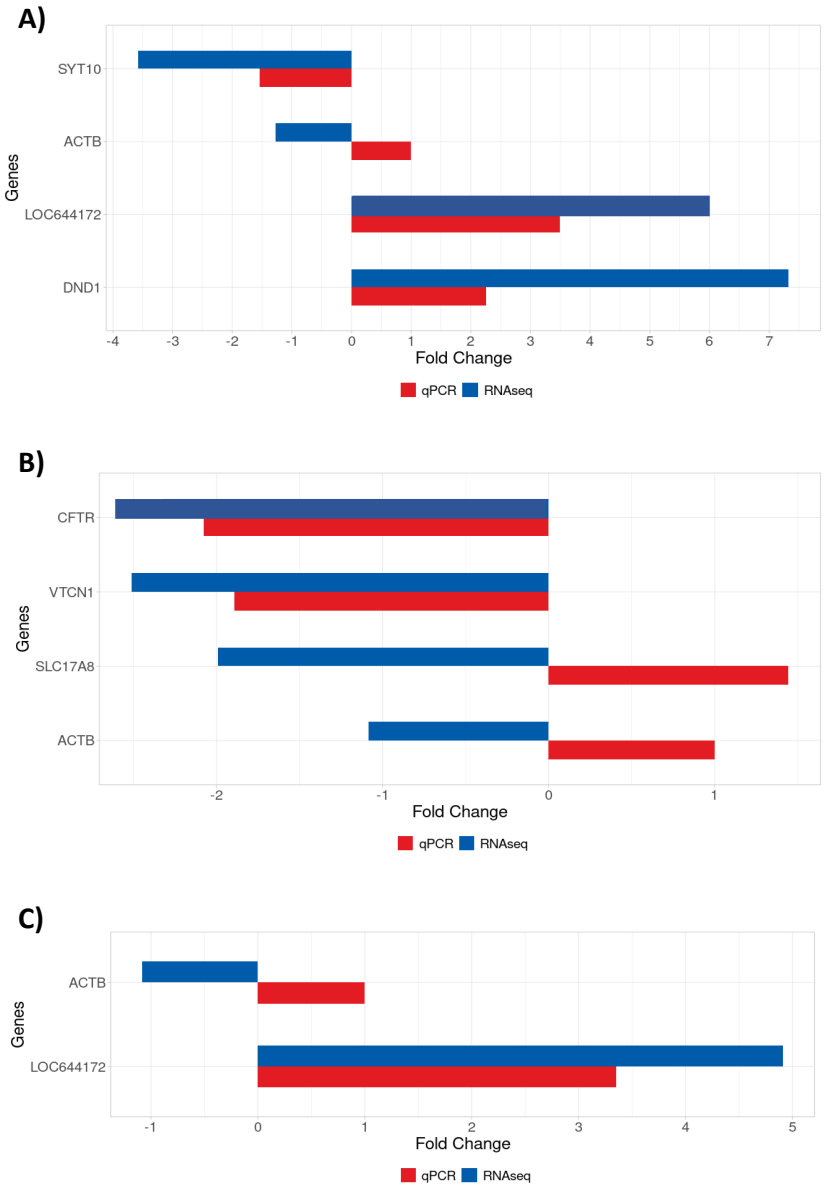
In comparison to Koot´s study (Koot et al., 2016), this project focusses the molecular and functional study on the transcriptomically-defined profiles, based on the probability of pathology computed by our balanced ML-based prediction model, rather than the broad clinical classification of endometrial RIF, providing an improved, robust, and personalized approach for patient care. In addition, the molecular and functional implications of each of our prognosis groups could set the foundation for the discovery of new biomarkers and/or therapeutic targets in the future.

4.3. Validation of potential biomarkers for the pathological WOI

Although RNA-Seq technology is sufficiently accurate, applying another technology to measure gene expression adds rigor, and elaborates study design in comparison with previous studies on endometrial pathologies (Koot et al., 2016; Sebastian-Leon et al., 2018). In this regard, six potential biomarkers were selected for qPCR validation, comprising the three DEGs between p1 and c1 (including *LOC644172*, which coincided with the comparison between c2 and c1), along with the three genes with the highest FC among the 13 DEGs that intersected between p2 vs. control, and p2 vs. c1 profiles. No batch effects related to experimental procedures or endometrial timing were observed.

Most of the validated genes corroborated the trends observed by RNA-Seq ***(Figure 34)***, with the exception of solute carrier family 17 member 8 (*SLC17A8*) ***(Figure 34B)***. Inconsistent expression measurements between different technologies may be due to small FCs [< 2; (Everaert et al., 2017)], as was the case of *SLC17A8* which had a FC of 1.44. Trends were maintained in the comparison between p1 vs. c1, for DNA microRNA-mediated repression inhibitor 1 (*DND1*), synaptotagmin (*SYT10*), and *LOC644172*, synonym of mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2 (*MAPK8IP1P2*) ***(Figure 34A)***; in the comparison between p2 and c1, for CF transmembrane conductance regulator (*CFTR*), and V-set domain containing T cell activation inhibitor 1 (*VTCN1*) ***(Figure 34B)***; and in the comparison between c2 and c1, *LOC644172* ***(Figure 34C)***. In all cases, the expression of beta-actin housekeeping gene (*ACTB*) showed a |FC| = 1, indicating it was an adequate constitutive gene ***(Figure 34)***. Therefore, *DND1*, *SYT10*, *LOC644172*, *CFTR,* and *VTCN1* were consistent between gene expression technologies, reinforcing their role as potential biomarkers for pathological endometrial function. Unlike *SLC17A8, CFTR,* and *VTCN1*, our 236 gene signature included *DND1*, *SYT10*, and *LOC644172*, suggesting they may be more relevant for endometrial RIF.

Notably, *SLC17A8* (with discordant results obtained from RNA-Seq and qPCR technologies) encodes a multifunctional transporter of L-glutamate and multiple ions (e.g., chloride, sodium and phosphate), whose main function is the transport of L-glutamate into the synaptic vesicles at the presynaptic nerve terminals of excitatory neural cells (GeneCards, 2022d). Accordingly, *DND1* (that is up-regulated in p1 with respect to c1) encodes a protein that binds to microRNA-targeting sequences of mRNAs, inhibiting microRNA-mediated repression and positively regulating gene expression (GeneCards, 2022a).



Gene	FDR	FC
SYT10	1.61E-02	-3.34
ACTB	1.00E+00	-1.00
LOC644172	1.61E-02	6.16
DND1	2.79E-03	10.5

Gene	FC
SYT10	-1.54
ACTB	1.00
LOC644172	3.48
DND1	2.25

Gene	FDR	FC
CFTR	1.36E-02	-3.24
VTCN1	1.40E-02	-2.73
SLC17A8	1.47E-02	-2.27
ACTB	9.99E-01	-1.01

Gene	FC
CFTR	-2.08
VTCN1	-1.89
SLC17A8	1.44
ACTB	1.00

Gene	FDR	FC
ACTB	9.99E-01	-1.01
LOC644172	1.84E-02	5.53

Gene	FC
ACTB	1.00
LOC644172	3.35

Figure 34. Validation of selected RNA-Sequencing results by qPCR.
Comparison of the relative gene expression of selected differentially expressed genes (DEGs) comparing p1 vs. c1 ***(A)****, p2 vs. c1* ***(B)*** *and c2 vs. c1 profiles* ***(C)****, obtained by qPCR (red bar) and RNA-Seq (blue bar). The Y axis of bar plots shows the name of each DEG assessed, while the X axis represents fold change. Numerical results are shown in the corresponding tables on the right. Gene expression was normalized to beta-actin (ACTB) housekeeping gene expression. FC, fold change; FDR, p-value adjusted by false discovery rate; qPCR, quantitative polymerase chain reaction; RNA-Seq, RNA sequencing.*

SYT10 (that is down-regulated in p1 vs. c1) encodes a calcium-sensing membrane component involved in a myriad of cellular functions, including exocytosis (in neurons from the olfactory bulb), phospholipid and syntaxin binding and protein dimerization (GeneCards, 2022e). *LOC644172* (that is up-regulated in p1 vs. c1 and c2 vs. c1), also

known as *MAPK8IP1P2*, is a pseudogene of *MAPK8IP1* with low information (GeneCards, 2022c). *MAPK8IP1* encodes a protein that mediates JNK signalling by aggregating specific components of the mitogen-activated protein kinase (MAPK) cascade. It has been associated to the regulation of beta cell function, cell signalling in mature and developing nerve terminals, and the differentiation of CD8+ T cells (GeneCards, 2022b). Furthermore, as previously described in Section 1.3 of this thesis, the proper functioning of the MAPK signalling pathway is essential for the embryo implantation process (Massimiani et al., 2020), and consequently, its alteration could lead to implantation failure. Finally, *VTCN1* (that is down-regulated in p2 with respect to c1) encodes a surface protein of antigen-presenting cells, that interacts with the ligands bound to receptors on the surface of T cells. This protein negatively regulates T-cell-mediated immune responses by inhibiting T-cell activation, proliferation, cytokine production and progression of cytotoxicity (GeneCards, 2022f). Nevertheless, more research is required to understand the molecular mechanism of action of all these genes in the human endometrium and to confirm their role as biomarkers of the pathological WOI.

On the other hand, the relationship of all these genes with pathological endometrial function has been described for first time in this study, with the exception of *CFTR*. *CFTR* encodes an epithelial ion channel that plays an important role in the regulation of epithelial ion and water transport across the cell membrane and its altered expression has been suggested to be associated with implantation failure in IVF patients (Ruan et al., 2014). The down-regulation of *CFTR* in p2 with respect to c1 could lead to the reduction of uterine fluid volume (required for implantation). While its up regulation, either by bacterial infection or hormonal disturbance, can result in abnormal uterine fluid

accumulation or increased endometrial apoptosis leading implantation failure, although further investigation is required (Ruan et al., 2014).

The main molecular and clinical findings related to each of the transcriptomically-defined prognosis groups established in this thesis are summarized in ***Figure 35.***

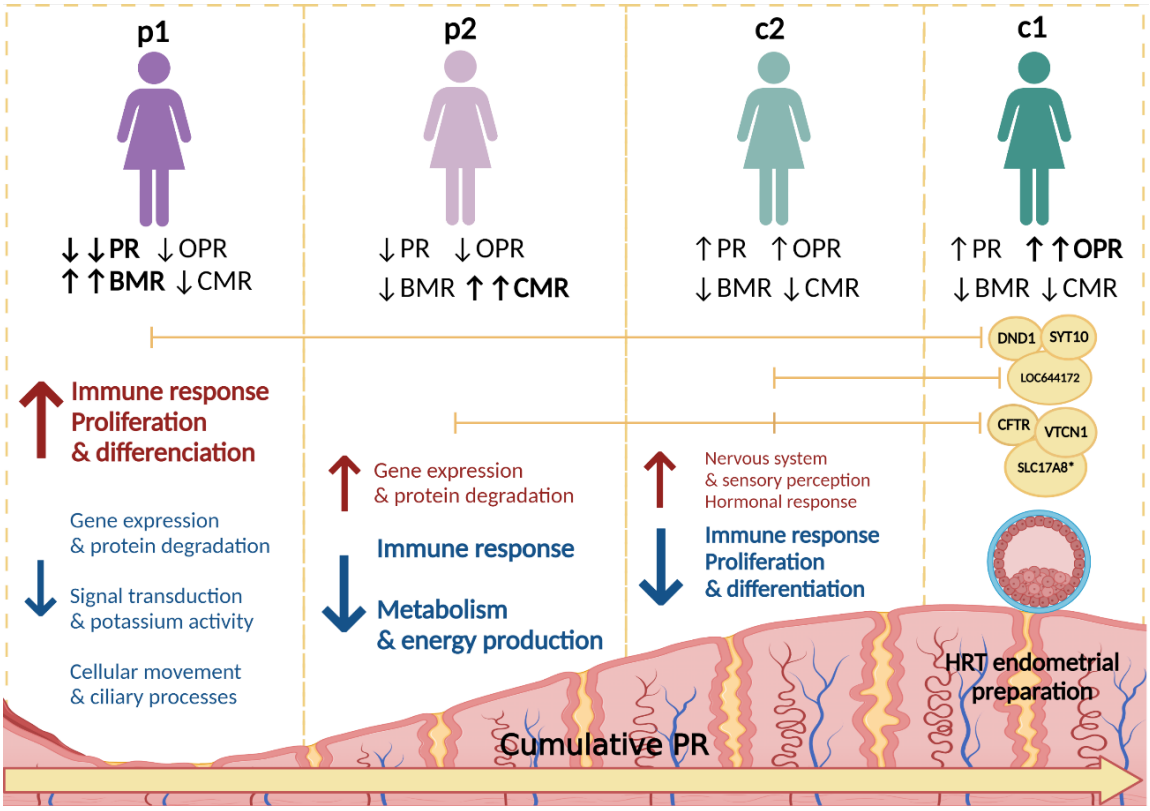


Figure 35. Main clinical and molecular differences among the transcriptomically-defined prognosis groups.

*Depiction of the most relevant clinical and molecular findings in the comparisons between distinct groups (i.e., p1, p2, c2 and c1). BMR, biochemical miscarriage rate; CMR, clinical miscarriage rate; OPR, ongoing pregnancy rate; PR, pregnancy rate. ↑ = high. ↑↑ = high and statistically significant in any comparison. ↓ = low. ↓↓ = low & statistically significant in any comparison. The DEGs between p1 and c1 included microRNA-mediated repression inhibitor 1 (DND1), synaptotagmin (SYT10) and LOC644172 [also known as the mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2 (MAPK8IP1P2)]. The only DEG between c2 and c1 was LOC644172. The DEGs with the highest fold change between p2 and control profiles or p2 and c1 included CF transmembrane conductance regulator (CFTR), V-set domain containing T cell activation inhibitor 1 (VTCN1) and solute carrier family 17 member 8 (SLC17A8). *The same gene expression trend of all indicated genes was corroborated by RNA-Sequencing and quantitative polymerase chain reaction (qPCR), with the exception of SLC17A8. Up-regulated enriched functions are shown in red while down-regulated enriched functions are represented in blue. The increased trend of cumulative pregnancy rate (PR) is displayed with an arrow. Created with BioRender.com (2022).*

5. Limitations and future perspectives

The limitations of this study should be considered when designing future studies that aim to improve the understanding of endometrial pathology, to ultimately personalize diagnostic and therapeutic strategies for patients facing RIF. Due to the transcriptomic heterogeneity of the study population, the sample size is the main limitation of this work. A larger patient cohort should be considered to further define each transcriptomic profile to train a prediction model with at least four classes. However, our novel prediction model robustly stratified patients into four transcriptomic profiles for first time, highlighting the complexity and multifactorial nature of endometrial RIF. Additionally, this study used a dichotomic model (RIF vs. control), based on the patients' clinical and transcriptomic data, to stratify ART patients into four transcriptomically-defined prognosis groups. The limiting time of this thesis did not allow us to modify the prediction model from binary to quaternary, however, our study sets a foundation for further refinement of our predictor, using our four groups and input from a larger patient population, prior to its potential clinical translation as a diagnostic strategy.

Furthermore, this study was focused on patients undergoing HRT at the moment of endometrial biopsy. Although HRT is one of the most widely employed protocols for endometrial preparation for IVF patients (Cagnacci & Venier, 2019), it could be interesting to assess the reproducibility of our findings using other types of cycles employed in the clinics. Previous works have shown similar findings in populations undergoing natural cycles (Koot et al., 2016; Sebastian-Leon et al., 2018), but the effect of modified cycles merits further investigation, to be able to standardized diagnostic procedures for infertile patients.

Finally, it could be interesting to study our molecular findings from a systems biology point of view. Systems biology is the discipline that looks for the holistic representation of interacting molecules in large networks or graphs (Assenov et al., 2008). In transcriptomics, gene co-expression analysis is very useful to study gene expression data at the systemic level, to elucidate how the genes are coordinated for a deeper functional understanding (Barabasi & Oltvai, 2004; Stuart et al., 2003). Further, systems biology makes it possible to develop predictive models of complex human diseases, prioritize new therapeutic targets, discover drugs and design relevant clinical trials. In this regard, developing a systemic pharmacological model of each of our transcriptomically-defined prognosis groups, could help predict novel/alternative treatments for infertile patients, and ultimately, pave the way for precision medicine in female reproduction (Butcher et al., 2004; Stephanou et al., 2018).

VII. CONCLUSIONS

"Science is not only a disciple of reason but, also, one of romance and passion."

Stephen Hawking

CONCLUSIONS

The following conclusions can be drawn from this PhD dissertation:

1. The pathological window of implantation, independent of endometrial progression, has an heterogeneous transcriptomic profile among patients undergoing hormone replacement therapy for endometrial preparation prior to embryo transfer (as previously reported for patients in natural cycles), that can be moderately identified with a binary model based on a biomarker signature of 236 genes. Therefore, the hormone replacement therapy is not useful to treat or correct the pathological window of implantation, because it remains detectable in our patient population.

2. Women suspected to have endometrial-factor infertility can be stratified into four transcriptomically-defined prognosis groups, according to a probability of pathology (computed by a robust artificial intelligence model). Accordingly, the profiles follow a gradient of prognosis, with two control profiles with better prognosis and two pathological profiles that have poor prognosis.

3. The control profiles are associated with moderate changes in endometrial function and clinical implications, while the pathological profiles are related to evident alterations in endometrial function and accordingly, compromised reproductive capacity. Specifically, patients stratified to the control group with the best prognosis had the highest ongoing pregnancy rate, while patients in the second control group had transcriptomic profiles closer to the pathology and related to excessive sensory perception and hormonal response. In contrast, patients stratified into the pathological

group with the worst prognosis have the highest biochemical miscarriages and lower pregnancy rates, due to poor feto-maternal immune tolerance, while patients with the poor prognosis have clinical miscarriages due to an initial immune tolerance followed by nutrient and energy deficiencies in subsequent stages of pregnancy.

4. The expression of the top differentially expressed genes in the transcriptomic profiles, *DND1*, *SYT10*, *LOC644172*, *CFTR* and *VTCN1* showed similar tendencies with RNA-Seq and qPCR analysis, reinforcing the reliability of these potential biomarkers for the pathological window of implantation in patients with recurrent implantation failure.

5. This study lays the groundwork for the design of additional next-generation tools based on this new transcriptomic stratification of patients with endometrial-factor infertility, to provide precise prognoses and facilitate the discovery of alternative and more personalized diagnoses and therapeutic strategies within the context of reproductive precision medicine.

REFERENCES

REFERENCES

Alahmari, F. (2020). A Comparison of Resampling Techniques for Medical Data Using Machine Learning. *Journal of Information and Knowledge Management*, *19*(1), 2040016. https://doi.org/10.1142/S021964922040016X

Alansari, L. M., & Wardle, P. (2012). Endometrial polyps and subfertility. *Human Fertility*, *15*(3), 129–133. https://doi.org/10.3109/14647273.2012.711499

Altmäe, S., Koel, M., Võsa, U., Adler, P., Suhorutšenko, M., Laisk-Podar, T., Kukushkina, V., Saare, M., Velthut-Meikas, A., Krjutškov, K., Aghajanova, L., Lalitkumar, P. G., Gemzell-Danielsson, K., Giudice, L., Simon, C., & Salumets, A. (2017). Meta-signature of human endometrial receptivity: A meta-analysis and validation study of transcriptomic biomarkers. *Scientific Reports*, *7*(1), 1–15. https://doi.org/10.1038/s41598-017-10098-3

Altmäe, S., Martinez-Conejero, J. A., Salumets, A., Simon, C., Horcajadas, J. A., & Stavreus-Evers, A. (2010). Endometrial gene expression analysis at the time of embryo implantation in women with unexplained infertility. *Molecular Human Reproduction*, *16*(3), 178–187. https://doi.org/10.1093/molehr/gap102

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S. (2020). *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Assenov, Y., Ramírez, F., Schelhorn, S. E. S. E., Lengauer, T., & Albrecht, M. (2008). Computing topological parameters of biological networks. *Bioinformatics*, *24*(2), 282–284. https://doi.org/10.1093/bioinformatics/btm554

Azem, F., Many, A., Yovel, I., Amit, A., Lessing, J. B., & Kupferminc, M. J. (2004). Increased rates of thrombophilia in women with repeated IVF failures. *Human Reproduction*, *19*(2), 368–370. https://doi.org/10.1093/humrep/deh069

Bala, R., Singh, V., Rajender, S., & Singh, K. (2021). Environment, Lifestyle, and Female Infertility. *Reproductive Sciences*, *28*(3), 617–638. https://doi.org/10.1007/s43032-020-00279-3

Barabasi, A. L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272

Bashiri, A., Halper, K. I., & Orvieto, R. (2018). Recurrent Implantation Failure-update overview on etiology, diagnosis, treatment and future directions. *Reproductive Biology and Endocrinology*, *16*(1), 1–18. https://doi.org/10.1186/s12958-018-0414-2

Bassil, R., Casper, R., Samara, N., Hsieh, T. B., Barzilay, E., Orvieto, R., & Haas, J. (2018). Does the endometrial receptivity array really provide personalized embryo transfer? *Journal of Assisted Reproduction and Genetics*, *35*(7), 1301–1305. https://doi.org/10.1007/s10815-018-1190-9

Bastu, E., Demiral, I., Gunel, T., Ulgen, E., Gumusoglu, E., Hosseini, M. K., Sezerman, U., Buyru, F., & Yeh, J. (2019). Potential Marker Pathways in the Endometrium That May Cause Recurrent Implantation Failure. *Reproductive Sciences*, *26*(7), 879–890. https://doi.org/10.1177/1933719118792104

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bergmann, S., Schindler, M., Munger, C., Penfold, C. A., & Boroviak, T. E. (2021). Building a stem cell-based primate uterus. *Communications Biology*, *4*(1), 1–12. https://doi.org/10.1038/s42003-021-02233-8

Bersinger, N. A., Wunder, D. M., Birkhäuser, M. H., & Mueller, M. D. (2008). Gene expression in cultured endometrium from women with different outcomes following IVF. *Molecular Human Reproduction*, *14*(8), 475–484. https://doi.org/10.1093/molehr/gan036

Bhagwat, S. S. R., Chandrashekar, D. D. S., Kakar, R., Davuluri, S., Bajpai, A. K., Nayak, S., Bhutada, S., Acharya, K., & Sachdeva, G. (2013). Endometrial Receptivity: A Revisit to Functional Genomics Studies on Human Endometrium and Creation of HGEx-ERdb. *PLoS ONE*, *8*(3), e58419. https://doi.org/10.1371/journal.pone.0058419

Billhaq, D. H., Lee, S. S. H., & Lee, S. S. H. (2020). The potential function of endometrial-secreted factors for endometrium remodeling during the estrous cycle. *Animal Science Journal*, *91*(1), e13333. https://doi.org/10.1111/asj.13333

Blagus, R., & Lusa, L. (2015). Joint use of over-and under-sampling techniques and cross-validation for the development and assessment of prediction models. *BMC Bioinformatics*, *16*(1), 1–10. https://doi.org/10.1186/s12859-015-0784-9

Blank, M., & Shoenfeld, Y. (2010). Antiphospholipid antibody-mediated reproductive failure in antiphospholipid syndrome. *Clinical Reviews in Allergy and Immunology*, *38*(2), 141–147. https://doi.org/10.1007/s12016-009-8146-x

Borthwick, J. M., Charnock-Jones, D. S., Tom, B. D., Hull, M. L., Teirney, R., Phillips, S. C., & Smith, S. K. (2003). Determination of the transcript profile of human endometrium. *Molecular Human Reproduction*, *9*(1), 19–33. https://doi.org/10.1093/molehr/gag004

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Butcher, E. C., Berg, E. L., & Kunkel, E. J. (2004). Systems biology in drug discovery. *Nature Biotechnology*, *22*(10), 1253–1259. https://doi.org/10.1038/nbt1017

Cagnacci, A., & Venier, M. (2019). The controversial history of hormone replacement therapy. *Medicina*, *55*(9), 602. https://doi.org/10.3390/medicina55090602

Canelon, S. P., & Boland, M. R. (2020). A Systematic Literature Review of Factors Affecting the Timing of Menarche: The Potential for Climate Change to Impact Women's Health. *International Journal of Environmental Research and Public Health*, *17*(5), 1703. https://doi.org/10.3390/IJERPH17051703

Cano, A., & Aliaga, R. (1995). Characteristics of urinary luteinizing hormone (LH) during the induction of LH surges of different magnitude in blood. *Human Reproduction*, *10*(1), 63–67. https://doi.org/10.1093/HUMREP/10.1.63

Carrasco-Ramiro, F., Peiró-Pastor, R., & Aguado, B. (2017). Human genomics projects and precision medicine. *Gene Therapy*, *24*(9), 551–561. https://doi.org/10.1038/gt.2017.77

Carson, D. D., Lagow, E., Thathiah, A., Al-Shami, R., Farach-Carson, M. C., Vernon, M., Yuan, L., Fritz, M. A., & Lessey, B. (2002). Changes in gene expression during the early to mid-luteal (receptive phase) transition in human endometrium detected by high-density microarray screening. *Molecular Human Reproduction*, *8*(9), 871–879. https://doi.org/10.1093/molehr/8.9.871

Catalini, L., & Fedder, J. (2020). Characteristics of the endometrium in menstruating species: Lessons learned from the animal kingdom. *Biology of Reproduction*, *106*(6), 1160–1169. https://doi.org/10.1093/biolre/ioaa029

Chen, Y., Lun, A. T. L., & Smyth, G. K. (2016). From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*, *5*. https://doi.org/10.12688/F1000RESEARCH.8987.2

Cimadomo, D., Craciunas, L., Vermeulen, N., Vomstein, K., & Toth, B. (2021). Definition, diagnostic and therapeutic options in recurrent implantation failure: An international survey of clinicians and embryologists. *Human Reproduction*, *36*(2), 305–317. https://doi.org/10.1093/humrep/deaa317

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Coughlan, C. (2018). What to do when good-quality embryos repeatedly fail to implant. *Best Practice & Research Clinical Obstetrics & Gynaecology*, *53*, 48–59. https://doi.org/10.1016/j.bpobgyn.2018.07.004

Coughlan, C., Ledger, W., Wang, Q., Liu, F., Demirol, A., Gurgan, T., Cutting, R., Ong, K., Sallam, H., & Li, T. C. (2014). Recurrent implantation failure: Definition and management. *Reproductive BioMedicine Online*, *28*(1), 14–38. https://doi.org/10.1016/j.rbmo.2013.08.011

Cozzolino, M., Diaz-Gimeno, P., Pellicer, A., & Garrido, N. (2020). Evaluation of the endometrial receptivity assay and the preimplantation genetic test for aneuploidy in overcoming recurrent implantation failure. *Journal of Assisted Reproduction and Genetics*, *37*(12), 2989–2997. https://doi.org/10.1007/s10815-020-01948-7

Cozzolino, M., Diáz-Gimeno, P., Pellicer, A., & Garrido, N. (2022). Use of the endometrial receptivity array to guide personalized embryo transfer after a failed transfer attempt was associated with a lower cumulative and per transfer live birth rate during donor and autologous cycles. *Fertility and Sterility*, *118*(4), 724–736. https://doi.org/10.1016/j.fertnstert.2022.07.007

Craciunas, L., Gallos, I., Chu, J., Bourne, T., Quenby, S., Brosens, J. J., & Coomarasamy, A. (2019). Conventional and modern markers of endometrial receptivity: A systematic review and meta-analysis. *Human Reproduction Update*, *25*(2), 202–223. https://doi.org/10.1093/humupd/dmy044

Craig Venter, J., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., … Zhu, X. (2001). The sequence of the human genome. *Science*, *29*(5507), 1304–1351. https://doi.org/10.1126/science.1058040

Critchley, H. O. D., Maybin, J. A., Armstrong, G. M., & Williams, A. R. W. (2020). Physiology of the endometrium and regulation of menstruation. *Physiological Reviews*, *100*(3), 1149–1179. https://doi.org/10.1152/physrev.00031.2019

Cruz, M., Gadea, B., Garrido, N., Pedersen, K. S., Martinez, M., Perez-Cano, I., Muñoz, M., & Meseguer, M. (2011). Embryo quality, blastocyst and ongoing pregnancy rates in oocyte donation patients whose embryos were monitored by time-lapse imaging. *Journal of Assisted Reproduction and Genetics*, *28*(7), 569–573. https://doi.org/10.1007/S10815-011-9549-1

Da Broi, M. G., Meola, J., Plaça, J. R., Peronni, K. C., Rocha, C. V., Silva, W. A., Ferriani, R. A., & Navarro, P. A. (2019). Is the profile of transcripts altered in the eutopic endometrium of infertile women with endometriosis during the implantation window? *Human Reproduction*, *34*(12), 2381–2390. https://doi.org/10.1093/humrep/dez225

Das, M., & Holzer, H. E. G. (2012). Recurrent implantation failure: gamete and embryo factors. *Fertility and Sterility*, *97*(5), 1021–1027. https://doi.org/10.1016/J.FERTNSTERT.2012.02.029

Devesa-Peiro, A., Sebastian-Leon, P., Garcia-Garcia, F., Arnau, V., Aleman, A., Pellicer, A., & Diaz-Gimeno, P. (2020). Uterine disorders affecting female fertility: what are the molecular functions altered in endometrium? *Fertility and Sterility*, *113*(6), 1261–1274. https://doi.org/10.1016/j.fertnstert.2020.01.025

Devesa-Peiro, A., Sebastian-Leon, P., Pellicer, A., & Diaz-Gimeno, P. (2021). Guidelines for biomarker discovery in endometrium: correcting for menstrual cycle bias reveals new genes associated with uterine disorders. *Molecular Human Reproduction*, *27*(4), gaab011. https://doi.org/10.1093/MOLEHR/GAAB011

Diaz-Gimeno, P., Horcajadas, J. A., Martinez-Conejero, J. A., Esteban, F. J., Alama, P., Pellicer, A., & Simon, C. (2011). A genomic diagnostic tool for human endometrial receptivity based on the transcriptomic signature. *Fertility and Sterility*, *95*(1), 50–60. https://doi.org/10.1016/j.fertnstert.2010.04.063

Diaz-Gimeno, P., Ruiz-Alonso, M., Blesa, D., Bosch, N., Martínez-Conejero, J. J. A., Alamá, P., Garrido, N., Pellicer, A., & Simón, C. (2013). The accuracy and reproducibility of the endometrial receptivity array is superior to histology as a diagnostic method for endometrial receptivity. *Fertility and Sterility*, *99*(2), 508–517. https://doi.org/10.1016/j.fertnstert.2012.09.046

Diaz-Gimeno, P., Ruiz-Alonso, M., Blesa, D., & Simon, C. (2014). Transcriptomics of the human endometrium. *International Journal of Developmental Biology*, *58*(2–4), 127–137. https://doi.org/10.1387/ijdb.130340pd

Diaz-Gimeno, P., Ruiz-Alonso, M., Sebastian-Leon, P., Pellicer, A., Valbuena, D., & Simon, C. (2017). Window of implantation transcriptomic stratification reveals different endometrial subsignatures associated with live birth and biochemical pregnancy. *Fertility and Sterility*, *108*(4), 703–710. https://doi.org/10.1016/j.fertnstert.2017.07.007

Diaz-Gimeno, P., Sebastian-Leon, P., Sanchez-Reyes, J. M., Spath, K., Aleman, A., Vidal, C., Devesa-Peiro, A., Labarta, E., Sánchez-Ribas, I., Ferrando, M., Kohls, G., García-Velasco, J. A., Seli, E., Wells, D., & Pellicer, A. (2021). Identifying and optimizing human endometrial gene expression signatures for endometrial dating. *Human Reproduction*, *37*(2), 284–296. https://doi.org/10.1093/HUMREP/DEAB262

Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, N. S., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., & Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, *14*(6), 671–683. https://doi.org/10.1093/bib/bbs046

Dimitriadis, E., Menkhorst, E., Saito, S., Kutteh, W. H., & Brosens, J. J. (2020). Recurrent pregnancy loss. *Nature Reviews Disease Primers*, *6*(1), 1–19. https://doi.org/10.1038/s41572-020-00228-z

Direito, A., Bailly, S., Mariani, A., & Ecochard, R. (2013). Relationships between the luteinizing hormone surge and other characteristics of the menstrual cycle in normally ovulating women. *Fertility and Sterility*, *99*(1), 279–285. https://doi.org/10.1016/j.fertnstert.2012.08.047

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dominguez, F., Remoh, J., Pellicer, A., & Simon, C. (2003). Human endometrial receptivity: a genomic approach. *Reproductive Biomedicine Online*, *6*(3), 332–338. https://doi.org/10.1016/S1472-6483(10)61853-6

Doyle, N., Combs, J. C., Jahandideh, S., Wilkinson, V., Devine, K., & O'Brien, J. E. (2022). Live birth after transfer of a single euploid vitrified-warmed blastocyst according to standard timing vs. timing as recommended by endometrial receptivity analysis. *Fertility and Sterility*, *118*(2), 314–321. https://doi.org/10.1016/j.fertnstert.2022.05.013

Doyle, N., Jahandideh, S., Hill, M. J., Widra, E. A., Levy, M., & Devine, K. (2022). Effect of Timing by Endometrial Receptivity Testing vs Standard Timing of Frozen Embryo Transfer on Live Birth in Patients Undergoing In Vitro Fertilization: A Randomized Clinical Trial. *JAMA*, *328*(21), 2117–2125. https://doi.org/10.1001/JAMA.2022.20438

Enciso, M., Carrascosa, J. P., Sarasa, J., Martínez-Ortiz, P. A., Munné, S., Horcajadas, J. A., & Aizpurua, J. (2018). Development of a new comprehensive and reliable endometrial receptivity map (ER Map/ER Grade) based on RT-qPCR gene expression analysis. *Human Reproduction*, *33*(2), 220–228. https://doi.org/10.1093/humrep/dex370

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115-118. https://doi.org/10.1038/nature21056

Everaert, C., Luypaert, M., Maag, J. L. V., Cheng, Q. X., DInger, M. E., Hellemans, J., & Mestdagh, P. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Scientific Reports*, *7*(1), 1–11. https://doi.org/10.1038/s41598-017-01617-3

Farquharson, R. G., Jauniaux, E., & Exalto, N. (2005). Updated and revised nomenclature for description of early pregnancy events. *Human Reproduction*, *20*(11), 3008–3011. https://doi.org/10.1093/humrep/dei167

Feero, W. G., Guttmacher, A. E., & Collins, F. S. (2010). Genomic Medicine — An Updated Primer. *New England Journal of Medicine*, *362*(21), 2001–2011. https://doi.org/10.1056/nejmra0907175

Ferraretti, A. P., La Marca, A., Fauser, B. C. J. M., Tarlatzis, B., Nargund, G., & Gianaroli, L. (2011). ESHRE consensus on the definition of 'poor response to ovarian stimulation for in vitro fertilization: The Bologna criteria. *Human Reproduction*, *26*(7), 1616–1624. https://doi.org/10.1093/humrep/der092

Ford, H. B., & Schust, D. J. (2009). Recurrent pregnancy loss: etiology, diagnosis, and therapy. *Reviews in Obstetrics & Gynecology*, *2*(2), 76.

Fox, C., & Lessey, B. A. (2018). Signaling between embryo and endometrium: Normal implantation. In *Recurrent Implantation Failure: Etiologies and Clinical Management* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-319-71967-2_1

Gambadauro, P., & Gudmundsson, J. (2017). Endometrial cancer in a woman undergoing hysteroscopy for recurrent IVF failure. *Gynecological Surgery*, *14*(1), 1–3. https://doi.org/10.1186/s10397-017-1009-1

GeneCards. (2022a). *DND1 Gene - GeneCards | DND1 Protein | DND1 Antibody*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=DND1&keywords=DND1

GeneCards. (2022b). *MAPK8IP1 Gene - GeneCards | JIP1 Protein | JIP1 Antibody*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MAPK8IP1&keywords=MAPK8IP1

GeneCards. (2022c). *MAPK8IP1P2 Gene - GeneCards | MAPK8IP1P2 Pseudogene*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=MAPK8IP1P2

GeneCards. (2022d). *SLC17A8 Gene - GeneCards | VGLU3 Protein | VGLU3 Antibody*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC17A8&keywords=SLC17A8

GeneCards. (2022e). *SYT10 Gene - GeneCards | SYT10 Protein | SYT10 Antibody*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=SYT10&keywords=SYT10

GeneCards. (2022f). *VTCN1 Gene - GeneCards | VTCN1 Protein | VTCN1 Antibody*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=VTCN1&keywords=VTCN1

Guerif, F., Bidault, R., Gasnier, O., Couet, M. L., Gervereau, O., Lansac, J., & Royere, D. (2004). Efficacy of blastocyst transfer after implantation failure. *Reproductive BioMedicine Online*, *9*(6), 630–636. https://doi.org/10.1016/S1472-6483(10)61773-7

Gui, C., & Chan, V. (2017). Machine learning in medicine. *University of Western Ontario Medical Journal*, *86*(2), 76–78. https://doi.org/10.5206/uwomj.v86i2.2060

Gurunath, S., Pandian, Z., Anderson, R. A., & Bhattacharya, S. (2011). Defining infertility-a systematic review of prevalence studies. *Human Reproduction Update*, *17*(5), 575–588. https://doi.org/10.1093/humupd/dmr015

Guzella, T. S., & Caminhas, W. M. (2009). A review of machine learning approaches to Spam filtering. *Expert Systems with Applications*, *36*(7), 10206–10222. https://doi.org/10.1016/j.eswa.2009.02.037

Hamamah, S. (2013). *WIN test* (Patent No. EP10305561.2).

Hansen, K. D., Irizarry, R. A., & Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics*, *13*(2), 204–216. https://doi.org/10.1093/biostatistics/kxr054

Haouzi, D., Mahmoud, K., Fourar, M., Bendhaou, K., Dechaud, H., De Vos, J., Rème, T., Dewailly, D., & Hamamah, S. (2009). Identification of new biomarkers of human endometrial receptivity in the natural cycle. *Human Reproduction*, *24*(1), 198–205. https://doi.org/10.1093/humrep/den360

Harada, T., Khine, Y. M., Kaponis, A., Nikellis, T., Decavalas, G., & Taniguchi, F. (2016). The Impact of Adenomyosis on Women's Fertility. *Obstetrical and Gynecological Survey*, *71*(9), 557. https://doi.org/10.1097/OGX.0000000000000346

Harper, J. C., Boelaert, K., Geraedts, J., Harton, G., Kearns, W. G., Moutou, C., Muntjewerff, N., Repping, S., SenGupta, S., Scriven, P. N., Traeger-Synodinos, J., Vesela, K., Wilton, L., & Sermon, K. D. (2006). ESHRE PGD Consortium data collection V: Cycles from January to December 2002 with pregnancy follow-up to October 2003. *Human Reproduction*, *21*(1), 3–21. https://doi.org/10.1093/humrep/dei292

Harper, M. J. (1992). The implantation window. *Bailliere's Clinical Obstetrics and Gynaecology*, *6*(2), 351–371. https://doi.org/10.1016/s0950-3552(05)80092-6

Hayssen, V., & Orr, T. (2017). *Reproduction in Mammals: The female perspective*. JHU Press.

Hernandez-Vargas, P., Muñoz, M., & Dominguez, F. (2020). Identifying biomarkers for predicting successful embryo implantation: Applying single to multi-OMICs to improve reproductive outcomes. *Human Reproduction Update*, *26*(2), 264–301. https://doi.org/10.1093/humupd/dmz042

Horne, A. W., & Critchley, H. O. D. (2007). The effect of uterine fibroids on embryo implantation. *Seminars in Reproductive Medicine*, *25*(6), 483–489. https://doi.org/10.1055/s-2007-991046

Hornik, K., Buchta, C., & Zeileis, A. (2008). Open-Source Machine Learning: R Meets Weka. *Computational Statistics*, *24*(2), 225–232. https://doi.org/10.1007/s00180-008-0119-7

Igenomix. (2021). *ERA - España*. https://www.igenomix.es/servicios-pacientes/era-analisis-receptividad-endometrial/

Illumina. (2021a). *AmpliSeq for Illumina Transcriptome Human Gene Expression Panel*. https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/ampliseq-transcriptome-gene-expression-panel.html

Illumina. (2021b). *FastQC*. https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/fastqc.html

Illumina. (2021c). *RNA Sequencing | RNA-Seq methods and workflows*. https://www.illumina.com/techniques/sequencing/rna-sequencing.html

Illumina. (2021d). *Sequencing Platforms | Compare NGS platform applications & specifications*. https://www.illumina.com/systems/sequencing-platforms.html

Illumina. (2022a). *AmpliSeq for Illumina Sequencing Solution*. https://www.illumina.com/products/by-brand/ampliseq.html

Illumina. (2022b). *Cluster density guidelines for Illumina sequencing platforms using non-patterned flow cells*. https://emea.support.illumina.com/bulletins/2016/10/cluster-density-guidelines-for-illumina-sequencing-platforms-.html

Illumina. (2022c). *NextSeq Series Specifications | Key performance parameters*. https://emea.illumina.com/systems/sequencing-platforms/nextseq/specifications.html

Illumina. (2022d). *Optimizing Cluster Density on Illumina Sequencing Systems*. https://www.illumina.com/content/dam/illumina-marketing/documents/products/other/miseq-overclustering-primer-770-2014-038.pdf

Jauniaux, E., Farquharson, R. G., Christiansen, O. B., & Exalto, N. (2006). Evidence-based guidelines for the investigation and medical treatment of recurrent miscarriage. *Human Reproduction*, *21*(9), 2216–2222. https://doi.org/10.1093/humrep/del150

Johnston-MacAnanny, E. B., Hartnett, J., Engmann, L. L., Nulsen, J. C., Sanders, M. M., & Benadiva, C. A. (2010). Chronic endometritis is a frequent finding in women with recurrent implantation failure after in vitro fertilization. *Fertility and Sterility*, *93*(2), 437–441. https://doi.org/10.1016/j.fertnstert.2008.12.131

Jollife, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Jones, R. E., & Lopez, K. H. (2013). *Human Reproductive Biology: Fourth Edition*. Academic Press.

Kalem, Z., Namlı Kalem, M., Bakirarar, B., Kent, E., & Gurgan, T. (2018). Natural cycle versus hormone replacement therapy cycle in frozen-thawed embryo transfer. *Saudi Medical Journal*, *39*(11), 1102. https://doi.org/10.15537/smj.2018.11.23299

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kao, L. C., Tulac, S., Lobo, S., Imani, B., Yang, J. P., Germeyer, A., Osteen, K., Taylor, R. N., Lessey, B. A., & Giudice, L. C. (2002). Global gene profiling in human endometrium during the window of implantation. *Endocrinology*, *143*(6), 2119–2138. https://doi.org/10.1210/endo.143.6.8885

Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: Current approaches and outstanding challenges. *PLoS Computational Biology*, *8*(2), e1002375. https://doi.org/10.1371/journal.pcbi.1002375

Khizroeva, J., Nalli, C., Bitsadze, V., Lojacono, A., Zatti, S., Andreoli, L., Tincani, A., Shoenfeld, Y., & Makatsariya, A. (2019). Infertility in women with systemic autoimmune diseases. *Best Practice and Research: Clinical Endocrinology and Metabolism*, *33*(66), 101369. https://doi.org/10.1016/j.beem.2019.101369

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), 1–13. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, S. M., & Kim, J. M. (2017). A Review of Mechanisms of Implantation. *Development & Reproduction*, *21*(4), 351. https://doi.org/10.12717/dr.2017.21.4.351

Kodaman, P. H., Arici, A., & Seli, E. (2004). Evidence-based diagnosis and management of tubal factor infertility. *Current Opinion in Obstetrics and Gynecology*, *16*(3), 221–229. https://doi.org/10.1097/00001703-200406000-00004

Koenig, I. R., Fuchs, O., Hansen, G., von Mutius, E., & Kopp, M. V. (2017). What is precision medicine? *The European Respiratory Journal*, *50*(4). https://doi.org/10.1183/13993003.00391-2017

Koler, M., Achache, H., Tsafrir, A., Smith, Y., Revel, A., & Reich, R. (2009). Disrupted gene pattern in patients with repeated in vitro fertilization (IVF) failure. *Human Reproduction*, *24*(10), 2541–2548. https://doi.org/10.1093/humrep/dep193

Kolte, A. M., Bernardi, L. A., Christiansen, O. B., Quenby, S., Farquharson, R. G., Goddijn, M., & Stephenson, M. D. (2015). Terminology for pregnancy loss prior to viability: a consensus statement from the ESHRE early pregnancy special interest group. *Human Reproduction*, *30*(3), 495–498. https://doi.org/10.1093/humrep/deu299

Koot, Y. E. M., Van Hooff, S. R., Boomsma, C. M., Van Leenen, D., Koerkamp, M. J. A. G., Goddijn, M., Eijkemans, M. J. C., Fauser, B. C. J. M., Holstege, F. C. P., & Macklon, N. S. (2016). An endometrial gene expression signature accurately predicts recurrent implantation failure after IVF. *Scientific Reports*, *6*(1), 1–12. https://doi.org/10.1038/srep19411

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, *13*, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005

Kuminski, E., George, J., Wallin, J., & Shamir, L. (2014). Combining Human and Machine Learning for Morphological Analysis of Galaxy Images. *Publications of the Astronomical Society of the Pacific*, *126*(944), 959. https://doi.org/10.1086/678977

Künzle, R., Mueller, M. D., Hänggi, W., Birkhäuser, M. H., Drescher, H., & Bersinger, N. A. (2003). Semen quality of male smokers and nonsmokers in infertile couples. *Fertility and Sterility*, *79*(2), 287–291. https://doi.org/10.1016/S0015-0282(02)04664-2

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), 1–17. https://doi.org/10.1186/gb-2014-15-2-r29

Lawrenz, B., & Fatemi, H. M. (2017). Effect of progesterone elevation in follicular phase of IVF-cycles on the endometrial receptivity. *Reproductive BioMedicine Online*, *34*(4), 422–428. https://doi.org/10.1016/j.rbmo.2017.01.011

Ledee, N., Munaut, C., Aubert, J., Serazin, V., Rahmati, M., Chaouat, G., Sandra, O., & Foidart, J. M. (2011). Specific and extensive endometrial deregulation is present before conception in IVF/ICSI repeated implantation failures (IF) or recurrent miscarriages. *Journal of Pathology*, *225*(4), 554–564. https://doi.org/10.1002/path.2948

Lessey, B. A., & Kim, J. J. (2017). Endometrial receptivity in the eutopic endometrium of women with endometriosis: it is affected, and let me show you why. *Fertility and Sterility*, *108*(1), 19–27. https://doi.org/10.1016/j.fertnstert.2017.05.031

Lessey, B. A., & Young, S. L. (2019). What exactly is endometrial receptivity? *Fertility and Sterility*, *111*(4), 611–617. https://doi.org/10.1016/j.fertnstert.2019.02.009

Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 1–16. https://doi.org/10.1186/1471-2105-12-323

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liu, K., Case, A., Cheung, A. P., Sierra, S., AlAsiri, S., Carranza-Mamane, B., Case, A., Dwyer, C., Graham, J., Havelock, J., Hemmings, R., Lee, F., Liu, K., Murdock, W., Senikas, V., Vause, T. D. R., & Wong, B. C.-M. (2011). Advanced Reproductive Age and Fertility. *Journal of Obstetrics and Gynaecology Canada*, *33*(11), 1165–1175. https://doi.org/10.1016/j.ijgo.2011.11.002

Long, E., Lin, H., Liu, Z., Wu, X., Wang, L., Jiang, J., An, Y., Lin, Z., Li, X., Chen, J., Li, J., Cao, Q., Wang, D., Liu, X., Chen, W., & Liu, Y. (2017). An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nature Biomedical Engineering*, *1*(2), 1–8. https://doi.org/10.1038/s41551-016-0024

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/s13059-014-0550-8

Maceachern, S. J., & Forkert, N. D. (2021). Machine learning for precision medicine. *Genome*, *64*(4), 416–425. https://doi.org/10.1139/GEN-2020-0131

Macklon, N. S. (2017). Recurrent implantation failure is a pathology with a specific transcriptomic signature. *Fertility and Sterility*, *108*(1), 9–14. https://doi.org/10.1016/j.fertnstert.2017.05.028

Macklon, N. S., Geraedts, J. P. M., & Fauser, B. C. J. M. (2002). Conception to ongoing pregnancy: The "black box" of early pregnancy loss. *Human Reproduction Update*, *8*(4), 333–343. https://doi.org/10.1093/humupd/8.4.333

Magic, Z., Radulovic, S., & Brankovic-Magic, M. (2007). cDNA microarrays: Identification of gene signatures and their application in clinical practice. *Journal of BU ON.: Official Journal of the Balkan Union of Oncology*, *12*, S39-44.

Manzoni, C., Kia, D. A., Vandrovcova, J., Hardy, J., Wood, N. W., Lewis, P. A., & Ferrari, R. (2018). Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Briefings in Bioinformatics*, *19*(2), 286–302. https://doi.org/10.1093/BIB/BBW114

Massimiani, M., Lacconi, V., La Civita, F., Ticconi, C., Rago, R., & Campagnolo, L. (2020). Molecular signaling regulating endometrium–blastocyst crosstalk. *International Journal of Molecular Sciences*, *21*(1), 23. https://doi.org/10.3390/ijms21010023

Mihm, M., Gangooly, S., & Muttukrishna, S. (2011). The normal menstrual cycle in women. *Animal Reproduction Science*, *124*(3–4), 229–236. https://doi.org/10.1016/j.anireprosci.2010.08.030

Miravet-Valenciano, J., Ruiz-Alonso, M., Gómez, E., & Garcia-Velasco, J. A. (2017). Endometrial receptivity in eutopic endometrium in patients with endometriosis: it is not affected, and let me show you why. *Fertility and Sterility*, *108*(1), 28–31. https://doi.org/10.1016/j.fertnstert.2017.06.002

Mirkin, S., Arslan, M., Churikov, D., Corica, A., Diaz, J. I., Williams, S., Bocca, S., & Oehninger, S. (2005). In search of candidate genes critically expressed in the human endometrium during the window of implantation. *Human Reproduction*, *20*(8), 2104–2117. https://doi.org/10.1093/humrep/dei051

Mitchell, T. (1997). *Machine Learning*. New York: McGraw Hill.

Miwa, I., Tamura, H., Takasaki, A., Yamagata, Y., Shimamura, K., & Sugino, N. (2009). Pathophysiologic features of "thin" endometrium. *Fertility and Sterility*, *91*(4), 998–1004. https://doi.org/10.1016/j.fertnstert.2008.01.029

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., & Köster, J. (2021). Sustainable data analysis with Snakemake. *F1000Research*, *10*. https://doi.org/10.12688/f1000research.29032.1

Moragianni, V. A., Jones, S. M. L., & Ryley, D. A. (2012). The effect of body mass index on the outcomes of first assisted reproductive technology cycles. *Fertility and Sterility*, *98*(1), 102–108. https://doi.org/10.1016/j.fertnstert.2012.04.004

Moustafa, S., & Young, S. (2020). Diagnostic and therapeutic options in recurrent implantation failure. *F1000Research*, *9*(208), 208. https://doi.org/10.12688/f1000research.22403.1

Mumusoglu, S., Polat, M., Ozbek, I. Y., Bozdag, G., Papanikolaou, E. G., Esteves, S. C., Humaidan, P., & Yarali, H. (2021). Preparation of the Endometrium for Frozen Embryo Transfer: A Systematic Review. *Frontiers in Endocrinology*, *12*, 831. https://doi.org/10.3389/fendo.2021.688237

Munro, M. G. (2019). Uterine polyps, adenomyosis, leiomyomas, and endometrial receptivity. *Fertility and Sterility*, *111*(4), 629–640. https://doi.org/10.1016/j.fertnstert.2019.02.008

Nagaoka, S. I., Hassold, T. J., & Hunt, P. A. (2012). Human aneuploidy: Mechanisms and new insights into an age-old problem. *Nature Reviews Genetics*, *13*(7), 493–504. https://doi.org/10.1038/nrg3245

Nakagawa, K., Kwak-Kim, J., Ota, K., Kuroda, K., Hisano, M., Sugiyama, R., & Yamaguchi, K. (2015). Immunosuppression with tacrolimus improved reproductive outcome of women with repeated implantation failure and elevated peripheral blood th1/th2 cell ratios. *American Journal of Reproductive Immunology*, *73*(4), 353–361. https://doi.org/10.1111/aji.12338

Navot, D., Bergh, P. A., Williams, M., Garrlsi, G. J., Guzman, I., Sandler, B., Fox, J., Schreiner-Engel, P., Hofmann, G. E., & Grunfeld, L. (1991). An Insight into Early Reproductive Processes through the in Vivo Model of Ovum Donation. *The Journal of Clinical Endocrinology & Metabolism*, *72*(2), 408–414. https://doi.org/10.1210/JCEM-72-2-408

Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567. https://doi.org/10.1038/nbt1206-1565

Norwitz, E. R., Schust, D. J., & Fisher, S. J. (2001). Implantation and the survival of early pregnancy. *The New England Journal of Medicine*, *345*(19), 1400–1408. https://doi.org/10.1056/NEJMRA000763

Noyes, R. W., Hertig, A. T., & Rock, J. (1975). Dating the endometrial biopsy. *American Journal of Obstetrics and Gynecology*, *122*(2), 262–263. https://doi.org/10.1016/S0002-9378(16)33500-1

Ochoa-Bernal, M. A., & Fazleabas, A. T. (2020). Physiologic events of embryo implantation and decidualization in human and non-human primates. *International Journal of Molecular Sciences*, *21*(6), 1973. https://doi.org/10.3390/ijms21061973

Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D., & Cox, L. A. (2019). The need for multi-omics biomarker signatures in precision medicine. *International Journal of Molecular Sciences*, *20*(19), 4781. https://doi.org/10.3390/ijms20194781

Orvieto, R., Meltcer, S., Nahum, R., Rabinson, J., Anteby, E. Y., & Ashkenazi, J. (2009). The influence of body mass index on in vitro fertilization outcome. *International Journal of Gynecology and Obstetrics*, *104*(1), 53–55. https://doi.org/10.1016/j.ijgo.2008.08.012

Palagiano, A., Cozzolino, M., Ubaldi, F. M., Palagiano, C., & Coccia, M. E. (2021). Effects of Hydrosalpinx on Endometrial Implantation Failures: Evaluating Salpingectomy in Women Undergoing in vitro fertilization. *Revista Brasileira de Ginecologia e Obstetricia : Revista Da Federacao Brasileira Das Sociedades de Ginecologia e Obstetricia*, *43*(4), 304–310. https://doi.org/10.1055/S-0040-1722155

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory*, *30*(2), 19–50. https://doi.org/10.2308/ajpt-50009

Pirtea, P., De Ziegler, D., Tao, X., Sun, L., Zhan, Y., Ayoubi, J. M., Seli, E., Franasiak, J. M., & Scott, R. T. (2021). Rate of true recurrent implantation failure is low: results of three successive frozen euploid single embryo transfers. *Fertility and Sterility*, *115*(1), 45–53. https://doi.org/10.1016/j.fertnstert.2020.07.002

Ponnampalam, A. P., Weston, G. C., Trajstman, A. C., Susil, B., & Rogers, P. A. W. (2004). Molecular classification of human endometrial cycle stages by transcriptional profiling. *Molecular Human Reproduction*, *10*(12), 879–893. https://doi.org/10.1093/molehr/gah121

Punyadeera, C., Dassen, H., Klomp, J., Dunselman, G., Kamps, R., Dijcks, F., Ederveen, A., De Goeij, A., & Groothuis, P. (2005). Oestrogen-modulated gene expression in the human endometrium. *Cellular and Molecular Life Sciences*, *62*(2), 239–250. https://doi.org/10.1007/s00018-004-4435-y

R Core Team, R. (2020). R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*. https://www.r-project.org/.

Ravel, J., Moreno, I., & Simon, C. (2021). Bacterial vaginosis and its association with infertility, endometritis, and pelvic inflammatory disease. *American Journal of Obstetrics and Gynecology*, *224*(3), 251–257. https://doi.org/10.1016/j.ajog.2020.10.019

Riesewijk, A., Martin, J., van Os, R., Horcajadas, J. A., Polman, J., Pellicer, A., Mosselman, S., & Simon, C. (2003). Gene expression profiling of human endometrial receptivity on days LH+2 versus LH+7 by microarray technology. *Molecular Human Reproduction*, *9*(5–6), 253–264. https://doi.org/10.1093/molehr/gag037

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Rossi, B. V., Berry, K. F., Hornstein, M. D., Cramer, D. W., Ehrlich, S., & Missmer, S. A. (2011). Effect of alcohol consumption on in vitro fertilization. *Obstetrics and Gynecology*, *117*(1), 136. https://doi.org/10.1097/AOG.0b013e31820090e1

Ruan, Y. C., Chen, H., & Chan, H. C. (2014). Ion channels in the endometrium: regulation of endometrial receptivity and embryo implantation. *Human Reproduction Update*, *20*(4), 517–529. https://doi.org/10.1093/humupd/dmu006

Ruiz-Alonso, M., Blesa, D., Diaz-Gimeno, P., Gomez, E., Fernandez-Sanchez, M., Carranza, F., Carrera, J., Vilella, F., Pellicer, A., & Simon, C. (2013). The endometrial receptivity array for diagnosis and personalized embryo transfer as a treatment for patients with repeated implantation failure. *Fertility and Sterility*, *100*(3), 818–824. https://doi.org/10.1016/j.fertnstert.2013.05.004

Santillan, I., Lozano, I., Illan, J., Verdu, V., Coca, S., Bajo-Arenas, J. M., & Martinez, F. (2015). Where and when should natural killer cells be tested in women with repeated implantation failure? *Journal of Reproductive Immunology*, *108*, 142–148. https://doi.org/10.1016/j.jri.2014.12.009

Sargent, I. L., Wilkins, T., & Redman, C. W. G. (1988). Maternal immune responses to the fetus in early pregnancy and recurrent miscarriage. *The Lancet*, *2*(8620), 1099–1104. https://doi.org/10.1016/S0140-6736(88)90522-3

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270*(5235), 467–470. https://doi.org/10.1126/science.270.5235.467

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101–1108. https://doi.org/10.1038/nprot.2008.73

Schoolcraft, W. B., Surrey, E. S., & Gardner, D. K. (2001). Embryo transfer: Techniques and variables affecting success. *Fertility and Sterility*, *76*(5), 863–870. https://doi.org/10.1016/S0015-0282(01)02731-5

Sebastian-Leon, P., Garrido, N., Remohí, J., Pellicer, A., & Diaz-Gimeno, P. (2018). Asynchronous and pathological windows of implantation: Two causes of recurrent implantation failure. *Human Reproduction*, *33*(4), 626–635. https://doi.org/10.1093/humrep/dey023

Segundo-Val, I. S., & Sanz-Lozano, C. S. (2016). Introduction to the gene expression analysis. In *Molecular genetics of asthma* (pp. 29–43). Human Press. https://doi.org/10.1007/978-1-4939-3652-6_3

Shalom-Paz, E., Anabusi, S., Michaeli, M., Karchovsky-Shoshan, E., Rothfarb, N., Shavit, T., & Ellenbogen, A. (2015). Can intra cytoplasmatic morphologically selected sperm injection (IMSI) technique improve outcome in patients with repeated IVF-ICSI failure? a comparative study. *Gynecological Endocrinology*, *31*(3), 247–251. https://doi.org/10.3109/09513590.2014.982085

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2017). DNA sequencing at 40: Past, present and future. *Nature*, *550*(7676), 345–353. https://doi.org/10.1038/nature24286

Shi, C., Han, H. J., Fan, L. J., Guan, J., Zheng, X. B., Chen, X., Liang, R., Zhang, X. W., Sun, K. K., Cui, Q. H., & Shen, H. (2018). Diverse endometrial mRNA signatures during the window of implantation in patients with repeated implantation failure. *Human Fertility*, *21*(3), 183–194. https://doi.org/10.1080/14647273.2017.1324180

Shi, L., Campbell, G., Jones, W. D., Campagne, F., Wen, Z., Walker, S. J., Su, Z., Chu, T. M., Goodsaid, F. M., Pusztai, L., Shaughnessy, J. D., Oberthuer, A., Thomas, R. S., Paules, R. S., Fielden, M., Barlogie, B., Chen, W., Du, P., Fischer, M., … Wolfinger, R. D. (2010). The Microarray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nature Biotechnology*, *28*(8), 827–838. https://doi.org/10.1038/nbt.1665

Sigurgeirsson, B., Åmark, H., Jemt, A., Ujvari, D., Westgren, M., Lundeberg, J., & Gidlöf, S. (2017). Comprehensive RNA sequencing of healthy human endometrium at two time points of the menstrual cycle. *Biology of Reproduction*, *96*(1), 24–33. https://doi.org/10.1095/biolreprod.116.142547

Simon, A., & Laufer, N. (2012). Repeated implantation failure: Clinical approach. *Fertility and Sterility*, *97*(5), 1039–1043. https://doi.org/10.1016/j.fertnstert.2012.03.010

Simon, C., Gomez, C., Cabanillas, S., Vladimirov, I., Castillon, G., Giles, J., Boynukalin, K., Findikli, N., Bahçeci, M., Ortega, I., Vidal, C., Funabiki, M., Izquierdo, A., López, L., Portela, S., Frantz, N., Kulmann, M., Taguchi, S., Labarta, E., … ERA-RCT Study Consortium Group. (2020). A 5-year multicentre randomized controlled trial comparing personalized, frozen and fresh blastocyst transfer in IVF. *Reproductive BioMedicine Online*, *41*(3), 402–415. https://doi.org/10.1016/j.rbmo.2020.06.002

Simon, C., Horcajadas, J. A., Garcia-Velasco, J. A., & Pellicer Martinez, A. (2009). *El Endometrio Humano: Desde la investigación a la clínica*. Editorial Medica Panamericana.

Somigliana, E., Vigano, P., Busnelli, A., Paffoni, A., Vegetti, W., & Vercellini, P. (2018). Repeated implantation failure at the crossroad between statistics, clinics and over-diagnosis. *Reproductive BioMedicine Online*, *36*(1), 32–38. https://doi.org/10.1016/j.rbmo.2017.09.012

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*(11), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Stephanou, A., Fanchon, E., Innominato, P. F., & Ballesta, A. (2018). Systems Biology, Systems Medicine, Systems Pharmacology: The What and The Why. *Acta Biotheoretica*, *66*(4), 345–365. https://doi.org/10.1007/S10441-018-9330-2

Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, *302*(5643), 249–255. https://doi.org/10.1126/science.1087447

Su, Z., Łabaj, P. P., Li, S. S., Thierry-Mieg, J., Thierry-Mieg, D., Shi, W., Wang, C., Schroth, G. P., Setterquist, R. A., Thompson, J. F., Jones, W. D., Xiao, W., Xu, W., Jensen, R. V., Kelly, R., Xu, J., Conesa, A., Furlanello, C., Gao, H. H., … Shi, L. (2014). A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology*, *32*(9), 903–914. https://doi.org/10.1038/nbt.2957

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Takahashi, K., Mukaida, T., Tomiyama, T., Goto, T., & Oka, C. (2004). GnRH antagonist improved blastocyst quality and pregnancy outcome after multiple failures of IVF/ICSI-ET with a GnRH agonist protocol. *Journal of Assisted Reproduction and Genetics*, *21*(9), 317–322. https://doi.org/10.1023/B:JARG.0000045470.68525.A4

Talbi, S., Hamilton, A. E., Vo, K. C., Tulac, S., Overgaard, M. T., Dosiou, C., Le Shay, N., Nezhat, C. N., Kempson, R., Lessey, B. A., Nayak, N. R., & Giudice, L. C. (2006). Molecular phenotyping of human endometrium distinguishes menstrual cycle phases and underlying biological processes in normo-ovulatory women. *Endocrinology*, *147*(3), 1097–1121. https://doi.org/10.1210/en.2005-1076

Tapia, A., Gangi, L. M., Zegers-Hochschild, F., Balmaceda, J., Pommer, R., Trejo, L., Pacheco, I. M., Salvatierra, A. M., Henríquez, S., Quezada, M., Vargas, M., Ríos, M., Munroe, D. J., Croxatto, H. B., & Velasquez, L. (2008). Differences in the endometrial transcript profile during the receptive period between women who were refractory to implantation and those who achieved pregnancy. *Human Reproduction*, *23*(2), 340–351. https://doi.org/10.1093/humrep/dem319

The Human Protein Atlas. (2022). *The human proteome in endometrium - The Human Protein Atlas*. https://www.proteinatlas.org/humanproteome/tissue/endometrium

Timeva, T., Shterev, A., & Kyurkchiev, S. (2014). Recurrent implantation failure: The role of the endometrium. *Journal of Reproduction and Infertility*, *15*(4), 173.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*(3), 562–578. https://doi.org/10.1038/nprot.2012.016

Turocy, J. M., & Rackow, B. W. (2019). Uterine factor in recurrent pregnancy loss. *Seminars in Perinatology*, *43*(2), 74–79. https://doi.org/10.1053/j.semperi.2018.12.003

Unuane, D., Tournaye, H., Velkeniers, B., & Poppe, K. (2011). Endocrine disorders & female infertility. *Best Practice and Research: Clinical Endocrinology and Metabolism*, *25*(6), 861–873. https://doi.org/10.1016/j.beem.2011.08.001

Van't Veer, L. J., Dai, H., Van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., Van Der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., & Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, *415*(6871), 530–536. https://doi.org/10.1038/415530a

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. CreateSpace.

Vander Borght, M., & Wyns, C. (2018). Fertility and infertility: Definition and epidemiology. *Clinical Biochemistry*, *62*, 2–10. https://doi.org/10.1016/j.clinbiochem.2018.03.012

Vartanyan, E., Tsaturova, K., & Devyatova, E. (2020). Thin endometrium problem in IVF programs. *Gynecological Endocrinology : The Official Journal of the International Society of Gynecological Endocrinology*, *36*(sup1), 24–27. https://doi.org/10.1080/09513590.2020.1816724

Verberg, M. F. G., Eijkemans, M. J. C., Macklon, N. S., Heijnen, E. M. E. W., Baart, E. B., Hohmann, F. P., Fauser, B. C. J. M., & Broekmans, F. J. (2009). The clinical significance of the retrieval of a low number of oocytes following mild ovarian stimulation for IVF: A meta-analysis. *Human Reproduction Update*, *15*(1), 5–12. https://doi.org/10.1093/humupd/dmn053

Wall, D. P., Kosmicki, J., Deluca, T. F., Harstad, E., & Fusaro, V. A. (2012). Use of machine learning to shorten observation-based screening and diagnosis of autism. *Translational Psychiatry*, *2*(4), e100–e100. https://doi.org/10.1038/tp.2012.10

Wang, X. Q., & Li, D. J. (2020). The mechanisms by which trophoblast-derived molecules induce maternal–fetal immune tolerance. *Cellular & Molecular Immunology*, *17*(11), 1204–1207. https://doi.org/10.1038/s41423-020-0460-5

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Waylen, A. L., Metwally, M., Jones, G. L., Wilkinson, A. J., & Ledger, W. L. (2009). Effects of cigarette smoking upon clinical outcomes of assisted reproduction: A meta-analysis. *Human Reproduction Update*, *15*(1), 31–44. https://doi.org/10.1093/humupd/dmn046

Wei, J. W., Huang, K., Yang, C., & Kang, C. S. (2017). Non-coding RNAs as regulators in epigenetics. *Oncology Reports*, *37*(1), 3–9. https://doi.org/10.3892/or.2016.5236

Wickham, H. (2011). Ggplot2. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*, 180–185. https://doi.org/10.1002/wics.147

Wilcox, A. J., Baird, D. D., & Weinberg, C. R. (1999). Time of Implantation of the Conceptus and Loss of Pregnancy. *New England Journal of Medicine*, *340*(23), 1796–1799. https://doi.org/10.1056/nejm199906103402304

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/C2015-0-02071-8

Yamada, H., Morikawa, M., Kato, E. H., Shimada, S., Kobashi, G., & Minakami, H. (2003). Pre-conceptional natural killer cell activity and percentage as predictors of biochemical pregnancy and spontaneous abortion with normal chromosome karyotype. *American Journal of Reproductive Immunology*, *50*(4), 351–354. https://doi.org/10.1034/j.1600-0897.2003.00095.x

Yang, F., Zheng, Q., & Jin, L. (2019). Dynamic Function and Composition Changes of Immune Cells During Normal and Pathological Pregnancy at the Maternal-Fetal Interface. *Frontiers in Immunology*, *10*, 2317. https://doi.org/10.3389/fimmu.2019.02317

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Yu, X. T., & Zeng, T. (2018). Integrative analysis of omics big data. In *Computational Systems Biology* (pp. 109–135). Humana Press. https://doi.org/10.1007/978-1-4939-7717-8_7

Zepiridis, L. I., Grimbizis, G. F., & Tarlatzis, B. C. (2016). Infertility and uterine fibroids. In *Best Practice and Research: Clinical Obstetrics and Gynaecology* (Vol. 34, pp. 66–73). Elsevier. https://doi.org/10.1016/j.bpobgyn.2015.12.001

Zhang, P. Y., & Yu, Y. (2020). Precise Personalized Medicine in Gynecology Cancer and Infertility. *Frontiers in Cell and Developmental Biology*, *7*, 382. https://doi.org/10.3389/fcell.2019.00382

Zhang, Z. (2016). Introduction to machine learning: K-nearest neighbors. *Annals of Translational Medicine*, *4*(11). https://doi.org/10.21037/atm.2016.03.37

Zini, A., Boman, J. M., Belzile, E., & Ciampi, A. (2008). Sperm DNA damage is associated with an increased risk of pregnancy loss after IVF and ICSI: systematic review and meta-analysis. *Human Reproduction*, *23*(12), 2663–2668. https://doi.org/10.1093/HUMREP/DEN321

APPENDIX

APPENDIX A. Supplemental Tables

Supplemental table 1. Primers employed for qPCR validation of potential biomarkers.

Selection criteria	Gene	Primer sequence (5´→ 3´)
6 DEGs	**DND1**	FW: CTCAAATTCAGCTCGCACCG RV: AGAGGTGTGACTGCCCTTCC
	SYT10	FW: AAGGCTCTGCACATCGTCA RV: AACCAACAAGGCCAGTCCAC
	MAPK8IP1P2	FW: TACGAGGCCTACAACATGCG RV: GACGACATCGCCCTTGTGAT
	CFTR	FW: GCTCCTACACCCAGCCATTT RV: AGAACACGGCTTGACAGCTT
	VTCN1	FW: GGCAAGGGGAATGCTAACCT RV: GAGAAGTTGGCTCCCTGGTC
	SLC17A8	FW: GGCATGGAGGCAACCTTACT RV: CCACTCCGTTTGAGATCCCC
HK	**ACTB**	FW: CGTACCACTGGCATCGTGAT RV: GTGTTGGCGTACAGGTCTTTG

Six differentially expressed genes (DEGs) amongst defined pathological (p) and control (c) groups and a housekeeping gene (HK), were selected to validate potential biomarkers of the pathological window of implantation using quantitative polymerase chain reaction (qPCR). MicroRNA-mediated repression inhibitor 1 (DND1), synaptotagmin (SYT10) and LOC644172, also known as mitogen-activated protein kinase 8 interacting protein 1 pseudogene 2 (MAPK8IP1P2), were differentially expressed between p1 and c1 transcriptomic profiles, with LOC644172 intersecting as a DEG between c2 and c1 profiles. CF transmembrane conductance regulator (CFTR), V-set domain containing T cell activation inhibitor 1 (VTCN1) and solute carrier family 17 member 8 (SLC17A8) were selected as intersected DEGs with the highest fold change between p2 and control or p2 and c1 profiles. Beta-actin (ACTB) was selected as a housekeeping gene. Forward (FW) and reverse (RV) primer sequences were designed for each gene using Primer-BLAST (NCBI).

Supplemental table 2. The predictive 236-gene signature for the pathological window of implantation related to recurrent implantation failure.

No.	Gene symbol	Correlation AttributeEval score	GeneCards name
1	DND1	0.3108	DND MicroRNA-Mediated Repression Inhibitor 1
2	COL8A1	0.3098	Collagen Type VIII Alpha 1 Chain
3	LINC00662	0.2763	Long Intergenic Non-Protein Coding RNA 662

4	FBXO44	0.2728	FBXO44
5	ELMOD1	0.2727	ELMO Domain Containing 1
6	MBL1P	0.2597	Mannose Binding Lectin 1, Pseudogene
7	CORIN	0.2576	Corin, Serine Peptidase
8	GLB1L2	0.2574	Galactosidase Beta 1 Like 2
9	TRIM38	0.2532	Tripartite Motif Containing 38
10	GTSE1	0.2529	G2 And S-Phase Expressed 1
11	OSR1	0.2513	Odd-Skipped Related Transcription Factor 1
12	SERPINA6	0.2499	Serpin Family A Member 6
13	IQCC	0.2487	IQ Motif Containing C
14	PROCA1	0.2476	Protein Interacting With Cyclin A1
15	LOC644172	0.2473	(MAPK8IP1P2) Mitogen-Activated Protein Kinase 8 Interacting Protein 1 Pseudogene 2
16	LYL1	0.2468	LYL1 Basic Helix-Loop-Helix Family Member
17	POF1B	0.2446	POF1B Actin Binding Protein
18	C15orf26	0.2445	(CFAP161) Cilia And Flagella Associated Protein 161
19	DNALI1	0.2444	Dynein Axonemal Light Intermediate Chain 1
20	S1PR1	0.2442	Sphingosine-1-Phosphate Receptor 1
21	PEX12	0.2427	Peroxisomal Biogenesis Factor 12
22	HIST1H2BH	0.2426	(H2BC9) H2B Clustered Histone 9
23	GAS2L2	0.2411	Growth Arrest Specific 2 Like 2
24	LRRC71	0.241	Leucine Rich Repeat Containing 71
25	RASSF1	0.2406	Ras Association Domain Family Member 1
26	C3orf18	0.2405	Chromosome 3 Open Reading Frame 18
27	KANSL1-AS1	0.2392	KANSL1 Antisense RNA 1
28	MAPK8IP1	0.2375	Mitogen-Activated Protein Kinase 8 Interacting Protein 1
29	ANKRD40	0.2369	Ankyrin Repeat Domain 40
30	ZNF713	0.235	Zinc Finger Protein 713
31	TRIM66	0.2334	Tripartite Motif Containing 66
32	OR1J2	0.2331	Olfactory Receptor Family 1 Subfamily J Member 2
33	DNAH6	0.2301	Dynein Axonemal Heavy Chain 6
34	UPK3BL	0.2281	(UPK3BL1) Uroplakin 3B Like 1
35	RSPH10B	0.2276	Radial Spoke Head 10 Homolog B
36	HHAT	0.2266	Hedgehog Acyltransferase
37	RASGRF2	0.2265	Ras Protein Specific Guanine Nucleotide Releasing Factor 2
38	SCAMP5	0.2263	Secretory Carrier Membrane Protein 5
39	KIAA1984	0.2258	(CCDC183) Coiled-Coil Domain Containing 183
40	PRRT1	0.2254	Proline Rich Transmembrane Protein 1
41	XKR6	0.2254	XK Related 6
42	CNGA4	0.2249	Cyclic Nucleotide Gated Channel Subunit Alpha 4
43	AGR3	0.2235	Anterior Gradient 3, Protein Disulphide Isomerase Family Member
44	TNFRSF14	0.2231	TNF Receptor Superfamily Member 14
45	TPPP	0.2231	Tubulin Polymerization Promoting Protein
46	LRRC18	0.223	Leucine Rich Repeat Containing 18
47	KRTAP4-12	0.2224	Keratin Associated Protein 4-12
48	MLF1IP	0.2222	(CENPU) Centromere Protein U
49	KIAA0408	0.2219	Uncharacterized Protein KIAA0408
50	DYRK3	0.2213	Dual Specificity Tyrosine Phosphorylation Regulated Kinase 3
51	CD22	0.22	CD22 Molecule
52	HBD	0.2194	Hemoglobin Subunit Delta
53	SLED1	0.2182	Proteoglycan 3, Pro Eosinophil Major Basic Protein 2 Pseudogene
54	IPO8	0.2181	Importin 8
55	LOC728024	0.2175	STING1 ER Exit Protein 1 Pseudogene
56	LOC645513	0.2169	(SEPTIN7P14) Septin 7 Pseudogene 14

57	RAB39B	0.2167	RAB39B, Member RAS Oncogene Family
58	IRAK1	0.2153	Interleukin 1 Receptor Associated Kinase 1
59	GPR52	0.214	G Protein-Coupled Receptor 52
60	FAM66D	0.2135	Family With Sequence Similarity 66 Member D
61	LINC00471	0.2135	Long Intergenic Non-Protein Coding RNA 471
62	OR1N2	0.2129	Olfactory Receptor Family 1 Subfamily N Member 2
63	DNM1P46	0.2122	Dynamin 1 Pseudogene 46
64	SPEF2	0.2118	Sperm Flagellar 2
65	SYT10	0.2117	Synaptotagmin 10
66	CPNE8	0.2115	Copine 8
67	MYL5	0.2113	Myosin Light Chain 5
68	P2RY14	0.2113	Purinergic Receptor P2Y14
69	NRF1	0.2107	Nuclear Respiratory Factor 1
70	LOC100128881	0.2102	(VPS9D1-AS1) VPS9D1 Antisense RNA 1
71	TOP2A	0.2102	DNA Topoisomerase II Alpha
72	ZNF738	0.2095	Zinc Finger Protein 738
73	EPHB2	0.209	EPH Receptor B2
74	TPPP3	0.2088	Tubulin Polymerization Promoting Protein Family Member 3
75	FAM183A	0.2086	Family With Sequence Similarity 183 Member A
76	SNHG7	0.208	Small Nucleolar RNA Host Gene 7
77	LOC100287718	0.2079	(ANKRD66) Ankyrin Repeat Domain 66
78	ZNF467	0.207	Zinc Finger Protein 467
79	TRAF3IP1	0.2069	TRAF3 Interacting Protein 1
80	BUB1	0.2067	BUB1 Mitotic Checkpoint Serine/Threonine Kinase
81	AKAP14	0.2064	A-Kinase Anchoring Protein 14
82	WDR67	0.2063	WD Repeat Domain 67 Pseudogene
83	PAGE4	0.2061	PAGE Family Member 4
84	RARRES2	0.2055	Retinoic Acid Receptor Responder 2
85	OR1J4	0.2053	Olfactory Receptor Family 1 Subfamily J Member 4
86	RSPH4A	0.2053	Radial Spoke Head Component 4A
87	HECW2	0.2052	HECT, C2 And WW Domain Containing E3 Ubiquitin Protein Ligase 2
88	H2AFZ	0.2051	(H2AZ1) H2A.Z Variant Histone 1
89	EMILIN3	0.205	Elastin Microfibril Interfacer 3
90	NLRP5	0.2046	NLR Family Pyrin Domain Containing 5
91	WDR38	0.2042	WD Repeat Domain 38
92	CDKL2	0.204	Cyclin Dependent Kinase Like 2
93	HEMK1	0.2039	HemK Methyltransferase Family Member 1
94	DIS3	0.2036	DIS3 Homolog, Exosome Endoribonuclease And 3'-5' Exoribonuclease
95	ABLIM3	0.2035	Actin Binding LIM Protein Family Member 3
96	NUSAP1	0.2032	Nucleolar And Spindle Associated Protein 1
97	FANK1	0.2031	Fibronectin Type III And Ankyrin Repeat Domains 1
98	LAG3	0.203	Lymphocyte Activating 3
99	LBH	0.2025	LBH Regulator Of WNT Signaling Pathway
100	CCDC88C	0.2022	Coiled-Coil Domain Containing 88C
101	CTBS	0.2021	Chitobiase
102	IFLTD1	0.2019	(LMNTD1) Lamin Tail Domain Containing 1
103	GTF2I	0.2013	General Transcription Factor IIi
104	CDKN2B	0.2012	Cyclin Dependent Kinase Inhibitor 2B
105	FAM8A1	0.2011	Family With Sequence Similarity 8 Member A1
106	LRRC56	0.201	Leucine Rich Repeat Containing 56
107	HPGDS	0.2009	Hematopoietic Prostaglandin D Synthase
108	LRP2BP	0.2008	LRP2 Binding Protein
109	HAPLN4	0.2006	Hyaluronan And Proteoglycan Link Protein 4
110	CCDC127	0.2002	Coiled-Coil Domain Containing 127

111	NACAP1	0.1995	(NACA4P) NACA Family Member 4, Pseudogene
112	MAPK8	0.1995	Mitogen-Activated Protein Kinase 8
113	KLK7	0.1994	Kallikrein Related Peptidase 7
114	AP2S1	0.1994	Adaptor Related Protein Complex 2 Subunit Sigma 1
115	SGPP1	0.199	Sphingosine-1-Phosphate Phosphatase 1
116	CAPZB	0.199	Capping Actin Protein Of Muscle Z-Line Subunit Beta
117	ANKRD53	0.199	Ankyrin Repeat Domain 53
118	LOC648987	0.1987	(ANXA2R-OT1) ANXA2R Overlapping Transcript
119	LOC100506343	0.1984	(INKA2-AS1) INKA2 Antisense RNA 1
120	ZNF805	0.1983	Zinc Finger Protein 805
121	NCAPH	0.1979	Non-SMC Condensin I Complex Subunit H
122	CHMP1A	0.1979	Charged Multivesicular Body Protein 1A
123	CEBPG	0.1977	CCAAT Enhancer Binding Protein Gamma
124	FUK	0.1977	(FCSK) Fucose Kinase
125	PDRG1	0.1977	P53 And DNA Damage Regulated 1
126	USP39	0.1976	Ubiquitin Specific Peptidase 39
127	SCAND2	0.1975	(SCAND2P) SCAN Domain Containing 2 Pseudogene
128	ZNF77	0.1969	Zinc Finger Protein 77
129	FAM228B	0.1968	Family With Sequence Similarity 228 Member B
130	ACVR2B	0.1967	Activin A Receptor Type 2B
131	ASPN	0.1963	Asporin
132	LOC100270804	0.1962	(LINC00653) Long Intergenic Non-Protein Coding RNA 653
133	ZNF252P.AS1	0.1961	ZNF252P Antisense RNA 1
134	DDO	0.1957	D-Aspartate Oxidase
135	CCDC39	0.1952	Coiled-Coil Domain Containing 39
136	PAOX	0.1952	Polyamine Oxidase
137	OR10C1	0.1952	Olfactory Receptor Family 10 Subfamily C Member 1
138	NSUN5P2	0.195	NSUN5 Pseudogene 2
139	FZD7	0.1949	Frizzled Class Receptor 7
140	LOC440925	0.1946	(LINC01124) Long Intergenic Non-Protein Coding RNA 1124
141	LOC643037	0.1944	(C11orf97) Chromosome 11 Open Reading Frame 97
142	EVC2	0.1939	EvC Ciliary Complex Subunit 2
143	ZNF517	0.1939	Zinc Finger Protein 517
144	ENO4	0.1939	Enolase 4
145	IL27RA	0.1938	Interleukin 27 Receptor Subunit Alpha
146	PCDHB8	0.193	Protocadherin Beta 8
147	PYROXD2	0.1929	Pyridine Nucleotide-Disulphide Oxidoreductase Domain 2
148	ZNF780A	0.1929	Zinc Finger Protein 780A
149	HPCAL4	0.1926	Hippocalcin Like 4
150	ATP6V1G2	0.1926	ATPase H+ Transporting V1 Subunit G2
151	C9orf9	0.1925	(SPACA9) Sperm Acrosome Associated 9
152	VWC2	0.1921	Von Willebrand Factor C Domain Containing 2
153	FAM86B1	0.1921	Family With Sequence Similarity 86 Member B1
154	OR6K6	0.1918	Olfactory Receptor Family 6 Subfamily K Member 6
155	TPX2	0.1917	TPX2 Microtubule Nucleation Factor
156	TPST2	0.1917	Tyrosylprotein Sulfotransferase 2
157	CENPO	0.1911	Centromere Protein O
158	PAFAH1B2	0.191	Platelet Activating Factor Acetylhydrolase 1b Catalytic Subunit 2
159	RACGAP1	0.1906	Rac GTPase Activating Protein 1
160	STK17B	0.1903	Serine/Threonine Kinase 17b
161	PRPF40A	0.1902	Pre-MRNA Processing Factor 40 Homolog A
162	HIF3A	0.19	Hypoxia Inducible Factor 3 Subunit Alpha
163	LNP1	0.1899	Leukemia NUP98 Fusion Partner 1
164	GAMT	0.1896	Guanidinoacetate N-Methyltransferase
165	OR2B2	0.1894	Olfactory Receptor Family 2 Subfamily B Member 2

166	SPEF1	0.1893	Sperm Flagellar 1
167	CGNL1	0.1889	Cingulin Like 1
168	ZNF737	0.1888	Zinc Finger Protein 737
169	LOC100505549	0.1886	(ATP8B1-AS1) ATP8B1 Antisense RNA 1
170	LRRN4CL	0.1885	LRRN4 C-Terminal Like
171	ZNF25	0.1885	Zinc Finger Protein 25
172	GFRA1	0.1884	GDNF Family Receptor Alpha 1
173	HIST1H2AD	0.1884	(H2AC7) H2A Clustered Histone 7
174	SETMAR	0.1883	SET Domain And Mariner Transposase Fusion Gene
175	KRT19	0.1882	Keratin 19
176	TTLL10	0.1882	Tubulin Tyrosine Ligase Like 10
177	NOM1	0.1882	Nucleolar Protein With MIF4G Domain 1
178	ST8SIA4	0.1881	ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 4
179	HMOX1	0.1879	Heme Oxygenase 1
180	SHROOM4	0.1878	Shroom Family Member 4
181	GNG12	0.1876	G Protein Subunit Gamma 12
182	STOX2	0.1875	Storkhead Box 2
183	CYP4X1	0.1873	Cytochrome P450 Family 4 Subfamily X Member 1
184	DPY19L1P1	0.1873	DPY19L1 Pseudogene 1
185	FBXL15	0.1873	F-Box And Leucine Rich Repeat Protein 15
186	GMNC	0.1871	Geminin Coiled-Coil Domain Containing
187	C14orf93	0.1863	Chromosome 14 Open Reading Frame 93
188	MYCBPAP	0.1862	MYCBP Associated Protein
189	LOC100129269	0.1862	(LINC01160) Long Intergenic Non-Protein Coding RNA 1160
190	ZNF675	0.1861	Zinc Finger Protein 675
191	ESM1	0.186	Endothelial Cell Specific Molecule 1
192	TMEM237	0.1859	Transmembrane Protein 237
193	TMEM176B	0.1859	Transmembrane Protein 176B
194	LPAR1	0.1855	Lysophosphatidic Acid Receptor 1
195	GPAT2	0.1854	Glycerol-3-Phosphate Acyltransferase 2, Mitochondrial
196	ZNF8	0.1854	Zinc Finger Protein 8
197	USF1	0.1852	Upstream Transcription Factor 1
198	RPL23AP7	0.1851	Ribosomal Protein L23a Pseudogene 7
199	GPR18	0.185	G Protein-Coupled Receptor 18
200	ACYP1	0.1849	Acylphosphatase 1
201	N4BP2L1	0.1849	NEDD4 Binding Protein 2 Like 1
202	SLC43A1	0.1848	Solute Carrier Family 43 Member 1
203	HMCN1	0.1847	Hemicentin 1
204	FBXL14	0.1844	F-Box And Leucine Rich Repeat Protein 14
205	TUBA4B	0.1842	Tubulin Alpha 4b
206	TADA2A	0.1841	Transcriptional Adaptor 2A
207	PPIL6	0.1839	Peptidylprolyl Isomerase Like 6
208	VDR	0.1839	Vitamin D Receptor
209	EFNA5	0.1838	Ephrin A5
210	SMU1	0.1838	SMU1 DNA Replication Regulator And Spliceosomal Factor
211	HMGB3P1	0.1838	High Mobility Group Box 3 Pseudogene 1
212	AP3S2	0.1835	Adaptor Related Protein Complex 3 Subunit Sigma 2
213	HIST1H2AL	0.1835	(H2AC16) H2A Clustered Histone 16
214	RASA4CP	0.1831	RAS P21 Protein Activator 4C, Pseudogene
215	DPYSL2	0.1826	Dihydropyrimidinase Like 2
216	NLRP1	0.1824	NLR Family Pyrin Domain Containing 1
217	GPSM3	0.1822	G Protein Signaling Modulator 3
218	GATSL3	0.1822	(CASTOR1) Cytosolic Arginine Sensor For MTORC1 Subunit 1
219	IDUA	0.182	Alpha-L-Iduronidase

220	NAT14	0.1818	N-Acetyltransferase 14 (Putative)
221	FAM84A	0.1815	(LRATD1) LRAT Domain Containing 1
222	CAPSL	0.1811	Calcyphosine Like
223	PTCHD4	0.181	Patched Domain Containing 4
224	FCHSD2	0.181	FCH And Double SH3 Domains 2
225	ALG1	0.1807	ALG1 Chitobiosyldiphosphodolichol Beta-Mannosyltransferase
226	NUDC	0.1804	Nuclear Distribution C, Dynein Complex Regulator
227	LOC100506233	0.1804	(RAB30-DT) RAB30 Divergent Transcript
228	SOS1	0.1801	SOS Ras/Rac Guanine Nucleotide Exchange Factor 1
229	CDHR3	0.1801	Cadherin Related Family Member 3
230	ZBBX	0.18	Zinc Finger B-Box Domain Containing
231	MGAT2	0.1798	Alpha-1,6-Mannosyl-Glycoprotein 2-Beta-N-Acetylglucosaminyltransferase
232	DUSP2	0.1795	Dual Specificity Phosphatase 2
233	NKD1	0.1795	NKD Inhibitor Of WNT Signaling Pathway 1
234	TDGF1	0.1793	Teratocarcinoma-Derived Growth Factor 1
235	ZNF542	0.1793	Zinc Finger Protein 542, Pseudogene
236	CCDC164	0.179	(DRC1) Dynein Regulatory Complex Subunit 1

The genes are numerically ordered according to the potential predictive power. The gene symbol, score from CorrelationAttributeEval, and complete gene name from GeneCards are listed. No., number of ordered genes.

Supplemental table 3. Enriched functions from gene set enrichment analysis between different prognosis groups.

Functional group	ID	Function Name	Comparisons between different prognosis groups											
			p1 vs. c1		p1 vs. c2		p1 vs. p2		p2 vs. c1		p2 vs. c2		c2 vs. c1	
			NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR	NES	FDR
Immune response	hsa04060	cytokine-cytokine receptor interaction	3.29	6.10 E-06	3.82	2.66 E-08	3.41	2.91 E-06	-2.60	2.40 E-03	2.70	1.50 E-03	-2.83	3.00 E-04
	hsa04650	natural killer cell mediated cytotoxicity	2.83	3.00 E-04	4.51	7.42 E-12			2.48	3.80 E-03	2.88	3.00 E-04		
	hsa04061	viral protein interaction with cytokine and cytokine receptor	2.55	6.30 E-03	2.69	1.60 E-03	2.23	3.42 E-02	-2.02	3.74 E-02	2.20	3.28 E-02		
	hsa04623	cytosolic DNA-sensing pathway	2.31	1.37 E-02										
	hsa04062	chemokine signaling pathway	2.22	1.77 E-02	2.09	3.77 E-02								
	hsa04610	complement and coagulation cascades	2.24	1.96 E-02			-2.59	1.90 E-03	2.52	2.40 E-03	2.41	1.07 E-02		
	hsa04670	leukocyte transendothelial migration	2.13	2.41 E-02					1.94	3.41 E-02				
	hsa04622	RIG-I-like receptor signaling pathway	2.07	3.50 E-02										

	hsa04620	Toll-like receptor signaling pathway	1.99	3.61 E-02										
	hsa04612	antigen processing and presentation			2.52	5.80 E-03	2.19	1.62 E-02	-2.22	1.94 E-02			-2.69	2.30 E-03
	hsa04660	T cell receptor signaling pathway			2.43	8.20 E-03								
	hsa04662	B cell receptor signaling pathway			2.05	4.84 E-02								
	hsa04672	intestinal immune network for IgA production					2.47	3.90 E-03	-2.19	1.94 E-02				
	hsa04657	IL-17 signaling pathway					2.52	3.90 E-03	-2.09	2.42 E-02	-2.22	2.72 E-02		
	hsa04145	phagosome					2.25	2.18 E-02	-2.18	1.94 E-02				
	GO:0002684	positive regulation of immune system process			2.64	1.67 E-02								
	GO:0002697	regulation of immune effector process							-2.34	4.04 E-02				
	GO:0002717	positive regulation of natural killer cell mediated immunity									2.71	2.09 E-02	-2.50	4.97 E-02
	GO:0042269	regulation of natural killer cell mediated cytotoxicity									2.55	3.68 E-02	-2.64	1.65 E-02
	GO:0003823	antigen binding											-2.37	3.97 E-02
Immune response / proliferation & differentiation	hsa04380	osteoclast differentiation			2.13	3.77 E-02								
	hsa04640	hematopoietic cell lineage					3.29	7.54 E-06	-2.43	5.20 E-03	-2.25	2.72 E-02	-2.22	4.47 E-02
	GO:0050672	negative regulation of lymphocyte proliferation							-2.65	2.83 E-02			-2.76	8.30 E-03
	GO:0042129	regulation of T cell proliferation							-2.60	2.96 E-02				
	GO:0030888	regulation of B cell proliferation											-2.67	9.50 E-03
	GO:0050671	positive regulation of lymphocyte proliferation											-2.61	3.12 E-02
Nervous system & sensory perception	hsa04740	olfactory transduction	14.10	8.06 E-205	8.87	2.58 E-55	5.85	6.08 E-24	12.87	5.00 E-168	3.67	1.68 E-07	14.42	2.64 E-200
	hsa04742	taste transduction	3.40	3.53 E-06					3.12	1.00 E-04			3.22	2.05 E-05

	GO:0050912	detection of chemical stimulus involved in sensory perception of taste	3.30	4.61 E-05					3.52	9.23 E-07			2.80	6.80 E-03
	GO:0050913	sensory perception of bitter taste	3.30	4.61 E-05					3.52	9.23 E-07			2.80	6.80 E-03
	GO:0008527	taste receptor activity	2.86	2.30 E-03					3.38	1.42 E-06			2.71	4.90 E-03
Hormonal response / nervous system & sensory perception	hsa04080	neuroactive ligand-receptor interaction	3.03	3.15 E-05	-2.56	3.50 E-03	-2.94	2.00 E-04	3.64	5.68 E-08	3.29	8.74 E-06	3.15	2.05 E-05
Hormonal response / metabolism & energy production	hsa00140	steroid hormone biosynthesis							-2.02	3.17 E-02				
Signal transduction	hsa04151	PI3K-Akt signaling pathway	3.30	3.33 E-06	2.10	3.77 E-02			2.06	2.72 E-02			2.20	3.80 E-02
	hsa04630	JAK-STAT signaling pathway	2.83	3.00 E-04	2.65	1.60 E-03			1.92	4.23 E-02				
	hsa04668	TNF signaling pathway	2.06	3.18 E-02										
	hsa04020	calcium signaling pathway			-2.09	4.84 E-02								
Signal transduction / potassium transport	GO:0006813	potassium ion transport	-2.46	4.64 E-02										
	GO:0005267	potassium channel activity			-2.47	2.95 E-02								
	GO:0034705	potassium channel complex	-2.94	1.00 E-03										
Metabolism & energy production	hsa04140	autophagy - animal	2.32	1.25 E-02					1.95	3.19 E-02	2.15	2.72 E-02		
	hsa00510	N-Glycan biosynthesis	2.27	1.77 E-02										
	hsa00430	taurine and hypotaurine metabolism			-2.07	4.84 E-02								
	hsa00830	retinol metabolism			2.01	4.84 E-02					2.15	3.66 E-02		
	hsa04144	endocytosis			2.06	4.84 E-02								
	hsa00982	drug metabolism - cytochrome P450			2.07	4.84 E-02								
	hsa00590	arachidonic acid metabolism							-2.54	2.40 E-03				

	hsa00051	fructose and mannose metabolism							-2.50	3.80 E-03	-2.21	2.72 E-02		
	hsa00730	thiamine metabolism							-2.25	1.93 E-02	-2.18	2.72 E-02		
	hsa04714	thermogenesis							-1.98	2.50 E-02				
	hsa00524	neomycin. kanamycin and gentamicin biosynthesis									-2.12	1.59 E-02		
	GO:0009065	glutamine family amino acid catabolic process	2.36	4.64 E-02										
	GO:0044255	cellular lipid metabolic process					-2.59	3.64 E-02						
	GO:0045937	positive regulation of phosphate metabolic process									2.64	3.40 E-02		
	GO:0004623	phospholipase A2 activity	-2.50	2.00 E-02			-2.50	4.19 E-02						
	GO:0005746	mitochondrial respirasome							-2.36	2.80 E-02	-2.21	4.93 E-02		
	GO:1990204	oxidoreductase complex							-2.31	2.80 E-02				
	GO:0045271	respiratory chain complex I							-2.21	3.66 E-02				
	GO:0098800	inner mitochondrial membrane protein complex							-2.57	3.80 E-03	-2.45	1.81 E-02		
	GO:1902494	catalytic complex							-2.13	2.13 E-02				
Absorption processes	hsa04978	mineral absorption	-2.27	1.25 E-02					-2.08	1.94 E-02				
	hsa04974	protein digestion and absorption			-2.26	3.74 E-02								
	GO:0005903	brush border									-2.12	4.93 E-02		
Cellular movement & ciliary processes	hsa04510	focal adhesion	2.32	7.90 E-03										
	hsa04810	regulation of actin cytoskeleton	2.18	2.41 E-02										
	GO:0035082	axoneme assembly	-2.63	3.68 E-02					-2.51	4.04 E-02				
	GO:0060294	cilium movement involved in cell motility	-2.47	3.68 E-02										

	GO:0097722	sperm motility	-2.47	3.68 E-02										
	GO:0060285	cilium-dependent cell motility	-2.46	3.68 E-02										
	GO:0007018	microtubule-based movement	-3.28	4.61 E-05					-3.72	2.45 E-07	-3.06	2.60 E-03	-2.97	3.90 E-03
	GO:0097014	ciliary plasm	-2.93	1.00 E-03					-3.39	1.75 E-06	-2.89	5.00 E-04		
Cellular adhesion & membranes	hsa04514	cell adhesion molecules					2.29	1.62 E-02	-2.14	2.14 E-02				
	GO:0050839	cell adhesion molecule binding	2.68	2.70 E-03									2.36	2.37 E-02
	GO:0031012	extracellular matrix					-3.47	1.04 E-06	3.76	3.16 E-08	2.96	4.00 E-04	3.29	1.06 E-05
	GO:0030312	external encapsulating structure					-2.71	1.30 E-03	2.55	3.90 E-03	2.79	2.10 E-03		
	GO:0045177	apical part of cell					2.50	1.42 E-02	-2.25	2.13 E-02	-2.85	5.00 E-04		
	GO:0098796	membrane protein complex					2.24	3.97 E-02						
	GO:0098590	plasma membrane region							-2.11	2.80 E-02				
	GO:0009897	external side of plasma membrane									2.07	4.93 E-02		
	GO:0031301	integral component of organelle membrane							2.22	3.66 E-02				
	GO:0090559	regulation of membrane permeability											2.48	4.53 E-02
	GO:0009925	basal plasma membrane									-2.13	4.93 E-02		
Gene expression & protein degradation	hsa04141	protein processing in endoplasmic reticulum	2.39	1.11 E-02	2.20	3.77 E-02			1.95	3.19 E-02	2.07	3.66 E-02		
	hsa03050	proteasome					2.37	5.70 E-03	-2.08	1.94 E-02	-2.32	1.59 E-02		
	hsa03010	ribosome					-2.36	1.62 E-02	2.11	3.75 E-02				
	hsa03040	spliceosome							-2.42	5.00 E-03				
	hsa03013	nucleocytoplasmic transport							1.90	4.48 E-02				
	GO:0044389	ubiquitin-like protein ligase binding											-2.72	4.90 E-03

	GO:0015934	large ribosomal subunit					-3.42	4.06 E-06	2.56	9.00 E-03	2.22	4.26 E-02		
	GO:0005840	ribosome					-3.29	1.34 E-05	3.05	3.00 E-04				
	GO:0005829	cytosol					-2.51	7.00 E-03	2.42	1.01 E-02	2.28	4.26 E-02		
	GO:1905369	endopeptidase complex					2.46	1.83 E-02			-2.12	4.93 E-02		
	GO:0140535	intracellular protein-containing complex									2.20	4.93 E-02	-2.44	3.14 E-02
	GO:0045293	mRNA editing complex									2.18	4.20 E-02		
	GO:0016604	nuclear body									2.55	2.10 E-03	-2.39	3.14 E-02
Proliferation & differenciation	hsa04110	cell cycle	2.11	3.10 E-02	2.10	4.20 E-02								
	GO:0007368	determination of left/right symmetry	-2.43	4.64 E-02										
Insulin secretion	hsa04911	insulin secretion			-2.10	3.77 E-02								
	GO:0061178	regulation of insulin secretion involved in cellular response to glucose stimulus							-2.47	2.96 E-02				
Longevity & senescence	hsa04211	longevity regulating pathway							2.13	2.09 E-02				
	hsa04218	cellular senescence	2.06	3.42 E-02										
Angiogenesis. coagulation & blood pressure	GO:0090049	regulation of cell migration involved in sprouting angiogenesis					-3.02	4.10 E-03						
	hsa04614	renin-angiotensin system							2.16	2.76 E-02				
	GO:0030193	regulation of blood coagulation					-2.88	4.10 E-03	3.00	1.00 E-04	2.65	2.09 E-02		

Enriched molecular functions (FDR < 0.05) from gene set enrichment analysis (GSEA) were organized in functional groups. Functions were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG; IDs starting with "-hsa") or Gene Ontology database (IDs starting with "GO"). The normalised enrichment score (NES) and false discovery rate (FDR) are represented for each comparison between the transcriptomically-defined prognosis groups, with red cells indicating functions with up-regulated genes, and blue cells indicating functions with down-regulated genes.

APPENDIX B. Scientific production PhD student

1. Scientific articles

- Sanchez-Reyes, J.M., Parraga-Leo A., Sebastian-Leon, P., Spath K., Vidal C., Devesa-Peiro A., … & Diaz-Gimeno P. (2023). Depicting endometrial function in the mid-secretory phase into four transcriptomic profiles with different prognosis. Pending submitted to American Journal of Obstetrics & Ginecology. IF: 8.661 (Q1).
- Diaz-Gimeno, P., Sebastian-Leon, P., Sanchez-Reyes, J. M., Spath, K., Aleman, A., Vidal, C., ... & Pellicer, A. (2021). Identifying and optimizing human endometrial gene expression signatures for endometrial dating. Human Reproduction, 37(2), 284-296. doi: 10.1093/humrep/deab262. IF: 6.353 (Q1).

2. Book chapters

- Devesa-Peiro A., Sanchez-Reyes J.M, Diaz-Gimeno P. (2020) Chapter 4: Molecular biology approaches utilized in preimplantation genetics: real-time PCR, microarrays, next-generation sequencing, karyomapping, and others. In Garcia-Velasco (1st Ed.) Human Reproductive Genetics. ELSEVIER.
- Sanchez-Reyes J.M y Diaz-Gimeno P. (2020) Capítulo 2: Expresión Génica Endometrial y sus Implicaciones en Medicina Reproductiva. In Espinós-Gomez (1ª Ed.) Avances en Fertilidad. Glosa.

3. Congress communications

- Sanchez-Reyes J.M., Parraga-Leo A., Sebastian-Leon P., Spath K., Vidal C., Gorriz M., Devesa-Peiro A., Remohi J., Wells D., Pellicer A., Diaz-Gimeno P. (2022) A new

endometrial transcriptomic stratification in the mid-secretory phase reveals a new taxonomy for endometrial prognosis in infertility. Presented at 78th ASRM Scientific Congress & Expo, 22-26/10/2022, Anaheim (California). Poster communication.

- Sanchez-Reyes J.M, Parraga-Leo A., Sebastian-Leon P., Spath K., Pellicer A., Remohi J., Wells D., Diaz-Gimeno P. (2022) Endometrial Dysregulated Functions In Recurrent Implantation Failure In *In Vitro* Fertilization Patients After Endometrial Preparation With Hormone Replacement Therapy. Presented at the 69th Annual Scientific Meeting of the Society for Reproductive Investigation, 15-19/03/2022, Denver (Colorado). Poster communication.
- Sanchez-Reyes J.M, Spath K., Sebastian-Leon P., Wells D., Diaz-Gimeno P. (2020) Determination of optimal housekeeping genes for transcriptional study of the endometrium. Presented at the 67th Annual Scientific Meeting of the Society for Reproductive Investigation, 10-14/03/2020, Vancouver (Canada). Poster communication.

4. Funded projects

- Searching for the pathological window of implantation and its therapeutic targets for its clinical translation for precision medicine in reproduction. Instituto de Investigación Carlos III, Spanish Government, Spain. PI19/00537. Principal Investigator: Patricia Díaz Gimeno. Research team: Alejandro Alemán, Patricia Sebastián León, Josefa María Sánchez Reyes and Imma Sánchez Ribas. Duration:1/2020 – 12/2022. 123,420 EUR.

IVI Foundation

IVIRMA Global

VNIVERSITAT DE VALÈNCIA

Instituto de Investigación Sanitaria La Fe