Genómica funcional y disección molecular de FOXD1 para la identificación de nuevos biomarcadores genéticos asociados a patologías de la reproducción de origen endometrial y placentario

Paula Juliana Quintero Ronderos MD

Doctorado en Ciencias Biomédicas
Universidad del Rosario
2018

Tesis de Doctorado

Doctorado en Ciencias Biomédicas

Genómica funcional y disección molecular de *FOXD1* para la identificación de nuevos biomarcadores genéticos asociados a patologías de la reproducción de origen endometrial y placentario

Estudiante
Paula Juliana Quintero Ronderos MD

Director de tesis
Paul Laissue MD, MSc, PhD, HDR
Director del Centro de Investigación en Genética y Genómica de la Universidad del Rosario (CIGGUR)
Profesor Titular Escuela de Medicina y Ciencias de la Salud, Universidad del Rosario

Universidad del Rosario
Doctorado en Ciencias Biomédicas
2018

Aprobado por el comité evaluador en cumplimiento de los requisitos exigidos para otorgar al título de Doctor en Ciencias Biomédicas de la Universidad del Rosario

_____

Diego Forero, MD, PhD
Facultad de Medicina
Universidad Antonio Nariño

_____

Gustavo Andrés Salguero, MD, PhD
Unidad de Terapia Celular
Hemocentro Distrital de Bogotá

_____

Camilo Salazar, MSc, PhD
Facultad de Ciencias Naturales y Matemáticas
Universidad del Rosario

Contenido

Lista de tablas V
Lista de figuras VI
Lista de anexos VIII
Lista de abreviaturas IX

Resumen 1
Summary 3
1. Introducción general 5
2. Capítulo I: Identificación de nuevas variantes potencialmente asociadas a la etiología del AER por NGS 10
2.1 Introducción 11
2.1.1 Aspectos moleculares de la implantación y el AER 13
2.1.2 Secuenciación de Sanger y NGS de genes candidato del AER 15
2.2 Metodología 19
2.2.1 Inclusión de casos no-familiares de AER idiopático y de controles 20
2.2.2 Preparación de la librería y secuenciación por NGS del exoma 21
2.2.3 Análisis bioinformático y depuración de los datos 22
2.2.4 Creación del subset de genes candidato 23
2.2.5 Análisis bioinformático del subset de genes 24
2.2.6 Análisis de las variantes candidatas 25
2.2.6.1 Secuenciación directa de las variantes candidatas 25
2.2.6.2 Análisis de fragmentación molecular orbital (FMO) de la estructura de proteínas con variantes candidatas 25
2.2.6.3 Secuenciación de la región codificante de *THBD* en pacientes colombianas con AER 26
2.2.6.4 Análisis bioinformático de las variantes de *THBD* 27

2.2.6.5 Análisis funcional *in vitro* de las variantes candidatas 28

2.2.6.5.1 Estudio de la expresión de THBD WT y mutante en células THP1 28

2.2.6.5.2 Estudio de la localización subcelular de THBD WT y mutante. 29

2.2.6.5.3 Estudio del efecto de la mutación THBD-p.Trp153Gly en la expresión de genes de la cascada corriente abajo de la citoquina HMGB1 30

2.2.6.5.4 Modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD..... 31

2.2.6.5.5 Estudio genético y funcional de las variantes de *EPAS1* 32

2.2.6.5.5.1 Inclusión de pacientes no relacionados con AER idiopático y de controles 32

2.2.6.5.5.2 Secuenciación directa de los exones 10, 11 y 12 del gen *EPAS1* 33

2.2.6.5.5.3 Análisis bioinformático de las variantes de *EPAS1*..... 33

2.2.6.5.5.4 Estudio gen reportero (bajo condiciones de hipoxia) de las mutaciones p.Leu504Val y p.Tyr488Cys de EPAS1..... 33

2.3 Resultados..... 34

2.3.1 Secuenciación NGS y de Sanger para la búsqueda de nuevas variantes de AER idiopático y análisis bioinformático 34

2.3.2 Fragmentación molecular orbital (FMO) de mutaciones en proteínas candidatas 39

2.3.3 Secuenciación directa y ensayos funcionales de las variantes de THBD 43

2.3.3.1 Secuenciación de la región codificante de *THBD* en pacientes colombianas con AER 43

2.3.3.2 Estudio de la expresión de *THBD* WT y mutante en células THP-145

2.3.3.3 Estudio de la localización subcelular de THBD WT y mutante 47

2.3.3.4 Estudio del efecto de la mutación THBD p.Trp153Gly en la expresión de genes de la cascada corriente abajo de la citoquina HMGB1 52

3.3.3.5 Modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD 55

2.3.4 Secuenciación directa y ensayos funcionales de las mutaciones de *EPAS1* 58

2.4 Discusión capítulo I 60

2.4.1 Abordaje NGS para la búsqueda de mutaciones en pacientes con AER 60

2.4.2 Mutaciones candidatas en la etiología del AER 63

2.4.3 Análisis *in silico* del potencial efecto deletéreo de las mutaciones candidatas de AER 64

2.4.4 Tamizaje de la mutación THBD p.Trp153Gly en la población colombiana afectada por AER 65

2.4.5 Estudio de la localización subcelular de THBD WT y mutante 66

2.4.6 Estudio del efecto de la mutación THBD p.Trp153Gly en la expresión de genes de la cascada corriente abajo de la citoquina HMGB1 67

2.4.7 Modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD. 69

2.4.8 Tamizaje de las mutaciones de *EPAS1* y estudio funcional in vitro 70

3. Capítulo II: *Estudios moleculares y funcionales del gen FOXD1* 72

3.1 Introducción 73

3.2 Metodología 78

3.2.1 Inclusión de pacientes y de controles con FIR, AER, PE, Y RCIU 80

3.2.2 Secuenciación directa del gen FOXD1 81

3.2.3 Preparación de la muestra de placenta para experimentos de ChIP-seq 81

3.2.4 ChIP-seq 82

3.2.5 Análisis bioinformático de la secuenciación y depuración de los datos . 83

3.2.6 Análisis de los picos encontrados por ChIP-seq 83

3.2.7 Ensayos gen reportero 84

3.3 Resultados 85

3.3.1 Estudio de *FOXD1* como un potencial marcador molecular de AER, en FI durante FIV (ART), RCIU y PE 85

3.3.2 Ensayo gen reportero 87

3.3.3 Identificación de nuevos genes blanco directos de FOXD1 88

3.4 Discusión capítulo II 93

3.4.1 Estudio de FOXD1 como un potencial marcador molecular de AER, en FI durante FIV (ART), RCIU y PE 93

3.4.2 Identificación de nuevos genes blanco directos de FOXD1 95

4. Conclusiones 98

5. Recomendaciones y perspectivas 99

6. Anexos 100

7. Bibliografía 105

Lista de tablas

Tabla 1. Etiología y prevalencia en el AER ……..……..….. 6
Tabla 2. Etiología y prevalencia en la FIR …..……..……..….. 8
Tabla 3. Principales moléculas implicadas en el proceso de implantación ……13
Tabla 4. Variantes asociadas al AER usando secuenciación de Sanger ……16
Tabla 5. Genes asociados a AER encontrados por la técnica de NGS ……18
Tabla 6. Variantes nuevas identificadas en genes candidato mediante NGS y confirmadas por secuenciación directa en pacientes con AER ……36
Tabla 7. Predictores de patogenicidad para las variantes confirmadas por secuenciación de Sanger ……...……37
Tabla 8. Características clínicas de las pacientes AER francesas y colombianas con variantes nuevas identificadas en genes candidato mediante NGS y confirmadas por secuenciación directa ……38
Tabla 9. Secuenciación del ORF de *THBD* en pacientes con AER, FIR e individuos control ……44
Tabla 10. Categorías encontradas por IFI en células CHO transfectadas con la versión WT y mutante de THBD ……48
Tabla 11. Categorías encontradas por IFI en células THP-1 transfectadas con la versión WT y mutante de THBD ……48
Tabla 12. Secuenciación directa de exones 10, 11 y 12 de *EPAS1* en pacientes francesas con AER ……59
Tabla 13. Construcciones realizadas con las mutaciones de FOXD1 ……84
Tabla 14. Resultados de la secuenciación del ORF de *FOXD1* en pacientes colombianas con AER, FIR e individuos control ……86

Lista de figuras

Figura 1. Principales moléculas implicadas en el proceso de implantación …….12
Figura 2. Metodología para la identificación de nuevas variantes potencialmente asociadas al AER y validación funcional ……21
Figura 3. Filtrado bioinformático de los datos de la secuenciación del exoma ……34
Figura 4. Resultados de FMO para FGA WT y mutante (p.Phe685Cys) ……41
Figura 5. Resultados de FMO para MMP-10 WT y mutante (p. Asp199Asn) ……42
Figura 6. Expresión endógena de THBD en la línea celular THP-1 ……45
Figura 7. Citometría de flujo de células THP-1 transfectadas con la versión THBD WT y mutante ……46
Figura 8. Perfiles de localización subcelular de la línea celular CHO transfectada con las construcciones de THBD WT y mutante (THBD- p.Trp153Gly) ……49
Figura 9. Perfiles de localización subcelular de la línea celular COS-7 transfectada con las construcciones de THBD WT y mutante (THBD- p.Trp153Gly) ……50
Figura 10. Perfiles de localización subcelular de la línea celular THP-1 transfectada con las construcciones de THBD WT y mutante (THBD- p.Trp153Gly) ……51
Figura 11. Expresión endógena del receptor RAGE en células THP-1 ……52
Figura 12. Expresión de los genes que participan en la cascada pro-inflamatoria de HMGB1 en células THP-1 transfectadas con la versión WT y mutante de THBD …54
Figura 13. Alineamiento de los dominios lectina analizados por BLASTp ……55
Figura 14. QMEAN4 normalizado ……..……56
Figura 15. Modelo por homología del dominio de lectina de THBD ……56
Figura 16. Mapas de potencial electrostático de los modelos WT y MT ……57
Figura 17. Distribución de las energías calculadas para los modelos THBD WT y THBD p.Trp153Gly ……58
Figura 18. Diagrama del gen *EPAS1* y de las mutaciones encontradas en el exoma y por secuenciación directa ……59
Figura 19. Resultados de la estandarización ensayo gen reportero de EPAS1 WT y mutantes con el promotor de *INHA* ……60

Figura 20. Modelo murino IRCS. Se realizó un cruce inicial entre una hembra C57BL/6J y un macho SEG/Pas (*Spretus*) ……74
Figura 21. Metodología para el estudio molecular y funcional de FOXD1 ……79
Figura 22. Ensayo gen reportero de FOXD1 WT y mutantes con el promotor de *C3* ……..88
Figura 23. Expresión de FOXD1 en tejido placentario sano a término ……89
Figura 24. Gráfica de distribución del conteo de los picos vs. valor de enriquecimiento ……90
Figura 25. Gráfica de distribución del conteo de los picos vs. valor de alineamiento ……..90
Figura 26. Gráfica de localización de los picos ……91
Figura 27. Gráfica de la distribución de la identificación de los genes blanco …….92
Figura 28. Ensayo gen reportero de FOXD1 WT con el promotor de *CD86* y *CTSC* ……...93

Lista de anexos

Anexo I. ***Quintero-Ronderos P***, Laissue P. *"Classic and novel sequencing methods to dissect recurrent pregnancy loss aetiology"*. Reproductive Science. Submitted. Under review ……100

Anexo II. ***Quintero-Ronderos P***, Mercier E, Fukuda M, González R, Suárez CF, Patarroyo MA, Vaiman D, Gris JC, Laissue P. *"Novel genes and mutations in patients affected by recurrent pregnancy loss"*. PLoS One. 2017 Oct 10;12(10):e0186149 ……..……101

Anexo III. ***Quintero-Ronderos P***, Mercier E, Gris JC, Esteban-Perez C, Moreno-Ortiz H, Fonseca DJ, Lucena E, Vaiman D, Laissue P. *"THBD sequence variants potentially related to recurrent pregnancy loss"*. Reprod Biol Endocrinol. 2017 Dec 1;15(1):92 ……..……...102

Anexo IV. ***Quintero-Ronderos P***, Laissue P*. "The multisystemic functions of FOXD1 in development and disease".* J Mol Med (Berl). 2018 Aug;96(8):725-739 ……..…….103

Anexo V. ***Quintero-Ronderos P***, Jiménez K, Esteban-Pérez C, Ojeda D, Bello S, Fonseca DJ, Coronel MA, Moreno-Ortiz H, Sierra-Díaz DC, Lucena E, Barbaux S, Vaiman D, Laissue P. *"FOXD1 mutations are related to repeated implantation failure, intra-uterine growth restriction and preeclampsia"*. Human Mutation. Submitted. Under review ……104

Lista de abreviaturas

AER: aborto espontáneo recurrente
AG: agregados-like
APBS: Adaptive Poisson-Boltzmann Solver
ART: assisted reproductive techniques - Técnicas de reproducción asistida
CGI: grado de intensidad para ChIP
ChIP-seq: secuenciación de la inmunoprecipitación de la cromatina
CMF: campo mutualmente consistente
COSMO: conductor-like screening model
CTLD: C-type lectin domain – dominio de tipo lectina
DFO: deferroxamina
EMSA: electrophoretic mobility shift assay
EPCR: receptor endotelial para la proteína C
EPO: eritropoyetina
eRNAs: enhancers RNAs
EU: epitelio uterino
FIR: falla de implantación recurrente
FIV: fertilización *in vitro*
FKHD: familia forkhead box
FMO: fragmentación molecular orbital
GWAS: estudios de asociación de genoma completo
HIF-1a: factor inducible de hipoxia 1a
ICSI: inyección intracitoplasmática de espermatozoides
IFI: inmunofluorescencia indirecta
IOP: insuficiencia ovárica premature
IRCS: interspecific recombinant congenic strain
IUGR: intrauterine growth restriction
IVF: *in vitro* fertilization
LIF: factor inhibitorio de la leucemia
lncRNAs: long non-coding RNAs – ARN largo no codificante
MAF: minor allele frequency – frecuencia alélica mínima
MCD: membrana citoplasmática delimitada
MMPs: metaloproteinasas
MT: mutante
MTHFR: metilen-tetrahidrofolato reductasa
NGS: secuenciación de siguiente generación
ns: nanosegundos
ODDD: dominio de degradación dependiente de oxígeno
OPN: osteopontina

ORF: open reading frame – marco abierto de lectura

pb: pares de bases

PE: preeclampsia

PTEN: homólogo de la tensina

RCIU: retardo del crecimiento intrauterino

RIF: recurrent implantation Failure

RMSD: root-mean-square deviation

RNA-seq: secuenciación de ARN

ROS: especies reactivas de oxígeno

RPL: recurrent pregnancy loss

SAF: síndrome antifosfolípido

SNPs: single nucleotide polymorphisms

TFBS: sitios potenciales de unión a factores de transcripción

THBD: trombomodulina

TSS: sitios de inicio de la transcripción

Tr: trofoblasto

Trd: trofoectodermo

uNK: células natural killer uterinas

UTRs: untranslated regions

WB: Western Blot

WES: whole-exome Sequencing – secuenciación de exoma completo

WT: wild type

X-Pol: teoría explicita de polarización

Resumen

El aborto espontáneo recurrente (AER) se define como dos o más pérdidas consecutivas de la gestación antes de la semana 20 del desarrollo intrauterino. Esta patología afecta a aproximadamente entre el 1% y el 5% de las parejas. La etiología del AER se puede dividir en causas no-genéticas como genéticas. Sin embargo, ~50% de los casos se considera idiopático. De manera análoga, la etiología molecular de la falla de implantación recurrente (FIR), definida como la falla de la implantación en al menos 2 o más ciclos consecutivos de fertilización *in vitro*, es poco conocida. La etiología molecular del AER y de la FIR está asociada potencialmente a variantes de secuencia en cientos de genes candidato que participan en las cascadas moleculares fisiológicas de la implantación y durante toda la gestación. La aproximación gen candidato, usando la secuenciación de Sanger, ha sido de utilidad para la descripción de pocos genes implicados en el AER. La secuenciación de siguiente generación (NGS) ha sido una herramienta eficiente puesto que permite el estudio simultáneo de múltiples genes relacionados con enfermedades complejas.

En la primera parte de este trabajo de tesis se utilizó la aproximación NGS-exoma para identificar nuevos genes y mutaciones potencialmente implicadas en el desarrollo del AER. Algunas de las mutaciones encontradas por este abordaje fueron estudiadas mediante ensayos funcionales *in vitro* para determinar su posible efecto deletéreo. La identificación de la variante THBD p.Trp153Gly en mujeres colombianas con AER y su validación mediante ensayos funcionales *in vitro* sugiere, por primera vez, una relación directa entre formas mutantes de esta proteína y la fisiopatología del AER, considerándose como un posible marcador molecular para el diagnóstico en pacientes colombianas con AER idiopático.

En la segunda parte del presente trabajo de tesis identificamos y estudiamos funcionalmente nuevas mutaciones de *FOXD1*, un gen relevante en la fisiología endometrial y placentaria, identificadas en pacientes FIR, AER, preeclampsia (PE)

y retardo del crecimiento intrauterino (RCIU). Los ensayos funcionales *in vitro* demostraron que las mutaciones FOXD1-p.His267Tyr y FOXD1-p.Arg57del. modifican la transactivación del promotor de *C3*, contribuyendo con el fenotipo. FOXD1 podría considerarse en consecuencia un marcador molecular diagnóstico para las pacientes con AER, FIR, RCIU y PE.

Por último, ensayos por inmunoprecipitación de la cromatina secuenciación NGS (ChIP-seq) permitieron determinar potenciales nuevos genes blanco directos de FOXD1 (CTSC, CD86, CMA1 y TRPC6) en un contexto placentario.

La información generada durante este trabajo de tesis aporta al conocimiento sobre el origen genético del AER, la FIR, y el RCIU/PE. Estos resultados podrían ser de utilidad para los especialistas clínicos en el contexto del desarrollo de la medicina traslacional.

Summary

Recurrent pregnancy loss (RPL) is defined as the loss of two or more consecutive and spontaneous miscarriages before the 20th week of gestation. This pathology affects ~1% to 5% of the couples. The RPL aetiology is classified as non-genetics and genetics. However, the 50% of the cases remains idiopathic. Similarly, the aetiology of the recurrent implantation failure (RIF), defined as the implantation failure in at least 2 or more consecutive cycles of *in vitro* fertilization (IVF), is poorly understood. To note, the molecular RPL and RIF aetiology is potentially associated with hundreds of genes which participate in the physiological molecular pathways related to the implantation and gestation. The candidate gene approach using Sanger sequencing has been useful to identify some genes associated with RPL. However, the next generation sequencing (NGS) has been an effective tool to overcome this limitation because it allows the simultaneous study of multiple genes related to complex diseases.

In first part of this thesis, we used the NGS-exome approach to identify new genes and mutation potentially implicated in the development of RPL. Some of the mutations found by this approach were tested by functional *in vitro* assays to determine their possible deleterious effect within the pathology. The identification of the variant THBD p.Trp153Gly in Colombian women with RPL and its validation through functional *in vitro* assays suggested, for the first time, a direct association with the RPL pathophysiology. Therefore, it may be considered as a molecular biomarker for the RPL diagnosis in Colombian patients.

In the second part of this thesis, we identified and studied the potential implication of new FOXD1 mutations, a relevant gene in the endometrium and placenta physiology, in patients with RIF, RPL, preeclampsia (PE) and intrauterine growth restriction (IUGR). The functional *in vitro* assays demonstrated that FOXD1 p.His267Tyr and FOXD1 p.Arg57del mutations modify the *C3* promoter transactivation, thus contributing to the phenotype. Therefore, FOXD1 could be

considered a molecular biomarker for the diagnosis of RPL, RIF, PE and IUGR patients.

The chromatin immunoprecipitation assays (ChIP) allowed to determine new direct FOXD1 target genes (CTSC, CD86, CMA1 y TRPC6) within the placenta context.

The information generated during this thesis contributes to the general knowledge about the origin and development of RPL, RIF, PE/IUGR. These results might be useful for the clinical specialist in a context of translational medicine.

1. Introducción general

El AER se define como dos o más (tres o más según las guías europeas) abortos espontáneos y consecutivos antes de la semana 20 de gestación [1,2]. Esta patología afecta a aproximadamente entre el 1% y el 5% de las parejas [3–5]. El AER constituye un problema de salud pública por su prevalencia, y por los problemas sociales y psicológicos que puede generar tanto en la mujer como en la pareja.

La etiología del AER se puede dividir en causas no-genéticas como genéticas. Sin embargo, el 50% de los casos se considera idiopático. Varias causas no - genéticas de AER se han descrito, las cuales pueden dividirse en diferentes grupos: asociadas a las anomalías estructurales uterinas, a infecciones, a defectos endocrinológicos, a alteraciones en el sistema inmune y al síndrome antifosfolípido (SAF). Una de las causas no - genéticas más comunes del AER es el SAF (15% – 20%) (**Tabla 1**), una patología autoinmune en la que se promueve un estado de hipercoagulabilidad que genera la trombosis y el infarto de los vasos uterinos y placentarios [2,6,7]. Además, se ha demostrado que los autoanticuerpos generados en esta enfermedad interfieren con la decidualización alterando el desarrollo y la funcionalidad del trofoblasto [8,9].

Alrededor del 1% al 5% de las mujeres con AER presentan anormalidades en la morfología uterina como el útero septado, el útero bicorne y los miomas. Estas anomalías se consideran causales de la enfermedad puesto que no permiten la implantación adecuada del blastocisto al epitelio endometrial [2,10–13]. Las alteraciones endocrinológicas también pueden ocasionar la aparición de AER (8% – 12%), como en el caso de la hiperprolactinemia y los ovarios poliquísticos por resistencia a la insulina [14–18]. Otra de las etiologías de AER que se ha descrito, en el <5% de los casos, es la disfunción inmunológica.

Tabla 1. Etiología y prevalencia en el AER	
Etiología	**Prevalencia (%)**
• Idiopática	50
• Síndrome antifosfolípido	15 - 20
• Alteraciones endocrinológicas	8 - 12
• Vaginosis bacteriana	15
• Cromosomopatías • Alteraciones inmunológicas • Anormalidades uterinas • Trombofilias	< 5

Durante el período de gestación debe haber una tolerancia materna hacia el feto en crecimiento. Varios estudios demostraron que mujeres con un aumento en el número de uNK y de la respuesta inmune tipo Th1 presentaban una mayor tasa de pérdidas de la gestación (5%) en comparación con el grupo de mujeres control generando el rechazo del tejido endometrial hacia el embrión en crecimiento [19–21]. En lo que respecta a la asociación entre las infecciones vaginales y el AER (15%) son hoy en día muy discutidas. Se ha reportado que la presencia de vaginosis bacteriana durante el primer trimestre de gestación aumenta el riesgo de AER y de nacimiento prematuro pues genera un ambiente pro-inflamatorio que interfiere con el adecuado proceso de implantación [22–25].

Las causas genéticas del AER, las cuales representan un 3% - 5% de los casos, se han asociado principalmente a alteraciones cromosómicas como las aneuploidías fetales, especialmente en el primer trimestre del embarazo [26,27]. Se han descrito alteraciones citogenéticas como las trisomías (13, 14, 15, 16, 21 y 22), la monosomía del cromosoma X y las traslocaciones no balanceadas [26,28–35]. La inactivación anormal del cromosoma X, que se genera durante la embriogénesis temprana también se ha propuesto como una causa genética del AER [36]. Se ha sugerido que las causas de esta inactivación preferencial sean mutaciones en el

gen *XIST*, las alteraciones crípticas del cromosoma X como traslocaciones o microdeleciones y las fallas durante el proceso de impronta genética [1,37].

Los polimorfismos en el locus *HLA-G* se han propuesto adicionalmente como factores de riesgo para el desarrollo de AER. Estudios demostraron que la presencia de los alelos HLA-G*0104, HLA-G*0105N y HLA-G*01013 aumentaban significativamente el riesgo de AER [38,39]. Mutaciones en genes asociados a defectos trombofílicos como las variantes c.1691 G>A del Factor V de coagulación (FV), c.20210 G>A del Factor II (protrombina) y la c.677 C>T en la enzima metilen-tetrahidrofolato reductasa (*MTHFR*) se asociaron estadísticamente al AER en un bajo porcentaje de los casos [2,40–42]. Recientemente, nuestro grupo describió las mutaciones de *FOXD1* como una causa importante del AER [43].

Aproximadamente entre el 20% y el 25% de las parejas fértiles consiguen un embarazo durante un único ciclo menstrual [44]. Para tratar la infertilidad se desarrollaron las técnicas de reproducción asistida (ART, Assisted Reproduction Treatment), las cuales son menos eficientes que la reproducción humana en condiciones fisiológicas. Las tasas de implantación son del 20% al 30% en ART y tienden a disminuir después del tercer tratamiento fallido [45,46]. La falla de implantación recurrente (FIR) se define como la falla de la implantación en al menos 2 o más ciclos consecutivos de fertilización *in vitro* (FIV) o inyección intracitoplasmática de espermatozoides (ICSI) en la que un embrión fresco o congelado de alta calidad es transferido en cada ciclo con niveles séricos negativos de β-HCG [44,47–52].

La FIR se puede clasificar, dependiendo de su etiología, en multifactorial, endometrial o idiopática (3%). Las causas de FIR multifactorial incluyen los factores anatómicos uterinos (40%), los factores masculinos (40%), las alteraciones cromosómicas del embrión, las alteraciones hormonales, infecciones, trombofilias y algunos factores inmunológicos. Éstas últimas corresponden al 10% de los casos (**Tabla 2**). La FIR idiopática, que corresponde al 3% de los casos está definida por la ausencia de evidencia de alteraciones anatómicas o histológicas en el endometrio, anomalías en los factores masculinos o en los embriones [49,53].

Tabla 2. Etiología y prevalencia en la FIR	
Etiología	**Prevalencia (%)**
• Alteraciones endometriales	40
• Factores masculinos	40
• Cromosomopatías del embrión • Alteraciones hormonales • Infecciones • Trombofilias • Alteraciones inmunológicas	10
• Idiopática	3

Es importante señalar que la etiología molecular del AER y de la FIR está asociada potencialmente a cientos de genes que participan en las cascadas moleculares fisiológicas de la implantación y durante toda la gestación [2,54–56].

En el caso del AER, para la identificación de nuevas moléculas potencialmente etiológicos de la enfermedad, se han usado varias aproximaciones como el análisis de ligamiento genético, los estudios de asociación de genoma completo (GWAS) y la secuenciación directa (Sanger) de genes candidato [57–62].

La aproximación gen candidato, usando la secuenciación de Sanger, ha sido de utilidad para la descripción de algunos de los genes implicados en el AER. Sin embargo, al tratarse de una patología compleja, esta técnica no es costo-efectiva para estudiar numerosos genes simultáneamente, sino que se enfoca en el estudio de regiones genómicas cortas o en un grupo pequeño de genes. La secuenciación de siguiente generación (NGS) ha sido una herramienta efectiva para sobrepasar algunas de estas limitaciones puesto que permite el estudio simultáneo de múltiples genes relacionados con enfermedades complejas [63–65].

Es importante destacar que en un estudio reciente efectuado por abordajes de QTL, de secuenciación y por ensayos funcionales *in vitro* se describió el gen *FOXD1* que codifica para un factor de transcripción, como un actor fundamental en la

implantación de roedores y de humanos, y potencialmente relevante en la fisiología placentaria [43]. Sus mutaciones fueron demostradas como causales del AER. Sin embargo, se conocen pocos genes blanco directos y se desconoce si las pacientes con FIR y patologías placentarias (preeclampsia y retraso de crecimiento intrauterino - RCIU). presentan mutaciones en FOXD1 que puedan estar asociadas a la enfermedad.

En este proyecto de tesis se utilizó la aproximación NGS-exoma para identificar nuevos genes y mutaciones potencialmente implicadas en el desarrollo del AER (**Capítulo 1**). Algunas de las mutaciones encontradas por este abordaje fueron estudiadas mediante ensayos funcionales *in vitro* para determinar su posible efecto deletéreo en la patología del AER.

Identificamos y estudiamos además la potencial implicación de nuevas mutaciones de *FOXD1* en pacientes colombianas con FIR, AER y, en pacientes francesas con preeclampsia y RCIU (**Capítulo 2**). Algunas variantes fueron evaluadas por ensayos funcionales *in vitro* buscando establecer nuevos marcadores moleculares que puedan ayudar al diagnóstico y al manejo de las mujeres afectadas. También se realizaron ensayos para determinar nuevos genes blanco directos de FOXD1.

2. Capítulo I: *Identificación de nuevas variantes potencialmente asociadas a la etiología del AER por NGS*

2.1 Introducción

La implantación en los mamíferos es el proceso mediante el cual el endometrio permite el paso de nutrientes y oxígeno para el adecuado crecimiento del blastocisto durante la gestación. El proceso desde la fertilización hasta la implantación se ha conservado evolutivamente entre las diferentes especies de mamíferos [66]. El período de pre-implantación es variable entre especies, por ejemplo, en los ratones ocurre durante los 4 días siguientes al coito, mientras que en los humanos ocurre 9 días después [67]. Es importante señalar que la implantación (y sus mecanismos) es diferente entre algunas especies de mamíferos y se puede dividir en tres tipos: extrínseca, central e intersticial. La implantación excéntrica ocurre cuando el epitelio luminal rodea al trofoblasto. Se presenta en mamíferos como los ratones. La implantación central sucede cuando el blastocisto crece hasta fusionarse con el epitelio sin penetrarlo. Se presenta en los bovinos, el cerdo, las ovejas y los conejos. Por último, la implantación intersticial ocurre cuando el trofoblasto pasa por el epitelio hasta invadir al estroma uterino como ocurre en los humanos [66,68]. El proceso de implantación en los humanos se puede dividir en 4 fases: la preparación del endometrio, la adhesión del blastocisto, la ruptura de la barrera epitelial y la invasión del blastocisto.

La complejidad molecular del AER está asociada a la potencial participación de cientos de genes pertenecientes a las diversas cascadas moleculares de la implantación (**Figura 1** y **Tabla 3**) [2,54–56].

A continuación, se describirán algunos de los procesos fisiológicos y las moléculas relacionadas con la implantación. Por su extensión, solo se mencionarán algunas de ellas. Posteriormente, se revisarán algunas de las aproximaciones que se han usado para el estudio de las moléculas implicadas en la implantación y el AER.

Figura 1. Principales moléculas implicadas en el proceso de implantación. **(A)** Las cascadas de señalización de LIF y el estrógeno durante la receptividad uterina. **(B)** La señalización de las integrinas durante la implantación y la expresión de quemoquinas en la aposición y la adhesión. **(C)** Las vías FasR/FasL y progesterona durante la apoptosis de las células endometriales. **(D)** Durante la decidualización se sobre-expresan proteínas relacionadas con la homeostasis del flujo sanguíneo como REN, AGT, IGF-1, TNF-α, TGF-α/β, BMP-2, ACE, IHH, PDGF, (P)RR, TIMPs, PAI-1 y THBD. **Tr:** trofoblasto. **EU:** epitelio uterino.

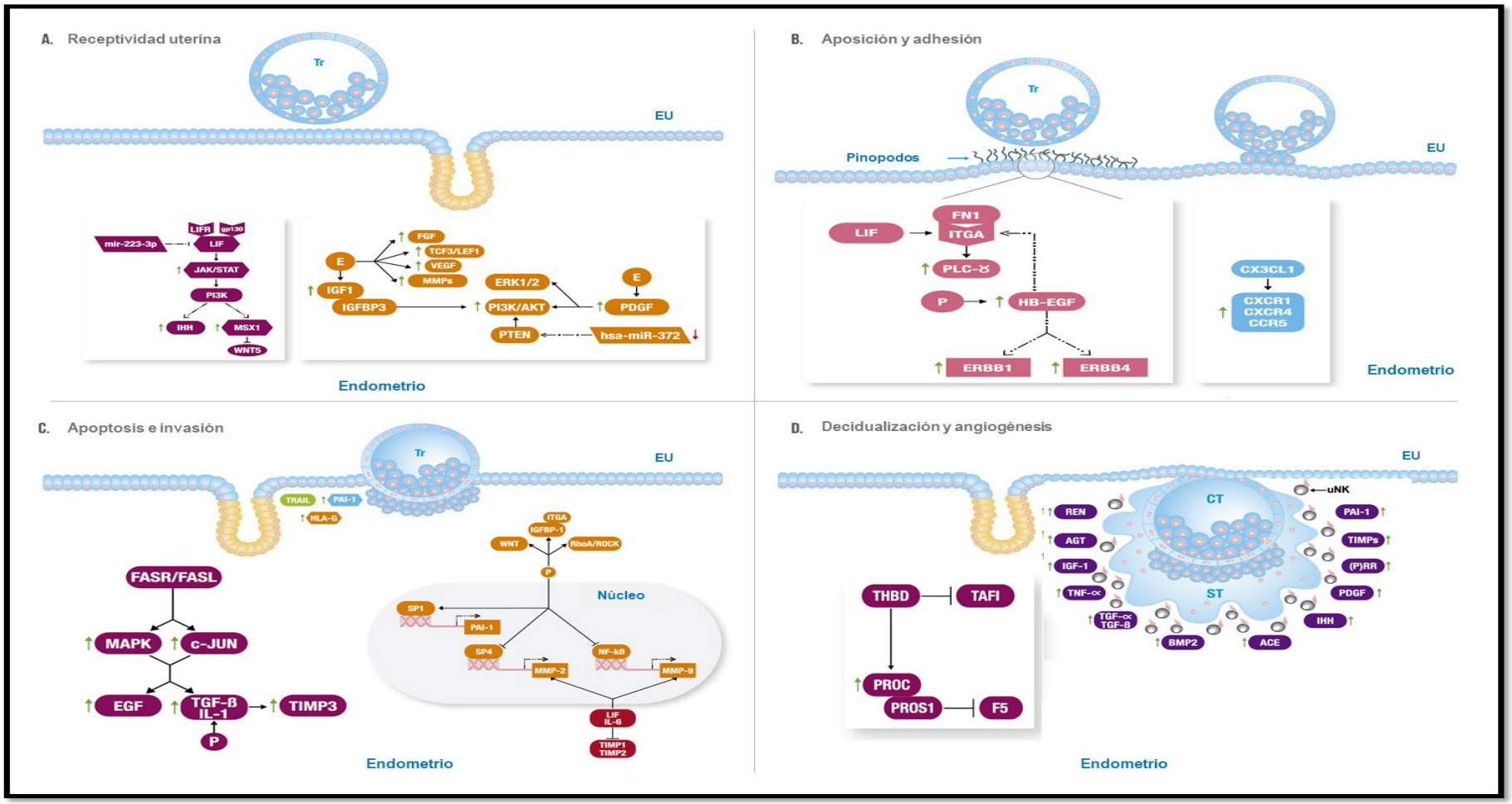


Tabla 3. Principales moléculas implicadas en el procesos de implantación	
Mólecula	**Función biológica**
Progesterona	• Diferenciación células estromales en deciduales • Expresión *IGFBP-1*, la prolactina, el factor tisular y citoquinas
Estrógeno	• Proliferación del endometrio y un aumento en la angiogénesis • Supervivencia celular • Inhibición de PTEN
IHH, MSX1, WNT5A, KLF5, cPLA2α, LIF, FGF-ERK, STAT3, AKT, BCL-2, IGF-I	• Diferenciación, proliferación y receptividad del epitelio
EGF, TGF-α, PDGF, FGF, MMPs, MIF	• Remodelación tisular
Integrinas, OPN, HB-EGF, LIF, HER3, HER4	• Formación de pinopodos
L-selectina, mucinas, integrinas, TGF-β, Fas-ligando, PPARγ	• Adhesión del blastocisto al endometrio
COX2, ADM, WNT4, BMP2, IL-11, HOXA10, HOXA11	• Decidualización
THBD, F5, FGA, FLT1, EPAS1, VEGF, NOS2, PGF, JAK2, EPCR	• Coagulación y angiogénesis
Citoquinas, IDO2, LIFR, TLR3	• Migración de monocitos y uNK • Disminución respuesta Th1 • Aumento respuesta Tregs

2.1.1 Aspectos moleculares de la implantación y el AER

Durante la implantación, las células endometriales sufren cambios morfológicos, metabólicos, bioquímicos y moleculares para que el tejido sea receptivo al blastocisto. Estos cambios son controlados fundamentalmente por las hormonas sexuales esteroideas: el estrógeno y la progesterona, las cuales son fundamentales para la activación de genes relacionados con los procesos de diferenciación, proliferación y receptividad del epitelio [69,70]. Durante el proceso de decidualización, aumenta la expresión de numerosos genes como *LIF, IGFBP-1, WNT4*, *HOXA10*. La expresión de citoquinas durante este proceso aumenta la migración de las células inmunes, como los monocitos y las células natural killer uterinas (uNK), hacia el endometrio para permitir la receptividad y la tolerancia al blastocisto. Además, a nivel inmunológico, disminuyen las moléculas de las vías de señalización asociadas a una respuesta Th1 mientras que aumenta la de las células Tregs [70–77].

Además de los cambios morfológicos celulares, durante las fases proliferativa y secretora, se produce la expresión diferencial de un gran número de genes que son regulados por factores intrínsecos y extrínsecos de cada célula [78–80]. El estrógeno durante la implantación produce un engrosamiento (proliferación) del endometrio, un aumento en la angiogénesis, mejora de la supervivencia celular e inhibe la expresión de genes apoptóticos como el homólogo de la tensina (*PTEN*) [81]. Adicionalmente, se observa un aumento en la expresión de genes asociados a la remodelación tisular y a la angiogénesis como las metaloproteinasas (MMPs) y *VEGF* [82–84].

Los cambios ocurridos durante la decidualización son los que permiten principalmente la receptividad del endometrio al blastocisto para iniciar el proceso de implantación. En esta fase, se produce un aumento en la expresión de los genes que permiten la formación de pinopodos para facilitar el proceso de invasión y la adhesión del blastocisto, lo que promueve la producción de integrinas, de la osteopontina (*OPN*) y del factor inhibitorio de la leucemia (*LIF*), entre otros (**Figura 1**) [82,85,86].

Para que el proceso de adhesión sea exitoso se genera una unión fuerte entre el trofoectodermo (Trd) y las células deciduales, por lo que se necesita de la actividad conjunta de numerosas moléculas (e.g. mucinas, integrinas, Fas-ligando) que permiten la ruptura de la barrera epitelial [87]. Además, se inicia la comunicación vascular entre el endometrio y el Trd, la cual es realizada por proteínas asociadas a los procesos de coagulación y angiogénesis como THBD y EPAS1. THBD es una proteína fundamental para la implantación puesto que actúa en la activación de la proteína C, la cual a su vez genera un ambiente anticoagulante y anti-inflamatorio (dominio tipo lectina) [88,89]. En pacientes con AER, se han evidenciado modificaciones en la función de esta proteína asociadas a alteraciones en la vía de la proteína C al igual que un estado de hipercoagulabilidad [90]. Es importante precisar que durante este trabajo de tesis se efectuaron experimentos relacionados con esta proteína.

2.1.2 Secuenciación de Sanger y NGS de genes candidato del AER

La aproximación gen candidato seguida de secuenciación de Sanger ha sido de mediana utilidad para la genotificación de genes implicados en el AER. Al día de hoy, se ha estudiado un número limitado de regiones codificantes de genes involucrados en el proceso de implantación (e.g. *JAK2, THBD, EPCR, AMN*) [61,62,91]. Sin embargo, una cantidad considerable de genes y de variantes frecuentes (SNPs) han sido asociados con el fenotipo. Únicamente un pequeño grupo de variantes ha sido validado por ensayos funcionales (**Tabla 4**).

Se ha evidenciado que las pacientes con desórdenes trombofílicos poseen una mayor frecuencia de abortos espontáneos, en especial aquellas con policitemia vera (22%) y trombocitemia esencial (31%) [61,92]. Un estudio de la región codificante de *JAK2* por secuenciación de Sanger demostró que la mayoría de pacientes (estudio en 103 pacientes) con desórdenes trombofílicos (65% de las pacientes con policitemia vera y 23% con trombocitemia esencial) eran portadoras de un cambio de una valina por una fenilalanina en la posición 617 (p.Val617Phe), la cual podría contribuir con la etiología de AER [93–95]. En un estudio de *Mercier et al.* (2007), de casos y de controles en mujeres francesas con AER idiopático, se demostró que el cambio JAK2 p.Val617Phe se encontraba más frecuentemente en pacientes con AER (1,06%) que en controles (0,20%). Además, se evidenció que estas mujeres tenían un riesgo relativo cinco veces mayor de presentar el AER en comparación con los controles (OR: 5,33, CI: 2,37 – 11,9, valor *p* <0,001) [61].

Otros genes potencialmente etiológicos han sido estudiados mediante la aproximación gen candidato para determinar si variantes de secuencia podrían estar asociadas a la patogénesis del AER. Por ejemplo, teniendo en cuenta los modelos murinos de letalidad embrionaria temprana, *Kaare et al.* (2007) propusieron la trombomodulina (*THBD*) y el receptor endotelial para la proteína C (*EPCR*) como genes candidato [60]. Estos autores, buscaron variantes por secuenciación directa en toda la región codificante de estos dos genes que pudieran estar asociadas al AER. De manera más precisa, estudiaron 46 pacientes de origen caucásico con historia de AER idiopático.

Tabla 4. Variantes asociadas al AER usando secuenciación de Sanger					
Gen	**Número de pacientes**	**cDNA/alelos**	**Proteína**	**Rs**	**Ensayo funcional**
JAK	3496	c.1849 G>A	p.Val617Phe	rs77375493	No
LIF	95	c.190 G>A	p.Val64Met	----	No
CX3CR1	200	c.745 G>A	p.Val249Ile	rs3732379	No
		c.839 C>T	p.Thr280Met	rs3732378	
IL1B	385	c.-598 T>C	----	rs16944	No
	775				
IL6	775	c.-634 G>C	----	rs1800796	No
TNFA	200	c.-418 G>A	----	rs361525	No
	204	c.-1211 T>C		rs1800629	No
PAI-1	60	c. -816 A>G	----	rs1799889	No
	308	c.*722 T>G		rs7242	No
		c.-989 A>G		rs2227631	No
HLA-G	100	c.-1179 G>A	----	rs1233335	Si
		c.*29-126 G>A		rs915670	
		c.-788 C>A		rs114252012	
	78	*0105N		----	No
	127	*01013		----	Si
C4BPA	962	c.359 G>A	p.Arg120His	rs867500835	Si
		c.671 T>C	p.Ile224Thr	rs116795518	
		c.1268 G>A	p.Gly423G	rs116700161	
C4BPB	962	c.694 A>G	p.Thr232Ala	rs141922788	Si
CD46	962	c.971 C>T	p.Pro324Leu	rs41317833	Si
		c.638 A>T	p.Asn213Ile	----	
CTL4	300	c.49 A>G	p.Thr17Ala	rs231775	Si
		c.*1384 G>A	----	rs3087243	No
FAS	134	c.-1378 G>A	----	rs2234767	No
BAX	67	c.-179 A>G	----	rs751678403	No
		c.26 G>A	p.Arg9lys	rs74422693	
ENOS	200	c.817-26 A>G	----	rs1007311	No
		c.894 T>G	p.Glu298Asp	rs1799983	
	145				
PROKR2	98	c.991 G>A	p.Val331Met	rs117106081	No
THBD	262	c.457 T>G	p.Trp153Gly	----	No
F5	51	c.2627 C>T	p.Ala876Val	rs765295717	Si
		c.2863 T>C	p.Ser955Gly	rs749384397	
FGB	98	c.–455 G>A	----	----	No
ANXA5	243	c.-64 T>G	----	rs1050606	No
TAFI	86	c.505 G>A	p.Ala169Thr	rs3742264	No
		c.*310 T>A	----	rs1087	
ALPP	100	c.265 A>T	p.Ile89Leu	rs13026692	Si
MTHFD	200	c.1958 G>A	p.Arg653Gln	rs2236225	No
MTR	136	c.2756 A>G	p.Asp919Gly	rs1805087	No
MTHFR	200	c.1286 A>C	p.Glu429Ala	rs1801131	No
	136				
	156				
	200	c.665 C>T	p.Ala222Val	rs1801133	No
	136				
	156				

Los autores encontraron dos variantes en el gen *THBD* (c.1418C>T y c.1728+ 23_+40del) y cuatro en *EPCR* (c.323–20T>C, c.323–9_336dup, c.655A>G y c.717+16G>C). Sin embargo, todas las variantes encontradas presentaban frecuencias similares tanto en los casos como en los controles por lo que se concluyó que no participaban en la patogénesis de la enfermedad [60].

Es importante recordar que cientos de genes son candidatos del AER idiopático por la complejidad molecular y la sutil regulación de los procesos de implantación y placentación.

En este marco, la aproximación gen candidato/secuenciación de Sanger ha sido relativamente ineficiente porque permite el análisis de únicamente 700 pares de bases en una sola reacción. La secuenciación de siguiente generación (NGS) ha sido una herramienta útil para sobrepasar estas limitaciones pues permite estudiar amplias regiones del genoma de manera simultánea. Además, el NGS permite efectuar estudios de transcriptómica a nivel global (RNA-seq). Esta técnica ha permitido identificar mutaciones causales de numerosas enfermedades monogénicas y poligénicas [65,96–105].

En este contexto, la NGS podría considerarse una aproximación potencialmente eficiente para el estudio de genes que puedan estar involucrados en el desarrollo del AER idiopático. Nuestro grupo ha abordado con éxito el estudio de varias enfermedades mono y poligénicas por técnicas NGS, y será utilizada en el presente trabajo de tesis [65,97,105–111]. Hasta la fecha se han publicado 5 reportes (incluyendo uno derivado del presente trabajo de tesis) que implementan la técnica de NGS tanto para estudios genómicos como transcriptómicos en pacientes con AER (**Tabla 5**) [112–116]. *Qiao et al.,* realizaron la secuenciación de exoma en las pérdidas de la gestación en 4 familias afectadas por AER. Los autores encontraron variantes de baja frecuencia en las regiones codificantes de los genes *DYNC2H1* (p.Tyr2016Cys y p.Asp2184Val) y *ALOX15* (p.Tyr139Cys y p.Thr560Met). La aproximación de NGS también se ha utilizado en el AER para determinar niveles diferenciales de expresión en los miRNAS y en el transcriptoma global entre las vellosidades coriónicas placentarias y la decidua de pacientes con AER e individuas sanas [113,114]. Algunos

Tabla 5. Genes asociados a AER encontrados por la técnica de NGS					
Estudio	**Técnica NGS**	**Muestra**	**Resultado**		**Proceso biológico**
Qiao et al., 2016	Secuenciación exoma completo	4 familias	*DYNC2H1* *ALOX15*	p.Tyr2016Cys p.Asp2184Val p.Tyr139Cys p.Thr560Met	Mantenimiento y síntesis de cilias Inflamación
Sõber et al., 2016	RNA-seq y miRNA-seq	2 muestras de vellosidades coriónicas	51 sobre-regulados	*ATF4* *C3* *GPX4* *CD74* *PHLDA2* *ICAM1* *LAPTM5* *EGR1* *PDLIM1* *MAPK3* miR-3168 miR-1260b miR-193a-3p miR-494 miR-142-3p	Proliferación celular y apoptosis Inmunomodulación Crecimiento placentario Adhesión celular Hematopoyesis Invasión del trofoblasto y proliferación del sincitiotrofoblasto Moduladores de los genes de proliferación e invasión
			138 sub-regulados	*MT-RNR2* *MTRNR2L9* *MTRNR2L8* *MTRNR2L10* *MTRNR2L3* *MTRNR2L1* *SNORA38B* *HNRPA0* *RN7SK* *SCARNA5* *SSTR5-AS1*	Función mitocondrial Spliceosoma y función ribosomal
Wang et al., 2016	Secuenciación de RNA pequeños	18 tejidos de decidua y vellosidades coriónicas	4 sobre-regulados en vellosidades 5 sub-regulados en vellosidades 32 sobre-regulados en decidua	hsa-miR-191-5p hsa-miR-24-3p hsa-miR-100-5p hsa-miR-146a-5p hsa-miR-1 hsa-miR-372 hsa-miR-371a-5p hsa-miR-376c-3p hsa-miR-486-5p hsa-miR-516-5p hsa-miR-517a-3 hsa-miR-519a-3p hsa-miR-519d	Proceso de apoptosis, ciclo celular y vía de señalización p53
Huang et al., 2017	RNA-seq	9 tejidos de pacientes con FIR y 11 con AER	661 sobre-regulados en FIR y 301 sobre-regulados en AER	*C3* *C4* *C4BP* *DAF* *DF* *SERPING1*	Vías de coagulación y complemento

de los transcritos sobre-expresados en el tejido de las pacientes fueron *ATF4, C3, PHLDA2, GPX4, ICAM1, SLC16A2, miR-3168, miR-142-3p, hsa-miR-191-5p, hsa-miR-24-3p*. Éstos están involucrados en la modulación de la función inmunológica, en el crecimiento placentario, en la invasión del trofoblasto y en la proliferación del sincitiotrofoblasto. Los genes que se encontraron sub-expresados estaban involucrados principalmente en la apoptosis, en el ciclo celular y en las funciones relacionadas con la mitocondria, spliceosoma y ribosoma (*MT-RNR2, MTRNR2L9, MTRNR2L8, SNORA38B, SCARNA5, hsa- miR-1, hsa-miR-372,* entre otros) [113,114]. Un estudio adicional de transcriptómica se realizó usando el tejido endometrial de pacientes con AER y FIR idiopática [115]. Los autores encontraron transcritos diferencialmente expresados en ambos grupos de pacientes, las cuales pertenecían a las vías de coagulación y del complemento, como *C3, C4, C4BP, DAF, DF* y *SERPING1* [115].

En síntesis, la aproximación de NGS es una herramienta eficiente para identificar nuevos genes candidato de relevancia dentro del proceso de implantación en AER. Estos genes después, de su validación funcional, podrían proponerse como marcadores moleculares para el diagnóstico y pronóstico del AER. Los aspectos antes mencionados fueron escritos, a manera de revisión, y sometidos a la revista **Reproductive Science** (IF: 2.5 – 2015) bajo el título ***"Classic and novel sequencing methods to dissect recurrent pregnancy loss aetiology"*** (Ver **Anexo I**).

2.2 Metodología

En este capítulo del trabajo de tesis se realizó un abordaje experimental que permitió responder a la siguiente pregunta de investigación:

1- ¿Qué nuevos genes y mutaciones, identificados por NGS y validados funcionalmente, podrían estar relacionados con la etiopatología del AER?

La secuenciación de exoma por NGS (WES) es una técnica que permite estudiar todas las regiones codificantes del genoma (exones). En este estudio, usamos la aproximación de WES con el objetivo de identificar las variantes candidatas en genes que posiblemente estuvieran asociadas a la fisiopatología del AER usando la siguiente metodología.

2.2.1 Inclusión de casos no-familiares de AER idiopático y de controles

Se incluyeron en el estudio 49 pacientes afectadas por AER no familiar. Las pacientes colombianas fueron remitidas de centros de reproducción asistida asociados a la Universidad del Rosario, de las consultas privadas de gineco-obstetras de Cecolfes y de los genetistas de Méderi y de la Universidad del Rosario (n=6). Las muestras de las pacientes francesas, que hacen parte de un banco de ADN de AER del Hospital de Nîmes (Francia), fueron también incluidas (n=43) [117]. El criterio de inclusión de las pacientes se determinó de acuerdo con la guía americana para AER que definió la enfermedad como 2 o más abortos espontáneos y consecutivos antes de la semana 20 de gestación. Se incluyeron las pacientes que cumplían con este criterio y que tenían AER idiopático. Los criterios de exclusión fueron: pacientes con síndrome antifosfolípido, alteraciones en el cariotipo materno, coagulopatías y anormalidades anatómicas en el útero. El grupo control incluyó 100 mujeres de los mismos orígenes étnicos de las pacientes, que tuvieran al menos 2 hijos vivos y sin antecedentes de AER, enfermedad autoinmune, coagulopatías o enfermedades ginecológicas. A las pacientes y a los controles se les explicó los objetivos del estudio y previa firma del consentimiento informado, se les tomó una muestra de sangre periférica para realizar la extracción del ADN por medio del método de salting-out. El diagrama de la metodología usada para la identificación de nuevas variantes por WES se encuentra resumido en la **Figura 2.**

Figura 2. Metodología para la identificación de nuevas variantes potencialmente asociadas al AER y validación funcional.

2.2.2 Preparación de la librería y secuenciación por NGS del exoma

Las muestras de las pacientes fueron enviadas a la plataforma internacional Genotypic Technology Pvt. Ltd. en Bangalore, India. Para la preparación de la librería, el ADN se fragmentó por medio enzimático y la librería se construyó usando el protocolo para SureSelectXT Reagent Kit, HSQ (Catalog # G9611A). 200 nanogramos de ADN genómico de cada muestra fue llevado a un volumen final de 50 µL con TE y se sonicó con el sistema Covaris S220 para fragmentar el ADN a una longitud de 150 pb a 400 pb. El tamaño de los fragmentos fue verificado con el equipo Agilent HS Bioanalyzer Chip. La librería de ADN genómico fue construida siguiendo el protocolo para el sistema “Target Enrichment for Illumina Paired-End Sequencing Library manual”.

A los fragmentos se les repararon los extremos, se fosforilaron, se agregaron adeninas en los extremos y se ligaron los adaptadores. Después de la ligación, se realizó una PCR de 10 ciclos para enriquecer los fragmentos ligados a los adaptadores. La librería preparada se hibridó siguiendo las instrucciones del sistema Target Enrichment for Illumina Paired-End Sequencing Library manual. Los fragmentos de la librería hibridados fueron aislados por medio de captura magnética usando perlas de estreptavidina. Las perlas con los fragmentos capturados fueron resuspendidas en 15 µL de agua libre de nucleasas y se realizaron máximo 12 ciclos de amplificación por PCR de la librería. La purificación de los fragmentos se realizó usando el sistema HighPrep PCR clean up. La librería capturada fue cuantificada usando el fluorometro Qubit y en el sistema Agilent High Sensitivity Bioanalyzer Chip.

La librería obtenida fue denaturada y secuenciada usando el equipo Illumina HiSeq 4000, que utiliza la tecnología de secuenciación por síntesis para fragmentos de ~150 pb. La librería fue diluida, denaturada e hibridada con oligonucleótidos inmovilizados sobre una superficie que fue usada como primers. Cada fragmento hibridado sirvió como molde en el proceso de amplificación en puente resultando en grupos de 1000 copias clonales. La secuenciación fue realizada por síntesis usando 4 nucleótidos fluorescentes. Durante cada ciclo de secuenciación, un dNTP con el fluorocromo fue agregado, luego junto con el bloqueador fueron clivados, y la siguiente base complementaria fue agregada a la cadena de nucleótidos. Se realizaron 150 ciclos que corresponden a 150 pb secuenciadas por cada fragmento.

2.2.3 Análisis bioinformático y depuración de los datos

Una vez completada la secuenciación, los datos fueron extraídos de un servidor usando el software Illumina pipeline para obtener los archivos FASTQ. Se realizó el control de calidad de los datos usando los scripts de la empresa Genotypic. El objetivo de este análisis fue remover los llamados de base indeseables del extremo 3´, los adaptadores de secuenciación y los extremos 3´de baja calidad. Los datos con alta calidad fueron alineados con las secuencias de referencia del genoma

humano GRCh37/hg19 con el programa BWA mem y las variantes fueron identificadas solo para las regiones que estaban presentes en la región blanco de todo el exoma Sureselect-Exome-50Mb-hg19_revA con las herramientas SAMtool y Bcftool. Las variantes que pasaron el umbral mínimo de cobertura de 30X y tuvieron un puntaje de calidad mínimo de Q30 (probabilidad de un llamado incorrecto de base 1 en 1000) fueron reportadas. Éstas se anotaron usando el programa Variant Studio para obtener la localización (intrónica/exónica/UTR), el nombre del gen, la variante proteica, la función, el identificador dbSNP y la predicción de patogenicidad (SIFT y Polyphen) de cada variante.

2.2.4 Creación del subset de genes candidato

La selección del subgrupo de genes candidato se realizó mediante una revisión sistemática de la literatura de genes que tuvieran una función importante dentro del proceso de implantación. Se usó la base de datos de PubMed y las palabras MeSH *Recurrent spontaneous abortion* OR *RSA* OR *implantation* AND *genetics*. Se incluyeron los artículos que mostraban asociación estadísticamente significativa entre variantes de genes candidato y el desarrollo de AER. También se incluyeron los artículos con ensayos funcionales que demostraran que las variantes en los genes candidatos tienen una implicación funcional. Adicionalmente, se buscaron modelos murinos KO en la base de datos del MGI del Jackson Laboratory (http://www.informatics.jax.org/) que tuvieran fenotipos asociados a *Abnormal embryo attachment*, *Failure of embryo implantation*, *Impaired embryo implantation*, *Abnormal spontaneous abortion rate*, *Abnormal decidualization*, *Abnormal postimplantation uterine environment*, *Endometrium inflammation* y *Abnormal uterine receptivity*. Con estos resultados se realizó una lista en Excel de los genes candidato y se determinó la función de cada uno usando las bases de datos de GeneCards (http://www.genecards.org/), UniProt (https://www.uniprot.org/), NCBI (https://www.ncbi.nlm.nih.gov/gene/) y Ensembl (http://www.ensembl.org/index.html). También se determinó la expresión tisular (endometrio) de cada uno de los genes usando la herramienta de Illumina NextBio (http://www.nextbio.com/b/nextbioCorp.nb) y la base de datos “The human protein

atlas” (https://www.proteinatlas.org/) . Se buscó la expresión y la función de cada gen durante la gestación usando la base de datos PubMed y las palabras MeSH *Pregnancy* OR *Gestation* OR *Embryo Implantation* OR *Decidualization* AND “nombre del gen o sus siglas”. Con esta metodología se obtuvo una lista de 234 genes candidato de AER.

2.2.5 Análisis bioinformático del subset de genes

El análisis bioinformático del subset de genes candidato se realizó a partir de los archivos con las variantes identificadas en la secuenciación de exoma. Los archivos del exoma de cada paciente en formato .xlsx se transformaron a archivos en formato .csv para poder ser analizados en el programa *R.* La programación en R para el filtrado de las variantes en los genes candidatos se realizó usando los siguientes scripts:

- ConstructTable(folderName): con esta función se creó una tabla única que contenía toda la información de cada uno de los archivos que se encontraban en la carpeta con formato .csv. De esta forma, el programa se optimizó para funcionar con una tabla unificada que contenía la información de todas las pacientes.
- FilterByConsequence(table): con esta función se filtraron las variantes missense, nonsense, frameshift o que estaban en los sitios de splicing.
- FilterByGene(table, folderName, geneSubset = "genes.txt"): buscó las variantes que se encontraban en la lista del subset de genes. De esta forma, se obtuvieron las variantes en los genes candidato.
- SaveTable(table, fileName): esta función generó una tabla en formato Excel con las variantes encontradas en el subset de genes candidato para su posterior análisis.

Utilizando la tabla de las variantes encontradas en el subset de genes candidato se realizó un filtro manual de todas las variantes nuevas (Novel SNPs y Novel INDELs). A las variantes missense encontradas se les realizó un alineamiento múltiple (ClustalW) con secuencias proteicas de distintas especies para determinar si se

encontraban dentro de regiones conservadas evolutivamente. Las variantes missense que estaban en regiones conservadas entre especies y las variantes nonsense, frameshift o que estaban en los sitios de splicing fueron escogidas para su confirmación, usando PCR y secuenciación directa a partir del ADN de las pacientes (ver más adelante 2.2.6.1). Las variantes confirmadas se tamizaron en el grupo control por PCR/secuenciación directa.

Se realizó un análisis estadístico comparativo usando el test χ^2 para determinar si había enriquecimiento de las variantes con cambios en la proteína en el grupo de genes candidato vs. todo el exoma [118,119]. La significancia estadística se definió como un valor de $p < 0,05$.

2.2.6 Análisis de las variantes candidatas

2.2.6.1 Secuenciación directa de las variantes candidatas

Para la validación de las variantes identificadas por NGS, se efectúo PCR/secuenciación directa. Se diseñaron primers flanqueantes de cada una de las variantes de interés. Se realizó una PCR para amplificar cada uno de los fragmentos que contenían las variantes candidatas. Los amplicones fueron purificados mediante exonucleasa I y fosfatasa alcalina. Se realizó secuenciación de Sanger con los primers diseñados en el interior de cada amplicón usando el programa Primer3 (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). Se efectuó un tamizaje en el grupo control (n=100) con las variantes que fueron confirmadas en cada paciente. Las variantes que no se encontraron dentro del grupo control fueron seleccionadas como potenciales candidatas para la validación mediante estudios funcionales (ver más adelante).

2.2.6.2 Análisis de fragmentación molecular orbital (FMO) de la estructura de proteínas con variantes candidatas

Se realizó un análisis de fragmentación molecular orbital (FMO) de algunas proteínas, mutantes, que tuvieran una estructura previamente cristalizada. Esto se realizó para determinar si algunas de las variantes candidatas, encontradas por

medio de NGS exoma en pacientes con AER, alteraban potencialmente las estructuras proteicas [120–123].

Se usó la función de swapaa del software UCSF Chimera para realizar las sustituciones de los aminoácidos por las mutaciones en la estructura de la proteína usando una librería de rotámeros. Se calcularon los puentes de hidrógeno con el método de Poisson-Boltzmann usando el servidor H++ web server y el mismo pH descrito en la cristalización. Se optimizaron los átomos de hidrógeno usando el modelo "conductor-like screening model (COSMO)". Para determinar el efecto de las sustituciones de aminoácidos en la estructura de las proteínas, se realizaron los cálculos de FMO. Éstos permitieron descomponer las interacciones energéticas en términos de electrostática, repulsión, transferencias de carga y dispersión. Se escogieron las regiones cerca del sitio catalítico de cada proteína para ser fragmentadas. Los residuos con un radio ≤ 10Å alrededor del residuo mutante fueron incluidos en el análisis para disminuir los costos computacionales. Las interacciones entre fragmentos ≥3kcal/mol fueron considerados como significativos y se tuvieron en cuenta para la creación del modelo de fragmentación 3D. Los resultados de este estudio se incluyeron en la publicación ***"Novel genes and mutations in patients affected by recurrent pregnancy loss"*** (Ver **Anexo II**).

2.2.6.3 Secuenciación de la región codificante de *THBD* en pacientes colombianas con AER

Se realizó un tamizaje para identificar si la variante encontrada por WES se encontraba de manera frecuente en pacientes colombianas con AER. Además, se realizó el tamizaje para identificar otras variantes en la región codificante de *THBD* que pudieran tener un efecto deletéreo. Para el tamizaje de las variantes, se incluyeron 233 y 29 pacientes francesas y colombianas, respectivamente, afectadas por AER idiopático no familiar. Las pacientes colombianas fueron remitidas de los centros de reproducción asociados a la Universidad del Rosario, de las consultas privadas de gineco-obstetras de Cecolfes y de los genetistas de Mederi y de la Universidad del Rosario. Las pacientes francesas hacen parte del banco de ADN de pacientes con AER del Hospital de Nîmes, Francia [117]. El criterio de inclusión de las

pacientes se efectuó de acuerdo con la guía americana para AER. Se incluyeron las pacientes que cumplían con este criterio y que tenían AER idiopático. El grupo control incluía 165 mujeres colombianas que tuvieran al menos 2 hijos vivos y sin antecedentes de AER, enfermedad autoinmune, coagulopatías o enfermedades ginecológicas. Para el grupo control de origen caucásico se utilizaron las bases de datos públicas de SNPs. A las pacientes y a los controles se les explicó los objetivos del estudio y previa firma del consentimiento informado, se les tomó una muestra de sangre periférica para realizar la extracción del ADN por medio del método de salting-out. Para la PCR/secuenciación directa se diseñaron primers flanqueantes de la región codificante del gen *THBD* (monoexónico). Se realizó una PCR para amplificar el ORF completo del este gen. Los amplicones fueron purificados mediante exonucleasa I y fosfatasa alcalina. Se realizó secuenciación de Sanger usando primers diseñados en el interior del amplicón. Las secuencias se compararon con la versión wild type (WT) de *THBD* (ENSG00000178726). Se realizó un tamizaje en el grupo control con las variantes raras (frecuencia alélica mínima - MAF <1%) que fueron confirmadas en cada paciente. Las variantes que no se encontraron dentro del grupo control fueron seleccionadas como potenciales candidatas para la validación mediante estudios funcionales (ver más adelante).

2.2.6.4 Análisis bioinformático de las variantes de *THBD*

Las frecuencias alélicas de las variantes candidatas obtenidas en pacientes francesas con AER fueron comparadas con aquellas reportadas en la base de datos de Ensembl. Las frecuencias alélicas de las variantes encontradas en pacientes colombianas con AER fueron comparadas con el grupo control del mismo origen étnico. Se realizó la predicción de la potencial patogenicidad de las variantes missense candidatas usando SIFT y PolyPhen. Adicionalmente, se realizó un alineamiento múltiple (ClustalW) con las secuencias proteicas de distintas especies para determinar si se encontraban dentro de regiones conservadas evolutivamente (*Pan troglodytes, Sus scrofa, Canis lupus, Mus musculus, Gallus gallus y Xenopus laevis*). Se realizaron pruebas estadísticas de χ^2 para calcular potenciales diferencias entre las frecuencias alélicas de casos y controles.

2.2.6.5 Análisis funcional *in vitro* de las variantes candidatas

Las variantes de THBD (p.Trp153Gly) y de EPAS1 (p.Leu504Val y p.Tyr488Cys) que se encontraron en regiones evolutivamente conservadas y que no se encontraron en la población control fueron escogidas para estudiarlas mediante ensayos *in vitro*. Se escogieron principalmente estas variantes teniendo en cuenta los dominios funcionales en los que se encontraban y la función de la proteína en el contexto de la enfermedad. La metodología para estas variantes se presentará a continuación.

2.2.6.5.1 Estudio de la expresión de THBD WT y mutante en células THP-1

Células THP-1 no transfectadas fueron lisadas usando el buffer RIPA y un coctel inhibidor de proteasas para determinar la expresión basal de THBD en esta línea celular. El extracto de proteínas totales fue cuantificado utilizando el kit Pierce BCA protein assay (Thermo Scientific). La expresión de la proteína se analizó por electroforesis en un gel al 12% de poliacrilamida y se transfirió a una membrana de nitrocelulosa. Ésta se bloqueó con leche al 5% y se incubó con una dilución 1:500 del anticuerpo primario anti-THBD (Santa Cruz Biotechnology Inc., sc-13164). Después de 5 lavados con PBS-Tween 0,1%, se incubó con una dilución 1:5000 de fosfatasa alcalina conjugada con el anticuerpo secundario anti-IgG de ratón (Vector Burlingame, CA). La expresión de THBD se reveló con el sistema de sustrato líquido BCIP/NBT (Promega, Madison, USA). Se realizaron 3 réplicas biológicas.

Para determinar la sobre-expresión de THBD en la membrana celular de la línea THP-1, se realizó la transfección de estas células con 500 ng de la versión WT y mutante de *THBD* en pcDNA 3.1/Zeo (+) y con el reactivo Fugene (Promega, Madison, WI, USA) en un medio libre de suero por 24 horas, en una caja de 24 pozos. Las células transfectadas fueron lavadas con PBS 1X. Luego de incubar con el anticuerpo primario anti-THBD (Santa Cruz Biotechnology, Inc, sc-13164) por 1 hora a 4°C, se realizaron 2 lavados con PBS 1X y se incubó con el anticuerpo secundario anti-IgG de ratón conjugado con FITC. Para permeabilizar las células,

se fijaron con paraformaldehído 1% por 30 minutos y se les adicionó saponina por 30 minutos. Las células se lavaron con PBS 1X y se analizaron por citometría de flujo utilizando el equipo BD FACSCanto™ II. La prueba de Fischer se utilizó para calcular las diferencias en las intensidades de fluorescencia entre las células transfectadas con el ADN plasmídico *THBD* WT y mutante. Se realizaron 4 réplicas biológicas por duplicado cada una.

2.2.6.5.2 Estudio de la localización subcelular de THBD WT y mutante

Se realizaron ensayos de localización subcelular por medio de proteínas marcadas (GFP e inmunofluorescencia-IFI) de las versiones WT y mutante de THBD (p.Trp153Gly). El ORF completo de *THBD* WT y mutante fueron clonados en el plásmido pcDNA3.1-CT/GFP-TA para su localización subcelular mediante GFP. Adicionalmente, se utilizaron los clonajes del ORF completo de *THBD* WT y mutante en el plásmido pcDNA 3.1/Zeo (+) para efectuar experimentos de IFI. Para determinar la localización subcelular por medio de GFP, se transfectaron células CHO y COS-7 con el ADN plasmídico GFP de *THBD* WT o mutante por 24 horas. Las células postransfección fueron fijadas con paraformaldehído (4%) y lavadas con PBS 1X. Se utilizó el líquido de montaje con DAPI UltraCruz® Mounting Medium (sc-24941). Para determinar la localización subcelular por medio de IFI, se usaron los clonajes efectuados en pcDNA 3.1/Zeo (+) (WT o mutante). Las células CHO, COS-7 y THP-1 postransfección fueron fijadas con paraformaldehído al 4%, permeabilizadas con Tritón 1X, bloqueadas con BSA 0,1% y lavadas con PBS 1X. Se incubaron las células toda la noche con el anticuerpo primario anti-THBD (Santa Cruz Biotechnology, Inc, sc-13164). Después de 5 lavados, las células se incubaron con el anticuerpo secundario anti- IgG de ratón CFL594 generado en cabra (Santa Cruz Biotechnology Inc., sc 362277) por 2 horas. Se utilizó el líquido de montaje con DAPI UltraCruz® Mounting Medium (Santa Cruz Biotechnology Inc., sc-24941). Se identificaron los diferentes perfiles de localización subcelular y se realizó un conteo celular para determinar potenciales diferencias estadísticamente significativas entre los perfiles encontrados en las células transfectadas con el ADN plasmídico *THBD* WT y mutante. La prueba de Fischer se utilizó para calcular las diferencias

estadísticas entre los perfiles encontrados. Se realizaron 3 réplicas biológicas por triplicado cada una.

2.2.6.5.3 Estudio del efecto de la mutación THBD-p.Trp153Gly en la expresión de genes de la cascada corriente abajo de la citoquina HMGB1

Se realizó qPCR de las células transfectadas con la versión THBD WT y mutante, y estimuladas con la citoquina HMGB1 para determinar si la mutación THBD-p.Trp153Gly, encontrada en pacientes colombianas con AER, generaba un ambiente pro-inflamatorio por potenciales alteraciones en la expresión de genes corriente abajo de HMGB1. La transfección se realizó en la línea celular THP-1, utilizando 500 ng de las construcciones efectuadas en pcDNA 3.1/Zeo (+) (WT o mutante) y el reactivo Fugene (Promega, Madison, WI, USA) en un medio libre de suero por 24 horas. Posteriormente, se realizó la estimulación de las células con la citoquina HMGB1 (Origene, cat#: TP720309) por 24 horas. Después de la incubación, se realizaron lavados con PBS 1X y se efectuó la extracción de ARN total utilizando Trizol y el kit PureLink® RNA (Life technologies). La síntesis de ADNc se realizó usando el kit SuperScript III First-Strand (Life Technologies) siguiendo las recomendaciones del fabricante. Con el ADNc obtenido, se efectuó una RT-PCR usando primers específicos para confirmar la expresión endógena de los transcritos de *RAGE* y *GAPDH* en esta línea celular. Además, se realizó la qPCR para determinar cambios en la expresión de los siguientes genes que se encuentran dentro de la cascada pro-inflamatoria corriente abajo de HMGB1: *IL-8, CCL2, NFKB1, TNF-α, IL-10, IL-1β, GM-CSF, NOS2, VEGFA, PGF, FLT-1, JAK2, IKBKB, MAPK8, V-CAM1, MAP2K1* y *STAT3.* Se utilizó el gen *GAPDH* para la normalización. Los primers fueron diseñados usando el software Primer3Plus [124] y se usó 0,4 µM de cada primer para la reacción de qPCR. El 1X SYBER Green PCR Master Mix (Thermo Fisher Scientific) y el QuantStudio 3 Real-Time PCR system (Thermo Fisher Scientific) fueron utilizados para el ensayo de qPCR. Se realizó una curva de melting para verificar la especificidad y la formación de los dímeros de primers. El análisis de expresión se realizó usando el método ΔΔCT. Se efectuó una

prueba de t-student para determinar diferencias estadísticas potencialmente significativas ($p < 0,05$). Se realizaron 5 réplicas biológicas por duplicado cada una.

2.2.6.5.4 Modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD

Se realizó un BLASTp utilizando la secuencia de aminoácidos del dominio lectina-like de THBD (aminoácidos 19-170) y usando como referencia la base de datos de estructuras del *protein data bank* (PDB). La estructura cristalográfica con mayor similitud fue el receptor de manosa CTLD3 humano - PDB ID: 5XTS [125]. Con esta estructura como plantilla, se modeló el dominio lectina de THBD1 utilizando el servidor I-TASSER [126–128]. El mejor modelo basado en homología obtenido (WT) fue usado como base para generar la estructura del domino mutante (MT, p.Trp153Gly) con el algoritmo swapaa del programa UCSF Chimera [129], utilizando la librería de rotámeros de Dunbrack [130]. Las estructuras WT y MT del dominio lectina fueron refinadas con dinámica molecular en el programa NAMD 2.12 con el campo de fuerza CHARMM36, usando frontera periódica, un ensamble isotérmico e isobárico (P=1 atm, T=300 K) y un modelo TIP3 de solvatación [131–133]. Se procedió con una minimización de 0,25 nanosegundos (ns), seguido de una dinámica de 1,0 ns y finalizando con una minimización de 0,25 ns. La calidad de las estructuras fue estimada usando los programas MolProbity y Qmean [134,135].

Las estructuras refinadas para WT y MT fueron alineadas estructuralmente y comparadas con el RMSD (root-mean-square deviation) usando el programa VMD 1.9.3 [136]. El volumen y área superficial de los modelos refinados fueron estimados con el programa 3V, usando un radio de prueba de 1 Å [137]. Adicionalmente, se estimó el pKa por residuo usando el programa PROPKA [138]. A partir de estos resultados, se utilizó el algoritmo APBS (Adaptive Poisson-Boltzmann Solver) para generar el potencial electrostático para las estructuras analizadas[139]. Se calculó para cada uno de los residuos la carga y el radio por átomo usando el programa PDB2PQR [140].

Se calculó el cambio de energía (kcal/mol) que causa la sustitución p.Trp153Gly usando los últimos 7 ns (7000 frames, a partir de la estabilización de la trayectoria) de una dinámica de 10 ns, a partir de las estructuras WT y MT refinadas. La dinámica fue calculada usando el programa NAMD 2.12 con los mismos parámetros previamente descritos en el proceso de mejoramiento estructural. Se evaluó la energía para los modelos WT y MT con el plugin NAMD Energy 1.4 implementado en el programa VMD 1.9.3. La diferencia en los valores de la moda y promedio de la energía libre de las dos trayectorias fue utilizada como criterio para el cambio de energía implicado en la mutación p.Trp153Gly. Las distribuciones de energía obtenidas fueron evaluadas con una prueba de normalidad de Kolmogorov-Smirnov, y se realizó una prueba *t* -student para comparar las distribuciones de energía de WT vs. MT usando el paquete *stats* del programa R 3.3.2.

2.2.6.5.5 *Estudio genético y funcional de las variantes de EPAS1*

2.2.6.5.5.1 Inclusión de pacientes no relacionados con AER idiopático y de controles

Se escogió el gen *EPAS1* para el tamizaje de las variantes en los exones 10, 11 y 12, ya que se encontraron en los experimentos de NGS mutaciones en 3 pacientes francesas con AER. Además, EPAS1 es un factor de transcripción fundamental durante el proceso de decidualización. Para este tamizaje se incluyeron 224 pacientes francesas afectadas por AER idiopático no familiar. Las pacientes francesas hacen parte del banco de pacientes de AER del Hospital de Nîmes, Francia [117]. El criterio de inclusión de las pacientes se efectuó de acuerdo con la guía americana para AER. Se incluyeron las pacientes que cumplían con este criterio y que tenían AER idiopático. Para el grupo control se utilizó la base de datos pública de SNPs Ensembl (https://www.ensembl.org/index.html). A las pacientes se les explicó los objetivos del estudio y previa firma del consentimiento informado, se les tomó una muestra de sangre periférica para realizar la extracción del ADN por medio del método de salting-out.

2.2.6.5.5.2 Secuenciación directa de los exones 10, 11 y 12 del gen *EPAS1*

Se diseñaron primers flanqueantes de la región codificante de los exones 10, 11 y 12 de *EPAS1*. Estos exones codifican para el dominio de degradación dependiente de oxígeno (ODDD) y para el dominio de N-transactivación de la proteína. Se realizó una PCR para amplificar cada exón. Los amplicones fueron purificados mediante exonucleasa I y fosfatasa alcalina. Se realizó secuenciación de Sanger usando primers diseñados en el interior del amplicón. Las secuencias se compararon con la versión WT de *EPAS1* (ENSG00000116016).

2.2.6.5.5.3 Análisis bioinformático de las variantes de *EPAS1*

Se realizó la predicción de la potencial patogenicidad de las variantes missense candidatas usando SIFT y PolyPhen-2. Adicionalmente, se les realizó un alineamiento múltiple (ClustalW) con secuencias proteicas de distintas especies para determinar si se encontraban dentro de regiones conservadas evolutivamente (*Mus musculus, Rattus norvegicus, Pan troglodytes, Equus caballus* y *Bos Taurus)*.

2.2.6.5.5.4 Estudio gen reportero (bajo condiciones de hipoxia) de las mutaciones p.Leu504Val y p.Tyr488Cys de EPAS1

Para el estudio gen reportero bajo condiciones de hipoxia de las mutaciones EPAS1 se realizaron las construcciones del ORF completo de *EPAS1* WT en pcDNA 3.1/Zeo (+). Esta construcción se usó de molde para realizar las mutagénesis de las posiciones c.1463A>G y c.1510C>G usando el Kit GeneArt® Site-Directed Mutagenesis System (Thermo Fischer). Los promotores de los genes blanco directos conocidos de EPAS1 (*VEGF, INHA, EPO* y *PAI-1*) fueron clonados en el plásmido pGL4.22[luc2CP/Puro], el cual incluye la secuencia del gen reportero de luciferasa (*luc2CP - Photinus pyralis*). Se usó también un plásmido que contenía la secuencia codificante de *Renilla* como normalizador de la transfección. Se realizó una co-transfección en células COS-7 bajo condiciones de hipoxia química con 100 μM deferroxamina (DFO) con las 3 construcciones: la que contiene el ORF WT o

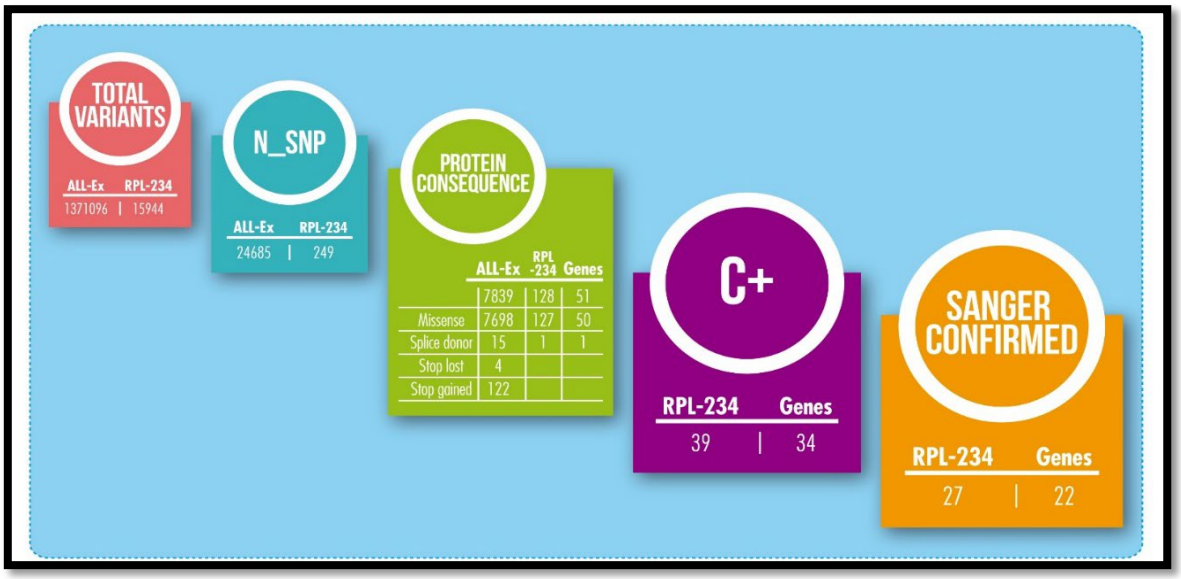
mutante de *EPAS1* (500 ng), la que incluye la secuencia de los promotores (700 ng) y la que codifica para la *Renilla* (30 ng). La magnitud de la transactivación/transrepresión se analizó usando el sistema Dual-Luciferase Reporter Assay (Promega, Madison, WI, USA). Previo al ensayo gen reportero, se efectuaron varios ensayos de puesta a punto de la hipoxia.

2.3 Resultados

2.3.1 Secuenciación NGS y de Sanger para la búsqueda de nuevas variantes de AER idiopático y análisis bioinformático

Los resultados de la secuenciación del exoma en las pacientes colombianas y francesas con AER idiopático mostraron una profundidad de exploración, para más del 85% de las variantes, de 30X en promedio. El total de las variantes encontradas en todos los exomas fue 1,371,096 y de éstas, 15,944 estaban en el subset de genes candidato (RPL-234) (**Figura 3**).

Figura 3. Filtrado bioinformático de los datos de la secuenciación del exoma. All-ex: datos del exoma completo excluyendo los resultados RPL-234. RPL-234: subset de genes candidato para AER. C+: variantes que involucran residuos estrictamente conservados durante la evolución.



Usando el filtro de variantes nuevas (N_SNP) se encontraron 24,685 en todos los exomas y 249 en el subset RPL-234. En el subset se identificaron 128 variantes en 51 genes candidato que presentaban cambios a nivel de la proteína: 127 missense y 1 splice donor. El análisis estadístico comparativo (test χ^2) entre las variantes con cambios en la proteína entre el grupo RPL-234 (51%) versus las de todo el exoma (31,7%) mostró que el subset de genes RPL-234 estaba enriquecido por variantes con cambios a nivel proteico (p=0,0005). Treinta y nueve variantes se encontraban en regiones conservadas evolutivamente (C+) y de éstas, 27 fueron confirmadas por medio de secuenciación directa y consideradas candidatas en la etiología del AER (**Figura 3, Tabla 6**). Se usaron 4 programas bioinformáticos (SIFT, PolyPhen-2, Mutation taster y MutPred) para predecir la potencial patogenicidad de las variantes C+. El 27% de las variantes candidatas presentaban un puntaje compatible con efectos potencialmente deletéreos en todos los programas: CDH1-p.Val55Gly, THBD-p.Trp153Gly, FGA-p.Phe685Cys, BMP7-p.Arg150Cys, IDO2-p.Phe180Cys, IDO2-p.Phe250Ser y MMP-10-p.Asp199Asn (**Tabla 7**).

Luego de la selección de las variantes y su verificación por secuenciación de Sanger identificamos que 20 de las 49 pacientes presentaron mutaciones en genes candidato del subset 234-RPL. El 80% de estas pacientes presentaron pérdidas tempranas de la gestación, 10% pérdidas tardías y el 10% restante ambos tipos de abortos (**Tabla 8**). En las pacientes colombianas se identificaron 4 variantes en genes candidato relevantes para la coagulación, la remodelación de la matriz extracelular y la modulación de la función inmunológica. En las pacientes francesas se identificaron 23 variantes en genes candidato relevantes en procesos biológicos como la adhesión celular trofoblasto-endometrio, la coagulación, la angiogénesis, la proliferación celular, la diferenciación, la migración, la apoptosis y la activación de los receptores nucleares esteroideos (**Tabla 6**). Ninguna de las 27 variantes fue encontrada en los grupos control respectivos (francesas y colombianas). Se escogieron algunas de las variantes, como THBD p.Trp153Gly, EPAS1 p.Tyr488Cys y EPAS1 p.Leu504Val para realizar estudios funcionales *in vitro* y determinar su posible causalidad en la fisiopatología del AER. Estas variantes

fueron aquellas que se encontraban en regiones conservadas evolutivamente de la proteína (ver más adelante).

Tabla 6. Variantes nuevas identificadas en genes candidato mediante NGS y confirmadas por secuenciación directa en pacientes con AER.

Paciente	Gen	Variante		Proceso biológico
		Posición nucleótido	Posición proteína	
Pt-37	***TRO***	c.2909G>A	p.Gly970Asp	Adhesión celular trofoblasto-endometrio
Pt-39	***CDH11***	c.1355C>G	p.Ala452Gly	
Pt-20	***CDH1***	c.164T>G	p.Val55Gly	
Pt-46	***THBD***	c.457T>G	p.Trp153Gly	Coagulación
Pt-22	***F5***	c.5932A>C	p.Thr1978Pro	
Pt-22	***F5***	c.4619A>C	p.Glu1540Ala	
Pt-16	***FGA***	c.2054T>G	p.Phe685Cys	
Pt-45	***MMP10***	c.595G>A	p. Asp199Asn	Remodelación de la matriz extracelular
Pt-26	***MMP9***	c.17C>T	p.Pro6Leu	
Pt-28	***COL6A3***	c.6859C>T	p.Arg2287Trp	
Pt-29	***ADAMTS1***	c.2068G>C	p.Val690Leu	
Pt-45	***TNC***	c.4063G>A	p.Val1355Met	
Pt-35	***FLT1***	c.953C>T	p.Ser318Leu	Angiogenesis
Pt-31	***FLT1***	c.2435G>A	p. Arg812Gln	
Pt-5	***EPAS1***	c.1510C>G	p.Leu504Val	
Pt-11 Pt-17	***EPAS1***	c.1463A>G	p.Tyr488Cys	
Pt-4	***LIFR***	c.2338C>T	p.Arg780Cys	Proliferación celular, diferenciación, migración y apoptosis
Pt-36	***FGFR2***	c.1088C>T	p.Ala363Val	
Pt-19	***BMP7***	c.448C>T	p.Arg150Cys	
Pt-19	***AMN***	c.207G>A	p.Met69Ile	Metabolismo
Pt-48	***IDO2***	c.749T>C	p.Phe250Ser	Modulación de la función inmunológica
Pt-5	***IDO2***	c.539T>G	p.Phe180Cys	
Pt-26	***CR1***	c.4501A>G	p.Thr1501Ala	
Pt-28	***TLR3***	c.2384C>T	p.Ala795Val	
Pt-20	***TRAF3IP1***	c.415C>T	p.Arg139Trp	
Pt-14	***NCOA1***	c.2011T>G	p.Ser671Ala	Activación de los receptores nucleares esteroideos

Tabla 7. Predictores de patogenicidad para las variantes candidatas confirmadas por secuenciación de Sanger

Proteína	Variante	SIFT	Polyphen	Mutation taster			Mutpred	Total
				PhyloP	PhastCons	Predicción		
TRO	p.Gly970Asp	Tolerada (0.07)	Probablemente dañina (0.999)	1.224	0.019	Polimorfismo	0.379	1 de 4
CDH11	p.Ala452Gly	Tolerada (0.38)	Benigna (0.197)	3.39	0.999	Causante de enfermedad	0.322	1 de 4
CDH1	p.Val55Gly	Deletérea (0)	Probablemente dañina (1)	3.938	1	Causante de enfermedad	0.821	4 de 4
THBD	p.Trp153Gly	Deletérea (0)	Probablemente dañina (0.976)	3.847	0.987	Causante de enfermedad	0.716	4 de 4
F5	p.Thr1978Pro	Deletérea (0)	Benigna (0.388)	3.052	1	Causante de enfermedad	0.889	3 de 4
	p.Glu1540Ala	Deletérea (0.02)	Benigna (0.185)	2.084	0.221	Polimorfismo	0.455	1 de 4
FGA	p.Phe685Cys	Deletérea (0)	Probablemente dañina (1)	5.158	1	Causante de enfermedad	0.960	4 de 4
MMP10	p. Asp199Asn	Deletérea (0)	Probablemente dañina (0.992)	5.238	1	Causante de enfermedad	0.711	4 de 4
MMP9	p.Pro6Leu	Tolerada (0.4)	Posiblemente dañina (0.824)	0.878	0.366	Polimorfismo	0.461	1 de 4
COL6A3	p.Arg2287Trp	Deletérea (0.02)	Probablemente dañina (0.985)	1.479	0.003	Polimorfismo	0.517	3 de 4
ADAMTS1	p.Val690Leu	Tolerada (0.1)	Benigna (0.113)	4.077	1	Causante de enfermedad	0.547	2 de 4
TNC	p.Val1355Met	Tolerada (0.06)	Benigna (0.316)	1.239	0.843	Causante de enfermedad	0.674	2 de 4
FLT1	p.Ser318Leu	Tolerada (0.16)	Posiblemente dañina (0.794)	2.463	0.019	Polimorfismo	0.472	1 de 4
	p.Arg812Gln	Tolerada (0.16)	Benigna (0.407)	6.18	1	Causante de enfermedad	0.334	1 de 4
EPAS1	p.Tyr488Cys	Tolerada (0.08)	Posiblemente dañina (0.687)	3.174	1	Causante de enfermedad	0.254	2 de 4
	p.Leu504Val	Deletérea (0)	Probablemente dañina (0.979)	5.684	1	Causante de enfermedad	0.456	3 de 4
LIFR	p.Arg780Cys	Tolerada (0.07)	Posiblemente dañina (0.542)	1.265	0.899	Polimorfismo	0.577	2 de 4
FGFR2	p.Ala363Val	Tolerada (0.34)	Posiblemente dañina (0.492)	4.295	1	Causante de enfermedad	0.258	2 de 4
BMP7	p.Arg150Cys	Deletérea (0)	Probablemente dañina (0.952)	2.863	0.998	Causante de enfermedad	0.658	4 de 4
AMN	p.Met69Ile	Tolerada (0.05)	Benigna (0.016)	2.147	0.98	Causante de enfermedad	0.614	2 de 4
IDO2	p.Phe180Cys	Deletérea (0)	Probablemente dañina (1)	3.591	0.978	Causante de enfermedad	0.936	4 de 4
	p.Phe250Ser	Deletérea (0)	Probablemente dañina (0.994)	2.707	1	Causante de enfermedad	0.795	4 de 4
CR1	p.Thr1501Ala	Tolerada (0.3)	Benigna (0.076)	-0.206	0	Polimorfismo	0.540	1 de 4
TLR3	p.Ala795Val	Tolerada (0.14)	Probablemente dañina (0.961)	4.188	1	Causante de enfermedad	0.716	3 de 4
TRAF3IP1	p.Arg139Trp	Tolerada (0.07)	Probablemente dañina (0.928)	0.261	0.006	Polimorfismo	0.603	2 de 4
NCOA1	p.Ser671Ala	Tolerada (0.64)	Benigna (0.041)	0.447	0.997	Polimorfismo	0.270	Ninguno

Tabla 8. Características clínicas de las pacientes AER francesas y colombianas con variantes nuevas identificadas en genes candidato mediante NGS y confirmadas por secuenciación directa.

Origen étnico	Código de paciente	Número total de pérdidas	Pérdidas tempranas (< 10 semanas de gestación)	Pérdidas tardías (≥ 10 semanas de gestación)
FRANCIA	Pt-4	3	3	
	Pt-5	3	3	
	Pt-11	3	3	
	Pt-14	3	3	
	Pt-16	3	3	
	Pt-17	3	3	
	Pt-19	3	3	
	Pt-20	3	3	
	Pt-22	3	3	
	Pt-26	3	3	
	Pt-28	3	3	
	Pt-29	3	3	
	Pt-31	4	4	
	Pt-35	2		2
	Pt-36	4	4	
	Pt-37	2		2
	Pt-39	4	4	
COLOMBIA	Pt-45	3	3	
	Pt-46	3	2	1
	Pt-48	4	3	1

2.3.2 Fragmentación molecular orbital (FMO) de mutaciones en proteínas candidatas

El análisis de FMO se realizó para determinar el efecto potencial que causaban las mutaciones de FGA p.Phe685Cys (**Figura 4**) y MMP-10 p. Asp199Asn (**Figura 5**) en la interacción con los residuos cercanos de la estructura cristalizada de las proteínas. Este análisis se realizó en estas proteínas porque eran las únicas que dentro de su estructura cristalizada incluían las mutaciones de interés. En la estructura de FGA Phe^{685} corresponde al amino ácido en la posición Phe^{666}. El análisis de FMO sugirió que el cambio por una Cys^{666} produjo un efecto estructural al desestabilizar las interacciones alrededor del sitio catalítico (**Figura 4A, D**). Dos interacciones importantes se encontraron en la estructura de la proteína WT: un puente de hidrógeno formado entre Phe^{666} y Tyr^{673}, y una interacción no clásica de un puente de hidrógeno y un CH-π entre Phe^{666} y Arg^{659}. También se evidenciaron otras interacciones formadas por puentes de hidrógeno no clásicos entre Phe^{666}, Trp^{689} y Phe^{677} (**Figura 4B, E**). La sustitución por una Cys^{666} resultó en un cambio significativo en la energía total de interacción de 39,67 (WT) a -2,38 (MT) kcal/mol. Adicionalmente, se perdieron 3 puentes de hidrógeno no clásicos (interacciones CH-π) y se generaron dos interacciones dominadas por efectos de repulsión conducentes a la desestabilización de la estructura de la proteína: interacción del grupo tiol de Cys^{666} con Tyr^{673} y con las cadenas laterales de Asn^{799} (**Figura 4C, F**).

Un análisis similar realizado en la estructura cristalizada de MMP-10 WT identificó 3 interacciones mayores con el residuo Asp^{199}: un puente de hidrógeno entre Asp^{199} y Lys^{165}, interacciones electrostáticas entre Asp^{199} y el ion zinc (Zn^{2+}) coordinado por 4 residuos (Bloque de Zn: His^{167}, His^{182}, His^{195} y Asp^{169}) y una interacción electrostática entre Asp^{199} y His^{178} (**Figura 5A, B, E**). Al realizar el análisis de FMO con la sustitución Asn^{199} MT solo se identificó una interacción y la pérdida de las interacciones mayores antes mencionadas (**Figura 5A, C, F**) lo que se reflejó en el cambio de la energía total de interacción entre la estructura WT y MT (-108.45 kcal/mol WT vs. -11.98 kcal/mol MT) generando la potencial desestabilización de la proteína. Estos resultados sugirieron que las mutaciones encontradas en estas dos

proteínas en pacientes con AER podrían generar efectos patogénicos por desestabilización de las estructuras proteicas debido a la pérdida de interacciones energéticas con los residuos cercanos a los sitios de las mutaciones.

Los resultados antes mencionados fueron publicados en la revista **PLOS One** (IF: 2.8 – 2016) bajo el título ***"Novel genes and mutations in patients affected by recurrent pregnancy loss"*** (Ver **Anexo II**). Además, fueron presentados en la modalidad de poster y presentación oral en el **"67th Annual Meeting of the American Society of Human Genetics"** y en el **"3rd World Congress on Recurrent Pregnancy Loss (WCRPL 2018)"**, respectivamente.

Figura 4. Resultados de FMO para FGA WT y mutante (p.Phe685Cys). **A)** amino ácidos interactuando con Phe^{666}. Interacciones conservadas entre WT y MT sombreadas en gris. **B)** Interacciones WT con Phe^{666}. **C)** Interacciones con Cys^{666} (MT). Barras describen el tipo de interacciones: electrostáticas (verde), repulsión (rojo), transferencia de carga (azul), dispersión (amarillo) y solvatación (cian). Valores positivos son considerados desestabilizantes y los negativos estabilizantes. **D)** Vista del sistema analizado. Cuadro rojo zona de la mutación. **E)** Interacciones WT con Phe^{666}. **F)** Interacciones con Cys^{666} (MT). Las energías son expresadas en kcal/mol.

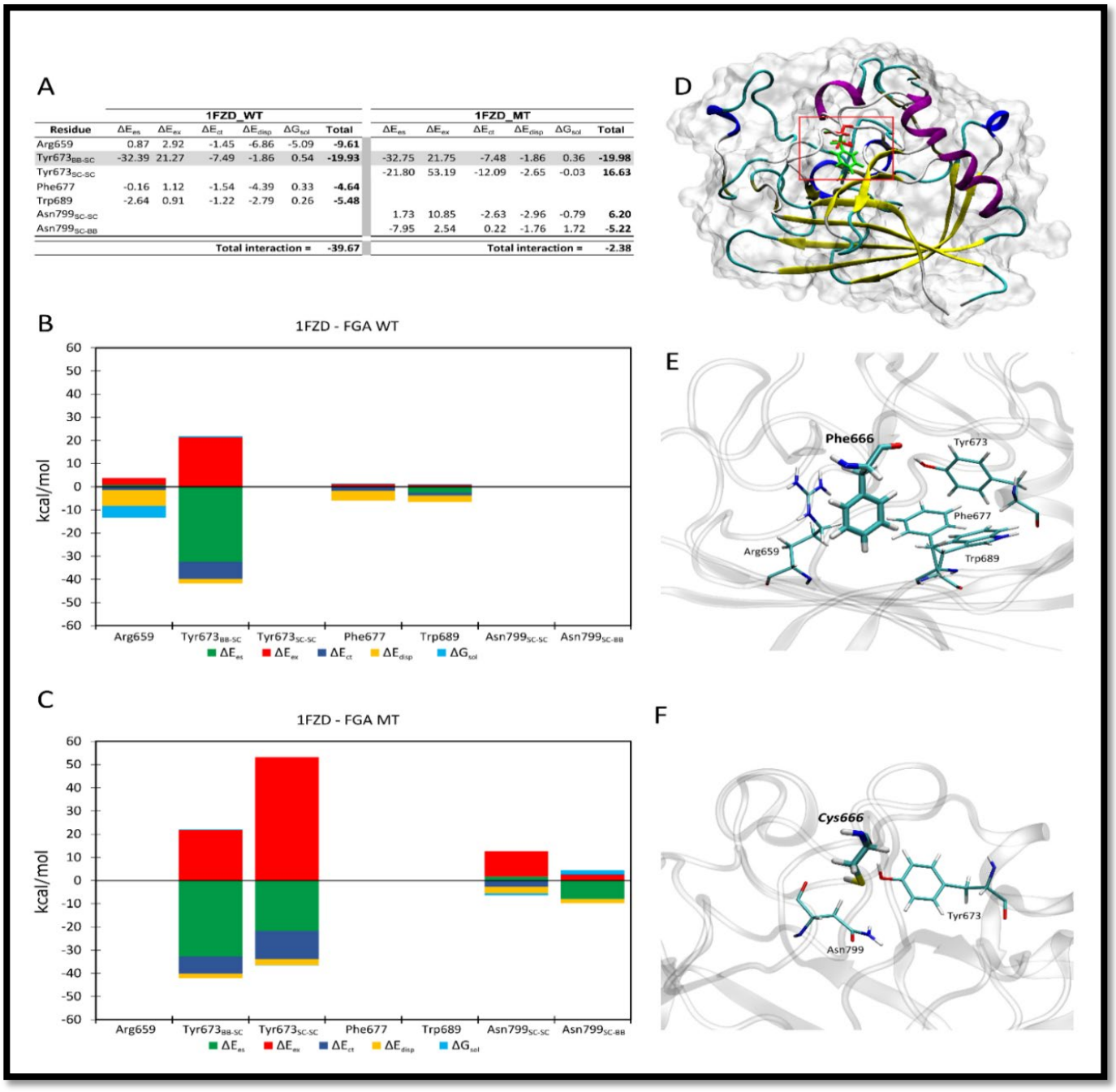
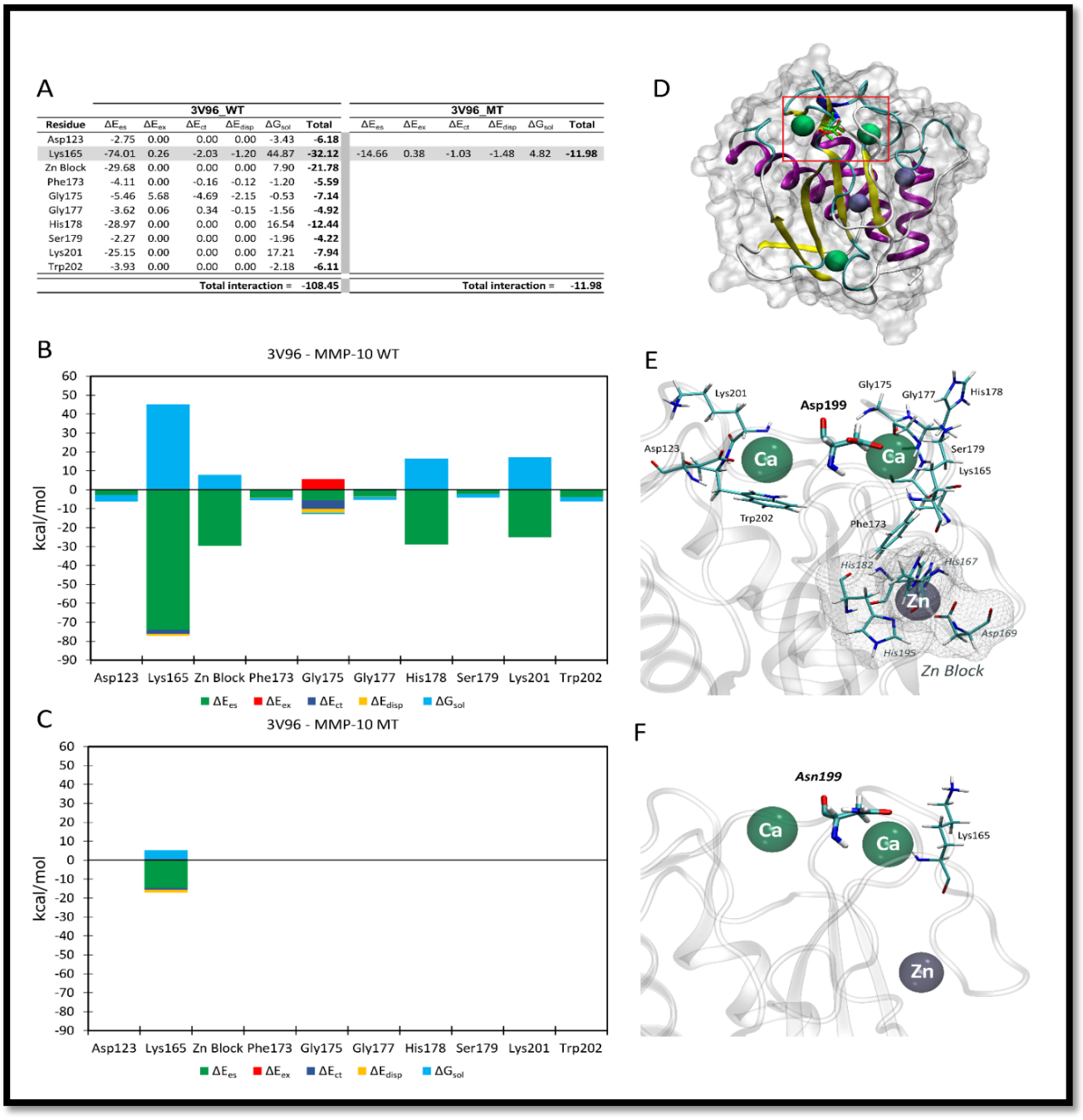


Figura 5. Resultados de FMO para MMP-10 WT y mutante (p. Asp199Asn). **A)** Amino ácidos interactuando con Asp[199]. Interacciones conservadas entre WT y MT sombreadas en gris. **B)** Interacciones WT con Asp[199]. **C)** Interacciones con Asn[199] (MT). Barras describen el tipo de interacciones: electrostáticas (verde), repulsión (rojo), transferencia de carga (azul), dispersión (amarillo) y solvatación (cian). Valores positivos son considerados desestabilizantes y los negativos estabilizantes. **D)** Vista del sistema analizado. Cuadro rojo zona de la mutación. **E)** Interacciones WT con Asp[199]. **F)** Interacciones con Asn[199] (MT). Las energías son expresadas en kcal/mol.



2.3.3 Secuenciación directa y ensayos funcionales de las variantes de THBD

2.3.3.1 Secuenciación de la región codificante de *THBD* en pacientes colombianas con AER

Las 262 pacientes con AER se dividieron en subgrupos teniendo en cuenta el número de AER que habían presentado: el subgrupo G1 correspondía a pacientes con 2 AER, el G2 a pacientes con 3 AER, el G3 a 4 AER y el G4 a >4 AER.

El 59,9% de las pacientes (n=157) pertenecían al G1, mientras que el 33,2% pertenecían al G2 y G3 (n=69 y n=18, respectivamente). El G4 lo conformaban 18 pacientes, representando el 6,9%. Los resultados de la secuenciación directa de toda la región codificante del gen *THBD* mostraron 2 variantes missense (c.457 T>G - p.Trp153Gly y c.1418 C>T- p.Ala473Val). La variante *THBD* c.457 T>G (p.Trp153Gly) fue encontrada únicamente en las pacientes colombianas y no había sido reportada previamente en las bases de datos públicas de SNPs. Para esta variante se encontraron diferencias significativas en las frecuencias alélicas entre los casos y los controles del mismo origen étnico ($p= 9 \times 10^{-6}$) (**Tabla 9**). La variante *THBD* c.1418 C>T (p.Ala473Val) presentaba un MAF de 0,16 en la base de datos de Ensembl. Se encontraron diferencias estadísticamente significativas ($p= 0,02$) entre el grupo de pacientes francesas con AER e individuos europeos reportados en la base de datos de Ensembl. Entre las pacientes colombianas y el grupo control no se encontraron diferencias estadísticamente significativas para esta variante la cual se encontraba con mayor frecuencia en el G1 (**Tabla 9**). Estos resultados sugieren que la variante *THBD* c.457 T>G (p.Trp153Gly) es específica de la población colombiana y podría ser causal del AER en mujeres de este origen étnico.

Los resultados antes mencionados fueron publicados en la revista **Reproductive Biology and Endocrinology** (IF: 2.8 – 2016) bajo el título ***"THBD sequence variants potentially related to recurrent pregnancy loss."*** (Ver **Anexo III**).

Tabla 9. Secuenciación del ORF de *THBD* en pacientes con AER y FIR e individuos control

Pacientes con AER francesas														
ADN	Proteína	Cigocidad (n)	Rs	Grupo AER	Pacientes	European controls (Ensembl)	MAF	Frecuencia alélica (casos)		Frecuencia alélica (controles)		*Valor P*	SIFT	PolyPhen	Conservación evolutiva
c.1418 C>T	p.Ala473Val	Homocigoto (3) Heterocigoto (29)	rs1042579	G1: 65,6% G2: 25% G3: 9,4%	32/233	173/503	0.16	C: 0,92	T: 0,08	C: 0,80	T: 0,20	0,02	Tolerado (0,42)	Beningno (0)	Si
Pacientes con AER colombianas															
ADN	Proteína	Cigosidad (n)	Rs	Grupo AER	Pacientes	Colombian Controls	MAF	Frecuencia alélica (casos)		Frecuencia alélica (controles)		*Valor P*	SIFT	PolyPhen	Conservación evolutiva
c. 457 T>G	p.Trp153Gly	Homocigoto (3) Heterocigoto (1)	-	G1: 50% G3: 25% G4: 25%	4/29	0/165	-	T: 0,88	G: 0,12	T: 1	G: 0	0,000009	Deletéreo (0)	Probablemente dañino(0.976)	Si
c.1418 C>T	p.Ala473Val	Heterocigoto (4)	rs1042579	G1: 50% G2: 25% G4: 25%	4/29	51/165	0.16	C: 0,93	T: 0,07	C: 0,95	T: 0,05	0,73	Tolerado (0,42)	Beningno (0)	Si

2.3.3.2 Estudio de la expresión de THBD WT y mutante en células THP-1

Para identificar la expresión endógena de THBD, se realizó un Western blot utilizando el extracto total de proteínas de células THP-1 no transfectadas (**Figura 6**). La expresión de las versiones de THBD WT y mutante (luego de su sobre-expresión) en la membrana celular de THP-1 se analizó por citometría de flujo. Se demostró que ambas versiones de la proteína se expresan en la membrana celular, sin diferencias estadísticamente significativas ente ellas. Cuando se permeabilizaron las células, se evidenció un aumento en las intensidades de fluorescencia en comparación con las intensidades de la membrana célular, tanto para las células transfectadas con la versión WT, como para la mutante. Sin embargo, no se encontraron diferencias estadísticamente significativas entre las intensidades de fluorescencia de las células permeabilizadas vs. las intensidades de la membrana celular con cada una de las versiones de THBD (WT y mutante) (**Figura 7**).

Figura 6. Expresión endógena de THBD en la línea celular THP-1. **M:** marcador de peso (Opti-Protein XL Marker). **1:** THBD (104 KDa). **2:** β-actina (42 KDa).

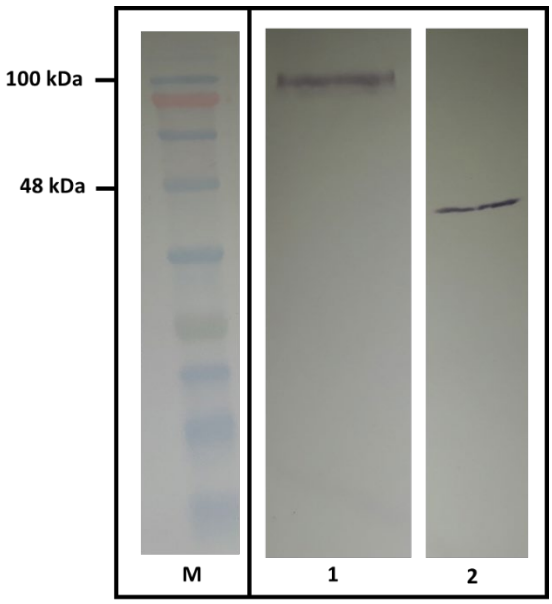
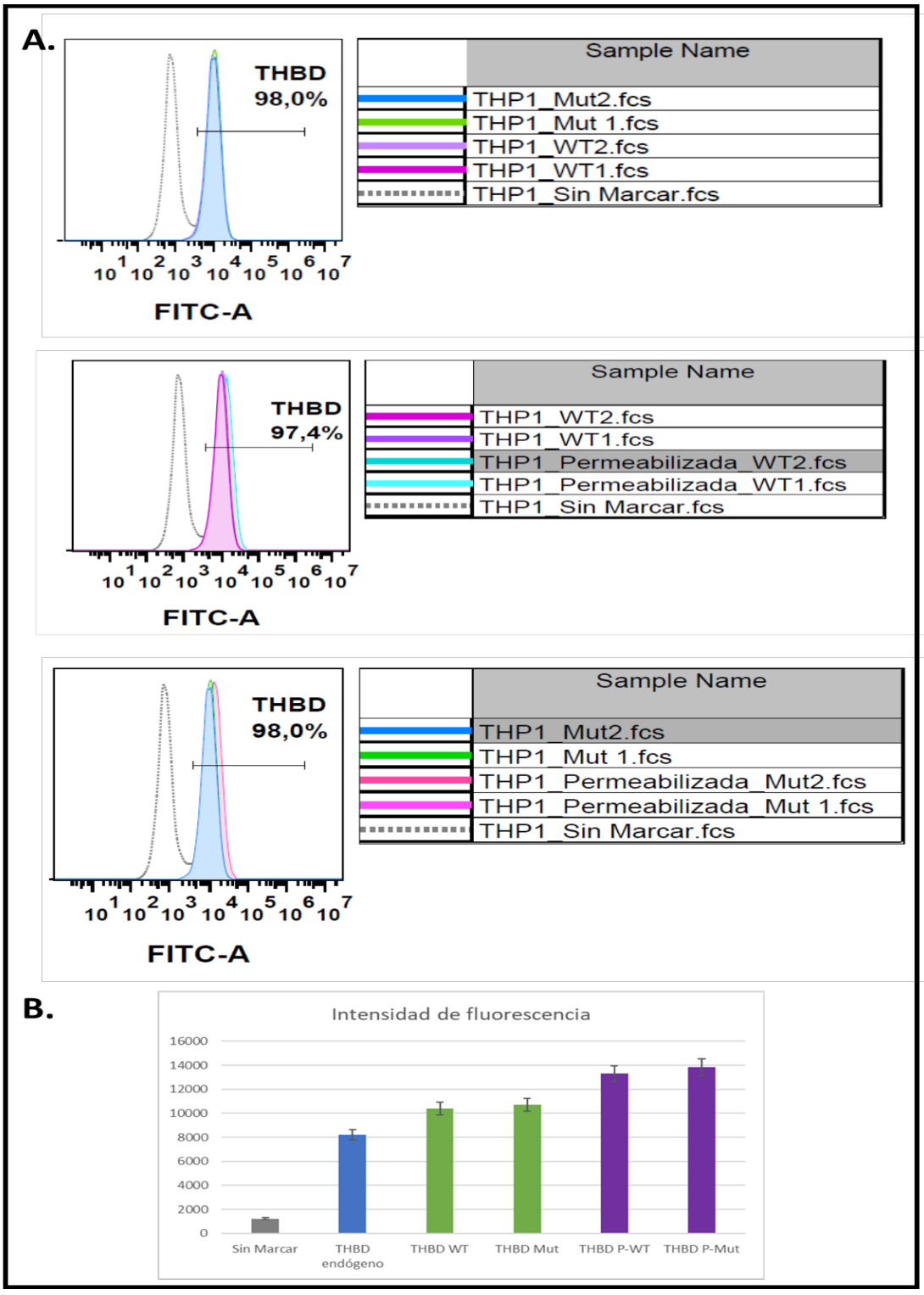


Figura 7. Citometría de flujo de células THP-1 transfectadas con la versión THBD WT y mutante. **A.** Expresión en la membrana celular de las células THP-1 transfectadas con la versión THBD WT y mutante. **B.** Intensidades de fluorescencia en células permeabilizadas y sin permeabilización.



2.3.3.3 Estudio de la localización subcelular de THBD WT y mutante

Para determinar la localización subcelular de THBD WT y mutante (THBD - p.Trp153Gly), se transfectaron células CHO y COS-7 usando una construcción con el ORF de la proteína en fase con GFP y se usó en paralelo la técnica de IFI para ambas líneas celulares y para las células THP-1. Se identificaron señales de fluorescencia, tanto en la membrana citoplasmática, como en el citoplasma de estas células mediante ambas aproximaciones.

Se describieron diferentes perfiles de localización subcelular: la delimitación o no de la membrana celular (MCD+ y MCD-, respectivamente), la presencia o no de agregados-like en la región perinuclear (AG+ y AG-, respectivamente) y, a su vez, esta categoría se dividió dependiendo de si los agregados-like eran masivos (+) o no (-) (**Figuras 8** y **9**). El conteo celular con los perfiles previamente descritos demostró que tanto la versión WT de la proteína como la mutante podían localizarse en la membrana celular. Sin embargo, se encontraron diferencias estadísticamente significativas con la presencia de agregados-like vs. la ausencia de estos agregados en la región perinuclear entre la versión WT (68,5% y 31,5%, respectivamente) y mutante (78% y 22%, respectivamente) (p=0,041). Para el perfil de agregados-like perinucleares masivo (+) y (-) se encontraron diferencias estadísticamente significativas entre las células transfectadas con la versión WT (21% y 47,5%, respectivamente) y mutante (38% y 40%, respectivamente) de la proteína (p= 0,0019). Para la combinación de los perfiles MCD(-)/AG(+)/Masivo(+) y MCD(-) /AG (+)/Masivo(-) se encontró también diferencias estadísticamente significativas (p=0,0237) entre la versión WT (13,5% y 32%, respectivamente) y mutante (22% y 25,5%, respectivamente) (ver **Tabla 10**).

En las células THP-1 transfectadas con las versiones WT y mutante de THBD, se identificó señal de fluorescencia en el citoplasma de las células, sin delimitación de la membrana celular. Los perfiles que se describieron de localización subcelular fueron la presencia o no de agregados-like citoplasmáticos (AG+ y AG-, respectivamente) (**Figura 10**). El conteo celular con los perfiles previamente

descritos demostró diferencias estadísticamente significativas con la presencia de agregados-like vs. la ausencia de éstos entre la versión WT (26% y 74%, respectivamente) y mutante (63% y 37%, respectivamente) (p=1,05 e^{-13}) (ver **Tabla 11**).

Tabla 10. Categorías encontradas por IFI en células CHO transfectadas con la versión WT y mutante de THBD

Categorías	WT - número de células (%)	MUT - número de células (%)	Valor *p* (Fisher´s test)
MCD (+) MCD (-)	79 (39,5%) 121 (60,5%)	87 (43,5%) 113 (56,5)	0,4775
AG (+) AG (-)	137 (68,5%) 63 (31,5%)	156 (78%) 44 (22%)	**0,0418**
Masivo (+) Masivo (-)	42 (21%) 95 (47,5%)	76 (38%) 80 (40%)	**0,0019**
MCD (+) / AG (+) MCD (+) / AG (-)	46 (23%) 33 (16,5%)	61 (30,5%) 26 (13%)	0,1438
MCD (+) /AG (+)/MASIVO (+) MCD (+) /AG (+)/MASIVO (-)	15 (7,5%) 31 (15,5%)	32 (16%) 29 (14,5%)	0,0501
MCD (-) / AG (+) MCD (-) / AG (-)	91 (45,5%) 30 (15%)	95 (47,5%) 18 (9%)	0,1065
MCD (-) /AG (+)/MASIVO (+) MCD (-) /AG (+)/MASIVO (-)	27 (13,5%) 64 (32%)	44 (22%) 51 (25,5%)	**0,0237**

Tabla 11. Categorías encontradas por IFI en células THP-1 transfectadas con la versión WT y mutante de THBD.

Categorías	WT Número de células (%)	Mut Número de células (%)	Valor *p* (Fisher´s test)
Agregados (+)	52 (26%)	126 (63%)	**1,05 e^{-13}**
Agregados (-)	148 (74%)	74 (37%)	

Figura 8. Perfiles de localización subcelular de la línea celular CHO transfectada con las construcciones de THBD WT y mutante (THBD- p.Trp153Gly).

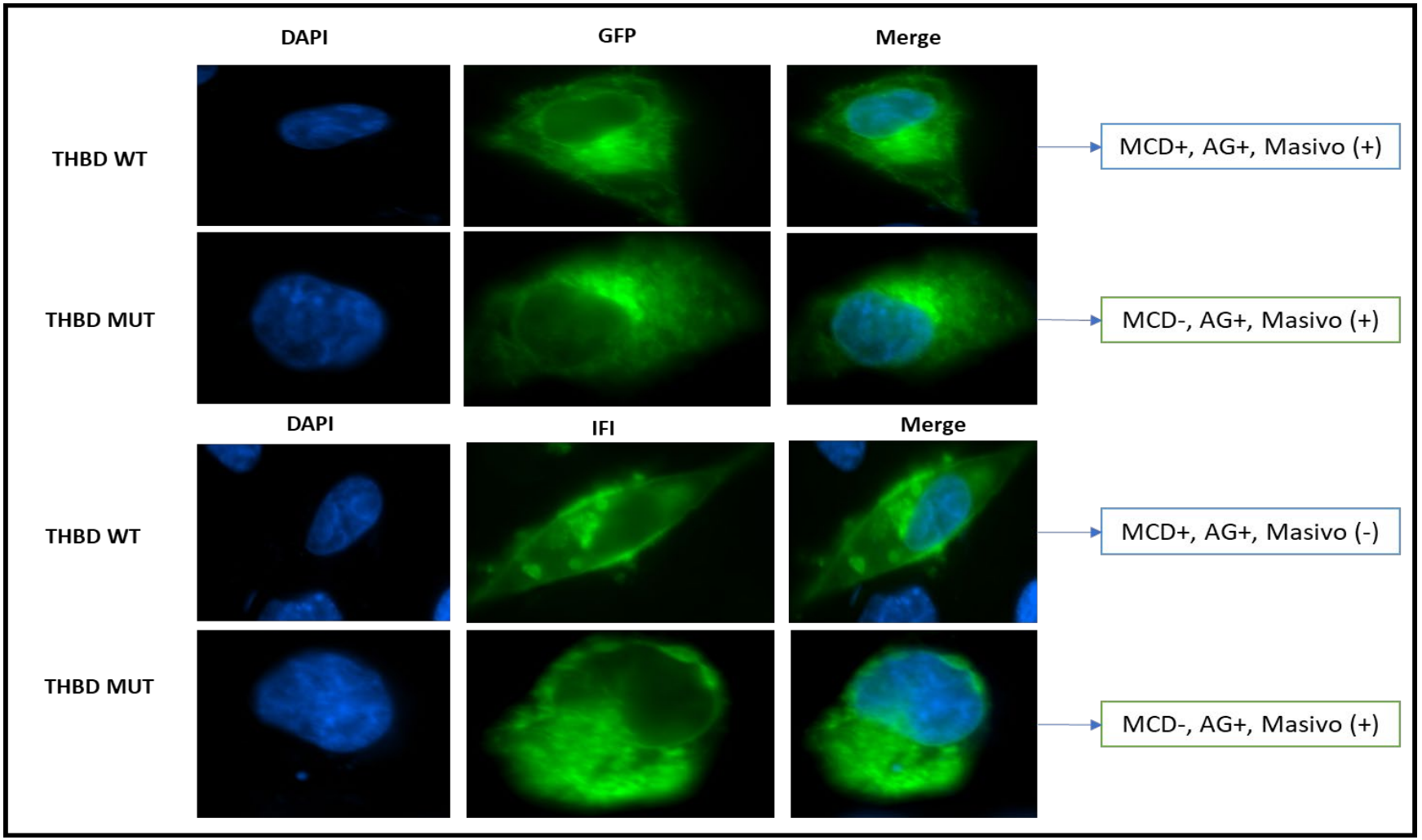


Figura 9. Perfiles de localización subcelular de la línea celular COS-7 transfectada con las construcciones de THBD WT y mutante (THBD- p.Trp153Gly).

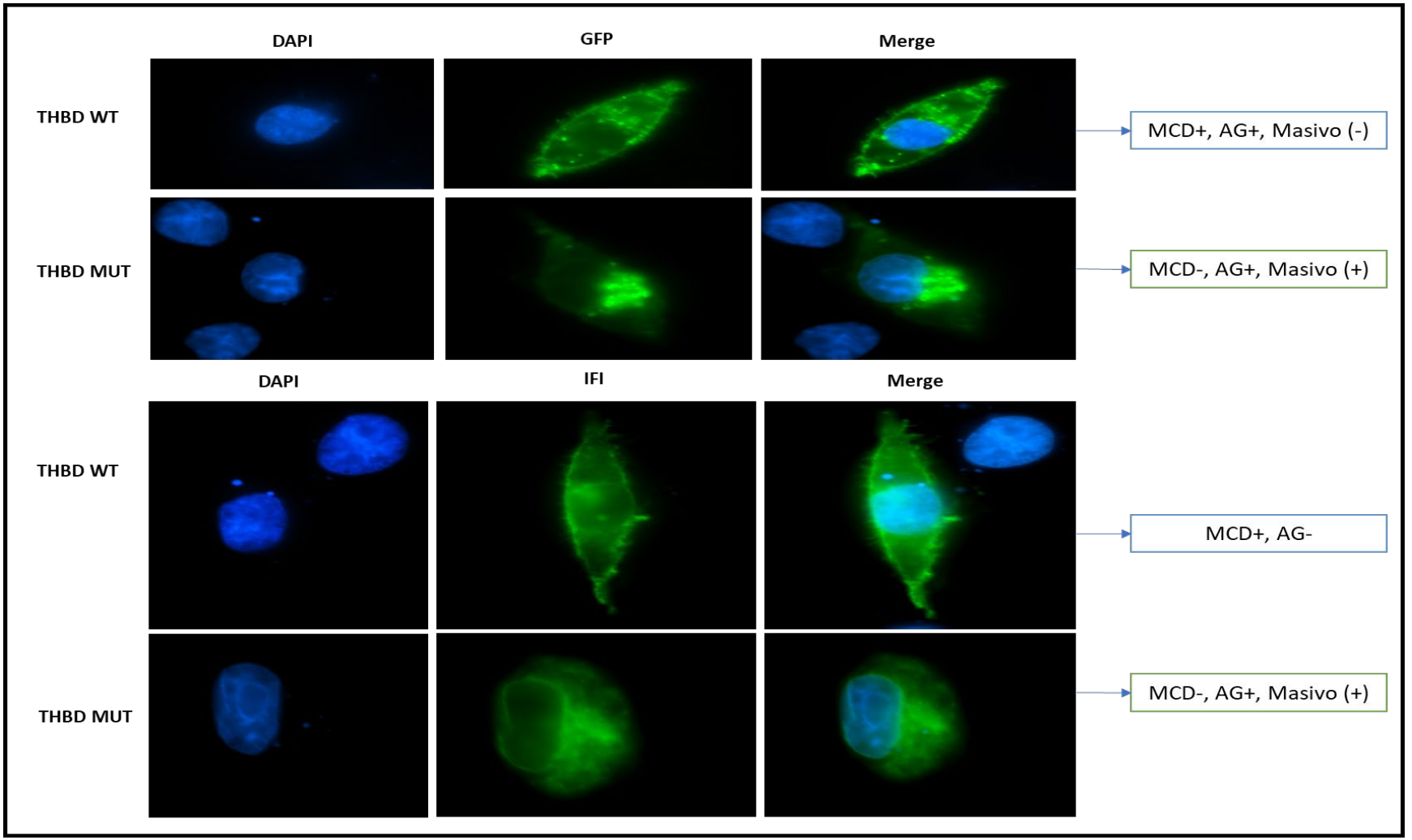
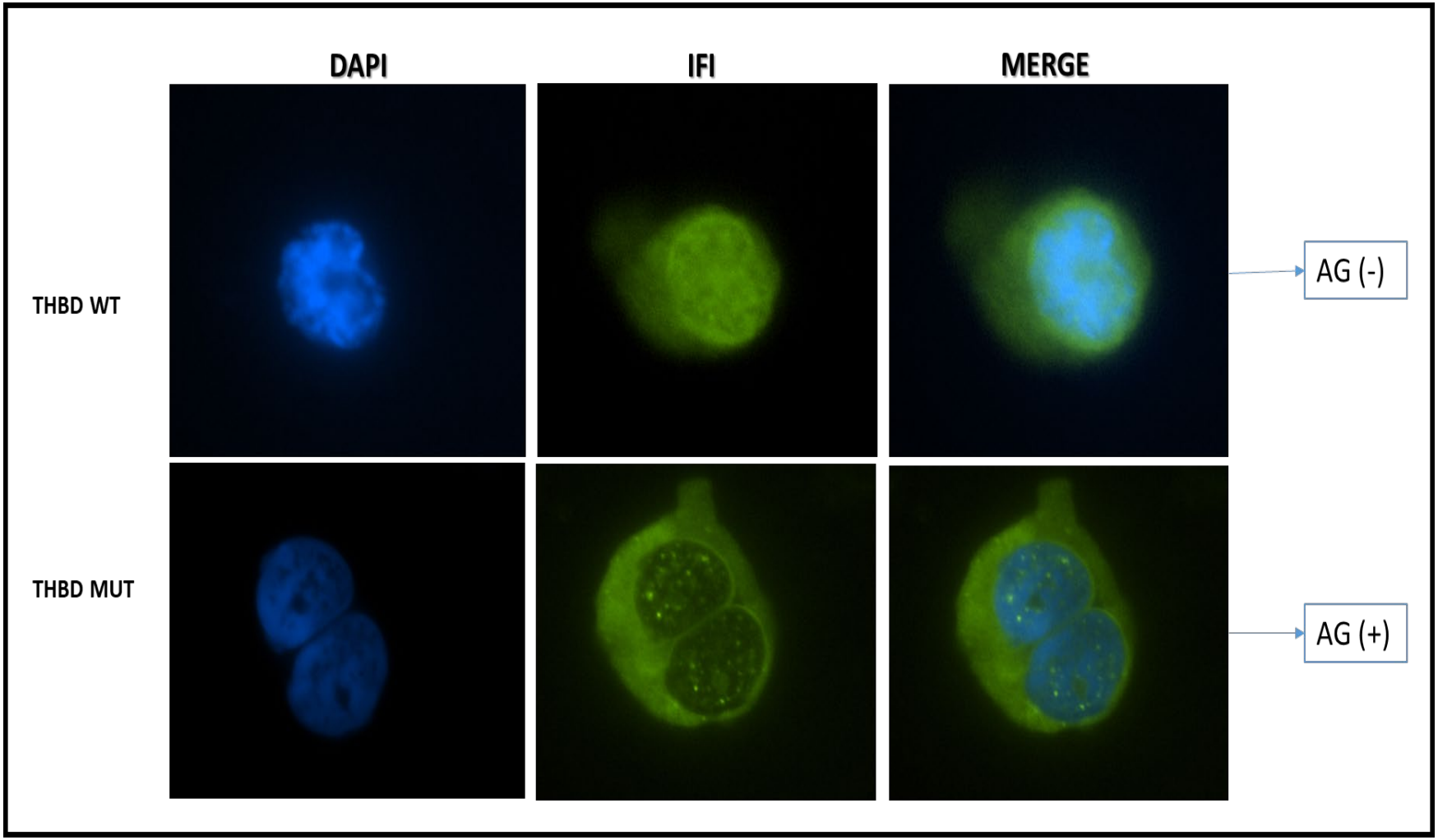


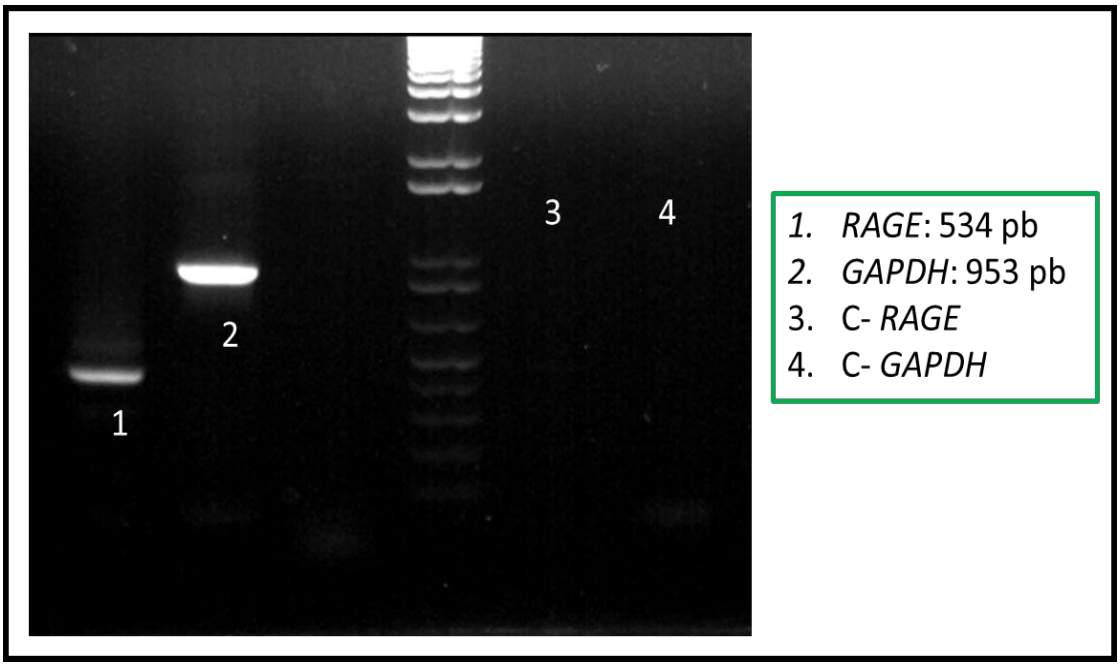
Figura 10. Perfiles de localización subcelular de la línea celular THP-1 transfectada con las construcciones de THBD WT y mutante (THBD- p.Trp153Gly).



2.3.3.4 Estudio del efecto de la mutación THBD p.Trp153Gly en la expresión de genes de la cascada corriente abajo de la citoquina HMGB1

La mutación de THBD-p.Trp153Gly se encuentra dentro del dominio tipo lectina (C-type lectin domain - CTLD) de la proteína, el cual se ha visto involucrado en procesos anti-inflamatorios por el secuestro de la citoquina HMGB1. Este secuestro permite la hidrólisis de la citoquina y evita su unión al receptor con función pro-inflamatoria RAGE. La expresión endógena de éste en las células THP-1 se evidenció por medio de RT-PCR (**Figura 11**).

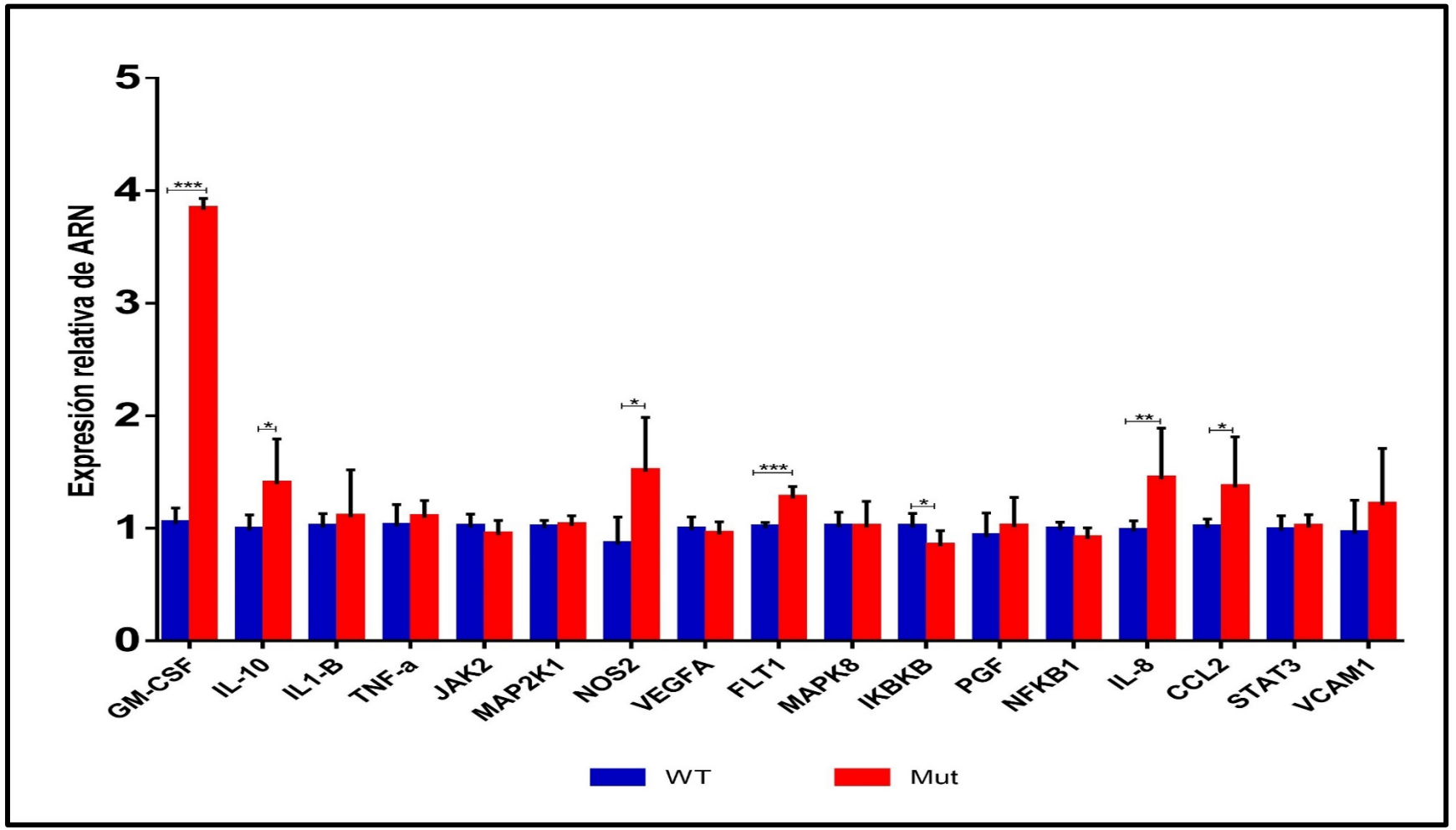
Figura 11. Expresión endógena del receptor RAGE en células THP-1.



Para determinar si la mutación generaba la expresión diferencial de algunos transcritos corriente-abajo de la cascada pro-inflamatoria de HMGB1, se realizó qPCR con el ADNc de las células THP-1 transfectadas con cada versión de la proteína y estimuladas con HMGB1.

Los resultados de este ensayo demostraron que los genes *GM-CSF* (aumento de 3,8 veces, *p* <0,0001)*, IL-10* (aumento de 1,4 veces, *p*=0,018)*, NOS2* (aumento de 1,8 veces, *p*=0,011)*, FLT1* (aumento de 1,3 veces, *p* <0,0001)*, IKBKB* (disminución de 1,2 veces, *p*=0,02)*, IL-8* (aumento de 1,5 veces, p=0,008) y *CCL2* (aumento de 1,3 veces, p=0,0229) se encontraban diferencialmente expresados con significancia estadística entre el WT y el mutante. Además, se demostró un aumento en la expresión de los genes *IL-1B, TNF-α, MAP2K1, PGF, STAT3* y *VCAM1* entre las células transfectadas con la versión WT y mutante, sin diferencias estadísticamente significativas (**Figura 12**).

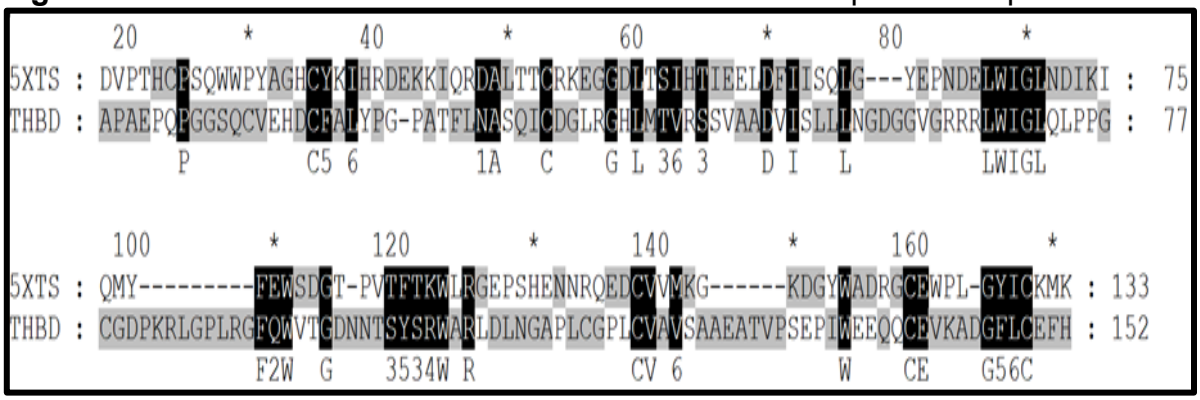
Figura 12. Expresión diferencial de los genes que participan en la cascada pro-inflamatoria de HMGB1 en células THP-1 transfectadas con la versión WT y mutante de THBD. * valor *p <0.05,* ** valor *p <0.01* y *** valor *p <0.001*.



3.3.3.5 Modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD

El BLASTp para el modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD identificó que la estructura cristalográfica con mayor similitud fue el receptor de manosa CTLD3 humano (PDB ID: 5XTS), con un 30% de similitud, y un 16% de identidad (BLOSUM62 matriz) (**Figura 13**).

Figura 13. Alineamiento de los dominios lectina analizados por BLASTp.



El mejor modelo calculado por homología en I-TASSER para el dominio lectina de THBD resulta tener una calidad estructural pobre (contactos a menos de 0,4 Å, rotámeros no favorecidos y divergencias en el mapa de Ramachandran), mostrando un puntaje para Molprobity de 3,19 (siendo 0 el óptimo) y de -6,73 para QMEAN4 (siendo valores < -4,0 indicadores de baja calidad). Después del proceso de refinamiento, la calidad de los modelos mejoró significativamente. El puntaje para Molprobity del modelo WT fue de 1,20 y para el modelo MT fue de 1,19. En QMEAN4, el modelo WT obtuvo un puntaje de -3,75 y el modelo MT de -3,80 (**Figura 14**).

La mutación p.Trp153Gly se encuentra en el interior de la proteína (**Figura 15**) y los modelos WT y MT tienen un RMSD de 3,032 Å considerando cadenas laterales y de 1,37 Å considerando únicamente el esqueleto. Esta diferencia implica un cambio de volumen de 186 $Å^3$ (WT=20958 $Å^3$ y MT=20772 $Å^3$), siendo mayor el volumen del modelo WT comparado con el MT.

Figura 14. QMEAN4 normalizado. Se comparan los modelos estudiados con estructuras cristalográficas reportadas. El refinamiento aproxima el modelo crudo a parámetros estructurales aceptables.

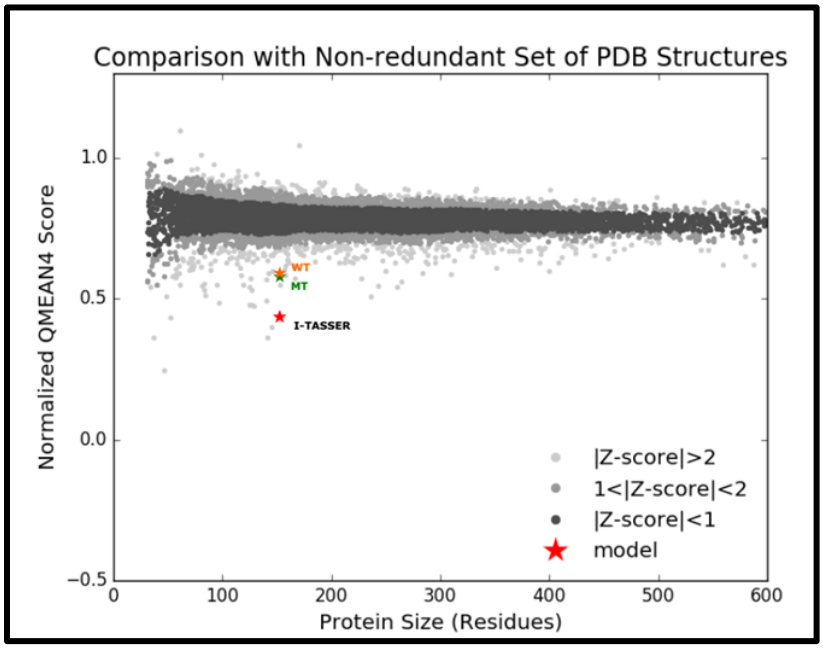
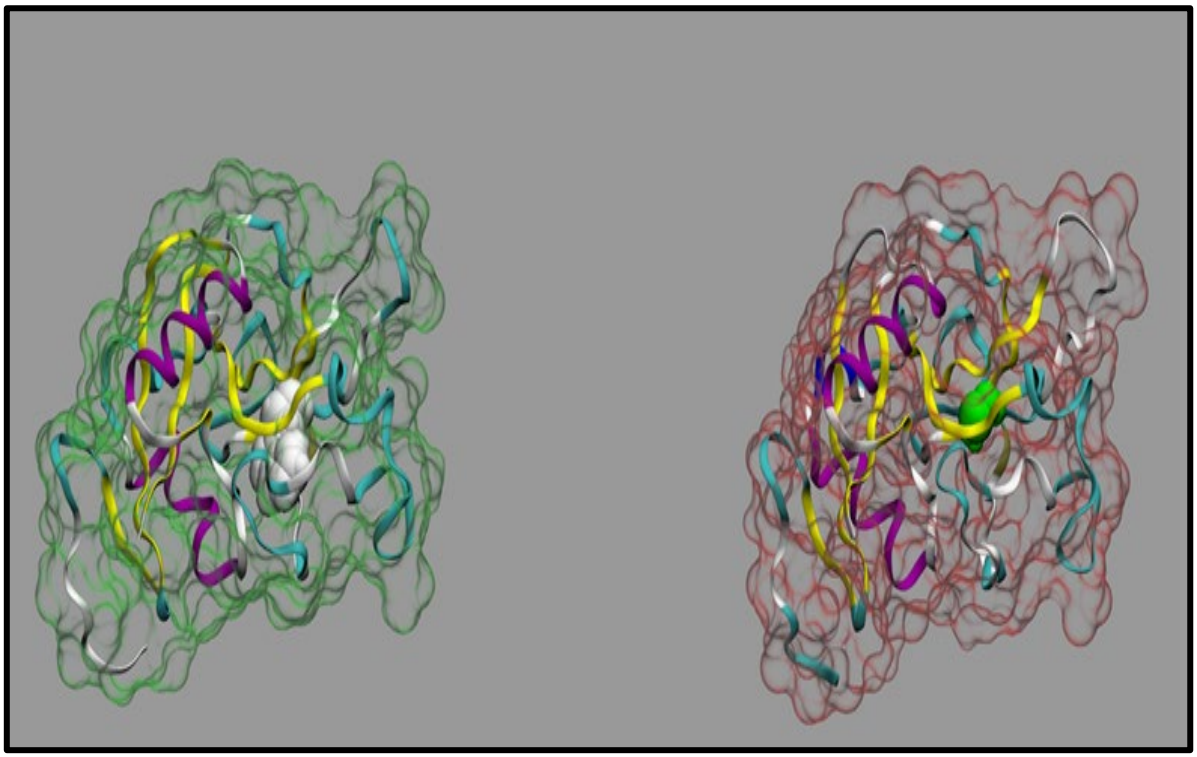
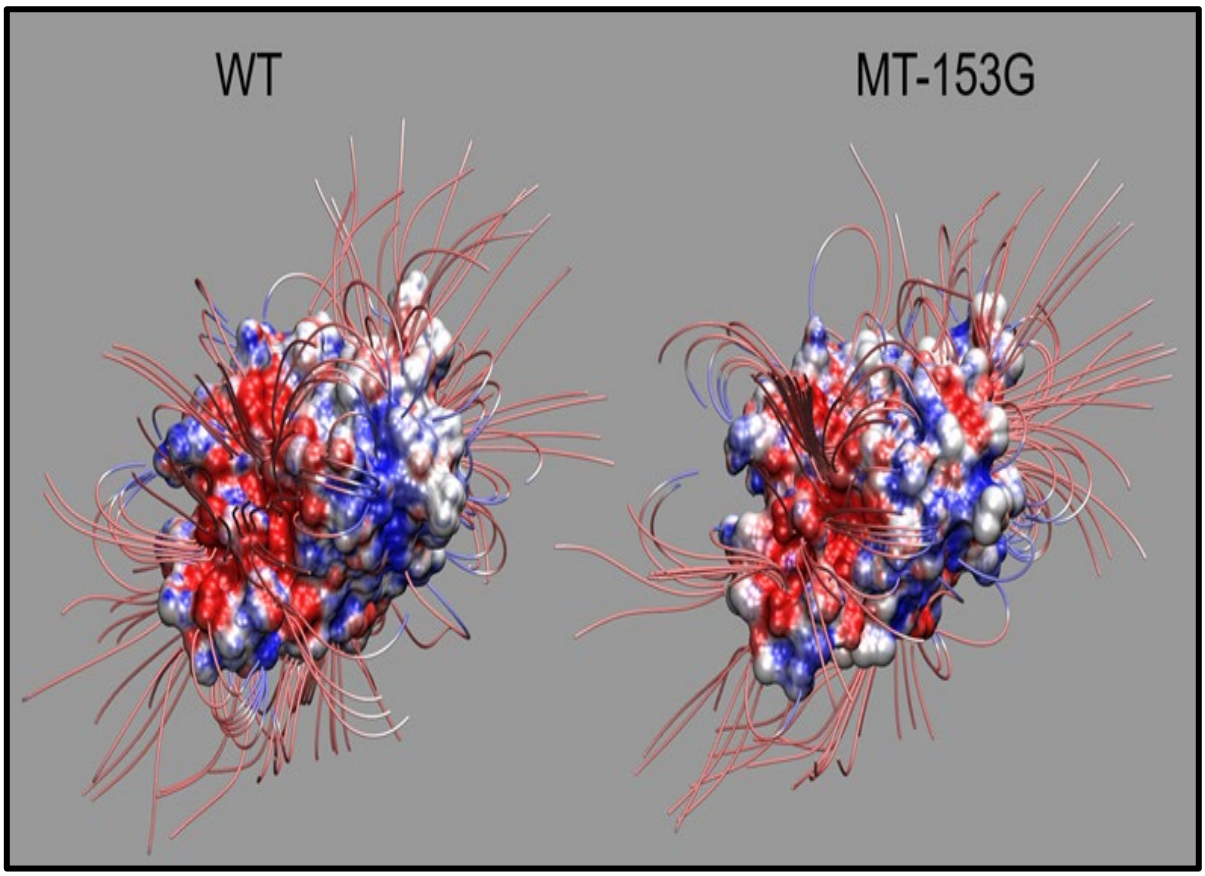


Figura 15. Modelo por homología del dominio de lectina de THBD. Modelo WT (verde) y MT (rojo) mostrando la posición 153 ubicada en el core hidrofóbico de la proteína, Trp (blanco) y Gly (verde).



Por otro lado, el cambio en el área superficial ocurre en sentido opuesto, siendo mayor el área superficial del modelo MT (8008 $Å^2$) comparada con el WT (7831 $Å^2$). Así mismo, al comparar el mapa de potencial electrostático para los modelos WT y MT, es notoria la diferencia en los rasgos ente éstos (**Figura 16**).

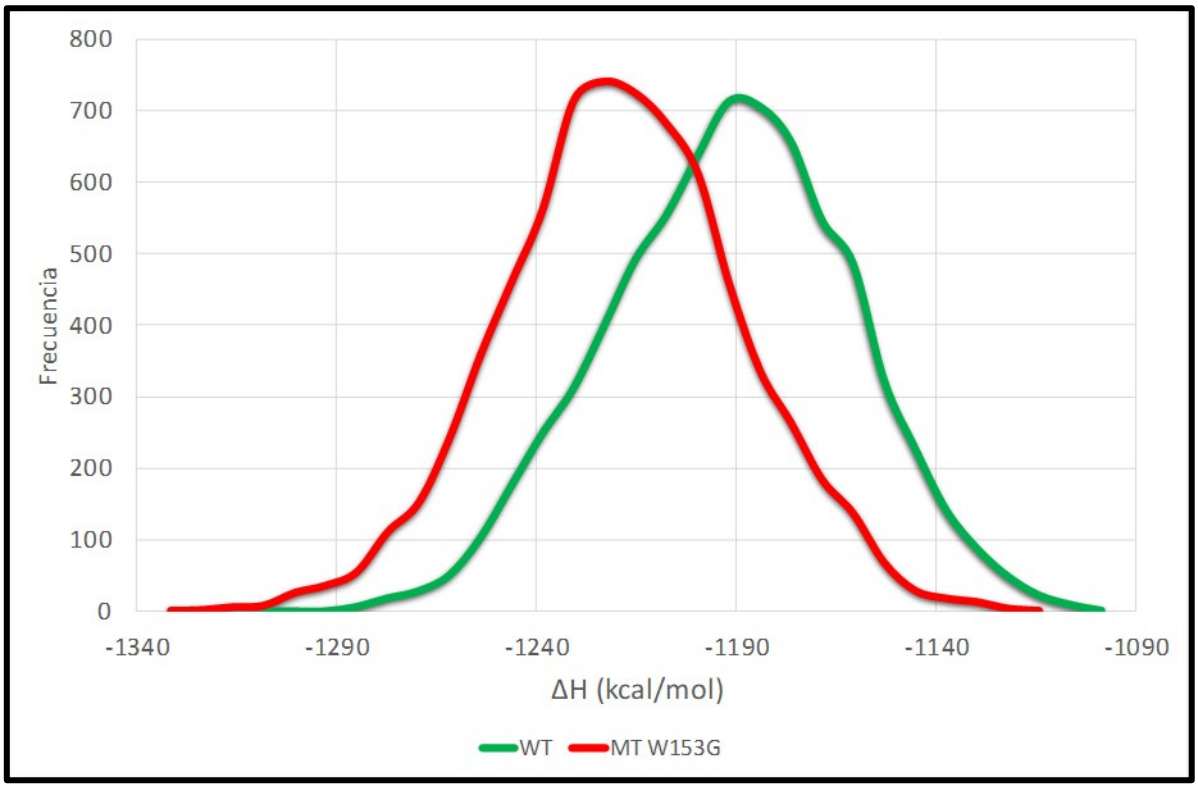
Figura 16. Mapas de potencial electrostático de los modelos WT y MT. Se evidencian diferencias en la distribución de cargas (azul=positivo, rojo=negativo) y en la disposición y concentración de las líneas de campo.



El cambio energético implicado en la mutación p.Trp153Gly muestra que, en promedio, el modelo MT (E= -1222,5 kcal/mol) es más estable que el WT (E= -1195,7 kcal/mol). La distribución de la energía calculada mostró una clara diferencia en la distribución para WT y MT (**Figura 17**). La diferencia entre las medias de las dos distribuciones es de 26,83 kcal/mol con un error estándar de 0,47 kcal/mol. Esta diferencia resulta significativa según el t-test efectuado (p=0,05). La diferencia

según la moda (valor más frecuente) de las distribuciones es de 34,89 kcal/mol (E_{WT}= -1196,01 kcal/mol y E_{MT}= -1230,90 kcal/mol).

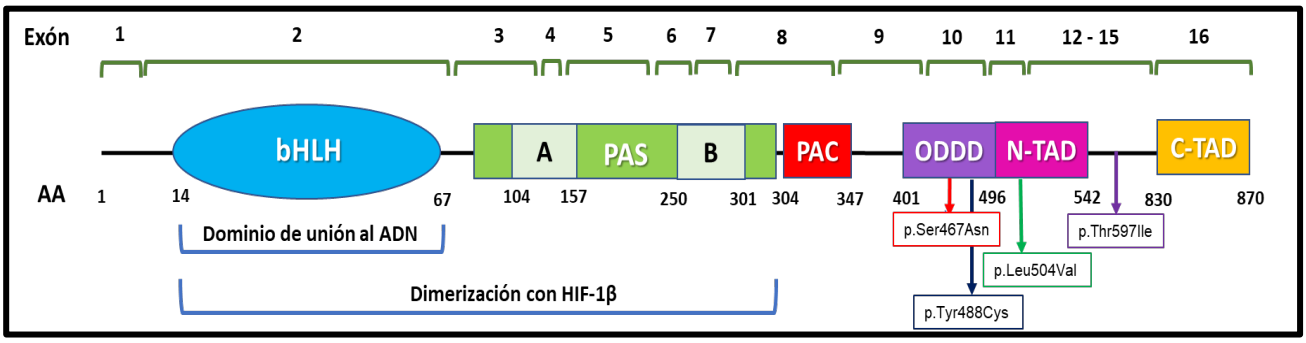
Figura 17. Distribución de las energías calculadas para los modelos THBD WT y THBD p.Trp153Gly.



2.3.4 Secuenciación directa y ensayos funcionales de las mutaciones de EPAS1

Los resultados de la secuenciación directa de las 224 pacientes francesas con AER para los exones codificantes para los dominios de degradación dependiente de oxígeno (exón 10) y para el de N-transactivación (exón 11 y 12) del gen *EPAS1* (**Figura 18**) mostraron 2 variantes missense (c.1400 G>A - p.Ser467Asn y c.1790 C>T- p.Thr597Ile).

Figura 18. Diagrama del gen *EPAS1* y de las mutaciones encontradas en el exoma y por secuenciación directa.



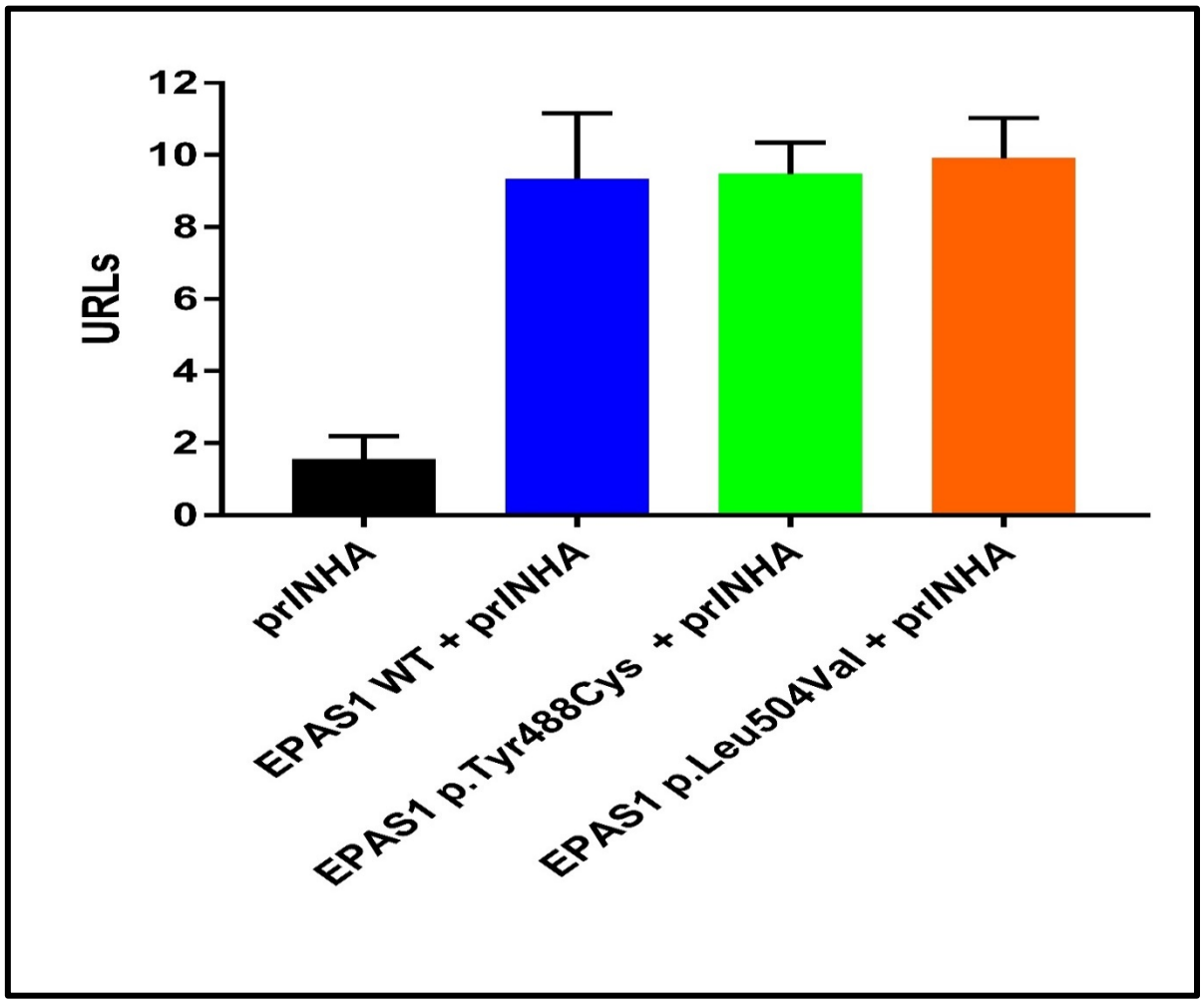
Cada variante fue encontrada en una paciente de las 224 tamizadas. Éstas habían sido reportadas previamente en las bases de datos públicas con un MAF <1%. Sin embargo, éstas no mostraron resultados potencialmente deletéreos en los predictores de patogenicidad SIFT y PolyPhen por lo cual no se consideraron con significancia funcional para realizar el screening en controles y para realizar estudios funcionales *in vitro* (**Tabla 12**).

Tabla 12. Secuenciación directa de exones 10, 11 y 12 de *EPAS1* en pacientes francesas con AER

EPAS1	Pacientes	Cambio Nt	Cambio AA	MAF	Rs	SIFT	PolyPhen	Conservación evolutiva
Exón 10	1/197	c. 1400 G>A	p.Ser467Asn	0.00003	rs772047070	Tolerado (0,42)	Benigno (0,004)	NO
Exón 11	2/216	c. 1512 C>T	p.Leu504Leu	___	___	___	___	SI
Exón 12	1/224	c.1790C>T	p.Thr597Ile	0.000008	rs777113349	Tolerado (0,19)	Benigno (0)	SI

Para el estudio gen reportero bajo condiciones de hipoxia de las mutaciones EPAS1 identificadas por NGS (p.Tyr488Cys y p.Leu504Val), hasta la fecha, se realizó un ensayo gen reportero de estandarización bajo condiciones de hipoxia química con el promotor del gen *INHA*. Los resultados de este primer ensayo no mostraron diferencias estadísticamente significativas entre la células transfectadas con la versión WT y las versiones mutantes (**Figura 19**).

Figura 19. Resultados de la estandarización ensayo gen reportero de EPAS1 WT y mutantes con el promotor de *INHA.*



2.4 Discusión capítulo I

2.4.1 Abordaje NGS para la búsqueda de mutaciones en pacientes con AER

La secuenciación de exoma que realizamos presentó una profundidad de secuenciación promedio de 30X. La profundidad de secuenciación es fundamental para disminuir los posibles falsos positivos en el llamado de variantes [141].

Se ha estimado que se necesita una profundidad de 40X a 80X en promedio para genotipificar con precisión >90% de los genes secuenciados por WES y para una sensibilidad adecuada en la detección de SNPs heterocigotos. Cuando se usa NGS con una profundidad promedio de 10X a 60X se recomienda realizar secuenciación directa para validar las variantes encontradas. Para fines diagnósticos es ideal una profundidad de 120X para identificar las variantes no-sinónimas potencialmente involucradas en una patología [97,142,143]. La profundidad de secuenciación de nuestro exoma fue suficiente para identificar numerosas variantes heterocigotas candidatas en genes de AER. Estas variantes fueron validadas por secuenciación directa.

El análisis de los exomas se efectuó principalmente en el subset de 234 genes candidato (234-RPL) que hacen parte de las cascadas moleculares involucradas en el proceso de implantación. Las variantes candidatas fueron seleccionadas mediante filtros dirigidos a la búsqueda de variantes nuevas (no reportadas en bases de datos), con cambios a nivel de proteína y en regiones evolutivamente conservadas. A diferencia de la aproximación de GWAS (que usa marcadores en desequilibrio de ligamiento y permite la identificación de variantes comunes), el exoma permite la identificación de variantes causales raras de enfermedades complejas. En nuestro estudio encontramos un enriquecimiento de las variantes missense en el subset de genes candidato (234-RPL) y escogimos las variantes con potenciales efectos deletéreos asociadas a la patología de AER. Este tipo de diseño permite el estudio clínico en pacientes individuales o en una población pequeña de 50-100 casos [104,144–146]. En cuanto al uso del subset de genes, *O'Roak et al.,* demostraron que en un subset de genes se identificó un exceso de variantes *de novo* asociadas al autismo, indicando que éstas tenían un efecto severo y que los genes de subset eran genes mayores que contribuían a la enfermedad [147]. Sin embargo, el exoma (en comparación con otras técnicas como el GWAS o el WGS) no permite identificar variantes en regiones promotoras, de splicing o UTRs. También, presenta baja cobertura en regiones repetitivas o ricas en GCs, lo que puede contribuir al fenómeno denominado heredabilidad perdida.

Con la metodología utilizada en el presente trabajo, se obtuvieron 27 variantes en 22 genes candidato que posiblemente podrían generar alteraciones funcionales (**Figura 3** y **Tabla 6**). Con esta aproximación encontramos que 20 de 49 pacientes presentaron variantes no sinónimas en los genes candidato y la mayoría presentaba pérdidas tempranas de la gestación. Siete de estos pacientes tenían más de una variante en el mismo o en diferentes genes, lo que podría sugerir un efecto aditivo epistático dentro de la patogénesis del AER. El proceso desde la fertilización hasta la implantación se ha conservado evolutivamente entre las diferentes especies de mamíferos (Lee and DeMayo, 2004). Por lo cual una gran cantidad de moléculas controlan e interactúan de manera regulada en las cascadas de fecundación, pre-implantación embrionaria, implantación y placentación. Esta complejidad molecular determina en parte la naturaleza poligénica de las patologías complejas de la reproducción (e.g. AER, insuficiencia ovárica prematura - IOP, FIR). *Fonseca D et al. (2015)* utilizaron un custom array de secuenciación NGS en pacientes no-relacionadas con IOP y demostraron que 25% de las pacientes presentaban mutaciones potencialmente deletéreas en diversos genes relacionados con esta fisiopatología [106]. En el caso de *Patiño L et al. (2017),* usando la aproximación NGS se evidenció que el 42% de las pacientes con IOP presentaban al menos 2 mutaciones en distintos genes, lo que demostró la característica poligénica de esta enfermedad [110]. Es importante señalar que recientemente nuestro grupo efectuó ensayos funcionales positivos de mutaciones identificadas en pacientes con IOP a partir de una metodología similar de selección de mutaciones [110,148]. Esto demostró que la aproximación de NGS y el estudio de un subset de genes (enriquecido por variantes con alteración a nivel de la proteína) podría considerarse como una herramienta útil para el tamizaje de pacientes con AER. Las variantes identificadas en nuestro NGS se encontraban en genes relacionados con procesos biológicos importantes durante la implantación como la interacción del trofoblasto al endometrio, las cascadas de la coagulación y la angiogénesis, la inmunomodulación, el metabolismo celular, la organización de la matriz extracelular, la proliferación y la migración celular, entre otros (**Tabla 6**). Dentro del análisis bioinformático, se realizó la predicción de patogenicidad de las variantes usando

programas con altos niveles de predicción positiva [149,150]. El 27% de la variantes fueron positivas para un efecto potencial deletéreo en los 4 programas incluyendo la mutación THBD-p.Trp153Gly, la cual fue seleccionada para los posteriores ensayos funcionales *in vitro* (**Tabla 7**).

2.4.2 Mutaciones candidatas en la etiología del AER

Entre las 27 variantes validadas por secuenciación de Sanger, se identificó una mutación en el gen *TRO* (c.2909G>A, p.Gly970Asp) el cual está implicado en la implantación ya que permite la interacción del trofoblasto con la membrana apical del endometrio (Fukuda and Sugihara, 2012, 2008). Esta mutación se encuentra en el dominio que contiene repeticiones en tándem de decapéptidos, por lo cual podría alterar las características fisicoquímicas de la proteína y su interacción con las proteínas tastina y bistina que en conjunto permiten la adhesión entre el trofoblasto y el endometrio. Además, se encontraron mutaciones en las proteínas implicadas en las cascadas de coagulación relevantes en el proceso de decidualización del endometrio como F5 (p.Glu1512Ala y p.Thr1950Pro), THBD (p.Trp153Gly) y FGA (p.Phe685Cys). Estudios previos han demostrado una asociación entre las variantes encontradas en el gen *F5* y el fenotipo de AER [151–153]. Las dos mutaciones encontradas en F5 se encuentran localizadas en dominios funcionales de la proteína. La mutación F5 p.Glu1512Ala se localiza en el dominio B de la proteína, el cual interactúa con F10 para activarla y es necesario para el clivaje realizado por la trombina para su inactivación [154–156]. La mutación F5 p.Thr1950Pro se localiza en el dominio C1 de la proteína que está involucrado en el ensamblaje del complejo protrombinasa y en la regulación de la trombina [157].

También se encontraron variantes relevantes en otros procesos biológicos de la implantación como la inmunomodulación (IDO2 p.Phe180Cys y p.Phe250Ser, TLR3 p.Ala795Val), la proliferación y la diferenciación endometrial (LIFR p.Arg780Cys), y la angiogénesis (EPAS1-p.Tyr488Cys y p.Leu504Val). La mutación EPAS1-p.Tyr488Cys se encontró en dos pacientes. Por esta razón, y debido a que es un factor de transcripción relevante en la implantación, se escogieron las mutaciones

en este gen para realizar un tamizaje en una cohorte de pacientes francesas con AER idiopático y para su posterior estudio funcional.

2.4.3 Análisis in silico del potencial efecto deletéreo de las mutaciones candidatas de AER

Se realizó un análisis de FMO con las proteínas que incluyeran dentro de su estructura la región donde se encontraban las mutaciones de interés. Las proteínas que cumplieron este criterio de selección fueron FGA y MMP-10. FMO es un método computacional químico-cuántico que analiza sistemas moleculares complejos incluyendo el ambiente atómico que lo rodea. El FMO se fundamenta en fragmentar las estructuras proteicas para calcular la descomposición de las interacciones energéticas en términos de electrostática, repulsión, transferencias de carga y dispersión [120]. Otras metodologías similares como la aproximación campo mutualmente consistente (CMF) y el método de correlación incremental permiten calcular la energía de interacción electrostática y de polarización, y la correlación energética entre moléculas, respectivamente [158,159]. La teoría explicita de polarización (X-Pol) es otra aproximación que se fundamenta en la estructura electrónica de macromoléculas en la que un sistema condensado es fragmentado en subsistemas individuales para determinar sus diferencias energéticas [160]. La diferencia entre estos métodos y el FMO es la aproximación que se utiliza para estimar las interacciones entre los fragmentos y que la FMO permite calcular diferentes tipos de interacciones energéticas entre éstos. La aproximación de FMO es más robusta en términos computacionales, ya que permite descomponer el ambiente atómico que rodea el residuo de interés en términos de electrostática, repulsión, transferencias de carga y dispersión. Además, nuestro abordaje es innovador pues es la primera vez que se ha usado una técnica química-cuántica para simular las interacciones energéticas asociadas a mutaciones presentes en enfermedades de la reproducción humana [110,116].

FGA es una glicoproteína que forma la subunidad alfa del fibrinógeno y es importante en el proceso de coagulación como la fibrinólisis, la agregación de

plaquetas y la interacción con la trombina. El análisis de FMO demostró que el cambio en FGA p.Phe685Cys produce inestabilidad por el aumento de la interacción energética de la región cerca al aminoácido de interés secundaria a la pérdida de 3 puentes de hidrógeno no clásicos (**Figura 4**). Esto sugiere que se puede producir una alteración en la estructura de la proteína que pueda conducir a alteraciones funcionales. MMP-10 es una proteína involucrada en la degradación de la matriz extracelular durante la invasión del sincitiotrofoblasto. El análisis de FMO demostró que la mutación MMP-10 p.Asp199Asn producía un aumento significativo en las energías totales de interacción secundarias a la pérdida de las interacciones con un puente de sal (Lys165) y el bloque de Zn^{2+} (**Figura 5**). Este bloque es funcional para la regulación de MMP-10, ya que es el sitio catalítico donde se unen los TIMPs para inhibir su función [161]. Este resultado sugiere que la mutación podría producir un efecto patogénico por inestabilidad de la proteína en las pacientes con AER.

2.4.4 Tamizaje de la mutación THBD p.Trp153Gly en la población colombiana afectada por AER

En este capítulo del trabajo de tesis también se realizó el tamizaje de la mutación THBD p.Trp153Gly y los ensayos *in vitro*. Es importante recordar que ésta es una glicoproteína que se localiza en la membrana celular y es expresada por el sincitiotrofoblasto durante la decidualización. Su función biológica es participar en el proceso de coagulación, activación del complemento, migración, adhesión celular, e inflamación [88,89,162–167]. THBD está compuesta por 5 dominios: el CTLD, el de tipo EGF, un dominio rico en serina/treonina, un dominio transmembrana y un dominio citoplasmático. El dominio tipo EGF es esencial para la unión con la trombina y para la activación de la proteína C sirviendo como intermediario en el proceso de anticoagulación (inactivación de los factores Va y VIIIa) y en la regulación de la fibrinogénesis [168,169]. Modelos murinos *Thbd*$^{-/-}$ demostraron letalidad embrionaria debido a un aumento en la actividad procoagulante y la generación de trombina. Esto generaba alteraciones en el desarrollo placentario [91,170]. Estudios han demostrado que las pacientes con AER presentan una disminución en la

expresión de THBD en el tejido de placenta generando un estado de hipercoagulabilidad [168,171].

La mutación THBD-p.Trp153Gly se identificó en el presente trabajo por medio de NGS y en el tamizaje de pacientes colombianas con AER, mostrando diferencias estadísticamente significativas en las frecuencias alélicas entre los casos y los controles (**Tabla 9**). Estos resultados sugieren que esta variante es específica de la población colombiana y podría ser causal del AER en mujeres con un fenotipo temprano (<10 semanas de gestación) [172]. Esta mutación afecta un residuo evolutivamente conservado. Los programas de predicción de patogenicidad identificaron que el cambio de un triptófano por una glicina generaba un efecto potencialmente deletéreo en la proteína. Esto podría considerarse en diferencias fisicoquímicas de los residuos. La mutación se encuentra localizada en el dominio CTLD, el cual se ha relacionado con procesos anti-inflamatorios, de regulación del complemento, y la migración y la adhesión celular [88,89,166,167,173].

2.4.5 Estudio de la localización subcelular de THBD WT y mutante

Los resultados de la localización subcelular usando el gen reportero GFP y la IFI, demostraron diferencias estadísticamente significativas con la presencia de agregados-like en la región perinuclear entre la versión WT y mutante (68,5% vs. 78%, p=0,041) y para el perfil de agregados-like perinucleares masivo (+) y (-) (21% vs. 38%, p= 0,0019) (**Tabla 10**) en las líneas celulares CHO y COS-7. En las células THP-1, se demostró diferencias estadísticamente significativas con la presencia de agregados-like vs. la ausencia de éstos entre la versión WT (26% y 74%, respectivamente) y mutante (63% y 37%, respectivamente) (p=1,05 e^{-13}) (**Tabla 11**). Estas líneas expresan endógenamente THBD, por lo que deberían generar las modificaciones post-transduccionales requeridas para el buen funcionamiento de la proteína sobre-expresada tanto WT como mutante. Los agregados de la proteína THBD-p.Trp153Gly en la región perinuclear y citoplasmática de estas líneas celulares podría sugerir que hay un plegamiento anormal de ésta, alterando su transporte hasta la membrana celular. En el dominio CTLD se encuentran algunos

sitios de N- y O-glicosilación y, por programas bioinformáticos, la posición 153 de THBD es un sitio de predicción de unión a carbohidratos. Esto podría sugerir que la mutación puede cambiar la afinidad a los carbohidratos que se encuentran en el retículo endoplasmático, ocasionando que no pasen los puntos de control de glicosilación por alteraciones en el plegamiento correcto de THBD, quedando secuestrada en el retículo y/o en el citoplasma [174,175].

2.4.6 Estudio del efecto de la mutación THBD-p.Trp153Gly en la expresión de genes de la cascada corriente abajo de la citoquina HMGB1

En el presente trabajo también se determinó el posible efecto funcional que generaba la mutación THBD-p.Trp153Gly en la actividad anti-inflamatoria del CTLD [88,89,165]. Li *et al.*, demostraron que, en el modelo murino con la región CTLD ausente de Thbd, se generaba aumento de la respuesta inflamatoria dada por el incremento en la expresión de citoquinas [88]. La actividad anti-inflamatoria de este dominio se encuentra relacionada con el secuestro de la citoquina HMGB1 (high mobility group box 1), que es secretada al espacio extracelular bajo condiciones de necrosis y en respuesta a un ambiente celular pro-inflamatorio [88]. El CTLD se une covalentemente a HMGB1, evitando la unión de esta citoquina al receptor pro-inflamatorio RAGE. El clivaje de HMGB1 se genera por la formación del complejo THBD/Trombina [176,177].

Se realizó qPCR para determinar si la mutación THBD-p.Trp153Gly podía generar la expresión diferencial de los genes de la cascada corriente debajo de HMGB1 aumentando la respuesta inflamatoria, que conlleve al desarrollo de AER. Los resultados mostraron un aumento estadísticamente significativo en la expresión de los genes *GM-CSF, IL-10, NOS2, FLT1, IL-8* y *CCL2,* al igual que una disminución estadísticamente significativa en la expresión de *IKBKB* (**Figura 12**). Se ha evidenciado que la estimulación con HMGB1 produce el aumento en la expresión de *GM-CSF* e *IL-10*, sugiriendo que esta citoquina podría actuar paracrinamente como factor de estimulación del crecimiento y como promotor la liberación de otras citoquinas [178,179].

GM-CSF induce el crecimiento y la diferenciación de los monocitos, contribuyendo a la estimulación de la respuesta inflamatoria. En el caso de IL-10, que actúa como una citoquina anti-inflamatoria, se ha evidenciado que su aumento durante la respuesta inflamatoria contribuye a la supresión de la producción de citoquinas pro-inflamatorias, generando un balance de esta respuesta [178,180]. Estudios también han demostrado que HMGB1 participa en proceso de estrés oxidativo por el aumento de las especies reactivas de oxígeno (ROS) y del óxido nítrico (NOS2) [177,181,182]. En células trofoblásticas y endoteliales de cordón umbilical, se ha evidenciado que HMGB1 actúa como mediador de la inflamación aumentando el estrés oxidativo por la vía RAGE/NF-κβ [181].

En los trofoblastos, se ha observado la sobre-expresión de *FLT1* por medio de la cascada de RAGE durante el primer trimestre de gestación, produciendo disfunción endotelial debido a la disminución en la expresión de *VEG*F, lo que promueve el desarrollo de enfermedades de la reproducción como la preeclampsia [183]. Por otra parte, IL-8 es una quemoquina que actúa durante los procesos inflamatorios para atraer neutrófilos. Se ha evidenciado que HMGB1 genera la sobre-expresión de ésta en enfermedades pro-inflamatorias como la osteoartritis y bajo condiciones de hipoxia [182,184]. Los niveles de esta citoquina junto con CCL2 se encuentran aumentadas en pacientes con preeclampsia, quienes presentan una respuesta inflamatoria aumentada por acumulación de monocitos y macrófagos alrededor de los trofoblastos [181]. CCL2 es una quemoquina que se produce vía HMGB1/RAGE y que actúa en la quimiotaxis y migración de los monocitos y macrófagos [185]. La producción de la respuesta pro-inflamatoria está regulada por la vía de señalización NF-κβ, la cual es estimulada por la unión de HMGB1 al receptor de RAGE [177]. La proteína IKBKB se encarga de fosforilar el inhibidor de NF-κβ (Iκβ), para permitir la translocación de este factor de transcripción al núcleo y transactivar genes involucrados en la respuesta inflamatoria [186]. Nuestros resultados demostraron una disminución en la expresión de *IKBKB* con la estimulación de HMGB1 en la versión mutante de THBD, lo cual sugiere que las células podrían estar regulando el ambiente inflamatorio, controlando la expresión de genes pro-inflamatorios.

Con estos resultados, concluimos que la mutación THBD-p.Trp153Gly genera el aumento en la expresión de genes asociados a la cascada pro-inflamatoria de HMGB1, propiciando un ambiente pro-inflamatorio en el endometrio, lo que podría contribuir al desarrollo del AER.

2.4.7 Modelamiento in silico del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD.

Para el modelamiento *in silico* del efecto de la sustitución p.Trp153Gly en el dominio lectina de la proteína THBD se realizó un BLASTp con el que se identificó la estructura cristalográfica del receptor de manosa CTLD3 humano (PDB ID: 5XTS), con un 16% de identidad (**Figura 13**). A pesar de la baja identidad, se preservan los residuos aromáticos que están relacionados con la conservación de la estructura terciaria en proteínas, bien sea por interacciones aromáticas o por reducción en la hidratación del esqueleto de la proteína [187,188]. El triptófano que sufre la sustitución por glicina en la posición 153 es conservado entre ambas proteínas. De igual forma, se observa la identidad de cisteínas en las posiciones 35, 52, 137, 158 y 167, que están involucradas en la generación de puentes disulfuro, importantes en el plegamiento. Los resultados obtenidos con el modelamiento *in silico* mostraron que, a pesar de que la mutación no ocurre en la superficie del dominio modelado, tiene un claro impacto tanto en la estabilidad de la proteína como en su potencial capacidad de interacción, ya que implica cambios en el volumen, el área superficial, los paisajes electrostáticos y energéticos. El modelo con la sustitución THBD p.Trp153Gly presentó un menor volumen, una mayor área superficial y un cambio energético estabilizante, comparado con el modelo THBD WT (**Figuras 16** y **17)**, implicando posibles diferencias en la capacidad de interacción con ligandos potenciales como HMGB1[189]. Es importante mencionar que esta citoquina es secretada al espacio extracelular por los monocitos, macrófagos, uNKs y células dendríticas. Durante la fase secretora del endometrio y en la gestación temprana, los macrófagos y los monocitos representan el 20% de los leucocitos que se encuentran en el endometrio. Estas células son estimuladas específicamente por HMGB1 [190]. Bhutada *et al.,* demostraron que HMGB1 es expresada por las células

del epitelio endometrial y que, a su vez, es secretada en la cavidad uterina [191]. Los niveles de esta citoquina disminuye durante la fase receptiva del endometrio que se caracteriza por un aumento en la expresión de THBD, lo que podría sugerir que los bajos niveles de HMGB1 se deben al secuestro por THBD para permitir la adecuada implantación del blastocisto [192]. La propagación de la respuesta pro-inflamatoria de HMGB1 se da por unión con los receptores RAGE y a los receptores toll like (TLR2 y TLR4). Esto produce la activación de la vía NF-kβ y la expresión de TNF-α, IL-6 e IFN-ɣ, las cuales están aumentada en pacientes con AER, sugiriendo que altos niveles de HMGB1 podrían estar asociados con el ambiente pro-inflamatorio de la cavidad uterina de estas pacientes [181,193,194].

Estos resultados sugieren que la mutación THBD-p.Trp153Gly participa en la generación de un ambiente pro-inflamatorio asociado a disfunciones en la implantación y AER. Esta mutación podría alterar la interacción con la citoquina HMGB1, la cual no se hidrolizaría de manera adecuada, promoviendo su unión con receptores de la respuesta inflamatoria como RAGE, TLR2 y TLR4.

2.4.8 Tamizaje de las mutaciones de EPAS1 y estudio funcional in vitro

Otras de las mutaciones escogidas para realizar ensayos funcionales fueron EPAS1-p.Leu504Val y EPAS1-p.Tyr488Cys. EPAS1 es un factor de transcripción que presenta una alta similitud con el factor inducible de hipoxia 1a (HIF-1a). Se encuentra involucrado en la adaptación al ambiente hipóxico por la inducción de genes como la eritropoyetina (*EPO*) y *VEGF*, entre otros. Este factor de transcripción actúa por heterodimerización con ARNT y su interacción con los sitios de unión a HIF-1. A diferencia de HIF-1a su expresión es específica en ciertos tejidos como la placenta y el sincitiotrofoblasto, los cuales están sometidos a condiciones de hipoxia durante la gestación [195,196]. Las mutaciones encontradas en el presente trabajo se encuentran localizadas entre el dominio de degradación dependiente de oxígeno (ODDD) y el dominio de N-transactivación (**Figura 18**). Estos dominios están sujetos a hidroxilación durante condiciones de hipoxia para regular la estabilidad y la activación transcripcional de EPAS1.

En el presente trabajo, realizamos un tamizaje en pacientes francesas con AER para identificar posibles mutaciones en los exones que codifican para estos dominios (exones 10, 11 y 12).

A las variantes encontradas en el tamizaje (p.Ser467Asn y p.Thr597Ile) no se les realizó ensayos funcionales. Sin embargo, *Maemura et al.,* realizaron deleciones seriadas de estos dominios y demostraron que la transactivación del promotor de *VEGF* se inhibe significativamente en comparación con la versión completa de EPAS1 [197]. En la presente tesis se efectuó un ensayo de estandarización gen reportero con la versión WT y las mutaciones de EPAS1 identificadas por NGS (p.Tyr488Cys y p.Leu504Val) y con uno de los promotores de los genes blanco de EPAS1 (*INHA*). Se escogió el promotor de *INHA* porque su expresión aumenta en el sincitiotrofoblasto sometido a condiciones de hipoxia. *Depoix et al.,* demostraron que EPAS1 regula la expresión de *INHA* bajo condiciones de hipoxia durante la diferenciación de los citotrofoblastos a sincitiotrofoblastos en el primer trimestre de gestación [198]. Nuestros experimentos no mostraron diferencias estadísticamente significativas entre la versión WT y las mutaciones (**Figura 19**).

En síntesis, los resultados presentados en este primer capítulo generan información relevante para la comprensión de las posibles alteraciones moleculares asociadas a la patogénesis del AER. Los estudios realizados de THBD demostraron por primera vez que mutaciones en el dominio CLTD pueden estar asociadas al AER al generar un ambiente pro-inflamatorio en el endometrio por alteración en su interacción con HMGB1. La mutación de THBD-p.Trp153Gly encontrada mediante el análisis bioinformático innovador del exoma completo y tamizada en pacientes colombianas no relacionadas con AER idiopático podrían servir como biomarcador para el diagnóstico y pronóstico de la enfermedad.

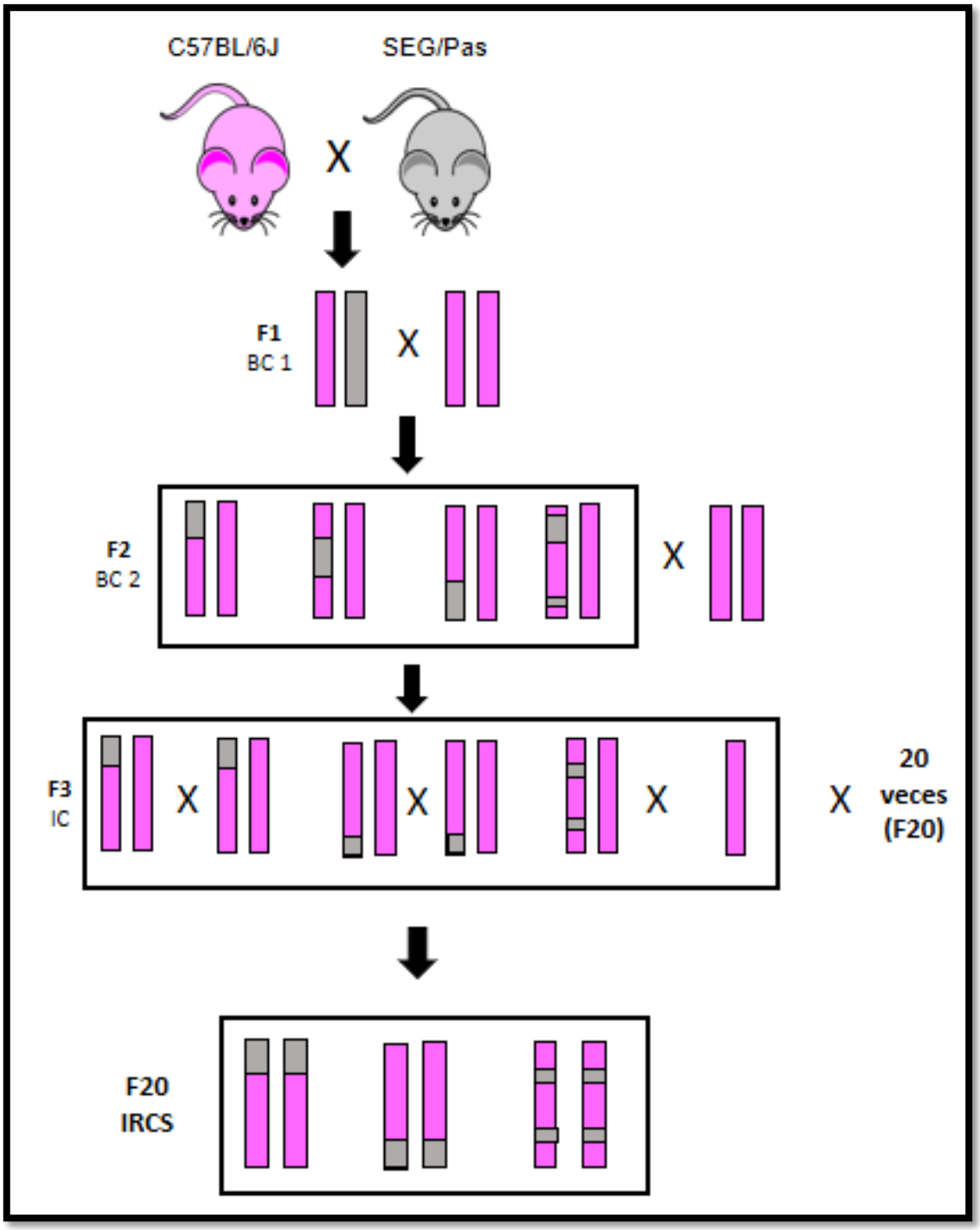
3. Capítulo II: *Estudios moleculares y funcionales del gen FOXD1*

3.1 Introducción

El uso del modelo murino *"interspecific recombinant congenic strain"* (IRCS) ha permitido el estudio de QTL asociados a la fertilidad femenina y masculina [199–201]. Con el modelo IRCS se han podido determinar algunos QTL responsables de la reabsorción embrionaria, un fenotipo análogo al AER en los humanos [43,201–203]. Este modelo ha sido usado eficientemente para estudiar fenotipos complejos pues una de sus ventajas es que permite medir rasgos de manera indefinida a lo largo del tiempo pues los fragmentos de estudio han sido fijados en el genoma de las líneas consanguíneas. Además, presenta frecuentemente diferencias fenotípicas estadísticamente significativas entre las líneas de estudio y los individuos parentales (lo que se traduce en altos valores de LOD score) y porque los fragmentos mapeados son relativamente cortos. Una desventaja de los IRCS es que son de difícil reproducción pues se cruzan dos especies de ratones evolutivamente distantes. Además, se facilita una selección negativa debido a un efecto de shock transcriptómico [200,204,205].

El modelo murino congénico IRCS se realizó a partir del cruce inicial entre dos subespecies de ratones (*M. musculus* y el *M. spretus*) separadas por aproximadamente 2 millones de años de evolución lo que representa diferencias fenotípicas y genotípicas muy relevantes entre las líneas parentales. El cruce inicial ("*outcross")* se realiza entre una hembra "*inbred"* de *M. musculus* (C57BL/6J) y un macho "*inbred"* de *M. spretus* (SEG/Pas). A partir del resultado del cruce inicial (F1), realizaron dos cruces consecutivos "*backcross"* entre las hembras obtenidas en F1 y F2 con un macho C57BL/6J de la línea parental para disminuir el genoma del *M. spretus* en el fondo genético del C57BL/6J hasta en un 12.5%. Después se realizaron 20 cruces adicionales entre hermanos/hermanas que permitieron fijar en estado homocigoto los fragmentos *M. spretus* en el fondo genético de *M. musculus* **(Figura 20**) [206]. Es importante mencionar que la línea de *spretus* contiene aproximadamente un SNP cada 80 – 100 pares de bases (comparado con el *M. musculus*) lo que permitió la utilización de mapeo de alta

Figura 20. Modelo murino IRCS. Se realizó un cruce inicial entre una hembra C57BL/6J y un macho SEG/Pas (*Spretus*). A partir de la F1, se realizaron dos cruces *backcross (BC)* entre las hembras F1 con un macho parental C57BL/6J y una hembra F2 con un macho parental C57BL/6J. Después se realizaron cruces entre hermanos/hermanas (F3 a F20) para fijar en estado homocigoto los fragmentos *M. spretus* en el fondo genético de *M. musculus.*



densidad para localizar con la precisión los fragmentos de *M. spretus* en el fondo genético de *M. musculus* [205]. Además, determinan en parte las diferencias fenotípicas interespecíficas resultantes de la divergencia entre las regiones codificantes (especialmente de los factores de transcripción) y de los promotores durante el proceso de especiación. Este fenómeno fue denominado en este modelo "shock transcriptómico" [204].

Laissue et al. (2009), por medio de ecografía de alta frecuencia en hembras IRCS gestantes (E8.5-E12.5) demostraron que algunas líneas de ratones con tres fragmentos de *spretus* presentaban un aumento en la reabsorción embrionaria en comparación con las líneas de ratones C57BL/6J x C57BL/6J (grupo control) [201]. En este estudio se analizaron 207 hembras gestantes de 39 líneas IRCS y 22 líneas del grupo control. De las 39 líneas IRCS estudiadas se identificaron 4 que mostraban un aumento en la reabsorción embrionaria (denominadas 66H, 103C, 135B y 135E) en comparación con el grupo control. La línea 66H presentaba tres fragmentos genómicos de *spretus* localizados en los cromosomas MMU1, MMU13 y MMU18. Para determinar con precisión la contribución del genotipo al fenotipo, se separaron los fragmentos genómicos de *spretus* y se generaron cinco sublíneas de las cuales 3 eran co-isogénicas (66HMMU1, 66HMMU13 y 66HMMU18). Este modelo se realizó mediante cruces entre hembras 66H con machos C57BL/6J para obtener una F2 con un único fragmento de *spretus* (co-isogénico) los cuales se fijaron en estado homocigoto en el fondo genético de *M. musculus* por medio de un intercross y genotipificaciones seriadas [201]. A las hembras gestantes de cada sublínea se les realizó nuevamente ultrasonografías las cuales evidenciaron que la mayor tasa de reabsorción embrionaria, en comparación con el grupo control, se presentaba en las sublíneas 66HMMU1 (24,6% p=0,013) y 66HMMU13 (14,7% p=0.01) [201].

En un estudio reciente de *Laissue et al.* (2016), que representa la continuidad del abordaje previamente descrito, se exploró de manera específica la sublínea IRCS 66HMMU13 la cual, por ultrasonografía, presentaba una tasa alta de reabsorción embrionaria [43,201]. Esta sublínea presentaba un fragmento único de *spretus* de

aproximadamente 2,6 Mb localizado entre los marcadores rs120693734 y D13Mit47 en el cromosoma MMU13. Este fragmento contenía 31 genes dentro de los cuales se encontraban genes codificantes para factores de transcripción [201]. Se realizaron experimentos de transcriptómica por microarreglos en 8 hembras de la sublínea 66HMMU13 cruzadas con machos C57BL6/J. Se extrajo el ARN total de la placenta y el endometrio de hembras gestantes de la sublínea 66HMMU13 (hembras 66HMMU13 x machos C57BL/6J) y de un grupo control C57BL/6J (C57BL/6J x C57BL/6J). En los experimentos de transcriptómica se compararon los niveles de expresión, por microarreglos, del endometrio y de la placenta entre la sublínea 66HMMU13 y el grupo C57BL/6J. Se demostró que en el endometrio de la sublínea 66HMMU13, 6,8% de los genes se encontraban sobreexpresados, mientras que el 4,3% se encontraban subexpresados en comparación con el grupo control. En la placenta de ratones 66HMMU13, 3,6% de los genes estaban sobreexpresados y 7,2% subexpresados en comparación con el grupo control. Luego, se seleccionaron los 50 genes mayormente sobreexpresados y los 40 genes más subexpresados. Mediante un análisis bioinformático de los sitios potenciales de unión a factores de transcripción (TFBS) usando Genomatix y un análisis estadístico se identificaron, en los promotores mayormente desregulados, una sobre-representación de TFBS de algunas familias de factores de transcripción. La familia forkhead box (FKHD) de factores de transcripción se encontró en el primer lugar de significancia estadística (p=0,0009). El fragmento de *spretus* en el cromosoma MMU13 de la sublínea 66HMMU13 contenía un único factor de transcripción perteneciente a la familia de los *forkhead* (*Foxd1*) [43]. Esto lo convirtió en un gen candidato en la etiología de la reabsorción embrionaria pues variantes en su secuencia, originadas en la divergencia evolutiva entre *M. musculus* y *M. spretus*, podrían estar relacionados con modificaciones en la regulación de genes blanco.

La secuenciación directa de la región codificante de *Foxd1* de *Mus spretus* y de la sublínea 66HMMU13 mostró 5 variantes no-sinónimas (p.Asp73Glu, p.Asn126Glu, p.Thr152Ala, p.Asp76_Leu77InsAsp y p.Pro319del) en comparación con el grupo control. La variante p.Thr152Ala se encontró en el sitio de unión del ADN de este factor de transcripción en una posición conservada evolutivamente. Los análisis de

la potencial patogenicidad, usando SIFT, mostraron que este cambio tenía un efecto potencialmente deletéreo. Para determinar el efecto funcional de la variante se realizó un ensayo gen reportero con los promotores de genes blanco directos para Foxd1 que se encontraron desregulados en el análisis transcriptómico (*Pgf* y *C3*) y que representaban moléculas importantes en el mantenimiento de la gestación en el ratón. Se determinó que la variante p.Thr152Ala disminuía la habilidad de *Foxd1* para transactivar el promotor de *Pgf* pero no afectaba la transactivación del promotor de *C3* [43]. Con estos resultados se determinó que la versión de *spretus,* representada principalmente por p.Thr152Ala, tenía un efecto funcional y podría ser la responsable de las camadas pequeñas de *spretus* observadas en la naturaleza.

En el mismo estudio se sugirió que las mutaciones en *FOXD1* podían estar implicadas en el AER. Por esta razón, se secuenció la región codificante de *FOXD1* en 556 pacientes y en 271 controles [43]. Se encontraron 10 variantes no-sinónimas en estado heterocigoto en las pacientes con AER (5,9%, *p*=0.00056). El riesgo relativo de tener AER al presentar variantes de *FOXD1* en estas pacientes fue de 10,3 (IC: 1,4–77,2). La variante que mostró mayor significancia estadística fue la inserción de dos residuos de alanina en la posición 429 de *FOXD1* (p.429AlaAla) (*p*=0,015). Los ensayos funcionales con esta variante y la variante p.Ala356Gly demostraron la incapacidad de FOXD1 mutante para transactivar el promotor de *PGF*, mientras que la transactivación del promotor de *C3* no mostró cambios con la variante p.429AlaAla, pero si una disminución significativa con la variante p.Ala356Gly [43]. La variante p.Ile364Met retuvo su capacidad de transactivar el promotor de *PGF* como la versión WT, pero aumentaba significativamente la transactivación del promotor *C3*. Con estos hallazgos se pudo concluir que *FOXD1* no solo cumple una función importante en la implantación de los ratones, sino que las mutaciones de *FOXD1* son una causa frecuente de AER.

En este contexto, es posible que las mutaciones de este gen estén asociadas a la falla de implantación recurrente (FIR) y a patologías de la placenta (el estudio de esta hipótesis se desarrolló en el presente trabajo de tesis).

En síntesis, el modelo IRCS permitió efectuar el clonaje posicional de *FOXD1*. Los resultados se aplicaron de una especie *inbred* a una *outbred* (humanos) con éxito. Estos aspectos hacen parte del review publicado en la revista **Journal of Molecular Medicine** (IF: 4.9 – 2016) bajo el título ***"The multisystemic functions of FOXD1 in development and disease"*** (Ver **Anexo IV**) [207]. Parte de este trabajo de tesis se focalizó en el estudio de la función de FOXD1 en la biología de la implantación y de la placentación.

3.2 Metodología

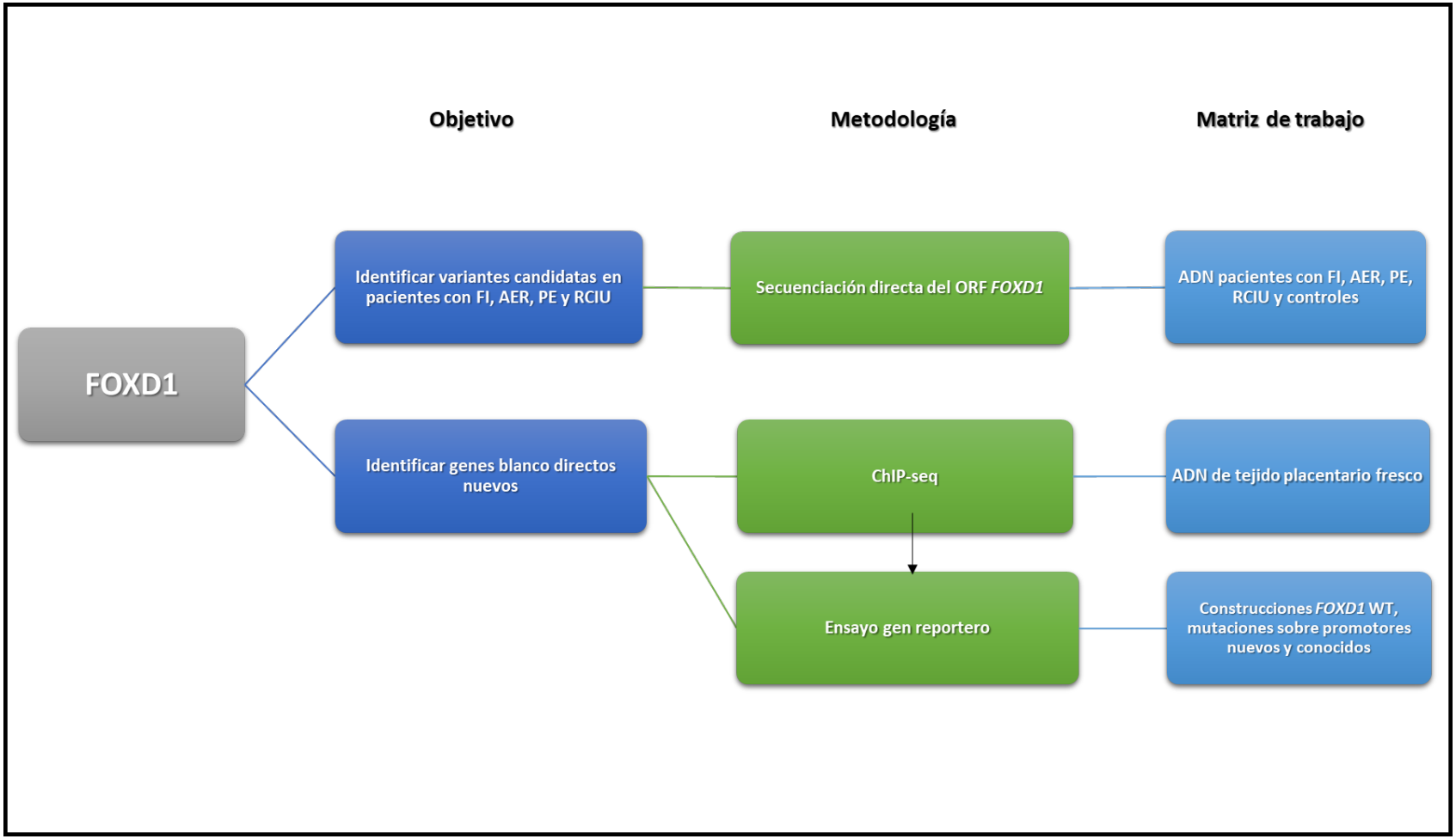
En este capítulo del trabajo de tesis se realizó un abordaje experimental que permitió responder a preguntas de investigación referentes a la función de FOXD1 (**Figura 21**):

1- ¿Cómo mutaciones nuevas de este factor de transcripción podrían modificar la transactivación de genes blanco directos conocidos y cuál sería su potencial relación funcional con la FIR, el IUGR y la PE ?

2- ¿Qué nuevos genes son blanco directos de FOXD1 en un contexto placentario?

Para responder a la pregunta 1 se desarrollaron los pasos metodológicos descritos en los puntos *3.2.1, 3.2.2* y *3.2.7*. Para responder a la pregunta 2 se desarrollaron los pasos metodológicos descritos en los puntos *3.2.3, 3.2.4, 3.2.5, 3.2.6* y *3.2.7*.

Figura 21. Metodología para el estudio molecular y funcional de FOXD1.



3.2.1 Inclusión de pacientes y de controles con FIR, AER, PE, Y RCIU

Para determinar si las pacientes con AER idiopático colombianas y con FIR presentaban variantes en el gen *FOXD1* que pudieran ser usadas como marcadores moleculares, se incluyeron 31 pacientes colombianas con AER y 30 pacientes con FIR. Las pacientes con AER fueron remitidas de los centros de reproducción asociados a la Universidad del Rosario, las consultas privadas de gineco-obstetras de Cecolfes y de los genetistas de Mederi y la Universidad del Rosario. Las pacientes con FIR fueron remitidas de centros de asistencia a la reproducción (e.g. Cecolfes) quienes presentaron dos o más fallas recurrentes de la implantación en al menos 2 ciclos consecutivos de FIV en los que un único embrión de alta calidad fue transferido en cada ciclo. A estas pacientes se les efectuó seguimiento de los niveles séricos de β-HCG para determinar si había o no implantación. Los criterios de exclusión para las pacientes con FIR fueron: pacientes con síndrome antifosfolípido, alteraciones en el cariotipo materno, coagulopatías, anormalidades anatómicas en el útero, miomatosis uterina, hidrosalpinx y alteraciones en los factores masculinos de las parejas. Los criterios de exclusión para las pacientes con AER y el grupo control fueron descritos en la metodología del Capítulo I del presente trabajo. El grupo control para FIR fue constituido por 100 pacientes del mismo origen étnico a quienes se les realizó un ciclo de FIV y tuvieron un embarazo exitoso. A las pacientes y a los controles se les explicó los objetivos del estudio y, previa firma del consentimiento informado, se les tomó una muestra de sangre periférica para realizar la extracción del ADN por medio del método de salting-out.

Las pacientes francesas con PE (n=31), con RCIU (n=39) y con PE/RCIU (n=27) fueron atendidas en el Instituto Cochin (Paris, France). Los criterios de inclusión para las pacientes con PE fueron: presión sistólica >140 mmHg, presión diastólica >90 mmHg y proteinuria >0,3 gr/día. Los criterios de inclusión para RCIU fueron reducción del crecimiento fetal durante la gestación con un peso al nacer por debajo del percentil 10. Los criterios de exclusión para ambos grupos fueron diabetes, malformaciones cromosómicas y fetales, infección materna y tratamiento con

aspirina. Los controles para este grupo de pacientes fueron los reportados en el estudio de Laissue et al, 2016 [43].

3.2.2 Secuenciación directa del gen FOXD1

La región codificante completa del gen *FOXD1* se amplificó en dos fragmentos [43]. Los amplicones fueron purificados mediante exonucleasa I y fosfatasa alcalina. Se realizó secuenciación de Sanger usando los primers diseñados en el interior del amplicón para cada fragmento. Se efectuó un análisis bioinformáticos (alineamientos múltiples interespecíficos, SIFT y Polyphen) y la búsqueda de las variantes en las bases de datos públicas de SNPs (NCBI, ExAC y Ensembl) para determinar su potencial patogenicidad. Las variantes candidatas fueron tamizadas en el grupo control descrito previamente. Las variantes nuevas que se encontraron en las pacientes y no en el grupo control fueron seleccionadas para su validación mediante estudios funcionales.

3.2.3 Preparación de la muestra de placenta para experimentos de ChIP-seq

Para determinar los genes blanco directos no conocidos de FOXD1 en la placenta humana se realizó ChIP-seq de las vellosidades coriónicas de tejido placentario sano a término. A partir de una placenta humana, previo lavado con solución salina, se disecaron las vellosidades placentarias evitando el corte de las membranas y los vasos maternos. Se tomaron muestras de aproximadamente 200 mg (1 cm), las cuales se almacenaron en nitrógeno líquido y en trizol. Antes del envío de la muestra para ChIP-seq (Epigentek Inc., USA) se realizó la extracción de ADN para la secuenciación directa de *FOXD1* (para determinar si la donante presentaba alguna variante en este gen). Además, se extrajo el ARN total para efectuar RT-PCR y determinar la expresión de *FOXD1* en el tejido placentario, al igual que la extracción de proteínas totales para confirmar la presencia de FOXD1 en la placenta por Western Blot (WB). Luego de la verificación de la expresión de FOXD1 en la placenta, una muestra en fresco de 200 mg fue enviada a Epigentek Inc. (USA) para realizar los experimentos de ChIP-seq.

3.2.4 ChIP-seq

Para determinar los blancos directos no conocidos de FOXD1 en el tejido placentario se efectuó la siguiente metodología en la plataforma externa Epigentek Inc. (NY, USA).

Se realizó la extracción de ADN de la biopsia de tejido placentario normal y se preparó según el protocolo de ChIP (inmunoprecipitación de la cromatina). Se aisló la cromatina de la muestra usando el kit ChromaFlash Chromatin Extraction (EpiGentek, Cat. #P-2001) y se realizó un cross-linking de FOXD1 al ADN con formaldehido. La cromatina se fragmentó entre 200 – 1000 pb usando el sistema EpiSonic 2000 Sonication (EpiGentek, Cat. #EQC-2000). La concentración de la cromatina fue medida por fluorescencia y 5 µl de la cromatina fueron usados para purificar el ADN que sirvió como “input”. Los tamaños de los fragmentos de ADN (100 – 600 pb) fueron determinados por el equipo Bioanalyzer Agilent. El anticuerpo anti-FOXD1 (invitrogen # PA5-35145) fue validado usando el kit Pre-Sure ChIP Antibody Validation (EpiGentek, Cat. #P-2031). El grado de intensidad para ChIP (CGI) del anticuerpo fue de 4,2 (referencia: CGI > 3). Para la reacción de ChIP se usaron 10 µg de cromatina con 5 µg del anticuerpo, 10 µl de perlas y 500 µl del buffer para ChIP. La reacción fue incubada a temperatura ambiente por 3,5 horas en rotación continua. Después de la incubación, las perlas fueron lavadas y el “ChIPed” ADN fue purificado y eluido en 20 µl de agua. Para la preparación de la librería se ligaron los adaptadores a los extremos de los fragmentos de ADN, se realizó una amplificación clonal usando primers universales y se purificó. La librería fue verificada usando el equipo Bioanalyzer y cuantificada por KAPA qPCR. 10nM de la muestra fueron secuenciados por NGS usando la plataforma Illumina HiSeq 4000.

3.2.5 Análisis bioinformático de la secuenciación y depuración de los datos

El análisis de los datos primarios de la ChIP-seq se realizó usando Bowtie2 versión 2.2.6 y MACS versión 2.1.0 como se ha reportado en la literatura [208–210]. El control de calidad de los datos generados por la secuenciación de Illumina fue realizado usando FastQC, version 0.10.1 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/). Estos datos fueron mapeados con la secuencia del genoma humano (hg19) usando Bowtie, fueron convertidos de archivos SAM a BAM y fueron ordenados por coordenadas usando SAMtools version 0.1.19 [211]. Las estadísticas del mapeo se realizaron usando Picard, version 1.90 (http://picard.sourceforge.net/). Estos resultados fueron sometidos al llamado de picos enriquecidos contra la muestra "input" usando los parámetros por defecto de MACS2 con un valor de corte para *p* de 0,1. Los picos encontrados fueron filtrados contra la lista de ENCODE para hg19 usando Bedtools, version v2.25.0 [212,213]. Este análisis se realizó en la plataforma externa Epigentek Inc. (NY, USA).

3.2.6 Análisis de los picos encontrados por ChIP-seq

Se filtraron los picos manualmente usando los valores de enriquecimiento ("fold_enrichment") y los valores de alineamiento ("pile_up"). Para determinar el punto de corte para el enriquecimiento, se realizó un conteo de los picos para cada valor y, así, realizar una gráfica de distribución para determinar el umbral de estos valores representativos para ~10,000 picos. Se realizó la misma metodología para determinar el umbral para los valores de alineamiento [208,214–216]. Los picos encontrados bajo estos umbrales fueron filtrados teniendo en cuenta las siguientes localizaciones: corriente arriba del gen ("Upstream"), dentro del gen ("Inside"), sobrelapado al inicio del gen ("OvelapStart") e incluyendo todo el gen ("IncludeFeature") [217]. Para determinar los picos cercanos a los sitios de inicio de la transcripción (TSS) de los genes blanco, se usó el valor de la distancia más corta con un valor de corte de ≤2,000 bp desde el TSS más cercano

("shortestDistance")[218]. Para correlacionar la localización del pico con un gen blanco, se filtró por la categoría RefSeq ("NM_id") y se visualizó usando la herramienta IGV. Se escogieron los genes blanco teniendo en cuenta la función biológica en la gestación y en la placenta usando las bases de datos de genes GeneCards (http://www.genecards.org/), UniProt (https://www.uniprot.org/), NCBI (https://www.ncbi.nlm.nih.gov/gene/) y Ensembl (http://www.ensembl.org/index.html).

Se utilizó la herramienta bioinformática Genomatix para determinar los sitios putativos de unión a FOXD1 en los promotores de estos genes blanco y se utilizó la información de los resultados generados por transcriptómica de los genes blanco diferencialmente expresados de Foxd1, obtenidos en el estudio de *Laissue et al.* [43], para escoger los genes blanco candidatos para los estudios funcionales posteriores.

3.2.7 Ensayos gen reportero

Se crearon construcciones plasmídicas (pcDNA 3.1 Zeo (+)) que contenían la región codificante de *FOXD1* WT y mutantes de acuerdo con las variantes encontradas en las pacientes con FIR, AER y RCIU (**Tabla 13**).

Tabla 13. Construcciones de las mutaciones FOXD1

Patología	Cambio proteína
AER y FIR	p.Gly103Asp
FIR	p.His267Tyr
RCIU	p.Arg57del
AER	p.Arg55Pro
	p.Ala356Gly
	p.Ile364Met

Se realizaron clonajes de las regiones promotoras blanco de FOXD1 en el plásmido pGL4.22[luc2CP/Puro], el cual incluye la secuencia del gen reportero de luciferasa (*luc2CP - Photinus pyralis*). En este plásmido se clonaron los promotores de los

genes blanco de interés identificados en el estudio previo de Laissue et al.: *PGF* y *C3* [43]. También se crearon las construcciones con los promotores que se encontraron por la técnica ChIP-seq descrita anteriormente (Ver resultados más adelante). Se usó un plásmido que contenía la secuencia codificante de *Renilla*, el cual se empleó como normalizador de la transfección. Se realizó una co-transfección en células COS-7 con las 3 construcciones: la que contenía el ORF WT o mutante de *FOXD1*, la que contenía la secuencia de alguno de los promotores blanco y la que codificaba para la *Renilla*. Los niveles de la transactivación/transrepresión se analizaron usando el sistema Dual-Luciferase Reporter Assay (Promega, Madison, WI, USA). La actividad de la luciferasa en cada replica fue dividida por la actividad reportada para la Renilla. Con los resultados se realizó un estudio estadístico usando la prueba t-student para determinar si existían diferencias significativas entre las condiciones WT y las mutantes. Se realizaron 5 réplicas biológicas con 6 réplicas técnicas cada una.

3.3 Resultados

3.3.1 Estudio de FOXD1 como un potencial marcador molecular de AER, en FI durante FIV (ART), RCIU y PE

El análisis de secuenciación identificó 8 variantes heterocigotas no-sinónimas (**Tabla 14**). Las variantes FOXD1 p.Arg57del y p.His267Tyr no se han reportado en las bases de datos públicas ni previamente, mientras que las variantes FOXD1 p.Ala356Gly y p.Ile364Met presentan un MAF bajo en las bases de datos y fueron reportadas previamente en pacientes con AER [43]. Ninguna de estas variantes fue encontrada en grupos control. La variante FOXD1 p.His267Tyr fue encontrada en pacientes colombianas con FIR, mientras que la variante FOXD1 p.Arg57del se encontró en una paciente francesa con RCIU.

Tabla 14. Resultados de la secuenciación del ORF de *FOXD1* en pacientes colombianas con AER, FIR, PE,RCIU e individuos control								
Pacientes con AER colombianas								
ADN	**Proteína**	**Rs**	**Pacientes**	**Controles**	**MAF**	**SIFT**	**PolyPhen**	**Conservación evolutiva**
c.263C>G	p.Ala88Gly	rs7705335	5/31	6/100	0,18	Tolerado (0,52)	Benigno (0)	Si
c.308G>A	p.Gly103Asp	rs1369199745	1/31	1/100	0,000008	Tolerado (0,46)	Benigno (0,334)	Si
Pacientes con FIR colombianas								
ADN	**Proteína**	**Rs**	**Pacientes**	**Controles**	**MAF**	**SIFT**	**PolyPhen**	**Conservación evolutiva**
c.263C>G	p.Ala88Gly	rs7705335	2/30	6/100	0,18	Tolerado (0,99)	Benigno (0)	Si
c.308G>A	p.Gly103Asp	rs1369199745	1/30	1/100	0,000008	Tolerado (0,46)	Benigno (0,334)	Si
c.799C>T	p.His267Tyr	___	1/30	0/100	__	Deletéreo (0,04)	Posiblemente deletéreo (0,69)	Si
c.976G>A	p.Ala326Thr	rs552595262	1/30	2/100	0,0142	Deletéreo (0,07)	Posiblemente deletéreo (0,69)	Si
Pacientes francesas con RCIU								
ADN	**Proteína**	**Rs**	**Pacientes**	**Controles**	**MAF**	**SIFT**	**PolyPhen**	**Conservación evolutiva**
c.168_170delGCG	**p.Arg57del**	**rs750578392**	**1/39**	**0/271**	**0.0065**	**__**	**__**	**Si**
c.263C>G	p.Ala88Gly	rs7705335	4/39	23/271	0,18	Tolerate (0,99)	Benign (0)	Si
c.976G>A	p.Ala326Thr	rs552595262	1/39	8/271	0.0142	Deleterious (0,07)	Posibly deleterious (0,69)	Si
c.1092C>G	p.Ile364Met	rs992724147	1/39	0/271	0.001	Deleterious (0)	Deleterious (0,99)	Si
c.1146_1158delCAGGCCGCCGCC	p.Gln383_Ala387delQAAA	rs771204220	4/39	13/271	0,006	__	__	NA
Pacientes francesas con PE								
ADN	**Proteína**	**Rs**	**Pacientes**	**Controles**	**MAF**	**SIFT**	**PolyPhen**	**Conservación evolutiva**
c.263C>G	p.Ala88Gly	rs7705335	6/31	23/271	0,18	Tolerate (0,99)	Benign (0)	Si
c.976G>A	p.Ala326Thr	rs552595262	2/31	8/271	0.0142	Deleterious (0,07)	Posibly deleterious (0,69)	Si
Pacientes francesas con PE+RCIU								
ADN	**Proteína**	**Rs**	**Pacientes**	**Controles**	**MAF**	**SIFT**	**PolyPhen**	**Conservación evolutiva**
c.263C>G	p.Ala88Gly	rs7705335	1/27	23/271	0,18	Tolerate (0,99)	Benign (0)	Si
c.1067C>G	p.Ala356Gly	rs917127030	1/27	0/271	0.0003	Deleterious (0,01)	Benign (0,001)	Si
c.1187C>T	p.Pro396Leu	rs540644822	1/27	1/271	0,0002	Tolerate (0,2)	Benign (0,1)	Si

En este grupo de pacientes una con RCIU/PE presentaba la variante FOXD1 p.Ala356Gly y una con RCIU presentaba la variante FOXD1 p.Ile364Met. El resto de las variantes se consideraron polimorfismos, ya que presentaban un MAF>1% en las bases de datos públicas y estaban presentes en la población control. Los análisis de la potencial patogenicidad, usando los predictores SIFT y PolyPhen, mostraron que la mutación FOXD1 p. His267Tyr tenía un efecto potencialmente deletéreo.

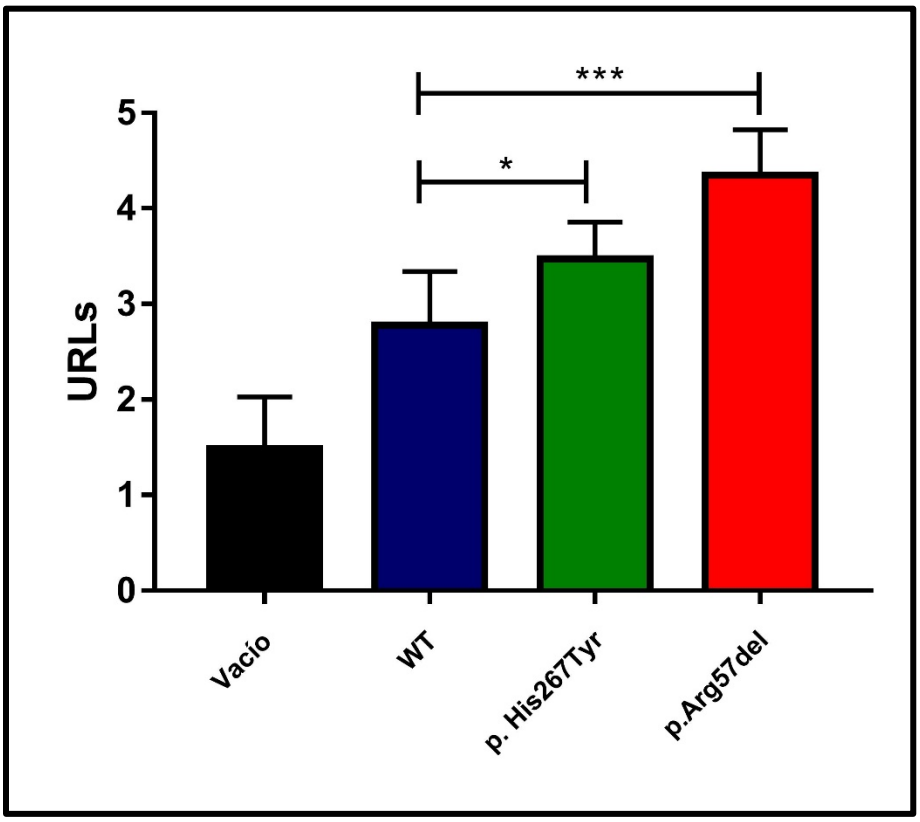
3.3.2 Ensayo gen reportero

Para estudiar el efecto funcional de algunas de las mutaciones previamente encontradas sobre las transactivación de promotores blanco ya conocidos de FOXD1 se realizaron ensayos gen reportero. Se usaron las construcciones plasmídicas FOXD1 WT (pcDNA3.1 Zeo (+)) y las que incluían las mutaciones encontradas en las pacientes con FIR y RCIU. Para el promotor del gen blanco C3, identificado en el estudio realizado por Laissue et al., se utilizó la construcción subclonadas en el plásmido pGL4.22 [43].

Los resultados del ensayo gen reportero demostraron que FOXD1 transactiva el promotor de *C3* cuando se compara con el vector vacío (1,9 veces, p=0,0024). La actividad transcripcional del promotor presentó un aumento estadísticamente significativo cuando se sobre-expresan las mutaciones de FOXD1 encontradas en las pacientes con FIR y RCIU en comparación con la versión WT. La mutación FOXD1 p.His267Tyr aumentó 1,25 veces (*p*=0,03) y FOXD1 p.Arg57del aumentó 1,5 veces (*p*=0,0004) (**Figura 22**).

Los resultados antes mencionados fueron sometidos en la revista **Human Mutation** (IF: 5,4 – 2017) bajo el título ***"FOXD1 mutations are related to repeated implantation failure, intra-uterine growth restriction and preeclampsia"*** (Ver **Anexo V**).

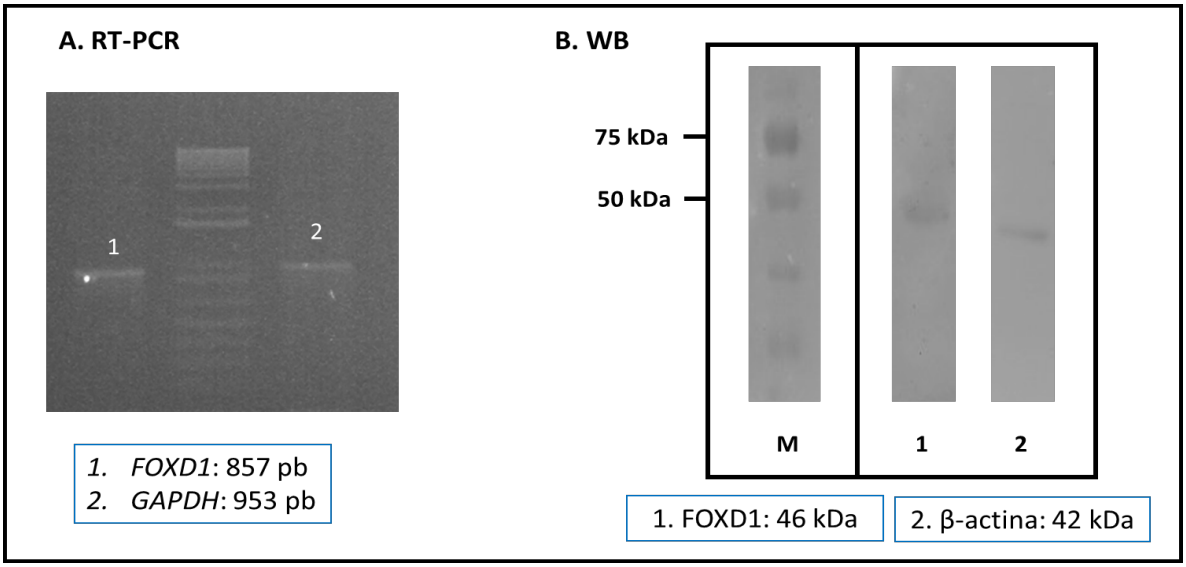
Figura 22. Ensayo gen reportero de FOXD1 WT y mutantes con el promotor de *C3.* * $p<0.05$, *** $p<0.001$



3.3.3 Identificación de nuevos genes blanco directos de FOXD1

Para determinar los genes blanco directos no conocidos de FOXD1 se realizó ChIP-seq de las vellosidades coriónicas de tejido placentario sano a término. Los análisis previos al envío mostraron que la paciente presentaba una variante sinónima frecuente en el ORF de FOXD1-p.Ser436Ser (rs2972191, MAF 0.3). Tanto la RT-PCR como el WB demostraron la expresión de FOXD1 en el tejido placentario (**Figura 23**).

Figura 23. Expresión de FOXD1 en tejido placentario sano a término. **A.** La RT-PCR muestra la presencia del transcrito en la placenta. Se usó el gen housekeeping *GAPDH*. **B.** El WB demuestra la expresión de FOXD1 en el extracto total de proteínas de placenta. Se utilizó la proteína housekeeping de β-actina. **M:** marcador de peso (Precision Plus Protein™ Dual Color Standards).



El control de calidad de los datos primarios de la ChIP-seq mostró una gran cantidad de valores con una distribución adecuada en el puntaje de calidad, sugiriendo que la secuenciación cumplió con los estándares técnicos. Se mapearon entre 79,72% y 95,32% de todos los datos.

Durante el llamado en la anotación de los picos se utilizaron los puntos de corte para el valor de *p* de 0,01 y 0,05 los cuales generaron 4 y 10 picos, respectivamente. Por esta razón, se usó el valor *p* de 0,1 para el análisis bioinformático subsiguiente. Debido a que el valor de *p* ajustado para cada pico no mostró diferencias estadísticamente significativas con el "input" se realizaron los filtros manuales usando los valores de enriquecimiento ("fold_enrichment") y de alineamiento ("pile_up"). La gráfica de distribución para el enriquecimiento mostró que ~20,000 picos se encontraban dentro del umbral de 1,27 (**Figura 24**). La mayoría de los picos mostraban valores de enriquecimiento bajos lo que sugiere ruido de fondo. Usando este filtro, se obtuvieron 19,744 picos de los 680,728 picos totales encontrados en el ChIP-seq. Se realizó el mismo análisis para determinar el valor de corte del alineamiento. El resultado de la distribución mostró que ~32,000 picos se

encontraban dentro del umbral de 0,41 (**Figura 25**). Se usó este umbral para filtrar los picos obtenidos con el umbral del enriquecimiento generando 8,878 picos.

Figura 24. Gráfica de distribución del conteo de los picos vs. valor de enriquecimiento. Se muestra que en el valor de enriquecimiento 1,27 (umbral) se encuentran ~20,000 picos que conforman la meseta de la distribución.

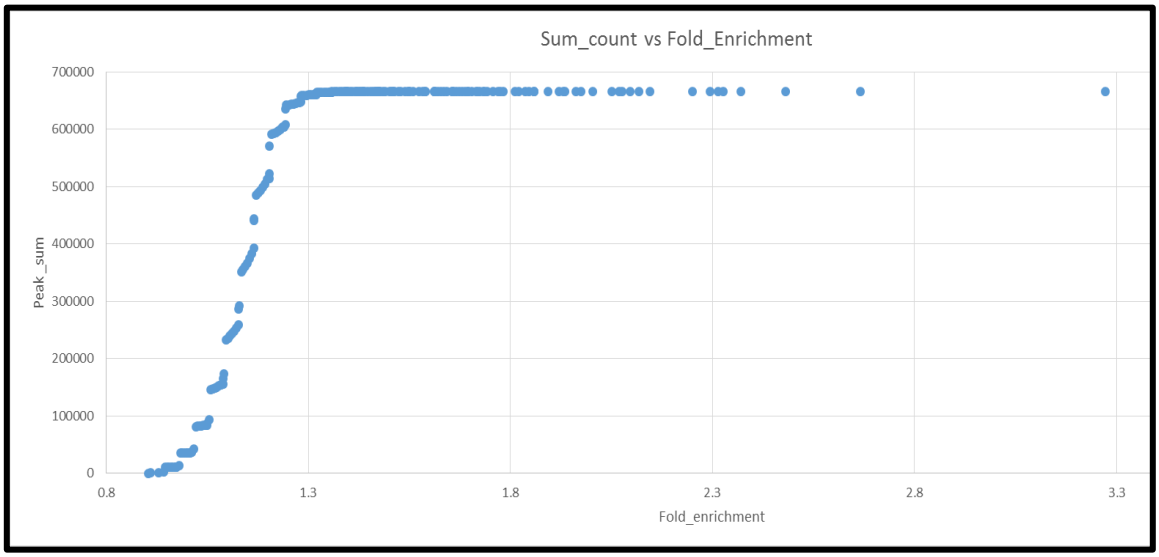
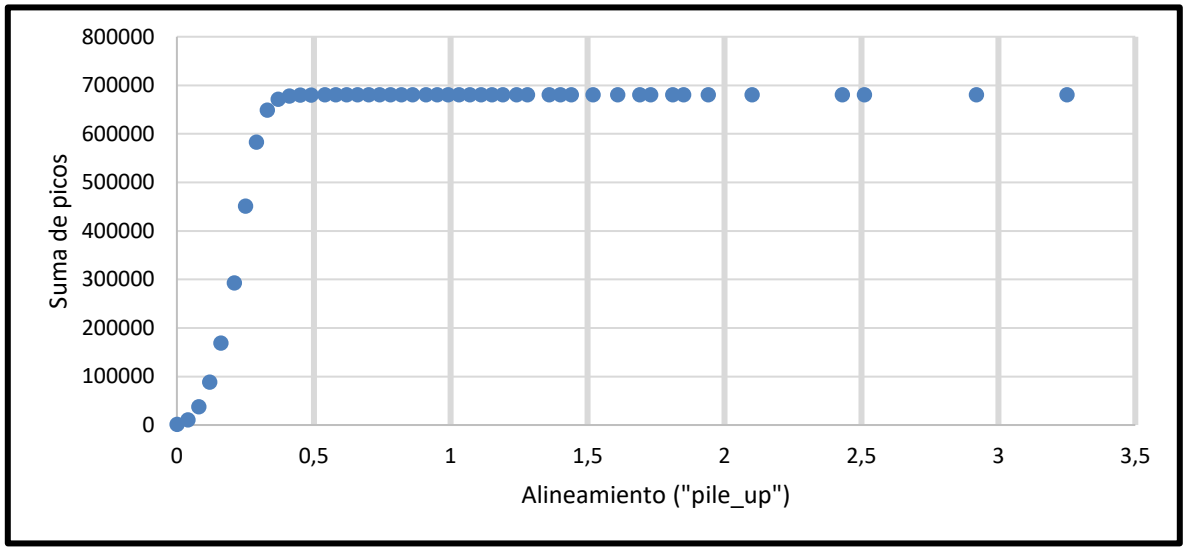


Figura 25. Gráfica de distribución del conteo de los picos vs. valor de alineamiento. Se muestra que en el valor de enriquecimiento 0,4 (umbral) se encuentran ~32,000 picos que conforman la meseta de la distribución.



Los 8,878 picos se filtraron usando las categorías de localización: corriente arriba del gen (“Upstream”), dentro del gen (“Inside”), sobrelapado el inicio del gen (“OvelapStart”) e incluyendo todo el gen (“IncludeFeature”). Se obtuvieron 5,903 picos de los cuales la mayoría se encontraban localizados corriente arriba del gen (69%) (**Figura 26**).

Para determinar los picos que se encontraban cerca a los TSS de los genes blanco se filtró por la distancia de ≤ 2,000 pb del TSS obteniendo 574 picos. Esta región fue escogida arbitrariamente asumiendo que contiene el promotor mínimo regulador de la transcripción. La correlación de estos picos con el gen blanco se realizó filtrando por la identificación de RefSeq con lo cual se obtuvieron 177 genes, el resto de los picos no tenían RefSeq asignado o pertenecían a RNAs no codificantes (**Figura 27**). 18,8% de los picos sin identificación fueron visualizados con la herramienta IGV obteniendo la identificación de su gen blanco.

Figura 26. Gráfica de localización de los picos.

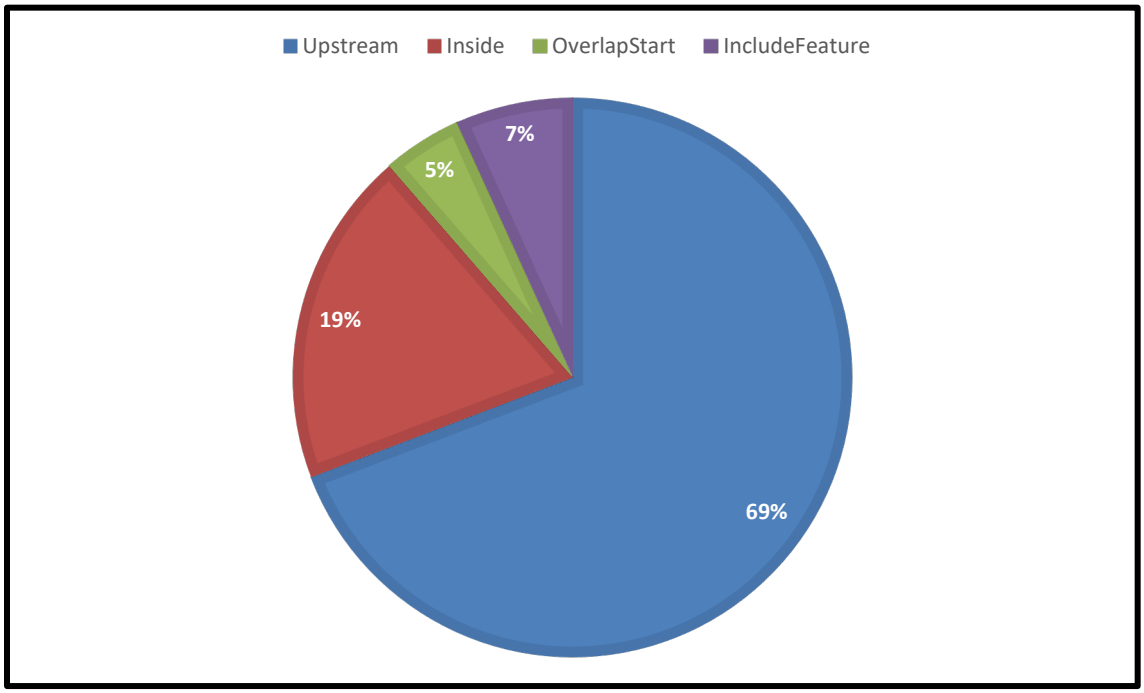
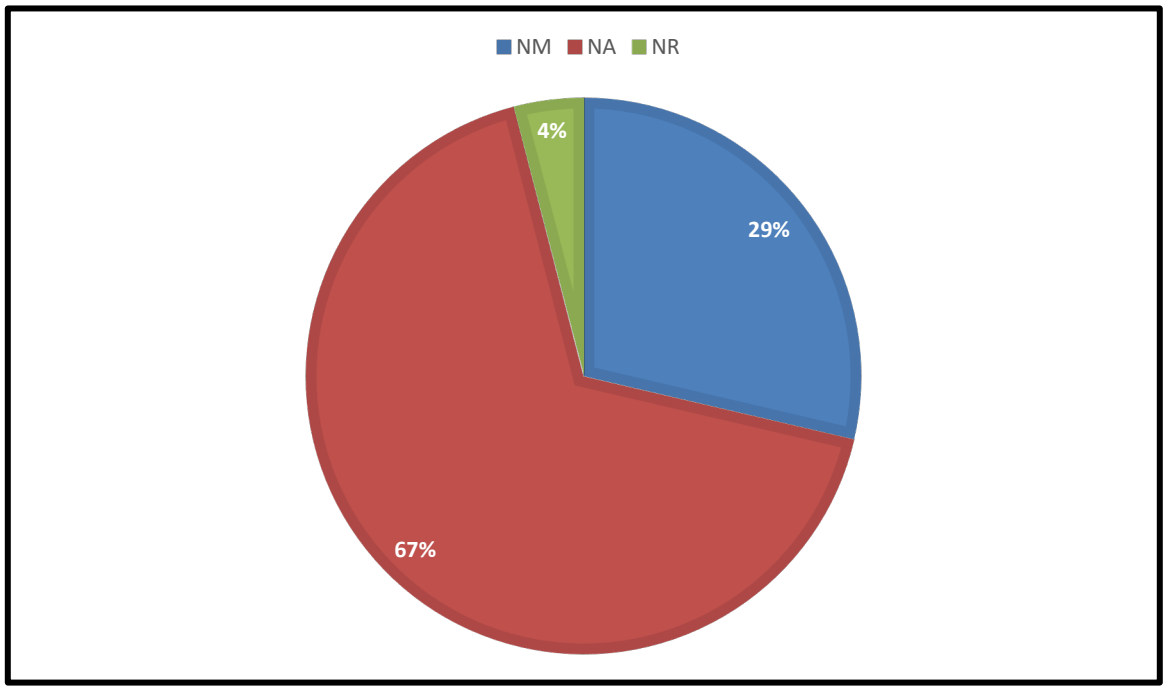


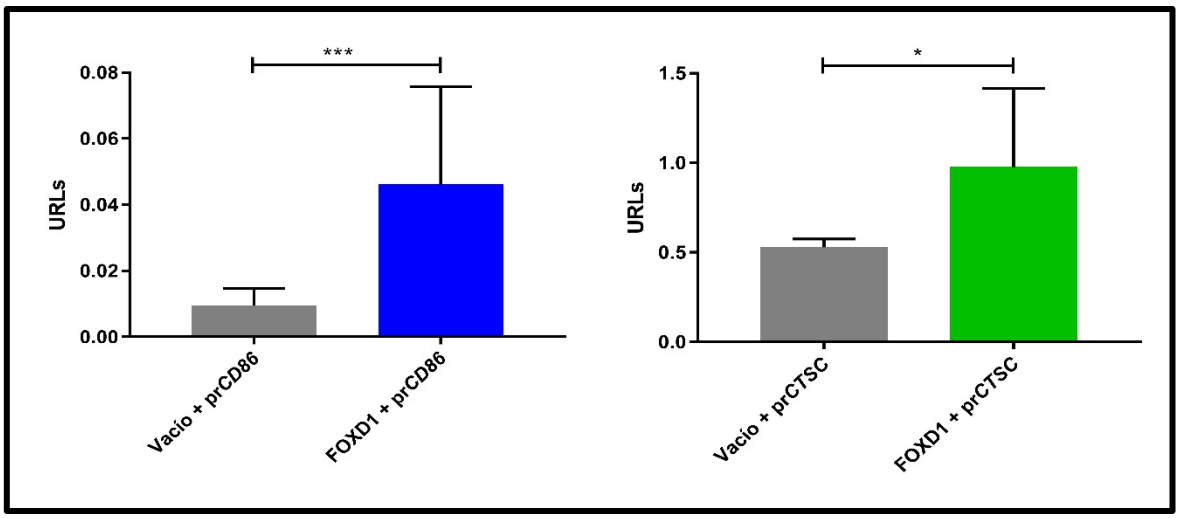
Figura 27. Gráfica de la distribución de la identificación de los genes blanco. **NM**: con RefSeq, **NA**: no identificado, **NR:** RNAs no codificantes.



Utilizando manualmente las bases de datos para determinar la plausibilidad biológica de los 177 picos de genes blanco dentro de la fisiología de la gestación y el desarrollo de la placenta, se encontraron 28 picos que estaban involucrados dentro de las cascadas moleculares de ambos procesos fisiológicos. Los promotores de estos genes se analizaron mediante la herramienta Genomatix para identificar los sitios putativos de unión a FOXD1. Se identificaron 12 genes con sitios putativos de unión. De éstos, 6 genes se encontraron sub-expresados en el estudio de transcriptómica realizado por *Laissue et al.,* (2016) [43]. Los genes blanco *CTSC, TRPC6, MMP-8, CD226* y *CD86* se encontraron sub-expresados en el tejido placentario de las hembras IRCS con la mutación Foxd1-p.Thr152Ala, mientras que el gen blanco *CMA1* se encontraba sub-expresado en el endometrio de este modelo murino. Por esta razón, se escogieron los genes blanco *CTSC, TRPC6, CD86* y *CMA1* para realizar las construcciones plasmídicas y los estudios funcionales posteriores a la validación de la ChIP-seq.

Se realizó una primera aproximación del ensayo gen reportero utilizando las construcciones de los promotores *CD86* y *CTSC*. Se evidenció que FOXD1 generaba un aumento estadísticamente significativo de la transactivación del promotor de *CD86* y de *CTSC* (4,9 y 1,8 veces, respectivamente) en comparación con el vector vacío (p=0,001 y p=0,026, respectivamente) (**Figura 28**). Estos resultados serán confirmados por EMSA/Supershift, puesto que el ensayo gen reportero es un aproximación indirecta.

Figura 28. Ensayo gen reportero de FOXD1 WT con el promotor de *CD86* y *CTSC.* * p<0.05, *** p<0.001



3.4 Discusión capítulo II

3.4.1 Estudio de FOXD1 como un potencial marcador molecular de AER, en FI durante FIV (ART), RCIU y PE

En este capítulo del trabajo de tesis se realizó el tamizaje de mutaciones de *FOXD1* en pacientes con FIR, AER, RCIU y PE, al igual que ensayos funcionales para las mutaciones FOXD1 p.His267Tyr y p.Arg57del. Los resultados del tamizaje realizado por secuenciación de Sanger en población colombiana y francesa sugirieron que estas mutaciones podrían ser causales de la FIR y del RCIU. Estas mutaciones afectaron residuos conservados evolutivamente.

Los ensayos funcionales demostraron que las mutaciones FOXD1 p.His267Tyr y FOXD1 p.Arg57del aumentaban significativamente (p<0,05 y p<0,001, respectivamente) la transactivación del promotor del gen blanco *C3* [43]. El incremento de los niveles de C3 podrían afectar el proceso de implantación y placentación en las pacientes con FIR/RCIU. Estudios realizados en pacientes con FIR y en vellosidades coriónicas demostraron, por medio de transcriptómica, que se presentaba una sobre-expresión de *C3*. Esta proteína está involucrada en la activación de la respuesta inmune innata que conduce a la apoptosis celular y potencia un ambiente pro-inflamatorio en el endometrio [113,115]. Modelos murinos deficientes de C3 han demostrado una mayor reabsorción embrionaria (fenotipo homólogo a AER) en comparación con los ratones WT debido a una alteración en el desarrollo de la placenta y del trofoblasto [219]. Por esta razón, se debe mantener un equilibrio entre los niveles de expresión de C3 durante la implantación, pues su incremento, como ocurre en los resultados previamente expuestos, puede conducir a disfunciones en la implantación como ocurre en el AER y la FIR, al igual que en la placentación como ocurre en la PE y el RCIU.

Con estos resultados establecemos que *FOXD1* es un gen implicado dentro de la patología de FIR, AER, RCIU y PE sirviendo de marcador molecular para estas enfermedades y para el pronóstico de los tratamientos de reproducción asistida (FIR). Es importante mencionar que el ensayo gen reportero no permite determinar la interacción directa entre el factor de transcripción y el promotor, por lo cual se necesita una verificación por ensayos de EMSA para determinar si las mutaciones encontradas en estas pacientes pueden alterar la unión de la proteína al promotor de *C3*. Por otra parte, los ensayos se realizaron sobre un modelo celular alejado a la células de los órganos en los que se desarrollan esta patologías, lo cual no ilustra de manera precisa lo que pudiera suceder en los tejidos asociados a la implantación y placentación. Modelos adicionales, fundamentados en la edición de genoma e implementados en modelos celulares y animales, se inscriben como una potencial alternativa para una mejor comprensión de mecanismos moleculares específicos asociados a estas mutaciones.

3.4.2 Identificación de nuevos genes blanco directos de FOXD1

Los resultados de la ChIP-seq obtenidos permitieron identificar nuevos posibles genes blanco directos de FOXD1, lo que genera nuevo conocimiento sobre la regulación transcripcional de este factor de transcripción en el proceso de implantación/mantenimiento de la gestación. En los experimentos de ChIP-seq entre el 79% y el 95% de las lecturas fueron mapeadas. Sin embargo, no se obtuvieron picos utilizando un valor de p=0,05 para el llamado de los picos. Por esta razón, la ChIP-seq no se consideró óptima en términos de especificidad. Los promotores de genes escogidos para la posterior verificación por ensayos gen reportero y EMSA fueron aquellos que tenían sitios putativos de unión a FOXD1 (tamizados por Genomatix) y que tenían una expresión diferencial en el endometrio o placenta, según los estudios de transcriptómica realizados por Laissue et al. [43]. En este marco, a pesar de que los genes blanco propuestos son candidatos teóricos relevantes desde una óptica fisiopatológica, las limitaciones técnicas evocadas refuerzan la necesidad de validar los resultados mediante técnicas adicionales.

Los genes blanco *CTSC*, *CD86* y *CMA1* participan en procesos inmunológicos durante la gestación. *CTSC* codifica para una cistein-proteasa involucrada en la degradación de proteínas intracelulares y en la activación de la cascada pro-inflamatoria producida por las células NK y por los linfocitos T citotóxicos [220–223]. Esta proteína se encuentra expresada en el tejido placentario y se ha sugerido que participa en la degradación de la matriz extracelular, en la invasión celular del trofoblasto y en la inflamación de la decidua [224].

El gen *CD86* codifica para un receptor involucrado en la co-estimulación de la proliferación de las células T y en la producción de IL-2 por unión a CTLA-4. Modelos murinos tratados con anti-CD86 han demostrado que la inhibición de CD86 facilita el cambio de una respuesta inmune, de Th1 a Th2, moderando la expresión de algunas citoquinas como el IFN-ɣ. Estos ratones también presentaban un incremento en las células $CD4^{+}CD25^{+}$ Tregs que evitan el rechazo del endometrio contra el embrión [225]. Estudios realizados en pacientes con preeclampsia

demostraron que alteraciones en la expresión de *VEGF* generaban la disminución en los niveles de CD86 en las células dendríticas, inhibiendo su maduración y alterando el balance inmunológico en la diferenciación de las células Th0 a Th1 [226].

El gen *CMA1* codifica para una proteasa secretada por los mastocitos durante la generación de péptidos vasoactivos y la degradación de la matriz extracelular. Se ha reportado que los mastocitos uterinos están involucrados en el remodelamiento de las arterias espirales. Alteraciones en este proceso se han relacionado con pérdidas de la gestación, y el desarrollo de preeclampsia o retardo del crecimiento intrauterino [227,228]. En este contexto, se ha postulado que CMA1 podría regular la tensión arterial actuando en conjunto con la angiotensina II y podría regular la activación de MMPs como MMP-9 y MMP-2 [229,230]. *Meyer et al.,* demostraron que la ablación de las células uNKs y de los mastocitos uterinos generaba alteraciones en el remodelamiento de las arterias espirales por disminución en la apoptosis, de manera secundaria a los bajos niveles de Cma1 (Mcpt-5) conduciendo a la restricción en el crecimiento fetal. Además, los autores demostraron que Cma1 estaba directamente involucrada en la migración de los trofoblastos extra-vellosos durante la gestación. En mujeres gestantes, los mastocitos uterinos se localizan cerca a los trofoblastos extra-vellosos sintetizando CMA1 y aumentando su capacidad migratoria y de invasión durante el remodelamiento de las arterias espirales [231–233]. En este marco, anormalidades en la expresión de esta proteasa podrían conducir al desarrollo de patologías reproductivas, especialmente asociadas a la implantación/placentación.

En lo que respecta *TRPC6*, codifica una proteína que conforma los canales de calcio que actúan directamente con IP_3 y DAG. *Clarson et al.,* demostraron que TRPC6 se encontraba expresada en el sincitiotrofoblasto durante el primer trimestre del embarazo y en placentas a término, lo que sugiere que esta proteína podría actuar como un operador de los canales para la entrada de calcio a la célula [234]. Estudios efectuados en modelos murinos demostraron que el aumento en la expresión de Trpc6 produce un incremento en la entrada de calcio en los vasos placentarios produciendo una mayor vasoconstricción y un aumento de la presión arterial [235].

Estos resultados sugieren que alteraciones en la expresión de esta proteína podría conducir a la aparición de preeclampsia y/o retardo del crecimiento intrauterino.

Otro de los resultados encontrados en los experimentos de ChIP-seq fue la presencia de picos dentro regiones asociadas a RNAs no codificantes. Algunos estudios han demostrado una relación entre los factores de transcripción de la familia *Forkhead* (FKHD) y los lncRNAs al formar un loop en la cromatina para crear un puente y acercar los promotores a los factores de transcripción y a los enhancers [236]. Además, se ha descrito que los ncRNAs se localizan cerca a algunos factores de transcripción FKHD en modelos murinos, como Foxa2 y Foxf1 actuando como enhancers RNAs (eRNAs) para reclutar complejos de proteínas o proteínas reguladoras en los sitios de transcripción [237]. Estos resultados podrían indicar que FOXD1 actúa en y sobre vías en las que participan ARNs no codificantes para regular la expresión de moléculas corriente-abajo.

4. Conclusiones

Los resultados de este trabajo de tesis doctoral permitieron identificar, mediante la técnica NGS-exoma, mutaciones en nuevos genes involucrados en diferentes vías moleculares de la implantación, potencialmente asociadas con la etiopatología del AER. El análisis bioinformático propuesto (focalizado en un subset de genes) puede considerarse como una aproximación eficiente para la identificación de mutaciones con un posible efecto funcional.

La identificación de la variante THBD-p.Trp153Gly en mujeres colombianas con AER y su validación mediante ensayos funcionales *in vitro* sugiere, por primera vez, una relación directa entre formas mutantes de esta proteína y la fisiopatología del AER. La mutación THBD-p.Trp153Gly se podría considerarse como un marcador molecular para el diagnóstico en pacientes colombianas con AER idiopático.

El tamizaje de variantes de secuencia en *FOXD1* en pacientes colombianas con FIR y AER y, en pacientes francesas con PE y RCIU permitió identificar 2 mutaciones nuevas en este factor de transcripción: FOXD1 p.His267Tyr y FOXD1 p.Arg57del. Los ensayos funcionales *in vitro* demostraron que estas mutaciones modifican la transactivación del promotor de *C3*, contribuyendo con el fenotipo. FOXD1 podría considerarse un marcador molecular diagnóstico para las pacientes con AER, FIR, RCIU y PE.

La identificación de los potenciales nuevos genes blanco directos de FOXD1 (*CTSC, CD86, CMA1* y *TRPC6*) aporta al conocimiento sobre la función regulatoria de este factor de transcripción y sobre las vías moleculares en las que podría actuar durante el proceso de la implantación/placentación. La información generada durante este trabajo de tesis aporta al conocimiento general sobre el origen del AER, la FIR y el RCIU/PE. Ésta puede ser de utilidad para los especialistas clínicos que trabajan en el área de la reproducción ya que genera nuevas herramientas potenciales para el diagnóstico de estas patologías en un contexto de medicina traslacional.

5. Recomendaciones y perspectivas

- La aproximación NGS debe implementarse en otras poblaciones para identificar posibles nuevas mutaciones asociadas a AER y de FIR.

- Otras formatos y aplicaciones de la NGS, como la secuenciación de genoma completo y secuenciación la de célula única, podrían ampliar el conocimiento sobre las vías moleculares que actúan en el proceso de implantación, del AER y la FIR.

- La validación funcional *in vitro* e *in vivo* de las variantes encontradas por NGS y en FOXD1, usando la técnica de edición del genoma CRISPR/Cas9 para generar modelos "*knock-in",* podría ampliar el conocimiento sobre las implicaciones funcionales de estas mutaciones durante el desarrollo de la del fenotipo. Esto permitiría describir nuevos biomarcadores de diagnóstico/pronóstico del AER, de la FIR, de la PE y del RCIU.

- Un tamizaje genético de *THBD* y *FOXD1* en grandes cohortes de pacientes con AER, FIR, PE y RCIU, de diferente origen étnico, permitiría identificar nuevas mutaciones específicas de cada población y establecer una relación genotipo-fenotipo.

- La validación de los resultados de la ChIP-seq de FOXD1 por medio de EMSA/Supershift es necesaria para generar nuevo conocimiento sobre los genes blanco directos regulados por FOXD1 en el contexto placentario.

6. Anexos

Anexo I: <u>Quintero-Ronderos</u> P, Laissue P. *"Classic and novel sequencing methods to dissect recurrent pregnancy loss aetiology".* Reproductive Science. Submitted. Under review

Anexo II: <u>Quintero-Ronderos P</u>, Mercier E, Fukuda M, González R, Suárez CF, Patarroyo MA, Vaiman D, Gris JC, Laissue P. *"Novel genes and mutations in patients affected by recurrent pregnancy loss".* PLoS One. 2017 Oct 10;12(10):e0186149

Anexo III: <u>Quintero-Ronderos P</u>, Mercier E, Gris JC, Esteban-Perez C, Moreno-Ortiz H, Fonseca DJ, Lucena E, Vaiman D, Laissue P. *"THBD sequence variants potentiallyrelated to recurrent pregnancy loss".* Reprod Biol Endocrinol. 2017 Dec 1;15(1):92

Anexo IV: <u>Quintero-Ronderos P,</u> Laissue P. *"The multisystemic functions of FOXD1 in development and disease".* J Mol Med (Berl). 2018 Aug;96(8):725-739

Anexo V**: **Quintero-Ronderos P, Jiménez K, Esteban-Pérez C, Ojeda D, Bello S, Fonseca DJ, Coronel MA, Moreno-Ortiz H, Sierra-Díaz DC, Lucena E, Barbaux S, Vaiman D, Laissue P. *"FOXD1 mutations are related to repeated implantation failure, intra-uterine growth restriction and preeclampsia".* Human Mutation. Submitted. Under review

7. Bibliografía

1. Li, T. C. *et al.* Recurrent miscarriage: aetiology, management and prognosis. *Hum. Reprod. Update* **8,** 463–81 (2002).
2. Rai, R. & Regan, L. Recurrent miscarriage. *Lancet* **368,** 601–611 (2006).
3. Stirrat, G. M. Recurrent miscarriage. *Lancet* **336,** 673–675 (1990).
4. Hogge, W. A., Byrnes, A. L., Lanasa, M. C. & Surti, U. The clinical use of karyotyping spontaneous abortions. *Am. J. Obstet. Gynecol.* **189,** 397-400; discussion 400–2 (2003).
5. Bricker Leanne & Farquharson Roy G. Recurring miscarriage. *Obstet. Gynaecol.* **2,** 17–23 (2000).
6. Rai, R. S., Clifford, K., Cohen, H. & Regan, L. High prospective fetal loss rate in untreated pregnancies of women with recurrent miscarriage and antiphospholipid antibodies. *Hum. Reprod.* **10,** 3301–3304 (1995).
7. Santos, T. da S. *et al.* Antiphospholipid syndrome and recurrent miscarriage: A systematic review and meta-analysis. *J. Reprod. Immunol.* **123,** 78–87 (2017).
8. Mak, I. Y. H. *et al.* Regulated expression of signal transducer and activator of transcription, Stat5, and its enhancement of PRL expression in human endometrial stromal cells in vitro. *J. Clin. Endocrinol. Metab.* **87,** 2581–2588 (2002).
9. Bose, P. *et al.* Heparin and aspirin attenuate placental apoptosis in vitro: Implications for early pregnancy failure. *Am. J. Obstet. Gynecol.* **192,** 23–30 (2005).
10. Salim, R., Regan, L., Woelfer, B., Backos, M. & Jurkovic, D. A comparative study of the morphology of congenital uterine anomalies in women with and without a history of recurrent first trimester miscarriage. *Hum. Reprod.* (2003). doi:10.1093/humrep/deg030
11. Grimbizis, G. F., Camus, M., Tarlatzis, B. C., Bontis, J. N. & Devroey, P. Clinical implications of uterine malformations and hysteroscopic treatment results. *Human Reproduction Update* **7,** 161–174 (2001).
12. Hart, R. *et al. A prospective controlled study of the effect of intramural uterine fibroids on the outcome of assisted conception. Human reproduction (Oxford, England)* **16,** (2001).
13. Rackow, B. W. & Taylor, H. S. Submucosal uterine leiomyomas have a global effect on molecular determinants of endometrial receptivity. *Fertil. Steril.* **93,** 2027–2034 (2010).
14. Hirahara, F. *et al.* Hyperprolactinemic recurrent miscarriage and results of randomized bromocriptine treatment trials. *Fertil. Steril.* **70,** 246–252 (1998).
15. Garzia, E. *et al.* Lack of expression of endometrial prolactin in early implantation failure: A pilot study. *Hum. Reprod.* **19,** 1911–1916 (2004).
16. Craig, L. B., Ke, R. W. & Kutteh, W. H. Increased prevalence of insulin resistance in women with a history of recurrent pregnancy loss. *Fertil. Steril.* **78,** 487–490 (2002).
17. Rai, R., Backos, M., Rushworth, F. & Regan, L. Polycystic ovaries and recurrent

miscarriage--a reappraisal. *Hum. Reprod.* **15,** 612–615 (2000).

18. Kaur, R. & Gupta, K. Endocrine dysfunction and recurrent spontaneous abortion: An overview. *Int. J. Appl. Basic Med. Res.* **6,** 79 (2016).

19. Clifford, K., Flanagan, A. M. & Regan, L. Endometrial CD56+ natural killer cells in women with recurrent miscarriage: a histomorphometric study. *Hum. Reprod.* **14,** 2727–2730 (1999).

20. Reinhard, G., Noll, A., Schlebusch, H., Mallmann, P. & Ruecker, A. Shifts in the TH1/TH2 balance during human pregnancy correlate with apoptotic changes. *Biochem. Biophys. Res. Commun.* **245,** 933–938 (1998).

21. Beaman, K. D. *et al.* Immune Etiology of Recurrent Pregnancy Loss and Its Diagnosis. *Am. J. Reprod. Immunol.* **67,** 319–325 (2012).

22. Ralph, S. G., Rutherford, A. J. & Wilson, J. D. Influence of bacterial vaginosis on conception and miscarriage in the first trimester: cohort study. *BMJ* **319,** 220–223 (1999).

23. Hay, P. E. *et al.* Abnormal bacterial colonisation of the genital tract and subsequent preterm delivery and late miscarriage. *BMJ* **308,** 295–298 (1994).

24. Isik, G., Demirezen, Ş., Dönmez, H. & Beksaç, M. Bacterial vaginosis in association with spontaneous abortion and recurrent pregnancy losses. *J. Cytol.* **33,** 135 (2016).

25. Giakoumelou, S. *et al.* The role of infection in miscarriage. *Hum. Reprod. Update* **22,** 116–133 (2016).

26. Jia, C.-W. *et al.* Aneuploidy in Early Miscarriage and its Related Factors. *Chin. Med. J. (Engl).* **128,** 2772 (2015).

27. Hyde, K. J. & Schust, D. J. Genetic Considerations in Recurrent Pregnancy Loss. *Cold Spring Harb. Perspect. Med.* **5,** a023119–a023119 (2015).

28. Sierra S & Stephenson M. Genetics of recurrent pregnancy loss. *Semin Reprod Med.* **24,** 17–24. (2006).

29. Stephenson, M. D., Awartani, K. A. & Robinson, W. P. Cytogenetic analysis of miscarriages from couples with recurrent miscarriage: a case-control study. *Hum. Reprod.* (2002). doi:10.1016/S0015-0282(01)02285-3

30. Hassold, T. & Hunt, P. To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* **2,** 280–291 (2001).

31. Rubio, C. *et al.* Chromosomal abnormalities and embryo development in recurrent miscarriage couples. *Hum. Reprod.* (2003). doi:10.1093/humrep/deg015

32. Choi, T. Y., Lee, H. M., Park, W. K., Jeong, S. Y. & Moon, H. S. Spontaneous abortion and recurrent miscarriage: A comparison of cytogenetic diagnosis in 250 cases. *Obstet. Gynecol. Sci.* **57,** 518 (2014).

33. Coelho, F. F. *et al.* Detection of aneuploidies in spontaneous abortions by quantitative fluorescent PCR with short tandem repeat markers: a retrospective study. *Genet. Mol. Res.* **15,** (2016).

34. ESHRE Capri Workshop Group. Genetic aspects of female reproduction. *Hum. Reprod. Update* **14,** 293–307. (2008).

35. Menasha, J., Levy, B., Hirschhorn, K. & Kardon, N. B. Incidence and spectrum of chromosome abnormalities in spontaneous abortions: New insights from a 12-year study. *Genet. Med.* **7,** 251–263 (2005).

36. Uehara, S. *et al.* Preferential X-chromosome inactivation in women with idiopathic recurrent pregnancy loss. in *Fertility and Sterility* **76,** 908–914 (2001).

37. Sangha, K. K., Stephenson, M. D., Brown, C. J. & Robinson, W. P. Extremely skewed X-chromosome inactivation is increased in women with recurrent spontaneous abortion. *American journal of human genetics* **65,** 913–917 (1999).

38. Aldrich, C. L. *et al. HLA-G genotypes and pregnancy outcome in couples with unexplained recurrent miscarriage. Molecular human reproduction* **7,** (2001).

39. Pfeiffer, K. A., Fimmers, R., Engels, G., van der Ven, H. & van der Ven, K. The HLA-G genotype is potentially associated with idiopathic recurrent spontaneous abortion. *Mol. Hum. Reprod.* **7,** 373–8 (2001).

40. Kovalevsky, G., Gracia, C. R., Berlin, J. A., Sammel, M. D. & Barnhart, K. T. Evaluation of the association between hereditary thrombophilias and recurrent pregnancy loss: a meta-analysis. *Arch. Intern. Med.* **164,** 558–563 (2004).

41. Rey, E., Kahn, S. R., David, M. & Shrier, I. Thrombophilic disorders and fetal loss: a meta-analysis. *Lancet (London, England)* **361,** 901–8 (2003).

42. Rai, R., Backos, M., Elgaddal, S., Shlebak, A. & Regan, L. Factor V Leiden and recurrent miscarriage-prospective outcome of untreated pregnancies. *Hum. Reprod.* **17,** 442–445 (2002).

43. Laissue, P. *et al.* Association of FOXD1 variants with adverse pregnancy outcomes in mice and humans. *Open Biol.* **6,** 160109 (2016).

44. Polanski, L. T. *et al.* What exactly do we mean by 'recurrent implantation failure'? A systematic review and opinion. *Reprod. Biomed. Online* **28,** 409–423 (2014).

45. Croucher, C. A., Lass, A., Margara, R. & Winston, R. M. Predictive value of the results of a first in-vitro fertilization cycle on the outcome of subsequent cycles. *Hum. Reprod.* **13,** 403–8 (1998).

46. Yang, R. *et al.* Biochemical pregnancy and spontaneous abortion in first IVF cycles are negative predictors for subsequent cycles: an over 10,000 cases cohort study. *Arch. Gynecol. Obstet.* **292,** 453–458 (2015).

47. Simon, A. & Laufer, N. Assessment and treatment of repeated implantation failure (RIF). *J. Assist. Reprod. Genet.* **29,** 1227–1239 (2012).

48. Rinehart, J. Recurrent implantation failure: Definition. *J. Assist. Reprod. Genet.* **24,** 284–287 (2007).

49. Timeva, T., Shterev, A. & Kyurkchiev, S. Recurrent implantation failure: The role of the endometrium. *J. Reprod. Infertil.* **15,** 173–183 (2014).

50. Simon, A. & Laufer, N. Repeated implantation failure: Clinical approach. *Fertil. Steril.* **97,** 1039–1043 (2012).

51. El-Toukhy, T. & Taranissi, M. Towards better quality research in recurrent implantation failure: standardizing its definition is the first step. *Reprod. Biomed. Online* **12,** 383–5 (2006).

52. Levi Setti, P. E. *et al.* Implantation failure in assisted reproduction technology and a critical approach to treatment. *Ann. N. Y. Acad. Sci.* **1034,** 184–199 (2004).

53. Ocal, P. *et al.* Recurrent implantation failure is more frequently seen in female patients with poor prognosis. *Int. J. Fertil. Steril.* **6,** 71–8 (2012).

54. Regan, L., Braude, P. R. & Trembath, P. L. Influence of past reproductive performance on risk of spontaneous abortion. *BMJ* **299,** 541–545 (1989).

55. Knudsen, U. B., Hansen, V., Juul, S. & Secher, N. J. Prognosis of a new pregnancy following previous spontaneous abortions. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **39,** 31–36 (1991).

56. Strobino, B. *et al.* Characteristics of women with recurrent spontaneous abortions and women with favorable reproductive histories. *Am. J. Public Health* **76,** 986–91 (1986).

57. Kolte, A. M. M. *et al.* A genome-wide scan in affected sibling pairs with idiopathic recurrent miscarriage suggests genetic linkage. *MHR Basic Sci. Reprod. Med.* **17,** 379–385 (2011).

58. Shekouhi, S. *et al.* Identification of Xq22.1-23 as a region linked with hereditary recurrent spontaneous abortion in a family. *Iran. J. Reprod. Med.* **11,** 659–64 (2013).

59. Li Wang, Zeng Chan Wang, Cui Xie, Xiao Feng Liu & Mao Sheng Yang. Genome-Wide Screening for Risk Loci of Idiopathic Recurrent Miscarriage in a Han Chinese Population: A Pilot Study. *Reprod. Sci.* **17,** 578–584 (2010).

60. Kaare, M. *et al.* Variations in the thrombomodulin and endothelial protein C receptor genes in couples with recurrent miscarriage. *Hum. Reprod.* **22,** 864–8 (2007).

61. Mercier, E., Lissalde-Lavigne, G. & Gris, J.-C. *JAK2* V617F Mutation in Unexplained Loss of First Pregnancy. *N. Engl. J. Med.* **357,** 1984–1985 (2007).

62. Kaare, M., Painter, J. N., Ulander, V. M., Kaaja, R. & Aittomäki, K. Variations of the amnionless gene in recurrent spontaneous abortions. *Mol. Hum. Reprod.* **12,** 25–29 (2006).

63. Neveling, K. *et al.* A Post-Hoc Comparison of the Utility of Sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases. *Hum. Mutat.* **34,** 1721–1726 (2013).

64. Baudhuin, L. M. *et al.* Confirming Variants in Next-Generation Sequencing Panel Testing by Sanger Sequencing. *J. Mol. Diagnostics* **17,** 456–461 (2015).

65. Laissue, P. Aetiological coding sequence variants in non-syndromic premature ovarian failure: From genetic linkage analysis to next generation sequencing. *Mol. Cell. Endocrinol.* **411,** 243–257 (2015).

66. Lee, K. Y. & DeMayo, F. J. Animal models of implantation. *Reproduction* **128,** 679–695 (2004).

67. Wilcox, A. J., Baird, D. D. & Weinberg, C. R. Time of Implantation of the Conceptus and Loss of Pregnancy. *N. Engl. J. Med.* **340,** 1796–1799 (1999).

68. Wimsatt, W. A. Some comparative aspects of implantation. *Biol. Reprod.* **12,** 1–40 (1975).

69. Aghajanova, L. *et al.* Comparative Transcriptome Analysis of Human Trophectoderm and Embryonic Stem Cell-Derived Trophoblasts Reveal Key Participants in Early Implantation. *Biol. Reprod.* **86,** 1–21 (2012).

70. Cha, J., Sun, X. & Dey, S. K. Mechanisms of implantation: strategies for successful pregnancy. *Nat. Med.* **18,** 1754–1767 (2012).

71. King, A. Uterine leukocytes and decidualization. *Hum. Reprod. Update* **6,** 28–36 (2000).

72. Christian, M. *et al.* Interferon-gamma modulates prolactin and tissue factor expression in differentiating human endometrial stromal cells. *Endocrinology* **142,** 3142–51 (2001).

73. Irwin, J. C. & Giudice, L. C. Insulin-like growth factor binding protein-1 binds to placental cytotrophoblast alpha5beta1 integrin and inhibits cytotrophoblast invasion into decidualized endometrial stromal cultures. *Growth Horm. IGF Res.* **8,** 21–31 (1998).

74. Kusama, K., Yoshie, M., Tamura, K., Imakawa, K. & Tachikawa, E. EPAC2-mediated calreticulin regulates LIF and COX2 expression in human endometrial glandular cells. *J. Mol. Endocrinol.* **54,** 17–24 (2014).

75. Li, Q. *et al.* WNT4 Acts Downstream of BMP2 and Functions via β-Catenin Signaling Pathway to Regulate Human Endometrial Stromal Cell Differentiation. *Endocrinology* **154,** 446–457 (2013).

76. Aghajanova, L. Leukemia inhibitory factor and human embryo implantation. *Ann N Y Acad Sci* **1034,** 176–183 (2004).

77. Kojima, K. *et al.* Expression of leukaemia inhibitory factor (LIF) receptor in human placenta: a possible role for LIF in the growth and differentiation of trophoblasts. *Hum. Reprod.* **10,** 1907–11 (1995).

78. Carson, D. D. *et al.* Changes in gene expression during the early to mid-luteal (receptive phase) transition in human endometrium detected by high-density microarray screening. *Mol. Hum. Reprod.* **8,** 871–879 (2002).

79. Kao, L. C. *et al.* Global gene profiling in human endometrium during the window of implantation. *Endocrinology* **143,** 2119–2138 (2002).

80. Popovici, R. M., Kao, L. C. & Giudice, L. C. Discovery of new inducible genes in in vitro decidualized human endometrial stromal cells using microarray technology. *Endocrinology* **141,** 3510–3513 (2000).

81. Uegaki, K. *et al.* PTEN is involved in the signal transduction pathway of contact inhibition in endometrial cells. *Cell Tissue Res.* **323,** 523–528 (2006).

82. Ejskjaer, K. *et al.* Expression of the epidermal growth factor system in human endometrium during the menstrual cycle. *Mol. Hum. Reprod.* **11,** 543–551 (2005).

83. Kats, R., Al-Akoum, M., Guay, S., Metz, C. & Akoum, A. Cycle-dependent expression of macrophage migration inhibitory factor in the human endometrium. *Hum. Reprod.* **20,** 3518–3525 (2005).

84. Print, C. *et al.* Soluble factors from human endometrium promote angiogenesis and regulate the endothelial cell transcriptome. *Hum. Reprod.* **19,** 2356–2366 (2004).

85. Stavreus-Evers, A. *et al.* Co-existence of heparin-binding epidermal growth factor-like growth factor and pinopodes in human endometrium at the time of implantation. *Mol. Hum. Reprod.* **8,** 765–769 (2002).

86. Nikas, G. & Aghajanova, L. Endometrial pinopodes: some more understanding on human implantation? *Reprod. Biomed. Online* **4 Suppl 3,** 18–23 (2002).

87. Jones, R. L., Stoikos, C., Findlay, J. K. & Salamonsen, L. A. TGF-?? superfamily expression and actions in the endometrium and placenta. *Reproduction* (2006). doi:10.1530/rep.1.01076

88. Li, Y.-H., Kuo, C.-H., Shi, G.-Y. & Wu, H.-L. The role of thrombomodulin lectin-like domain in inflammation. *J. Biomed. Sci.* **19,** 34 (2012).

89. Conway, E. M. *et al.* The lectin-like domain of thrombomodulin confers protection from neutrophil-mediated tissue damage by suppressing adhesion molecule expression via nuclear factor kappaB and mitogen-activated protein kinase pathways. *J. Exp. Med.* **196,** 565–77 (2002).

90. Van Dreden, P., Woodhams, B., Rousseau, A., Favier, M. & Favier, R. Comparative evaluation of Tissue factor and Thrombomodulin activity changes during normal and idiopathic early and late foetal loss: The cause of hypercoagulability? *Thromb. Res.* **129,** 787–792 (2012).

91. Isermann, B. *et al.* The thrombomodulin–protein C system is essential for the maintenance of pregnancy. *Nat. Med.* **9,** 331–337 (2003).

92. Griesshammer, M., Struve, S. & Harrison, C. M. Essential thrombocythemia/polycythemia vera and pregnancy: The need for an observational study in Europe. *Semin. Thromb. Hemost.* **32,** 422–429 (2006).

93. Kralovics, R. *et al.* A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N. Engl. J. Med.* **352,** 1779–90 (2005).

94. Melillo, L. *et al.* Outcome of 122 pregnancies in essential thrombocythemia patients: A report from the Italian registry. *Am. J. Hematol.* **84,** 636–640 (2009).

95. Passamonti, F. *et al.* Increased risk of pregnancy complications in patients with essential thrombocythemia carrying the JAK2 (617V>F) mutation. *Blood* **110,** 485–9 (2007).

96. Barber, L. J. *et al.* Comprehensive genomic analysis of a BRCA2 deficient human pancreatic cancer. *PLoS One* (2011). doi:10.1371/journal.pone.0021639

97. Laissue, P. The molecular complexity of primary ovarian insufficiency aetiology and the use of massively parallel sequencing. *Mol. Cell. Endocrinol.* **460,** 170–180 (2018).

98. Çalişkan, M. *et al.* Exome sequencing reveals a novel mutation for autosomal recessive non-syndromic mental retardation in the TECR gene on chromosome 19p13. *Hum. Mol. Genet.* (2011). doi:10.1093/hmg/ddq569

99. Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. *Nature Reviews Genetics* (2012). doi:10.1038/nrg3241

100. Hrdlickova, B., de Almeida, R. C., Borek, Z. & Withoff, S. Genetic variation in the non-coding genome: Involvement of micro-RNAs and long non-coding RNAs in disease. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1842,** 1910–1922 (2014).

101. Huang, Q. Genetic Study of Complex Diseases in the Post-GWAS Era. *J. Genet. Genomics* **42,** 87–98 (2015).

102. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13,** 135–145 (2012).

103. Lettre, G. Rare and low-frequency variants in human common diseases and other complex traits. *J. Med. Genet.* **51,** 705–14 (2014).

104. Zuk, O. *et al.* Searching for missing heritability: Designing rare variant association studies. *Proc. Natl. Acad. Sci.* **111,** E455–E464 (2014).

105. Mitropoulos, K. *et al.* Success stories in genomic medicine from resource-limited countries. *Hum. Genomics* **9,** 11 (2015).

106. Fonseca, D. J. *et al.* Next generation sequencing in women affected by nonsyndromic premature ovarian failure displays new potential causative genes and mutations. *Fertil. Steril.* **104,** 154–62.e2 (2015).

107. Patiño, L. C. *et al.* Exome Sequencing Is an Efficient Tool for Variant Late-Infantile Neuronal Ceroid Lipofuscinosis Molecular Diagnosis. *PLoS One* **9,** e109576 (2014).

108. Ortega-Recalde, O. *et al.* Whole-Exome Sequencing Enables Rapid Determination of Xeroderma Pigmentosum Molecular Etiology. *PLoS One* **8,** e64692 (2013).

109. Diggle, C. P. *et al.* Prostaglandin transporter mutations cause pachydermoperiostosis with myelofibrosis. *Hum. Mutat.* **33,** 1175–1181 (2012).

110. Patiño, L. C. *et al.* New mutations in non-syndromic primary ovarian insufficiency patients identified via whole-exome sequencing. *Hum. Reprod.* **32,** 1512–1520 (2017).

111. Carlosama, C. *et al.* A homozygous donor splice-site mutation in the meiotic gene MSH4 causes primary ovarian insufficiency. *Hum. Mol. Genet.* **26,** 3161–3166 (2017).

112. Qiao, Y. *et al.* Whole exome sequencing in recurrent early pregnancy loss. *Mol. Hum. Reprod.* **22,** 364–72 (2016).

113. Sõber, S. *et al.* RNA sequencing of chorionic villi from recurrent pregnancy loss patients reveals impaired function of basic nuclear and cellular machinery. *Sci. Rep.* **6,** 38439 (2016).

114. Wang, J. mei *et al.* Deep-sequencing identification of differentially expressed miRNAs in decidua and villus of recurrent miscarriage patients. *Arch. Gynecol. Obstet.* **293,** 1125–1135 (2016).

115. Huang, J. *et al.* A comparison of transcriptomic profiles in endometrium during window of implantation between women with unexplained recurrent implantation failure and recurrent miscarriage. *Reproduction* **153,** 749–758 (2017).

116. Quintero-Ronderos, P. *et al.* Novel genes and mutations in patients affected by recurrent pregnancy loss. *PLoS One* **12,** e0186149 (2017).

117. Lissalde-Lavigne, G. *et al.* Factor V Leiden and prothrombin G20210A polymorphisms as risk factors for miscarriage during a first intended pregnancy: the matched case-control 'NOHA first' study. *J. Thromb. Haemost.* **3,** 2178–84 (2005).

118. Wang, Q., Shashikant, C. S., Jensen, M., Altman, N. S. & Girirajan, S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci. Rep.* **7,** 885 (2017).

119. Lee, S. M., Wu, B. & Kersey, J. H. Likelihood-Based Approach to Gene Set Enrichment Analysis with a Finite Mixture Model. *Stat. Biosci.* **6,** 38–54 (2014).

120. Heifetz, A. *et al.* The Fragment Molecular Orbital Method Reveals New Insight into the Chemical Nature of GPCR–Ligand Interactions. *J. Chem. Inf. Model.* **56,** 159–172 (2016).

121. Fedorov, D. G., Nagata, T. & Kitaura, K. Exploring chemistry with the fragment molecular orbital method. *Phys. Chem. Chem. Phys.* **14,** 7562 (2012).

122. Stewart, J. J. MOPAC: a semiempirical molecular orbital program. *J. Comput. Aided. Mol. Des.* **4,** 1–105 (1990).

123. Hitaoka, S., Chuman, H. & Yoshizawa, K. A QSAR study on the inhibition mechanism of matrix metalloproteinase-12 by arylsulfone analogs based on molecular orbital calculations. *Org. Biomol. Chem.* **13,** 793–806 (2015).

124. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132,** 365–86 (2000).

125. Hu, Z. *et al.* Structural Insights into the pH-Dependent Conformational Change and Collagen Recognition of the Human Mannose Receptor. *Structure* **26,** 60–71.e3 (2018).

126. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5,** 725–738 (2010).

127. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9,** 40 (2008).

128. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12,** 7–8 (2015).

129. Pettersen, E. F. *et al.* UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

130. Dunbrack, R. L. Rotamer libraries in the 21st century. *Curr. Opin. Struct. Biol.* **12,** 431–40 (2002).

131. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).

132. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34,** 2135–2145 (2013).

133. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).

134. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27,** 343–350 (2011).

135. Chen, V. B. *et al.* MolProbity : all-atom structure validation for macromolecular

crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

136. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14,** 33–38 (1996).

137. Voss, N. R. & Gerstein, M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res.* **38,** W555–W562 (2010).

138. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *J. Chem. Theory Comput.* **7,** 2284–2295 (2011).

139. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.* **98,** 10037–10041 (2001).

140. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32,** W665–W667 (2004).

141. Sims, D., Sudbery, I., Ilott, N. E., Heger, A. & Ponting, C. P. Sequencing depth and coverage: key considerations in genomic analyses. *Nat. Rev. Genet.* **15,** 121–132 (2014).

142. Meynert, A. M., Ansari, M., FitzPatrick, D. R. & Taylor, M. S. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics* **15,** 247 (2014).

143. Kim, K. *et al.* Effect of Next-Generation Exome Sequencing Depth for Discovery of Diagnostic Variants. *Genomics Inform.* **13,** 31 (2015).

144. Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* **44,** 623–630 (2012).

145. Goldstein, D. B. *et al.* Sequencing studies in human genetics: Design and interpretation. *Nat. Rev. Genet.* **14,** 460–470 (2013).

146. Tennessen, J. A. *et al.* Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Sci. (New York, NY)* **337,** 64–69 (2012).

147. O'Roak, B. J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* **485,** 246–250 (2012).

148. Delcour, C. *et al.* ATG7 and ATG9A loss-of-function variants trigger autophagy impairment and ovarian failure. *Genet. Med.* (2018). doi:10.1038/s41436-018-0287-y

149. Thusberg, J., Olatubosun, A. & Vihinen, M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum. Mutat.* **32,** 358–368 (2011).

150. Walters-Sen, L. C. *et al.* Variability in pathogenicity prediction programs: impact on clinical diagnostics. *Mol. Genet. Genomic Med.* **3,** 99–110 (2015).

151. Dawood, F., Mountford, R., Farquharson, R. & Quenby, S. Genetic polymorphisms on the factor V gene in women with recurrent miscarriage and acquired APCR. *Hum. Reprod.* **22,** 2546–2553 (2007).

152. Altintas, A. *et al.* Factor V Leiden and G20210A prothrombin mutations in patients

with recurrent pregnancy loss: data from the southeast of Turkey. *Ann. Hematol.* **86,** 727–731 (2007).

153. Aytekin, E., Ergun, S. G., Ergun, M. A. & Percin, F. E. Evaluation of GenoFlow Thrombophilia Array Test Kit in Its Detection of Mutations in Factor V Leiden (G1691A), Prothrombin G20210A, MTHFR C677T and A1298C in Blood Samples from 113 Turkish Female Patients. *Genet. Test. Mol. Biomarkers* **18,** 717–721 (2014).

154. Asselta, R., Tenchini, M. L. & Duga, S. Inherited defects of coagulation factor V: the hemorrhagic side. *J. Thromb. Haemost.* **4,** 26–34 (2006).

155. Steen, M. & Dahlbäck, B. Thrombin-mediated proteolysis of factor V resulting in gradual B-domain release and exposure of the factor Xa-binding site. *J. Biol. Chem.* **277,** 38424–30 (2002).

156. Erdogan, E., Bukys, M. A. & Kalafatis, M. The contribution of amino acid residues 1508-1515 of factor V to light chain generation. *J. Thromb. Haemost.* **6,** 118–24 (2008).

157. Peng, W., Quinn-Allen, M. A. & Kane, W. H. Mutation of hydrophobic residues in the factor Va C1 and C2 domains blocks membrane-dependent prothrombin activation. *J. Thromb. Haemost.* **3,** 351–4 (2005).

158. Otto, P. & Ladik, J. Investigation of the interaction between molecules at medium distances. *Chem. Phys.* **8,** 192–200 (1975).

159. Fiedler, B. & Friedrich, J. in 132–190 doi:10.1039/9781782626862-00132

160. Xie, W. & Gao, J. Design of a Next Generation Force Field: The X-POL Potential. *J. Chem. Theory Comput.* **3,** 1890–1900 (2007).

161. Batra, J. *et al.* Matrix Metalloproteinase-10 (MMP-10) Interaction with Tissue Inhibitors of Metalloproteinases TIMP-1 and TIMP-2: BINDING STUDIES AND CRYSTAL STRUCTURE. *J. Biol. Chem.* **287,** 15935–15946 (2012).

162. Maruyama, I., Bell, C. E. & Majerus, P. W. Thrombomodulin is found on endothelium of arteries, veins, capillaries, and lymphatics, and on syncytiotrophoblast of human placenta. *J. Cell Biol.* **101,** 363–71 (1985).

163. Martin, F. A., Murphy, R. P. & Cummins, P. M. Thrombomodulin and the vascular endothelium: insights into functional, regulatory, and therapeutic aspects. *Am. J. Physiol. Heart Circ. Physiol.* **304,** H1585-97 (2013).

164. Masini, S. *et al.* Thrombin-activatable fibrinolysis inhibitor polymorphisms and recurrent pregnancy loss. *Fertil. Steril.* **92,** 694–702 (2009).

165. Ito, T., Kakihana, Y. & Maruyama, I. Thrombomodulin as an intravascular safeguard against inflammatory and thrombotic diseases. *Expert Opin. Ther. Targets* **20,** 151–158 (2016).

166. Delvaeye, M. *et al.* Thrombomodulin Mutations in Atypical Hemolytic–Uremic Syndrome. *N. Engl. J. Med.* **361,** 345–357 (2009).

167. Huang, H.-C. *et al.* Thrombomodulin-mediated cell adhesion: involvement of its lectin-like domain. *J. Biol. Chem.* **278,** 46750–9 (2003).

168. Stortoni, P. *et al.* Placental thrombomodulin expression in recurrent miscarriage.

Reprod. Biol. Endocrinol. **8,** 1 (2010).

169. Suzuki, K. *et al.* A domain composed of epidermal growth factor-like structures of human thrombomodulin is essential for thrombin binding and for protein C activation. *J. Biol. Chem.* **264,** 4872–4876 (1989).

170. Sood, R. *et al.* Maternal Par4 and platelets contribute to defective placenta formation in mouse embryos lacking thrombomodulin. *Blood* **112,** 585–91 (2008).

171. de Saint Martin, L. *et al.* Increased thrombin generation measured in the presence of thrombomodulin in women with early pregnancy loss. *Fertil. Steril.* **95,** 1813–5.e1 (2011).

172. Quintero-Ronderos, P. *et al.* THBD sequence variants potentially related to recurrent pregnancy loss. *Reprod. Biol. Endocrinol.* **15,** 92 (2017).

173. Li, Y.-H., Shi, G.-Y. & Wu, H.-L. The role of thrombomodulin in atherosclerosis: from bench to bedside. *Cardiovasc. Hematol. Agents Med. Chem.* **4,** 183–7 (2006).

174. Parodi, A., Cummings, R. D. & Aebi, M. in *Essentials of Glycobiology* **3rd Editio,** (2017).

175. Suzuki, K. *et al.* Structure and expression of human thrombomodulin, a thrombin receptor on endothelium acting as a cofactor for protein C activation. *EMBO J.* **6,** 1891–7 (1987).

176. Ito, T. *et al.* Proteolytic Cleavage of High Mobility Group Box 1 Protein by Thrombin-Thrombomodulin Complexes. *Arterioscler. Thromb. Vasc. Biol.* **28,** 1825–1830 (2008).

177. Abeyama, K. *et al.* The N-terminal domain of thrombomodulin sequesters high-mobility group-B1 protein, a novel antiinflammatory mechanism. *J. Clin. Invest.* **115,** 1267–1274 (2005).

178. Rossini, A. *et al.* HMGB1-stimulated human primary cardiac fibroblasts exert a paracrine action on human and murine cardiac stem cells. *J. Mol. Cell. Cardiol.* **44,** 683–693 (2008).

179. Yoshihara-Hirata, C. *et al.* Anti-HMGB1 Neutralizing Antibody Attenuates Periodontal Inflammation and Bone Resorption in a Murine Periodontitis Model. *Infect. Immun.* **86,** (2018).

180. Alsousi, A. A. & Igwe, O. J. Redox-active trace metal-induced release of high mobility group box 1(HMGB1) and inflammatory cytokines in fibroblast-like synovial cells is Toll-like receptor 4 (TLR4) dependent. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864,** 3847–3858 (2018).

181. Shirasuna, K. *et al.* AGEs and HMGB1 Increase Inflammatory Cytokine Production from Human Placental Cells, Resulting in an Enhancement of Monocyte Migration. *Am. J. Reprod. Immunol.* **75,** 557–568 (2016).

182. Amin, A. R. & Islam, A. B. M. M. K. Genomic Analysis and Differential Expression of HMG and S100A Family in Human Arthritis: Upregulated Expression of Chemokines, IL-8 and Nitric Oxide by HMGB1. *DNA Cell Biol.* **33,** 550–565 (2014).

183. Huang, Q. T. *et al.* Advanced glycation end products as an upstream molecule triggers ROS-induced sFlt-1 production in extravillous trophoblasts: A novel bridge between oxidative stress and preeclampsia. *Placenta* **34,** 1177–1182 (2013).

184. Min, H. J. *et al.* ROS-dependent HMGB1 secretion upregulates IL-8 in upper airway epithelial cells under hypoxic condition. *Mucosal Immunol.* **10,** 685–694 (2017).

185. Sagheddu, R. *et al.* Targeting RAGE as a potential therapeutic approach to Duchenne muscular dystrophy. *Hum. Mol. Genet.* **27,** 3734–3746 (2018).

186. Patel, V. *et al.* The stretch responsive microRNA miR-148a-3p is a novel repressor of IKBKB, NF- κ B signaling, and inflammatory gene expression in human aortic valve cells. *FASEB J.* **29,** 1859–1868 (2015).

187. Anjana, R. *et al.* Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Bioinformation* **8,** 1220–1224 (2012).

188. Bignucolo, O., Leung, H. T. A., Grzesiek, S. & Bernèche, S. Backbone Hydration Determines the Folding Signature of Amino Acid Residues. *J. Am. Chem. Soc.* **137,** 4300–4303 (2015).

189. Zhang, Z., Witham, S. & Alexov, E. On the role of electrostatics in protein–protein interactions. *Phys. Biol.* **8,** 035001 (2011).

190. Andersson, U. *et al.* High Mobility Group 1 Protein (Hmg-1) Stimulates Proinflammatory Cytokine Synthesis in Human Monocytes. *J. Exp. Med.* **192,** 565–570 (2000).

191. Bhutada, S. *et al.* High mobility group box 1 (HMGB1) protein in human uterine fluid and its relevance in implantation. *Hum. Reprod.* **29,** 763–780 (2014).

192. Riesewijk, A. *et al.* Gene expression profiling of human endometrial receptivity on days LH+2 versus LH+7 by microarray technology. *Mol. Hum. Reprod.* **9,** 253–64 (2003).

193. Kwak-Kim, J. Y. H., Gilman-Sachs, A. & Kim, C. E. in *Immunology of Gametes and Embryo Implantation* 64–79 (KARGER, 2005). doi:10.1159/000087821

194. Bates, M. D., Quenby, S., Takakuwa, K., Johnson, P. M. & Vince, G. S. Aberrant cytokine production by peripheral blood mononuclear cells in recurrent pregnancy loss? *Hum. Reprod.* **17,** 2439–44 (2002).

195. Tian, H., McKnight, S. L. & Russell, D. W. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. *Genes Dev.* **11,** 72–82 (1997).

196. Takeda, N. *et al.* Endothelial PAS domain protein 1 gene promotes angiogenesis through the transactivation of both vascular endothelial growth factor and its receptor, Flt-1. *Circ. Res.* **95,** 146–53 (2004).

197. Maemura, K. *et al.* Generation of a dominant-negative mutant of endothelial PAS domain protein 1 by deletion of a potent C-terminal transactivation domain. *J. Biol. Chem.* **274,** 31565–70 (1999).

198. Depoix, C. L. L., de Selliers, I., Hubinont, C. & Debieve, F. HIF1A and EPAS1 potentiate hypoxia-induced upregulation of inhibin alpha chain expression in human term cytotrophoblasts in vitro. *Mol. Hum. Reprod.* **23,** gax002 (2017).

199. L'Hôte, D. *et al.* Centimorgan-range one-step mapping of fertility traits using interspecific recombinant congenic mice. *Genetics* **176,** 1907–21 (2007).

200. Laissue, P., L'Hôte, D., Serres, C. & Vaiman, D. Mouse models for identifying genes modulating fertility parameters. *animal* **3,** 55 (2009).

201. Laissue, P. *et al.* Identification of Quantitative Trait Loci responsible for embryonic lethality in mice assessed by ultrasonography. *Int. J. Dev. Biol.* **53,** 623–629 (2009).

202. Vatin, M. *et al.* Refined Mapping of a Quantitative Trait Locus on Chromosome 1 Responsible for Mouse Embryonic Death. *PLoS One* **7,** e43356 (2012).

203. Vatin, M. *et al.* Polymorphisms of human placental alkaline phosphatase are associated with in vitro fertilization success and recurrent pregnancy loss. *Am. J. Pathol.* **184,** 362–368 (2014).

204. L'Hôte, D. *et al.* Interspecific resources: a major tool for quantitative trait locus cloning and speciation research. *BioEssays* **32,** 132–42 (2010).

205. Guénet, J. L. & Bonhomme, F. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* **19,** 24–31 (2003).

206. Burgio, G. *et al.* Interspecific Recombinant Congenic Strains Between C57BL/6 and Mice of the Mus spretus Species: A Powerful Tool to Dissect Genetic Control of Complex Traits. *Genetics* **177,** 2321–33 (2007).

207. Quintero-Ronderos, P. & Laissue, P. The multisystemic functions of FOXD1 in development and disease. *J. Mol. Med.* **96,** 725–739 (2018).

208. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7,** 1728–1740 (2012).

209. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).

210. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

211. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

212. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).

213. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26,** 841–842 (2010).

214. He, Q., Johnston, J. & Zeitlinger, J. ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat. Biotechnol.* **33,** 395–401 (2015).

215. Welch, R. P. *et al.* Chipenrich: Gene Set Enrichment For ChIP-seq Peak Data. (2017). at <https://bioconductor.org/packages/devel/bioc/vignettes/chipenrich/inst/doc/chipenrich-vignette.html>

216. Liu, T. MACS2 pileup value in xls file. (2014). at <https://groups.google.com/d/msg/macs-announcement/c__KCotHsok/Wp1VdX3XgAUJ>

217. Elemento, O. & Giannopoulou, E. G. ChIP sequencing analysis in R/Bioconductor. (2014). at

<http://physiology.med.cornell.edu/faculty/elemento/lab/data/courses/2014/CSHLseq/ChIP-seq.pdf>

218. Sahu, P. Biostars: Bioinformatics explained. (2015). at <https://www.biostars.org/p/198970/#210571>

219. Chow, W.-N. *et al.* Complement 3 deficiency impairs early pregnancy in mice. *Mol. Reprod. Dev.* **76,** 647–655 (2009).

220. Pham, C. T. & Ley, T. J. Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8627–32 (1999).

221. Adkison, A. M., Raptis, S. Z., Kelley, D. G. & Pham, C. T. N. Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis. *J. Clin. Invest.* **109,** 363–371 (2002).

222. Méthot, N. *et al.* Inhibition of the Activation of Multiple Serine Proteases with a Cathepsin C Inhibitor Requires Sustained Exposure to Prevent Pro-enzyme Processing. *J. Biol. Chem.* **282,** 20836–20846 (2007).

223. Meade, J. L. *et al.* Proteolytic Activation of the Cytotoxic Phenotype during Human NK Cell Development. *J. Immunol.* **183,** 803–813 (2009).

224. Menkhorst, E. M. *et al.* Decidual-Secreted Factors Alter Invasive Trophoblast Membrane and Secreted Proteins Implying a Role for Decidual Cell Regulation of Placentation. *PLoS One* **7,** e31418 (2012).

225. Zhu, X.-Y. *et al.* Blockade of CD86 Signaling Facilitates a Th2 Bias at the Maternal-Fetal Interface and Expands Peripheral CD4+CD25+ Regulatory T Cells to Rescue Abortion-Prone Fetuses1. *Biol. Reprod.* **72,** 338–345 (2005).

226. Wang, J. *et al.* Vascular endothelial growth factor affects dendritic cell activity in hypertensive disorders of pregnancy. *Mol. Med. Rep.* **12,** 3781–3786 (2015).

227. Whitley, G. S. J. & Cartwright, J. E. Trophoblast-mediated spiral artery remodelling: a role for apoptosis. *J. Anat.* **215,** 21–26 (2009).

228. Khong, T. Y., De Wolf, F., Robertson, W. B. & Brosens, I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *Br. J. Obstet. Gynaecol.* **93,** 1049–59 (1986).

229. Urata, H., Kinoshita, A., Misono, K. S., Bumpus, F. M. & Husain, A. Identification of a highly specific chymase as the major angiotensin II-forming enzyme in the human heart. *J. Biol. Chem.* **265,** 22348–57 (1990).

230. Tchougounova, E. *et al.* A Key Role for Mast Cell Chymase in the Activation of Pro-matrix Metalloprotease-9 and Pro-matrix Metalloprotease-2. *J. Biol. Chem.* **280,** 9291–9296 (2005).

231. Marx, L. *et al.* Decidual mast cells might be involved in the onset of human first-trimester abortion. *Am. J. Reprod. Immunol.* **41,** 34–40 (1999).

232. Garfield, R. E., Irani, A.-M., Schwartz, L. B., Bytautiene, E. & Romero, R. Structural and functional comparison of mast cells in the pregnant versus nonpregnant human uterus. *Am. J. Obstet. Gynecol.* **194,** 261–267 (2006).

233. Meyer, N. *et al.* Chymase-producing cells of the innate immune system are required

for decidual vascular remodeling and fetal growth. *Sci. Rep.* **7,** 45106 (2017).

234. Clarson, L. H., Roberts, V. H. J., Hamark, B., Elliott, A. C. & Powell, T. Store-operated Ca 2+ entry in first trimester and term human placenta. *J. Physiol.* **550,** 515–528 (2003).

235. Jacobs, B. E., Liu, Y., Pulina, M. V., Golovina, V. A. & Hamlyn, J. M. Normal pregnancy: mechanisms underlying the paradox of a ouabain-resistant state with elevated endogenous ouabain, suppressed arterial sodium calcium exchange, and low blood pressure. *Am. J. Physiol. Circ. Physiol.* **302,** H1317–H1329 (2012).

236. Dykes, I. M. & Emanueli, C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. *Genomics. Proteomics Bioinformatics* **15,** 177–186 (2017).

237. Herriges, M. J. *et al.* Long noncoding RNAs are spatially correlated with transcription factors and regulate lung development. *Genes Dev.* **28,** 1363–1379 (2014).