



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Análisis de la variabilidad y expresión génica en el parto prematuro espontáneo

Mag. Silvana María Pereyra Lepre

Directora: Dra. Rossana Sapiro

Co-director: Dr. Bernardo Bertoni

PROGRAMA DE DESARROLLO DE CIENCIAS BÁSICAS
(PEDECIBA) – ÁREA BIOLOGÍA
DOCTORADO EN CIENCIAS BIOLÓGICAS

Montevideo, Uruguay
28 de Junio de 2019



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Análisis de la variabilidad y expresión génica en el parto prematuro espontáneo

Mag. Silvana María Pereyra Lepre

Tesis de Doctorado presentada al
Programa de postgrado PEDECIBA
Biología, como parte de los requisitos
necesarios para la obtención del título de
Doctor en Ciencias Biológicas.

Directora: Dra. Rossana Sapiro

Co-director: Dr. Bernardo Berton

PROGRAMA DE DESARROLLO DE CIENCIAS BÁSICAS
(PEDECIBA) – ÁREA BIOLOGÍA
DOCTORADO EN CIENCIAS BIOLÓGICAS

Montevideo, Uruguay
28 de Junio de 2019

Autor:

Mag. Silvana María Pereyra Lepre
*Departamento de Genética,
Facultad de Medicina, UdelaR.*

Directora:

Dra. Rossana Sapiro
*Departamento de Histología y Embriología,
Facultad de Medicina, UdelaR.*

Co-director:

Dr. Bernardo Bertoni
*Departamento de Genética,
Facultad de Medicina, UdelaR.*

Tribunal:

Dr. José Tort
*Departamento de Genética,
Facultad de Medicina, UdelaR.*

Dr. Alfonso Cayota
*Departamento Básico de Medicina,
Facultad de Medicina, UdelaR.*

Dra. Carolina Bonilla
*Departamento de Medicina Preventiva,
Faculdade de Medicina, Universidade de São Paulo.*

AGRADECIMIENTOS

A Rossana Sapiro, que con energía incansable fue guía y compañera en todas las etapas de este trabajo de tesis, desde la colecta de muestras en el hospital hasta las discusiones con los revisores. Por las enseñanzas académicas y de las otras.

A Bernardo Bertoni, por apoyar de forma constante todo el desarrollo de la tesis y por confiar en que podía llevarla a cabo. Por ser mentor y guiarme pacientemente en el mundo de la genética humana.

A los Dres. Alfonso Cayota, José Tort y Carolina Bonilla, miembros del tribunal de evaluación, por aceptar evaluar esta tesis y por los valiosos aportes brindados.

Quiero agradecer enormemente a todas las madres que aceptaron participar en los estudios realizados en esta tesis, participando de forma desinteresada en este proyecto.

A los Dres. Justo Alonso y Daniel Borbonet, que nos abrieron las puertas de las clínicas que dirigen: Clínica Ginecotocológica B y Clínica de Neonatología del Centro Hospitalario Pereira Rossell, respectivamente. La colecta de muestras fue realizada con la colaboración invaluable de sus médicos, y en especial por los Dres. Claudio Sosa, Manuela DeMaria, Elena Lepre, Elsa Arocena, Agustina Bado, Patricia Cardozo, Grazzia Rey, Diego Greif, Daniel Grasso, Ariel Diaz, Natalia Cabrera y Pamela Grimaldi. A todos ellos, muchas gracias.

Al PEDECIBA, SNI (ANII) y CSIC, que financiaron parcialmente esta investigación, la última mediante un proyecto de iniciación a la investigación. Este apoyo fue fundamental para el desarrollo de la tesis.

A todos los integrantes del Departamento de Histología y Embriología, especialmente a Jimena Mas de Ayala y a Mariana Ford, que acompañaron muchas horas de mesada, aprendiendo juntas, haciendo mucho más entretenida la tarea, además de ayudar a solucionar varios problemas de laboratorio en el camino. A

Frances Evans, que esterilizó repetidamente los morteros necesarios para las extracciones de ARN, siempre con una disposición excelente.

A Claudia Chiale y Mercedes Rodríguez-Teja, que proveyeron oligonucleótidos para uno de los genes de referencia usados en la cuantificación de ARN.

A todos los compañeros del Departamento de Genética, muy especialmente a los isleños Mónica, Valentina, Lucía B., Lucía F., Patricia, Soledad, Ismael, por el apoyo y compañerismo de todos los días. Un placer trabajar con un grupo tan humano. Gracias especiales a Lucía B., que leyó la tesis de principio a fin, buscando errores aquí y allá.

El apoyo de los Dres. Titus Brown (UC Davis, EEUU) y Rosario Uriarte (AEPSM) fue fundamental para que pudiera asistir al curso *Next-Gen Sequence Analysis Workshop* (Michigan, EEUU), que fue fundamental para realizar los análisis bioinformáticos de mi tesis.

A mis padres, por el apoyo incondicional siempre. A mis hermanos, otros constantes, por el compañerismo diario.

A Matías y a Martina, por la vida compartida, por la alegría de todos los días, por la contención y porque esta tesis también es trabajo de ellos.

A todos ellos, muchas gracias.

TABLA DE CONTENIDOS

AGRADECIMIENTOS	IV
TABLA DE CONTENIDOS.....	VI
ABREVIATURAS	1
LISTA DE TABLAS	3
LISTA DE FIGURAS	4
RESUMEN	7
ABSTRACT	9
1. INTRODUCCIÓN GENERAL	11
1.1 PARTO PRETÉRMINO; UNA ENFERMEDAD GENÉTICA COMPLEJA	11
1.1.1. Clasificación del parto prematuro	12
1.1.2. Fisiopatología del PPT	13
1.1.3. Contribución genética al PPT	16
1.1.4. Contribución de los genomas materno y fetal	17
1.2 PARTO PRETÉRMINO ESPONTÁNEO Y EL ROL DE LA INFLAMACIÓN.....	17
1.3 ESTRATEGIAS PARA EL ANÁLISIS GENÉTICO DE ENFERMEDADES COMPLEJAS	20
1.3.1. Estudios de variantes génicas en PPT	24
1.4 EFECTO DE LAS VARIANTES GÉNICAS EN LA EXPRESIÓN GÉNICA	26
1.4.1. Expresión génica en parto a término y parto prematuro.....	27
1.5 APORTES DE ESTE TRABAJO DE TESIS	28
2. OBJETIVOS Y ESTRATEGIA DE TRABAJO	29
2.1 HIPÓTESIS DE TRABAJO.....	30
2.2 OBJETIVOS	30
2.2.1. Objetivo general	30
2.2.2. Objetivos específicos.....	30
2.3 ESTRATEGIA DE TRABAJO	31
3. INTERACCIONES ENTRE FACTORES AMBIENTALES Y VARIANTES GENÉTICAS EN PARTO PRETÉRMINO.....	32
3.1 ANTECEDENTES Y ESTRATEGIA	33

3.2 METODOLOGÍA	34
3.2.1. Muestras usadas en este estudio	34
3.2.2. Colecta de las muestras, extracción de ADN genómico y PCR en tiempo real	35
3.2.3. Análisis estadísticos.....	36
3.2.4. Aspectos bioéticos	37
3.3 RESULTADOS.....	37
3.3.1. Características de la muestra estudiada	37
3.3.2. Análisis de asociación génica	40
3.4 DISCUSIÓN	44
4. EL TRANSCRIPTOMA DE MEMBRANAS FETALES REVELA VÍAS INVOLUCRADAS EN EL PARTO PRETÉRMINO	48
4.1 ANTECEDENTES Y ESTRATEGIA	49
4.2 METODOLOGÍA	50
4.2.1. Selección de las muestras	50
4.2.2. Aspectos bioéticos	51
4.2.3. Colecta de la muestra.....	51
4.2.4. Extracción de ARN total	52
4.2.5. Secuenciación masiva de ARN	53
4.2.6. Limpieza y filtrado de las secuencias obtenidas.....	54
4.2.7. Alineamiento de las secuencias al genoma de referencia y conteo de lecturas.....	54
4.2.8. Expresión diferencial.....	56
4.2.9. Análisis de términos de ontología génica	59
4.2.10. Validación de resultados de ARN-seq por PCR en tiempo real...	60
4.3 RESULTADOS	62
4.3.1. Colecta de muestras y secuenciado de ARN	62
4.3.2. Análisis de expresión diferencial.....	67
4.3.3. Análisis de Ontología génica.....	76
4.3.4. Validación de la expresión génica por PCR en tiempo real cuantitativa	79
4.3.5. Validación de los resultados de ARN-seq comparándolos con datos de estudios previos de transcriptomas de prematuros	82
4.4 DISCUSIÓN	83
5. IDENTIFICACIÓN DE VARIANTES GÉNICAS CANDIDATAS A CONTRIBUIR AL PARTO PRETÉRMINO.....	93
5.1 ANTECEDENTES Y ESTRATEGIA	94
5.2 METODOLOGÍA	95

5.2.1. Mapeo a genoma de referencia.....	95
5.2.2. Asignación, filtrado y anotación de SNVs	96
5.2.3. Análisis de variantes génicas ubicadas en los genes diferencialmente expresados	98
5.2.4. Detección de variantes que afectan sitios de unión de miARN.....	99
5.3 RESULTADOS.....	100
5.3.1. Variantes génicas encontradas en secuenciado de ARN	100
5.3.2. Variantes novedosas y raras encontradas en genes DE	101
5.3.3. Comparación de las variantes encontradas en genes DE en bases de datos genómicas.....	102
5.3.4. Diferencias en el conteo de genotipos de genes DE entre casos y controles	104
5.3.5. Efectos funcionales y sobre sitios de unión de miARN de variantes de los 9 genes.....	109
5.3.6. Reconstrucción de haplotipos para los genes con diferencias genotípicas entre casos y controles.....	109
5.4 DISCUSIÓN	115
5.4.1. Variantes novedosas	116
5.4.2. Presencia de variantes génicas raras.....	117
5.4.3. Genes con diferencias genotípicas entre casos y controles	119
5.4.4. Efectos funcionales de las variantes en los 9 genes	122
5.4.5. Uso del ARN-seq para identificar variantes génicas	123
6. CONCLUSIONES GENERALES	125
6.1 PERSPECTIVAS DE TRABAJO	127
7. REFERENCIAS.....	130
8. APÉNDICES	159
8.1 CUESTIONARIO DEL SISTEMA INFORMÁTICO PERINATAL.....	160
8.2 TABLAS SUPLEMENTARIAS.....	162
8.2.1. Tabla suplementaria 1.....	162
8.2.2. Tabla suplementaria 2.....	163
8.2.3. Tabla suplementaria 3.....	168
8.3 ARTÍCULOS PUBLICADOS:	174

ABREVIATURAS

ADN	Ácido desoxiribonucleico
ADNc	ADN copia
ARN	Ácido ribonucleico
ARN-Seq	Secuenciado masivo de ARN
ARNm	ARN mensajero
ARNnc	ARN no codificante
BH	Benjamini y Hochberg
CHPR	Centro Hospitalario Pereira Rossell
CV ²	Cuadrado del coeficiente de variación
DE	Diferencialmente expresados
DEPC	Dietil pirocarbonato
DP	Profundidad de lectura
EG	Edad gestacional
EHW	Equilibrio Hardy-Weinberg
eQTL	Loci de expresión de características cuantitativas
FDR	Tasa de descubrimientos falsos
FPKM	Fragmentos Por Kilobase de exón por cada Millón de fragmentos mapeados
GB	Gigabases
GO	Ontología de genes
GWAS	Estudio de asociación a escala genómica
HGNC	Comité de nomenclatura génica de la Organización del Genoma Humano
HIV	Virus de inmunodeficiencia humana

HRM	Desnaturalización en alta resolución
HSC	Células madre hematopoyéticas
IC	Intervalos de confianza
IL	Interleuquina
Kb	Kilobases
KEGG	Enciclopedia de Genes y Genomas Kyoto
log2FC	Logaritmo en base 2 de la tasa de cambio
LRT	Test de cociente de verosimilitud
MAF	Frecuencia del alelo alternativo
MDS	Análisis de escalamiento multidimensional
miARN	microARN
OR	Razón de probabilidades u "odds ratio"
PCA	Análisis de componentes principales
PPT	Parto pretérmino
qRT-PCR	PCR en tiempo real cuantitativa
RCIU	Retraso en el crecimiento intrauterino
RIN	Números de Integridad de ARN
rlog	Logaritmo regularizado
RPKM	Lecturas Por Kilobase de exón por cada Millón de fragmentos mapeados
RPM	Rotura prematura de membranas
SE	Error estándar
SNP	Polimorfismo de nucleótido simple
SNV	Variante de nucleótido simple
TLR	Receptor de tipo Toll
UTR	Región no traducida

LISTA DE TABLAS

Tabla 3.1. Características generales de los SNPs estudiados.	36
Tabla 3.2. Características sociodemográficas y clínicas de madres y sus recién nacidos de la población de estudio.	38
Tabla 3.3. Conteos de genotipos de los SNPs analizados para madres y recién nacidos de la muestra. Se presentan las frecuencias genotípicas entre paréntesis.	39
Tabla 3.4. Análisis de regresión logística entre casos y controles para el polimorfismo materno rs16944 (<i>IL1β</i>) y el de recién nacidos rs1800795 (<i>IL6</i>) ajustado por el otro polimorfismo, así como su interacción.	41
Tabla 3.5. Estadísticos de tests exactos de Hardy-Weinberg y tests de chi-cuadrado para los genotipos de madres y recién nacidos analizados.	42
Tabla 3.6. Análisis de regresión logística de los genotipos maternos ajustados por madres que estuvieron expuestas a humo de tabaco durante el embarazo, y el correspondiente término de interacción con exposición al humo de tabaco.	43
Tabla 4.1. Genes DE seleccionados para la validación de expresión génica diferencial.	60
Tabla 4.2. Oligonucleótidos utilizados para las reacciones de qRT-PCR.	62
Tabla 4.3. Características demográficas y clínicas de los grupos término y pretérmino severo.	64
Tabla 4.4. Características clínicas y demográficas de los 4 controles (C01, C02, C03 y C04) y 4 pacientes pretérminos severos (P01, P02, P03 y P04) analizados por ARN-seq.	65
Tabla 4.5. Estadísticos del alineamiento de secuencias contra el genoma de referencia.	66
Tabla 4.6. Genes DE simultáneamente encontrados en Bukowski et al. (2017) y en este estudio.	83
Tabla 5.1. Conteo y porcentaje de variantes génicas detectadas según el contexto genómico en el que se encuentran. Ordenada según el porcentaje de variantes en total de ARN-seq.	101
Tabla 5.2. SNPs en genes DE que están reportados como eQTLs en útero por GTEx v7.	103
Tabla 5.3. Genes con diferencias significativas en la distribución de los genotipos de sus variantes para casos y controles.	105
Tabla 5.4. Frecuencia de haplotipos en poblaciones de 1000 genomas y las muestras de este estudio para los SNVs analizados en los 9 genes con diferencias significativas en la distribución de sus genotipos entre casos y controles.	111

LISTA DE FIGURAS

Figura 1.1. El desencadenamiento del trabajo de parto. Se caracteriza por la activación miometrial a un estado contráctil, la maduración del cérvix uterino (por cambios en las proteínas de la matriz extracelular (E), así como en la barrera epitelial e inmune) y por la activación de la decidua (D) y membranas fetales (C, A) (que facilita la separación de las membranas corioamnióticas y placenta del útero). Adaptado de Romero et al. (2014).	14
Figura 1.2. Posibilidad de identificar variantes génicas por su frecuencia alélica de riesgo y tamaño del efecto genético (“ <i>odds ratio</i> ”). Adaptada de Manolio et al. (2009).	22
Figura 3.1. Distribución de los genotipos combinados de las díadas madre-recién nacido para el genotipo materno IL1 β (rs16944) y el del recién nacido IL6 (rs1800795). Sólo la combinación de los genotipos IL6 G/G en madres y IL6 G/C en recién nacidos tuvo diferencias entre casos y controles ($p < 0,05$; test de χ^2). 43	
Figura 4.1. Flujo de trabajo de análisis informático de expresión diferencial. ...	55
Figura 4.2. Porcentaje de las lecturas alineadas distribuidas según su localización genómica para las muestras secuenciadas.	66
Figura 4.3. (A) Gráfico de análisis de componentes principales (PCA), señalando prematuridad y RPM. (B) Dendrograma de las muestras a partir de las distancias euclidianas de los conteos normalizados por <i>rlog</i>	68
Figura 4.4. Los 15 genes más significativamente sobre- (A) y sub- (B) expresados en parto pretérmino severo según su valor p según DESeq2 ajustado por BH. ...	69
Figura 4.5. Gráfico de volcán con los genes analizados en el DESeq2, mostrando la selección de genes expresados diferencialmente de forma significativa. Los puntos arriba de la línea roja muestran los genes significativos (valor $p < 0,05$) y las líneas azules muestran valores de log2FC de 2 y -2.	69
Figura 4.6. Visualización de los datos según la tasa de cambio y el promedio de los conteos de lecturas. Los puntos en rojo representan genes con $p < 0,05$	70
Figura 4.7. Histograma de los valores de significancia obtenidos para los análisis de expresión diferencial con DESeq2.	71
Figura 4.8. Gráfico de escalamiento multidimensional. Las muestras del grupo control están en azul y las de pretérmino severo en rojo.	71
Figura 4.9. Cuadrado del coeficiente de variación (CV^2) entre genes (A) y entre isoformas (B), en función del logaritmo en base 10 de FPKM.	73
Figura 4.10. Gráfico de Venn con la intersección de los genes reportados como DE por los métodos DESeq2, edgeR y Cufflinks2.	73

Figura 4.11. Heatmap de valores normalizados (*rlog*) de los niveles de expresión y agrupamiento jerárquico para los 270 genes DE (FDR < 0,05 y una tasa absoluta de cambio logarítmica > 2). La condición de cada muestra se denota en la parte superior de la figura. Los valores de *rlog* están codificados en una escala roja (alta expresión) a verde (baja expresión). En el dendrograma de genes se denotan con colores los 3 grupos mayores de acuerdo a los patrones de expresión de los genes. 75

Figura 4.12. Heatmap de valores de expresión génica normalizados por *rlog* y agrupamiento jerárquico del grupo de 18 genes DE que discriminan controles de parto pretérmino, para las 8 muestras. Se muestran los códigos de los genes en la derecha del gráfico. Se usa la misma escala roja a verde de la Figura 4.11 para denotar el nivel de expresión génica. 76

Figura 4.13. Términos de ontología génica enriquecidos entre los genes DE para las categorías Procesos Biológicos (**A**), Componentes Celulares (**B**) y Función Molecular (**C**). Se muestran los 10 términos más significativamente enriquecidos, con un valor p ajustado (BH) < 0,05. El eje x representa el valor de enriquecimiento, como el logaritmo de los valores de p ajustados por BH. El número de genes DE identificados para cada término se muestra entre paréntesis. 78

Figura 4.14. Vías KEGG enriquecidas entre los genes DE. Se muestran las 10 vías más significativamente enriquecidas, con un valor p ajustado (BH) < 0,05. El eje x representa el valor de enriquecimiento, como el logaritmo de los valores de p ajustados por BH. El número de genes DE identificados para cada vía se muestra entre paréntesis. 79

Figura 4.15. Diagrama KEGG de la vía de señalización de las quimioquinas. Los genes sobreexpresados en las membranas fetales de partos prematuros se muestran en rojo, y los subexpresados en verde. 80

Figura 4.16. Validación de los niveles de expresión génica. Tasas de cambio de seis genes DE medidos por ARN-seq (rojo) versus qRT-PCR (azul). Se muestran los valores p de significancia para las tasas de cambio de las qRT-PCR. *p<0,05; **p<0,01; ***p<0,001. ns: no significativo (valor p ajustado > 0,05). 81

Figura 5.1. Flujo de trabajo usado para el procesamiento de ARN-seq para la identificación de variantes génicas 96

Figura 5.2. Frecuencia del alelo de referencia de las variantes génicas encontradas en cada uno de los 9 genes que presentan diferencias significativas entre la distribución de sus genotipos entre casos y controles. La lista de variantes por gen se presenta en la Tabla suplementaria 3. Se presentan las frecuencias alélicas de nuestros datos y la de tres poblaciones de 1000 genomas, como referencia. 106

Figura 5.3. Análisis de componentes principales de los genotipos de SNPs seleccionados para 9 genes diferencialmente expresados entre controles y

prematuros severos. AFR: africanos. AMR: latinoamericanos. EAS: asiáticos.
EUR: europeos..... 115

RESUMEN

El parto pretérmino (PPT) es un problema de salud mundial, y la principal causa de mortalidad y morbilidad neonatal. La comprensión de los mecanismos patológicos que desencadenan el PPT es limitado por su naturaleza compleja y multicausal, donde factores genéticos, poblacionales y ambientales interaccionan determinando el fenotipo patológico (PPT), definido como el desencadenamiento de un parto antes de las 37 semanas de edad gestacional. La vía inflamatoria es una de las causas del PPT más estudiadas. El objetivo principal de esta tesis es identificar el contexto genético en el que se produce el PPT en nuestra población utilizando herramientas genómicas y transcriptómicas que incluyan factores tanto maternos como fetales. Para lograr este objetivo, se desarrollaron tres estrategias. Primero realizamos un estudio de genes candidatos inflamatorios para identificar factores genéticos que contribuyen al PPT. Encontramos evidencia de interacción entre polimorfismos del genoma materno y fetal, así como de interacción entre polimorfismos de genoma materno cuando se controla por variables ambientales.

En segundo lugar, examinamos los perfiles de expresión génica en membranas fetales de recién nacidos pretérminos severos (<34 semanas de edad gestacional) y nacidos a término, usando secuenciación masiva de ARN. Encontramos 270 genes diferencialmente expresados entre ambos grupos: 252 sobreexpresados y 18 subexpresados. Nuestros resultados confirman la importancia de vías inflamatorias en el desencadenamiento del PPT y sugieren que una activación del inflamasoma puede estar implicada en la inducción del mismo. También se detectaron vías sobreexpresadas no vinculadas previamente al PPT, como la hematopoyética. Por otro lado, el grupo de los 18 genes subexpresados potencialmente caracteriza el patrón de expresión del PPT severo.

Por último, usamos los datos transcriptómicos para estudiar variantes génicas vinculadas al PPT, centrándonos en el análisis de los genes diferencialmente expresados encontrados en la sección anterior. Encontramos 125 variantes

novedosas, entre las cuales una presenta un cambio de sentido con efecto deletéreo putativo. También encontramos 9 genes con diferencias significativas en el conteo de genotipos entre casos y controles, todos involucrados en mecanismos de inflamación y respuestas inmunes. Encontramos variantes con potencial para controlar la expresión génica, como variantes en sitios blanco de micro ARNs.

En suma, esta tesis comprueba la utilidad de una aproximación múltiple al estudio de las enfermedades de base compleja, aportando nuevos datos genéticos y transcriptómicos al estudio de la prematurez. Los datos surgidos de este análisis contribuyen a generar un perfil genotípico y de expresión génica de la enfermedad, estableciendo los factores de riesgo potenciales de PPT específicos en nuestra población.

Palabras clave: Parto prematuro, transcriptómica, variantes de nucleótido simple, expresión génica.

ABSTRACT

Preterm birth (PTB) is a global health problem, and the leading cause of neonatal mortality and morbidity. The understanding of the pathological mechanisms that trigger PTB is limited by its complex and multicausal nature, where genetic, population and environmental factors interact determining the pathological phenotype (PTB) which is the birth before 37 weeks of gestational age. The inflammatory pathway is one of the most studied causes of PTB. The main objective of this thesis is to identify the genetic context in which PTB occurs in our population using genomic and transcriptomic tools that include both maternal and fetal factors. To achieve this goal, three strategies were developed. First, we conducted a study of inflammatory candidate genes to identify genetic factors that contribute to PTB. We found evidence of interaction between maternal and fetal genome polymorphisms, as well as interaction between maternal genetic polymorphisms when controlled by environmental variables.

Secondly, we examined the gene expression profiles in fetal membranes of severe preterm newborns (<34 weeks of gestational age) and term infants, using RNA sequencing. We found 270 genes differentially expressed between both groups: 252 overexpressed and 18 underexpressed. Our results confirm the role of inflammatory pathways in PTB induction, and suggest that an activation of the inflammasome may be involved. It also detects overexpressed pathways not previously linked to PTB, such as the hematopoietic pathway. On the other hand, the group of 18 underexpressed genes potentially characterizes the expression pattern of severe PTB.

Finally, we use these transcriptomic data to study gene variants linked to PTB, focusing on the analysis of the differentially expressed genes found in the previous section. We find 125 novel variants, among which one presents a change of direction with putative deleterious effect. We also found 9 genes with significant differences in genotype counts between cases and controls, all involved in

mechanisms of inflammation and immune responses. We found variants with potential to control gene expression, such as variants in a target site of microRNAs.

In sum, this thesis proves the usefulness of a multiple approach to the study of complex diseases, providing new genetic and transcriptomic data to the study of prematurity. The data generated from this analysis contribute to generate a genotypic and gene expression profile of the disease, establishing the potential risk factors of specific PTB in our population.

Keywords: Preterm birth, transcriptomics, single nucleotide variants, gene expression.



INTRODUCCIÓN GENERAL

1.1 PARTO PRETÉRMINO; UNA ENFERMEDAD GENÉTICA COMPLEJA

El parto pretérmino (PPT) es definido como aquel que se produce antes de las 37 semanas de edad gestacional. Se trata de un problema de salud mundial, siendo la principal causa de muerte neonatal y de muerte infantil en menores de 5 años (Chawanpaiboon et al., 2019; Vogel et al., 2018). Desafortunadamente, y a pesar de los muchos esfuerzos realizados en investigación y en salud pública en el tema, la frecuencia del PPT ha ido aumentando en los últimos años (Chawanpaiboon et al., 2019; Martin et al., 2018). Las tasas de PPT ascienden a 10,6% (intervalo de confianza 8,7–11,9) a nivel global (Chawanpaiboon et al., 2019). En Latinoamérica, 9,8% de los recién nacidos lo hacen en forma prematura, y la incidencia es entre 10,2% y 12,2% en Uruguay entre 2014 y 2018 (MSP, 2019). Además de las complicaciones inmediatas o a corto plazo durante el primer mes de vida de los neonatos, el PPT tiene consecuencias de por vida para los recién nacidos prematuros, sus familias y la población general. Los efectos a largo plazo

más evidentes de los recién nacidos prematuros son deficiencias cognitivas (retraso en el desarrollo neurológico), físicas (retraso en desarrollo motor, enfermedades respiratorias) y sensoriales (problemas oftálmicos y auditivos) (Saigal & Doyle, 2008).

1.1.1. CLASIFICACIÓN DEL PARTO PREMATURO

El PPT se clasifica generalmente de acuerdo a su causa obstétrica: (a) el parto prematuro espontáneo y (b) el parto prematuro por indicación médica, ya sea mediante cesárea o inducción del parto. El PPT por indicación médica usualmente representa el 30% del PPT total (Eidem et al., 2015; Myatt et al., 2012), y puede ser indicado por condiciones maternas (como preeclampsia, eclampsia, placenta previa) o fetales (restricción del crecimiento intrauterino o sufrimiento fetal) (Goldenberg et al., 2008). Dentro del grupo de PPT espontáneo o idiopático, aproximadamente el 45% de los casos ocurren sin ruptura prematura de membranas (RPM), mientras que el 25% restante es consecuencia de la RPM a pretérmino (Eidem et al., 2015; Goldenberg et al., 2008; Moutquin, 2003; Vogel et al., 2018). La presentación de cada subtipo no es homogénea en su presentación clínica, indicando que el PPT es una patología heterogénea, lo que dificulta identificar factores causales.

El PPT también se clasifica de acuerdo a la edad gestacional (EG) en la que se desencadenan: neonatos nacidos entre las 24 y 33 semanas (pretérminos severos) presentan mayor riesgo de muerte y enfermedades que los pretérminos moderados (EG entre 34 y 36 semanas). Los PPT moderados constituyen un 60-70% de los PPT totales (Goldenberg et al., 2008). La mayor incidencia de complicaciones neonatales y casi de la mitad de las muertes perinatales se producen en los PPT severos, dando cuenta por lo tanto de la mayor parte de la morbilidad perinatal (Saigal & Doyle, 2008). Se plantea que PPT de diferentes grupos de EG tienen diversas causas y/o mecanismos patológicos (Moutquin, 2003).

Independientemente del tipo de PPT, los tratamientos actuales no logran prolongar el tiempo al nacimiento una vez que el trabajo de parto ha sido iniciado (Moutquin, 2003; Romero et al., 2014).

1.1.2. FISIOPATOLOGÍA DEL PPT

El PPT es un síndrome, causado por varios procesos patológicos (Romero et al., 2014). Si bien se conocen varios componentes de su fisiopatología, no hay certeza sobre los mecanismos biológicos que lo desencadenan. Esto se debe a que es un fenotipo molecular complejo con muchos procesos moleculares en el que interactúan varios tipos de tejidos gestacionales. Estos tejidos además provienen de dos individuos genéticamente diferentes (madre y feto). Los mismos incluyen decidua, miometrio, cérvix y sangre proveniente de la madre, y membranas fetales (corion y amnios), cordón umbilical y sangre provenientes del feto. El estudio del PPT también se complica por las dificultades en determinar clínicamente el fenotipo, dada la variabilidad de presentaciones del PPT. Igualmente, se conocen factores de predisposición de origen materno, fetal y placentario.

Una asunción tácita que se hace en los estudios de prematuridad es que el parto a término y pretérmino comparten un mismo mecanismo, conocido como la “vía común” del parto. Es decir, se asume que el PPT es simplemente un trabajo de parto que se inicia demasiado pronto, una visión que genera controversia y aún no está completamente establecida. Esta idea se basa en que el parto a término y el PPT implican eventos clínicos similares, que se consideran la vía común del parto: aumento de contractilidad uterina, dilatación del cérvix uterino y ruptura de membranas corioamnióticas (Romero et al., 2006).

Durante el embarazo la progesterona inhibe la contractilidad uterina, manteniendo su estado quiescente. El inicio del trabajo de parto está marcado por la activación miométrica, desencadenada por la activación del eje hipotálamo-hipófisis-

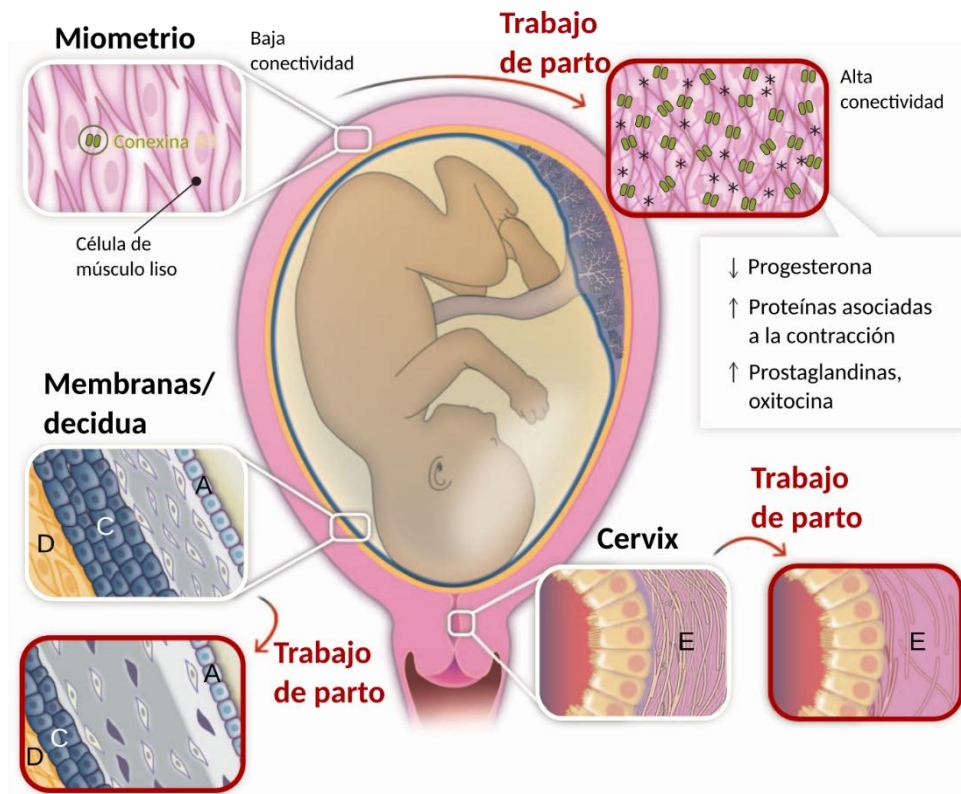


Figura 1.1. El desencadenamiento del trabajo de parto. Se caracteriza por la activación miométrial a un estado contráctil, la maduración del cérvix uterino (por cambios en las proteínas de la matriz extracelular (E), así como en la barrera epitelial e inmune) y por la activación de la decidua (D) y membranas fetales (C, A) (que facilita la separación de las membranas corioamnióticas y placenta del útero). Adaptado de Romero et al. (2014).

suprarrenal fetal, que a su vez está estimulado por la producción endógena de hormona liberadora de corticotropina placentaria (Behrman et al., 2007) (Figura 1.1). Esto resulta en una ausencia de progesterona y activación de estrógeno, que en consecuencia causa la activación y expresión de quimioquinas, citoquinas y proteínas asociadas a la contracción muscular: oxitocina y prostaglandinas. La activación miométrial está acompañada entonces de un cambio de un estado anti-inflamatorio a uno pro-inflamatorio (Romero et al., 2014). Esta cascada biológica desemboca en una vía común que involucra maduración de cervix uterino, contractibilidad uterina, activación de decidua y membranas fetales. La maduración cervical está mediada por cambios en proteínas de la matriz extracelular, así como en la barrera epitelial e inmune, que disminuyen las

tensiones en el cuello del útero para permitir su dilatación (Romero et al., 2014). La activación de la decidua y membranas refiere a la separación de las membranas corioamnióticas de la decidua y eventualmente a la ruptura de membranas. En este proceso están involucradas una mayor expresión de citoquinas inflamatorias y quimioquinas, mayor actividad de proteasas, disolución de adhesivos celulares y apoptosis (Menon & Fortunato, 2007).

La diferencia fundamental entre el parto a término y el PPT es que el primero es resultado de una activación fisiológica de esta cascada, mientras que en el PPT procesos patológicos activan uno o más componentes de la “vía común” del parto (Figura 1.1). Estos procesos patológicos no son mutuamente excluyentes, pudiendo actuar en conjunto. Así, el PPT es definido como un síndrome, iniciado por múltiples mecanismos diferentes (Goldenberg et al., 2008; Romero et al., 2006). Los mecanismos propuestos para el desencadenamiento del PPT incluyen la activación del eje hipotálamo-hipófisis-suprarrenal materno o fetal, hemorragia anteparto, factores mecánicos como sobredistensión uterina e incompetencia cervical, e inflamación e infección bacteriana (Lockwood & Kuczynski, 2001; Villar et al., 2012). Cada uno de estos factores puede estar afectado por factores genéticos y/o ambientales.

Debe considerarse que las causas del PPT pueden variar de acuerdo a la edad gestacional. Se plantea por ejemplo que los mecanismos de infección son responsables de desencadenar la mayoría de los PPT severos (Behrman et al., 2007). Posiblemente una de las dificultades en entender el síndrome ha sido la tendencia a estudiar los nacimientos PPT de forma agrupada, asumiendo erróneamente uniformidad en los mecanismos que lo desencadenan (Zhang et al., 2018).

1.1.3. CONTRIBUCIÓN GENÉTICA AL PPT

Estudios epidemiológicos identificaron factores de riesgo de PPT ambientales y constitucionales, como edad materna, infecciones maternas, bajo nivel socioeconómico, o hábitos de la madre como el consumo de alcohol, tabaco y drogas, entre otros. A pesar de que el parto prematuro es heterogéneo en su etiología (Wen et al., 2004; Berkowitz & Papiernik, 1993), evidencia epidemiológica indica que los factores genéticos juegan un papel significativo en la etiología del PPT espontáneo (Anum et al., 2009). Por otro lado, los factores genéticos contribuyen sustancialmente en la determinación de la duración de la gestación y en el riesgo de PPT, incluso cuando se controlan los factores socioeconómicos (Goldenberg et al., 2008).

Uno de los predictores más fuertes de riesgo de PPT es la existencia de antecedentes familiares de PPT o el antecedente de un parto previo pretérmino (Goldenberg et al., 2008; Winkvist et al., 1998). Aproximadamente el 20% de las madres que tienen un PPT tienen otro hijo nacido prematuramente (Bakketeig et al., 1979), lo que sugiere que los factores que son constantes a lo largo del tiempo, como los genéticos, afectan la edad gestacional. También se encuentra mayor concordancia entre la edad gestacional de hijos de gemelas monocigóticas que en hermanas di-cigóticas (Kistka et al., 2008). La heredabilidad del PPT se estima entre 17 a 40% (Clausson et al., 2000; Treloar et al., 2000; York et al., 2013). Se reporta además una asociación con consanguinidad (Mumtaz et al., 2010). En modelos animales se comprobó el rol de factores genéticos regulando la edad gestacional, obteniéndose cepas de ratones endogámicas que se diferencian en su tiempo de gestación (Murray et al., 2010).

Otro argumento que apoya la presencia de un factor genético son las diferencias significativas en la incidencia de PPT en diferentes grupos étnicos. Las mujeres afroamericanas tienen un riesgo casi dos veces mayor de parto prematuro al de las mujeres de ascendencia europea con características demográficas similares (Goldenberg, 2002). Estas diferencias son aún más marcadas cuando se considera

únicamente el parto prematuro severo (*i.e.* edad gestacional menor a 34 semanas), en cuyo caso el riesgo aumenta a casi 3 veces en mujeres afroamericanas en comparación con las europeas (Porter et al., 1997).

1.1.4. CONTRIBUCIÓN DE LOS GENOMAS MATERNO Y FETAL

Uno de los aspectos que agrega complejidad al estudio del PPT es que debe considerarse el efecto y la interacción de dos genomas: el materno y el fetal. La evidencia apoya la idea de que ambos genomas contribuyen al desencadenamiento del PPT (Zhang et al., 2017). Cuando se estima la contribución relativa de cada genoma, se calcula que los factores genéticos fetales contribuyen entre 11-35% al PPT mientras que la contribución del genoma materno sería de entre 13-20% (York et al., 2013; York et al., 2014). Esta contribución mixta dificulta la interpretación de los estudios de asociación génica, para lo cual es recomendado genotipar tanto a las madres como a los hijos.

Así, la susceptibilidad al PPT puede estar afectada tanto por el genoma materno y/o fetal, como también por interacciones gen-gen (epístasis) o gen-ambiente. En este sentido, el parto prematuro es considerado una enfermedad de base genética compleja.

1.2 PARTO PRETÉRMINO ESPONTÁNEO Y EL ROL DE LA INFLAMACIÓN

Existe evidencia abundante de que mediadores inflamatorios participan en los mecanismos del PPT. De los procesos que desencadenan el trabajo de parto, la infección y la inflamación uterina son factores etiológicos que explican casi el 40% de PPT (Romero et al., 2003). La infección intrauterina es el único proceso patológico con una relación causal establecida con la prematuridad, y para el cual hay una patofisiología molecular conocida (Romero et al., 2003). El proceso patológico inflamatorio puede ser causado por microorganismos invadiendo el

espacio amniótico (*i.e.* infección intra-amniótica) o por estrés o muerte celular, que libere señales de alarma/alarminas (*i.e.* inflamación intra-amniótica estéril). En ambos casos, las citoquinas son esenciales en la iniciación y regulación del PPT (Bowen et al., 2002). Por ejemplo, la Interleuquina (IL)-1 β es un mediador central en el desencadenamiento del PPT, ya que puede estimular la expresión y liberar otros mediadores de trabajo de parto, como las prostaglandinas (Brown et al., 1998). De hecho, la administración de IL1 β provoca PPT en ratones, que puede ser detenido con antagonistas de IL1 β (Romero & Tartakovsky, 1992), confirmando su papel en el PPT. También, frecuentemente la infección maternal (*e.g.* corioamnionitis) es seguida de una respuesta inflamatoria por parte del feto, con altos niveles de citoquinas proinflamatorias en su circulación (Vrachnis et al., 2010). Monocitos y neutrófilos infiltran el miometrio y cérvix durante el trabajo de parto a término y a pretérmino, lo que se asocia con un aumento significativo en la expresión de IL1 β , Interleuquina 6 (IL6) e Interleuquina 8 (IL8) (Osman et al., 2003). Las células del estroma de tejidos maternos y fetales también liberan citoquinas proinflamatorias. Las contracciones miométriales son estimuladas por las prostaglandinas, así como por el aumento de IL1 β y el factor de necrosis tumoral alfa (TNF α), que aumentan la entrada de calcio en las células musculares lisas del miometrio. Las citoquinas proinflamatorias estimulan la expresión de metaloproteinasas de la matriz, que promueven la degradación de la matriz extracelular y la remodelación cervical (Boyle et al., 2017a).

El mantenimiento del embarazo está fuertemente asociado a la regulación de los niveles de citoquinas proinflamatorias en la interfaz materno-fetal. Si bien se han descrito algunos roles fisiológicos para las mismas, en general se apoya que una producción excesiva o anormal en membranas maternas o fetales es nociva para la continuación del embarazo (Peltier, 2003).

Las citoquinas pueden ser usadas como marcadores de inflamación. Se han encontrado marcadores de inflamación en el líquido amniótico de embarazadas asintomáticas en el segundo trimestre asociados con parto prematuro, tales como

la metaloproteasa de la matriz 8 (Yoon et al., 2001) e IL-6 (Wenstrom et al., 1998) analizados en el líquido amniótico en amniocentesis del segundo trimestre. Esto sugiere que un proceso inflamatorio intra-amniótico crónico está asociado con el parto prematuro.

Un grupo importante de receptores involucrados en las respuestas inflamatorias relacionadas con la inmunidad innata son los receptores de tipo Toll (TLR). Los TLRs son una familia de receptores involucrados en el reconocimiento de patógenos y en la subsecuente activación de cascadas de señalización que inducen una respuesta inmune. Tienen un rol crucial en la respuesta inmune innata y en la iniciación de la inflamación (Noguchi et al., 2010). El TLR4 es uno de los más estudiados, cuyo ligando principal es el liposacárido bacteriano (Akira et al., 2006). La activación de TLR4 por el liposacárido bacteriano desencadena una cascada de señalización que involucra quinasas de serina/treonina que median la fosforilación y ubiquitinación de varios sustratos, resultando en la activación del factor transcripcional nuclear NF- κ B, el cual regula la expresión de genes de diversas citoquinas proinflamatorias (Kawai & Akira, 2006). En particular, la estimulación de TLRs con sus ligandos induce una liberación de citoquinas proinflamatorias en las células epiteliales del útero, membranas fetales y placenta, que pueden estar implicadas en la patogénesis de la contracción uterina y ruptura de membranas en el proceso del PPT (Patni et al., 2007; Schaefer et al., 2004).

El análisis de las contribuciones de estas moléculas inflamatorias al PPT es complejo debido en parte a las redundancias existentes en la red de las citoquinas (Yadav & Sarvetnick, 2003). A pesar de tener diferentes funciones biológicas, las citoquinas pueden compartir subunidades o receptores celulares entre sí, lo que puede atenuar el efecto patológico de un cambio en una de ellas. Por esto, el estudio de genes individuales puede subestimar sus efectos sobre el fenotipo, por lo que se recomienda realizar análisis genéticos de mayor escala que involucren varios actores de la vía involucrada.

1.3 ESTRATEGIAS PARA EL ANÁLISIS GENÉTICO DE ENFERMEDADES COMPLEJAS

La arquitectura genética describe el espectro de contribuciones genéticas que son responsables de la variabilidad fenotípica heredable. El entendimiento de las mismas ha evolucionado desde genes únicos influenciando desórdenes Mendelianos a genes múltiples (herencia poligénica) afectando características comunes complejas (Badano & Katsanis, 2002). A su vez, la interacción entre loci (epistasia) o entre genes y factores ambientales puede incidir sustancialmente en el fenotipo de características complejas, y deben ser considerados en estudios genéticos (Carlborg & Haley, 2004).

Los estudios de asociación a escala genómica (GWAS) buscan encontrar una asociación por ligamiento genético entre una variante común conocida y el gen involucrado en la enfermedad. Durante su apogeo esta técnica mostró que estas asociaciones pueden ocurrir porque estos marcadores están ligados al gen involucrado. Además, en algunos casos estas variantes pueden afectar la expresión génica (Emilsson et al., 2008). Las variantes de nucleótido simple (SNVs) son alteraciones genéticas en una única base en comparación con el contexto de la población en la que se presenta. Además de generar cambios en la estructura proteica y la función debido a mutaciones en las secuencias codificantes de transcritos, los SNVs pueden tener una variedad de efectos funcionales en la regulación y expresión génica. Las variantes de este último tipo se conocen como loci de expresión de características cuantitativas (eQTL). Hay evidencia de que estos polimorfismos pueden explicar variación en la expresión génica en las poblaciones (Cookson et al., 2009; Morley et al., 2004). Por ejemplo, variantes en regiones intrónicas pueden tener efectos funcionales en la expresión mediante la regulación del corte y empalme alternativo (Xiong et al., 2015). También pueden afectar la vida media del ARN mensajero (ARNm) y resultar en menores niveles de proteína mediante interacciones ARNm–microARN (miARN), ya sea por

disminución de niveles del transcripto o por represión transcripcional (Moszyńska et al., 2017).

Los GWAS se consideran como una herramienta eminente en la asociación de polimorfismos genéticos con la susceptibilidad a enfermedades y características complejas. Estos estudios se realizan bajo el supuesto de un modelo “variantes comunes-enfermedades comunes”, en el que las enfermedades complejas están gobernadas por un número moderado de variantes comunes, cada una de las cuales explica un porcentaje considerable del riesgo en una población (Lander, 1996). Sin embargo, se encontró que las variantes comunes asociadas a enfermedades tienen efectos muy pequeños en las mismas, no explicando la mayoría de la contribución genética a las enfermedades (Figura 1.2). Esto se conoce como el paradigma de la “heredabilidad perdida” (Maher, 2008; Manolio et al., 2009). Las variantes raras no son suficientemente frecuentes para ser capturadas por los GWAS, ni tampoco tienen efectos suficientemente grandes como para ser detectados en análisis de ligamiento clásicos en estudios familiares (Figura 1.2). El problema de la heredabilidad perdida dio lugar al planteamiento de nuevos paradigmas. En ellos se atribuye la predisposición genética a: 1) muchas variantes comunes (frecuencia del alelo alternativo (MAF) $> 5\%$) con efectos muy débiles, que están por debajo de los niveles de significancia de los GWAS (Shi et al., 2016), 2) variantes de baja frecuencia (MAF 1-5%) y raras (MAF $< 1\%$) con efectos moderados a grandes (Cirulli & Goldstein, 2010; Marouli et al., 2017) o 3) una combinación de interacciones genéticas, ambientales y epigenéticas (Eichler et al., 2010). Se ha visto que las variantes raras pueden contribuir a cambios grandes de expresión en diferentes tejidos (Li et al., 2017a).

Las variantes raras son abundantes en el genoma humano, como consecuencia de la rápida expansión demográfica de la población humana hace aproximadamente 10.000 años (Keinan & Clark, 2012). Se predice que estas variantes raras serán en su mayoría población-específicas debido a los efectos fundadores causados por deriva génica, destacando la importancia de estudiarlas en cada población (Sirugo

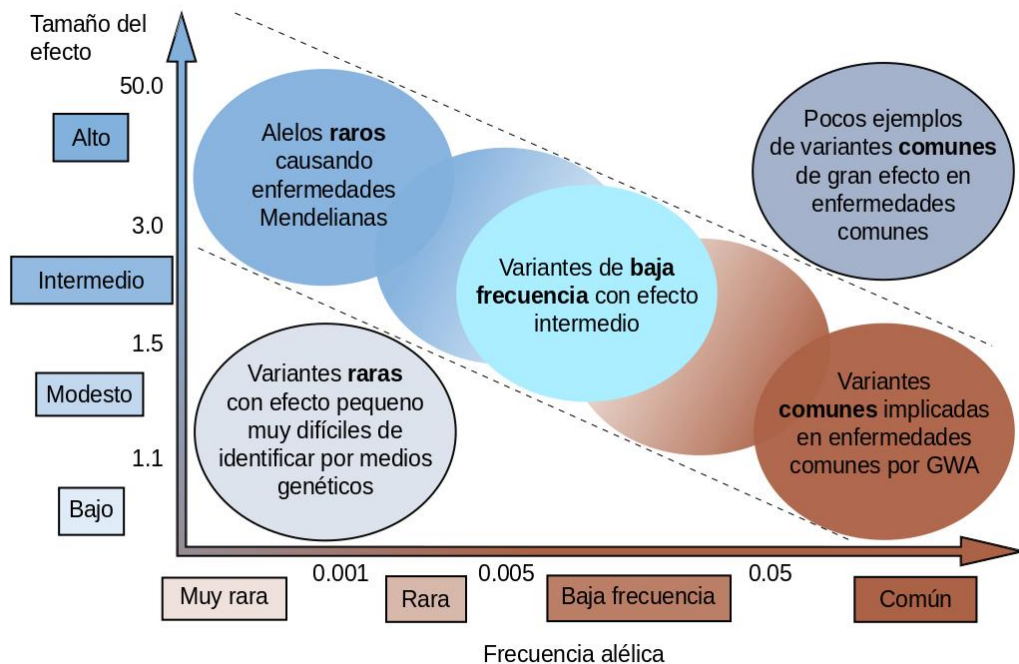


Figura 1.2. Posibilidad de identificar variantes génicas por su frecuencia alélica de riesgo y tamaño del efecto genético (“*odds ratio*”). Adaptada de Manolio et al. (2009).

et al., 2019). Según la teoría evolutiva, las variantes que favorecen las enfermedades serían sujetas a selección negativa, por ser perjudiciales para el *fitness* evolutivo, y por lo tanto se presentan con baja frecuencia alélica. La existencia de estas variantes refleja entonces el balance entre la creación de nuevas variantes de susceptibilidad por mutación, y la selección evitando que se establezcan en la población (Gibson, 2012).

Por otra parte, Boyle et al. (2017b) sugirieron un modelo “omnigénico” para describir la arquitectura de las características y enfermedades complejas. Este modelo sugiere que las mismas están gobernadas por un gran número de variantes génicas con efectos pequeños: posiblemente basado en una arquitectura génica de redes regulatorias compuestas de pocos genes clave que afectan directamente la característica y a la vez con muchos genes fuera de ese núcleo que la afectan indirectamente. Las vías de regulación génica estarían suficientemente

interconectadas para permitir que todos los genes expresados en una célula relevante para la enfermedad contribuyan a su fenotipo. Así, miles de genes aparentemente no relacionados con la característica pueden tener efecto sobre el fenotipo de la enfermedad. Las enfermedades genéticas complejas estarían causadas por un contexto genético diverso, que puede incluir interacciones en *cis* y en *trans*, así como expresión y regulación génica a diferentes niveles.

Adicionalmente, se ha propuesto la existencia de “asociaciones sintéticas”, en las cuales la asociación de una variante común con una patología se debe en realidad al desequilibrio de ligamiento entre la misma y variantes raras que confieren susceptibilidad a la enfermedad segregando en el mismo haplotipo (Dickson et al., 2010). Esta idea aún es discutida (Hernandez et al., 2017). Las asociaciones sintéticas probablemente no expliquen la mayoría de la heredabilidad perdida de las características complejas, pero es un efecto que debe ser considerado cuando surgen asociaciones aparentes con variantes comunes.

También se propone que un porcentaje significativo de la susceptibilidad heredada a enfermedades comunes se debe a la suma de los efectos de un conjunto de variantes raras independientes en varios genes, cada una confiriendo un aumento moderado en el riesgo de la patología (Bodmer & Bonilla, 2008). Este modelo es plausible para la determinación de la EG, ya que variantes de efecto mayor serían seleccionadas en contra en la población por su reducida viabilidad.

Aunque el impacto relativo de las variantes comunes y raras en las enfermedades comunes es aún discutido, es posible que una parte importante de las bases genéticas de las enfermedades complejas se deba a variantes raras que tengan un efecto importante en la susceptibilidad individual a patologías.

1.3.1. ESTUDIOS DE VARIANTES GÉNICAS EN PPT

Entender las bases genéticas del PPT ayudará a determinar acciones para prevenirlo. Hay muchos esfuerzos de investigación en prematuridad humana, que incluso se nuclean en varias bases de datos públicas específicas para la patología (Kim et al., 2016; Sirota et al., 2018; Uzun et al., 2012).

El estudio del PPT ha usado varias estrategias. Se han realizado análisis de asociación basados en el uso de genes candidatos como posibles desencadenantes del parto a término o pretérmino, así como también estudios de GWAS. Varios estudios de genes candidatos, usando casi exclusivamente un diseño caso-control, identificaron algunos polimorfismos asociados a PPT (Myking et al., 2011; Romero et al., 2010a; Ryckman et al., 2010; Velez Edwards et al., 2008, Velez Edwards et al., 2009). Sólo entre 2007 y 2015 se habían encontrado 119 genes candidatos presumiblemente asociados con PPT (Sheikh et al., 2016). Los genes encontrados con estas aproximaciones están en su mayoría relacionados con mecanismos de inflamación y remodelación de la matriz extracelular (Crider et al., 2005; Sheikh et al., 2016; Strauss et al., 2018). Estos resultados apoyan la hipótesis de la existencia de una base inflamatoria en el desencadenamiento del parto y del PPT. En particular, nuestro grupo de investigación previamente detectó asociación entre un polimorfismo de nucleótido simple (SNP) en el gen *TLR4* de los recién nacidos y el riesgo de PPT en prematuros severos, sugiriendo diferencias en los procesos inflamatorios en este estado (Rey et al., 2008). En otra cohorte de prematuros severos analizamos díadas de madre e hijo y no encontramos asociación de algunos polimorfismos vinculados con la vía inflamatoria en hijos (Pereyra et al., 2012).

Los GWAS realizados en PPT hasta el momento no fueron informativos. Una excepción es el estudio GWAS de mayor tamaño muestral realizado hasta el momento en PPT, en el que se identificaron variantes significativamente asociadas al PPT espontáneo únicamente en tres genes: *EBF1*, *EEFSEC*, y *AGTR2* (Zhang et al., 2017). Sin embargo, y a pesar de los esfuerzos de investigación en el tema,

hasta el momento las variantes genéticas identificadas mediante análisis de genes candidatos y GWAS muestran una asociación leve y variable con el PPT, además de poco reproducible en diferentes estudios y poblaciones. Esto sugiere que las SNV asociadas con la duración de la gestación tienen efectos pequeños en el fenotipo (Zhang et al., 2018). La falta de reproducibilidad de los resultados también puede deberse a variabilidad en la determinación de los fenotipos, tamaños de muestra pequeños, y/o diferencias genéticas entre las poblaciones (Sirugo et al., 2019). La dificultad en encontrar variantes asociadas con el PPT puede deberse a la contribución de variantes raras, que presenten mayor penetrancia (Bodmer & Bonilla, 2008). Para estudiar su rol en el PPT se requieren estudios con tecnología adecuada para su detección, tales como la secuenciación masiva.

Por otro lado, por definición, el PPT es una variable dependiente del tiempo. Es posible que las variantes génicas puedan conferir diferentes efectos en diferentes EG, lo que complica aún más su análisis (Zhang et al., 2018). En este sentido, análisis estratificados de las EG (*e.g.* pretérminos severos, tardíos, término) podrían revelar efectos hasta ahora no detectados.

Tal como ocurre generalmente con las enfermedades genéticas humanas, la mayoría de los estudios de PPT ocurren en poblaciones con ancestría europea (Sirugo et al., 2019). En particular, los estudios genéticos de PPT en poblaciones latinoamericanas se limitan a Argentina (Gimenez et al., 2017; Kaluarachchi et al., 2016; Kim, 2012; Mann et al., 2013), Chile (Lee et al., 2013; Romero et al., 2010b) y a los realizados por nuestro grupo en Uruguay (Pereyra et al., 2012; Rey et al., 2012, Rey et al. 2008). Las diferentes incidencias de PPT en poblaciones con diferente ancestría (Goldenberg et al., 2008) sugieren que el mestizaje genético incide en el PPT. Las bases genéticas del PPT pueden variar entre las poblaciones, por variación genética específica, así como por resultado de deriva génica, selección local, o ambas. Las conclusiones de estudios genéticos en otras poblaciones no necesariamente son transferibles a la clínica en la nuestra, ya que pueden ser incompletas o no aplicables (Petrovski & Goldstein, 2016).

1.4 EFECTO DE LAS VARIANTES GÉNICAS EN LA EXPRESIÓN GÉNICA

La transcriptómica es el estudio de moléculas de ARN mediante técnicas de gran escala. En contraste con el ADN, que es mayormente idéntico en todas las células del organismo, el ARN transcrito activamente es altamente dinámico, y refleja la diversidad de tipos y estados celulares. La investigación en el transcriptoma humano permitió identificar bases moleculares de muchas enfermedades y procesos biológicos (Montgomery & Dermitzakis, 2011). La abundancia del transcripto de un gen puede ser regulada por varios mecanismos, a través de modificaciones epigenéticas e interacciones con factores de transcripción y otros modificadores. A su vez, polimorfismos genéticos en las regiones reguladoras pueden afectar estos mecanismos de regulación (Pai et al., 2015; Skeeles et al., 2013). Por lo tanto, las variantes genéticas que afectan la regulación génica juegan un papel importante en la susceptibilidad a enfermedades complejas, considerándose la expresión génica como un determinante de la variación fenotípica.

El estudio de la variación en la expresión génica ha arrojado luz sobre los mecanismos de acción de loci asociados con enfermedades, permitiendo relacionar marcadores genéticos de la enfermedad con niveles de expresión de uno o más genes (Cookson et al., 2009; Fraser & Xie, 2009). El secuenciado masivo de ARN (ARN-Seq) permite medir prácticamente sin sesgo los niveles de expresión génica de un transcripto (Wang et al., 2009), revelar variantes de secuencia en las regiones transcritas, así como también determinar eventos de corte y empalme alternativos y expresión alelo específica (Marioni et al., 2008; Pan et al., 2008). Por otro lado, el análisis de variantes genéticas en estos datos puede ser usado para identificar genes o regiones reguladoras que controlan la expresión génica. Así, la combinación del análisis de la variación en la expresión génica y de la variación en la secuencia de ADN se puede usar para caracterizar enfermedades complejas en humanos (Montgomery & Dermitzakis, 2011).

1.4.1. EXPRESIÓN GÉNICA EN PARTO A TÉRMINO Y PARTO PREMATURO

Aunque ha habido varios estudios transcriptómicos relacionados a la gestación y al PPT, la mayoría se basan en comparar RPM vs. no RPM o preeclampsia y muy pocos analizaron el PPT espontáneo sin ruptura de membranas. En su mayoría se basan en muestras de placenta (Eidem et al., 2015).

La expresión génica del trabajo de parto a término ha sido más estudiada que la del PPT. Estudios en miometrio, placenta y membranas fetales demuestran que el perfil de expresión génica es región-específico y tejido-específico (Bukowski et al., 2006; Haddad et al., 2006; Kim et al., 2012; Nhan-Chang et al., 2010). Durante el trabajo de parto, las membranas corioamnióticas tienen un perfil de expresión génica diferente en el sitio de ruptura de membranas al de un lugar alejado, lo cual es consistente con los cambios histológicos observados en esa región previo al parto (Mittal et al., 2010; Nhan-Chang et al., 2010). Estos cambios pueden responder a la separación de las diferentes capas de las membranas fetales y/o a la exposición de las membranas a citoquinas y enzimas que degradan matriz presentes en la vagina y en el líquido amniótico. El perfil de expresión génica en este tejido y el resto de los tejidos gestacionales estudiados es consistente con una activación de las vías de inflamación y de adhesión celular (Bollopragada et al., 2009; Mittal et al., 2010; Nhan-Chang et al., 2010).

Si bien existen algunos datos de transcriptomas de PPT espontáneo, los resultados aún son limitados y se limitan a poblaciones de ascendencia europea. Estudios en placenta (Chim et al., 2012), miometrio y membranas fetales (Bukowski et al., 2017) encuentran una sobreexpresión de genes relacionados con respuesta inmune. Por otro lado, las mayores diferencias en expresión entre PPT y el reposo ocurren en la interfaz materno-fetal: decidua, corion y amnios (Bukowski et al., 2017). Ambos análisis se realizaron con microarreglos de ARN, en poblaciones de China y de Estados Unidos. No se conocen marcadores de expresión propios de nuestra población, la cual tiene un contexto genético y ambiental propio, ni tampoco de ninguna otra población latinoamericana. Además, el uso de microarreglos para el

análisis de expresión génica tiene limitaciones metodológicas conocidas, por lo cual es recomendado emplear ARN-seq. A diferencia del primero, el ARN-seq no tiene ruido por artefactos de hibridación cruzada, tiene mayor rango de resolución y la habilidad de detectar transcriptos nuevos en forma no sesgada.

1.5 APORTES DE ESTE TRABAJO DE TESIS

El objetivo de los estudios genéticos del PPT es el de identificar vías biológicas que expliquen las diferencias entre el fenotipo patológico y fisiológico. Esta información contribuye a un mejor entendimiento del parto y podría usarse para desarrollar modelos predictivos y terapias adecuadas para reducir la incidencia del parto pretérmino. La complejidad y heterogeneidad del PPT requieren un estudio con diferentes aproximaciones metodológicas para dilucidar los procesos conduciendo a este fenotipo. En este trabajo de tesis usamos un enfoque genético y transcriptómico para identificar genes vinculados al PPT en una población uruguaya.

2

OBJETIVOS Y ESTRATEGIA DE TRABAJO

2.1 HIPÓTESIS DE TRABAJO

En base a lo expuesto en el capítulo anterior, nuestra hipótesis plantea que el parto prematuro espontáneo es desencadenado por genes vinculados a la vía inflamatoria, que determinan un perfil de expresión propio del PPT. Este perfil es dependiente de la presencia de variantes génicas tanto en el genoma materno como el fetal. Planteamos que para identificar dicho perfil se deben combinar metodologías y tomar en cuenta las características genéticas de la población.

2.2 OBJETIVOS

2.2.1. OBJETIVO GENERAL

Aportar a la identificación del contexto genético en el que se produce el PPT en nuestra población utilizando herramientas genómicas y transcriptómicas, que incluyan factores tanto maternos como fetales.

2.2.2. OBJETIVOS ESPECÍFICOS

- Evaluar las contribuciones de los genomas materno y fetal al fenotipo PPT, buscando asociaciones entre el PPT y SNPs en genes pertenecientes a la vía inflamatoria.
- Caracterizar el transcriptoma de las membranas corioamnióticas de los recién nacidos prematuros severos.
- Identificar variantes génicas candidatas, potencialmente relacionadas con la expresión génica diferencial en PPT severo.

2.3 ESTRATEGIA DE TRABAJO

Para alcanzar estos objetivos usamos tres estrategias. Primero, realizamos un estudio de genes candidatos, seleccionando genes relacionados a vías inflamatorias (Capítulo 3). Comparamos la presencia de SNPs en un grupo de madres e hijos nacidos a término y nacidos a pretérmino, y lo evaluamos en el contexto de variables ambientales que puedan estar contribuyendo al fenotipo.

En segundo lugar, examinamos los perfiles de expresión génica en membranas fetales de recién nacidos pretérminos severo y nacidos a término, usando secuenciación masiva de ARN (Capítulo 4). Debido a que el PPT severo parece depender más que el PPT moderado de factores genéticos y constitucionales, postulamos que los cambios de expresión génica serán más marcados en las membranas corioamnióticas de recién nacidos de menos de 34 semanas de edad gestacional. Buscamos caracterizar el perfil de expresión del PPT e identificar genes vinculados con el fenotipo extremo.

Por último, usamos los datos transcriptómicos para estudiar variantes génicas vinculadas al PPT (Capítulo 5), centrándonos en el análisis de los genes diferencialmente expresados encontrados en la sección anterior. Buscamos variantes (tanto reportadas como no reportadas) que puedan estar modulando la expresión génica en nuestra población.

3

INTERACCIONES ENTRE FACTORES AMBIENTALES Y VARIANTES GENÉTICAS EN PARTO PRETÉRMINO

Publicado en:

Pereyra S, Bertoni B, Sapiro R. (2016) **Interactions between environmental factors and maternal-fetal genetic variations: strategies to elucidate risks of preterm birth.** European Journal of Obstetrics and Gynecology and Reproductive Biology, 202:20-25. <https://doi.org/10.1016/j.ejogrb.2016.04.030>

3.1 ANTECEDENTES Y ESTRATEGIA

Como se indicó previamente, el PPT puede desencadenarse por la activación en forma prematura de la cascada inflamatoria. Esto condujo a que muchos estudios de asociación genética se basen en la búsqueda de asociaciones con genes candidatos de la vía inflamatoria. Incluso, estudios previos de nuestro laboratorio demostraron que el SNPs rs4986790 en el gen *TLR4* se asocia a PPT en pretérminos severos con RPM (Rey et al., 2008). Además, con el objetivo de poder genotipar varios SNPs simultáneamente en forma eficiente y con un bajo costo, hemos generado un sistema multiplex que detecta 4 SNPs de la vía inflamatoria en un único ensayo. Este sistema identifica simultáneamente mutaciones en cuatro genes involucrados en vías inflamatorias, mediante un análisis de desnaturalización en alta resolución (HRM): *TLR4*, Interleuquina 6 (*IL6*), Interleuquina 1B (*IL1β*) y Receptor beta de Interleuquina 12 (*IL12RB*) (Pereyra et al., 2012). Es decir que se centra en SNPs relacionados a vías inflamatorias involucradas en mecanismos de defensa, activación de la inmunidad innata e infección, tal como se indica en la Enciclopedia de Genes y Genomas Kyoto (KEGG, Kanehisa & Goto, 2000). Previamente, lo usamos para determinar la frecuencia alélica de dichas variaciones en nuestra población.

Sin embargo, como ya se observó, el PPT tiene contribuciones potenciales de dos genomas, el materno y el fetal. Por lo que, junto con otros autores, postulamos que además de las contribuciones individuales, las interacciones entre ambos genomas pueden afectar el riesgo de presentar PPT (Myking et al., 2011). Al ser una característica multifactorial, condiciones como el stress materno, embarazos multiples y exposición a tóxicos durante el embarazo podrían interaccionar con la predisposición genética y desarrollar así el PPT (Berkowitz & Papiernik, 1993; Goldenberg et al., 2008; Wen et al., 2004).

En este capítulo se describe la búsqueda de asociaciones entre el PPT y SNPs candidatos utilizando nuestra técnica de detección de SNPs (Pereyra et al., 2012).

Para ello usamos una cohorte de díadas de recién nacidos y sus madres, mediante un estudio de tipo caso-control. En segundo lugar, combinamos estos datos genotípicos con un análisis de factores de riesgo ambientales al PPT, de forma de aportar datos sobre la interacción genotipo-ambiente en esta patología.

3.2 METODOLOGÍA

3.2.1. MUESTRAS USADAS EN ESTE ESTUDIO

Realizamos un diseño experimental de tipo caso-control. La colecta de muestras fue realizada en la Clínica Ginecitológica “C” del Centro Hospitalario Pereira Rossell (CHPR), Montevideo, Uruguay, entre febrero de 2012 y marzo de 2014. Se seleccionaron madres y sus hijos recibiendo atención obstétrica en el CHPR. Los casos fueron díadas ($n = 143$) de madres y sus neonatos con embarazos complicados por PPT espontáneo, de edad gestacional menor a 37 semanas. Los controles ($n = 108$) fueron madres y sus neonatos nacidos a término (edad gestacional ≥ 37 semanas), que no presentaron complicaciones en su parto vaginal o cesárea electiva a término. Los criterios de exclusión para las muestras comprendieron: embarazos múltiples, consumo de drogas, patologías y/o infecciones crónicas, malformaciones fetales y no otorgar su consentimiento para participar del estudio. Las infecciones que no fueron consideradas como criterios de exclusión para participar del estudio fueron infecciones del tracto urinario, genitales bajas y sífilis, pero no el virus de inmunodeficiencia humana (HIV), ya que las pacientes con HIV fueron excluidas de la muestra.

Con la aprobación de las participantes se realizó una encuesta en persona sobre datos clínicos, familiares y sociodemográficos, con un cuestionario específicamente diseñado para este fin. Se colectaron datos clínicos de las madres y sus neonatos mediante el Sistema Informático Perinatal, que consiste en registros

de la clínica perinatal desarrollados por el Centro Latinoamericano de Perinatología de la Organización Panamericana de la Salud (Simini, 1999), el cual se presenta en el Apéndice.

El tamaño de la muestra se decidió al inicio del estudio, según cálculos del poder estadístico del mismo (Hong & Park, 2012). Para ello se consideró la frecuencia del alelo alternativo de los SNPs que tuvieran un efecto inflamatorio posible en poblaciones europeas (estimado en un 20-40% aproximadamente). Además, se estimó que la razón de proporciones o “*odds ratio*” esperada sería de 1,5 a 2 entre casos y controles. Esto permitió calcular el tamaño de muestra necesario para poder tener un buen poder estadístico.

3.2.2. COLECTA DE LAS MUESTRAS, EXTRACCIÓN DE ADN GENÓMICO Y PCR EN TIEMPO REAL

Se colectaron muestras de sangre de los recién nacidos (a partir de sangre del cordón umbilical colectada en el momento del nacimiento) y raspados de mucosa yugal de las madres. En ambos tipos de muestras se realizó la extracción de ADN con el kit comercial DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). El análisis de HRM se realizó en el instrumento Rotor-Gene 6000 real-time (Corbett Life Science, Sydney, Australia) con colorante de tecnología saturante Eva Green (Type-it HRM PCR Kit, Qiagen, Hilden, Germany) tal como describimos en Pereyra et al. (2012). Brevemente, se analizan los SNPs rs4986790 (*TLR4*), rs1800795 (*IL6*), rs16944 (*IL1 β*) y rs375947 (*IL12RB*) en un ensayo de PCR multiplex (Tabla 3.1). Los SNPs candidatos se seleccionaron por estar en genes con efecto inflamatorio posible. Se realizaron las PCRs en un volumen final de 10 μ l, conteniendo 1X de buffer Qiagen HRM-typing, 2,46 μ M de cada uno de los primers del gen *TLR4*, 0,46 μ M de cada uno de los primers del gen *IL6*, 0,26 μ M de cada uno de los primers del gen *IL1 β* , 0,9 μ M de cada uno de los primers del gen *IL12RB* y 25 ng de ADN molde. La reacción de PCR se realizó con una

desnaturalización inicial a 95°C por 5 minutos, seguido de 35 ciclos de 95°C por 15 segundos y 60°C por 1 minuto, y luego las curvas de HRM se generaron adquiriendo fluorescencia con una rampa de temperatura de 72 a 88°C. Se incluyeron muestras con genotipos homocigotas y heterocigotas como controles positivos en todos los experimentos.

Las curvas de HRM se normalizaron con las regiones pre- y post-desnaturalización, y se asignó el genotipo a cada muestra según la forma de la curva de HRM en el software Rotor-Gene 1.7 (build 73) e inspección visual. Las curvas de desnaturalización obtenidas fueron analizadas separadamente para cada amplicón, para asignar genotipos a cada muestra.

Tabla 3.1. Características generales de los SNPs estudiados.

Ubicación genómica	Identificador SNP	Gen	Alelo referencia	Alelo alternativo	Contexto genómico
2: 112837290	rs16944	<i>IL1β</i>	A	G	Promotor
7: 22727026	rs1800795	<i>IL6</i>	C	G	Intrón
9: 117713024	rs4986790	<i>TLR4</i>	A	G	Exón: cambio no sinónimo Asp299Gly
19: 18069641	rs375947	<i>IL12RB</i>	A	G	Exón: cambio no sinónimo Met365Thr

3.2.3. ANÁLISIS ESTADÍSTICOS

Las frecuencias alélicas, genotípicas y test exactos de equilibrio Hardy-Weinberg se calcularon en el programa PLINK v1.07 (Purcell et al., 2007). Se compararon las frecuencias alélicas usando el test Chi-cuadrado o el test exacto de Fisher, y los promedios usando el test *t* de Student. Se evaluó la asociación entre los genotipos y los factores ambientales con el PPT, mediante un análisis de regresión logística.

Se evaluó la interacción de las diferentes combinaciones de SNPs sobre el PPT. Se exploraron efectos diferenciales de los genotipos sobre el riesgo de PPT mediante la inclusión de términos de interacción o condicionales usando el software Epi Info 2000 (Centers for Disease Control and Prevention, Atlanta, GA, USA). En todos los casos, se consideró significativo un valor de p menor o igual a 0,05.

Las variables ambientales y clínicas estudiadas se categorizaron de forma dicotómica para su análisis. Se consideró que las madres tuvieron exposición al humo de tabaco cuando fue fumadora activa y/o pasiva durante la gestación.

3.2.4. ASPECTOS BIOÉTICOS

Todos los datos y muestras utilizados fueron tratados de forma anónima e identificados por un código que no identifica a la persona participante. Los pacientes participantes fueron informados de la finalidad y el alcance de este estudio. El protocolo de trabajo se adecua a la Declaración de Helsinki y fue aprobado por el Comité de Ética de la Facultad de Medicina, Universidad de la República (Consejo de la Facultad de Medicina, Exp. N° 071140-001476-08). Se obtuvo el consentimiento informado de todas las madres que participaron en el estudio.

3.3 RESULTADOS

3.3.1. CARACTERÍSTICAS DE LA MUESTRA ESTUDIADA

Al estimar el tamaño muestral necesario para poder tener un buen poder estadístico, calculamos que una muestra de 500 alelos es suficiente para detectar un OR entre 1,5-2,0 con un poder estadístico de 80% y un intervalo de confianza de 95%.

En total, se incluyeron 251 díadas madre-neonato. Las características generales de la muestra estudiada se presentan en la Tabla 3.2. Comparando casos y controles, no se detectaron diferencias significativas con respecto a edad materna y paridad. La edad promedio y error standard de la edad gestacional fue $33,6 \pm 0,2$ semanas en casos vs. $39,1 \pm 0,1$ en controles (valor $p < 0,01$). La edad gestacional de los casos varió entre 24 a 36 semanas, mientras que la de los controles lo hizo de 37 a 41 semanas. No se encontraron diferencias significativas entre ambos grupos con respecto a sexo del recién nacido (51% neonatos varones en casos vs. 54% en controles). Siete por ciento de las mujeres que tuvieron un parto a término y 17% de las madres que lo hicieron a pretérmino tuvieron al menos un PPT previamente,

Tabla 3.2. Características sociodemográficas y clínicas de madres y sus recién nacidos de la población de estudio.

Característica	Pacientes a término (%) n=108	Pacientes a pretérmino (%) n=143	OR (IC 95%)	valor p
Edad materna (<19 y >35)	16	22	1,6 (0,8-2,9)	0,2
Nivel educativo ^(a)	35	43	0,7 (0,4-1,2)	0,2
Estado civil ^(b)	17	25	1,6 (0,8-3,2)	0,1
Recién nacidos masculinos	51	54	1,15 (0,6-1,9)	0,6
Exposición a humo de tabaco	44	65	2,6 (1,5-4,3)	4 x 10⁻⁴
Hipertensión	7	4	0,4 (0,14-1-4)	0,2
Hemorragias durante el embarazo	1	4	4,7 (0,5-39)	0,2
Infecciones	19	27	1,6 (0,8-3)	0,2
RCIU	2	1	2,3 (0,6-8,9)	0,2
RPM	13	45	5,5 (2,9-10,6)	1 x 10⁻⁶
Anemia	14	12	0,8 (0,4-1,7)	0,5

Se presenta la razón de probabilidades u "odds ratio" (OR), los intervalos de confianza (IC) y los valores de p entre pacientes a término y pretérmino. La significancia estadística se marca con valores en negrita.

^(a) al menos un año vs. menos de un año de educación secundaria

^(b) madres solteras vs. madres casadas o en concubinato

RPM: ruptura prematura de membranas; RCIU: retraso en el crecimiento intrauterino

Tabla 3.3. Conteos de genotipos de los SNPs analizados para madres y recién nacidos de la muestra. Se presentan las frecuencias genotípicas entre paréntesis.

SNP (Símbolo del gen)	Genotipo	Madres		Recién nacidos	
		Controles	Casos	Controles	Casos
rs4986790 (<i>TLR4</i>)	A A	92 (0,89)	124 (0,87)	90 (0,87)	128 (0,91)
	A G	11 (0,11)	18 (0,13)	13 (0,13)	13 (0,09)
	G G	0	0	0	0
rs1800795 (<i>IL6</i>)	C C	0	0	0	0
	C G	38 (0,35)	64 (0,45)	28 (0,26)	45 (0,31)
	G G	70 (0,65)	79 (0,55)	80 (0,74)	98 (0,69)
rs16944 (<i>IL1β</i>)	A A	16 (0,16)	17 (0,12)	15 (0,14)	19 (0,13)
	A G	61 (0,61)	81 (0,57)	59 (0,55)	89 (0,63)
	G G	23 (0,23)	44 (0,31)	33 (0,31)	33 (0,23)
rs375947 (<i>IL12RB</i>)	A A	49 (0,47)	56 (0,40)	42 (0,38)	59 (0,42)
	A G	37 (0,36)	61 (0,44)	51 (0,47)	67 (0,47)
	G G	18 (0,17)	22 (0,16)	15 (0,14)	16 (0,11)

sin diferencias estadísticas entre los grupos. El nivel educativo alcanzado por la madre no afectó la incidencia del PPT; 35% de las madres que completaron al menos 1 año o más de educación secundaria tuvieron partos a término, y 43% lo hicieron a pretérmino. La exposición al humo de tabaco (activa y/o pasivamente) sí se encontró asociada al PPT (valor $p < 0,01$): 44% de las madres con partos a término estuvieron expuestas, mientras 65% de las que tuvieron PPT lo estuvieron (Tabla 3.2). Por otro lado, no se encontraron diferencias entre madres que tuvieron partos a término y pretérmino con respecto a la presencia de hipertensión arterial en las madres (7% vs 4%, respectivamente), hemorragias durante el embarazo (1% vs 4%, respectivamente), anemia (14% vs 12%, respectivamente), infecciones maternas (19% vs 27%, respectivamente), ni restricción del crecimiento

intrauterino (2% vs 1%, respectivamente). Se encontró que la RPM está significativamente asociada con PPT (Tabla 3.2).

3.3.2. ANÁLISIS DE ASOCIACIÓN GÉNICA

Las frecuencias genóticas de madres y recién nacidos para cada SNP se presentan en la Tabla 3.3. Al igual que lo reportado previamente en (Pereyra et al., 2012), los genotipos C/C para rs1800795 (*IL6*) y G/G para rs4986790 (*TLR4*) no se encontraron en la población estudiada. Se realizaron tests de Chi-cuadrado sobre las frecuencias alélicas para evaluar diferencias entre los grupos. Ninguno de los SNPs analizados en las madres estuvo asociado significativamente con el PPT. No se encontró asociación significativa entre los genotipos analizados de los recién nacidos nacidos a término y a pretérmino ($p > 0,05$; Tabla 3.5).

Los test exactos de equilibrio Hardy-Weinberg indican que los SNPs rs4986790 (gen *TLR4*) y rs375947 (gen *IL12RB*) están en equilibrio Hardy-Weinberg para madres del grupo a término y a pretérmino ($p > 0,05$), mientras que rs1800795 (gen *IL6*) y rs16944 (gen *IL1 β*) no lo están. Se encontró la misma situación cuando se analizaron los genotipos del grupo de los recién nacidos a pretérmino (Tabla 3.5). En el caso de los recién nacidos a término se encontraron los 4 SNPs analizados en equilibrio Hardy-Weinberg.

El análisis de regresión logística detectó una interacción significativa entre el polimorfismo *IL1 β* en muestras maternas y el polimorfismo *IL6* en muestras de neonatos (Tabla 3.4). Cuando se analizó el efecto del SNP *IL6* en los recién nacidos mientras que al mismo tiempo se controla por las variantes *IL1 β* de la madre, se encontró una asociación significativa con el PPT. Esto podría estar indicando un posible efecto escondido de cada SNP separadamente. De las combinaciones genóticas entre madres y recién nacidos para estos SNPs, solamente la combinación *IL1 β* G/G en madres y *IL6* G/C en recién nacidos fue diferente entre casos y controles ($p < 0,05$; test de Fisher; Figura 3.1). No se detectaron otras asociaciones significativas entre los genotipos de madres e hijos y PPT.

Tabla 3.4. Análisis de regresión logística entre casos y controles para el polimorfismo materno rs16944 (*IL1 β*) y el de recién nacidos rs1800795 (*IL6*) ajustado por el otro polimorfismo, así como su interacción.

Genotipo	Genotipo de riesgo	OR (95% IC)	valor p
rs16944 (<i>IL1β</i>) madre	GG	1,14 (1,0-1,3)	0,0997
rs1800795 (<i>IL6</i>) recién nacido	CG	12,3 (2,0 -76,7)	0,0072
Interacción		0,3 (0,1-0,8)	0,0122
OR: razón de probabilidades o " <i>odds ratio</i> ". IC: intervalo de confianza. Los valores significativos están marcados en negrita.			

La población de estudio es bastante homogénea (Tabla 3.2). Sin embargo, dos variables tuvieron diferencias significativas entre PPT y controles: exposición al humo de tabaco y RPM. Cuando se analizaron los SNPs genotipados controlando por el efecto de cada una de estas dos variables ambientales, se reveló el efecto de la exposición del humo de tabaco en el PPT. Se encontró que los polimorfismos rs16944 (*IL1 β*) y rs1800795 (*IL6*) en madres están significativamente asociados con PPT cuando se controla por la exposición del humo de tabaco durante el embarazo (Tabla 3.6).

Dentro de las 143 díadas con parto prematuro, 64 de ellas tuvieron parto pretérmino severo (con menos de 34 semanas de edad gestacional). Cuando se consideran únicamente los prematuros severos se encuentra que el polimorfismo rs4986790 en los recién nacidos (*TLR4*) y la RPM están significativamente asociados con PPT ($p = 0,028$ y $p = 0,030$, respectivamente). Esto confirma datos previamente reportados por Rey et al. (2008).

Tabla 3.5. Estadísticos de tests exactos de Hardy-Weinberg y tests de chi-cuadrado para los genotipos de madres y recién nacidos analizados.

SNP (Símbolo génico)	Alelo alternativo	Madres						Recién nacidos					
		MAF en casos	MAF en controles	Valor p EHW (casos)	Valor p EHW (controles)	OR	valor p χ^2	MAF en casos	MAF en controles	Valor p EHW (casos)	Valor p EHW (controles)	OR	valor p χ^2
rs4986790 (TLR4)	G	0,06	0,05	1	1	1,2	0,64	0,05	0,06	1	1	0,72	0,41
rs1800795 (IL6)	C	0,22	0,18	<0,01	<0,05	1,35	0,19	0,16	0,13	<0,05	0,21	1,25	0,38
rs16944 (IL1 β)	A	0,4	0,47	<0,05	<0,05	0,78	0,19	0,45	0,42	<0,01	0,23	1,15	0,44
rs375947 (IL12RB)	G	0,38	0,35	0,31	0,48	1,12	0,55	0,35	0,38	0,71	1	0,89	0,54
MAF: Frecuencia del alelo alternativo; EHW: Equilibrio Hardy-Weinberg. OR: <i>odds ratio</i> . Valor p χ^2 : casos vs. controles.													

Tabla 3.6. Análisis de regresión logística de los genotipos maternos ajustados por madres que estuvieron expuestas a humo de tabaco durante el embarazo, y el correspondiente término de interacción con exposición al humo de tabaco.

	OR ajustado (95% IC)	valor p	Interacción (95% IC)	valor p
rs4986790 (<i>TLR4</i>)	1 (0,3–3,5)	0,985	2,1 (0,3–13,1)	0,416
rs1800795 (<i>IL6</i>)	2,8 (1,2–6,4)	0,013	0,3 (0,1–0,9)	0,039
rs16944 (<i>IL1β</i>)	0,4 (0,2–0,8)	0,006	3,3 (1,3–7,9)	0,009
rs375947 (<i>IL12RB</i>)	1,2 (0,7–2,0)	0,427	0,8 (0,4–1,7)	0,589

OR: razón de probabilidades o "odds ratio". IC: intervalo de confianza. Los valores significativos están marcados en negrita.

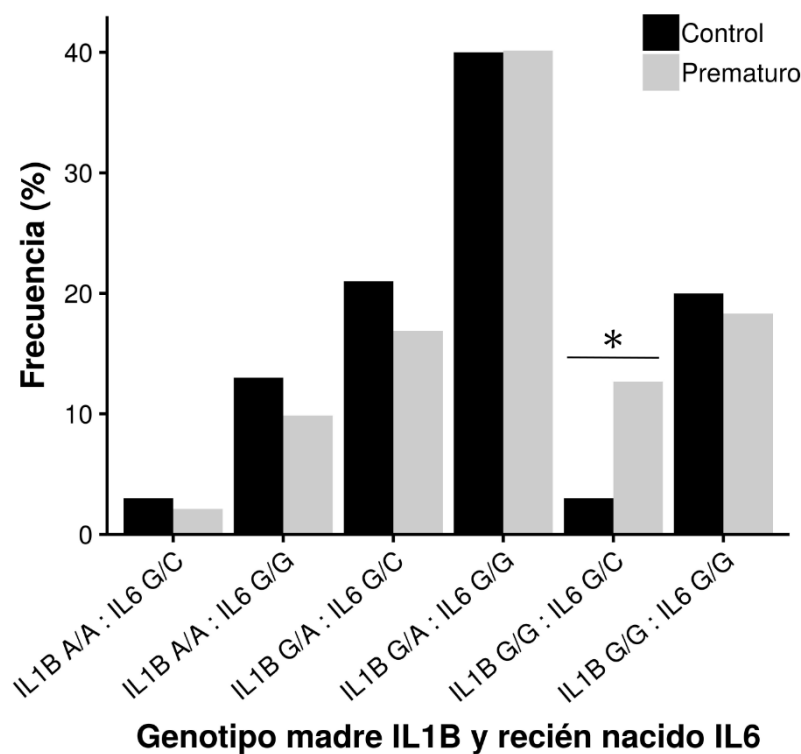


Figura 3.1. Distribución de los genotipos combinados de las díadas madre-recién nacido para el genotipo materno *IL1 β* (rs16944) y el del recién nacido *IL6* (rs1800795). Sólo la combinación de los genotipos *IL6* G/G en madres y *IL6* G/C en recién nacidos tuvo diferencias entre casos y controles ($p < 0,05$; test de χ^2).

3.4 DISCUSIÓN

En este capítulo, combinamos un diseño de díadas caso-control, el análisis de varios SNPs pertenecientes a una vía metabólica común, y el estudio de algunos factores ambientales de riesgo a PPT. Este tipo de análisis combinado permitió la discriminación de las contribuciones maternas y fetales al PPT y las condiciones ambientales relacionadas.

Se analizó el riesgo del PPT asociado con los genotipos de madre e hijo, lo que permitió detectar un riesgo aumentado de PPT cuando la madre y el recién nacido son portadores de una combinación específica de SNPs. La presencia del genotipo *IL1 β* G/G en madres y el genotipo *IL6* G/C en el genoma fetal está asociado con PPT (Figura 3.1, Tabla 3.4). Esto señala la importancia del rol de la interacción de los genomas materno y fetal en la determinación del PPT. Cuando se consideran ambos genes separadamente, el efecto no se observa, indicando que un SNP puede enmascarar el efecto del otro. Eso puede deberse a efectos de acción opuestos en la cascada que determina el PPT. Por otro lado, el alto valor de “*odds ratio*” observado (12,3; Figura 3.1) sugiere que portar la combinación específica de SNPs mencionada conlleva un riesgo asociado con PPT mucho mayor que las otras.

El efecto combinado de los genomas materno y fetal en PPT ha sido discutido en la literatura (Myking et al., 2011). Algunos estudios favorecen una contribución del genoma materno al tiempo de nacimiento (Kalish et al., 2006; Svensson et al., 2009), otros el del genoma fetal (Lui et al., 2018), mientras que otros apoyan el efecto de los dos genomas materno y fetal separadamente (Lunde et al., 2007; York et al., 2013; York et al., 2009). Se sabe poco acerca de las contribuciones genéticas de los genomas materno y fetal en relación a la duración del embarazo. En general se reconoce el rol de ambos genomas, con una contribución relativa discutida (York et al., 2013; Zhang et al., 2017). Los datos presentados en este capítulo sugieren que el momento del nacimiento puede cambiar como consecuencia de portar una combinación particular de variantes génicas entre madre y recién nacido.

Una consecuencia importante de esta interacción es que el riesgo de una madre de tener otro hijo prematuro luego de haber tenido un PPT depende no sólo de su propio genoma sino también del de su futuro hijo.

El efecto encontrado de los polimorfismos en los genes *IL1 β* y *IL6* sobre el PPT es concordante con la hipótesis de una base inflamatoria del PPT. La mayoría de los polimorfismos de los genes inflamatorios elegidos en el diseño múltiplex fueron previamente asociados con PPT. Por ejemplo, el SNP rs4986790 del gen *TLR4* fue previamente asociado con diferencias en la respuesta a liposacáridos y la predisposición al PPT (Arbour et al., 2000; Lorenz et al., 2002; Rey et al., 2008). También, los portadores del alelo G del polimorfismo rs1800795 (*IL6*) producen más IL6 que aquellos con genotipo C/C (Fishman et al., 1998). El genotipo C/C es protector contra PPT en mujeres de ascendencia europea, aunque esto no se observa en otras poblaciones heterogéneas o mestizadas, ni en los genotipos del feto (Wu et al., 2013). Por otro lado, el SNP rs16944, que se encuentra en la región promotora del gen *IL1 β* , está asociado con un riesgo aumentado de tener PPT en una población europea (Hollegaard et al., 2008).

Los datos presentados en este capítulo apoyan la idea de una posible contribución de la vía inflamatoria al PPT, dado que se encontró una asociación de una combinación específica de SNPs en genes inflamatorios con el PPT. Sin embargo, no se encontró asociación con el PPT para otras combinaciones de los SNPs analizados, ni tampoco cuando se analizaron separadamente.

Las diferencias entre nuestros resultados y los de otros estudios podrían deberse a diferencias genéticas entre las diferentes poblaciones humanas. La mayoría de los estudios de asociación génica con el PPT se realizaron en Estados Unidos o Europa, lo que dificulta la comparación de los resultados con poblaciones de diferente ancestría (Menon et al., 2009; Svensson et al., 2009). La población uruguaya ha sido descrita fundamentalmente como de origen europeo. Sin embargo, análisis de mestizaje genético demostraron que es una población tri-híbrida, con una contribución genética también de nativos americanos y africanos, que tienen una

contribución aproximada de 10% y 6% respectivamente (Hidalgo et al., 2005). Al analizar los linajes maternos mediante ADN mitocondrial, se revela una mayor contribución ancestral de 34% de nativo americanos, mientras la contribución africana varía entre 8% y 21% (Bonilla et al., 2015; Gascue et al., 2005; Pagano et al., 2005). Estas diferencias en ancestría podrían estar explicando las diferencias observadas con otras poblaciones en la contribución genética al PPT. También enfatizan la importancia de hacer estudios específicos a cada población para evaluar la contribución genética a diferentes patologías.

En este capítulo también evaluamos la posibilidad de que los factores ambientales y genéticos puedan interactuar para conducir al PPT. De las variables analizadas, sólo el consumo de tabaco durante el embarazo y RPM estuvieron distribuidas diferencialmente entre casos y controles, lo cual es concordante con estudios previos (Ananth & Vintzileos, 2006; Berkowitz & Papiernik, 1993; Goldenberg et al., 2008; Wen et al., 2004; York et al., 2014). La exposición al tabaco durante el embarazo es un factor bien documentado de riesgo de complicaciones durante el embarazo (Ko et al., 2014; Shah & Bracken, 2000), por lo cual podría enmascarar una predisposición genética al PPT. Encontramos que cuando se usan métodos estadísticos que controlan el efecto de la exposición al humo de tabaco, es posible revelar la asociación de los polimorfismos en los genes *IL1 β* y *IL6* en el genoma materno con el PPT. Así, incorporando datos ambientales revelamos que ambos SNPs están en sí mismos asociados con la patología, algo que no se encontró cuando se analizaron únicamente los datos genéticos. Esta hipótesis fue manejada en algunos estudios de otros autores en relación a recién nacidos y PPT. Por ejemplo, hay genotipos maternos que modifican la asociación entre el consumo de tabaco materno durante el embarazo y peso del recién nacido (Wang et al., 2002a). También se encontró que la presencia de vaginosis bacteriana en portadores de determinados genotipos del gen de Factor de Necrosis de Tumores 2 predispone al PPT espontáneo (Macones et al., 2004). Por lo tanto, consideramos que los factores ambientales deben ser considerados al realizar estudios de asociación genética de

tipo caso-control en PPT ya que estos contribuyen sustancialmente a la variabilidad fenotípica, para evitar una mala interpretación de los resultados.

Esta puede ser una de las razones por las que los resultados de asociación génica en PPT en estudios caso-control son replicados sólo raramente. Como se mencionó, la etiología del PPT es heterogénea: es resultado de varias causas subyacentes diferentes (Dolan et al., 2010; Romero et al., 2014). Además, se conocen varios factores ambientales que afectan el nacimiento, que deben ser considerados en la interpretación de datos genéticos. Por lo tanto, al momento de diseñar estudios de investigación en PPT, tanto las variables genéticas como las ambientales deben ser incorporadas.

Nuestro diseño experimental usó una estrategia combinada, analizando datos genéticos de las díadas madre-hijo en conjunción con variables ambientales. La misma usa una población bien caracterizada y un menor tamaño muestral en comparación a los necesarios para realizar análisis de tipo GWAS, así como también un número reducido de genes candidatos. Existen sólo unos pocos estudios de genes candidatos en PPT con este tipo de diseño.

En contrapartida, estas mismas ventajas también pueden ser consideradas como debilidades del estudio. Una de las limitaciones de nuestro estudio es el pequeño tamaño muestral, en comparación con análisis de GWAS, y el pequeño número de genes candidatos estudiados. Esto hace que la información obtenida y la significancia estadística de los datos sea menor a la lograda con los GWAS.

Nuestros datos de este capítulo sugieren que la vía inflamatoria está involucrada en el PPT, que los genotipos maternos y fetales contribuyen a esta patología, y que los factores ambientales (*e.g.* exposición al humo de tabaco) pueden enmascarar los efectos de las variantes genéticas en el PPT.

4

EL TRANSCRIPTOMA DE MEMBRANAS FETALES REVELA VÍAS INVOLUCRADAS EN EL PARTO PRETÉRMINO

Publicado en:

Pereyra S, Sosa C, Bertoni B, Sapiro R. (2019) **Transcriptomic analysis of fetal membranes reveals pathways involved in preterm birth.** BMC Medical Genomics. 12:53. <https://doi.org/10.1186/s12920-019-0498-3>

Pereyra S, Sosa C, Bertoni B, Sapiro R. (2018) **Transcriptomic analysis of fetal membranes reveals pathways involved in preterm birth.** bioRxiv. <https://doi.org/10.1101/358945>

4.1 ANTECEDENTES Y ESTRATEGIA

Como vimos en el capítulo 3, el PPT es resultado de la interacción de componentes genéticos y ambientales, y constituye un síndrome multifactorial complejo (Henderson et al., 2012; Romero et al., 2014). Sin embargo, para comprender la arquitectura genética del embarazo y el PPT se debe tener en cuenta, no sólo las causas conocidas del PPT sino también otros factores. Por ej. la gran cantidad de tejidos gestacionales maternos y fetales que deben interactuar para facilitar el parto (Iams, 2014; Romero et al., 2014), cada uno con un perfil propio de expresión génica. Estudios transcriptómicos en estos tejidos pueden ayudar a desarrollar un escenario molecular del PPT y mejorar el entendimiento de la fisiología y patología del parto a término y pretérmino. En particular, el estudio del transcriptoma de las membranas fetales en el sitio de ruptura en PPT puede ayudar a dilucidar cuales son los genes que se expresan en el PPT.

El ARN-seq es una tecnología potente para el análisis del transcriptoma que permite una caracterización profunda de la expresión génica (Kukurba & Montgomery, 2015). Los estudios de ARN-seq publicados en trabajo de parto humano han sido en general limitados a embarazos a término sin complicaciones y a placenta con diferentes EG (Kim et al., 2012; Mikheev et al., 2008; Söber et al., 2015). Existen más datos de expresión génica con experimentos basados en microarreglos, aunque la mayoría están focalizados en preeclampsia y RPM (Söber et al., 2015). Tal como para otras enfermedades, es necesario estudiar marcadores de expresión propios de nuestra población, la cual tiene un contexto genético y ambiental diferente.

Como mencionamos, el PPT podría ser desencadenado por diversos mecanismos patológicos dependiendo de la EG (Moutquin, 2003). El contexto genético parece jugar un rol más relevante en neonatos pretérminos severos que en los pretérminos moderados (Rey et al., 2008). Las modificaciones de la expresión del corioamnios que llevan a PPT severo deberían ser más drásticas que las que conducen a parto a

término o incluso a PPT moderado. Por lo tanto, decidimos focalizarnos en estudiar el transcriptoma de tejidos corioamnióticos de PPT severos para encontrar genes diferencialmente expresados (DE) con mayor significancia biológica. Además, para validar las vías metabólicas principales encontradas en este estudio, comparamos nuestros resultados con datos reportados en la literatura.

Basado en los trabajos previos de nuestro grupo de investigación (Pereyra et al., 2012; Pereyra et al., 2016; Rey et al., 2008; Rey et al., 2012) y el estado actual del conocimiento del trabajo de parto a término y pretérmino (Challis et al., 2009; Nhan-Chang et al., 2010; Romero et al., 2007; Romero et al., 2015; Strauss et al., 2018; Vrachnis et al., 2010), esperamos modificaciones en vías inflamatorias en el transcriptoma pretérmino en comparación al de término. En general, el objetivo principal de este estudio es obtener un patrón de expresión génica del PPT.

4.2 METODOLOGÍA

4.2.1. SELECCIÓN DE LAS MUESTRAS

La colecta de muestras fue realizada en la Clínica Ginecitológica “C” del Centro Hospitalario Pereira Rossell (CHPR), Montevideo, Uruguay, entre octubre de 2013 y febrero de 2015. Se colectaron muestras de tejido corioamniótico de mujeres en trabajo de parto recibiendo atención obstétrica en el CHPR. Los tejidos corioamnióticos de parto pretérmino se colectaron de embarazos complicados por parto prematuro espontáneo antes de las 33 semanas de EG, mientras que los de parto a término se muestrearon de embarazos sin complicaciones que iniciaron trabajo de parto luego de las 37 semanas de edad gestacional.

Se consideraron criterios de exclusión para participar del estudio el tener un embarazo múltiple, ser consumidor de drogas, presencia de anomalías fetales, alumbramiento por cesárea y/o tener alguna de las siguientes condiciones médicas

preexistentes: diabetes, hipertensión, desórdenes autoinmunes, infección crónica (HIV, hepatitis). Se incluyeron a las pacientes del grupo control considerando que no hubieran tenido embarazos anteriores a pretérmino. Debido a la alta incidencia de la ruptura prematura de membranas (Rey et al., 2008; Vogel et al., 2018), no se excluyeron los pacientes con RPM de casos ni controles, y se incorporó esta característica como cofactor en los análisis estadísticos (ver abajo).

Se colectaron las características demográficas de las madres mediante entrevista con las pacientes luego del parto. Se obtuvieron los datos clínicos y obstétricos de la historia médica del Sistema de Información Perinatal (De Mucio et al., 2016).

4.2.2. ASPECTOS BIOÉTICOS

Todos los datos y muestras utilizados fueron tratados de forma anónima y denominados por un código que no identifica a la persona participante. Los pacientes participantes fueron informados de la finalidad y el alcance de este estudio. El procedimiento está aprobado por el Comité de Ética de la Facultad de Medicina, Universidad de la República (Consejo de la Facultad de Medicina, Exp. N° 071140-000907-11) y por el Comité de Ética del Hospital de la Mujer del CHPR. En todos los casos se obtuvo un consentimiento informado de la paciente, previo a la recolección del material biológico.

4.2.3. COLECTA DE LA MUESTRA

Las muestras de tejido corioamniótico se colectaron en un plazo de hasta 30 minutos luego del alumbramiento. El amnios y corion se obtuvieron de las membranas extraplacentarias, que proveen una fuente más limpia de las membranas fetales. Para cada paciente, se extrajo una porción de 1 cm² de membrana corioamniótica que rodea el sitio de ruptura de la misma, de acuerdo al protocolo descrito por Nhan-Chang et al. (2010). El amnios y el corion se

procesaron juntos. Se congeló el tejido inmediatamente en nitrógeno líquido o se lo sumergió en solución RNAlater (Qiagen, Hilden, Germany), para estabilizar y almacenar el ácido ribonucleico (ARN). Todas las muestras se almacenaron a -80°C hasta su procesamiento en el laboratorio.

4.2.4. EXTRACCIÓN DE ARN TOTAL

Para la extracción de ARN total, se optimizó un protocolo para asegurar buena calidad y concentración de ARNm. Para cada muestra, se molieron muestras de aproximadamente 50 mg de membrana corioamniótica con nitrógeno líquido en un mortero pre-enfriado. El polvo resultante se pasó a un tubo con 2 ml de reactivo Trizol (Invitrogen, Carlsbad, CA, Estados Unidos), mezclando por inversión y luego incubando 5 minutos a temperatura ambiente. Se centrifugó 10 minutos a 10.000xG a 4°C y se pasó el sobrenadante a un tubo limpio. Se agregó 200 µl de cloroformo, se incubó 2 minutos a temperatura ambiente y se centrifugó 15 minutos a 12.000xG a 4°C. Se traspasó la fase acuosa resultante a un tubo limpio, en el cual se le agregó 1 volumen de etanol 75%, mezclando por inversión. Se continuó el protocolo de extracción usando el kit RNeasy Mini (Qiagen, Hilden, Germany). Para ello se cargaron 700 µl de esta solución en una columna del kit, se centrifugó 15 segundos a 8.000xG a temperatura ambiente y se descartó el sobrenadante, según las instrucciones del fabricante. Este paso se repitió con el remanente de la solución tantas veces como necesario, cargando hasta 700 µl por vez. Se lavó la columna con 700 µl de buffer RW1 del kit, centrifugando 15 segundos a 8.000xG a temperatura ambiente y descartando el sobrenadante. Se realizaron 2 lavados más, cargando 500 µl de buffer RPE en la columna, y centrifugando 15 segundos en el primero y 2 minutos en el segundo, ambos a 8.000xG a temperatura ambiente y descartando el sobrenadante. Se colocó la columna RNeasy en un tubo de colecta y se cargó 50 µl de agua previamente tratada con DEPC (dietil pirocarbonato). Se centrifugó 1 minuto a 8.000xG para eluir el ARN.

Se evaluó la calidad y concentración del ARN total obtenido mediante espectrofotometría en Nanodrop 2000 (Nanodrop technologies Inc. Wilmington, DE, USA), midiendo la absorbancia a 260 y a 280 nm. Se evaluó la integridad del ARN en corridas electroforéticas y/o usando un equipo Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, Estados Unidos). Para ello se realizaron geles de agarosa 1%, con buffer de corrida TBE 0,5X, desnaturalizando las muestras 5 minutos a 60°C antes de cargarlas en el gel. Se incluyó en cada corrida 2 µl de ssRNA (#NO3625, New England Biolabs, Ipswich, MA, Estados Unidos), como marcador de peso molecular de ARN. Seleccionamos muestras de ARN que no mostraran degradación en los geles y/o que tuvieron Números de Integridad de ARN (RIN) 8 o mayores. Las muestras seleccionadas se enviaron al servicio de secuenciación MacroGen Inc. (Seúl, Corea del Sur) precipitadas según sus recomendaciones: se precipitaron las muestras de ARN en 0,1 volumen de Acetato de Sodio 3M y luego en 2 volúmenes de Etanol absoluto. Al llegar a MacroGen Inc., la integridad del ARN fue evaluada nuevamente en un Bioanalyzer 2100.

4.2.5. SECUENCIACIÓN MASIVA DE ARN

De las muestras que pasaron este control de calidad de ARN, se seleccionaron 4 muestras provenientes de partos a término y 4 muestras de partos a pretérmino severo para su secuenciación masiva de ARN. En la selección de muestras se incluyeron datos maternos (edad materna, paridad, etc) intentando parear las características de casos y controles. La preparación y secuenciación de la muestra fueron realizados por MacroGen Inc. En cada muestra, el ARNm fue purificado a partir del ARN total con el kit PolyAtract mRNA Isolation System II (Promega Inc., Madison, WI, Estados Unidos) y copiado a moléculas de ADN copia (ADNc) empleando el kit TruSeq RNA Sample Preparation Kit v2 (Illumina Inc.; San Diego, CA, Estados Unidos) para construir una biblioteca de ADNc por muestra. Las librerías fueron luego secuenciadas de forma masiva según el protocolo para la plataforma Illumina HiSeq2000 (Illumina Inc.), secuenciando *lecturas* de 100

pares de bases (pb) secuenciando cada fragmento por sus dos extremos (“*paired-end*”). Los datos crudos se depositaron en la base de datos NCBI’s Short Read Archive a la cual se puede acceder a través del código SRP139931.

4.2.6. LIMPIEZA Y FILTRADO DE LAS SECUENCIAS OBTENIDAS

El flujo de trabajo del análisis de expresión diferencial se resume en la Figura 4.1. La calidad de los resultados de la secuenciación en sus archivos en formato Fastq fueron evaluados con el programa FastQC (Andrews, 2010). Se usó el programa Trimmomatic 0.32 (Bolger et al., 2014) para remover las secuencias de adaptadores usados en la secuenciación, removiendo lecturas de menos de 36pb y reteniendo sólo aquellas bases en el extremo 3’ que tengan un score de calidad Phred mayor a 30. Una vez filtrados, los archivos se volvieron a analizar en el programa FastQC para evaluar su calidad. Se utilizó el programa MultiQC (Ewels et al., 2016) para resumir los resultados obtenidos.

4.2.7. ALINEAMIENTO DE LAS SECUENCIAS AL GENOMA DE REFERENCIA Y CONTEO DE LECTURAS

Para los siguientes análisis se usó como genoma humano de referencia el GRCh37/hg19 de Ensembl (Yates et al., 2016), versión 75. Se alinearon las secuencias filtradas de cada muestra al genoma de referencia mediante al programa de alineamiento TopHat 2.1.0 (Trapnell et al., 2009). Este programa de alineamiento es el recomendado para ARN-seq al alinear contra genomas, por considerar el corte y empalme del genoma para alinear (“*splice-aware*”). Se estableció una distancia esperada entre pares de lecturas de 160 pb y se usó la anotación genómica de Ensembl para la versión GRCh37/hg19. El resto de los parámetros fueron corridos con sus valores por defecto. Nuevamente se utilizó el

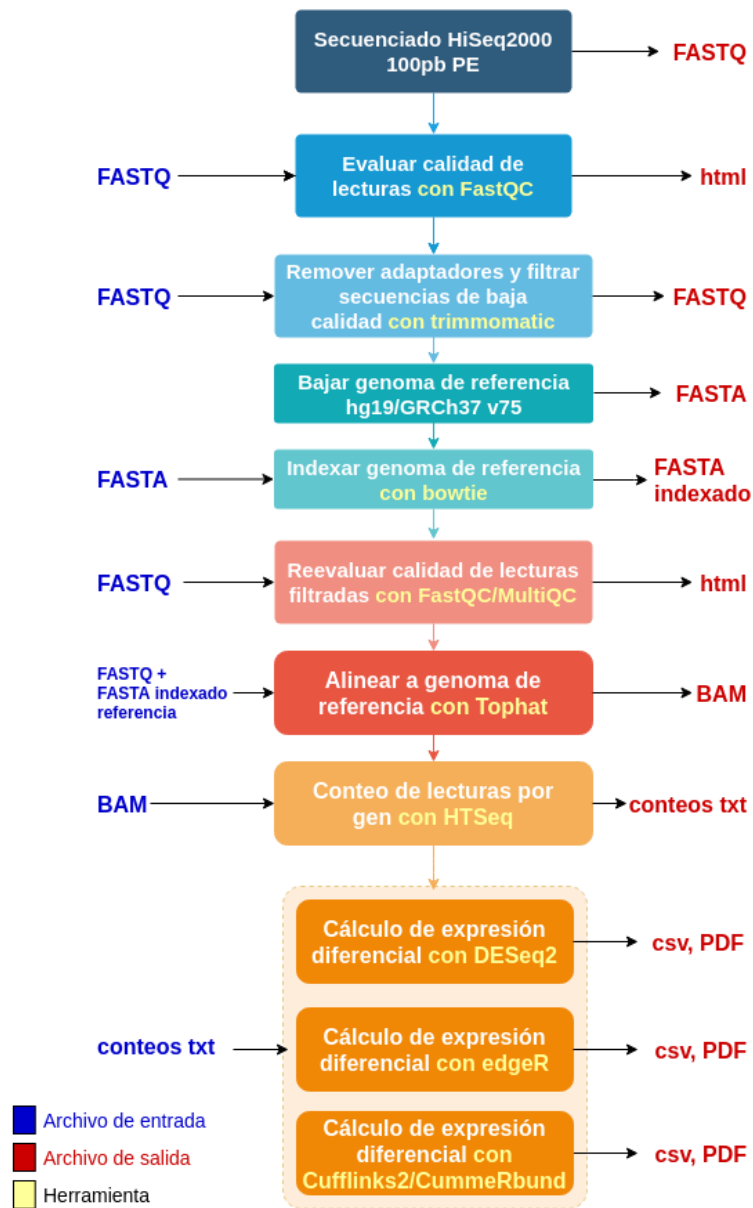


Figura 4.1. Flujo de trabajo de análisis informático de expresión diferencial.

programa MultiQC v0.8 (Ewels et al., 2016) para resumir las estadísticas de los resultados obtenidos.

Con los archivos de lecturas alineadas se realizó el conteo de las lecturas asignándolas a genes, con el paquete de Python HTSeq (Anders et al., 2015). Se

usó el modo “unión” para asignar las lecturas, según el cual no es necesario que una lectura esté alineada en toda su longitud a un gen para asignarla al mismo. En el caso que una lectura se alinee con más de un gen, no se consideró el alineamiento (opción `--nonunique none`). Este programa arroja una matriz de conteo de lecturas para cada gen y cada muestra, el cual es una buena aproximación de la abundancia del transcripto.

Se calcularon los estadísticos de los resultados del alineamiento con la herramienta CollectRnaSeqMetrics de Picard (<http://broadinstitute.github.io/picard/>).

4.2.8. EXPRESIÓN DIFERENCIAL

Se utilizaron 3 metodologías diferentes para estudiar la expresión diferencial de genes entre el grupo pretérmino severo y el grupo a término: el programa Cufflinks2 v2.2.1 (Trapnell et al., 2010) y los paquetes de R Differential Expression analysis for Sequence count data (DESeq2) v1.16.1 (Love et al., 2014) y Empirical analysis of digital gene expression in R (edgeR) v3.18.1 (Robinson et al., 2010). Para analizar los casos y controles se usó un modelo lineal con la condición pretérmino y con la RPM como cofactor. Todos los análisis estadísticos fueron realizados utilizando R environment for statistical computing version 3.4.1 (<https://www.R-project.org/>). Se utilizaron paquetes adicionales de Bioconductor (Huber et al., 2015) para graficar ('ggplot2' v2.2.1) y realizar agrupamientos y heatmaps ('pheatmap' v1.0.8).

Cufflinks2 es un programa que ensambla lecturas de ARN en transcriptos, cuantifica abundancia de transcriptos y testea expresión diferencial de genes. Cufflinks2 mide las abundancias de transcriptos en Fragmentos Por Kilobase de exón por cada Millón de fragmentos mapeados (FPKM). Esta medida es análoga a las Lecturas Por Kilobase de exon por cada Millón de fragmentos mapeados (RPKM) para lecturas simples (“single-reads”). RPKM es un método para cuantificar la expresión génica de los datos de ARN-seq usando la normalización

del largo total de lecturas y el número de lecturas secuenciadas (Mortazavi et al., 2008). Se calcula sumando las lecturas que se alinean a exones y exones candidatos y normalizándolas por el largo del ARNm esperado.

El análisis con Cufflinks2 v2.2.1 se realizó siguiendo parte del protocolo de trabajo sugerido en Trapnell et al. (2012). Brevemente, las lecturas alineadas se ensamblaron en transcriptos usando *cufflinks2* y se unieron con *cuffmerge* para crear una anotación única del transcriptoma. La expresión diferencial se estimó utilizando *cuffdiff* y los resultados se analizaron en el entorno del programa R v3.4.3, usando el paquete CummeRbund v2.20.0 (Goff et al., 2013).

Para el análisis de DESeq2 y de edgeR se provee como insumo la matriz de conteo de lecturas realizada por el programa HTSeq. Para el análisis con DESeq2 se realizó un pre-filtrado de datos, eliminando los genes sin ninguna lectura. En todos los casos se usó como grupo de referencia al grupo de parto a término. En el diseño se consideró además si la muestra presentó RPM como covariable, para aumentar la sensibilidad para encontrar diferencias debidas a la condición (prematureo severo o control a término). Así, se definió un modelo multifactorial considerando la condición (pretérmino severo o control a término) y la presencia de RPM en la muestra.

Para evaluar la expresión diferencial se utilizaron los conteos crudos de genes, pero para la visualización y agrupamiento (“*clustering*”) se realizó la transformación de los conteos de lecturas mediante el logaritmo regularizado (*rlog*) (Love et al., 2014). Esta transformación permite remover la dependencia de la varianza de la media, en particular varianzas grandes de los logaritmos de conteos cuando la media es baja.

Para el análisis de expresión diferencial se usan los conteos crudos y se define como criterio de corte arbitrario tener una tasa de descubrimientos falsos (FDR)<0,05 y tener al menos un conteo de 5 lecturas alineadas. El modelo usado por DESeq2 usa un modelo lineal generalizado donde los conteos para cada gen y muestra son ajustados a una distribución binomial negativa, de la cual se evalúa si

son significativamente diferentes entre casos y controles según el test de Wald (Love et al., 2014). Asume entonces una relación no lineal entre la varianza y la media de los niveles de expresión, lo que permite que la varianza ajustada usando datos de genes con niveles de expresión similares. Eso es útil porque un bajo número de réplicas hacen que la estimación de la varianza sea difícil cuando se usa únicamente datos disponibles para un único gen.

El paquete edgeR también usa un modelo de distribución binomial negativa para incorporar la variabilidad biológica y técnica. Usa luego métodos bayesianos empíricos para estimar el grado de dispersión a través de los transcriptos, lo que permite la estimación de variación biológica para cada gen, incluso para experimentos con bajos niveles de réplicas biológicas. Luego, se estima la expresión diferencial entre casos y controles para cada gen usando un test exacto análogo a un test exacto de Fisher, pero adaptado a datos sobre-dispersos. Una vez que se obtuvo la lista de genes diferencialmente expresados, como un objeto DGEList en R, se filtraron los datos para remover conteos bajos, como se había hecho en el análisis de DESeq2. En este caso se eliminaron los genes con lecturas que tuvieron menos de 10 conteos por millón (CPM) por librería. Para normalizar la composición de ARN total en las muestras, se usó la función `calcNormFactors` de edgeR. Esta normalización es especialmente importante cuando hay un pequeño número de genes con expresión muy alta en sólo unas muestras. Los genes de muy alta expresión consumirían gran parte del secuenciado de la librería, y consecuentemente los genes restantes quedarían submuestreados. Entonces, si no se ajusta por composición del ARN, estos genes restantes serían categorizados como subexpresados en esas muestras.

Para el análisis de expresión diferencial en edgeR se ajustó una distribución binomial negativa a los conteos de lecturas para cada gen, mediante el test de cociente de verosimilitud (likelihood ratio test, LRT).

Los valores p de significancia se ajustaron para pruebas múltiples usando la aproximación de Benjamini y Hochberg ("BH") (Benjamini & Hochberg, 1995)

para controlar la FDR. Se consideraron los genes como expresados si tuvieron como mínimo 5 lecturas alineadas. En cada uno de los 3 métodos, se identificó un gen expresado como diferencialmente entre casos y controles si cumplían lo siguiente: un valor absoluto de logaritmo en base 2 de la tasa de cambio ($\log_2FC$) mayor a 2 y un valor p ajustado por hipótesis múltiples (BH) FDR menor a 0,05. Adicionalmente, para reducir los potenciales falsos negativos, se requirió que un gen fuera identificado como significativo por los tres métodos para ser considerado como diferencialmente expresado.

Por otro lado, para analizar si la RPM tiene un efecto en la expresión génica, se realizó de la misma manera un análisis complementario de expresión génica contrastando tejidos con RPM (n=3) y sin RPM (n=5), usando prematurez como cofactor.

Adicionalmente, se investigó si los genes identificados como genes DE para pretérmino severo habían sido previamente asociados con PPT. Se generó una lista explorando bases de datos públicas y asociaciones encontradas en meta-análisis entre estos genes y PPT. Estos datos incluyeron una base de datos de asociaciones entre SNPs y PPT (Uzun et al., 2012), dos meta-análisis de estudios transcriptómicos de varios tejidos de PPTs (Eidem et al., 2015; Vora et al., 2018), un estudio transcriptómico en corioamnios (Bukowski et al., 2017) y datos de expresión génica públicos (Heng et al., 2014; Kim et al., 2016; Lee et al., 2013; Ngo et al., 2018).

4.2.9. ANÁLISIS DE TÉRMINOS DE ONTOLOGÍA GÉNICA

Para entender los procesos biológicos subyacentes, se realizaron anotaciones funcionales de los genes diferencialmente expresados mediante Ontología de genes (GO). Se usaron los paquetes GAGE v 2.26.3 (Luo et al., 2009) y GOSec v1.28.0 (Young et al., 2010) de Bioconductor 3.5. GOSec considera el efecto del sesgo de selección en los datos de ARN-seq que puede aparecer debido a diferencias en el

largo de los genes. Cuando fue necesario, los genes seleccionados fueron mapeados usando el paquete de genes biomaRt (Durinck et al., 2009). Buscamos enriquecimiento de asociaciones génicas con vías de KEGG y términos de GO. Se corrigió la significancia por pruebas múltiples usando la aproximación de BH, y se consideró que había enriquecimiento si el FDR era menor a 0,05.

4.2.10. VALIDACIÓN DE RESULTADOS DE ARN-SEQ POR PCR EN TIEMPO REAL

Se seleccionaron seis genes basados en su mayor valor de log2FC y menor valor p , para validar los datos de expresión génica por una técnica independiente: PCR en tiempo real cuantitativa (qRT-PCR). En la Tabla 4.7 se presentan el detalle de los genes seleccionados. Se usaron como genes de referencia los genes housekeeping Proteína de unión a caja TATA (*TBP*) (ID Ensembl ENSG00000112592) y Gliceraldehído-3-fosfato deshidrogenasa (*GAPDH*) (ID Ensembl ENSG00000111640).

Tabla 4.1. Genes DE seleccionados para la validación de expresión génica diferencial.

Símbolo de gen	Nombre	Ubicación cromosómica	ID gen Ensembl	HGNC ID
<i>CASP5</i>	caspase 5	11q22.3	ENSG00000137757	HGNC:1506
<i>IL1β</i>	interleukin 1 beta	2q14.1	ENSG00000125538	HGNC:5992
<i>LCN2</i>	lipocalin 2	9q34.11	ENSG00000148346	HGNC:6526
<i>MARCO</i>	macrophage receptor with collagenous structure	2q14.2	ENSG0000019169	HGNC:6895
<i>SERPINA1</i>	serpin family A member 1	14q32.13	ENSG00000197249	HGNC:8941
<i>TNFSF15</i>	TNF superfamily member 15	9q32	ENSG00000181634	HGNC:11931

Se usó 1µg de ARN total de cada muestra para sintetizar la primera hebra de ADNc usando el kit de Superscript II RT (Invitrogen, Carlsbad, CA, Estados Unidos) y oligonucleótidos al azar (*random primers*, 0,5µg/µl). Las reacciones de qRT-PCR se realizaron en un termociclador de tiempo real Corbett (Qiagen, Hilden, Germany), usando un kit de SYBER Green Biotools (Biotools, Madrid, España). La expresión relativa de un gen blanco se calculó según el método 2-DDCt (Livak & Schmittgen, 2001), usando la media geométrica de los genes *GADPH* y *TBP* como gen de referencia, usando el grupo de pacientes a término como grupo control.

Cada reacción se realizó con 10 µl de kit QUANTIMIX EASY kit (Biotools; #10606-4153), 0,5 µM de cada oligonucleótido directo y reverso, 1 µl de ADNc de la muestra, y agua destilada hasta llegar a un volumen final de 20 µl. Se usaron las siguientes condiciones de amplificación: 95°C por 5 minutos y 40 ciclos a 95°C por 15 segundos, seguido de 1 minuto a 63°C. Luego de la amplificación se realizó un paso de desnaturalización mediante el aumento de temperatura de 72 a 90°C con una adquisición de fluorescencia continua. En todas las reacciones se incluyó un control positivo y negativo (sin ADN molde). Cada muestra se corrió por duplicado y el coeficiente de variación entre duplicados fue menor a 2%. Los oligonucleótidos fueron diseñados para incluir uniones exón-exón, para reducir la probabilidad de amplificar ADN genómico (Tabla 4.8). Todos fueron diseñados en nuestro laboratorio excepto para los genes *IL1β*, *GAPDH* (Wang et al., 2002b) y *TBP* (C. Chiale, comunicación personal). Se evaluó la eficiencia de los oligonucleótidos para verificar que fuera entre 90-110%, y su especificidad se evaluó revisando que las curvas de desnaturalización presentarán un único pico de disociación.

Tabla 4.2. Oligonucleótidos utilizados para las reacciones de qRT-PCR.

Símbolo de gen	Oligonucleótido directo 5' → 3'	Oligonucleótido reverso 5' → 3'	Largo del amplicón
<i>CACNA1E</i>	TGCTGTGCTGACTGTCTTCC	CAATTCCAGGTGGCTCCTAA	93 pb
<i>CASP5</i>	CAAGGCCCTGATCTTGGTAG	TCTGCTGACTCAGGTGGTCC	144 pb
<i>IL1β</i>	CTCGCCAGTGAAATGATGGCT	GTCGGAGATTTCGTAGCTGGAT	144 pb
<i>IL10</i>	GCTGTCATCGATTTCTTCCC	CTCATGGCTTTGTAGATGCCT	103 pb
<i>KCNJ16</i>	ATTCTCTACCATTGCCCAGC	TGCCAGTCTGCCAGTTAGAA	131 pb
<i>LCN2</i>	CAAAGACCCGCAAAAGATGT	GCAACCTGGAACAAAAGTCC	127 pb
<i>MARCO</i>	TGGAGCCTTTTCGAAATCAAT	CGCCTGCAGATTCAGAACTT	140 pb
<i>SERPINA1</i>	CCCTCTGGATCCACTGCTT	ACGAGACAGAAGACGGCATT	111 pb
<i>TNFSF15</i>	CACATACCTGCTTGTCAGCC	TGTGAAGGTGCAAACTCCTG	90 pb
<i>TBP</i>	TGAGCCAGAGTTATTTCTTGG	AGGGTAGATGTTTTCAAATGCTT	142 pb
<i>GADPH</i>	GGGAAGGTGAAGGTCGG	TGGACTCCACGACGTACTCAG	292 pb

4.3 RESULTADOS

4.3.1. COLECTA DE MUESTRAS Y SECUENCIADO DE ARN

En el muestreo de membranas corioamnióticas de partos a término y pretérmino severo recolectamos un total de 36 membranas fetales. El método de colecta que dio mejores resultados para mantener la integridad del ARN fue la conservación de la membrana fetal en RNAlater® (Qiagen) y almacenamiento a 4°C hasta la extracción de ARN. Se extrajo ARN de aquellas membranas que estaban en condiciones de calidad adecuadas, lográndose ARN de 24 pacientes en total: 15 controles y 9 pretérminos severos. Las características de las muestras utilizadas en este estudio se presentan en la Tabla 4.3. De ellos, seleccionamos 8 para proseguir al análisis de secuenciado masivo de ARN: 4 prematuros severos y 4 controles a término (Tabla 4.4). Estas 8 muestras fueron seleccionadas de forma de tener

grupos caso y control pareados en cuanto a características demográficas de las madres. Consecuentemente, existen diferencias significativas entre casos y controles en edad gestacional, peso al nacer y el número de consultas prenatales, mientras que el resto de las características relevadas no presentaron diferencias significativas entre los grupos (Tabla 4.3).

Todos los controles tuvieron terminación del parto espontánea por vía vaginal. De los prematuros, todos excepto 3 de los casos tuvieron terminación vaginal espontánea. En esos 3 casos de los prematuros severos, la terminación fue por cesárea luego de que el trabajo de parto estaba iniciado. Las 8 muestras de ARN seleccionadas para el secuenciado masivo de ARN tuvieron terminación del parto vaginal y espontánea. Estas 8 muestras tuvieron un rango de la edad gestacional de 38 a 40 semanas para los controles, mientras que para los casos fue de 26 a 33 semanas.

Las extracciones de ARN de estas 8 muestras tuvieron valores de RIN entre 6,1 – 8,5 (promedio 7,5). La secuenciación masiva logró un promedio de 4,89 Gigabases (GB) resultantes por muestra (rango 3,98 – 6,13 GB), y en promedio un 50,2% de contenido GC por muestra (rango 46,57 – 51,54%). Se obtuvo un promedio de 47.841.094 lecturas por muestra (Tabla 4.5). Luego del filtrado de lecturas se removieron en promedio un 3,87% de las lecturas.

El alineamiento de las secuencias obtenidas mediante TopHat2 tuvo un alto porcentaje de alineamiento, con un promedio de 93,66% ($\pm 5,61$) de lecturas alineadas. Los estadísticos del alineamiento se presentan en la Tabla 4.5. Cuando se estudia la ubicación genómica en la cual se alinea cada lectura, se ve que la mayoría se alinea en regiones ribosomales o codificantes (exónicas) (Figura 4.2).

Tabla 4.3. Características demográficas y clínicas de los grupos término y pretérmino severo.

	Parto a término	Parto pretérmino severo	
	Total (n=15)	Total (n=9)	Valor p^a
Edad materna, años $\pm$ SE	24,1 $\pm$ 1,6	20,6 $\pm$ 1,0	0,146
Edad gestacional, semanas $\pm$ SE	39,3 $\pm$ 0,3	30,7 $\pm$ 0,9	1,9 x 10 ⁻¹⁰
Nivel educativo: primaria o menos, n (%)	9 (60,0)	3 (33,3)	0,399
Estado civil, % madres solteras	13,3	22,2	0,472
RPM (%)	4 (26,7)	2 (22,2)	1
Preeclampsia (%)	0	1 (11,1)	0,238
RCIU (%)	0	1 (11,1)	0,238
Paridad, n (%)			0,335
Nulípara	5 (33,3)	4 (44,4)	
≥ 2	5 (33,3)	2 (22,2)	
Peso del recién nacido (g), Promedio $\pm$ SE	3492 $\pm$ 120	1696 $\pm$ 205	1,6 x 10 ⁻⁰⁷
Sexo recién nacido, % femenino	47	55	0,472
Consultas prenatales $\pm$ SE	8,8 $\pm$ 0,8	4,3 $\pm$ 0,9	0,002

^aLa significancia se evaluó con test t de Student o de Chi-cuadrado, según correspondió.
RPM: ruptura prematura de membranas, RCIU: retraso del crecimiento intrauterino, SE: error estándar.

Tabla 4.4. Características clínicas y demográficas de los 4 controles (C01, C02, C03 y C04) y 4 pacientes pretérminos severos (P01, P02, P03 y P04) analizados por ARN-seq.

Muestra	Edad gestacional (semanas)	Peso (gramos)	Sexo recién nacido	RPM	Edad materna (años)	Educación (mayor nivel)	Consultas prenatales	Estado civil (soltera)	Gravidez	Paridad	Fumadora	Tipo sanguíneo	Anemia
C01	38	3135	Masculino	No	18	Secundaria	6	Si	1	Nulípara	No	O Rh+	Si
C02	40	3300	Femenino	No	31	Primaria	5	No	4	Múltipara	Si	NA	NA
C03	37	3295	Femenino	Si	20	Primaria	8	No	2	Primípara	No	A Rh+	No
C04	40	3460	Masculino	Si	19	Primaria	5	No	1	Nulípara	No	O Rh+	No
P01	26	940	Masculino	No	16	Secundaria	4	No	1	Nulípara	No	O Rh+	No
P02	28	1100	Masculino	Si	22	Secundaria	5	Si	1	Nulípara	No	O Rh+	No
P03	32	2175	Femenino	No	25	Primaria	2	Si	3	Múltipara	No	A Rh+	No
P04	33	1950	Femenino	No	19	Primaria	1	No	2	Primípara	Si	O Rh+	Si

Tabla 4.5. Estadísticos del alineamiento de secuencias contra el genoma de referencia.

Muestra	Lecturas secuenciadas totales	Lecturas alineadas	Porcentaje de lecturas alineadas	Lecturas alineadas de forma única	Porcentaje de lecturas alineadas de forma única
C01	59.547.019	55.351.967	92,96	54.266.450	91,13
C02	41.730.621	40.612.060	97,32	39.799.483	95,37
C03	46.097.365	44.827.668	97,25	43.988.573	95,43
C04	39.886.026	38.714.828	97,06	37.956.434	95,16
P01	38.550.493	31.441.879	81,56	30.154.534	78,22
P02	42.545.119	41.238.702	96,93	40.376.868	94,9
P03	58.187.494	56.282.003	96,73	54.724.954	94,05
P04	48.337.388	43.267.353	89,51	42.323.079	87,56

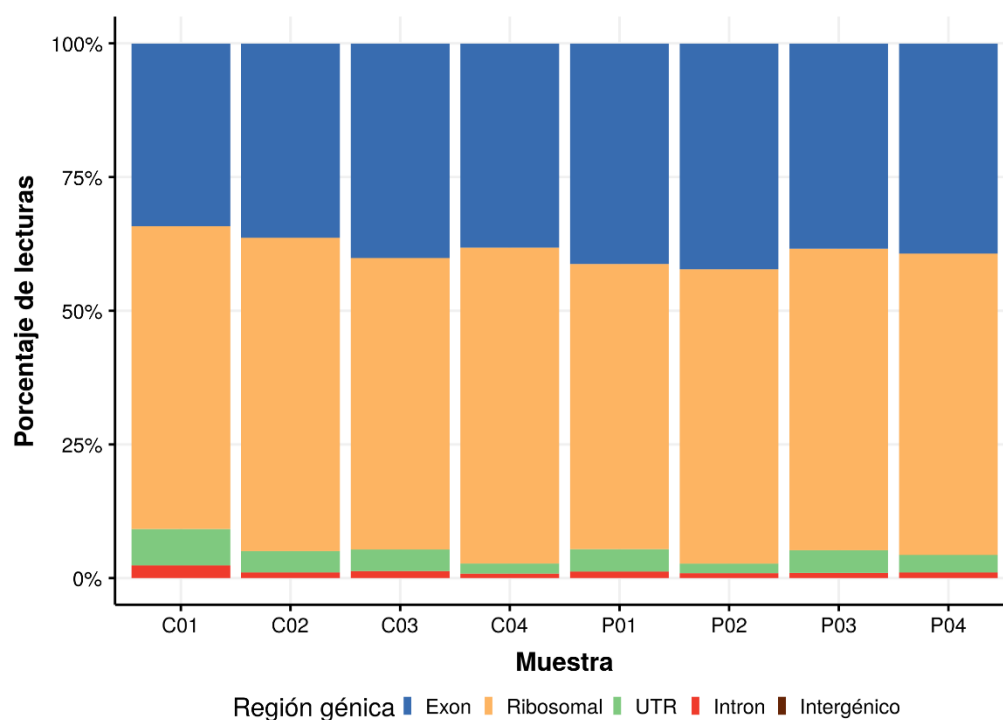


Figura 4.2. Porcentaje de las lecturas alineadas distribuidas según su localización genómica para las muestras secuenciadas.

4.3.2. ANÁLISIS DE EXPRESIÓN DIFERENCIAL

Los análisis de expresión diferencial se realizaron con 3 paquetes diferentes.

Para el análisis de expresión diferencial se usó como insumo la tabla de conteo de lecturas por gen producida por el programa HTSeq. El programa encontró que las secuencias alinearon contra 32.731 genes con al menos una lectura cada uno. Se realizó un análisis de componentes principales sobre los conteos de lecturas normalizados por el método *rlog* (Figura 4.3A). Este método ayuda a reducir las dimensiones de conjuntos de datos para determinar características claves en datos de gran escala. En datos de ARN-seq, el PCA agrupa muestras de a grupos según los genes más significativamente desregulados. El método también permite detectar efectos de lote (conocido como “batch effects”), que pueden deberse a condiciones de laboratorio, lotes de reactivos y/o manipulación (Leek et al., 2010).

En nuestros datos, los primeros dos componentes principales del modelo contuvieron la gran mayoría de la varianza: 64 y 22%. El análisis muestra que las muestras de partos a término se forman un grupo compacto, mientras que las de parto pretérmino severo muestran gran variabilidad. El análisis de agrupamiento jerárquico basado en la distancia euclidiana entre la matriz de conteo de lecturas normalizadas por *rlog* muestra una situación similar, con los controles como un grupo homogéneo y compacto (Figura 4.3B). No se observaron efectos de lote en las muestras, por lo que no fue necesario controlarlos en los análisis subsiguientes.

Se realizaron tres análisis de expresión diferencial independientes, con los programas DESeq2, edgeR y cufflinks2. El análisis con el programa DESeq2 encontró que de los 18.675 genes analizados (*i.e.* con conteos con más de 5 lecturas), 1.436 genes estaban diferencialmente expresados entre casos y pretérminos severos, con un valor *p* ajustado por BH menor a 0,05. Los 15 genes más significativamente sobre y subexpresados por DESeq2 se presentan en la Figura 4.4. De los 1.436 genes diferencialmente expresados, 1.154 genes tienen

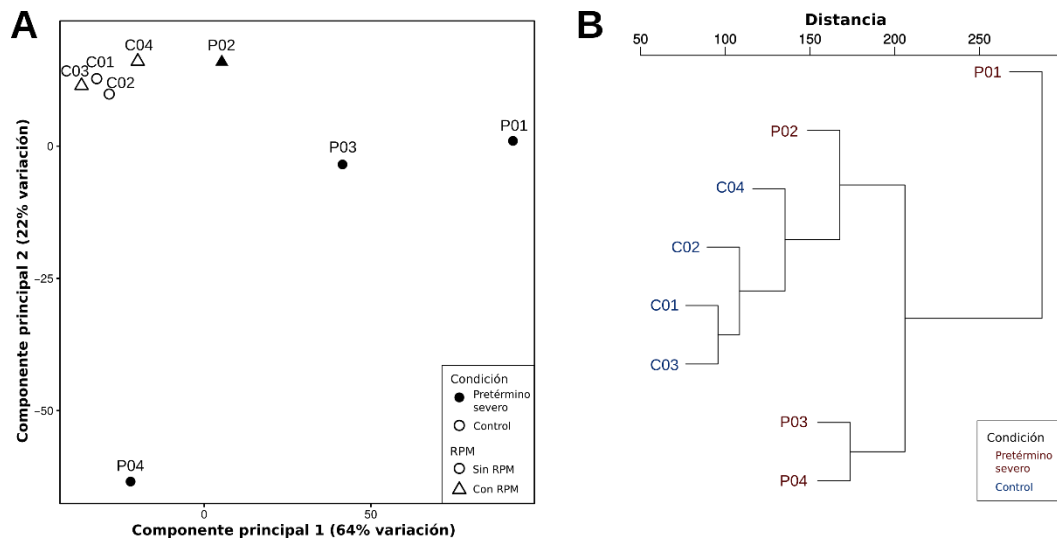


Figura 4.3. (A) Gráfico de análisis de componentes principales (PCA), señalando prematuridad y RPM. **(B)** Dendrograma de las muestras a partir de las distancias euclidianas de los conteos normalizados por *rlog*.

además una tasa de cambio logarítmica absoluta mayor a 2: siendo 935 genes sobreexpresados en prematuros severos ($\log_2FC > 2$) y 219 subexpresados ($\log_2FC < -2$). Su distribución se presenta en un gráfico de volcán (Figura 4.5), que combina el nivel de significancia estadística y la tasa de cambio de los genes respecto al control. Consideramos como genes regulados de forma diferencial a aquellos que tienen un valor de $p < 0,05$ y un $\log_2FC$ absoluto mayor a 2.

Se realizó un gráfico MA (Dudoit et al., 2002), que permite ver la distribución de los coeficientes estimados en el modelo para todos los genes (Figura 4.6). En el eje y se grafica el logaritmo de la razón entre los conteos para cada gen entre casos y controles (“M”) y en el eje x, el promedio de conteos para cada gen para todas las muestras (“A”). En este gráfico, genes con niveles de expresión similares entre los dos grupos aparecen cerca de la línea horizontal e igual a 0. Los genes con menor ajuste (en rojo) son los que tienen un valor p menor a 0,05, y representan posibles tendencias con respecto a la expresión génica. Aquí no se observan tendencias

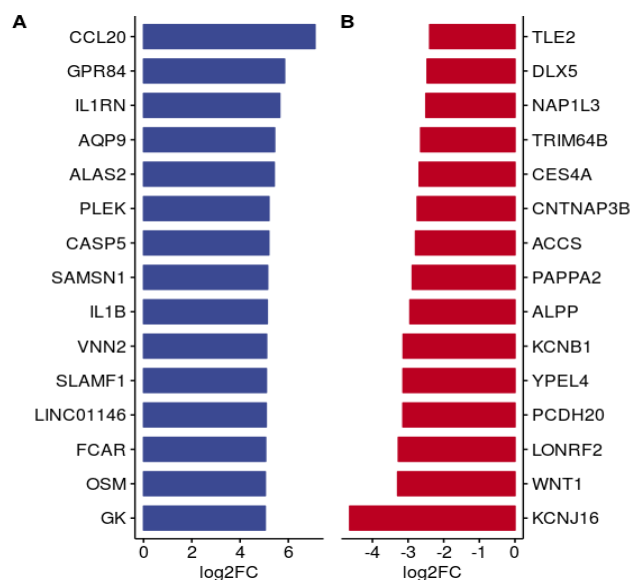


Figura 4.4. Los 15 genes más significativamente sobre- (A) y sub- (B) expresados en parto pretérmino severo según su valor p según DESeq2 ajustado por BH.

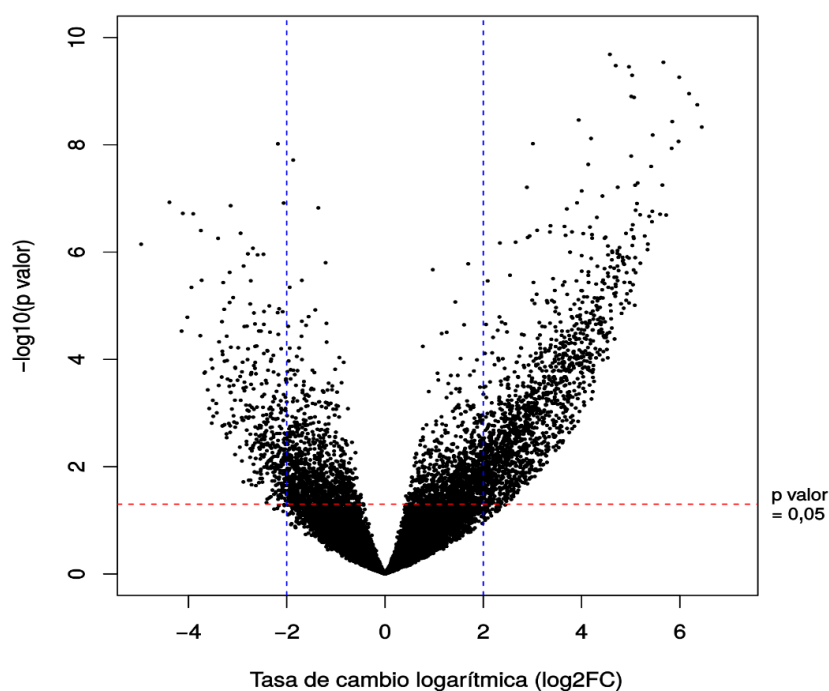


Figura 4.5. Grafico de volcán con los genes analizados en el DESeq2, mostrando la selección de genes expresados diferencialmente de forma significativa. Los puntos arriba de la línea roja muestran los genes significativos (valor $p < 0,05$) y las líneas azules muestran valores de $\log_2FC$ de 2 y -2.

anormales en los datos, indicando que la normalización realizada a los datos fue apropiada.

Como un control diagnóstico del análisis, se graficó también la distribución de los valores p obtenidos, y vemos que siguen una distribución uniforme entre 0 y 1 (Figura 4.7), sugiriendo que no hay problemas con la distribución de los datos.

El paquete edgeR, analizando el conteo de genes realizado con HTSeq, encontró 820 genes diferencialmente expresados entre casos y controles (p ajustado $< 0,05$, $|\log_2FC| > 2$). Cuando se realiza un análisis de escalamiento multidimensional (MDS), nuevamente se observa un agrupamiento compacto de controles y mayor dispersión en los pretérminos severos (Figura 4.8). El MDS se realiza usando el promedio de las mayores tasas de cambio logarítmicas absolutas entre cada par de muestras.

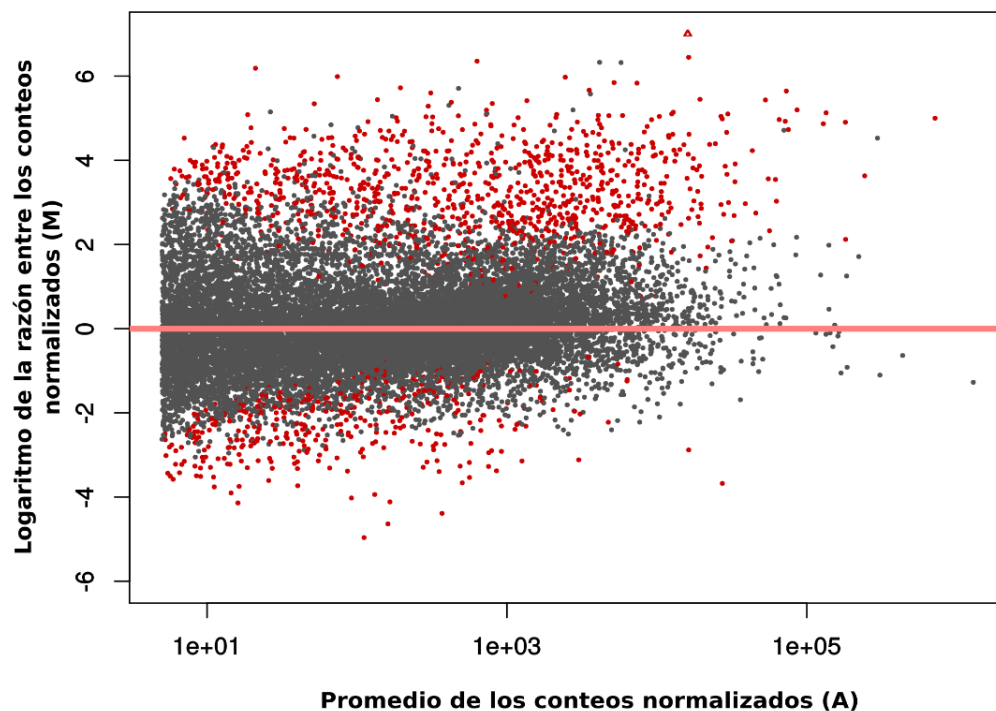


Figura 4.6. Visualización de los datos según la tasa de cambio y el promedio de los conteos de lecturas. Los puntos en rojo representan genes con $p < 0,05$.

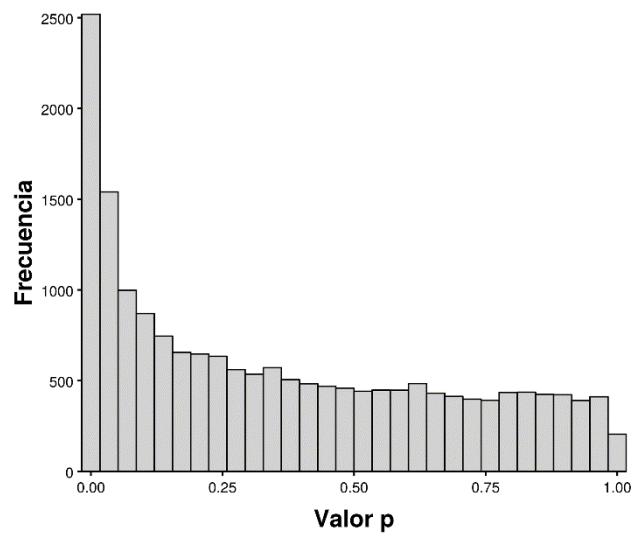


Figura 4.7. Histograma de los valores de significancia obtenidos para los análisis de expresión diferencial con DESeq2.

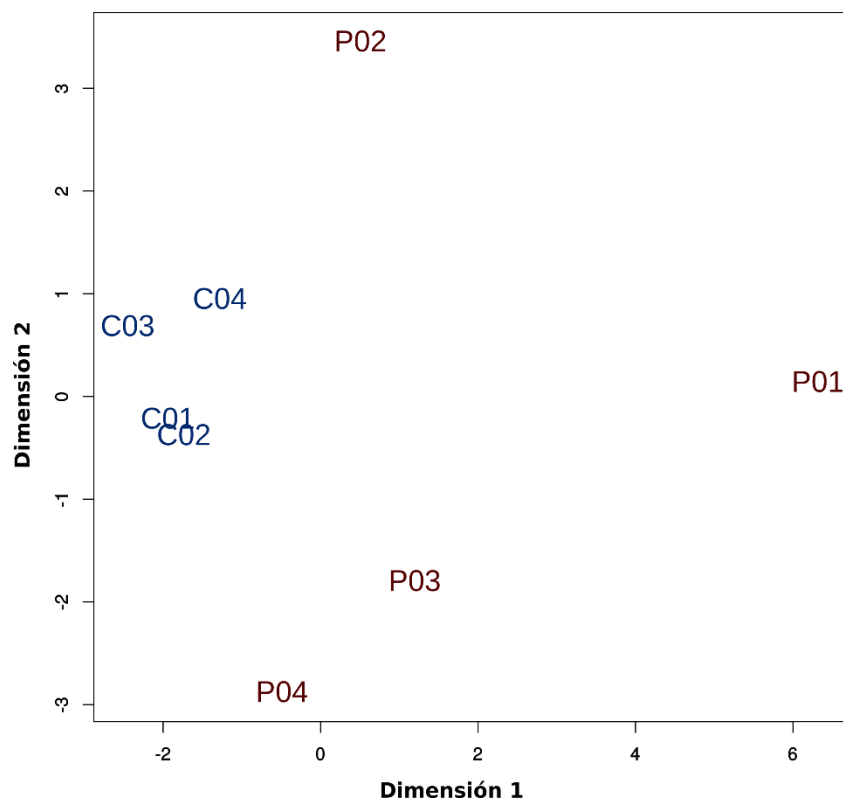


Figura 4.8. Gráfico de escalamiento multidimensional. Las muestras del grupo control están en azul y las de pretérmino severo en rojo.

Paralelamente, los análisis con el paquete Cufflinks2 se realizaron también a partir del alineamiento realizado con el TopHat2, usando los programas *cufflinks2*, *cuffmerge* y *cuffdiff* para el ensamblado y cuantificación de transcriptomas, la fusión de los mismos en un único archivo y el análisis de expresión diferencial de genes y transcriptos, respectivamente.

Para evaluar la calidad de los datos de ARN-seq es útil el cuadrado del coeficiente de variación (CV^2), una medida normalizada de la variabilidad entre los grupos analizados. En nuestro caso, el grupo de pretérmino severo tiene mayor variabilidad (*i.e.* CV^2) que el de términos, concordantemente con resultados de otros programas (Figura 4.9). Esta diferencia en la variación entre los dos grupos se ve tanto a nivel de genes como a nivel de isoformas.

Comparando los resultados de los 3 métodos, vemos que confluyen en 270 genes que son reportados como DE entre casos y controles ($p < 0,05$ y $|\log_2FC| > 2$) (Figura 4.10). Entre ellos, 252 fueron sobre expresados y 18 fueron subexpresados en PPT severo en relación con parto a término. La lista completa de genes DE detectada se presenta en la Tabla suplementaria 1-a. También se presentan referencias a estudios previos que encontraron asociación entre PPT y genes DE encontrados en este trabajo (Tabla suplementaria 1-d). Aproximadamente el 50% de los genes DE fueron previamente reportados como asociados al PPT.

En un heatmap con los valores normalizados por *rlog* de los conteos para cada uno de los 270 genes DE se puede observar su patrón de expresión (Figura 4.11). Un análisis de agrupamiento jerárquico de los 270 genes DE los agrupa en 3 grandes clusters. Este agrupamiento usa el método completo y distancias euclidianas. De los 3 grupos, el grupo más pequeño tiene un patrón discordante entre controles y pretérminos severos, y por tanto constituyen un grupo de genes de interés para estudiar el parto prematuro. No hubo patrón de expresión distintivo asociado con RPM. Un agrupamiento jerárquico de muestras no agrupa los grupos controles y pretérmino severo cuando se consideran los 270 genes DE. Tampoco los agrupa por si presentan o no RPM.

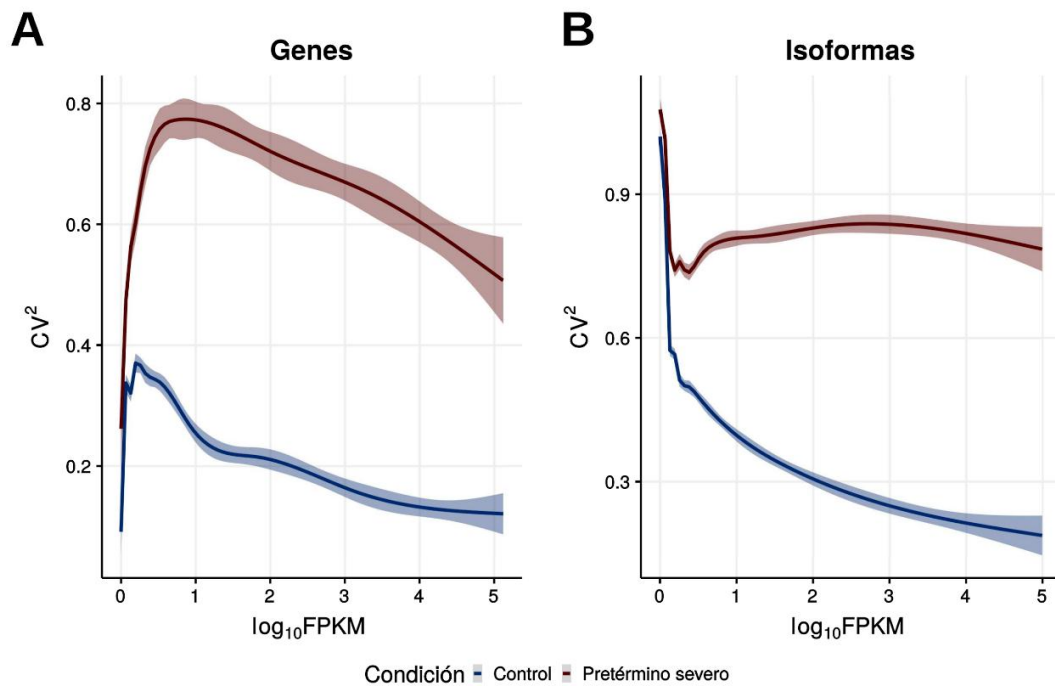


Figura 4.9. Cuadrado del coeficiente de variación (CV^2) entre genes (A) y entre isoformas (B), en función del logaritmo en base 10 de FPKM.

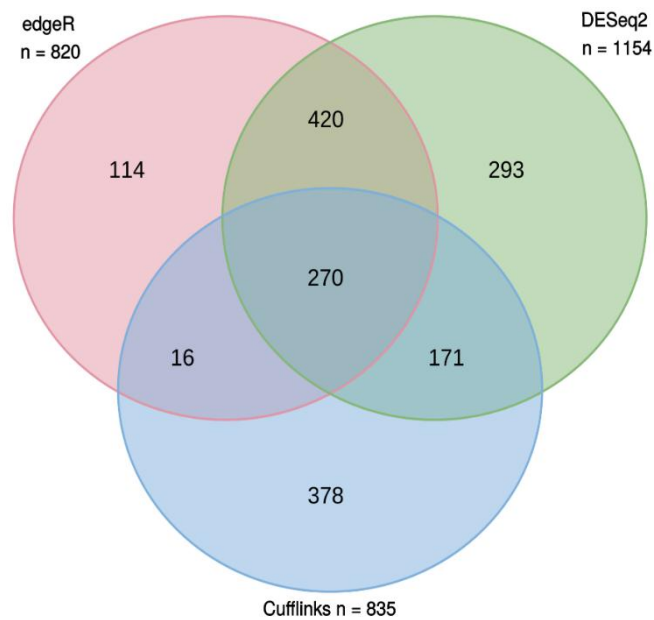


Figura 4.10. Gráfico de Venn con la intersección de los genes reportados como DE por los métodos DESeq2, edgeR y Cufflinks2.

El grupo más pequeño, en rojo en la Figura 4.11, está conformado por 18 genes DE. Cuando se consideran únicamente estos genes, las muestras se pueden separar con un análisis de agrupamiento jerárquico en grupo Control y grupo Pretérmino severo (Figura 4.12). Las muestras no se agrupan por su condición de presentar RPM. Los 18 genes de este grupo están subexpresados en Pretérminos severos con respecto al grupo Control. Incluyen genes miembros de la subfamilia de canales de potasio, genes con funciones de adhesión celular, la fosfatasa alcalina placentaria y un miembro de la familia Wnt. En análisis de Ontología génica, que veremos a continuación, este grupo de genes se asocia a términos GO relacionados a sistema nervioso, morfogénesis de órganos y complejos de canales iónicos, pero ninguno de ellos está significativamente enriquecido luego de corregir los valores p mediante BH.

Los dos grupos restantes (Figura 4.11) están conformados por 147 y 105 genes. En su mayoría tienen genes relacionados con citoquinas, miembros de la familia TNF, factores de coagulación y proteínas transmembrana. Una búsqueda de términos GO encontró 508 y 301 términos significativamente enriquecidos, respectivamente (valor p corregido por BH $< 0,05$). Estos términos son mayormente relacionados a respuesta inmune e inflamatoria y al sistema immune.

Al comparar los 270 genes DE con genes previamente asociados con trabajo de parto identificados con microarreglos de ARN (El-Azzamy et al., 2017; Haddad et al., 2006; Hamilton et al., 2013; Stephen et al., 2015), se encuentran coincidencias. En total, 58 de los 270 genes DE encontrados en esta tesis también fueron identificados como DE entre tejidos de nacimientos con trabajo de parto y sin trabajo de parto (Tabla suplementaria 1-b).

El análisis del efecto de la RPM sobre la expresión génica encontró sólo 14 genes diferencialmente expresados entre grupos con y sin RPM (FDR $< 0,05$ y tasa de cambio logarítmica absoluta > 2) (Tabla suplementaria 1-c). Únicamente 2 de estos genes fueron también detectados como DE con respecto a su condición de prematurez: *KCNJ16* y *TRPV6*.

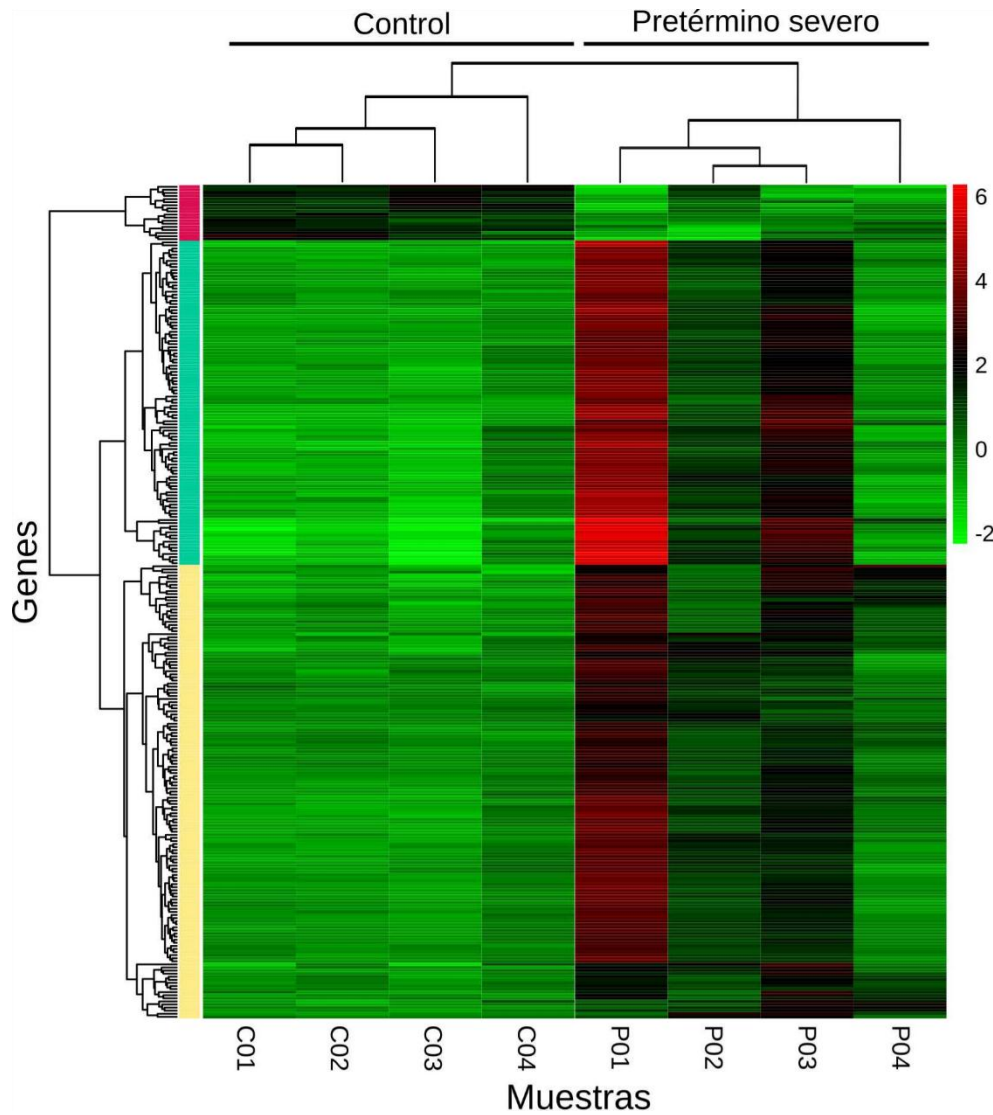


Figura 4.11. Heatmap de valores normalizados (*rlog*) de los niveles de expresión y agrupamiento jerárquico para los 270 genes DE ($FDR < 0,05$ y una tasa absoluta de cambio logarítmica > 2). La condición de cada muestra se denota en la parte superior de la figura. Los valores de *rlog* están codificados en una escala roja (alta expresión) a verde (baja expresión). En el dendrograma de genes se denotan con colores los 3 grupos mayores de acuerdo a los patrones de expresión de los genes.

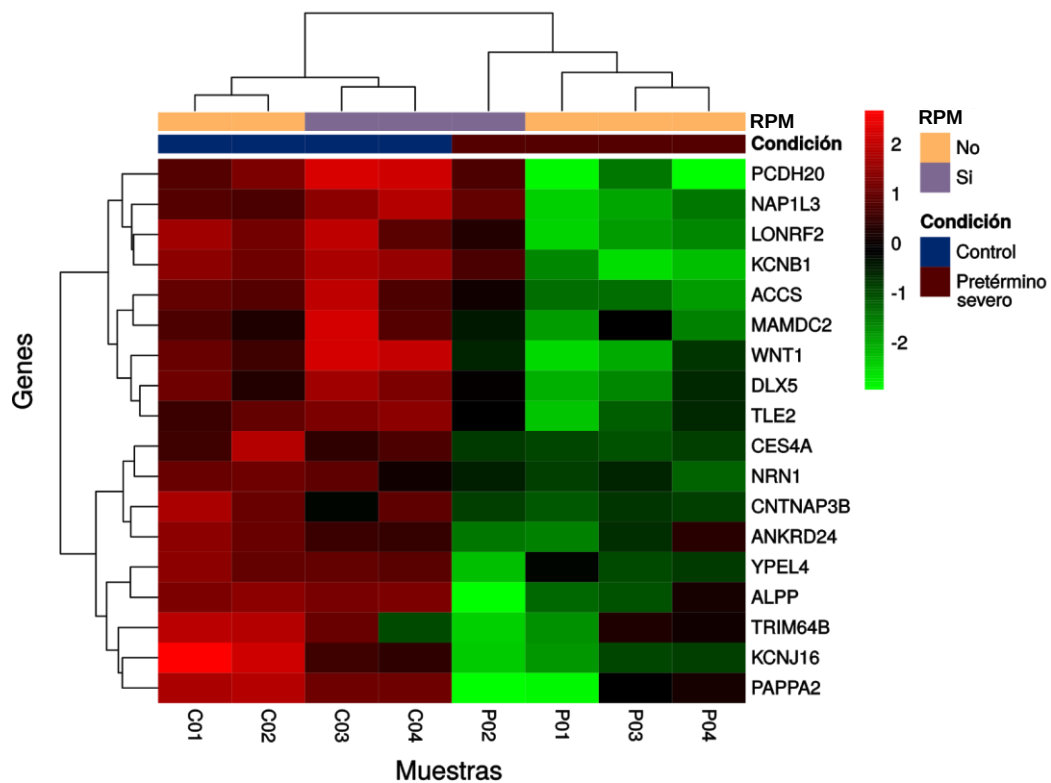


Figura 4.12. Heatmap de valores de expresión génica normalizados por *rlog* y agrupamiento jerárquico del grupo de 18 genes DE que discriminan controles de parto pretérmino, para las 8 muestras. Se muestran los códigos de los genes en la derecha del gráfico. Se usa la misma escala roja a verde de la Figura 4.11 para denotar el nivel de expresión génica.

4.3.3. ANÁLISIS DE ONTOLOGÍA GÉNICA

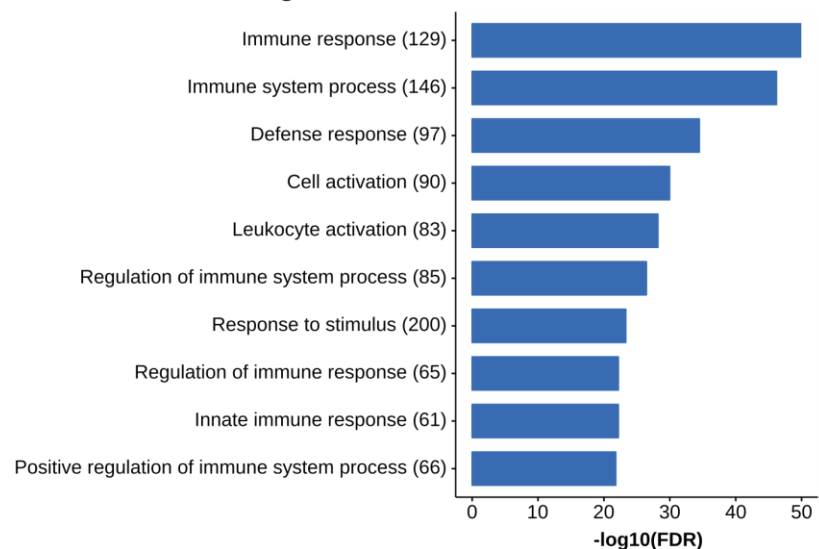
Se realizó un análisis de enriquecimiento de grupos de genes para identificar procesos y vías metabólicas que pudieran estar diferencialmente reguladas entre pretérminos severos y controles, en el paquete de R *goseq*. Aquellos genes con un $FDR < 0,05$ en el análisis de expresión génica se evaluaron contra un panel de fondo (*'background'*) de todos los genes expresados en el análisis de ARN-seq con sus anotaciones ENSEMBL. Así, se evaluaron los 270 genes DE entre pretérmino severo y controles contra un total de 18.405 genes expresados en las muestras

analizadas. Se evaluó la significancia de los términos de ontología génica usando la aproximación de Wallenius y aplicando corrección por hipótesis múltiples (BH). Únicamente se consideraron los términos con 3 o más genes significativos.

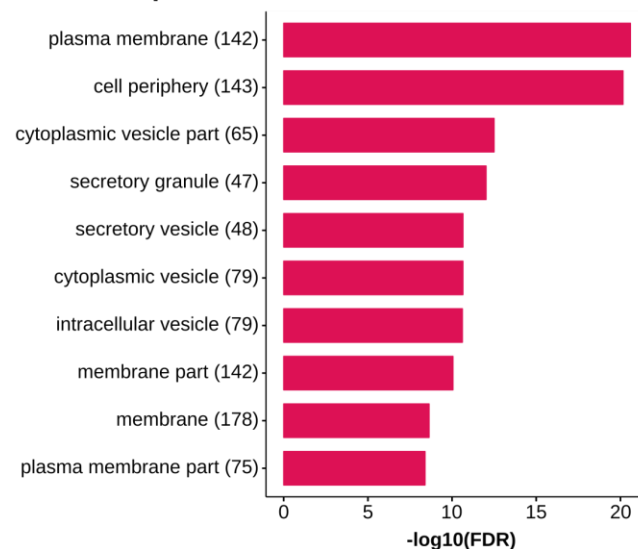
Se encontraron 1886 términos GO enriquecidos entre los genes DE analizados ($FDR < 0,05$), comprendidos en las categorías Procesos Biológicos (562 términos), Componentes Celulares (49) y Función Molecular (35) (Figura 4.13). Dentro de la categoría Procesos Biológicos, los términos *respuesta inmune*, *procesos de sistema inmune* y *respuesta defensiva* fueron los más significativamente enriquecidos dentro de los genes DE. Dentro de la categoría Componentes Celulares los términos más representados fueron membrana plasmática, periferia celular y vesículas citoplasmáticas. En la categoría Función Molecular los términos principales fueron: *actividad de receptor de señales*, *actividad de receptores* y *actividad de transductores*.

Se analizaron las vías metabólicas KEGG con el programa GAGE. Las vías más sobreexpresadas son la vía de señalización de las Citoquinas (“Chemokine signaling pathway”) y la del linaje de células hematopoyéticas (“Hematopoietic cell lineage”), ambas con $FDR = 7.038 \times 10^{-09}$ (Figura 4.14). Se encontraron varios procesos relevantes para el parto prematuro sobreexpresados, por ejemplo relacionados a la señalización intracelular: vías de señalización de tipo Toll, de tipo NOD, de células T y de Jak-STAT (Figura 4.14). La Figura 4.15 muestra los genes sobre y subexpresados encontrados en la vía de la señalización de las quimioquinas. No se encontraron vías metabólicas subexpresadas de forma significativa ($p < 0,05$).

A Procesos biológicos



B Componentes celulares



C Función molecular

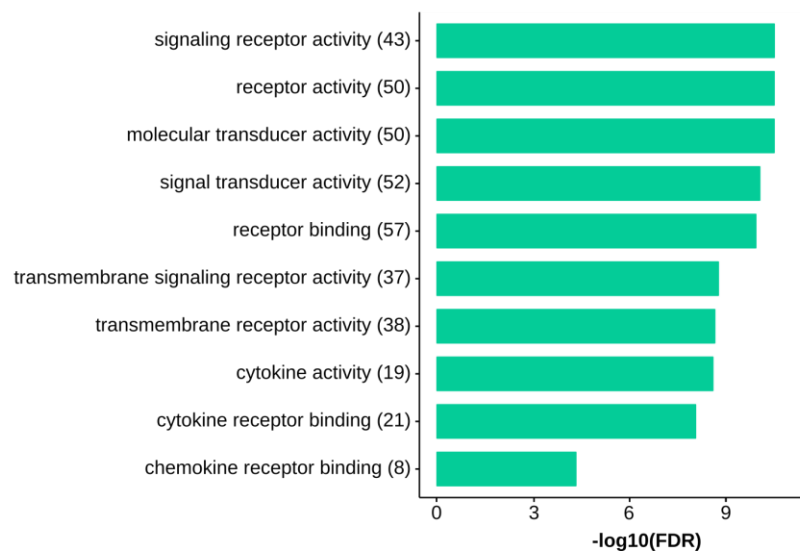


Figura 4.13. Términos de ontología génica enriquecidos entre los genes DE para las categorías Procesos Biológicos (A), Componentes Celulares (B) y Función Molecular (C). Se muestran los 10 términos más significativamente enriquecidos, con un valor p ajustado (BH) < 0,05. El eje x representa el valor de enriquecimiento, como el logaritmo de los valores de p ajustados por BH. El número de genes DE identificados para cada término se muestra entre paréntesis.

4.3.4. VALIDACIÓN DE LA EXPRESIÓN GÉNICA POR PCR EN TIEMPO REAL CUANTITATIVA

Se analizó la expresión de seis genes diferencialmente expresados por qRT-PCR, con el fin de validar los genes que se encontraron significativos en el análisis de ARN-seq. Los genes elegidos fueron: *IL1 β* , lipocalina 2 (*LCN2*), receptor de macrófagos con estructura colágena (*MARCO*), caspasa 5 (*CASP5*), serpinina familia A miembro 1, (*SERPINA1*) y miembro 15 de la superfamilia TNF (*TNFSF15*). Se analizaron 15 muestras de partos a término y 9 de partos pretérminos severos (Tabla 4.3). El análisis de ARN-seq mostró que todos ellos fueron sobreexpresados en parto pretérmino. El análisis de expresión génica por qRT-PCR confirmó el

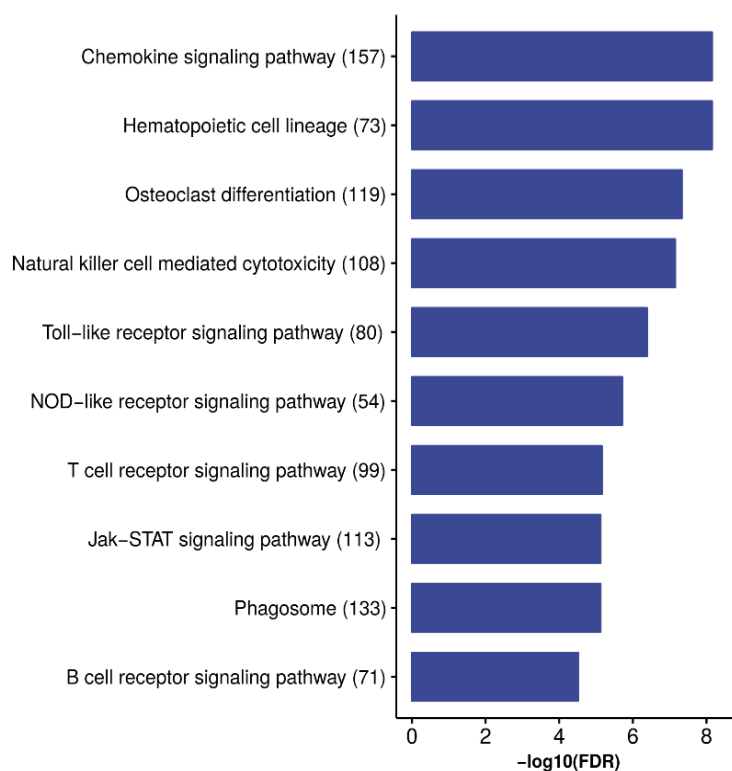


Figura 4.14. Vías KEGG enriquecidas entre los genes DE. Se muestran las 10 vías más significativamente enriquecidas, con un valor p ajustado (BH) < 0,05. El eje x representa el valor de enriquecimiento, como el logaritmo de los valores de p ajustados por BH. El número de genes DE identificados para cada vía se muestra entre paréntesis.

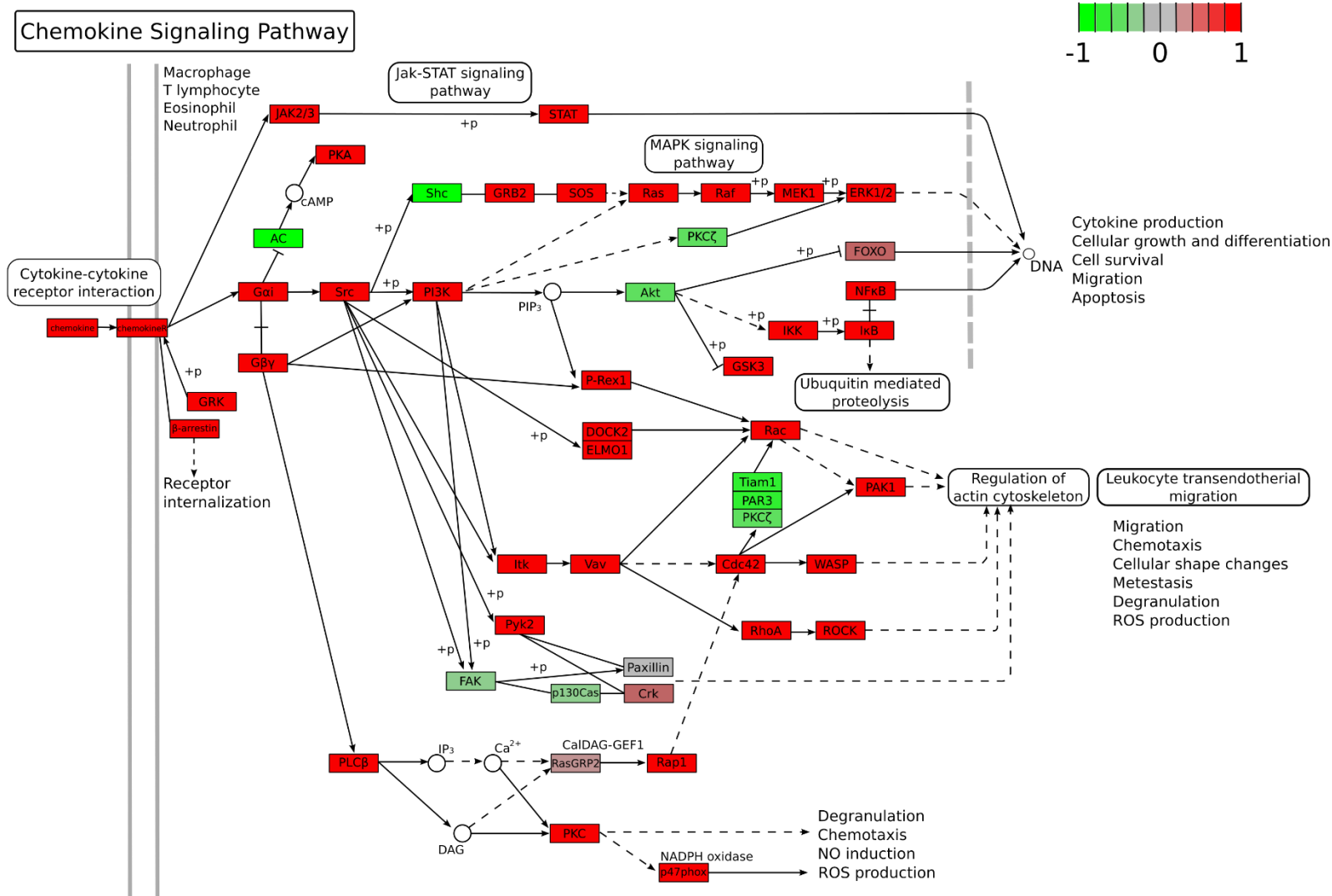


Figura 4.15. Diagrama KEGG de la vía de señalización de las quimioquinas. Los genes sobreexpresados en las membranas fetales de partos prematuros se muestran en rojo, y los subexpresados en verde.

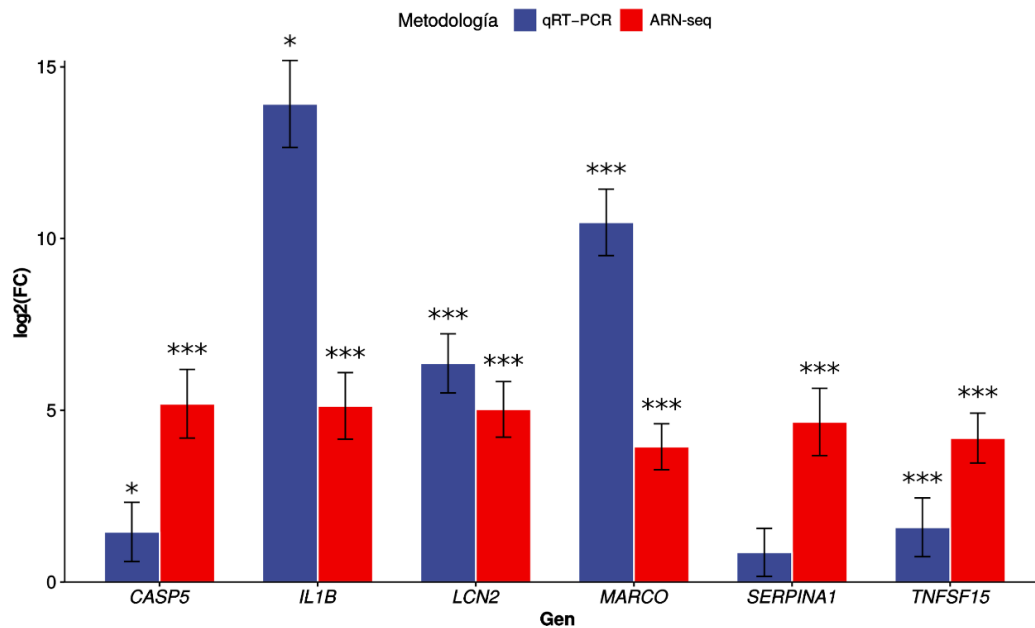


Figura 4.16. Validación de los niveles de expresión génica. Tasas de cambio de seis genes DE medidos por ARN-seq (rojo) versus qRT-PCR (azul). Se muestran los valores p de significancia para las tasas de cambio de las qRT-PCR. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$. ns: no significativo (valor p ajustado $> 0,05$).

sentido del cambio para los 6 genes analizados. La Figura 4.16 muestra las diferencias en los logaritmos de las tasas de cambio entre casos y controles para ambos métodos. Específicamente, 5 genes mostraron diferencias significativas entre pretérminos y controles cuando se analizaron los datos de qRT-PCR: *LCN2*, *MARCO* y *TNFSF15* (valor $p < 0,001$) y *CASP5* y *IL1 β* (valor $p < 0,05$). *SERPINA1* también está sobreexpresado según los datos de qRT-PCR, pero las diferencias entre grupos no fueron significativas.

4.3.5. VALIDACIÓN DE LOS RESULTADOS DE ARN-SEQ COMPARÁNDOLOS CON DATOS DE ESTUDIOS PREVIOS DE TRANSCRIPTOMAS DE PREMATUROS

Comparamos nuestros resultados con datos previamente reportados en la literatura. Una búsqueda en PubMed realizada el 19 de marzo de 2018 usando los términos “[preterm AND (transcriptome OR transcriptomic)]” resultó en 101 artículos. Recolectamos y analizamos los 101 resúmenes de las publicaciones, y encontramos que la mayoría de los estudios se centran en preeclampsia, tejido de vellosidades placentarias, modelos animales o sangre materna o fetal. Sólo ocho estudios usan la metodología de ARN-seq: de ellos, dos estudian sangre materna (Chim et al., 2017; Paquette et al., 2018), uno líquido amniótico (Kamath-Rayne et al., 2015), tres estudian modelos animales (Migale et al., 2016; Salaets et al., 2015; Willcockson et al., 2018) y dos únicamente el transcriptoma del trabajo de parto a término (Kim et al., 2012; Lui et al., 2018). Ninguno de estos es directamente comparable con nuestros resultados, por usar otras metodologías y/o diseños experimentales.

De los 101 hits de nuestra búsqueda, nos focalizamos en los trabajos que comparan parto pretérmino y a término en humanos y que estudien el transcriptoma de membranas coriónicas o amnióticas. Sólo un estudio de la búsqueda cumple con estos criterios (Bukowski et al., 2017), por lo que se consideró para compararlo con nuestros resultados. En este estudio, que realizaron mediante un análisis de microarreglos, encontraron 50 genes que identificaron el fenotipo de trabajo de parto pretérmino, de los cuales 7 están incluidos en los genes DE en nuestro estudio (Tabla 4.6). La dirección del cambio de expresión génica es la misma para los 7 genes mencionados.

Tabla 4.6. Genes DE simultáneamente encontrados en Bukowski et al. (2017) y en este estudio.

Símbolo génico	Gen	HGNC ID	Dirección de cambio
<i>BCL2A1</i>	BCL2 related protein A1	HGNC:991	sobreexpresado en PPT
<i>CCRL2</i>	C-C motif chemokine receptor-like 2	HGNC:1612	sobreexpresado en PPT
<i>CHI3L1</i>	chitinase 3 like 1	HGNC:1932	sobreexpresado en PPT
<i>CXCL2</i>	C-X-C motif chemokine ligand 2	HGNC:4603	sobreexpresado en PPT
<i>GPR84</i>	G protein-coupled receptor 84	HGNC:4535	sobreexpresado en PPT
<i>SAMSN1</i>	SAM domain, SH3 domain and nuclear localization signals 1	HGNC:10528	sobreexpresado en PPT
<i>SLC16A10</i>	solute carrier family 16 member 10	HGNC:17027	sobreexpresado en PPT

4.4 DISCUSIÓN

El poder establecer un perfil de expresión que identifique al PPT permitiría entender el contexto génico en el que se produce el evento, objetivo general de nuestra tesis. A pesar los esfuerzos por determinar factores de expresión específicos, ningún estudio previo ha permitido crear un perfil de expresión distintivo de la patología, que pueda mejorar su prevención, diagnóstico y tratamiento. Para contribuir en este aspecto, analizamos la expresión génica de las membranas corioamnióticas de PPT severos y recién nacidos a término por ARN-seq.

En este estudio, presentamos un patrón transcriptómico distintivo del PPT severo. Nuestro estudio reveló 270 genes diferencialmente expresados entre ambos grupos. Un número de esta magnitud, incluso aún luego de filtros exigentes, puede explicar parcialmente la ausencia de señales moleculares características del PPT, así como la dificultad de identificar biomarcadores e intervenciones terapéuticas potenciales

(Bukowski et al., 2017; Myatt et al., 2012). Hay varias vías y genes involucrados en el trabajo de parto, y muchos de ellos podrían ser activados o inactivados en condiciones patológicas (Blencowe et al., 2012; Buckberry et al., 2014; Myatt et al., 2012; Willcockson et al., 2018).

Como ya se dijo, se propone que el embarazo es mantenido por la subexpresión de citoquinas en la interfaz materno-fetal (Bukowski et al., 2017). La eliminación de esta subexpresión desencadena el parto a término, mientras que en el PPT la activación de múltiples vías del sistema inmune contrarresta los efectos de esta subregulación (Behrman & Butler, 2007; Bukowski et al., 2017; Challis et al., 2009; Flood & Malone, 2012; Migale et al., 2016; Romero et al., 2014; Vitoratos et al., 2006). Concordantemente, encontramos que los genes sobreexpresados están en su mayoría relacionados con la respuesta inmune (*e.g.* genes de interleuquina 24, ligandos de quimioquinas a motivos CXC o Factores de Necrosis de Tumores). A su vez, encontramos que el parto pretérmino severo tiene un enriquecimiento significativo de procesos inmunes en comparación con parto a término. Por ejemplo, nosotros observamos que la vía de las células T está sobreexpresada en prematuros severos. Se ha propuesto que los procesos de parto a término y el pretérmino están acompañados con un aumento en la expresión de las células T (Gomez-Lopez et al., 2016; Tarca et al., 2019). Los análisis de ontología génica y de enriquecimiento de vías KEGG indicaron que las vías inflamatorias e inmunes están enriquecidas en los pacientes pretérminos severos (Figura 4.13, Figura 4.14). Estos resultados son concordantes con los estudios mencionados, demostrando que el PPT está asociado a una activación extensa del sistema inmune en varios tejidos gestacionales. De hecho, la incidencia del PPT en asociación con la inflamación es mayor en el embarazo temprano (Korebrits et al., 1998), por lo que la sobreexpresión de vías inflamatorias e inmunes en el PPT severo es consistente con los resultados reportados en la bibliografía.

Apoyando esta hipótesis inflamatoria, nosotros detectamos que el gen *IL1 β* se encontró sobreexpresado en los partos pretérminos severos. *IL1 β* es un mediador

clave en los procesos patológicos de inflamación dado que puede estimular la expresión y liberar otros mediadores de parto (Heng et al., 2014; Romero et al., 2007, Romero et al., 2015; Watari et al., 2000; Watari et al., 1999). Evidencia reciente sugiere que *IL1 β* puede estar involucrado con el trabajo de parto mediante la activación del inflamasoma (Gotsch et al., 2008; Strauss et al., 2018). El inflamasoma es un complejo multiproteico ubicado en el citoplasma celular, cuya activación induce la transformación de pro-IL1 β inmadura en formas activas de IL1 β , mediante la acción de Caspasa-1 (CASP1) activa (Franchi et al., 2009; Petrilli et al., 2005; Shao et al., 2015; van de Veerdonk et al., 2011). La proteína 3 de la familia NLR contenedora de dominios pirina (NLRP3) es un receptor reclutador del inflamasoma (Jin & Flavell, 2010; Shao et al., 2015). Las concentraciones en líquido amniótico de algunos mediadores del inflamasoma (CASP1, IL1 β , NLRP3, IL18) son mayores en mujeres con parto a término y pretérmino espontáneo con inflamación/infección intraamniótica que en aquellas sin esta condición clínica (Gomez-Lopez et al., 2017; Gotsch et al., 2008; Pacora et al., 2000; Panaitescu et al., 2018; Plazyo et al., 2016; Romero et al., 2018; Strauss et al., 2018). Algunos genes del inflamasoma, incluyendo el gen contenedor de dominio CARD 4 (*NLRP3*), caspasa 5 (*CASP5*), ausente en melanoma 2 (*AIM2*) y el ya mencionado *NLRP3*, estuvieron sobreexpresados en nuestros datos. Además, el término de GO “inflammasome complex” (GO:0061702) está significativamente enriquecido en el grupo pretérmino severo. Si consideramos todos estos datos en conjunto, es posible hipotetizar que la activación del inflamasoma puede estar implicada en la inducción del parto pretérmino. Esto tendría importantes implicaciones médicas. El inflamasoma desempeña un papel fundamental en la patogénesis de diversas enfermedades inflamatorias humanas, como la diabetes, la enfermedad de Alzheimer y la aterosclerosis. Recientemente, se han analizado varias sustancias que promueven o inhiben la activación del inflamasoma NLRP3 con el objetivo farmacológico de lograr soluciones terapéuticas de enfermedades impulsadas por NLRP3 (Tang et al., 2018). Los componentes del inflamasoma pueden ser blancos de drogas que detengan el PPT

una vez que exista sospecha clínica y antes de que la respuesta inmunológica completa comience.

Como se puede ver en la Tabla suplementaria 1-d, varios de los genes identificados como DE en este trabajo fueron asociados previamente con PPT. En particular, dos de los genes sobreexpresados encontrados en este estudio fueron previamente propuestos como biomarcadores del PPT, por tener mayores concentraciones en el líquido amniótico en PPT que en parto a término: ligando 10 de quimioquinas a motivos CXC (*CXCL10*) y matriz de metalopeptidasa 8 (*MMP8*) (Holt et al., 2011; Kim et al., 2016; Lee et al., 2013; Maymon et al., 2000). Por lo tanto, estos biomarcadores pueden ser analizados en el líquido amniótico y dar información para prevenir el PPT. El *CXCL10* está relacionado al rechazo anti-fetal, de la misma forma en la que lo está al rechazo de aloinjertos (Lee et al., 2013). En este contexto, la sobreexpresión de *CXCL10* en PPT podría estar conduciendo al parto prematuro desencadenando una respuesta inflamatoria en el feto. *MMP8* disminuye la integridad de la matriz extracelular en el cuello del útero y membranas fetales (Arechavaleta-Velasco et al., 2004). *MMP8* puede alterar la matriz extracelular de las membranas y permitir el pasaje de las bacterias al endometrio, provocando el PPT inducido por bacterias (Witkin, 2015).

Entre los genes DE, un grupo pequeño de 18 genes se agrupan y permiten discriminar entre pacientes con parto a término y a pretérmino severo (Figura 4.12). Todos ellos están subexpresados en prematuros severos. Estos genes potencialmente caracterizan un patrón de expresión de pretérminos severos y por lo tanto son genes candidatos para entender el síndrome. Ninguno de estos genes subexpresados han sido propuestos previamente en la literatura como marcadores de PPT. Sin embargo, los genes fosfatasa alcalina placentaria (*ALPP*) y papalisina 1 (*PAPPA*) están incluidos en un grupo de 9 genes placentarios que juntos podrían predecir la edad gestacional en un estudio piloto realizado recientemente (Ngo et al., 2018). Los genes *ALPP* y papalisina 2 (*PAPPA2*), este último estrechamente relacionado a *PAPPA*, están entre los 18 genes subexpresados encontrados en este

trabajo. Por otro lado, PAPP2, un tipo de metaloproteinasa, típicamente con alta expresión génica en placenta (Farr et al., 2000; Uhlén et al., 2015), fue encontrada sobreexpresada en placentas de PPT versus a término (Chim et al., 2012).

El bajo número de genes subexpresados dificulta la identificación de vías estadísticamente significativas en este grupo: sólo dos de los 18 genes subexpresados están incluidos en una vía KEGG en común. Los genes Miembro 1 de la familia Wnt (*WNT1*) y Homeobox sin-distal 5 (*DLX5*) pertenecen a la vía “Signaling pathways regulating pluripotency of stem cells” (hsa04550). En partos a término, esta vía aumenta su expresión sobre el final del embarazo (Winn et al., 2007). El gen *WNT1* pertenece a la vía de señalización WNT, que está involucrada en proliferación celular, desarrollo y progresión de tumores. Esta vía podría contribuir a abortos y PPT, mediante la disminución de la proliferación e invasión de trofoblastos (Zmijanac Partl et al., 2018). Notoriamente, el gen *WNT2*, que está típicamente sobreexpresado en placenta, no estuvo diferencialmente expresado en nuestros datos. *DLX5* es un miembro de la familia génica de factores de transcripción de tipo homeobox. Los genes homeobox regulan el desarrollo tanto a nivel embrionario como a nivel placentario (Novakovic et al., 2017). Están ampliamente expresados en la placenta humana. Una función potencial de *DLX5* puede estar relacionada con el desarrollo neuroconductual y con la regulación funcional de células madre, ambos importantes para un desarrollo placentario normal (Novakovic et al., 2017). También, *DLX5* es sobreexpresado en placentas con preeclampsia, presumiblemente por una metilación alterada en el locus *DLX5* en preeclampsia que resulta en la pérdida de imprinting (Zadora et al., 2017). Se registró una disminución en la expresión para este gen en placenta humana en etapas tardías del embarazo, como una respuesta a un aumento en la metilación en el cuerpo del gen a medida que avanza la gestación (Novakovic et al., 2017). Nuestros resultados indican que la expresión de *WNT1* y *DLX5* es menor en las membranas fetales de PPT severos que en aquellas a término. En este contexto, especulamos que los genes subexpresados pueden actuar como genes de maduración que establezcan el ritmo de la gestación.

Estos genes de maduración putativos pueden ser muy difíciles de discriminar de aquellos conduciendo al parto pretérmino. Un gran problema en el estudio transcriptómico del PPT en humanos es la imposibilidad de obtener tejido sano de control a la misma EG para comparar con el tejido patológico a pretérmino. Consecuentemente, los genes DE entre PPT y término reflejan diferencias tanto de la EG como del evento patológico del parto pretérmino. Sabemos que, a consecuencia de nuestro diseño experimental, genes responsables del PPT y de la EG están solapados. Eidem et al. (2016) identificaron un grupo pequeño de genes candidatos específicos de PPT espontáneo o potencialmente relacionados a efectos de la EG. Para ello, compararon transcriptomas humanos de PPT espontáneos y a término con transcriptomas control pareados según su EG de una especie estrechamente relacionada (macacos). Encontraron funciones similares entre los genes de PPT y de EG, por lo que especulan que los efectos de ambos factores no se corresponden con procesos biológicamente distintivos (Eidem et al., 2016).

Con la intención de revelar genes responsables del PPT, comparamos los genes DE encontrados en PPT severo con genes asociados a trabajo de parto descritos en trabajos previos (El-Azzamy et al., 2017; Haddad et al., 2006; Hamilton et al., 2013; Stephen et al., 2015). Cincuenta y ocho de los genes DE encontrados en esta tesis habían sido identificados como genes relacionados con trabajo de parto, haciéndolos candidatos para el trabajo de parto independientemente de la EG. En el grupo restante, es posible que genes relacionados al PPT se solapen con los relacionados a la EG. Sin embargo, la pregunta de si el PPT es un parto que ocurre demasiado pronto o si representa un evento patológico con un nuevo grupo de genes expresados permanece abierta. Los análisis de los transcriptomas de membranas corioamnióticas que presentamos aquí no resuelven la controversia, aunque agregan información que podrá ayudar a completar el escenario de expresión del PPT. Por un lado, nuestros resultados confirmaron que las vías de respuesta inmune están sobreexpresadas en PPT. Por otro lado, la aproximación usando ARN-seq permitió identificar genes y vías sobreexpresadas no reportadas previamente como asociadas con PPT (*e.g.* vías de linaje celular hematopoyético,

diferenciación de osteoclastos). También encontramos que algunos de los genes subexpresados están involucrados en el sistema nervioso, morfogénesis (*WNT1*, *DLX5*, *PAPPA2*) y complejos de canales iónicos (canales de potasio mediados por voltaje, subfamilia J, miembro 16 (*KCNJ16*) y subfamilia B, miembro 1 (*KCNB1*)). Es difícil especular sobre la importancia de estos genes y vías en el PPT; por ejemplo, los canales iónicos están involucrados en un amplio espectro de funciones diferentes que eventualmente conducen al parto o PPT incluyendo implantación (Davidson & Coward, 2016), señalización de Ca^{++} (Wray et al., 2015), contracción de músculo liso (Buxton et al., 2011; Wray et al., 2015) e inflamación (Vaeth & Feske, 2018). Con respecto a la posibilidad que las vías del linaje celular hematopoyético estén conectadas con PPT, se sabe que la placenta es un órgano hematopoyético importante, que produce células madre hematopoyéticas (HSC) (Gekas et al., 2005; Khodadi et al., 2016). La placenta puede generar HSCs y causar su expansión, pero no su diferenciación. En principio se creyó que el acervo de las HSC placentarias aparecía temprano en el embarazo y luego disminuye, mientras que el desarrollo de HSC ocurría en el hígado (Alvarez-Silva et al., 2003; Khodadi et al., 2016). Estudios posteriores demostraron que la placenta en el medio de la gestación también posee gran cantidad de células madre hematopoyéticas pluripotenciales con capacidad de autoregeneración (Gekas et al., 2005, 2010). Adicionalmente, se encontraron progenitores hematopoyéticos en placentas a término con un bajo porcentaje de células comprometidas a linajes específicos (Kuchma et al., 2015). Dado que el linaje de macrófagos y monocitos hematopoyéticos produce osteoclastos, las vías de linaje celular hematopoyético y de diferenciación de osteoclastos están relacionadas. Aunque *a priori* ambas vías parecen lejanas del parto a término o pretérmino, las mismas pueden representar mecanismos inexplorados hacia el PPT. De hecho, todos los genes y procesos biológicos mencionados abren nuevas líneas de investigación, haciéndolos buenos candidatos como biomarcadores del PPT.

Una búsqueda bibliográfica para cotejar los genes encontrados con los postulados previamente, reveló que hay pocos estudios de gran escala que analicen tejidos

corioamnióticos, y que ninguno usa la tecnología de ARN-seq. Este es el primer estudio con ARN-seq de tejidos corioamnióticos en trabajo de parto término y pretérmino. Cuando se compararon los resultados con Bukowski et al. (2017), sólo 7 genes se solapan con los de nuestro estudio. Todos estos genes pueden estar relacionados con el PPT de algún modo: los receptores de quimioquinas *CCRL2* y *CXCL2* pertenecen a vías inflamatorias (Kobayashi, 2008; Mattern et al., 2014) y el receptor acoplado a proteína G 84 (*GPR84*) es un receptor de ácidos grasos que también se relaciona a inflamación (Mahmud et al., 2017). Los genes de la proteína relacionada a BCL2 A1 (*BCL2A1*) y la proteína 1 de tipo quitinasa 3 (*CHI3L1*) se asociaron recientemente con corioamnionitis en monos y humanos (Mahmud et al., 2017; Presicce et al., 2018); el gen miembro 10 de la familia portadora de solutos 16 (*SLC16A10*) está sobreexpresado en la placenta a mitad de la gestación (Uusküla et al., 2012); y el gen dominio SAM, dominio SH3 y señales de localización nuclear 1 (*SAMSN1*) es un gen poco caracterizado que puede estar involucrado en la diferenciación de los linfocitos B (Zhu et al., 2004). El hecho de que este grupo de genes se solape en ambos estudios, los cuales usaron diferentes técnicas experimentales, enfatiza que deben ser blanco de futuras investigaciones relacionadas al PPT.

Uno de los resultados más interesantes es la homogeneidad en la expresión génica de las muestras control. Luego de una reducción lineal de la variabilidad en un contexto de más de 30.000 genes expresados, las muestras de partos a término se agrupan en un grupo compacto (Figura 4.3). La dispersión de las muestras en el análisis de PCA es menor comparándola con perfiles de expresión de otros tejidos usando los mismos análisis exploratorios (Rahmioglu et al., 2017). Por el contrario, el análisis de la expresión corioamniótica de muestras prematuras severas reveló un grupo más disperso, resaltando la heterogeneidad de esta condición. Esto puede ser consecuencia de la complejidad de la caracterización del síndrome de parto prematuro, como fue previamente reportado con respecto a la expresión de otros tejidos relacionados al PPT, como el miometrio (Ackerman et al., 2018). En concordancia con esto, Ngo et al. (2018b) desarrolló un modelo de expresión

génica basada en EG que predice el tiempo hasta el alumbramiento para partos a término, pero que no lo hace para partos a pretérmino. Esto sugiere que para predecir el PPT el modelo debe incorporar al menos parte de los varios otros eventos fisiológicos que conducen al PPT (Ngo et al., 2018), *e.g.* RPM.

La ruptura prematura de membranas es una complicación frecuente del parto y, más importante aún, del PPT (Menon & Richardson, 2017). El transcriptoma del corioamnios puede ser diferente con y sin RPM. Para considerar este hecho, realizamos análisis estadísticos incluyendo la RPM como una covariable.

Adicionalmente, cuando analizamos las diferencias en expresión génica de la RPM, no encontramos diferencias sustanciales entre muestras de recién nacidos con y sin RPM. Detectamos únicamente 14 genes DE entre los grupos. Por otro lado, el PCA no mostró agrupamiento de las muestras basado en esta característica (Figura 4.3). Este bajo número de genes DE relacionados con RPM probablemente sea consecuencia de nuestro diseño experimental, que se focalizó en muestrear grupos discriminando por la condición de prematurez severa, mientras buscábamos aparear las variables restantes.

Finalmente, este estudio se realizó en una población mestizada donde el PPT es frecuente (10% aproximadamente), indicando un problema de salud pública serio. Recalcamos la importancia de estudios específicos de poblaciones, dado que el contexto genético puede afectar de diferente manera a las diferentes poblaciones, *e.g.* se sabe que la ancestría africana aumenta el riesgo de PPT (Crider et al., 2005; Plunkett et al., 2011; Plunkett & Muglia, 2008). Existe un vacío de información con respecto a determinadas poblaciones, excepto las provenientes de Norteamérica y Europa. En la búsqueda bibliográfica mencionada anteriormente encontramos sólo 3 de 101 análisis transcriptómicos basados en poblaciones de países latinoamericanos (Kim, 2012; Lee et al., 2013; Romero et al., 2010b), apoyando el hecho de que la mayoría de los estudios no incluyen poblaciones de esta parte del continente.

Todos los estudios mencionados que analizan poblaciones Latinoamericanas emplearon análisis de microarreglos, por lo que los datos de este trabajo son los primeros generados con ARN-seq en una población Latinoamericana disponibles para contrastar con futuros estudios. En particular, la población uruguaya es mestizada, con una contribución predominantemente europea, y con contribuciones nativo americanas y africanas de 10,4% y 5,6% respectivamente (Hidalgo et al., 2005; Sans et al., 1997).

El uso de análisis de ARN-seq identificó genes y vías metabólicas previamente reportados, así como también nuevos candidatos involucrados con el PPT. La combinación de genes sobre- y sub-expresados permitió desarrollar un clasificador de la patología basado en varios genes. Estos marcadores se generaron analizando una población mestizada en el cual el PPT tiene una incidencia alta. Serán necesarios más estudios que evalúen estos clasificadores, usando una muestra independiente de mayor tamaño para validar los resultados encontrados aquí. Estos marcadores también deberían ser comparados en otras poblaciones para obtener un panorama global del PPT severo.

5

IDENTIFICACIÓN DE VARIANTES GÉNICAS CANDIDATAS A CONTRIBUIR AL PARTO PRETÉRMINO

5.1 ANTECEDENTES Y ESTRATEGIA

La mutación es una de las principales fuentes de variación genética. Identificar variantes genómicas es un paso crítico para desentrañar la relación entre genotipo y fenotipo, y para entender su efecto potencial en características complejas. Como ya se mencionó, es posible que un conjunto de variantes contribuya significativamente a la susceptibilidad heredada de enfermedades complejas, cada una confiriendo un aumento moderado en el riesgo de la patología (Bodmer & Bonilla, 2008).

Como se indicó, se ha propuesto que el PPT espontáneo puede ser consecuencia del efecto de muchas variantes raras. Estas variantes raras no son detectadas en GWAS, debido a falta de poder estadístico, lo que es concordante con lo reportado en los estudios de asociación hasta el momento. Algunos estudios de secuenciado de genoma y exoma completos reportaron genes fetales asociados con PPT (Huusko et al., 2018; Li et al., 2017b; McElroy et al., 2013), mostrando que las aproximaciones de secuenciación masiva revelan asociaciones robustas de variantes raras que de otra forma se perderían en GWAS.

Así, estas herramientas permiten examinar el rol de las variantes frecuentes, novedosas y raras en la heredabilidad de las características complejas. El ARN-seq se usa principalmente para cuantificar la expresión génica del transcriptoma, pero también tiene el potencial de responder otras preguntas, entre ellas la identificación de variantes génicas en los genes expresados (Paul et al., 2016; Sheng et al., 2016). Este uso presenta algunos desafíos, dados por la complejidad intrínseca del transcriptoma. Por un lado, no se pueden detectar variantes en genes no expresados y para poder detectar variantes en genes pobremente expresados, el método de detección debe ser preciso y sensible en regiones con pocas lecturas (Piskol et al., 2013). Por ello, en comparación a la identificación de variantes a partir de ADN, al usar el transcriptoma es necesario aplicar criterios adicionales para considerar sitios de edición de ARN y para lidiar con artefactos vinculados al proceso de corte

y empalme del ARN. En particular, en datos transcriptómicos los SNVs son más fáciles de detectar que las inserciones, mientras que las deleciones pequeñas son más difíciles de detectar que los SNVs (Sun et al., 2017). Por otro lado, esta metodología no detecta correctamente variantes con expresión monoalélica. Teniendo estas consideraciones en cuenta, el ARN-seq es una técnica útil para la asignación de variantes en regiones expresadas del genoma (Chepelev et al. 2009; Piskol et al., 2013).

En este trabajo nos propusimos realizar un estudio *in silico* para identificar mutaciones puntuales en las regiones codificantes expresadas en membranas fetales humanas, a partir de los datos de ARN-seq generados en el capítulo anterior. Dado que las regiones expresadas constituyen sólo una pequeña parte del genoma humano, es posible obtener cobertura suficiente con esta tecnología. Buscamos identificar SNVs en el ARN secuenciado candidatos a modificar el fenotipo de parto prematuro severo. Para ello se compararon los SNVs obtenidos en ambos grupos. Esperamos identificar variantes raras y/o no reportadas que puedan estar asociadas a PPT.

5.2 METODOLOGÍA

5.2.1. MAPEO A GENOMA DE REFERENCIA

Para el análisis de variantes génicas se utilizó como insumo las lecturas limpias y filtradas con Trimmomatic en el capítulo anterior para las 8 muestras secuenciadas. Las mismas fueron alineadas al genoma humano de referencia GRCh37.75 de Ensembl (Yates et al., 2016) usando el programa STAR (Dobin et al., 2013), empleando el método de dos pasos. Este método es el que logra mayor sensibilidad para detectar SNPs e inserciones o deleciones (Engström et al., 2013). En la metodología de dos pasos de STAR se hace un primer alineamiento en donde se

detectan uniones de corte y empalme nuevas, que se agregan a los índices del genoma de referencia. Luego se hace un segundo alineamiento, en el cual se usa la información de las uniones previamente anotadas del archivo de anotación con formato GTF, y las nuevas, detectadas en el primer paso. Este procedimiento aumenta la sensibilidad de alineamiento sobre las uniones de empalme (Dobin et al., 2013). Se presenta un resumen del flujo de trabajo en la Figura 5.1.

5.2.2. ASIGNACIÓN, FILTRADO Y ANOTACIÓN DE SNVS

El alineamiento resultante es luego depurado con herramientas de Picard (<http://broadinstitute.github.io/picard/>) y GATK v3.7.0 (McKenna et al., 2010), siguiendo sus recomendaciones para identificación de variantes (Van der Auwera et al., 2013), con adaptaciones para datos de secuenciación de ARN (Van der Auwera, 2014). Así, se marcaron las lecturas duplicadas (herramienta ‘MarkDuplicates’) y se dividieron las lecturas para que queden asignadas a segmentos de exones, eliminando las bases que estuvieran alineadas a regiones intrónicas (herramienta ‘SplitNCigarReads’). Este último paso evita que se identifiquen variantes falsas en lecturas que estén alineadas a uniones exón-intrón

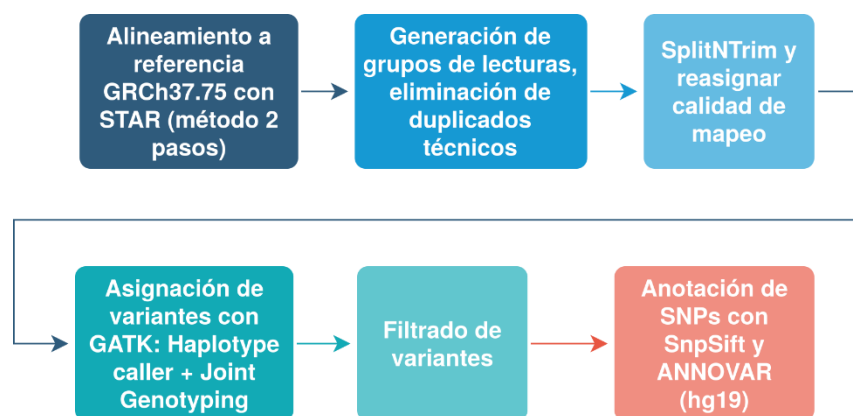


Figura 5.1. Flujo de trabajo usado para el procesamiento de ARN-seq para la identificación de variantes génicas

debido a un procesamiento de ARNm incorrecto. Se aplicaron también las herramientas ‘base quality score recalibration’, ‘indel realignment’ de GATK.

Para la asignación de variantes génicas se usó el *HaplotypeCaller* y *Joint Genotyping* de GATK, usando como referencia la anotación de SNPs del dbSNP para el genoma humano de referencia GRCh37 (dbsnp_138.b37.vcf). Al hacer la asignación de SNVs (‘HaplotypeCaller’) se usó el modo GVCF, que genera un archivo de variantes genómicas gVCF intermedio, que luego se usó para genotipar de forma conjunta varias muestras (‘Joint Genotyping’), usando la opción ‘GenotypeGVCFs’ del programa GATK.

Las variantes detectadas fueron separadas según si eran SNVs o inserciones o deleciones y cada grupo fue filtrado usando el módulo *VariantFiltration* del GATK. Los parámetros usados para filtrar las variantes se eligieron mirando las curvas de densidad de variables y seleccionando el mejor punto de corte para eliminar falsos positivos. Así se filtraron los SNVs según: la calidad de la asignación normalizada por la profundidad de secuenciado (‘QualByDepth’ < 10,0), el sesgo de hebra según test de Fisher (‘FisherStrand’ > 60,0), la raíz cuadrada de la calidad de mapeo de las lecturas (‘RMSMappingQuality’ < 40,0), la relación entre la calidad de mapeo de las lecturas que apoyan el alelo de referencia versus las que apoyan el alelo alternativo (MQRankSum < -5,0) y la evaluación de sesgo en la posición del alelo dentro de la lectura (ReadPosRankSum < -4,0). Para disminuir el número de falsos positivos, no se consideran variantes asignadas con baja cobertura de secuenciado. Para ello se filtraron las variantes según el número de lecturas totales cubriendo el sitio analizado (profundidad de lectura (DP) > 30 lecturas).

Se anotaron los SNVs filtrados con ANNOVAR, usando como referencia el genoma humano hg19. Para anotar las variantes encontradas, se usaron bases de datos génicas de dbSNP y de frecuencias alélicas de referencia de los proyectos NHLBI GO Exome Sequencing Project (NHLBI-ESP, <http://evs.gs.washington.edu/EVS/>), Exome Aggregation Consortium Data (ExAC) (Lek et al., 2016) y 1000

genomas (1000 Genomes Project Consortium et al., 2015). ANNOVAR define como variantes intergénicas a aquellas que estén al menos a 2 Kb de una secuencia codificante. Asimismo, incluye en región ARN no codificante (ARNnc) a variantes que no se solapen con transcritos codificantes anotados: incluye así tanto ARN no codificantes anotados (miARNs y ARNs no codificantes largos conocidos) como loci no anotados.

5.2.3. ANÁLISIS DE VARIANTES GÉNICAS UBICADAS EN LOS GENES DIFERENCIALMENTE EXPRESADOS

El análisis de los SNVs filtrados se prosiguió en el programa R, usando el paquete VariantAnnotation v1.22.3 (Obenchain et al., 2014). Se filtraron aquellos SNVs ubicados en los intervalos genómicos correspondientes a los 270 genes identificados como diferencialmente expresados en el capítulo anterior. Definimos las variantes del ARN-seq que también estuvieran en dbSNP (version 150) como “conocidas”. Contrariamente, aquellas variantes que no se pudieron encontrar en dbSNP fueron denominadas “novedosas”.

Para caracterizar las variantes génicas encontradas en los genes DE de este trabajo, se interrogó el catálogo de GWAS de NHGRI-EBI (Buniello et al., 2019) mediante el paquete de R ‘gwascat’ (Carey, 2019) y la base de GTExAnalysis version 7 de útero (GTEx Consortium, 2013) (‘Uterus.v7.signif_variant_gene_pairs.txt’).

Para los análisis subsiguientes, se seleccionaron los SNVs sin datos faltantes: *i.e.*, que fueron genotipados en las 8 muestras secuenciadas, y además que estuvieran anotados como pertenecientes a regiones exónicas, región no traducida (UTR) 5’ o 3’UTR. Se eliminaron de esta lista los SNVs ubicados en sitios conocidos de edición de ARN, según la base de datos RADAR (Ramaswami & Li, 2014). Con este subgrupo de SNVs, se analizaron sus genotipos sumando los conteos de homocigotas de referencia, heterocigotas y homocigotas alternativos para todos los SNVs de los controles y de los prematuros. Por otro lado, se sumaron los conteos

de alelos de referencia y alternativos entre ambos grupos. Se evaluó la diferencia entre los conteos de genotipos y entre los conteos alélicos entre casos y controles mediante tests de chi cuadrado.

A continuación, se analizó la suma de los conteos de genotipos y también de alelos, discriminando por gen. Para ello usamos los genes diferencialmente expresados en los análisis de ARN-seq que además tuvieran lecturas para todas las muestras. Trabajamos entonces con 150 genes que cumplían estas condiciones.

Para cada gen con diferencias significativas en sus conteos genotípicos entre casos y controles, se reconstruyeron los haplotipos mediante el programa PHASE v2.1 (Stephens & Scheet, 2005; Stephens et al., 2001). Los haplotipos estimados se compararon con los obtenidos del proyecto 1000 genomas para las siguientes super-poblaciones: europeas (eurodescendientes norteamericanos, toscanos y españoles), asiáticas (etnia Han en China), latinoamericanas (mexicanos, peruanos, colombianos y puertorriqueños) y africanas (Yorubas, Mende y Esan). La comparación de la distribución de los haplotipos reconstruidos entre casos y controles para cada gen se realizó mediante un test exacto de Fisher.

Se realizó un análisis de componentes principales con los genotipos de los SNVs para este grupo de genes, con las muestras secuenciadas en este trabajo y las muestras seleccionadas del proyecto 1000 genomas detalladas arriba. Este análisis provee una representación de nuestros datos en un menor número de dimensiones, capturando la variabilidad entre individuos a lo largo de los SNVs.

5.2.4. DETECCIÓN DE VARIANTES QUE AFECTAN SITIOS DE UNIÓN DE MIARN

Para evaluar las consecuencias funcionales de las variantes no sinónimas, usamos las bases de datos SIFT y PolyPhen2. En general, ambas usan algoritmos que estiman si un cambio de aminoácido es capaz de alterar la función proteica

basándose en la conservación de la proteína con respecto a secuencias homólogas, así como la similitud fisicoquímica con los aminoácidos alternativos (Vaser et al., 2016). Ambas califican el efecto del cambio de aminoácido en categorías que van desde benigna o tolerada, a posiblemente dañina o deletérea.

Para los SNVs encontrados en genes con diferencias significativas en sus conteos genotípicos entre casos y controles que están en regiones 3'UTR, se buscó si alguno afectaba la unión de micro ARN. Se buscó la base de datos PolymiRTS 3.0 (Bhattacharya et al., 2014), que analiza efectos de polimorfismos en miARN y en sus sitios blancos. Solamente se consideran aquellos SNVs que afectan el apareamiento a la región semilla del miARN.

5.3 RESULTADOS

5.3.1. VARIANTES GÉNICAS ENCONTRADAS EN SECUENCIADO DE ARN

En el protocolo de asignación de variantes a partir de la secuenciación del ARN, se encontraron 75.743 variantes únicas. De ellas, la mayoría se encuentran en un contexto de tipo 3'UTR (35,43%) y exónico (24,18%) (Tabla 5.1). Un menor porcentaje está en regiones intrónicas e intergénicas (18.55 y 7.29%, respectivamente).

Cuando nos restringimos únicamente a las variantes que se encuentran en los genes diferencialmente expresados entre casos y controles (Capítulo 3), vemos que se asignan 1887 variantes que están en 243 de los 270 genes diferencialmente expresados. Cuando éstos se clasifican según su ubicación genómica, 700 (37,1%) se presentan en intrones, 548 (29,04%) en 3'UTR y 452 (23,95%) en exones (Tabla 5.1). Las 452 variantes exónicas que están dentro de genes DE se clasifican en 238 sinónimas (52,65%), 212 no sinónimas (47,12%) y 1 sin sentido (0.22%).

Tabla 5.1. Conteo y porcentaje de variantes génicas detectadas según el contexto genómico en el que se encuentran. Ordenada según el porcentaje de variantes en total de ARN-seq.

Contexto genómico	Variantes en total de ARN-seq		Variantes en genes diferencialmente expresados	
	Conteo	Porcentaje	Conteo	Porcentaje
3'UTR	26875	35,43	548	29,04
exón	18341	24,18	452	23,95
intrón	14062	18,55	700	37,1
intergénico	5527	7,29	28	1,48
ARNnc_intrón	3203	4,23	46	2,44
5'UTR	2712	3,55	69	3,6
ARNnc_exón	2707	3,57	11	0,58
río_abajo	1763	2,27	30	1,54
río_arriba	559	0,68	2	0,11
Codón de terminación	132	0,17	0	0
sitio de corte	56	0,06	1	0,05
ARNnc_sitio de corte	5	0,01	0	0
3'UTR: región no traducida 3'. 5'UTR: región no traducida 5'. ARNnc: ARN no codificante.				

5.3.2. VARIANTES NOVEDOSAS Y RARAS ENCONTRADAS EN GENES DE

Dentro de las variantes en los genes DE hay 125 variantes novedosas, no reportadas en dbSNP ni ExAC, que se encuentran en 33 genes. Treinta y cuatro de ellas presentan su alelo alternativo en un único individuo (“*singleton*”), de las cuales sólo 6 cuentan con lecturas en todos los individuos. Estos 6 singletons no presentan diferencias de conteo significativas entre casos y controles (test de chi-cuadrado: $\chi^2 = 1,71$; $p = 0,19$).

Dos de las variantes novedosas son mutaciones con cambio de sentido, de las cuales una (7:149462528_A/G, en el gen *ZNF467*) tiene un efecto deletéreo según

SIFT (Tabla suplementaria 2). El alelo alternativo de esta variante se presenta únicamente en el individuo P04 (prematureo severo), en homocigosis, y causa un cambio de una Fenilalanina por una Prolina. La variante 14:88478084_G/A en el gen *GPR65* también genera un cambio de sentido (Arginina por Lisina) pero se predice que tiene un efecto tolerado. Se presenta en heterocigosis únicamente en el prematureo severo P03. Por otro lado, 41 de las variantes novedosas están en regiones regulatorias, como 5'UTR, 3'UTR y ARNs no codificantes (Tabla suplementaria 2).

También encontramos variantes en genes DE previamente identificadas como raras. Al evaluar las frecuencias alélicas reportadas en bases de datos genómicas 1000 genomas, ExAC y NHLBI-ESP, vemos que de las 1887 variantes encontradas en genes DE, hay 93 variantes con $MAF < 0,01$ en al menos una de ellas. Se distribuyen un 40% en exones, 28% en intrones, 24,7% en 3'UTR y 6,4% en 5'UTR. La mayoría de las variantes se presentaron en heterocigosis, con la excepción de 10 variantes que lo hicieron en homocigosis. También la mayoría lo hicieron en los individuos prematureos severos (51 variantes).

5.3.3. COMPARACIÓN DE LAS VARIANTES ENCONTRADAS EN GENES DE EN BASES DE DATOS GENÓMICAS

Dentro de las 1887 variantes encontradas en los genes DE, se encontraron 11 variantes que están reportadas en la base de datos de cis-eQTL GTEx en útero (Tabla 5.2). Las 11 variantes eQTL de útero se encuentran en los genes *ACCS* (ENSG00000110455) y *MARCH1* (ENSG00000145416). De ellas, una es de tipo codificante, perteneciente al gen *MARCH1* (rs13130399). Las 10 restantes se encontraron en regiones 3'UTR. La variante rs13130399 es una variante sinónima que actúa como cis-eQTL en varios tejidos: sangre entera, útero, cerebro, pulmón, piel, entre otros. Presenta menor expresión génica en los homocigotas del alelo

Tabla 5.2. SNPs en genes DE que están reportados como eQTLs en útero por GTEx v7.

ID dbSNP	Ubicación génica	Gen
rs1053209	3'UTR	<i>MARCH1</i>
rs1054683	3'UTR	<i>MARCH1</i>
rs1054682	3'UTR	<i>MARCH1</i>
rs1054681	3'UTR	<i>MARCH1</i>
rs1441724	3'UTR	<i>MARCH1</i>
rs1044722	3'UTR	<i>MARCH1</i>
rs1044721	3'UTR	<i>MARCH1</i>
rs1133505	3'UTR	<i>MARCH1</i>
rs6816526	3'UTR	<i>MARCH1</i>
rs13130399	Exón	<i>MARCH1</i>
rs178525	Intrón	<i>ACCS</i>

alternativo (GTEx Analysis Release V7 (dbGaP Accession phs000424.v7.p2), (GTEx Consortium, 2013)). Para este SNP, en las muestras secuenciadas hay una mayor frecuencia del alelo alternativo T en controles (MAF = 1), mientras que es menor en los prematuros severos (MAF = 0,25). Esto es consistente con la expresión génica diferencial evidenciada en el capítulo anterior, donde se encontró una mayor expresión génica en prematuros severos para *MARCH1* (logFC = 2,37; valor p ajustado = 0,02; ver Tabla suplementaria 1-a).

El catálogo de GWAS de NHGRI-EBI tiene 14 variantes de nucleótido simple asociadas a parto prematuro. Ninguna de ellas coincide con las variantes encontradas en los genes DE de prematuros.

5.3.4. DIFERENCIAS EN EL CONTEO DE GENOTIPOS DE GENES DE ENTRE CASOS Y CONTROLES

Dentro de las 1887 variantes de nucleótido simple encontradas en los análisis de ARN-seq, hay 85 que se encuentran en un sitio de edición de ARN conocido, por lo que se excluyeron de los análisis siguientes. De las variantes restantes, hay 554 que se encontraron en todas las muestras analizadas (*i.e.* tienen al menos 30 lecturas secuenciadas, según los criterios de filtrado detallados en la metodología) y que además tienen una ubicación en una posición génica de tipo exónica, 5'UTR o 3'UTR.

Con ellas, en un estudio de asociación entre casos y controles, no se encontraron variantes con diferencias significativas entre ambos grupos, considerando como umbral de significancia un FDR de 5×10^{-8} (Dudbridge & Gusnanto, 2008). No hubo SNPs significativamente asociados a la característica estudiada.

Por otro lado, encontramos diferencias significativas entre la suma del conteo de sus genotipos homocigotas de referencia, heterocigotas y homocigotas alternativos entre casos y controles (test de chi-cuadrado: $\chi^2 = 16,34$; $p = 0,00012$). También se encontraron diferencias significativas entre la suma del conteo de alelos entre casos y controles para las mismas variantes (test de chi-cuadrado $\chi^2 = 5,46$; $p = 0,019$). Se detecta un exceso significativo de heterocigotas en el grupo de prematuros severos frente a controles (32,3% y 26,9%, respectivamente; test chi-cuadrado de Pearson $\chi^2 = 15,07$; $p = 0,0001$).

Estas variantes codificantes seleccionadas dentro de los genes diferencialmente expresados se encuentran ubicadas en 150 genes. Cuando evaluamos las frecuencias genotípicas en cada uno de ellos, hay 9 genes que tienen diferencias significativas entre la suma de los conteos genotípicos entre casos y controles (Tabla 5.3).

Estos 9 genes contienen un total de 82 variantes. De ellas, 17 se encuentran en exones, siendo 7 sinónimas y 10 no sinónimas. Hay 4 variantes presentes en estos 9 genes que son variantes raras (*i.e.* MAF en Proyecto 1000 genomas < 0,01): rs570062682, rs550284671 (en gen *PLEK*), rs140080026 (en gen *KRAS*) y rs181468783 (en gen *LRRC25*).

De las variantes identificadas en este estudio para estos genes, sólo en *KRAS* encontramos variantes reportadas en la base de datos ClinVar (Landrum et al., 2014, reportado por Annovar), clasificándose todas como “Probablemente benigna”. El resto de las variantes identificadas no están categorizadas en esta base de datos.

Tabla 5.3. Genes con diferencias significativas en la distribución de los genotipos de sus variantes para casos y controles.

Símbolo del gen	Nombre del gen	HGCN ID	Posición cromosómica	Número de variantes por gen	Valor p (Chi-cuadrado)
<i>FAM129A</i>	family with sequence similarity 129 member A	HGNC:16784	1q25.3	12	$2,40 \times 10^{-03}$
<i>PLEK</i>	pleckstrin	HGNC:9070	2p14	13	$7,34 \times 10^{-07}$
<i>MARCH1</i>	membrane associated ring-CH-type finger 1	HGNC:26077	4q32.2-q32.3	4	$5,17 \times 10^{-04}$
<i>PIK3AP1</i>	phosphoinositide-3-kinase adaptor protein 1	HGNC:30034	10q24.1	12	$1,64 \times 10^{-02}$
<i>KRAS</i>	KRAS proto-oncogene, GTPase	HGNC:6407	12p12.1	15	$3,42 \times 10^{-02}$
<i>GNG2</i>	G protein subunit gamma 2	HGNC:4404	14q22.1	8	$2,94 \times 10^{-02}$
<i>CD300E</i>	CD300e molecule	HGNC:28874	17q25.1	5	$3,57 \times 10^{-02}$
<i>ICAM3</i>	intercellular adhesion molecule 3	HGNC:5346	19p13.2	3	$1,24 \times 10^{-02}$
<i>LRRC25</i>	leucine rich repeat containing 25	HGNC:29806	19p13.11	10	$4,88 \times 10^{-03}$

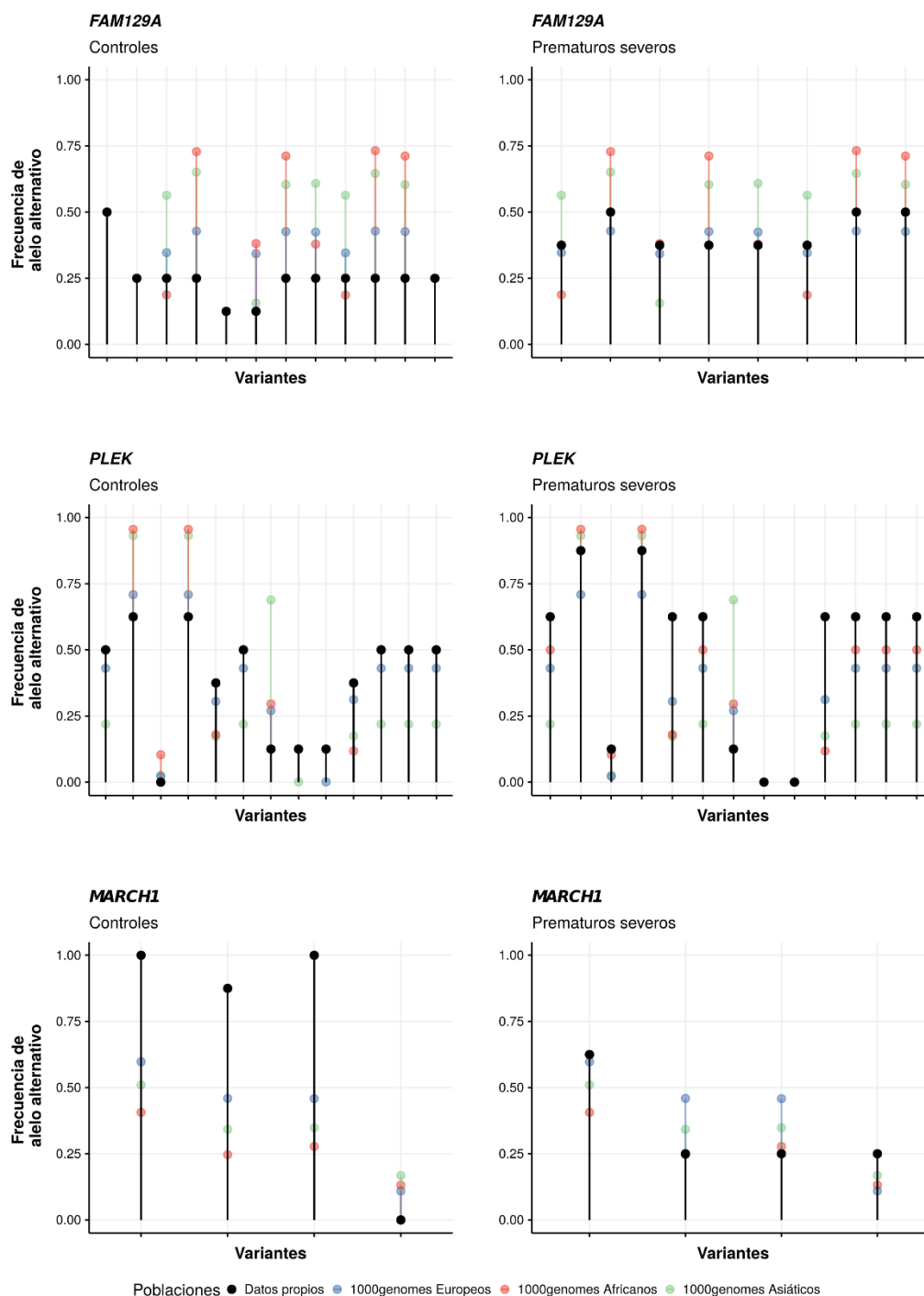


Figura 5.2. Frecuencia del alelo de referencia de las variantes génicas encontradas en cada uno de los 9 genes que presentan diferencias significativas entre la distribución de sus genotipos entre casos y controles. La lista de variantes por gen se presenta en la Tabla suplementaria 3. Se presentan las frecuencias alélicas de nuestros datos y la de tres poblaciones de 1000 genomas, como referencia.

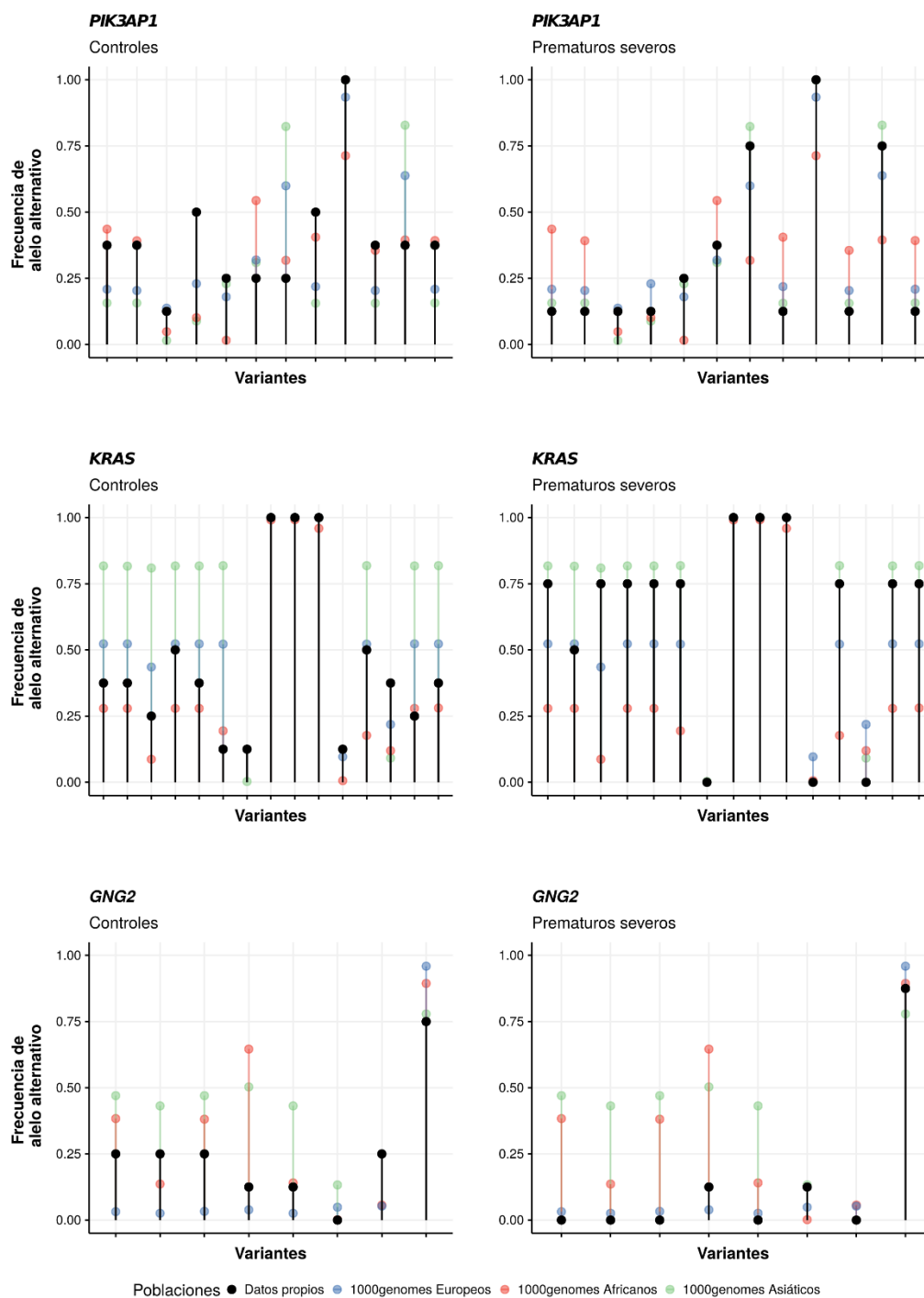


Figura 5.2. (Continuación)

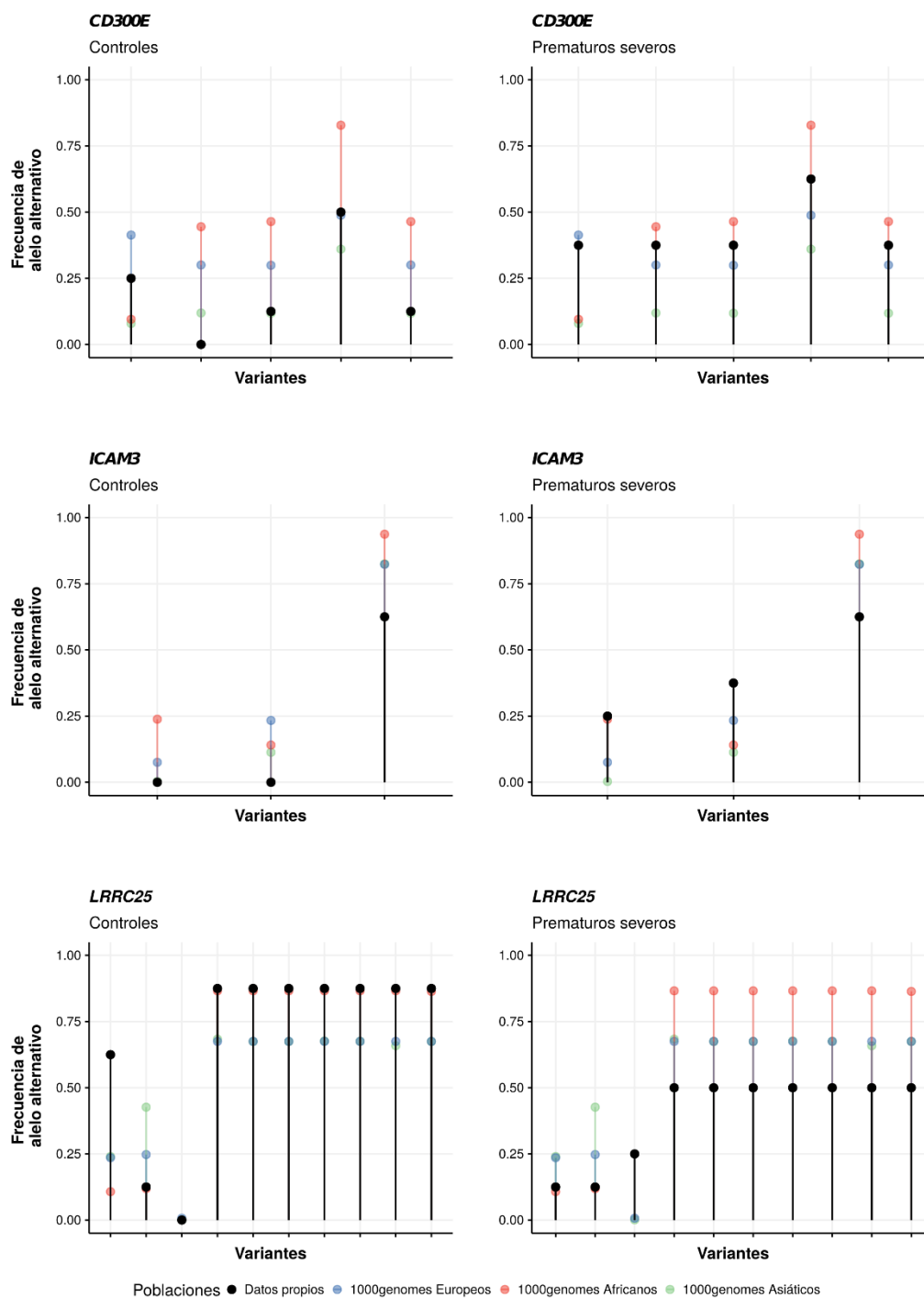


Figura 5.2. (Continuación)

5.3.5. EFECTOS FUNCIONALES Y SOBRE SITIOS DE UNIÓN DE MIARN DE VARIANTES DE LOS 9 GENES

Se evaluaron las consecuencias funcionales de los SNVs presentes en los genes con diferencias genotípicas en nuestras muestras que cayeran en exones. Así, se predice que la variante rs17035364 (gen *PLEK*) tiene efectos deletéreos según SIFT y PolyPhen2. A su vez, la variante rs3816281 (gen *PLEK*) tiene un efecto tolerado según SIFT pero posiblemente deletéreo según PolyPhen2. Ambas variantes se ubican en el gen *PLEK*, y generan efecto en las posiciones iniciales de la proteína: aminoácidos 5 y 97 respectivamente (de una proteína de 350 aminoácidos). En cuanto a su distribución, rs3816281 está representado por un alelo alternativo en casos y uno en controles, mientras que rs17035364 sólo tiene un alelo alternativo en los prematuros severos. El resto de las variantes exónicas tienen efectos benignos o tolerados.

Dentro de los 9 genes con diferencias genotípicas en las muestras analizadas, se encontraron 21 SNVs ubicados en 3'UTRs que confieren cambios en la unión de miARNs (Tabla suplementaria 3). Estos SNVs se encuentran en los genes *KRAS* (9 SNVs), *FAM129A* (5), *PLEK* (4) y *PIK3AP1* (3). Varios de ellos crean sitios de unión nuevos para miARNs.

5.3.6. RECONSTRUCCIÓN DE HAPLOTIPOS PARA LOS GENES CON DIFERENCIAS GENOTÍPICAS ENTRE CASOS Y CONTROLES

Estimamos las frecuencias alélicas de casos y controles para estos 9 genes (Tabla 5.3). Las frecuencias alélicas entre casos y controles también tienen diferencias significativas para estos 9 genes (Figura 5.2). Las variantes génicas presentadas en los gráficos (eje x) se presentan en la Tabla suplementaria 2.

La reconstrucción de los haplotipos más probable para cada uno de estos 9 genes para casos y controles se presenta en la Tabla 5.4. Algunos de los genes presentan

haplotipos estimados no encontrados en una submuestra de 2015 individuos pertenecientes a 3 poblaciones del proyecto 1000 genomas. Los genes *FAM129A*, *MARCH1*, *GNG2* y *KRAS* presentan haplotipos no reportados en el proyecto 1000 genomas en una o dos (en el caso del gen *KRAS*) muestras de prematuros severos cada uno. El gen *GNG2* presenta dos haplotipos no reportados en 1000 genomas, presentes en dos individuos diferentes del grupo control. En el caso del gen *KRAS*, el programa PHASE utilizado para la reconstrucción no pudo reconstruir de forma confiable la fase de algunos haplotipos de las muestras control, por lo que no se reportan. Igualmente, se puede afirmar que este gen presenta 3 haplotipos no reportados en 1000 genomas que se encontraron en 3 individuos del grupo control. La muestra control restante presenta los haplotipos con fase indefinida pero que tampoco coinciden con ninguno de los haplotipos reportados por dicho proyecto.

Solo el gen *MARCH1* presentó diferencias significativas entre la distribución de los haplotipos reconstruidos entre casos y controles (test exacto de Fisher, valor $p = 0,0009$). En este gen los haplotipos presentes en los prematuros severos son más diversos y se corresponden a haplotipos menos frecuentes en las poblaciones reportadas. Debemos considerar que el tamaño pequeño de la muestra limita el poder estadístico de nuestros análisis, por lo que estos resultados deben ser verificados en una muestra de mayor tamaño.

Los genotipos de los SNP analizados para los 9 genes (Tabla 5.4) se utilizaron para realizar un análisis de componentes principales, de forma de representar cómo se distribuye la variación de los datos (Figura 5.3). Según este análisis los primeros dos ejes explican el 16,95% y el 11,19% de la variación. Se grafican como referencia algunas subpoblaciones del proyecto 1000 genomas. El grupo de las muestras control se agrupa sobre el centroide de la distribución de puntos, de forma relativamente agrupada. Las muestras a pretérmino severo se disponen hacia la

Tabla 5.4. Frecuencia de haplotipos en poblaciones de 1000 genomas y las muestras de este estudio para los SNVs analizados en los 9 genes con diferencias significativas en la distribución de sus genotipos entre casos y controles.

<i>FAM129A</i>	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs492126, rs682331, rs526024, rs14426, rs14023, rs581100					
CGATACC	(0.572)	(0.427)	(0.542)	(0.243)	4 (0.5)	6 (0.75)
TAGAGTG	(0.331)	(0.490)	(0.365)	(0.211)	3 (0.375)	2 (0.25)
TAGAGCC	(0.088)	(0.039)	(0.062)	(0.178)	...	...
CGGAGCC	...	...	...	(0.019)	...	...
CGGTGCC	...	(0.034)	...	...	...	...
TAGTGCC	...	...	(0.027)	(0.348)	...	...
TGGTGCC	...	...	...	...	1 (0.125)	...

<i>PLEK</i>	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs3816281, rs2070171, rs1063479, rs1050181, rs6713721, rs6728995, rs6742200, rs1867313, rs3732044, rs8761					
GGGGCTATTG	(0.304)	(0.058)	(0.180)	(0.029)	1 (0.125)	3 (0.375)
GAACTCGCCA	(0.294)	(0.189)	(0.197)	(0.085)	5 (0.625)	3 (0.375)
TGAGCTACTG	(0.267)	(0.665)	(0.406)	(0.321)	1 (0.125)	1 (0.125)
GGACCCGCCA	(0.110)	(0.058)	(0.115)	(0.351)	...	1 (0.125)
GGAGCTACTG	(0.013)	(0.029)	(0.077)	(0.09)	1 (0.125)	...
GAAGCTACTG	...	...	...	(0.087)	...	...
GGA CTGCCA	...	...	...	(0.028)	...	...

Tabla 5.4 (Continuación)

MARCH1	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs17576350, rs1054681, rs1133505, rs13130399					
GTCC	(0.428)	(0.514)	(0.256)	(0.565)	3 (0.375)	...
GCTT	(0.402)	(0.319)	(0.581)	(0.226)	...	7 (0.875)
ACCC	(0.110)	(0.152)	(0.124)	(0.122)	2 (0.25)	...
GCTC	(0.023)	...	...	...	1 (0.125)	...
GCCT	(0.021)	...	(0.016)	...	1 (0.125)	1 (0.125)
GCCC	(0.016)	...	(0.010)	(0.027)	...	...
GTCT	...	...	...	(0.053)	...	...
ACTT	...	...	...	...	1 (0.125)	...
PIK3API	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque 1	Marcadores: rs7448, rs8578, rs1051450, rs1063605, rs7074516, rs3748236, rs12784975, rs3748234					
TATTCATT	(0.583)	(0.854)	(0.500)	(0.238)	6 (0.75)	2 (0.25)
CGGCTGCC	(0.217)	(0.087)	(0.327)	(0.078)	1 (0.125)	3 (0.375)
CATTCATC	(0.126)	...	(0.087)	(0.161)	2 (0.25)	1 (0.125)
TATTCATC	(0.022)	(0.019)	(0.020)	(0.113)	...	1 (0.125)
TATTCACC	(0.019)	...	...	...	...	...
CATTCGCC	(0.014)	...	...	...	...	1 (0.125)
CAGTCATC	...	...	...	(0.013)	...	...
CAGTCGTC	...	...	...	(0.024)	...	...
CGGCCGTC	...	...	...	(0.050)	...	...
CGGCTATC	...	...	...	(0.031)	...	...
CGGCTGTC	...	(0.019)	(0.029)	(0.235)	...	...
CGGCTGTT	...	...	...	(0.016)	...	...
Bloque 2	Marcadores: rs3748229, rs17112076					
CC	(0.481)	(0.447)	(0.490)	(0.454)	3 (0.375)	4 (0.5)
AC	(0.329)	(0.262)	(0.406)	(0.541)	3 (0.375)	2 (0.25)
CT	(0.190)	(0.291)	(0.101)	...	2 (0.25)	2 (0.25)

Tabla 5.4. (Continuación)

KRAS	EUR	EAS	AMR	AFR	Prematuros severos, n *	Controles, n *
Bloque	Marcadores: rs12245, rs12587, rs8720, rs1137196, rs1137189, rs1137188, rs13096, rs61764370, rs7973623, rs9266, rs712					
TGCGTACAGGC	(0.436)	(0.738)	(0.473)	(0.063)	4	...
ATTTAGTAAAA	(0.235)	(0.092)	(0.187)	(0.117)	...	...
ATTTAGTAGAA	(0.144)	(0.146)	(0.231)	(0.615)	2	3
ATTTAGTCGAA	(0.101)	...	(0.071)	...	...	...
TGCTTACAGGC	(0.083)	(0.024)	(0.032)	(0.096)	...	...
TGCTTATAGGA	...	...	...	(0.075)	...	...
TGCTTACAGGA	...	...	...	(0.019)	...	...
TGCGAACAGGC	...	...	...	...	2	...
TTTTAGTCGAA	...	...	...	...	...	1
TGTTTACAAAC	...	...	...	...	...	1
TGCGTATAGGC	...	...	...	...	...	1

GNG2	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs2000006, rs3825596, rs3204006, rs3204008, rs8014838, rs1540703, rs45595035, rs2357306					
GGGCTGCA	(0.807)	(0.370)	(0.577)	(0.272)	5 (0.625)	3 (0.375)
GGGCTGTA	(0.061)	...	(0.042)	(0.068)	...	2 (0.25)
GAGCTGCA	(0.051)	(0.154)	(0.034)	(0.040)	1 (0.125)	...
GGGCAGCA	(0.033)	...	(0.018)	...	1 (0.125)	1 (0.125)
CGCGTCCT	(0.016)	(0.233)	(0.132)	(0.078)	...	...
GGCCTGCA	...	...	(0.078)	(0.283)	...	...
GGCCTCCT	...	(0.034)	(0.019)	(0.191)	...	...
GGA CTGCA	...	...	...	...	1 (0.125)	...
CGGCTCCT	...	...	...	...	...	1 (0.125)
CGCGACCT	...	(0.199)	(0.094)	(0.057)	...	...
CGAGACCT	...	...	...	...	...	1 (0.125)

Tabla 5.4. (Continuación)

CD300E	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs554592, rs555542, rs578460, rs479782					
ATAC	(0.514)	(0.624)	(0.571)	(0.125)	3 (0.375)	4 (0.5)
TCGT	(0.302)	(0.138)	(0.222)	(0.467)	3 (0.375)	...
ACAC	(0.182)	(0.233)	(0.205)	(0.390)	2 (0.25)	3 (0.375)
TCGC	...	...	...	(0.017)	...	1 (0.125)
ICAM3	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs2230399, rs2304237, rs2304240					
CTG	(0.507)	(0.699)	(0.469)	(0.668)	1 (0.125)	5 (0.625)
CCG	(0.254)	(0.146)	(0.169)	(0.049)	3 (0.375)	...
CTA	(0.167)	(0.150)	(0.327)	(0.052)	2 (0.25)	3 (0.375)
GTG	(0.052)	...	(0.024)	(0.151)	1 (0.125)	...
GTA	(0.020)	...	...	...	1 (0.125)	...
GCG	...	...	...	(0.077)	...	...
LRRC25	EUR	EAS	AMR	AFR	Prematuros severos, n (frecuencia)	Controles, n (frecuencia)
Bloque	Marcadores: rs12461382, rs3848647, rs3848648, rs4328570, rs116399833, rs6512262, rs6512263, rs6512264, rs6512265					
CACCCGACG	(0.364)	(0.282)	(0.294)	(0.128)	4 (0.5)	1 (0.125)
CGTTACCAA	(0.243)	(0.272)	(0.166)	(0.090)	1 (0.125)	5 (0.625)
TGTTCCCAA	(0.222)	(0.388)	(0.354)	(0.110)	1 (0.125)	1 (0.125)
CGTTCCCAA	(0.171)	(0.015)	(0.179)	(0.669)	2 (0.25)	1 (0.125)
CGCCCGACG	...	(0.019)	...	...	...	...
CGTTACCCA	...	(0.019)	...	...	...	...
Super-poblaciones 1000 genomas: EUR: europeos, EAS: asiáticos, AMR: latinoamericanos, AFR: africanos.						
*Se presenta conteo y frecuencia para todos los genes excepto <i>KRAS</i> , en el cual sólo se presenta el conteo de algunos haplotipos por no poder estimarlos con confianza en algunas muestras.						

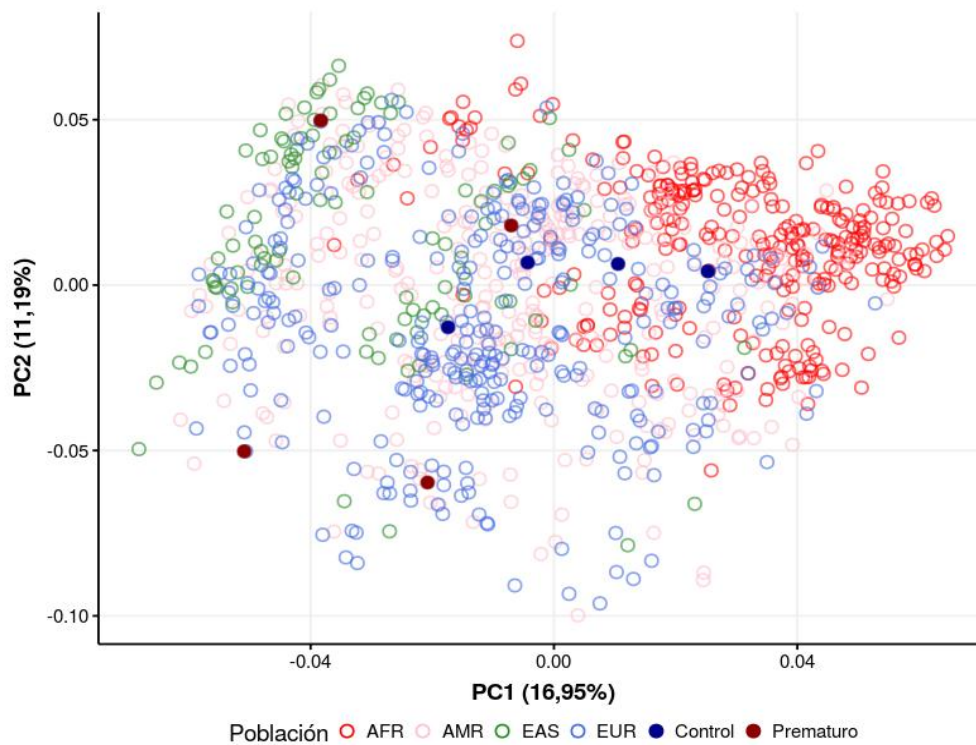


Figura 5.3. Análisis de componentes principales de los genotipos de SNPs seleccionados para 9 genes diferencialmente expresados entre controles y prematuros severos. AFR: africanos. AMR: latinoamericanos. EAS: asiáticos. EUR: europeos.

periferia de la distribución, mostrando tener más variabilidad. Para los 9 genes analizados, el PCA generó agrupamientos esperados según cada una de las superpoblaciones.

5.4 DISCUSIÓN

La búsqueda de variantes génicas asociadas al PPT data de la década de los '90, y si bien se han reportado muchas asociaciones positivas, muy pocas han sido reproducidas en nuevos estudios. En este capítulo, identificamos variantes génicas candidatas a afectar el PPT a partir de datos de ARN-seq en un diseño caso-control,

centrando el análisis en los 270 genes diferencialmente expresados en el capítulo anterior. Para ello realizamos un estudio exploratorio *in silico* centrándonos en los probables actores clave en las enfermedades genéticas complejas: las variantes raras.

La mayoría de las variantes génicas encontradas en nuestros datos se encuentran en regiones de tipo 3'UTR o exónicas. Encontramos en menor proporción variantes en regiones intergénicas. Este resultado, no esperado para ARN-seq, se puede deber a la presencia de ARNm con procesamiento postranscripcional incompleto, exones no anotados o transcritos que exceden la región exónica hacia los intrones. Cualquiera de estas situaciones resultaría en una baja cobertura de secuencia (*i.e.* bajo número de lecturas) de la región en el ARN-seq y, por lo tanto, mayor dificultad en asignar variantes. En consecuencia, esto podría resultar en más cantidad de falsos positivos. Por esta razón centramos nuestro análisis en variantes exónicas y en regiones UTRs.

5.4.1. VARIANTES NOVEDOSAS

Una de las contribuciones más importantes de este estudio es la identificación de 125 variantes no reportadas en las bases de datos genómicas, ubicadas en diferentes regiones del genoma. Aproximadamente un tercio de ellas están en regiones regulatorias. Entre estas variantes novedosas se detectaron dos variantes exónicas con cambio de sentido, de las cuales una tiene efecto deletéreo. Notablemente, esta variante se encuentra en homocigosis en un individuo prematuro y se encuentra en el gen *ZNF467*. *ZNF467* estimula la trans-activación de STAT3, un transductor de señal y activador de la transcripción que media la respuesta celular a interleuquinas (Nakayama et al., 2002; Sanders et al., 2012). Así STAT3 actúa como un regulador de la respuesta inflamatoria. Puede translocar al núcleo para reconocer un elemento de respuesta blanco en el ADN o también reclutar varios factores de transcripción o modificadores epigenéticos para alterar la regulación génica. Se sugiere que

STAT3 forma parte del mecanismo de determinación del inicio del parto, regulando positivamente la expresión de genes pro-trabajo de parto (Yu et al., 2015). Es factible proponer que una mutación deletérea en *ZNF467* impida que se active *STAT3*, alterando la señalización a interleuquinas y por lo tanto la respuesta inflamatoria, que puede causar el PPT.

Por otro lado, *ZNF467* regula la vía de señalización de Wnt en células madre derivadas de adipocitos (You et al., 2015). Como mencionamos en el capítulo anterior, esta vía podría contribuir a abortos y PPT, mediante la disminución de la proliferación e invasión de trofoblastos (Zmijanac Partl et al., 2018).

Estas variantes deben ser confirmadas en el ADN genómico de estos pacientes, para descartar un fenómeno de haplo expresión o expresión monoalélica, en el cual sólo una de las dos copias del gen es activa, elegida al azar. Este fenómeno es frecuente en mamíferos (12 a 24%) (Deng et al., 2014).

5.4.2. PRESENCIA DE VARIANTES GÉNICAS RARAS

Las variantes génicas comunes explican sólo una porción de la heredabilidad de las características complejas. Desde hace ya una década, se propone que son las variantes raras en genes candidatos de efecto mayor quienes contribuyen de forma sustancial al desarrollo de las patologías (Bodmer & Bonilla, 2008; Keinan & Clark, 2012). Esto es consistente con la perspectiva evolutiva, que predice que, dado que las patologías son perjudiciales para el *fitness* evolutivo, las variantes que promueven la enfermedad deberían ser seleccionadas en contra, y por lo tanto deberían ser poco frecuentes (Gibson, 2012).

Identificamos 93 variantes raras en nuestros datos, donde la mayoría se presenta en los individuos prematuros severos. La asignación de variantes genéticas se realizó usando niveles exigentes para su filtrado, por lo que consideramos poco probable que sean artefactos técnicos. Las variantes descritas podrían explicar una

parte de la heredabilidad del parto prematuro. Para poder determinar el grado en el que estas variantes contribuyen a la heredabilidad de este síndrome es necesario primero corroborarlas en el ADN de las membranas fetales que analizamos en este estudio, y luego estudiarlas en una muestra de mayor tamaño, analizando no sólo el transcriptoma sino también la secuencia de exomas y/o genoma completo.

En nuestro trabajo nos centramos en los genes DE, por lo que únicamente detectamos efectos de las variantes en *cis*. De hecho, la detección de variantes que tengan efectos en *trans* es difícil con los tamaños de muestra actuales (Westra et al., 2013), pero se estima que ~70% de la heredabilidad del ARNm está determinada por factores en *trans* (Price et al., 2011).

Debido al pequeño tamaño de la muestra con que se cuenta, las variantes raras y/o novedosas encontradas en estos transcriptomas pueden tener dos orígenes: pueden ser variantes presentes en la población uruguaya, derivadas de las poblaciones ancestrales que la conformaron, o bien pueden ser *singletons*, propios de cada individuo. En el primer caso las variantes están presentes con una frecuencia considerable (mayor al 1%) y son el resultado del aporte de las poblaciones nativo americanas, de las cuales no se cuenta con información genómica disponible, o por la deriva génica producto de las inmigraciones europeas o africanas. Con respecto a los *singletons*, sabemos que contribuyen en casi un 23% de la variación fenotípica humana, al menos en las poblaciones europeas (Hernandez et al., 2017). Esto está en concordancia con la idea de que variantes raras en genes candidatos de efecto mayor son una de las fuentes más importantes para explicar la heredabilidad perdida (Bodmer & Bonilla 2008; Gibson 2012). En cualquiera de los dos casos es necesario caracterizar estas variantes identificadas en una muestra mayor de la población uruguaya, ya que no existen referencias en la literatura hasta el momento.

5.4.3. GENES CON DIFERENCIAS GENOTÍPICAS ENTRE CASOS Y CONTROLES

El tamaño muestral reducido de nuestro diseño no permitió encontrar diferencias significativas entre casos y controles que puedan asociarse a la prematuridad. Sin embargo, sí se encuentran diferencias significativas entre grupos cuando se compara la suma de los conteos de genotipos y cuando se compara la suma del conteo de alelos.

Encontramos 9 genes con diferencias en el conteo de genotipos entre casos y controles. Tal como se esperaba, el PCA para esos genes separa las superpoblaciones europeas, asiáticas y africanas definidas en el proyecto 1000 genomas, que se agrupan de forma definida. Las poblaciones latinoamericanas (grupo AMR) se disponen de forma dispersa, esperable por su origen mestizado. Estos agrupamientos reflejan la arquitectura genética propia de cada población. Las muestras de controles a término se ubican de forma aglomerada, y su disposición en el PCA es concordante con la ancestría mestizada de nuestra población. La población uruguaya tiene una contribución predominantemente europea y contribuciones nativo americanas y africanas de 10,4 y 5,6%, respectivamente (Hidalgo et al., 2005; Sans et al., 1997). El hecho de que las muestras de prematuros severos presentan más variabilidad en su composición genotípica, como se observa en el PCA y en el análisis transcriptómico del capítulo anterior, indica una composición genética diferente, probablemente propia de la patología.

Uno de estos genes (*MARCH1*) presentó diferencias significativas en la distribución de haplotipos entre casos y controles. Para la mayoría de los 9 genes encontramos varios haplotipos no reportados en el proyecto 1000 genomas. La distribución de haplotipos y de alelos para una región determinada es característica de cada población. En nuestro caso, el bajo número de muestras nos impide hacer una estimación de las frecuencias haplotípicas para la población uruguaya, pero podemos reportar la presencia de haplotipos propios en la misma. La diversidad

haplotípica y genética existente entre las poblaciones hace que las bases genéticas de las características puedan ser diferentes, algo probablemente más importante aún en enfermedades complejas y heterogéneas como el PPT.

Todos estos genes tienen en común que participan en mecanismos de inflamación y respuestas inmunes. Cinco de ellos ya han sido asociados con PPT según la base de datos GeneStation (Kim et al., (2015), ver Tabla suplementaria 1.d): *FAM129A*, *PLEK*, *MARCH1*, *KRAS* e *ICAM3*.

El gen *FAM129A* codifica la proteína Niban, que está sobreexpresada en varios tipos de cáncer y presumiblemente protege a las células de apoptosis (Matsumoto et al., 2006). Niban regula positivamente traducción mediante la fosforilación de proteínas involucradas en este proceso (Sun et al., 2007). También promueve la respuesta inflamatoria en células epiteliales aéreas (King et al., 2013; McGeachie et al., 2018).

El gen *PLEK* se encuentra sobreexpresado en varias enfermedades de base inflamatoria (Lundmark et al., 2015). *PLEK* es un sustrato de proteína quinasa C expresado por macrófagos. Se sugiere que es un intermediario importante en la secreción y activación de vías de las citoquinas pro-inflamatorias $TNF\alpha$ y $IL1\beta$ (Ding et al., 2007).

El gen *MARCH1* cuenta con una variante exónica que está identificada como un eQTL (rs13130399). *MARCH1* puede activar la vía Wnt/ β -catenina y la vía NF- κ B (Meng et al., 2016) y se considera un actor clave en la patogénesis de varias enfermedades inflamatorias, mediante la regulación de la inmunidad innata (Galbas et al., 2017). La frecuencia del alelo alternativo de este eQTL en casos y controles es concordante con la expresión génica diferencial evidenciada en el capítulo anterior, donde se encontró una mayor expresión génica en membranas fetales en prematuros severos para *MARCH1*.

El adaptador *PIK3AP1* es necesario como nexo principal entre la vía de señalización TLR y la activación de la vía PI3K para controlar respuestas

inflamatorias. Por su habilidad de activar *PIK3*, *PIK3AP1* regula negativamente la producción de citoquina inflamatoria inducida por *TLR* (Ni et al., 2012).

KRAS es un proto-oncogen, que está mutado en varios tipos de cáncer. Las proteínas Ras pueden promover diferentes tipos de fenotipos, incluyendo proliferación descontrolada, pérdida de inhibición por contacto celular, motilidad aumentada y pérdida de integridad genómica. Este rango de fenotipos dependerá del número de vías efectoras que Ras active, siendo las más validadas las vías Raf/MAPK y PI3K (Yuan et al., 2018).

El gen *GNG2* codifica una de las subunidades gamma de una proteína de unión al nucleótido guanina, que está involucrada en el mecanismo de señalización a través de membranas. Está involucrado en proliferación celular, diferenciación, invasión y angiogénesis (Schwindinger & Robishaw, 2001). Cuando dimeriza puede activar efectores intracelulares (*e.g.* PI3K) o el mecanismo de activación de la proteína Ras.

CD300E es un receptor de tipo inmunoglobulina que transmite una señal de activación inmune (Brckalo et al., 2010). Este tipo de receptores activadores regulan la movilización de calcio, activación transcripcional producción de citoquinas, y/o proliferación y diferenciación celular (Ravetch & Lanier, 2000).

ICAM3 es uno de los ligandos más importantes del antígeno de linfocitos 1 (LFA1) en la iniciación de la respuesta inmune (Montoya et al., 2002). Es una molécula de adhesión primaria y señalización, involucrada en interacciones entre células inmunes y activación de linfocitos T (Montoya et al., 2002). Media la vía de señalización PI3K-Akt para promover la inflamación (Sheng et al., 2016) y para estimular la migración de células cancerosas (Park et al., 2010).

LRRC25 pertenece a la familia de proteínas conteniendo repetidos ricos en Leucina, que tienen funciones en la inmunidad innata (Ng et al., 2011). Participa en la inhibición de la vía NF- κ B activada por el liposacárido bacteriano y TLR4, y su consecuente respuesta inflamatoria (Feng et al., 2017).

Por las diferencias en las frecuencias genotípicas encontradas y su vinculación con vías metabólicas ya vinculadas al PPT, esos genes son buenos candidatos para estudios futuros que evalúen su efecto sobre este síndrome.

5.4.4. EFECTOS FUNCIONALES DE LAS VARIANTES EN LOS 9 GENES

Encontramos que los motivos de secuencia y estructurales de las UTRs posiblemente sean relevantes en el control de la expresión génica en nuestros datos. Las variantes génicas en regiones reguladoras (SNPs “reguladores”) juegan un papel importante en la susceptibilidad de enfermedades (Pai et al., 2015). Las regiones UTRs tienen motivos de secuencia y estructurales que pueden regular la eficiencia transcripcional y traduccional, así como la estabilidad, transporte y localización de los ARNm (Skeeles et al., 2013). Esta regulación funcional se efectúa mediante elementos en cis presentes en las UTRs 5' y 3', a los cuales se le unen proteínas de unión al ARN que actúan en trans u otros ARN no codificantes. Por ejemplo, los miARNs son ARNs no codificantes (de 22 nucleótidos de largo) que interaccionan con sitios blanco usualmente presentes en los 3'UTR, resultando en el clivado del ARNm blanco o en la represión de su traducción (Kim & Bartel, 2009). Un SNV en una región UTR puede alterar estos sitios de unión, alterar la regulación post-transcripcional y potencialmente alterar el fenotipo. En nuestros datos, encontramos 21 variantes génicas que se encuentran en un sitio blanco de miARNs en ARN mensajeros, y cuya variación puede alterar la regulación traduccional. Ninguna de ellas se encuentra en alguno de los sitios blanco de unión de miARNs presumiblemente asociados con PPT reportados en la bibliografía (Cook et al., 2019; Elovitz et al., 2014; Sanders et al., 2015).

A pesar de que los sitios de unión de miARNs son evolutivamente conservados y las SNV en estas regiones son raras, hay varias patologías que se han asociado a ellas (Jazdzewski et al., 2008; Mencía et al., 2009; Nicoloso et al., 2010). Por ejemplo, los SNP rs712 y rs61764370, del gen *KRAS*, están asociados a

susceptibilidad de varios tipos de cáncer (Jiang et al., 2015) y se proponen como biomarcadores de los mismos, así como de endometriosis e incluso como predictor de respuesta a fármacos (Kim & Slack, 2014). Se propone que la presencia del alelo alternativo en estos SNPs disrumpe un sitio de unión conservado para varios miARN, afectando así la afinidad del ARNm con miARNs (Jiang et al., 2015). En particular, cada uno altera un nucleótido del sitio 1 de complementariedad a la familia de miARN *let-7* y cambia la estructura secundaria del ARNm (Kim et al., 2014). Consecuentemente, la expresión del gen *KRAS* cambia, contribuyendo a la progresión de la enfermedad. En nuestros análisis las frecuencias genotípicas de los SNVs del gen *KRAS*, incluyendo a rs712 y rs61764370, tienen diferencias significativas entre prematuros severos y controles. En particular, rs712 tiene una frecuencia alélica del alelo alternativo que es mayor en casos que en controles, lo que es coherente con el aumento de la expresión en el gen *KRAS* en prematuros severos observado en los análisis de expresión génica.

Dos de las variantes génicas encontradas en estos 9 genes son deletéreas: rs17035364 y rs3816281, ambas en el gen *PLEK*. Estas constituyen dos buenas candidatas para estudiar su efecto en el parto prematuro.

5.4.5. USO DEL ARN-SEQ PARA IDENTIFICAR VARIANTES GÉNICAS

La técnica de ARN-seq permite evaluar los niveles de expresión génica e información de variantes génicas al mismo tiempo. Una ventaja de su uso para identificar variantes génicas es que permite focalizar el estudio de variantes en genes que efectivamente se están expresando en el momento de colecta de la muestra. Como las regiones expresadas constituyen una fracción pequeña del genoma humano, es posible lograr cobertura suficiente de secuenciado en estas regiones que permita la asignación de variantes génicas.

Esto mismo puede ser considerado también una limitante del estudio, dado que el poder de identificar variantes se basa en una buena expresión génica: *e.g.* genes

pobrementemente expresados no tendrán buena cobertura y por lo tanto no se detectarán variantes, si las hubiera. Tampoco reconoceremos variantes existentes que tengan expresión monoalélica. También hay que tener en cuenta que el protocolo de retrotranscripción necesario para realizar el secuenciado de ARN se basa en técnicas de PCR, y por lo tanto puede introducir errores. Otros sesgos propios del ARN-seq también pueden dificultar la asignación de variantes: sesgos en la unión de los hexámeros al azar, sesgo de regiones 3' por el uso de un protocolo de enriquecimiento poli-A, degradación de ARN, etc. Por estas razones, su mayor sensibilidad se alcanza cuando se integra la información genómica y transcriptómica (Wilkerson et al., 2014). Teniendo estas consideraciones en cuenta, igualmente comprobamos que el ARN-seq permite encontrar variantes génicas.

En conclusión, en este capítulo la aproximación utilizada nos permitió reportar variantes génicas novedosas de interés potencial para el PPT. Encontramos 9 genes en los cuales hay diferencias genotípicas entre parto a término y a pretérmino severo, con SNVs con efectos funcionales potenciales sobre su expresión. Estos SNV son buenos candidatos para estudiar en una muestra mayor su efecto sobre el PPT. Al ser un síndrome de base compleja, el estudio del PPT requiere la integración de diferentes tipos de datos. Los SNVs encontrados en este grupo de genes DE potencialmente modulan la expresión génica. Es necesario confirmar su presencia en el ADN genómico de estas membranas fetales, y luego estudiarlos en una muestra de la población uruguaya de mayor tamaño.

6

CONCLUSIONES GENERALES

Uno de los mayores desafíos de la genética humana es poder analizar las características genéticas que determinan que algunos individuos de una población sufran determinadas patologías, de forma que podamos entender qué causa las enfermedades. Para capturar la complejidad de las enfermedades humanas, no alcanza con estudiar sólo un nivel de complejidad biológica (genoma, transcriptoma, proteoma) sino que es necesario entender los mecanismos biológicos en forma holística y en su interacción con el medio social y ambiental del individuo. Con esto en mente, nos planteamos varias estrategias para estudiar una característica compleja como es el PPT, con las que identificamos una serie de genes vinculados con éste síndrome.

En primer lugar, nuestra aproximación de estudio de genes candidatos permitió demostrar que es necesario analizar tanto el genoma materno como el fetal para entender su contribución al PPT. La estrategia combinada utilizando datos genéticos y ambientales permitió desentrañar los efectos genéticos en el síndrome

estudiado, que no hubieran sido detectados utilizando un diseño experimental clásico. Aportamos evidencia nueva de que las variables ambientales necesariamente deben ser tenidas en cuenta en estudios genéticos. Esta aproximación también permitió detectar una interacción entre los polimorfismos genéticos maternos y fetales, que potencian su efecto para desarrollar el PPT. Detectamos interacciones génicas entre loci maternos y fetales. Esto forma parte de la complejidad del estudio del PPT, en el que participan e interaccionan dos genomas diferentes. En estudios futuros, ambos deben ser incluidos en estudios genéticos del PPT. Esto además dará la posibilidad de estudiar más profundamente la contribución de interacciones epistáticas materno-fetal, así como interrogantes con respecto a la incompatibilidad entre sus genotipos.

En segundo lugar, caracterizamos el transcriptoma del PPT severo en membranas fetales en nuestra población. Nuestro estudio es el primero hasta el momento en realizar ARN-seq en membranas fetales para el PPT espontáneo. El transcriptoma del PPT severo tiene una sobreexpresión de genes participantes en vías inflamatorias. Identificamos genes y vías metabólicas previamente reportados, así como también nuevos candidatos involucrados con el PPT. Hasta el momento, la mayoría de los estudios de expresión génica del PPT fueron realizados en placenta (Eidem et al., 2015). El hallazgo de vías inesperadas para PPT en nuestro trabajo resalta la utilidad de continuar estudiando tejidos gestacionales de origen exclusivamente fetal, que puede tener patrones de expresión diferentes a los encontrados en las madres.

Mediante el análisis *in silico* de variantes génicas en el transcriptoma detectamos variantes raras y variantes novedosas en los genes diferencialmente expresados entre casos y controles. Entre ellas, identificamos una mutación nueva de interés en una de las muestras prematuras severas, con cambio de sentido y predicción deletérea (7:149462528_A/G), así como variantes con potencial para controlar la expresión génica, como sitios blanco de micro ARNs (por ejemplo, rs712 y rs61764370), y 9 genes diferencialmente expresados en las que sus variantes

génicas también presentan diferencias en las frecuencias genotípicas entre casos y controles.

Los resultados obtenidos están en concordancia con nuestra hipótesis del desencadenamiento del PPT como un proceso inflamatorio, habiendo encontrado asociación génica entre genes candidatos inflamatorios con el PPT, sobreexpresión en prematuros severos de vías inflamatorias y diferencias genotípicas entre casos y controles en genes inflamatorios. Incluso las variantes encontradas también están relacionadas con vías de inflamación. Estos genes constituyen candidatos fuertes para usar como base en futuros estudios del PPT severo.

En esta tesis presentamos el primer transcriptoma mediante ARN-seq del PPT en una población latinoamericana. La mayoría de los estudios genéticos en enfermedades humanas se realizan en poblaciones de origen europeo, lo que muy probablemente limita nuestro entendimiento de la arquitectura genética de las mismas. En este sentido, nuestro aporte en esta población mestizada constituye un antecedente importante en el estudio del PPT en poblaciones Latinoamericanas, con el propósito final de mejorar la calidad de la asistencia médica en ellas.

6.1 PERSPECTIVAS DE TRABAJO

Los aportes realizados en esta tesis son la base de futuros análisis de las vías que contribuyen en el desencadenamiento del PPT. Los resultados obtenidos a partir de los estudios transcriptómicos pueden ser comparados con otras tecnologías para poder ser interpretados mejor. En particular, los genes diferencialmente expresados entre PPT severo y partos a término son buenos candidatos para analizar en el ADN de una muestra poblacional de mayor tamaño, de manera de confirmar su rol en el desencadenamiento del síndrome. En este estudio, los 9 genes DE con diferencias

en las frecuencias genotípicas entre casos y controles serán de interés especial, así como la variante novedosa con cambio de sentido encontrada.

Por otro lado, será de gran interés evaluar el papel de las variantes raras encontradas en el PPT en una muestra independiente. Para ello, primeramente, las variantes raras y variantes novedosas identificadas deberán ser corroboradas en el ADN de las membranas fetales secuenciadas. Luego, las variantes confirmadas serán estudiadas en una muestra uruguaya de mayor tamaño de díadas madre-hijo, ya disponible en nuestro laboratorio. Esperamos así determinar el grado en el que contribuyen a la heredabilidad del PPT.

Los diferentes niveles de complejidad del embarazo y el parto requieren la integración de varios tipos de datos de gran escala. Esta tesis se focaliza en la genética y transcriptómica del PPT, pero la epigenética y la proteómica también pueden jugar un papel importante en su patofisiología. La epigenética contribuye a la variación en la expresión génica, y así a la variación en los fenotipos de enfermedades genéticas complejas. Los genes diferencialmente expresados encontrados en este trabajo son buenos candidatos para estudiar diferencias en marcas epigenéticas en sus promotores, que puedan explicar las diferencias observadas entre PPT severo y controles. Con este objetivo, ya tenemos en marcha estudios epigenéticos en PPT. Seleccionamos 15 genes con islas CpG y sitios de unión a factores de transcripción en sus promotores, por ser mejores candidatos a encontrar señales de metilación relevantes para modular la expresión génica. Realizamos la amplificación por PCR de muestras ADN genómico tratadas con bisulfito de sodio provenientes de membranas fetales de recién nacidos PPT severos y a término. Los productos de PCR fueron secuenciados masivamente mediante una plataforma MinION en nuestro laboratorio y estamos actualmente analizando los resultados de la secuenciación. Esperamos confirmar la hipótesis de que la epigenética modula la expresión génica de genes involucrados en el parto, e incide así en el PPT.

En resumen, en esta tesis usamos diferentes estrategias para aportar conocimiento sobre el papel de la genética en el PPT para una muestra poblacional de nuestro país. Confirmamos el papel de la inflamación en el PPT e identificamos vías metabólicas no vinculadas previamente con prematurez. Identificamos un grupo de genes y de variantes génicas de interés para su estudio en relación al PPT, que serán estudiados para confirmar su contribución en el PPT en nuestra población.

7

REFERENCIAS

1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., ... Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74.

Ackerman, W. E., 4th, Buhimschi, I. A., Brubaker, D., Maxwell, S., Rood, K. M., Chance, M. R., ... Buhimschi, C. S. (2018). Integrated microRNA and mRNA network analysis of the human myometrial transcriptome in the transition from quiescence to labor. *Biology of Reproduction*, 98(6), 834–845.

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*, 124(4), 783–801.

Alvarez-Silva, M., Belo-Diabangouaya, P., Salaün, J., & Dieterlen-Lièvre, F. (2003). Mouse placenta is a major hematopoietic organ. *Development*, 130(22), 5437–5444.

Ananth, C. V., & Vintzileos, A. M. (2006). Epidemiology of preterm birth and its clinical subtypes. *The Journal of Maternal-Fetal & Neonatal Medicine*, 19(12), 773–782.

- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169.
- Andrews, S. (2010) FastQC: A quality control tool for high throughput sequence data. Obtenido de <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Anum, E. A., Springel, E. H., Shriver, M. D., & Strauss, J. F., 3rd (2009). Genetic contributions to disparities in preterm birth. *Pediatric research*, 65(1), 1–9.
- Arbour, N. C., Lorenz, E., Schutte, B. C., Zabner, J., Kline, J. N., Jones, M., ... Schwartz, D. A. (2000). TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nature Genetics*, 25(2), 187–191.
- Arechavaleta-Velasco, F., Marciano, D., Díaz-Cueto, L., & Parry, S. (2004). Matrix metalloproteinase-8 is expressed in human chorion during labor. *American Journal of Obstetrics and Gynecology*, 190(3), 843–850.
- Badano, J. L., & Katsanis, N. (2002). Human genetics and disease: Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Reviews. Genetics*, 3(10), 779.
- Bakketeig, L. S., Hoffman, H. J., & Harley, E. E. (1979). The tendency to repeat gestational age and birth weight in successive births. *American Journal of Obstetrics and Gynecology*, 135(8), 1086–1103.
- Behrman, R. E., & Butler, A. S. (2007). *Preterm birth: Causes, consequences, and prevention* (p. 791). Washington DC: National Academies Press.
- Behrman, R. E., Butler, A. S., & Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes. (2007). *Biological Pathways Leading to Preterm Birth*. National Academies Press (US).
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 57(1), 289–300.
- Berkowitz, G. S., & Papiernik, E. (1993). Epidemiology of Preterm Birth. *Epidemiologic Reviews*, 15(2), 414–443.
- Bhattacharya, A., Ziebarth, J. D., & Cui, Y. (2014). PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Research*, 42(Database issue), D86–D91.
- Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.-B., Narwal, R., ... Lawn, J. E. (2012). National, regional, and worldwide estimates of preterm

birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet*, 379(9832), 2162–2172.

Bodmer, W., & Bonilla, C. (2008). Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics*, 40(6), 695–701.

Bollopargada, S., Youssef, R., Jordan, F., Greer, I., Norman, J., & Nelson, S. (2009). Term labor is associated with a core inflammatory response in human fetal membranes, myometrium, and cervix. *American Journal of Obstetrics and Gynecology*, 200(1): 104.e1-104.e11.

Bonilla, C., Bertoni, B., Hidalgo, P. C., Artagaveytia, N., Ackermann, E., Barreto, I., ... Kittles, R. A. (2015). Breast cancer risk and genetic ancestry: a case-control study in Uruguay. *BMC Women's Health*, 15(1), 11.

Bowen, J. M., Chamley, L., Keelan, J. A., & Mitchell, M. D. (2002). Cytokines of the placenta and extra-placental membranes: roles and regulation during human pregnancy and parturition. *Placenta*, 23(4), 257–273.

Boyle, A. K., Rinaldi, S. F., Norman, J. E., & Stock, S. J. (2017a). Preterm birth: Inflammation, fetal injury and treatment strategies. *Journal of Reproductive Immunology*, 119, 62–66.

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017b). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, 169(7), 1177–1186.

Brckalo, T., Calzetti, F., Pérez-Cabezas, B., Borràs, F. E., Cassatella, M. A., & López-Botet, M. (2010). Functional analysis of the CD300e receptor in human monocytes and myeloid dendritic cells. *European Journal of Immunology*, 40(3), 722–732.

Brown, N. L., Alvi, S. A., Elder, M. G., Bennett, P. R., & Sullivan, M. H. (1998). Regulation of prostaglandin production in intact fetal membranes by interleukin-1 and its receptor antagonist. *The Journal of Endocrinology*, 159(3), 519–526.

Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A., & Roberts, C. T. (2014). Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. *Molecular Human Reproduction*, 20(8), 810–819.

Bukowski, R., Hankins, G. D. V., Saade, G. R., Anderson, G. D., & Thornton, S. (2006). Labor-associated gene expression in the human uterine fundus, lower segment, and cervix. *PLoS Medicine*, 3(6), e169.

- Bukowski, R., Sadovsky, Y., Goodarzi, H., Zhang, H., Biggio, J. R., Varner, M., ... Baldwin, D. A. (2017). Onset of human preterm and term birth is related to unique inflammatory transcriptome profiles at the maternal fetal interface. *PeerJ*, 5, e3685.
- Buniello, A., MacArthur, J. A. L., Cerezo, M., Harris, L. W., Hayhurst, J., Malangone, C., ... Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1), D1005–D1012.
- Buxton, I. L. O., Heyman, N., Wu, Y.-Y., Barnett, S., & Ulrich, C. (2011). A role of stretch-activated potassium currents in the regulation of uterine smooth muscle contraction. *Acta Pharmacologica Sinica*, 32(6), 758–764.
- Carey, V. J. (2019). *gwascat: representing and modeling data in the EMBL-EBI GWAS catalog*. R package version 2.16.0. Obtenido de <https://bioconductor.riken.jp/packages/release/bioc/manuals/gwascat/man/gwascat.pdf>
- Carlborg, O., & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nature Reviews. Genetics*, 5(8), 618–625.
- Challis, J. R., Lockwood, C. J., Myatt, L., Norman, J. E., Strauss, J. F., 3rd, & Petraglia, F. (2009). Inflammation and pregnancy. *Reproductive Sciences*, 16(2), 206–215.
- Chawanpaiboon, S., Vogel, J. P., Moller, A.-B., Lumbiganon, P., Petzold, M., Hogan, D., ... Gülmezoglu, A. M. (2019). Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *The Lancet. Global Health*, 7(1), e37–e46.
- Chepelev, I., Wei, G., Tang, Q., & Zhao, K. (2009). Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Research*, 37(16), e106.
- Chim, S. S. C., Lee, W. S., Ting, Y. H., Chan, O. K., Lee, S. W. Y., & Leung, T. Y. (2012). Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma. *PloS One*, 7(4), e34328.
- Chim, S. S. C., Wong, K. K. W., Chung, C. Y. L., Lam, S. K. W., Kwok, J. S. L., Lai, C.-Y., ... Leung, T.-Y. (2017). Systematic Selection of Reference Genes for the Normalization of Circulating RNA Transcripts in Pregnant Women Based on RNA-Seq Data. *International Journal of Molecular Sciences*, 18(8), 1709.

- Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews. Genetics*, 11(6), 415–425.
- Clausson, B., Lichtenstein, P., & Cnattingius, S. (2000). Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *BJOG: An International Journal of Obstetrics and Gynaecology*, 107(3), 375–381.
- Cook, J., Bennett, P. R., Kim, S. H., Teoh, T. G., Sykes, L., Kindinger, L. M., ... Terzidou, V. (2019). First Trimester Circulating MicroRNA Biomarkers Predictive of Subsequent Preterm Delivery and Cervical Shortening. *Scientific Reports*, 9(1), 5861.
- Cookson, W. O. C., Liang, L., Abecasis, G. R., Moffatt, M. F., & Lathrop, M. (2009). Mapping complex disease traits with global gene expression. *Nature Reviews. Genetics*, 10(3), 184–194.
- Crider, K. S., Whitehead, N., & Buus, R. M. (2005). Genetic variation associated with preterm birth: a HuGE review. *Genetics in Medicine*, 7(9), 593–604.
- Davidson, L. M., & Coward, K. (2016). Molecular mechanisms of membrane interaction at implantation. *Birth Defects Research. Part C, Embryo Today: Reviews*, 108(1), 19–32.
- De Mucio, B., Abalos, E., Cuesta, C., Carroli, G., Serruya, S., Giordano, D., ... Latin American Near Miss Group (LANe-MG). (2016). Maternal near miss and predictive ability of potentially life-threatening conditions at selected maternity hospitals in Latin America. *Reproductive Health*, 13(1), 134.
- Deng, Q., Ramsköld, D., Reinius, B., & Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science*, 343(6167), 193–196.
- Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H., & Goldstein, D. B. (2010). Rare variants create synthetic genome-wide associations. *PLoS Biology*, 8(1), e1000294.
- Ding, Y., Kantarci, A., Badwey, J. A., Hasturk, H., Malabanan, A., & Van Dyke, T. E. (2007). Phosphorylation of pleckstrin increases proinflammatory cytokine secretion by mononuclear phagocytes in diabetes mellitus. *Journal of Immunology*, 179(1), 647–654.

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21.
- Dolan, S. M., Hollegaard, M. V., Merialdi, M., Betran, A. P., Allen, T., Abelow, C., ... Menon, R. (2010). Synopsis of preterm birth genetic association studies: the preterm birth genetics knowledge base (PTBGene). *Public Health Genomics*, 13(7-8), 514–523.
- Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology*, 32(3), 227–234.
- Dudoit, S., Yang, Y. H., Callow, M. J., & Speed, T. P. (2002). Statistical methods for identifying differentially expressed genes in replicated cdna microarray experiments. *Statistica Sinica*, 12(1), 111–139.
- Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8), 1184–1191.
- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews. Genetics*, 11(6), 446–450.
- Eidem, H. R., Ackerman, W. E., McGary, K. L., Abbot, P., & Rokas, A. (2015). Gestational tissue transcriptomics in term and preterm human pregnancies: a systematic review and meta-analysis. *BMC Medical Genomics*, 8(1), 27.
- Eidem, H. R., Rinker, D. C., Ackerman, W. E., 4th, Buhimschi, I. A., Buhimschi, C. S., Dunn-Fletcher, C., ... Rokas, A. (2016). Comparing human and macaque placental transcriptomes to disentangle preterm birth pathology from gestational age effects. *Placenta*, 41, 74–82.
- El-Azzamy, H., Balogh, A., Romero, R., Xu, Y., LaJeunesse, C., Plazyo, O., ... Than, N. G. (2017). Characteristic Changes in Decidual Gene Expression Signature in Spontaneous Term Parturition. *Journal of Pathology and Translational Medicine*, 51(3), 264–283.
- Elovitz, M. A., Brown, A. G., Anton, L., Gilstrap, M., Heiser, L., & Bastek, J. (2014). Distinct cervical microRNA profiles are present in women destined to have a preterm birth. *American Journal of Obstetrics and Gynecology*, 210(3), 221.e1–e11.

- Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A. S., Zink, F., Zhu, J., ... Stefansson, K. (2008). Genetics of gene expression and its effect on disease. *Nature*, 452(7186), 423–428.
- Engström, P. G., Steijger, T., Sipos, B., Grant, G. R., Kahles, A., Räscher, G., ... RGASP Consortium. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods*, 10(12), 1185–1191.
- Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048.
- Farr, M., Strübe, J., Geppert, H. G., Kocourek, A., Mahne, M., & Tschesche, H. (2000). Pregnancy-associated plasma protein-E (PAPP-E). *Biochimica et Biophysica Acta*, 1493(3), 356–362.
- Feng, Y., Duan, T., Du, Y., Jin, S., Wang, M., Cui, J., & Wang, R.-F. (2017). LRRC25 Functions as an Inhibitor of NF- κ B Signaling Pathway by Promoting p65/RelA for Autophagic Degradation. *Scientific Reports*, 7(1), 13448.
- Fishman, D., Faulds, G., Jeffery, R., Mohamed-Ali, V., Yudkin, J. S., Humphries, S., & Woo, P. (1998). The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. *The Journal of Clinical Investigation*, 102(7), 1369–1376.
- Flood, K., & Malone, F. D. (2012). Prevention of preterm birth. *Seminars in Fetal & Neonatal Medicine*, 17(1), 58–63.
- Franchi, L., Eigenbrod, T., Muñoz-Planillo, R., & Nuñez, G. (2009). The inflammasome: a caspase-1-activation platform that regulates immune responses and disease pathogenesis. *Nature Immunology*, 10(3), 241–247.
- Fraser, H. B., & Xie, X. (2009). Common polymorphic transcript variation in human disease. *Genome Research*, 19(4), 567–575.
- Galbas, T., Raymond, M., Sabourin, A., Bourgeois-Daigneault, M.-C., Guimont-Desrochers, F., Yun, T. J., ... Thibodeau, J. (2017). MARCH1 E3 Ubiquitin Ligase Dampens the Innate Inflammatory Response by Modulating Monocyte Functions in Mice. *Journal of Immunology*, 198(2), 852–861.
- Gascue, C., Mimbacas, A., Sans, M., Gallino, J. P., Bertoni, B., Hidalgo, P., & Cardoso, H. (2005). Frequencies of the four major Amerindian mtDNA

haplogroups in the population of Montevideo, Uruguay. *Human Biology*, 77(6), 873–878.

Gekas, C., Dieterlen-Lièvre, F., Orkin, S. H., & Mikkola, H. K. A. (2005). The placenta is a niche for hematopoietic stem cells. *Developmental Cell*, 8(3), 365–375.

Gekas, C., Rhodes, K. E., Van Handel, B., Chhabra, A., Ueno, M., & Mikkola, H. K. A. (2010). Hematopoietic stem cell development in the placenta. *The International Journal of Developmental Biology*, 54(6-7), 1089–1098.

Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews. Genetics*, 13(2), 135–145.

Gimenez, L. G., Momany, A. M., Poletta, F. A., Krupitzki, H. B., Gili, J. A., Busch, T. D., ... Lopez-Camelo, J. S. (2017). Association of candidate gene polymorphisms with clinical subtypes of preterm birth in a Latin American population. *Pediatric Research*, 82(3), 554–559.

Goff, L., Trapnell, C., & Kelley, D. (2013). cummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. *R Package Version*, 2(0).

Goldenberg, R. L. (2002). The management of preterm labor. *Obstetrics and Gynecology*, 100(5 Pt 1), 1020–1037.

Goldenberg, R. L., Culhane, J. F., Iams, J. D., & Romero, R. (2008). Epidemiology and causes of preterm birth. *The Lancet*, 371(9606), 75–84.

Gomez-Lopez, N., Romero, R., Arenas-Hernandez, M., Ahn, H., Panaitescu, B., Vadillo-Ortega, F., ... Hassan, S. S. (2016). In vivo T-cell activation by a monoclonal α CD3 ϵ antibody induces preterm labor and birth. *American Journal of Reproductive Immunology*, 76(5), 386–390.

Gomez-Lopez, N., Romero, R., Xu, Y., Plazyo, O., Unkel, R., Leng, Y., ... Hassan, S. S. (2017). A Role for the Inflammasome in Spontaneous Preterm Labor With Acute Histologic Chorioamnionitis. *Reproductive Sciences*, 24(10), 1382–1401.

Gotsch, F., Romero, R., Chaiworapongsa, T., Erez, O., Vaisbuch, E., Espinoza, J., ... Hassan, S. S. (2008). Evidence of the involvement of caspase-1 under physiologic and pathologic cellular stress during human pregnancy: A link between the inflammasome and parturition. *The Journal of Maternal-Fetal & Neonatal Medicine*, 21(9), 605–616.

GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585.

Haddad, R., Tromp, G., Kuivaniemi, H., Chaiworapongsa, T., Kim, Y. M., Mazor, M., & Romero, R. (2006). Human spontaneous labor without histologic chorioamnionitis is characterized by an acute inflammation gene expression signature. *American Journal of Obstetrics and Gynecology*, 195(2), 394.e1–e24.

Hamilton, S. A., Tower, C. L., & Jones, R. L. (2013). Identification of chemokines associated with the recruitment of decidual leukocytes in human labour: potential novel targets for preterm labour. *PloS One*, 8(2), e56946.

Henderson, J. J., McWilliam, O. A., Newnham, J. P., & Pennell, C. E. (2012). Preterm birth aetiology 2004-2008. Maternal factors associated with three phenotypes: spontaneous preterm labour, preterm pre-labour rupture of membranes and medically indicated preterm birth. *The Journal of Maternal-Fetal & Neonatal Medicine*, 25(6), 642–647.

Heng, Y. J., Liong, S., Permezel, M., Rice, G. E., Di Quinzio, M. K. W., & Georgiou, H. M. (2014). The interplay of the interleukin 1 system in pregnancy and labor. *Reproductive Sciences*, 21(1), 122–130.

Heng, Y. J., Pennell, C. E., Chua, H. N., Perkins, J. E., & Lye, S. J. (2014). Whole blood gene expression profile associated with spontaneous preterm birth in women with threatened preterm labor. *PloS One*, 9(5), e96901.

Hernandez, R. D., Uricchio, L. H., Hartman, K., Ye, J., Dahl, A., & Zaitlen, N. (2017). *Singleton Variants Dominate the Genetic Architecture of Human Gene Expression*. *bioRxiv*. <https://doi.org/10.1101/219238>.

Hidalgo, P. C., Bengochea, M., Abilleira, D., Cabrera, A., & Alvarez, I. (2005). Genetic Admixture Estimate in the Uruguayan Population Based on the Loci LDLR, GYPA, HBGG, GC and D7S8. *International Journal of Human Genetics*, 5(3), 217–222.

Hollegaard, M. V., Grove, J., Thorsen, P., Wang, X., Mandrup, S., Christiansen, M., ... Hougaard, D. M. (2008). Polymorphisms in the tumor necrosis factor alpha and interleukin 1-beta promoters with possible gene regulatory functions increase the risk of preterm birth. *Acta Obstetrica et Gynecologica Scandinavica*, 87(12), 1285–1290.

Holt, R., Timmons, B. C., Akgul, Y., Akins, M. L., & Mahendroo, M. (2011). The molecular mechanisms of cervical ripening differ between term and preterm birth. *Endocrinology*, 152(3), 1036–1046.

- Hong, E. P., & Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & Informatics*, 10(2), 117–122.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., ... Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2), 115–121.
- Huusko, J. M., Karjalainen, M. K., Graham, B. E., Zhang, G., Farrow, E. G., Miller, N. A., ... Muglia, L. J. (2018). Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. *PLoS Genetics*, 14(7), e1007394.
- Iams, J. D. (2014). Preterm birth categories—labels with consequences. *American Journal of Obstetrics and Gynecology*, 210(2), 97–98.
- Jazdzewski, K., Murray, E. L., Franssila, K., Jarzab, B., Schoenberg, D. R., & de la Chapelle, A. (2008). Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. *Proceedings of the National Academy of Sciences of the United States of America*, 105(20), 7269–7274.
- Jiang, Q.-H., Peng, H.-X., Zhang, Y., Tian, P., Xi, Z.-L., & Chen, H. (2015). rs712 polymorphism within let-7 microRNA-binding site might be involved in the initiation and progression of colorectal cancer in Chinese population. *OncoTargets and Therapy*, 8, 3041–3045.
- Jin, C., & Flavell, R. A. (2010). Molecular mechanism of NLRP3 inflammasome activation. *Journal of Clinical Immunology*, 30(5), 628–631.
- Kalish, R. B., Vardhana, S., Normand, N. J., Gupta, M., & Witkin, S. S. (2006). Association of a maternal CD14 -159 gene polymorphism with preterm premature rupture of membranes and spontaneous preterm birth in multi-fetal pregnancies. *Journal of Reproductive Immunology*, 70(1-2), 109–117.
- Kaluarachchi, D. C., Momany, A. M., Busch, T. D., Gimenez, L. G., Saleme, C., Cosentino, V., ... Murray, J. C. (2016). Polymorphisms in NR5A2, gene encoding liver receptor homolog-1 are associated with preterm birth. *Pediatric Research*, 79(5), 776–780.
- Kamath-Rayne, B. D., Du, Y., Hughes, M., Wagner, E. A., Muglia, L. J., DeFranco, E. A., ... Xu, Y. (2015). Systems biology evaluation of cell-free amniotic fluid transcriptome of term and preterm infants to detect fetal maturity. *BMC Medical Genomics*, 8, 67.

- Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30.
- Kawai, T., & Akira, S. (2006). TLR signaling. *Cell Death and Differentiation*, 13(5), 816–825.
- Keinan, A., & Clark, A. G. (2012). Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science*, 336(6082), 740–743.
- Khodadi, E., Shahrabi, S., Shahjahani, M., Azandeh, S., & Saki, N. (2016). Role of stem cell factor in the placental niche. *Cell and Tissue Research*, 366(3), 523–531.
- Kim, C. J., Romero, R., Chaemsaitong, P., & Kim, J.-S. (2015). Chronic inflammation of the placenta: definition, classification, pathogenesis, and clinical significance. *American Journal of Obstetrics and Gynecology*, 213(4 Suppl), S53–S69.
- Kim, J. (2012). Identification of genes contributing to preterm birth: insights from genetic, transcriptomic, and epigenetic analyses. PhD (Doctor of Philosophy) thesis, University of Iowa. <https://doi.org/10.17077/etd.p3zuvpa3>
- Kim, J., & Bartel, D. P. (2009). Allelic imbalance sequencing reveals that single-nucleotide polymorphisms frequently alter microRNA-directed repression. *Nature Biotechnology*, 27(5), 472–477.
- Kim, J., Zhao, K., Jiang, P., Lu, Z.-X., Wang, J., Murray, J. C., & Xing, Y. (2012). Transcriptome landscape of the human placenta. *BMC Genomics*, 13, 115.
- Kim, M., Chen, X., Chin, L. J., Paranjape, T., Speed, W. C., Kidd, K. K., ... Slack, F. J. (2014). Extensive sequence variation in the 3' untranslated region of the KRAS gene in lung and ovarian cancer cases. *Cell Cycle*, 13(6), 1030–1040.
- Kim, M., Cooper, B. A., Venkat, R., Phillips, J. B., Eidem, H. R., Hirbo, J., ... McGary, K. L. (2016). GENE-STATION 1.0: a synthetic resource of diverse evolutionary and functional genomic data for studying the evolution of pregnancy-associated tissues and phenotypes. *Nucleic Acids Research*, 44(D1), D908–D916.
- Kim, M., & Slack, F. J. (2014). MicroRNA-mediated regulation of KRAS in cancer. *Journal of Hematology & Oncology*, 7, 84.
- Kim, S. M., Romero, R., Lee, J., Chaemsaitong, P., Lee, M.-W., Chaiyasit, N., ... Yoon, B. H. (2016). About one-half of early spontaneous preterm deliveries can be identified by a rapid matrix metalloproteinase-8 (MMP-8) bedside test at the time

of mid-trimester genetic amniocentesis. *The Journal of Maternal-Fetal & Neonatal Medicine*, 29(15), 2414–2422.

King, E. M., Chivers, J. E., Rider, C. F., Minnich, A., Giembycz, M. A., & Newton, R. (2013). Glucocorticoid repression of inflammatory gene expression shows differential responsiveness by transactivation- and transrepression-dependent mechanisms. *PloS One*, 8(1), e53936.

Kistka, Z. A.-F., DeFranco, E. A., Ligthart, L., Willemsen, G., Plunkett, J., Muglia, L. J., & Boomsma, D. I. (2008). Heritability of parturition timing: an extended twin design analysis. *American Journal of Obstetrics and Gynecology*, 199(1), 43.e1–e5.

Ko, T.-J., Tsai, L.-Y., Chu, L.-C., Yeh, S.-J., Leung, C., Chen, C.-Y., ... Hsieh, W.-S. (2014). Parental smoking during pregnancy and its association with low birth weight, small for gestational age, and preterm birth offspring: a birth cohort study. *Pediatrics and Neonatology*, 55(1), 20–27.

Kobayashi, Y. (2008). The role of chemokines in neutrophil biology. *Frontiers in Bioscience: A Journal and Virtual Library*, 13(13), 2400.

Korebrits, C., Ramirez, M. M., Watson, L., Brinkman, E., Bocking, A. D., & Challis, J. R. (1998). Maternal corticotropin-releasing hormone is increased with impending preterm birth. *The Journal of Clinical Endocrinology and Metabolism*, 83(5), 1585–1591.

Kuchma, M. D., Kyryk, V. M., Svitina, H. M., Shablii, Y. M., Lukash, L. L., Lobyntseva, G. S., & Shablii, V. A. (2015). Comparative Analysis of the Hematopoietic Progenitor Cells from Placenta, Cord Blood, and Fetal Liver, Based on Their Immunophenotype. *BioMed Research International*, 2015, 418752.

Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 2015(11), 951–969.

Lander, E. S. (1996). The new genomics: global views of biology. *Science*, 274(5287), 536–539.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, 42(Database issue), D980–D985.

Lee, J., Romero, R., Chaiworapongsa, T., Dong, Z., Tarca, A. L., Xu, Y., ... Kim, C. J. (2013). Characterization of the fetal blood transcriptome and proteome in

maternal anti-fetal rejection: evidence of a distinct and novel type of human fetal systemic inflammatory response. *American Journal of Reproductive Immunology*, 70(4), 265–284.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., ... Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews. Genetics*, 11(10), 733–739.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536(7616), 285–291.

Li, X., Kim, Y., Tsang, E. K., Davis, J. R., Damani, F. N., Chiang, C., ... Montgomery, S. B. (2017a). The impact of rare variation on gene expression across tissues. *Nature*, 550(7675), 239–243.

Li, J., Oehlert, J., Snyder, M., Stevenson, D. K., & Shaw, G. M. (2017b). Fetal de novo mutations and preterm birth. *PLoS Genetics*, 13(4), e1006689.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods*, 25(4), 402–408.

Lockwood, C. J., & Kuczynski, E. (2001). Risk stratification and pathological mechanisms in preterm delivery. *Paediatric and Perinatal Epidemiology*, 15 Suppl 2, 78–89.

Lorenz, E., Hallman, M., Marttila, R., Haataja, R., & Schwartz, D. A. (2002). Association between the Asp299Gly polymorphisms in the Toll-like receptor 4 and premature births in the Finnish population. *Pediatric Research*, 52(3), 373–376.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.

Lui, S., Duval, C., Farrokhnia, F., Girard, S., Harris, L. K., Tower, C. L., ... Jones, R. L. (2018). Delineating differential regulatory signatures of the human transcriptome in the choriodecidua and myometrium at term labor. *Biology of Reproduction*, 98(3), 422–436.

Lunde, A., Melve, K. K., Gjessing, H. K., Skjaerven, R., & Irgens, L. M. (2007). Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. *American Journal of Epidemiology*, 165(7), 734–741.

- Lundmark, A., Davanian, H., Båge, T., Johannsen, G., Koro, C., Lundeberg, J., & Yucel-Lindberg, T. (2015). Transcriptome analysis reveals mucin 4 to be highly associated with periodontitis and identifies pleckstrin as a link to systemic diseases. *Scientific Reports*, 5, 18475.
- Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., & Woolf, P. J. (2009). GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*, 10, 161.
- Macones, G. A., Parry, S., Elkousy, M., Clothier, B., Ural, S. H., & Strauss, J. F., 3rd. (2004). A polymorphism in the promoter region of TNF and bacterial vaginosis: preliminary evidence of gene-environment interaction in the etiology of spontaneous preterm birth. *American Journal of Obstetrics and Gynecology*, 190(6), 1504–1508.
- Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature*, 456(7218), 18–21.
- Mahmud, Z. A., Jenkins, L., Ulven, T., Labéguère, F., Gosmini, R., De Vos, S., ... Milligan, G. (2017). Three classes of ligands each bind to distinct sites on the orphan G protein-coupled receptor GPR84. *Scientific Reports*, 7(1), 17953.
- Mann, P. C., Cooper, M. E., Ryckman, K. K., Comas, B., Gili, J., Crumley, S., ... Murray, J. C. (2013). Polymorphisms in the fetal progesterone receptor and a calcium-activated potassium channel isoform are associated with preterm birth in an Argentinian population. *Journal of Perinatology*, 33(5), 336–340.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753.
- Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, 18(9), 1509–1517.
- Marouli, E., Graff, M., Medina-Gomez, C., Lo, K. S., Wood, A. R., Kjaer, T. R., ... Lettre, G. (2017). Rare and low-frequency coding variants alter human adult height. *Nature*, 542(7640), 186–190.
- Martin, J. A., Hamilton, B. E., Osterman, M. J. K., Driscoll, A. K., & Drake, P. (2018). Births: Final Data for 2017. *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 67(8), 1–50.

- Matsumoto, F., Fujii, H., Abe, M., Kajino, K., Kobayashi, T., Matsumoto, T., ... Hino, O. (2006). A novel tumor marker, Niban, is expressed in subsets of thyroid tumors and Hashimoto's thyroiditis. *Human Pathology*, 37(12), 1592–1600.
- Mattern, A., Zellmann, T., & Beck-Sickinger, A. G. (2014). Processing, signaling, and physiological function of chemerin. *IUBMB Life*, 66(1), 19–26.
- Maymon, E., Romero, R., Pacora, P., Gomez, R., Athayde, N., Edwin, S., & Yoon, B. H. (2000). Human neutrophil collagenase (matrix metalloproteinase 8) in parturition, premature rupture of the membranes, and intrauterine infection. *American Journal of Obstetrics and Gynecology*, 183(1), 94–99.
- McElroy, J. J., Gutman, C. E., Shaffer, C. M., Busch, T. D., Puttonen, H., Teramo, K., ... Muglia, L. J. (2013). Maternal coding variants in complement receptor 1 and spontaneous idiopathic preterm birth. *Human Genetics*, 132(8), 935–942.
- McGeachie, M. J., Clemmer, G. L., Hayete, B., Xing, H., Runge, K., Wu, A. C., ... Weiss, S. (2018). Systems biology and in vitro validation identifies family with sequence similarity 129 member A (FAM129A) as an asthma steroid response modulator. *The Journal of Allergy and Clinical Immunology*, 142(5), 1479–1488.e12.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.
- Mencía, A., Modamio-Høybjør, S., Redshaw, N., Morín, M., Mayo-Merino, F., Olavarrieta, L., ... Moreno-Pelayo, M. A. (2009). Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. *Nature Genetics*, 41(5), 609–613.
- Meng, Y., Hu, J., Chen, Y., Yu, T., & Hu, L. (2016). Silencing MARCH1 suppresses proliferation, migration and invasion of ovarian cancer SKOV3 cells via downregulation of NF- κ B and Wnt/ β -catenin pathways. *Oncology Reports*, 36(5), 2463–2470.
- Menon, R., & Fortunato, S. J. (2007). Infection and the role of inflammation in preterm premature rupture of the membranes. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, 21(3), 467–478.
- Menon, R., Pearce, B. D., Velez Edwards, D. R., Merialdi, M., Williams, S. M., Fortunato, S. J., & Thorsen, P. (2009). Racial disparity in pathophysiologic

pathways of preterm birth based on genetic variants. *Reproductive Biology and Endocrinology: RB&E*, 7(1), 62.

Menon, R., & Richardson, L. S. (2017). Preterm prelabor rupture of the membranes: A disease of the fetal membranes. *Seminars in Perinatology*, 41(7), 409–419.

Migale, R., MacIntyre, D. A., Cacciatore, S., Lee, Y. S., Hagberg, H., Herbert, B. R., ... Bennett, P. R. (2016). Modeling hormonal and inflammatory contributions to preterm and term labor using uterine temporal transcriptomics. *BMC Medicine*, 14(1), 86.

Mikheev, A. M., Nabekura, T., Kaddoumi, A., Bammler, T. K., Govindarajan, R., Hebert, M. F., & Unadkat, J. D. (2008). Profiling gene expression in human placentae of different gestational ages: an OPRU Network and UW SCOR Study. *Reproductive Sciences*, 15(9), 866–877.

Mittal, P., Romero, R., Tarca, A. L., Gonzalez, J., Draghici, S., Xu, Y., ... Kim, C. J. (2010). Characterization of the myometrial transcriptome and biological pathways of spontaneous human labor at term. *Journal of Perinatal Medicine*, 38, 617–643.

Montgomery, S. B., & Dermitzakis, E. T. (2011). From expression QTLs to personalized transcriptomics. *Nature Reviews. Genetics*, 12(4), 277–282.

Montoya, M. C., Sancho, D., Bonello, G., Collette, Y., Langlet, C., He, H. T., ... Sánchez-Madrid, F. (2002). Role of ICAM-3 in the initial interaction of T lymphocytes and APCs. *Nature Immunology*, 3(2), 159–168.

Morley, M., Molony, C. M., Weber, T. M., Devlin, J. L., Ewens, K. G., Spielman, R. S., & Cheung, V. G. (2004). Genetic analysis of genome-wide variation in human gene expression. *Nature*, 430(7001), 743–747.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, 5(7), 621–628.

Moszyńska, A., Gebert, M., Collawn, J. F., & Bartoszewski, R. (2017). SNPs in microRNA target sites and their potential role in human disease. *Open Biology*, 7(4)

Moutquin, J.-M. (2003). Classification and heterogeneity of preterm birth. *BJOG: An International Journal of Obstetrics and Gynaecology*, 110 Suppl 20, 30–33.

- MSP (2019). Informe de Gestión y Sistema Informático Perinatal Año 2018. Hospital de la Mujer “Dra. Paulina Luisi”, Centro Hospitalario Pereira Rossell. Obtenido de <http://www.asse.com.uy/contenido/Informe-de-Gestion-y-Sistema-Informatico-Perinatal--ano-2018-11412>.
- Mumtaz, G., Nassar, A. H., Mahfoud, Z., El-Khamra, A., Al-Choueiri, N., Adra, A., ... Yunis, K. A. (2010). Consanguinity: a risk factor for preterm birth at less than 33 weeks’ gestation. *American Journal of Epidemiology*, 172(12), 1424–1430.
- Murray, S. A., Morgan, J. L., Kane, C., Sharma, Y., Heffner, C. S., Lake, J., & Donahue, L. R. (2010). Mouse gestation length is genetically determined. *PloS One*, 5(8), e12418.
- Myatt, L., Eschenbach, D. A., Lye, S. J., Mesiano, S., Murtha, A. P., Williams, S. M., & Pennell, C. E. (2012). A Standardized Template for Clinical Studies in Preterm Birth. *Reproductive Sciences*, 19(5), 474–482.
- Myking, S., Myhre, R., Gjessing, H. K., Morken, N.-H., Sengpiel, V., Williams, S. M., ... Jacobsson, B. (2011). Candidate gene analysis of spontaneous preterm delivery: new insights from re-analysis of a case-control study using case-parent triads and control-mother dyads. *BMC Medical Genetics*, 12, 174.
- Nakayama, K., Kim, K.-W., & Miyajima, A. (2002). A novel nuclear zinc finger protein EZI enhances nuclear retention and transactivation of STAT3. *The EMBO Journal*, 21(22), 6174–6184.
- Ng, A. C. Y., Eisenberg, J. M., Heath, R. J. W., Huett, A., Robinson, C. M., Nau, G. J., & Xavier, R. J. (2011). Human leucine-rich repeat proteins: a genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl 1, 4631–4638.
- Ngo, T. T. M., Moufarrej, M. N., Rasmussen, M.-L. H., Camunas-Soler, J., Pan, W., Okamoto, J., ... Quake, S. R. (2018). Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science*, 360(6393), 1133–1136.
- Nhan-Chang, C.-L., Romero, R., Tarca, A. L., Mittal, P., Kusanovic, J. P., Erez, O., ... Kim, C. J. (2010). Characterization of the transcriptome of chorioamniotic membranes at the site of rupture in spontaneous labor at term. *American Journal of Obstetrics and Gynecology*, 202(5), 462.e1–e41.

- Ni, M., MacFarlane, A. W., 4th, Toft, M., Lowell, C. A., Campbell, K. S., & Hamerman, J. A. (2012). B-cell adaptor for PI3K (BCAP) negatively regulates Toll-like receptor signaling through activation of PI3K. *Proceedings of the National Academy of Sciences of the United States of America*, 109(1), 267–272.
- Nicoloso, M. S., Sun, H., Spizzo, R., Kim, H., Wickramasinghe, P., Shimizu, M., ... Calin, G. A. (2010). Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Research*, 70(7), 2789–2798.
- Noguchi, T., Sado, T., Naruse, K., Shigetomi, H., Onogi, A., Haruta, S., ... Kobayashi, H. (2010). Evidence for activation of Toll-like receptor and receptor for advanced glycation end products in preterm birth. *Mediators of Inflammation*, 2010, 490406.
- Novakovic, B., Fournier, T., Harris, L. K., James, J., Roberts, C. T., Yong, H. E. J., ... Murthi, P. (2017). Increased methylation and decreased expression of homeobox genes TLX1, HOXA10 and DLX5 in human placenta are associated with trophoblast differentiation. *Scientific Reports*, 7(1), 4523.
- Obenchain, V., Lawrence, M., Carey, V., Gogarten, S., Shannon, P., & Morgan, M. (2014). VariantAnnotation: a Bioconductor package for exploration and annotation of genetic variants. *Bioinformatics*, 30(14), 2076–2078.
- Osman, I., Young, A., Ledingham, M. A., Thomson, A. J., Jordan, F., Greer, I. A., & Norman, J. E. (2003). Leukocyte density and pro-inflammatory cytokine expression in human fetal membranes, decidua, cervix and myometrium before and during labour at term. *Molecular Human Reproduction*, 9(1), 41–45.
- Pacora, P., Romero, R., Maymon, E., Gervasi, M. T., Gomez, R., Edwin, S. S., & Yoon, B. H. (2000). Participation of the novel cytokine interleukin 18 in the host response to intra-amniotic infection. *American Journal of Obstetrics and Gynecology*, 183(5), 1138–1143.
- Pagano, S., Sans, M., Pimenoff, V., Cantera, A. M., Alvarez, J. C., Lorente, J. A., ... Sajantila, A. (2005). Assessment of HV1 and HV2 mtDNA variation for forensic purposes in an Uruguayan population sample. *Journal of Forensic Sciences*, 50(5), 1239–1242.
- Pai, A. A., Pritchard, J. K., & Gilad, Y. (2015). The genetic and mechanistic basis for variation in gene regulation. *PLoS Genetics*, 11(1), e1004857.
- Panaitescu, B., Romero, R., Gomez-Lopez, N., Xu, Y., Leng, Y., Maymon, E., ... Hsu, C.-D. (2018). In vivo evidence of inflammasome activation during

spontaneous labor at term. *The Journal of Maternal-Fetal & Neonatal Medicine*, 1–14.

Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genetics*, 40(12), 1413–1415.

Paquette, A. G., Shynlova, O., Kibschull, M., Price, N. D., Lye, S. J., & Global Alliance to Prevent Prematurity and Stillbirth Systems Biology of Preterm Birth Team. (2018). Comparative analysis of gene expression in maternal peripheral blood and monocytes during spontaneous preterm labor. *American Journal of Obstetrics and Gynecology*, 218(3), 345.e1–e345.e30.

Park, J. K., Park, S. H., So, K., Bae, I. H., Yoo, Y. D., & Um, H.-D. (2010). ICAM-3 enhances the migratory and invasive potential of human non-small cell lung cancer cells by inducing MMP-2 and MMP-9 via Akt and CREB. *International Journal of Oncology*, 36(1), 181–192.

Patni, S., Flynn, P., Wynen, L. P., Seager, A. L., Morgan, G., White, J. O., & Thornton, C. A. (2007). An introduction to Toll-like receptors and their possible role in the initiation of labour. *BJOG: An International Journal of Obstetrics and Gynaecology*, 114(11), 1326–1334.

Paul, M. R., Levitt, N. P., Moore, D. E., Watson, P. M., Wilson, R. C., Denlinger, C. E., ... Anderson, P. E. (2016). Multivariate models from RNA-Seq SNVs yield candidate molecular targets for biomarker discovery: SNV-DA. *BMC Genomics*, 17, 263.

Peltier, M. R. (2003). Immunology of term and preterm labor. *Reproductive Biology and Endocrinology: RB&E*, 1, 122.

Pereyra, S., Bertoni, B., & Sapiro, R. (2016). Interactions between environmental factors and maternal–fetal genetic variations: strategies to elucidate risks of preterm birth. *European Journal of Obstetrics, Gynecology, and Reproductive Biology*, 202, 20–25.

Pereyra, S., Velazquez, T., Bertoni, B., & Sapiro, R. (2012). Rapid multiplex high resolution melting method to analyze inflammatory related SNPs in preterm birth. *BMC Research Notes*, 5(1), 69.

Petrilli, V., Papin, S., & Tschopp, J. (2005). The inflammasome. *Current Biology: CB*, 15(15), R581.

- Petrovski, S., & Goldstein, D. B. (2016). Unequal representation of genetic variation across ancestry groups creates healthcare inequality in the application of precision medicine. *Genome Biology*, 17(1), 157.
- Piskol, R., Ramaswami, G., & Li, J. B. (2013). Reliable identification of genomic variants from RNA-seq data. *American Journal of Human Genetics*, 93(4), 641–651.
- Plazyo, O., Romero, R., Unkel, R., Balancio, A., Mial, T. N., Xu, Y., ... Gomez-Lopez, N. (2016). HMGB1 Induces an Inflammatory Response in the Chorioamniotic Membranes That Is Partially Mediated by the Inflammasome. *Biology of Reproduction*, 95(6), 130.
- Plunkett, J., Doniger, S., Orabona, G., Morgan, T., Haataja, R., Hallman, M., ... Muglia, L. (2011). An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genetics*, 7(4), e1001365.
- Plunkett, J., & Muglia, L. J. (2008). Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. *Annals of Medicine*, 40(3), 167–195.
- Porter, T. F., Fraser, A. M., Hunter, C. Y., Ward, R. H., & Varner, M. W. (1997). The risk of preterm birth across generations. *Obstetrics and Gynecology*, 90(1), 63–67.
- Presicce, P., Park, C.-W., SenthamaraiKannan, P., Bhattacharyya, S., Jackson, C., Kong, F., ... Kallapur, S. G. (2018). IL-1 signaling mediates intrauterine inflammation and chorio-decidua neutrophil recruitment and activation. *JCI Insight*, 3(6). <https://doi.org/10.1172/jci.insight.98306>
- Price, A. L., Helgason, A., Thorleifsson, G., McCarroll, S. A., Kong, A., & Stefansson, K. (2011). Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genetics*, 7(2), e1001317.
- Purcell, S., Neale, B., ToddBrown, K., Thomas, L., Ferreira, M., Bender, D., ... Daly, M. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559–575.
- Rahmioglu, N., Drong, A. W., Lockstone, H., Tapmeier, T., Hellner, K., Saare, M., ... Zondervan, K. T. (2017). Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues. *Epigenetics*, 12(10), 897–908.

- Ramaswami, G., & Li, J. B. (2014). RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Research*, 42(Database issue), D109–D113.
- Ravetch, J. V., & Lanier, L. L. (2000). Immune inhibitory receptors. *Science*, 290(5489), 84–89.
- Rey, G., Pereyra, S., Velazquez, T., Grasso, D., Alonso, J., Bertoni, B., & Sapiro, R. (2012). The Effect of Inflammation on Preterm Birth. In *Preterm Birth - Mother and Child*. Rijeka, Croatia: Intech.
- Rey, G., Skowronek, F., Alciaturi, J., Alonso, J., Bertoni, B., & Sapiro, R. (2008). Toll receptor 4 Asp299Gly polymorphism and its association with preterm birth and premature rupture of membranes in a South American population. *Molecular Human Reproduction*, 14(9), 555–559.
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139–140.
- Romero, R., Chaiworapongsa, T., & Espinoza, J. (2003). Micronutrients and Intrauterine Infection, Preterm Birth and the Fetal Inflammatory Response Syndrome. *The Journal of Nutrition*, 133 (5), 1668S – 1673S.
- Romero, R., Dey, S. K., & Fisher, S. J. (2014). Preterm labor: one syndrome, many causes. *Science*, 345(6198), 760–765.
- Romero, R., Espinoza, J., Gonçalves, L. F., Kusanovic, J. P., Friel, L. A., & Hassan, S. (2007). The role of inflammation and infection in preterm birth. *Seminars in Reproductive Medicine*, 25(1), 21–39.
- Romero, R., Espinoza, J., Kusanovic, J. P., Gotsch, F., Hassan, S., Erez, O., ... Mazar, M. (2006). The preterm parturition syndrome. *BJOG: An International Journal of Obstetrics and Gynaecology*, 113 Suppl, 17–42.
- Romero, R., Grivel, J.-C., Tarca, A. L., Chaemsathong, P., Xu, Z., Fitzgerald, W., ... Margolis, L. (2015). Evidence of perturbations of the cytokine network in preterm labor. *American Journal of Obstetrics and Gynecology*, 213(6), 836.e1–e836.e18.
- Romero, R., & Tartakovsky, B. (1992). The natural interleukin-1 receptor antagonist prevents interleukin-1-induced preterm delivery in mice. *American Journal of Obstetrics and Gynecology*, 167(4 Pt 1), 1041–1045.
- Romero, R., Velez Edwards, D. R., Kusanovic, J. P., Hassan, S. S., Mazaki-Tovi, S., Vaisbuch, E., ... Menon, R. (2010a). Identification of fetal and maternal single

nucleotide polymorphisms in candidate genes that predispose to spontaneous preterm labor with intact membranes. *American Journal of Obstetrics and Gynecology*, 202(5), 431.e1–e34.

Romero, R., Mazaki-Tovi, S., Vaisbuch, E., Kusanovic, J. P., Chaiworapongsa, T., Gomez, R., ... Beecher, C. (2010b). Metabolomics in premature labor: a novel approach to identify patients at risk for preterm delivery. *The Journal of Maternal-Fetal & Neonatal Medicine*, 23(12), 1344–1359.

Romero, R., Xu, Y., Plazyo, O., Chaemsaitong, P., Chaiworapongsa, T., Unkel, R., ... Gomez-Lopez, N. (2018). A Role for the Inflammasome in Spontaneous Labor at Term. *American Journal of Reproductive Immunology*, 79(6), e12440.

Ryckman, K. K., Morken, N.-H., White, M. J., Velez Edwards, D. R., Menon, R., Fortunato, S. J., ... Jacobsson, B. (2010). Maternal and fetal genetic associations of PTGER3 and PON1 with preterm birth. *PloS One*, 5(2), e9040.

Saigal, S., & Doyle, L. W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, 371(9608), 261–269.

Salaets, T., Richter, J., Brady, P., Jimenez, J., Nagatomo, T., Deprest, J., & Toelen, J. (2015). Transcriptome Analysis of the Preterm Rabbit Lung after Seven Days of Hyperoxic Exposure. *PloS One*, 10(8), e0136569.

Sanders, A. P., Burris, H. H., Just, A. C., Motta, V., Svensson, K., Mercado-Garcia, A., ... Baccarelli, A. A. (2015). microRNA expression in the cervix during pregnancy is associated with length of gestation. *Epigenetics*, 10(3), 221–8.

Sanders, Y. Y., Ambalavanan, N., Halloran, B., Zhang, X., Liu, H., Crossman, D. K., ... Hagood, J. S. (2012). Altered DNA methylation profile in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 186(6), 525–535.

Sans, M., Salzano, F. M., & Chakraborty, R. (1997). Historical genetics in Uruguay: estimates of biological origins and their problems. *Human Biology*, 69(2), 161–170.

Schaefer, T. M., Desouza, K., Fahey, J. V., Beagley, K. W., & Wira, C. R. (2004). Toll-like receptor (TLR) expression and TLR-mediated cytokine/chemokine production by human uterine epithelial cells. *Immunology*, 112(3), 428–436.

Schwindinger, W. F., & Robishaw, J. D. (2001). Heterotrimeric G-protein betagamma-dimers in growth and differentiation. *Oncogene*, 20(13), 1653–1660.

- Shah, N. R., & Bracken, M. B. (2000). A systematic review and meta-analysis of prospective studies on the association between maternal cigarette smoking and preterm delivery. *American Journal of Obstetrics and Gynecology*, 182(2), 465–472.
- Shao, B.-Z., Xu, Z.-Q., Han, B.-Z., Su, D.-F., & Liu, C. (2015). NLRP3 inflammasome and its inhibitors: a review. *Frontiers in Pharmacology*, 6, 262.
- Sheikh, I. A., Ahmad, E., Jamal, M. S., Rehan, M., Assidi, M., Tayubi, I. A., ... Al-Qahtani, M. (2016). Spontaneous preterm birth and single nucleotide gene polymorphisms: a recent update. *BMC Genomics*, 17(Suppl 9), 759.
- Sheng, Q., Zhao, S., Li, C.-I., Shyr, Y., & Guo, Y. (2016). Practicability of detecting somatic point mutation from RNA high throughput sequencing data. *Genomics*, 107(5), 163–169.
- Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *American Journal of Human Genetics*, 99(1), 139–153.
- Simini, F. (1999). Perinatal information system (SIP): a clinical database in Latin America and the Caribbean. *The Lancet*, 354(9172), 75.
- Sirota, M., Thomas, C. G., Liu, R., Zuhl, M., Banerjee, P., Wong, R. J., ... March of Dimes Prematurity Research Centers. (2018). Enabling precision medicine in neonatology, an integrated repository for preterm birth research. *Scientific Data*, 5, 180219.
- Sirugo, G., Williams, S. M., & Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell*, 177(1), 26–31.
- Skeeles, L. E., Fleming, J. L., Mahler, K. L., & Toland, A. E. (2013). The impact of 3'UTR variants on differential expression of candidate cancer susceptibility genes. *PloS One*, 8(3), e58609.
- Söber, S., Reiman, M., Kikas, T., Rull, K., Inno, R., Vaas, P., ... Laan, M. (2015). Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes. *Scientific Reports*, 5, 13336.
- Stephen, G. L., Lui, S., Hamilton, S. A., Tower, C. L., Harris, L. K., Stevens, A., & Jones, R. L. (2015). Transcriptomic profiling of human choriondecidua during term labor: inflammation as a key driver of labor. *American Journal of Reproductive Immunology*, 73(1), 36–55.

- Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *American Journal of Human Genetics*, 76(3), 449–462.
- Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68(4), 978–989.
- Strauss, J. F., 3rd, Romero, R., Gomez-Lopez, N., Haymond-Thornburg, H., Modi, B. P., Teves, M. E., ... Schenkein, H. A. (2018). Spontaneous preterm birth: advances toward the discovery of genetic predisposition. *American Journal of Obstetrics and Gynecology*, 218(3), 294–314.e2.
- Sun, G. D., Kobayashi, T., Abe, M., Tada, N., Adachi, H., Shiota, A., ... Hino, O. (2007). The endoplasmic reticulum stress-inducible protein Niban regulates eIF2 α and S6K1/4E-BP1 phosphorylation. *Biochemical and Biophysical Research Communications*, 360(1), 181–187.
- Sun, Z., Bhagwate, A., Prodduturi, N., Yang, P., & Kocher, J.-P. A. (2017). Indel detection from RNA-seq data: tool evaluation and strategies for accurate detection of actionable mutations. *Briefings in Bioinformatics*, 18(6), 973–983.
- Svensson, A. C., Sandin, S., Cnattingius, S., Reilly, M., Pawitan, Y., Hultman, C. M., & Lichtenstein, P. (2009). Maternal effects for preterm birth: a genetic epidemiologic study of 630,000 families. *American Journal of Epidemiology*, 170(11), 1365–1372.
- Tang, T., Gong, T., Jiang, W., & Zhou, R. (2018). GPCRs in NLRP3 Inflammasome Activation, Regulation, and Therapeutics. *Trends in Pharmacological Sciences*, 39(9), 798–811.
- Tarca, A. L., Romero, R., Xu, Z., Gomez-Lopez, N., Erez, O., Hsu, C.-D., ... Carey, V. J. (2019). Targeted expression profiling by RNA-Seq improves detection of cellular dynamics during pregnancy and identifies a role for T cells in term parturition. *Scientific Reports*, 9(1), 848.
- Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, 25(9), 1105–1111.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3), 562–578.

- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., ... Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5), 511–515.
- Treloar, S. A., Macones, G. A., Mitchell, L. E., & Martin, N. G. (2000). Genetic influences on premature parturition in an Australian twin sample. *Twin Research*, 3(2), 80–82.
- Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., ... Pontén, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, 347(6220), 1260419.
- Uusküla, L., Männik, J., Rull, K., Minajeva, A., Kõks, S., Vaas, P., ... Laan, M. (2012). Mid-gestational gene expression profile in placenta and link to pregnancy complications. *PloS One*, 7(11), e49248.
- Uzun, A., Laliberte, A., Parker, J., Andrew, C., Winterrowd, E., Sharma, S., ... Padbury, J. F. (2012). dbPTB: a database for preterm birth. *Database (Oxford)*, 2012, bar069.
- Vaeth, M., & Feske, S. (2018). Ion channelopathies of the immune system. *Current Opinion in Immunology*, 52, 39–50.
- Van der Auwera, G. (2014, April 16). Best Practices for Variant Discovery in RNAseq. Accedido 3 de mayo de 2019, en <https://gatkforums.broadinstitute.org/gatk/discussion/4067/best-practices-for-variant-discovery-in-rnaseq>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 11(1110): 11.10.1–11.10.33.
- van de Veerdonk, F. L., Netea, M. G., Dinarello, C. A., & Joosten, L. A. B. (2011). Inflammasome activation and IL-1 β and IL-18 processing during infection. *Trends in Immunology*, 32(3), 110–116.
- Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols*, 11(1), 1–9.
- Velez Edwards, D. R., Fortunato, S. J., Thorsen, P., Lombardi, S. J., Williams, S. M., & Menon, R. (2008). Preterm birth in Caucasians is associated with coagulation and inflammation pathway gene variants. *PloS One*, 3(9), e3283.

- Velez Edwards, D. R., Fortunato, S. J., Thorsen, P., Lombardi, S. J., Williams, S. M., & Menon, R. (2009). Spontaneous preterm birth in African Americans is associated with infection and inflammatory response gene variants. *American Journal of Obstetrics and Gynecology*, 200(2), 209.e1–e27.
- Villar, J., Papageorghiou, A. T., Knight, H. E., Gravett, M. G., Iams, J., Waller, S. A., ... Goldenberg, R. L. (2012). The preterm birth syndrome: a prototype phenotypic classification. *American Journal of Obstetrics and Gynecology*, 206(2), 119–123.
- Vitoratos, N., Papadias, K., Makrakis, E., Christodoulakos, G., Panoulis, K., & Creatsas, G. (2006). Association between serum tumor necrosis factor-alpha and corticotropin-releasing hormone levels in women with preterm labor. *The Journal of Obstetrics and Gynaecology Research*, 32(5), 497–501.
- Vogel, J. P., Chawanpaiboon, S., Moller, A.-B., Watananirun, K., Bonet, M., & Lumbiganon, P. (2018). The global epidemiology of preterm birth. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, 52, 3–12.
- Vora, B., Wang, A., Kostı, I., Huang, H., Paranjpe, I., Woodruff, T. J., ... Sirota, M. (2018). Meta-Analysis of Maternal and Fetal Transcriptomic Data Elucidates the Role of Adaptive and Innate Immunity in Preterm Birth. *Frontiers in Immunology*, 9, 993.
- Vrachnis, N., Vitoratos, N., Iliodromiti, Z., Sifakis, S., Deligeoroglou, E., & Creatsas, G. (2010). Intrauterine inflammation and preterm delivery. *Annals of the New York Academy of Sciences*, 1205, 118–122.
- Wang, X., Zuckerman, B., Pearson, C., Kaufman, G., Chen, C., Wang, G., ... Xu, X. (2002a). Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. *JAMA: The Journal of the American Medical Association*, 287(2), 195–202.
- Wang, H.-W., Sharp, T. V., Koumi, A., Koentges, G., & Boshoff, C. (2002b). Characterization of an anti-apoptotic glycoprotein encoded by Kaposi's sarcoma-associated herpesvirus which resembles a spliced variant of human survivin. *The EMBO Journal*, 21(11), 2602–2615.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, 10(1), 57–63.
- Watari, M., Watari, H., DiSanto, M. E., Chacko, S., Shi, G. P., & Strauss, J. F., 3rd. (1999). Pro-inflammatory cytokines induce expression of matrix-metabolizing

enzymes in human cervical smooth muscle cells. *The American Journal of Pathology*, 154(6), 1755–1762.

Watari, M., Watari, H., & Strauss, J. (2000). Lipopolysaccharide and pro-inflammatory cytokines induce expression of matrix metabolizing enzymes in human cervical smooth muscle cells-implication in the mechanism of cervical ripening. *International Journal of Gynecology & Obstetrics*, 70, B56–B56.

Wen, S. W., Smith, G., Yang, Q., & Walker, M. (2004). Epidemiology of preterm birth and neonatal outcome. *Seminars in Fetal & Neonatal Medicine*, 9(6), 429–435.

Wenstrom, K. D., Andrews, W. W., Hauth, J. C., Goldenberg, R. L., DuBard, M. B., & Cliver, S. P. (1998). Elevated second-trimester amniotic fluid interleukin-6 levels predict preterm delivery. *American Journal of Obstetrics and Gynecology*, 178(3), 546–550.

Westra, H.-J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., ... Franke, L. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics*, 45(10), 1238–1243.

Wilkerson, M. D., Cabanski, C. R., Sun, W., Hoadley, K. A., Walter, V., Mose, L. E., ... Hayes, D. N. (2014). Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Research*, 42(13), e107.

Willcockson, A. R., Nandu, T., Liu, C.-L., Nallasamy, S., Kraus, W. L., & Mahendroo, M. (2018). Transcriptome signature identifies distinct cervical pathways induced in lipopolysaccharide-mediated preterm birth. *Biology of Reproduction*, 98(3), 408–421.

Winkvist, A., Mogren, I., & Högberg, U. (1998). Familial patterns in birth characteristics: impact on individual and population risks. *International Journal of Epidemiology*, 27(2), 248–254.

Winn, V. D., Haimov-Kochman, R., Paquet, A. C., Jean Yang, Y., Madhusudhan, M. S., Gormley, M., ... Fisher, S. J. (2007). Gene Expression Profiling of the Human Maternal-Fetal Interface Reveals Dramatic Changes between Midgestation and Term. *Endocrinology*, 148(3), 1059–1079.

Witkin, S. S. (2015). The vaginal microbiome, vaginal anti-microbial defence mechanisms and the clinical challenge of reducing infection-related preterm birth. *BJOG: An International Journal of Obstetrics and Gynaecology*, 122(2), 213–218.

- Wray, S., Burdyga, T., Noble, D., Noble, K., Borysova, L., & Arrowsmith, S. (2015). Progress in understanding electro-mechanical signalling in the myometrium. *Acta Physiologica*, 213(2), 417–431.
- Wu, W., Clark, E. A. S., Stoddard, G. J., Watkins, W. S., Esplin, M. S., Manuck, T. A., ... Jorde, L. B. (2013). Effect of interleukin-6 polymorphism on risk of preterm birth within population strata: a meta-analysis. *BMC Genetics*, 14(1), 30.
- Xiong, H. Y., Alipanahi, B., Lee, L. J., Bretschneider, H., Merico, D., Yuen, R. K. C., ... Frey, B. J. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science*, 347(6218), 1254806.
- Yadav, D., & Sarvetnick, N. (2003). Cytokines and autoimmunity: redundancy defines their complex nature. *Current Opinion in Immunology*, 15(6), 697–703.
- Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., ... Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, 44(D1), D710–D716.
- Yoon, B. H., Oh, S. Y., Romero, R., Shim, S. S., Han, S. Y., Park, J. S., & Jun, J. K. (2001). An elevated amniotic fluid matrix metalloproteinase-8 level at the time of mid-trimester genetic amniocentesis is a risk factor for spontaneous preterm delivery. *American Journal of Obstetrics and Gynecology*, 185(5), 1162–1167.
- York, T. P., Eaves, L. J., Lichtenstein, P., Neale, M. C., Svensson, A., Latendresse, S., ... Strauss, J. F., 3rd. (2013). Fetal and maternal genes' influence on gestational age in a quantitative genetic analysis of 244,000 Swedish births. *American Journal of Epidemiology*, 178(4), 543–550.
- York, T. P., Eaves, L. J., Neale, M. C., & Strauss, J. F., 3rd. (2014). The contribution of genetic and environmental factors to the duration of pregnancy. *American Journal of Obstetrics and Gynecology*, 210(5), 398–405.
- York, T. P., Strauss, J. F., 3rd, Neale, M. C., & Eaves, L. J. (2009). Estimating fetal and maternal genetic contributions to premature birth from multiparous pregnancy histories of twins using MCMC and maximum-likelihood approaches. *Twin Research and Human Genetics*, 12(4), 333–342.
- You, L., Chen, L., Pan, L., Gu, W.-S., & Chen, J.-Y. (2015). Zinc finger protein 467 regulates Wnt signaling by modulating the expression of sclerostin in adipose derived stem cells. *Biochemical and Biophysical Research Communications*, 456(2), 598–604.

- Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, 11(2), R14.
- Yu, L. J., Wang, B., Parobchak, N., Roche, N., & Rosen, T. (2015). STAT3 cooperates with the non-canonical NF- κ B signaling to regulate pro-labor genes in the human placenta. *Placenta*, 36(5), 581–586.
- Yuan, T. L., Amzallag, A., Bagni, R., Yi, M., Afghani, S., Burgan, W., ... McCormick, F. (2018). Differential Effector Engagement by Oncogenic KRAS. *Cell Reports*, 22(7), 1889–1902.
- Zadora, J., Singh, M., Herse, F., Przybyl, L., Haase, N., Golic, M., ... Izsvák, Z. (2017). Disturbed Placental Imprinting in Preeclampsia Leads to Altered Expression of DLX5, a Human-Specific Early Trophoblast Marker. *Circulation*, 136(19), 1824–1839.
- Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L. M., Juodakis, J., ... Muglia, L. J. (2017). Genetic Associations with Gestational Duration and Spontaneous Preterm Birth. *The New England Journal of Medicine*, 377(12), 1156–1167.
- Zhang, G., Srivastava, A., Bacelis, J., Juodakis, J., Jacobsson, B., & Muglia, L. J. (2018). Genetic studies of gestational duration and preterm birth. *Best Practice & Research. Clinical Obstetrics & Gynaecology*, 52, 33–47.
- Zhu, Y. X., Benn, S., Li, Z. H., Wei, E., Masih-Khan, E., Trieu, Y., ... Keith Stewart, A. (2004). The SH3–SAM Adaptor HACS1 is Up-regulated in B Cell Activation Signaling Cascades. *The Journal of Experimental Medicine*, 200(6), 737–747.
- Zmijanac Partl, J., Karin, V., Skrtic, A., Nikuseva-Martic, T., Serman, A., Mlinarec, J., ... Serman, L. (2018). Negative regulators of Wnt signaling pathway SFRP1 and SFRP3 expression in preterm and term pathologic placentas. *The Journal of Maternal-Fetal & Neonatal Medicine*, 31(22), 2971–2979.

8

APÉNDICES

8.1 CUESTIONARIO DEL SISTEMA INFORMÁTICO PERINATAL

Anverso

HISTORIA CLINICA PERINATAL-CLAP-OPS/OMS

NOMBRE: _____ DOMICILIO: _____ LOCALIDAD: _____ TELEFONO: _____

FECHA DE NACIMIENTO: día mes año _____

RAZA: blanca ☐ indígena ☐ mestiza ☐ negra ☐ otra ☐

ALFA BETA: ☐ ninguno ☐ primaria ☐ secundaria ☐ univers. ☐ años en el mayor nivel _____

ESTUDIOS: ☐ ninguno ☐ primaria ☐ secundaria ☐ univers. ☐ años en el mayor nivel _____

ESTADO CIVIL: ☐ casado ☐ unión estable ☐ soltero ☐ otro _____

CONTROL PRENATAL EN: _____ PARTO EN: _____

NUMERO IDENTID.: _____

ANTECEDENTES

FAMILIARES: ☐ no ☐ si ☐ TBC ☐ diabetes ☐ hipertensión ☐ preeclampsia ☐ eclampsia ☐ otros _____

PERSONALES: ☐ no ☐ si ☐ cirugía tracto reprod. ☐ infertilidad ☐ VIH + ☐ cardiop. nefrop. ☐ cond. médica grave _____

OBSTETRICOS: gestas previas _____ abortos _____ vaginales _____ nacidos vivos _____ viven _____

ULTIMO PREVIO: <2500g _____ >4500g _____ preeclampsia-eclampsia _____

3 espont. consecutivos _____

partos _____ cesáreas _____

nacidos muertos _____ muertos 1 sem. _____ después 1 sem. _____

FIN EMBARAZO ANTERIOR: día mes año _____ menos de 6 meses más de 6 meses _____

EMBARAZO PLANEADO: si ☐ no ☐

FRACASO METODO ANTICONCEP.: ☐ no usaba ☐ barrera ☐ DIU ☐ s-lora ☐ depo ☐ natural _____

GESTACION ACTUAL

PESO ANTERIOR: _____ Kg. TALLA (cm): 1 _____

FUM: ☐ no ☐ si _____

EG CONFIABLE por FUM: ☐ no ☐ si _____

MOVIM. FETALES desde: _____

CIGARRILLOS POR DIA: ☐ no ☐ si _____

ALCOHOL: ☐ no ☐ si _____

ANTITETANICA: ☐ no ☐ si _____

ANTIRUBEOLA: ☐ no ☐ si _____

EX. NORMAL: ☐ no ☐ si _____

BACTERIURIA: ☐ no ☐ si _____

GRUPO Rh: ☐ no ☐ si _____

PAPANICOLAOU: ☐ no ☐ si _____

VIH solicitado: ☐ no ☐ si _____

VDRU/RPR <20 sem: ☐ no ☐ si _____

SIFILIS confirmada por FTA: ☐ no ☐ si _____

VDRU/RPR >20 sem: ☐ no ☐ si _____

Hb <20 sem: ☐ no ☐ si _____

Fe/FOLATOS indicados: ☐ no ☐ si _____

Hb >20 sem: ☐ no ☐ si _____

ESTREPTOCOCCO B 35-37 semanas: ☐ no ☐ si _____

PELVIANA intento versión externa: ☐ no ☐ si _____

TRABAJO DE PARTO

FECHA DE INGRESO: día mes año _____

CARNE: ☐ no ☐ si _____

CONSULTAS PRENATALES total: _____

HOSPITALIZ. en EMBARAZO días: _____

CORTICOIDES ANTENATALES CICLO UNICO: ☐ completo ☐ múltiples ☐ incompl. ☐ ninguna _____

INICIO: ☐ espontáneo ☐ inducido ☐ cesar. elect. _____

RUPTURA MEMBRANAS ANTEPARTO: ☐ no ☐ si _____

EG sem.: ☐ no ☐ si _____

Tiempo hasta el parto: _____

EDAD GEST. al parto: _____

PRESENTACION: ☐ cefálica ☐ pelviana ☐ transversa _____

COMPANIA: ☐ no ☐ si _____

POSICION: _____

Contr. / 10': _____

DILATACION: _____

ALTURA: _____

FCF / DIPS: _____

NOTAS: _____

ENFERMEDADES

ninguna ☐ no ☐ si ☐ corioamionitis ☐ HTA previa ☐ HTA inducida ☐ infección urinaria ☐ amenaza parto preter. ☐ preeclampsia ☐ eclampsia ☐ R.C.I.U. ☐ rotura prem. de membranas ☐ infección puerperal ☐ diabetes ☐ cardiop. nefrop. ☐ otra cond. grave _____

NACIMIENTO

VIVO: ☐ no ☐ si _____

MUERTO: ☐ parto ☐ ignora momento _____

ACOMPANANTE: ☐ pareja ☐ otro ☐ familiar ☐ ninguno _____

POSICION PARTO: ☐ sentada ☐ cuclillas ☐ acostada _____

EPISIOTOMIA: ☐ no ☐ si _____

DESGARRO Grado: _____

OCITOCICOS EN ALUMBRAMIENTO: ☐ ocitocina ☐ otro ☐ ninguno _____

PLACENTA COMPLETA: ☐ si ☐ no _____

LIGADURA CORDON: ☐ precoz (<30s) ☐ tardía (>30s) _____

INDICACION PRINCIPAL DE INDUCCION O PARTO OPERATORIO: _____

RECIEENACIDO

SEXO: ☐ f ☐ m _____

PESO AL NACER: _____ g

PER. CEFALICO: _____ cm

LONGITUD: _____ cm

EG CONFIABLE sem.: _____

PESO E.G.: _____

APGAR 1' min.: _____

APGAR 5' min.: _____

REANIMACION: ☐ no ☐ si _____

FALLECE en SALA de PARTO: ☐ no ☐ si _____

REFERIDO: ☐ otro ☐ alq. conj. ☐ hosp. _____

ATENDIO: ☐ médico ☐ enf./obst. ☐ auxil. ☐ estud./empir. ☐ otro _____

PARTO: ☐ no ☐ si _____

NEONATO: ☐ no ☐ si _____

PUERPERIO

hora: _____ min: _____

T°C: _____

pulso: _____

P A: _____

invol.: _____

loquios: _____

EGRESO RN

vivo: ☐ no ☐ si _____

EDAD días completos: _____

LACTANCIA exclusiva: ☐ no ☐ si _____

PESO AL EGRESO (g): _____

EGRESO MATERNO

viva: ☐ no ☐ si _____

ANTIRUBEOLA POST PARTO: ☐ no ☐ si _____

ANTICONCEPCION: ☐ barrera ☐ DIU ☐ hormonal ☐ natural _____

NUMERO IDENTID.: _____

Nombre Recién Nacido: _____

Responsable: _____

Responsable: _____

8.2 TABLAS SUPLEMENTARIAS

8.2.1. TABLA SUPLEMENTARIA 1

Esta tabla acompaña el artículo de Pereyra et al. (2019) relacionado al capítulo 4 de esta tesis. Se encuentra en <https://doi.org/10.6084/m9.figshare.7936187.v1>. Tiene la siguiente información:

- Genes diferencialmente expresados ($FDR < 0.05$ y tasa de cambio logarítmica absoluta > 2) detectados en la comparación entre parto pretérmino severo vs. parto a término.
- Subgrupo de genes DE que también se encontraron diferencialmente expresados entre trabajo de parto vs. no trabajo de parto en la literatura.
- Lista completa de genes diferencialmente expresados ($FDR < 0.05$ y tasa de cambio logarítmica absoluta > 2) detectados en la comparación RPM vs. no RPM.
- Lista de genes diferencialmente expresados previamente identificados como asociados con PPT en la literatura.

8.2.2. TABLA SUPLEMENTARIA 2

Variantes novedosas reportadas en esta tesis.

Variantes	Alelo refe- rencia	Alelo alter- nativo	Ubicación genómica	Simbolo génico	Función exónica	Cambio amino- ácido	SIFT_ pred
1:184761123_T/C	T	C	UTR3	<i>FAM129A</i>	.	.	.
1:184761129_T/C	T	C	UTR3	<i>FAM129A</i>	.	.	.
1:184761350_T/C	T	C	UTR3	<i>FAM129A</i>	.	.	.
1:184762676_T/C	T	C	UTR3	<i>FAM129A</i>	.	.	.
1:198711178_A/G	A	G	Intrón	<i>PTPRC</i>	.	.	.
1:7980539_T/C	T	C	UTR3	<i>TNFRSF9</i>	.	.	.
2:100897332_A/G	A	G	UTR3	<i>LONRF2</i>	.	.	.
2:113589330_G/A	G	A	Intrón	<i>IL1β</i>	.	.	.
2:113593620_T/A	T	A	Intrón	<i>IL1β</i>	.	.	.
2:113593621_T/A	T	A	Intrón	<i>IL1β</i>	.	.	.
2:113891592_T/A	T	A	UTR3	<i>IL1RN</i>	.	.	.
3:46417220_A/G	A	G	ncRNA_I ntrón	LOC10272 4297	.	.	.
3:46417221_A/G	A	G	ncRNA_I ntrón	LOC10272 4297	.	.	.
3:46417272_A/G	A	G	ncRNA_I ntrón	LOC10272 4297	.	.	.
3:71827179_T/C	T	C	Intrón	<i>PROK2</i>	.	.	.
4:185702134_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702175_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702209_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702240_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702266_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702323_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702333_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702830_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702831_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702901_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702902_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.

Tabla suplementaria 2 (Continuación)

Variantes	Alelo refe- rencia	Alelo alter- nativo	Ubicación genómica	Simbolo génico	Función exónica	Cambio amino- ácido	SIFT_ pred
4:185702928_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702931_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702960_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702961_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702968_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702983_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185702984_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185703037_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185703040_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185703046_T/C	T	C	Intrón	<i>ACSL1</i>	.	.	.
4:185703498_G/A	G	A	Intrón	<i>ACSL1</i>	.	.	.
4:84215324_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215333_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215391_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215479_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215718_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215719_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215963_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215973_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84215990_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216029_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216037_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216048_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216065_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216067_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216135_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216169_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
4:84216199_T/C	T	C	UTR3	<i>HPSE</i>	.	.	.
5:169701559_T/C	T	C	Intrón	<i>LCP2</i>	.	.	.
6:29695991_G/T	G	T	ncRNA_I ntrón	<i>HLA-F-AS1</i>	.	.	.
7:105897806_T/C	T	C	Intrón	<i>NAMPT</i>	.	.	.
7:105897818_T/C	T	C	Intrón	<i>NAMPT</i>	.	.	.

Tabla suplementaria 2 (Continuación)

Variantes	Alelo refe- rencia	Alelo alter- nativo	Ubicación genómica	Simbolo génico	Función exónica	Cambio amino- ácido	SIFT_ pred
7:149462528_A/G	A	G	Exón	<i>ZNF467</i>	nonsynon ymous_S NV	ZNF467: NM_207 336:exon 5:c.T106 3C:p.F35 5L	D
8:133722263_A/G	A	G	UTR3	<i>TMEM71</i>	.	.	.
8:22546823_T/C	T	C	UTR3	<i>EGR3</i>	.	.	.
9:117552560_G/A	G	A	UTR3	<i>TNFSF15</i>	.	.	.
10:70859296_A/G	A	G	Intrón	<i>SRGN</i>	.	.	.
10:70863052_A/G	A	G	Intrón	<i>SRGN</i>	.	.	.
10:98469207_T/C	T	C	Intrón	<i>PIK3AP1</i>	.	.	.
10:98469297_T/C	T	C	Intrón	<i>PIK3AP1</i>	.	.	.
10:98469298_T/C	T	C	Intrón	<i>PIK3AP1</i>	.	.	.
11:102208583_C/T	C	T	UTR3	<i>BIRC3</i>	.	.	.
12:108984227_T/C	T	C	UTR3	<i>TMEM119</i>	.	.	.
12:108984979_T/C	T	C	UTR3	<i>TMEM119</i>	.	.	.
12:25205767_A/G	A	G	UTR5	<i>LRMP</i>	.	.	.
12:25205769_A/G	A	G	UTR5	<i>LRMP</i>	.	.	.
14:88478084_G/A	G	A	Exón	<i>GPR65</i>	nonsynon ymous_S NV	GPR65: NM_003 608:exon 2:c.G893 A:p.R29 8K	T
15:58468556_A/G	A	G	Intrón	<i>AQP9</i>	.	.	.
15:58468631_A/G	A	G	Intrón	<i>AQP9</i>	.	.	.
15:58468642_A/G	A	G	Intrón	<i>AQP9</i>	.	.	.
15:58468990_A/G	A	G	Intrón	<i>AQP9</i>	.	.	.
15:58469003_T/C	T	C	Intrón	<i>AQP9</i>	.	.	.
15:80255054_T/C	T	C	Intrón	<i>BCL2A1</i>	.	.	.
16:21656608_T/C	T	C	Intrón	<i>IGSF6</i> <i>METTL9</i>	.	.	.
16:3292879_A/G	A	G	UTR3	<i>MEFV</i>	.	.	.

Tabla suplementaria 2 (Continuación)

Variantes	Alelo refe- rencia	Alelo alter- nativo	Ubicación genómica	Simbolo génico	Función exónica	Cambio amino- ácido	SIFT_ pred
17:1476222_T/C	T	C	UTR3	<i>SLC43A2</i>	.	.	.
17:1476223_T/C	T	C	UTR3	<i>SLC43A2</i>	.	.	.
17:1476237_T/C	T	C	UTR3	<i>SLC43A2</i>	.	.	.
17:1476246_T/C	T	C	UTR3	<i>SLC43A2</i>	.	.	.
17:1488395_T/C	T	C	Intrón	<i>SLC43A2</i>	.	.	.
19:15771652_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:15771656_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:15771666_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:15771951_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:15771952_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:15771961_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:15772035_A/G	A	G	Río abajo	<i>CYP4F3</i>	.	.	.
19:4217209_C/T	C	T	Exón	<i>ANKRD24</i>	synonymo us_SNV	ANKRD 24:NM_ 133475:e xon18:c. C2052T: p.S684S	T
20:30679518_A/G	A	G	Intrón	<i>HCK</i>	.	.	.
20:30679732_A/G	A	G	Intrón	<i>HCK</i>	.	.	.
20:30679733_A/G	A	G	Intrón	<i>HCK</i>	.	.	.
20:30680822_A/G	A	G	Intrón	<i>HCK</i>	.	.	.
20:30682241_A/G	A	G	Intrón	<i>HCK</i>	.	.	.
22:39350102_A/G	A	G	Intergénic a	<i>CBX6</i> lx3bA <i>POBEC3A</i>	.	.	.
22:39350262_A/G	A	G	Intergénic a	<i>CBX6</i> lx3bA <i>POBEC3A</i>	.	.	.
22:39350321_A/G	A	G	Intergénic a	<i>CBX6</i> lx3bA <i>POBEC3A</i>	.	.	.
22:39350607_A/G	A	G	Intergénic a	<i>CBX6</i> lx3bA <i>POBEC3A</i>	.	.	.
22:39350614_A/G	A	G	Intergénic a	<i>CBX6</i> lx3bA <i>POBEC3A</i>	.	.	.

Tabla suplementaria 2 (Continuación)

Variantes	Alelo refe- rencia	Alelo alter- nativo	Ubicación genómica	Simbolo génico	Función exónica	Cambio amino- ácido	SIFT_ pred
22:39350645_A/G	A	G	Intergénic a	<i>CBX6</i> <i>POBEC3A</i>	.	.	.
22:39350646_A/G	A	G	Intergénic a	<i>CBX6</i> <i>POBEC3A</i>	.	.	.
X:30716643_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30716700_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30716701_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30716727_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30716885_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30716889_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30716928_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717563_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717576_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717594_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717599_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717611_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717621_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717728_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30717755_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30741779_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30741788_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30741958_A/G	A	G	Intrón	<i>GK</i>	.	.	.
X:30741959_A/G	A	G	Intrón	<i>GK</i>	.	.	.

8.2.3. TABLA SUPLEMENTARIA 3

Lista de variantes incluidas en la Figura 5.2 de frecuencias alélicas para 9 genes seleccionados.

Gen	ID dbSNP	Alelo referenci a	Alelo alternativ o	Frecuencia alelo de alternativo Controles	Frecuencia alelo de alternativo Prematuros	Ubicación génica	Función en exón	Unión con sitio blanco de micro ARN(s)	SIFT_ pred	Polyphen2_H VAR_pred
<i>CD300E</i>	rs1878061	C	T	0.25	0.375	Exón	SNV no sinónimo	.	T	B
<i>CD300E</i>	rs479782	C	T	0	0.375	UTR3	.	.	.	.
<i>CD300E</i>	rs554592	A	T	0.125	0.375	UTR3	.	.	.	.
<i>CD300E</i>	rs555542	T	C	0.5	0.625	UTR3	.	.	.	.
<i>CD300E</i>	rs578460	A	G	0.125	0.375	UTR3	.	.	.	.
<i>FAM129A</i>	1:184762676_T/C	T	C	0.5	0.625	UTR3	.	.	.	.
<i>FAM129A</i>	rs1139069	T	C	0.25	0.5	UTR3	.	Si	.	.
<i>FAM129A</i>	rs14023	C	T	0.25	0.375	UTR3	.	.	.	.
<i>FAM129A</i>	rs14426	A	G	0.25	0.5	UTR3	.	Si	.	.
<i>FAM129A</i>	rs3187911	T	C	0.125	0.125	UTR3	.	.	.	.
<i>FAM129A</i>	rs487675	T	C	0.125	0.375	Exón	SNV sinónimo	.	.	.
<i>FAM129A</i>	rs492126	G	A	0.25	0.375	UTR3	.	.	.	.
<i>FAM129A</i>	rs526024	T	A	0.25	0.375	UTR3	.	Si	.	.
<i>FAM129A</i>	rs581100	C	G	0.25	0.375	UTR3	.	Si	.	.
<i>FAM129A</i>	rs682331	A	G	0.25	0.5	UTR3	.	.	.	.

Tabla suplementaria 3 (Continuación)

Gen	ID dbSNP	Alelo referenci a	Alelo alternativ o	Frecuencia alelo de alternativo Controles	Frecuencia alelo de alternativo Prematuros	Ubicación génica	Función en exón	Unión con sitio blanco de micro ARN(s)	SIFT_ pred	Polyphen2_H VAR_pred
<i>FAM129A</i>	rs7360	C	T	0.25	0.5	UTR3	.	Si	.	.
<i>FAM129A</i>	rs940823216	T	C	0.25	0.5	UTR3	.	.	.	.
<i>GNG2</i>	rs1540703	G	C	0.25	0	UTR3	.	.	.	.
<i>GNG2</i>	rs2000006	G	C	0.25	0	UTR3	.	.	.	.
<i>GNG2</i>	rs2357306	A	T	0.25	0	UTR3	.	.	.	.
<i>GNG2</i>	rs3204006	G	C	0.125	0.125	UTR3	.	.	.	.
<i>GNG2</i>	rs3204008	C	G	0.125	0	UTR3	.	.	.	.
<i>GNG2</i>	rs3825596	G	A	0	0.125	UTR3	.	.	.	.
<i>GNG2</i>	rs45595035	C	T	0.25	0	UTR3	.	.	.	.
<i>GNG2</i>	rs8014838	A	T	0.75	0.875	UTR3	.	.	.	.
<i>ICAM3</i>	rs2230399	C	G	0	0.25	Exón	SNV no sinónimo	.	T	B
<i>ICAM3</i>	rs2304237	T	C	0	0.375	Exón	SNV no sinónimo	.	T	B
<i>ICAM3</i>	rs2304240	A	G	0.625	0.625	Exón	SNV sinónimo	.	.	.
<i>KRAS</i>	rs1137188	G	A	0.375	0.75	UTR3	.	.	.	.
<i>KRAS</i>	rs1137189	A	T	0.375	0.5	UTR3	.	.	.	.
<i>KRAS</i>	rs1137196	T	G	0.25	0.75	UTR3	.	Si	.	.

Tabla suplementaria 3 (Continuación)

Gen	ID dbSNP	Alelo referenci a	Alelo alternativ o	Frecuencia alelo de alternativo Controles	Frecuencia alelo de alternativo Prematuros	Ubicación génica	Función en exón	Unión con sitio blanco de micro ARN(s)	SIFT_ pred	Polyphen2_H VAR_pred
<i>KRAS</i>	rs12245	A	T	0.5	0.75	UTR3	.	.	.	.
<i>KRAS</i>	rs12587	T	G	0.375	0.75	UTR3	.	Si	.	.
<i>KRAS</i>	rs13096	T	C	0.125	0.75	UTR3	.	.	.	.
<i>KRAS</i>	rs140080026	A	G	0.125	0	UTR3	.	Si	.	.
<i>KRAS</i>	rs4285970	G	A	1	1	UTR3	.	.	.	.
<i>KRAS</i>	rs4597149	T	C	1	1	UTR3	.	Si	.	.
<i>KRAS</i>	rs4963858	T	C	1	1	UTR3	.	.	.	.
<i>KRAS</i>	rs61764370	A	C	0.125	0	UTR3	.	Si	.	.
<i>KRAS</i>	rs712	A	C	0.5	0.75	UTR3	.	Si	.	.
<i>KRAS</i>	rs7973623	G	A	0.375	0	UTR3	.	Si	.	.
<i>KRAS</i>	rs8720	T	C	0.25	0.75	UTR3	.	Si	.	.
<i>KRAS</i>	rs9266	A	G	0.375	0.75	UTR3	.	Si	.	.
<i>LRRC25</i>	rs116399833	C	A	0.625	0.125	UTR3	.	.	.	.
<i>LRRC25</i>	rs12461382	C	T	0.125	0.125	UTR3	.	.	.	.
<i>LRRC25</i>	rs181468783	C	T	0	0.25	UTR5	.	.	.	.
<i>LRRC25</i>	rs3848647	A	G	0.875	0.5	UTR3	.	.	.	.
<i>LRRC25</i>	rs3848648	C	T	0.875	0.5	UTR3	.	.	.	.
<i>LRRC25</i>	rs4328570	C	T	0.875	0.5	UTR3	.	.	.	.
<i>LRRC25</i>	rs6512262	G	C	0.875	0.5	UTR3	.	.	.	.

Tabla suplementaria 3 (Continuación)

Gen	ID dbSNP	Alelo referenci a	Alelo alternativ o	Frecuencia alelo de alternativo Controles	Frecuencia alelo de alternativo Prematuros	Ubicación génica	Función en exón	Unión con sitio blanco de micro ARN(s)	SIFT_ pred	Polyphen2_H VAR_pred
<i>LRRC25</i>	rs6512263	A	C	0.875	0.5	UTR3	.	.	.	.
<i>LRRC25</i>	rs6512264	C	A	0.875	0.5	UTR3	.	.	.	.
<i>LRRC25</i>	rs6512265	G	A	0.875	0.5	Exón	SNV no sinónimo	.	T	B
<i>MARCH1</i>	rs1054681	T	C	1	0.625	UTR3	.	.	.	.
<i>MARCH1</i>	rs1133505	C	T	0.875	0.25	UTR3	.	.	.	.
<i>MARCH1</i>	rs13130399	C	T	1	0.25	Exón	SNV sinónimo	.	.	.
<i>MARCH1</i>	rs17576350	G	A	0	0.25	UTR3	.	.	.	.
<i>PIK3AP1</i>	rs1051450	T	G	0.375	0.125	UTR3	.	.	.	.
<i>PIK3AP1</i>	rs1063605	T	C	0.375	0.125	UTR3	.	Si	.	.
<i>PIK3AP1</i>	rs11592969	A	G	0.125	0.125	UTR3	.	Si	.	.
<i>PIK3AP1</i>	rs12784975	T	C	0.5	0.125	Exón	SNV no sinónimo	.	T	B
<i>PIK3AP1</i>	rs17112076	C	T	0.25	0.25	Exón	SNV no sinónimo	.	T	B
<i>PIK3AP1</i>	rs3748229	C	A	0.25	0.375	Exón	SNV no sinónimo	.	T	B

Tabla suplementaria 3 (Continuación)

Gen	ID dbSNP	Alelo referenci a	Alelo alternativ o	Frecuencia alelo de alternativo Controles	Frecuencia alelo de alternativo Prematuros	Ubicación génica	Función en exón	Unión con sitio blanco de micro ARN(s)	SIFT_ pred	Polyphen2_H VAR_pred
<i>PIK3AP1</i>	rs3748234	C	T	0.25	0.75	Exón	SNV sinónimo	.	.	.
<i>PIK3AP1</i>	rs3748236	A	G	0.5	0.125	Exón	SNV sinónimo	.	.	.
<i>PIK3AP1</i>	rs4344416	G	A	1	1	Exón	SNV sinónimo	.	.	.
<i>PIK3AP1</i>	rs7074516	C	T	0.375	0.125	UTR3	.	.	.	.
<i>PIK3AP1</i>	rs7448	C	T	0.375	0.75	UTR3	.	.	.	.
<i>PIK3AP1</i>	rs8578	A	G	0.375	0.125	UTR3	.	Si	.	.
<i>PLEK</i>	rs1050181	G	C	0.5	0.625	UTR3	.	.	.	.
<i>PLEK</i>	rs1063479	G	A	0.625	0.875	Exón	SNV no sinónimo	.	T	B
<i>PLEK</i>	rs17035364	C	T	0	0.125	Exón	SNV no sinónimo	.	D	P
<i>PLEK</i>	rs1867313	T	C	0.625	0.875	UTR3	.	.	.	.
<i>PLEK</i>	rs2070171	G	A	0.375	0.625	Exón	SNV sinónimo	.	.	.
<i>PLEK</i>	rs3732044	T	C	0.5	0.625	UTR3	.	Si	.	.

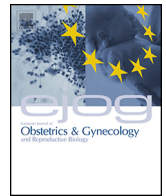
Tabla suplementaria 3 (Continuación)

Gen	ID dbSNP	Alelo referenci a	Alelo alternativ o	Frecuencia alelo de alternativo Controles	Frecuencia alelo de alternativo Prematuros	Ubicación génica	Función en exón	Unión con sitio blanco de micro ARN(s)	SIFT_ pred	Polyphen2_H VAR_pred
<i>PLEK</i>	rs3816281	G	T	0.125	0.125	Exón	SNV no sinónimo	.	T	D
<i>PLEK</i>	rs550284671	G	A	0.125	0	UTR3	.	.	.	.
<i>PLEK</i>	rs570062682	C	T	0.125	0	UTR3	.	.	.	.
<i>PLEK</i>	rs6713721	C	T	0.375	0.625	UTR3	.	Si	.	.
<i>PLEK</i>	rs6728995	T	C	0.5	0.625	UTR3	.	.	.	.
<i>PLEK</i>	rs6742200	A	G	0.5	0.625	UTR3	.	Si	.	.
<i>PLEK</i>	rs8761	G	A	0.5	0.625	UTR3	.	Si	.	.

8.3 ARTÍCULOS PUBLICADOS:

Pereyra S, Bertoni B, Sapiro R. (2016) Interactions between environmental factors and maternal-fetal genetic variations: strategies to elucidate risks of preterm birth. *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 202:20-25. <https://doi.org/10.1016/j.ejogrb.2016.04.030>.

Pereyra S, Sosa C, Bertoni B, Sapiro R. (2019) Transcriptomic analysis of fetal membranes reveals pathways involved in preterm birth. *BMC Medical Genomics*. 12:53. <https://doi.org/10.1186/s12920-019-0498-3>.



Interactions between environmental factors and maternal–fetal genetic variations: strategies to elucidate risks of preterm birth



Silvana Pereyra^a, Bernardo Bertoni^a, Rossana Sapiro^{b,*}

^a Departamento de Genética, Facultad de Medicina, Universidad de la República, Av. General Flores 2125, C.P. 11800 Montevideo, Uruguay

^b Departamento de Histología y Embriología, Facultad de Medicina, Universidad de la República, Av. General Flores 2125, C.P. 11800 Montevideo, Uruguay

ARTICLE INFO

Article history:

Received 30 December 2015

Received in revised form 6 April 2016

Accepted 23 April 2016

Keywords:

Preterm birth

Multifactorial disease

Genetic association

Inflammatory pathways

Labor

ABSTRACT

Context: Preterm birth (PTB) is a complex disease in which medical, social, cultural, and hereditary factors contribute to the pathogenesis of this adverse event. Interactions between genes and environmental factors may complicate our understanding of the relative influence of both effects on PTB. To overcome this, we combined data obtained from a cohort of newborns and their mothers with multiplex analysis of inflammatory-related genes and several environmental risk factors of PTB to describe the environmental–genetic influence on PTB.

Objective: The study aimed to investigate the association between maternal and fetal genetic variations in genes related to the inflammation pathway with PTB and to assess the interaction between environmental factors with these variations.

Study design: We conducted a case–control study at the Pereira Rossell Hospital Center, Montevideo, Uruguay. The study included 143 mother–offspring dyads who delivered at preterm (gestational age < 37 weeks) and 108 mother–offspring dyads who delivered at term. We used real-time PCR followed by a high-resolution melting analysis to simultaneously identify gene variations involved in inflammatory pathways in the context of environmental variables. The genes analyzed were: Toll-like receptor 4 (TLR4), Interleukin 6 (IL6), Interleukin 1 beta (IL1B) and Interleukin 12 receptor beta (IL12RB). **Results:** We detected a significant interaction between IL1B rs16944 polymorphism in maternal samples and IL6 rs1800795 polymorphism in newborns, emphasizing the role of the interaction of maternal and fetal genomes in PTB. In addition, smoke exposure and premature rupture of membranes (PROM) were significantly different between the premature group and controls. IL1B and IL6 polymorphisms in mothers were significantly associated with PTB when controlling for smoke exposure. TLR4 polymorphism and PROM were significantly associated with PTB when controlling for PROM, but only in the case of severe PTB.

Conclusions: Interactions between maternal and fetal genomes may influence the timing of birth. By incorporating environmental data, we revealed genetic associations with PTB, a finding not found when we analyzed genetic data alone. Our results stress the importance of studying the effect of genotype interactions between mothers and children in the context of environmental factors because they substantially contribute to phenotype variability.

© 2016 Elsevier Ireland Ltd. All rights reserved.

Introduction

Preterm birth (PTB), defined as a live birth occurring between 20 and 37 weeks of gestation, complicates 12.2% of pregnancies [1], contributes to more than one-third of infant deaths in the United

States annually [2], and is associated with serious medical complications, including neurodevelopmental delay, hearing disabilities, retinopathy, and chronic lung disease [3,4]. PTB is etiologically heterogeneous [5,6] but epidemiological evidence indicates that genetic factors play a significant role in the etiology of spontaneous PTB [7,8]. A number of candidate gene studies, almost exclusively using case–control design, have identified some genes that associate with PTB [9–13]. However, the results have rarely been replicated. Of importance for this phenotype are the possible effects of two genomes, maternal and fetal [13–15]. In addition,

* Corresponding author. Tel.: +598 2924 3414x3505; fax: +598 2924 2703.

E-mail addresses: spereyra@fmed.edu.uy (S. Pereyra), bbertoni@fmed.edu.uy (B. Bertoni), rsapiro@fmed.edu.uy (R. Sapiro).

interactions between maternal and fetal genomes may affect PTB risk [13]. As a multifactorial trait, maternal stress, multiple pregnancies, and exposure to toxics during pregnancy could interact with genetic predisposition and develop in PTB [5,6,16]. Interactions between genes and environmental factors may make it difficult to elucidate and discriminate both effects; for example, some maternal genotypes modify the association between maternal cigarette smoking and infant birth weight [17], as well as the presence of bacterial vaginosis in Tumor Necrosis Factor-2 carriers predisposes to spontaneous PTB [18]. Not considering environmental factors when performing a case–control study of genetic association and PTB may lead to misinterpretation of the results. Genome-wide association studies (GWAS) are promising but require a very large number of well-characterized subjects to overcome the challenge of multiple statistical comparisons. Given the practical limitations of assembling very large sample sizes in complex phenotypes, such as PTB, there is a need for creative research strategies that seek to resolve the tension between the hypothesis-directed candidate gene and the relatively unbiased GWAS approach. It is generally accepted that inflammation in pregnancy plays an important role [10,11,19–22]. Inflammatory pathways involve hundreds of mediators, and each of them could present several single nucleotide polymorphisms (SNPs). We previously reported a high resolution melting (HRM) analysis to simultaneously identify mutations in four genes involved in inflammatory pathways Toll-like receptor 4 (TLR4), Interleukin 6 (IL6), Interleukin 1 beta (IL1B), and Interleukin 12 receptor beta (IL12RB) [23].

In this study, our first aim was to look for associations between PTB and four SNPs of inflammatory genes TLR4, IL6, IL1B, and IL12RB in a relatively small cohort of dyads of newborns and their mothers, using a case–control study design. Secondly, we aimed to combine this genotype data with an analysis of environmental risk factors of PTB to enlighten the environmental–genetic influence on PTB.

Materials and methods

Subjects

A case–control study was conducted. Subjects were mothers and their offspring, receiving obstetrical care at the Pereira Rossell Hospital Center (CHPR), Montevideo, Uruguay, collected between February 2012 and March 2014. CHPR is the main gynecologic and obstetric center in Uruguay, and it has both secondary and tertiary care units. Cases were dyads ($n = 143$) of mothers and their neonates from pregnancies complicated by spontaneous PTB (gestational age < 37 weeks). Controls ($n = 108$) were mothers and their neonates delivered at term (gestational age ≥ 37 weeks), who experienced uncomplicated term vaginal delivery or elective cesarean. Exclusion criteria included multiple births, drug consumption, chronic pathologies and/or infections, fetal malformations, and failure to give consent to participate in the study. Infections included here were urinary tract infections, lower genital infections, and syphilis but not HIV because patients with HIV infections were excluded from the study.

Mothers were questioned regarding sociodemographic characteristics and obstetric history through a specifically designed questionnaire, including clinically diagnosed PROM, tobacco consumption, previous PTB, and education level. Medical data from mothers and newborns were collected in the perinatal informatics system [24]. Sample size was estimated at the beginning of the study, taking into consideration the frequency of minor allele of SNPs that have a possible inflammatory effect in European populations (approximately 20–40%). The odds ratio

(OR) was estimated to be approximately 1.5–2 between cases and controls.

Sample collection, DNA extraction, real-time PCR

Whole blood samples from newborns and cheek swabs from mothers were collected, and DNA was isolated using DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany). HRM analysis was performed on the Rotor-Gene 6000 real-time instrument (Corbett Life Science, Sydney, Australia) with a saturating dye technology (Type-it HRM PCR Kit, Qiagen, Hilden, Germany), as previously described in which SNPs rs4986790 (TLR4), rs1800795 (IL6), rs16944 (IL1B), and rs375947 (IL12RB) were typed [23]. DNA from individuals with heterozygous and both homozygous genotypes were included as controls in all experiments. HRM curves were normalized, and genotypes were assigned according to HRM curve shape by the Rotor-Gene software and visual inspection. Melting curves were analyzed separately for each amplicon.

Statistical analyses

Genotypes, allele frequencies, and Hardy–Weinberg (H–W) equilibrium exact tests were calculated in PLINK v1.07 [25]. Comparisons of allele frequencies were performed using the χ^2 test or Fisher's exact test and means using Student's *t*-test. Association between genotypes and environmental factors was examined by logistic regression analysis. Differential effects of genotypes on the risk of PTB were explored by the inclusion of interaction or conditional terms using Epi Info 2000 software (Centers for Disease Control and Prevention, Atlanta, GA, USA). Probability ≤ 0.05 was considered significant.

Ethics

The study protocol conforms to the Declaration of Helsinki and was approved by the School of Medicine Ethics Committee of the Republic University, Uruguay, in September 2011. Informed consent was obtained from the mothers in all cases.

Results

A total of 251 dyads of mothers and their newborn babies were included in the study. The main characteristics of the study population are shown in Table 1. No differences were detected in both groups regarding mother's age and parity. Mean $\pm$ standard

Table 1

Sociodemographic and medical characteristics of mothers and their newborns in study population.

Condition	Term (%) <i>n</i> = 108	Preterm (%) <i>n</i> = 143	OR (95%CI)	<i>p</i> -value
Maternal age (<19 and >35)	16	22	1.6 (0.8–2.9)	0.2
Education level ^a	35	43	0.7 (0.4–1.2)	0.2
Marital status ^b	17	25	1.6 (0.8–3.2)	0.1
Male newborn	51	54	1.15 (0.6–1.9)	0.6
Smoke exposure	44	65	2.6 (1.5–4.3)	<0.01
Hypertension	7	4	0.4 (0.14–1.4)	0.2
Pregnancy bleeding	1	4	4.7 (0.5–39)	0.2
Infections	19	27	1.6 (0.8–3)	0.2
IUGR	2	1	2.3 (0.6–8.9)	0.2
PROM	13	45	5.5 (2.9–10.6)	<0.01
Anemia	14	12	0.8 (0.4–1.7)	0.5

ORs: odds ratios, CI: confidence interval, *p*-values between term and preterm births.

^a At least 1 year of high school vs. less than 1 year of high school.

^b Single mothers vs. married or cohabiting mothers.

PROM: premature rupture of membranes; IUGR: intrauterine growth restriction. Statistical significance is marked as bold values.

error of gestational age was 33.6 ± 0.2 in cases vs. 39.1 ± 0.1 in controls ($p < 0.01$). Cases ranged from 24 to 36 weeks, whereas controls ranged from 37 to 41 weeks of gestational age. No statistical differences were found between both groups regarding the newborn sex (51% vs. 54%, male newborns at term vs. preterm, respectively). Seven percent of mothers who delivered at term and 17% of mothers who delivered preterm had at least one previous PTB, with no statistical differences observed between groups. The mothers' education level did not affect the incidence of PTB; 35% of mothers who had completed at least 1 year of high school or more delivered at term, and 43% had preterm newborns. Smoke exposure (active or passive smoker) was associated to PTB ($p < 0.01$): 44% of mothers who delivered their newborns at term were exposed to smoke, whereas 66% had a PTB (Table 1). No statistical differences were found between mothers who delivered at term and preterm regarding maternal hypertension (7% vs. 4%, respectively), pregnancy bleeding (1% vs. 4%, respectively), anemia (14% vs. 12%, respectively), and intrauterine growth restriction (2% vs. 1%, respectively). PROM was statistically significantly associated with PTB (Table 1). Maternal infections were not associated to PTB (19% vs. 27%, respectively).

Maternal and newborn allelic frequencies for each SNP are shown in Table 2. Similarly to our previous findings [23], genotypes C/C for rs1800795 (IL6) and G/G for rs4986790 (TLR4) were absent in the population. None of the SNPs analyzed from mothers were significantly associated with PTB, when analyzed with a Chi-squared test for the alleles. In addition, no significant association was found between term and preterm newborns for the analyzed SNPs ($p > 0.05$; Table 3).

The H–W equilibrium exact test indicates that SNPs from both groups of mothers rs4986790 (gene TLR4) and rs375947 (gene IL12RB) are in H–W equilibrium ($p > 0.05$), whereas gene rs1800795 (in gene IL6) and rs16944 (in gene IL1B) are not. When analyzing newborn genotypes, the same scenario was detected (Table 3).

Logistic regression analysis detected a significant interaction between IL1B polymorphism in maternal samples and IL6 polymorphism in newborns (Table 4). When analyzing the effect of the SNP IL6 in the newborn while controlling mother IL1B variations, a significant association was found with PTB, indicating a possible hidden effect when the SNPs are considered separately. Only the genotype combination of IL1B G/G genotype in mothers and IL6 G/C genotype in newborns was different between controls and PTB cases ($p < 0.05$; Fisher test; Fig. 1). No other significant associations were detected between maternal and newborn genotypes and PTB.

Table 2
Genotype SNP counts for mothers and newborns of study population. Frequencies are shown between parentheses.

SNP (Gene symbol)	Genotype	Mothers		Newborns	
		Controls	Cases	Controls	Cases
rs4986790 (TLR4)	A/A	92 (0.89)	124 (0.87)	90 (0.87)	128 (0.91)
	A/G	11 (0.11)	18 (0.13)	13 (0.13)	13 (0.09)
	G/G	0	0	0	0
rs1800795 (IL6)	C/C	0	0	0	0
	C/G	38 (0.35)	64 (0.45)	28 (0.26)	45 (0.31)
	G/G	70 (0.65)	79 (0.55)	80 (0.74)	98 (0.69)
rs16944 (IL1B)	A/A	16 (0.16)	17 (0.12)	15 (0.14)	19 (0.13)
	A/G	61 (0.61)	81 (0.57)	59 (0.55)	89 (0.63)
	G/G	23 (0.23)	44 (0.31)	33 (0.31)	33 (0.23)
rs375947 (IL12RB)	A/A	49 (0.47)	56 (0.40)	42 (0.38)	59 (0.42)
	A/G	37 (0.36)	61 (0.44)	51 (0.47)	67 (0.47)
	G/G	18 (0.17)	22 (0.16)	15 (0.14)	16 (0.11)

Cases vs. control ($p \geq 0.05$; χ^2 test).

Even though the studied population was quite homogeneous, two variables were significantly different between premature births and controls: smoke exposure and PROM. When analyzing the genotyped SNPs while controlling the effect of each of these environmental variables, we uncovered the effect of smoke exposure on PTB. IL1B and IL6 polymorphisms in mothers were significantly associated with PTB when controlling for smoke exposure (Table 5). TLR4 polymorphism and PROM were significantly associated with PTB when controlling for PROM but only in the case of severe PTB ($p < 0.05$; data not shown), confirming previously reported data [26].

Discussion

In this study, we combined a control–case dyad design, the analysis of several SNPs from a common pathway, and the study of some risk environmental factors to PTB. This approach is seldom used in association studies, in which only genotype data is incorporated. This type of analysis was fruitful because it allowed for discriminating differences between mother and fetal contribution to PTB; we could also determine if the variables tended to be heritable or lost between generations, as well as explore a hypothetical pathway of the trait.

By analyzing the risk of PTB associated with the genotypes of both mother and child, we detected a higher risk of PTB when the mother and the newborn carried a specific combination of SNPs. The presence of the IL1B G/G genotype in mothers and IL6 G/C genotype in the fetal genome is associated with PTB, emphasizing the role of the interaction of maternal and fetal genomes in the determination of PTB (Fig. 1, Table 4). When both genes are considered separately, this effect is not observed, indicating that one SNP may hide the effect of the other, probably due to opposite actions on the cascade that determines PTB. Furthermore, the high OR (12.3; Table 4) suggests that carrying the specific combination of SNPs indicated above is a much higher risk than the others. The combined effect of maternal and fetal genomes on PTB began to be discussed in the literature relatively recently [13]. Some studies suggest a maternal genetic contribution to the timing of birth [27,28], whereas some others support the effect of both maternal and fetal genomes separately [14,15,29]. To the best of our knowledge, no other studies have focused on the relationship between maternal and fetal genes in PTB. The data presented in this work suggest that the consequence of carrying a particular combination of variations between mother and the newborn may affect the time of birth. This significant association needs to be further studied in a larger sample to assess its potential clinical implications with PTB and to bring awareness to the issue that the mother's risk to have another child after one PTB is reliant not only of her own genotype but also of her future child's genotype.

The quadruplex assay developed in our laboratory focused on SNPs related to inflammatory pathways involved in host defense mechanisms, innate immunity activation, and infection, as indicated in some of the web-based reference databases that link the genome to biologic systems (Kyoto Encyclopedia of Genes and Genomes, <http://www.genome.jp/kegg/pathway.html>; http://www.genome.jp/kegg-bin/show_pathway?map04620). The effect of these genes in PTB is not surprising because both genes are in close relation regarding immune responses during pregnancy. PTB may reflect early activation of the normal parturition cascade, in which proinflammatory mediators, such as prostanooids and cytokines, are typically induced [22,30,31]. Therefore, polymorphisms in genes regulating the adaptive or innate immune system of both genomes could modify the pregnancy immune response and affect parturition timing. For instance, the non-synonymous TLR4 4 SNP (rs4986790) has been previously associated with differences in lipopolysaccharide

Table 3Hardy–Weinberg and χ^2 statistics for mother and newborn genotypes for the analyzed SNPs discriminated between cases and controls.

SNP (Gene symbol)	Minor allele	Mothers						Newborns					
		MAF in cases	MAF in controls	p-value HWE (cases)	p-value HWE (controls)	OR	p-value χ^2	MAF in cases	MAF in controls	p-value HWE (cases)	p-value HWE (controls)	OR	p-value χ^2
rs4986790 (TLR4)	G	0.06	0.05	1	1	1.2	0.64	0.05	0.06	1	1	0.72	0.41
rs1800795 (IL6)	C	0.22	0.18	<0.01	<0.05	1.35	0.19	0.16	0.13	<0.05	0.21	1.25	0.38
rs16944 (IL1B)	A	0.4	0.47	<0.05	<0.05	0.78	0.19	0.45	0.42	<0.01	0.23	1.15	0.44
rs375947 (IL12RB)	G	0.38	0.35	0.31	0.48	1.12	0.55	0.35	0.38	0.71	1	0.89	0.54

MAF: Minimum allele frequency, HWE: Hardy–Weinberg equilibrium, OR: odds ratio, p-value χ^2 : cases vs. control.

responsiveness and predisposition to PTB [26,32,33]. Interleukin 6 polymorphism 174 (rs1800795) G-allele carriers produce higher levels of IL-6 than those with the C/C genotype. Genotype C/C is protective against PTB in women of European descent, but it is not significant in other heterogeneous or admixed populations or in fetal genotype analysis [34]. In addition, the analyzed SNP in the IL1B gene promoter (rs16944) is associated with increased risk of PTB in a European population [35]. The protein encoded by the interleukin 12 receptor B (IL12RB) gene is a type I transmembrane protein that belongs to the hemopoietin receptor superfamily. The lack of expression of this gene was found to result in the immunodeficiency of patients with severe mycobacterial infections [36], and particularly the rs375947 SNP is associated with atopic dermatitis [37] but the association with PTB has not been particularly analyzed.

The data presented in this work support the idea of a possible contribution of the inflammatory pathway to PTB because we found an association of a specific combination of inflammatory SNPs (as in the case of IL1B and IL6 mentioned above) that predispose to PTB. However, no association was found neither in the other analyzed SNP combinations nor when these genotypes were analyzed separately. Ethnic differences may account for part of the discordances between our results and those of other studies. Most of the genetic association analyses of PTB have been performed in the USA or Europe, which makes it difficult to compare the reported results with populations with different ancestral backgrounds [27,38,39]. The non-European genetic contribution to the population of Uruguay has been estimated approximately 10% Native American and 6% African [40]. Maternal lineages assessed using mitochondrial DNA revealed a higher Native American ancestral proportion of 34%, whereas the African contribution varies between 8% and 21% [41–43]. Taking all these considerations together the observed differences between genetic risk to PTB with other populations were unsurprising.

In this work, we also tested the possibility that environmental and genetic factors may interact to lead to PTB. Only smoking and PROM were differentially distributed in cases and controls, as expected according to previous studies [5,6,16,44–46]. Exposure to tobacco during pregnancy is a well-documented risk factor for pregnancy complications [44,46,47] so it may hide the genetic predisposition to PTB. We found that when accounting for the effect of smoke exposure, it was possible to uncover the association

of both IL1B and IL6 polymorphisms carried by the mother with PTB. Thus, by incorporating environmental data, we revealed that both these SNPs are associated with PTB, a finding not found when we analyzed genetic data alone. This finding further stresses the importance of studying the effect of genotype in the context of environmental factors because they substantially contribute to phenotype variability. This may be a reason why results are rarely replicated in case–control studies in the genetics of PTB. The etiology of PTB itself is heterogeneous. PTB is likely the result of various different underlying causes [48,49]. In addition, several environmental factors are known to affect birth outcome, and consequently they must be considered when interpreting genetic data. Thus, when designing research studies on PTB, both genetic and environmental variables must be incorporated.

The combined strategy used here, analyzing genotype data of mother–child dyads in conjunction with environmental variables, requires a well-characterized population and a smaller sample size compared with modern GWAS level analysis, as well as a limited number of candidate genes. Few other candidate gene studies on PTB have been performed using this design, which offers protection against bias due to population stratification.

The same advantages discussed here may also be considered as weaknesses of the study. Our study was limited in that the sample size was small compared with modern GWAS-level analyses and in that it was based on a limited number of candidate genes.

Table 4

Logistic regression analysis of cases vs. controls for maternal rs16944 (IL1B) and newborn rs1800795 (IL6) adjusted to the other polymorphism as well as their interactions.

Genotype	OR (95%CI)	p-value
rs16944 (IL1B) mother	1.14 (1.0–1.3)	0.09
rs1800795 (IL6) newborn	12.30 (2.0–76.7)	<0.01
interaction	0.3 (0.1–0.8)	0.01

OR: odds ratio, CI: confidence interval.

Statistical significance is marked as bold values.

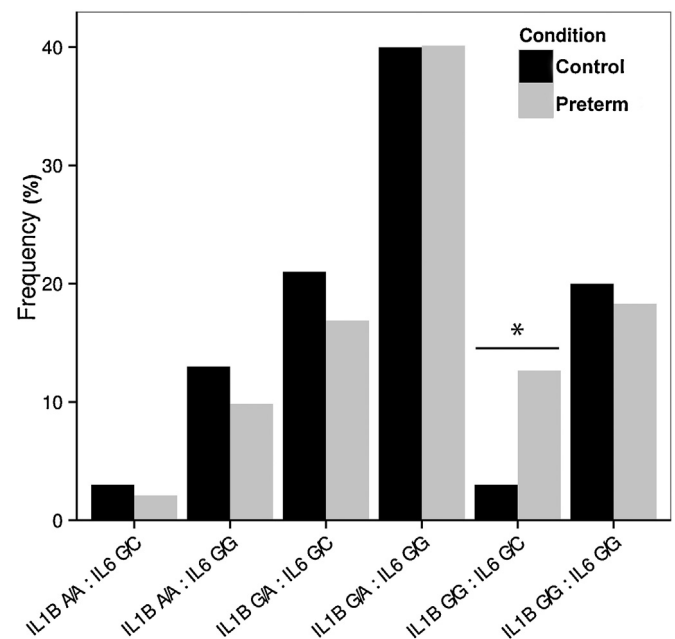


Fig. 1. Distribution of the combined genotypes from the mother–newborn dyads for the mother's IL1B (rs16944) and newborn's IL6 (rs1800795). Only the combination of IL6 G/G genotype in mothers and IL6 G/C genotype in newborns was different between case and control groups ($p < 0.05$, χ^2 statistics).

Table 5

Logistic regression analysis on the maternal genotypes adjusted by mothers who had smoke exposure and the corresponding interaction with smoke exposure.

	Adjusted OR (95%CI) p-value		Interaction (95%CI) p-value	
rs4986790 (TLR4),	1 (0.3–3.5)	>0.05	2.1 (0.3–13.1)	>0.05
rs1800795 (IL6)	2.8 (1.2–6.4)	<0.05	0.3 (0.1–0.9)	<0.05
rs16944 (IL1B)	0.4 (0.2–0.8)	<0.01	3.3 (1.3–7.9)	<0.01
rs375947 (IL12RB)	1.2 (0.7–2.0)	>0.05	0.8 (0.4–1.7)	>0.05

OR: odds ratio, CI: confidence interval.

Statistical significance is marked as bold values.

Consequently, the amount of information and statistical significance of the data are less compared with GWAS. Another weakness of this study is that we did not have an external replication sample to corroborate our findings. The findings should thus be regarded as exploratory, although the prior plausibility of the genes studied provides increased confidence in our results.

Conclusion

Although exploratory, our data suggest that an inflammatory pathway is involved in PTB, maternal genotypes affect the expression of fetal genes involved in delivery, and environmental factors (e.g., smoking) may hide the effect of genotype variations.

Funding

This work was funded by Comision Sectorial de Investigacion Cientifica-CSIC and Sistema Nacional de Investigadores-ANII (Uruguay). Sponsors had no involvement in the collection, analysis and interpretation of data nor in the writing of the manuscript.

Acknowledgements

We acknowledge Drs. Justo Alonso, Grazzia Rey, Daniel Borbonet, Manuela DeMaria, Elsa Arocena, Elena Lepre, Agustina Bado, Patricia Cardozo and Ariel Diaz for their help on patient recruitment.

References

- [1] Kochanek KD, Kirmeyer SE, Martin JA, Strobino DM, Guyer B. Annual summary of vital statistics: 2009. *Pediatrics* 2012;129(2):338–48.
- [2] MacDorman MF, Callaghan WM, Mathews TJ, Hoyert DL, Kochanek KD. Trends in preterm-related infant mortality by race and ethnicity, United States, 1999–2004. *Int J Health Serv* 2007;37(4):635–41.
- [3] Landry JS, Menzies D. Occurrence and severity of bronchopulmonary dysplasia and respiratory distress syndrome after a preterm birth. *Paediatr Child Health* 2011;16(7):399–403.
- [4] Sharma PK, Sankar MJ, Sapra S, et al. Growth and neurosensory outcomes of preterm very low birth weight infants at 18 months of corrected age. *Indian J Pediatr* 2011;78(12):1485–90.
- [5] Wen SW, Smith G, Yang Q, Walker M. Epidemiology of preterm birth and neonatal outcome. *Semin Fetal Neonatal Med* 2004;9(6):429–35.
- [6] Berkowitz G, Papiernik E. Epidemiology of preterm birth. *Epidemiol Rev* 1993;15(2):414–43.
- [7] Anum EA, Springel EH, Shriver MD, Strauss 3rd JF. Genetic contributions to disparities in preterm birth. *Pediatr Res* 2009;65(1):1–9.
- [8] Wu W, Witherspoon D, Fraser A, et al. The heritability of gestational age in a two-million member cohort: implications for spontaneous preterm birth. *Hum Genet* 2015;134(7):803–8.
- [9] Romero R, Velez Edwards DR, Kusanovic JP, Hassan SS, Mazaki-Tovi S, Vaisbuch E. Identification of fetal and maternal single nucleotide polymorphisms in candidate genes that predispose to spontaneous preterm labor with intact membranes. *Am J Obstet Gynecol* 2010;202.
- [10] Velez D, Fortunato S, Thorsen P, Lombardi S, Williams S, Menon R. Spontaneous preterm birth in African Americans is associated with infection and inflammatory response gene variants. *Am J Obstet Gynecol* 2009;200(2):209.
- [11] Velez D, Fortunato S, Thorsen P, Lombardi S, Williams S, Menon R. Preterm birth in Caucasians is associated with coagulation and inflammation pathway gene variants. *PLoS ONE* 2008;3(9):e3283.
- [12] Ryckman KK, Morken NH, White MJ, et al. Maternal and fetal genetic associations of PTGER3 and PON1 with preterm birth. *PLoS ONE* 2010;5(2):e9040.

- [13] Myking S, Myhre R, Gjessing H, et al. Candidate gene analysis of spontaneous preterm delivery: new insights from re-analysis of a case-control study using case-parent triads and control-mother dyads. *BMC Med Genet* 2011;12(1):174.
- [14] York TP, Eaves LJ, Lichtenstein P, et al. Fetal and maternal genes' influence on gestational age in a quantitative genetic analysis of 244,000 Swedish births. *Am J Epidemiol* 2013;178(4):543–50.
- [15] York TP, Strauss 3rd JF, Neale MC, Eaves LJ. Estimating fetal and maternal genetic contributions to premature birth from multiparous pregnancy histories of twins using MCMC and maximum-likelihood approaches. *Twin Res Hum Genet: Off J Int Soc Twin Stud* 2009;12(4):333–42.
- [16] Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet* 2008;371(9606):75–84.
- [17] Wang X, Zuckerman B, Pearson C, et al. Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. *JAMA* 2002;287(2):195–202.
- [18] Macones G, Parry S, Elkousy M, Clothier B, Ural S, Strauss J. A polymorphism in the promoter region of TNF and bacterial vaginosis: preliminary evidence of gene-environment interaction in the etiology of spontaneous preterm birth. *Am J Obstet Gynecol* 2004;190(6):1504–8.
- [19] Hartel C, Finas D, Ahrens P, et al. Polymorphisms of genes involved in innate immunity: association with preterm delivery. *Mol Hum Reprod* 2004;10(12):911–915.
- [20] Menon R, Fortunato SJ. Infection and the role of inflammation in preterm premature rupture of the membranes. *Best Pract Res Clin Obstet Gynaecol* 2007;21(3):467–78 [Epub 2007/04/24].
- [21] Nace J, Fortunato SJ, Maul H, Menon R. The expression pattern of two novel cytokines (IL-24 and IL-29) in human fetal membranes. *J Perinat Med* 2010;38(6):665–70 [Epub 2010/08/17].
- [22] Romero R, Espinoza J, Goncalves LF, Kusanovic JP, Friel L, Hassan S. The role of inflammation and infection in preterm birth. *Semin Reprod Med* 2007;25(1):21–39.
- [23] Pereyra S, Velazquez T, Bertoni B, Sapiro R. Rapid multiplex high resolution melting method to analyze inflammatory related SNPs in preterm birth. *BMC Res Notes* 2012;5:69.
- [24] Sosa CG, Althabe F, Belizan JM, Buekens P. Use of oxytocin during early stages of labor and its effect on active management of third stage of labor. *Am J Obstet Gynecol* 2011;204(3):238 e1–5 [Epub 2010/12/15].
- [25] Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81(3):559–75.
- [26] Rey G, Skowronek F, Alciaturi J, Alonso J, Bertoni B, Sapiro R. Toll receptor 4 Asp299Gly polymorphism and its association with preterm birth and premature rupture of membranes in a South American population. *Mol Hum Reprod* 2008;14(9):555–9.
- [27] Svensson AC, Sandin S, Cnattingius S, et al. Maternal effects for preterm birth: a genetic epidemiologic study of 630,000 families. *Am J Epidemiol* 2009;170(11):1365–1372.
- [28] Kalish RB, Vardhana S, Normand NJ, Gupta M, Witkin SS. Association of a maternal CD14-159 gene polymorphism with preterm premature rupture of membranes and spontaneous preterm birth in multi-fetal pregnancies. *J Reprod Immunol* 2006;70(1–2):109–17 [Epub 2006/01/24].
- [29] Lunde A, Melve KK, Gjessing HK, Skjaerven R, Irgens LM. Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. *Am J Epidemiol* 2007;165(7):734–41.
- [30] Kim CJ, Romero R, Chaemsaithong P, Kim J-S. Chronic inflammation of the placenta: definition, classification, pathogenesis, and clinical significance. *Am J Obstet Gynecol* 2015;213(Suppl. (4)):S53–69.
- [31] Cappelletti M, Della Bella S, Ferrazzi E, Mavilio D, Divanovic S. Inflammation and preterm birth. *J Leukocyte Biol* 2016;99(1):67–78 [Epub 2015/11/06].
- [32] Arbour NC, Lorenz E, Schutte BC, et al. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet* 2000;25(2):187–191.
- [33] Lorenz E, Hallman M, Marttila R, Haataja R, Schwartz DA. Association between the Asp299Gly polymorphisms in the Toll-like receptor 4 and premature births in the Finnish population. *Pediatr Res* 2002;52(3):373–6.
- [34] Wu W, Clark EAS, Stoddard GJ, et al. Effect of interleukin-6 polymorphism on risk of preterm birth within population strata: a meta-analysis. *BMC Genet* 2013;14:30–40.
- [35] Hollegaard MV, Grove J, Thorsen P, et al. Polymorphisms in the tumor necrosis factor alpha and interleukin 1-beta promoters with possible gene regulatory functions increase the risk of preterm birth. *Acta Obstet Gynecol Scand* 2008;87(12):1285–90 [Epub 2008/10/28].
- [36] van de Vosse E, Lichtenauer-Kaligis EG, van Dissel JT, Ottenhoff TH. Genetic variations in the interleukin-12/interleukin-23 receptor (beta1) chain, and implications for IL-12 and IL-23 receptor structure and function. *Immunogenetics* 2003;54(12):817–29.
- [37] Namkung JH, Lee JE, Kim E, et al. Association of single nucleotide polymorphisms in the IL-12 (IL-12A and B) and IL-12 receptor (IL-12Rbeta1 and beta2) genes and gene-gene interactions with atopic dermatitis in Koreans. *J Dermatol Sci* 2010;57(3):199–206.
- [38] Menon R, Velez DR, Thorsen P, et al. Ethnic differences in key candidate genes for spontaneous preterm birth: TNF-alpha and its receptors. *Hum Hered* 2006;62(2):107–18 [Epub 2006/10/19].

- [39] Menon R, Pearce B, Velez DR, Meriardi M, Williams SM, Fortunato SJ. Racial disparity in pathophysiologic pathways of preterm birth based on genetic variants. *Reprod Biol Endocrinol* 2009;7(1):1–16 [Epub 2009/06/17].
- [40] Hidalgo P, Bengochea M, Abilleira D, Cabrera A, Alvarez I. Genetic admixture estimate in the Uruguayan population based on the loci LDLR, GYP A, HBGG, GC and D7S8. *Int J Hum Genet* 2005;5(3):217.
- [41] Gascue C, Mimbacas A, Sans M, et al. Frequencies of the four major Amerindian mtDNA haplogroups in the population of Montevideo, Uruguay. *Hum Biol* 2005;77(6):873–8.
- [42] Bonilla C, Bertoni B, Hidalgo PC, et al. Breast cancer risk and genetic ancestry: a case–control study in Uruguay. *BMC Womens Health* 2015;15:11.
- [43] Pagano S, Sans M, Pimenoff V, et al. Assessment of HV1 and HV2 mtDNA variation for forensic purposes in an Uruguayan population sample. *J Forensic Sci* 2005;50(5):1239.
- [44] Ko TJ, Tsai LY, Chu LC, et al. Parental smoking during pregnancy and its association with low birth weight, small for gestational age, and preterm birth offspring: a birth cohort study. *Pediatr Neonatol* 2014;55(1):20–7 [Epub 2013/07/16].
- [45] Ananth CV, Vintzileos AM. Epidemiology of preterm birth and its clinical subtypes. *J Matern Fetal Neonatal Med* 2006;19(12):773–82.
- [46] York TP, Eaves LJ, Neale MC, Strauss 3rd JF. The contribution of genetic and environmental factors to the duration of pregnancy. *Am J Obstet Gynecol* 2014;210(5):398–405.
- [47] Shah NR, Bracken MB. A systematic review and meta-analysis of prospectivestudies on the association between maternal cigarette smoking and pretermdelivery. *Am J Obstet Gynecol* 2000;182(2):465–72 [Epub 2000/02/29].
- [48] Dolan SM, Hollegaard MV, Meriardi M, et al. Synopsis of preterm birth genetic association studies: the preterm birth genetics knowledge base (PTBGene). *Public Health Genom* 2010;13(7–8):514–23.
- [49] Romero R, Dey SK, Fisher SJ. Preterm labor: one syndrome, many causes. *Science* 2014;345(6198):760–5 [Epub 2014/08/16].

RESEARCH ARTICLE

Open Access



Transcriptomic analysis of fetal membranes reveals pathways involved in preterm birth

Silvana Pereyra¹, Claudio Sosa², Bernardo Bertoni¹ and Rossana Sapiro^{3*} 

Abstract

Background: Preterm birth (PTB), defined as infant delivery before 37 weeks of completed gestation, results from the interaction of both genetic and environmental components and constitutes a complex multifactorial syndrome. Transcriptome analysis of PTB has proven challenging because of the multiple causes of PTB and the numerous maternal and fetal gestational tissues that must interact to facilitate parturition. The transcriptome of the chorioamnion membranes at the site of rupture in PTB and term fetuses may reflect the molecular pathways of preterm labor.

Methods: In this work, chorioamnion membranes from severe preterm and term fetuses were analyzed using RNA sequencing. Functional annotations and pathway analysis of differentially expressed genes were performed with the GAGE and GOSeq packages. A subset of differentially expressed genes in PTB was validated in a larger cohort using qRT-PCR and by comparing our results with genes and pathways previously reported in the literature.

Results: A total of 270 genes were differentially expressed (DE): 252 were upregulated and 18 were down-regulated in severe preterm births relative to term births. Inflammatory and immunological pathways were upregulated in PTB. Both types of pathways were previously suggested to lead to PTB. Pathways that were not previously reported in PTB, such as the hemopoietic pathway, appeared upregulated in preterm membranes. A group of 18 downregulated genes discriminated between term and severe preterm cases. These genes potentially characterize a severe preterm transcriptome pattern and therefore are candidate genes for understanding the syndrome. Some of the downregulated genes are involved in the nervous system, morphogenesis (*WNT1*, *DLX5*, *PAPPA2*) and ion channel complexes (*KCNJ16*, *KCNB1*), making them good candidates as biomarkers of PTB.

Conclusions: The identification of this DE gene pattern will help with the development of a multi-gene disease classifier. These markers were generated in an admixed South American population in which PTB has a high incidence. Since the genetic background may differentially impact different populations, it is necessary to include populations such as those from South America and Africa, which are usually excluded from high-throughput approaches. These classifiers should be compared to those in other populations to obtain a global landscape of PTB.

Keywords: RNA-Seq, Preterm birth, Gestational age, Biomarkers

Background

Preterm birth (PTB), defined as the delivery of an infant before 37 weeks of completed gestation, is a worldwide health problem and remains the leading cause of global perinatal morbidity and mortality [1–3]. PTB is a complex, multifactorial syndrome comprised of multiple clinical subtypes that can be defined as either idiopathic or ‘medically indicated’ (by caesarean or labor induction). Idiopathic or

spontaneous PTB (sPTB) accounts for 70% of total PTB, while medically indicated PTB usually represents 30% of total PTB [4, 5]. Within the sPTB group, 45% of cases may occur without preterm rupture of membranes, while the remaining 25% are the consequence of the preterm premature rupture of membranes (PPROM) [5–8]. PTB has also been stratified according to gestational age (GA); neonates born between 24 and 33 weeks (severe PTB) are at higher risk of death and diseases later in life than moderate PTB (GA between 34 and 36 weeks). It has been speculated that PTB from different GA groups has diverse causes and/or pathological

* Correspondence: rsapiro@fmed.edu.uy

³Departamento de Histología y Embriología, Facultad de Medicina, Universidad de la República, Av. General Flores 2125, C.P. 11800 Montevideo, Uruguay

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

mechanisms [8]. Regardless of the PTB subtype, current therapies are not successful in prolonging time to birth once labor has been initiated [3].

PTB occurs as a result of the interaction of both genetic and environmental components, and it constitutes a complex multifactorial syndrome [3, 9]. The genetic architecture of pregnancy and PTB has proven challenging not only because of the multiple causes of PTB but also because of the numerous maternal and fetal gestational tissues that must interact to facilitate parturition [3, 10]. These tissues include the decidua, myometrium, cervix, maternal blood originating from the mother and villous placenta, fetal membranes (chorion and amnion), umbilical cord, and fetal blood originating from the fetus [5].

The main etiological factors related to PTB are inflammation, hemorrhage, activation of the maternal or fetal hypothalamic-pituitary axis, immune dysregulation, distension of the myometrium and cervical insufficiency [1, 3, 9, 11]. All these processes have diverse and distinctive ways of initiating labor but may share a common pathway that ends in the release of mediators that stimulate myometrial contraction, degradation of extracellular matrix components, inflammation and apoptosis. Consequently, these processes promote membrane rupture, cervical ripening, and uterine emptying, resulting in PTB [1, 11]. Studies of the transcriptomes of these tissues can help with the development of a molecular landscape of preterm labor and improve understanding of the physiology and pathology of term and preterm parturition. Specifically, study of the transcriptome of the membranes at the site of rupture in PTB may indicate shared genes of those pathways.

RNA sequencing (RNA-Seq) is a potent technology for transcriptome analysis that allows for a comprehensive characterization of gene expression [12]. The published RNA-Seq studies on human labor have been restrained to normal term pregnancies and confined to the placenta at different GAs [13–15]. More placental gene expression data are available from experiments based on microarrays [16], but most of these are concentrated on preeclampsia [13].

As mentioned, GA determines the diverse pathological mechanisms of PTB [8]. Genetic background seems to play a more relevant role in severe PTB neonates than in moderate PTB [17]. Modifications to chorioamniotic expression that end in severe PTB should be more drastic than those ending in term delivery or even moderate PTB. Therefore, we decided to focus on the transcriptome of severe PTB chorioamniotic tissues in an attempt to find DE genes with more biological significance. To validate the principal pathways found in this study, our results were compared with data previously reported in the literature.

Based on our previous work in the field [17–20] and the current status of knowledge of term and preterm

labor [21–26], we anticipate modifications in inflammatory pathways in the preterm transcriptome compared to those in the term transcriptome. Nevertheless, the ultimate goal of the study is to obtain a PTB expression signature.

Methods

Patient recruitment

Controls and cases were term and severe preterm deliveries from unrelated offspring of women receiving obstetrical care at the Pereira Rossell Hospital Center, Montevideo, Uruguay. Preterm tissues were collected immediately after labor from pregnancies complicated by birth before 33 weeks of gestational age (GA). Term chorioamnion tissues were obtained from uncomplicated pregnancies delivered after 37 weeks of GA. Multiple gestations, fetal anomalies, eclampsia and C-surgery delivery were excluded in both groups, as well as medically indicated PTB. Due to the high incidence of premature rupture of membranes (PROM) [6, 17], we did not exclude PROM from either cases or controls, and instead we used PROM as a cofactor in the statistical analyses (see below).

Maternal demographic characteristics were collected through questionnaires filled out by the mothers after delivery. Clinical and obstetric data were obtained from the Perinatal Information System, which consists of basic perinatal clinical records developed by the Latin American Center for Perinatology (CLAP) from WHO/PAHO [27].

Chorioamnion tissue collection, RNA extraction, and sequencing

Samples were collected within 30 min post-delivery. The amnion and chorion were obtained from the extraplacental membranes (reflected membranes), which provide a purer source of the fetal membranes. A 1 cm² portion of chorioamniotic membrane surrounding the exact place of membrane rupture for each subject was collected following the procedure described by Nhan-Chang et al. [24] and immediately frozen in liquid nitrogen or submerged in RNAlater solution (Qiagen, Hilden-Germany). The amnion-chorion was processed together. All samples were later stored at –80 °C until laboratory procedures were performed. Samples were ground into a fine powder in liquid nitrogen with a precooled pestle and mortar and subjected to RNA extraction.

Total RNA was extracted from each sample using a TRIzol® (Invitrogen, Carlsbad, California, USA) RNA extraction protocol that produces messenger RNA-enriched purification. The quality and concentration of RNA products were determined by UV-absorbance spectrophotometry (Nanodrop Technologies Inc., Wilmington, DE, USA). The integrity of the RNA molecules was checked using a

2100 Bioanalyzer platform (Agilent Technologies, Santa Clara, California, USA). We selected RNA samples from eight tissues (4 cases and 4 controls) for paired-end sequencing that were matched based on the mother's age, fetus sex and socioeconomic status and the lack of other medical complications, such as preeclampsia or intrauterine growth restriction (IUGR). All samples had RNA integrity numbers of 8 or higher. RNA samples were shipped to Macrogen Inc. (Seoul, South Korea) under the recommended RNA submission conditions, simultaneously, to avoid batch effects [28]. Upon arrival, samples were further assessed for RNA integrity using a 2100 Bioanalyzer platform (Agilent Technologies). Messenger-RNA (mRNA) content was purified from total RNA with the PolyATract mRNA Isolation System II (Promega Inc., Madison, Wisconsin, USA) and copied into cDNA molecules using an Illumina TruSeq RNA Sample Preparation Kit v2. One cDNA library was constructed for each specimen; libraries were subjected to massive sequencing following the Illumina HiSeq2000 protocol (Illumina Inc. San Diego California, USA). A read was defined as a 100 bp cDNA fragment sequenced from both ends (paired-end). The data have been deposited in the Sequence Read Archive (NCBI) and are accessible through SRA Series accession number SRP139931. A subset of high-quality extracted RNAs, comprising a total of 15 term and 9 severe PTB samples, including the sequenced samples, were used to validate the results by real-time quantitative-PCR (see below).

Mapping reads to the reference genome

The obtained reads were trimmed and clipped for quality control in Trimmomatic v0.32 [29] and checked for quality using FastQC v0.11.2 [30]. Reads were then aligned to the GRCh37 reference genome using Tophat v2.1.0 [31] and Ensembl annotations [32], as derived from Ensembl Release 75.

For data visualization purposes only, gene read counts were transformed by the regularized logarithm (*rlog*) [33]. This transformation removes the dependence of the variance on the mean and normalizes count data with respect to library size.

Identification of differentially expressed (DE) genes and gene set enrichment analysis

HTSeq-count with the parameters *m* = union, *s* = no, and *t* = exon was used to produce raw read counts for the expression of each gene [34]. Differential expression analysis on the gene level was performed with the R packages DESeq2 v1.16.1 [33], edgeR v3.18.0 [35] and Cuffdiff v2.2.1 [36]. A linear model with preterm condition and PROM as a cofactor was employed to analyze case-control conditions. A complementary differential expression analysis contrasting RNAseq data from tissues that presented PROM (*n* = 3) vs. those without

PROM (*n* = 5) using preterm as a cofactor was also performed. *P*-values were adjusted for multiple testing using the Benjamini and Hochberg ("BH") approach [37] to control the false discovery rate (FDR) [32]. A gene was considered expressed if it had more than 5 aligned reads. Genes were identified as DE with the following criteria: absolute logarithm 2-fold change > 2 and FDR-adjusted *p*-value < 0.05 for multiple hypothesis testing (method = "BH"). With the aim of reducing false-positive hits, we required a gene to be selected with these two criteria by the three mentioned algorithms to be considered DE.

To understand the underlying biological processes, functional annotations of DE genes were performed using the GAGE and GOSeq packages of Bioconductor 3.5 [38, 39]. The GOSeq method considers the effect of selection bias in RNA-Seq data that can arise as a result of gene length differences [39]. All gene mapping was performed using the biomaRt R package [40]. We looked for enrichment via genetic associations with KEGG pathways and Gene Ontology (GO) terms; GO terms were supported by at least 3 analyzed genes. Multiple testing was adjusted using the BH approach, and enrichment was declared if the BH adjusted *p*-value was less than 0.05. Principal component analysis (PCA) was used to explore the efficiency of DE genes to explain preterm delivery.

Additionally, we investigated whether the identified severe preterm DE genes were previously associated with PTB. A reference list was generated by exploring available public databases and recent meta-analysis associations between those genes and PTB (years 2012 to 2018). Those data included a database of associations between SNPs and PTB [41], two recent meta-analysis of transcriptomic studies of several PTB tissues [5, 42], the previously mentioned transcriptomic study of chorioamnion [43] and publicly available gene expression data (<http://www.genetation.org/analysis/gene/expression/> and [44–46]).

Validation of RNA-Seq results by assessing gene expression via quantitative real-time PCR (qRT-PCR)

Six genes were selected based on greater absolute logarithm fold changes and higher statistical significance for the confirmation of DE gene data by qRT-PCR: interleukin 1 beta (*IL1B*), lipocalin 2 (*LCN2*), macrophage receptor with collagenous structure (*MARCO*), caspase 5 (*CASP5*), serpin family A member 1 (*SERPINA1*), and TNF superfamily member 15 (*TNFSF15*). The expression values of these genes were evaluated in the 15 term birth and 9 severe PTB samples.

Total RNA from each sample (1 µg) was used to synthesize first-strand cDNA using a Superscript II RT Reagent Kit (Invitrogen, Carlsbad, California, USA) and random primers (0.5 µg/µl). qRT-PCR amplifications were performed in a Corbett Real-time Thermocycler (Qiagen, Hilden-Germany) using the Biotools SYBR

Green Kit (Biotools, Madrid, Spain). The relative expression ratio of a target gene was calculated as described in the 2-DDCt method [32].

Reaction mixtures contained 10 µl of QUANTIMIX EASY (Biotools; #10606–4153), each forward and reverse primer at 0.5 µM, 1 µl of cDNA sample, and deionized water to a final volume of 10 µl. The following conditions were used: 95 °C for 5 min and 40 cycles at 95 °C for 15 s, followed by 1 min at 63 °C. After amplification, a melting step was performed, with a rise in temperature from 72 to 90 °C with continuous acquisition of fluorescence.

A positive and a negative (non-template control) control were added to each PCR reaction. Each sample was assessed in duplicate, and the %CV between the duplicates was < 2%. All primer sequences for the validated genes were designed to span exon-exon junctions to minimize the potential of amplifying genomic DNA (Additional file 1). Amplification efficiencies of primers were within a range of 90 to 110% efficiency, and primer specificity was assessed by the presence of a single temperature dissociation peak. The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene and the TATA-box protein (*TBP*) gene were chosen as the reference genes to estimate relative quantification. The geometric mean of the *GAPDH* and *TBP* genes was used as the reference. The relative gene expression was calculated using the $2^{-\Delta\Delta C_t}$ method with the term group as the control group [47].

Validation of RNA seq results by comparison with previous preterm transcriptomic studies

To determine if the results presented here reflect the PTB mechanisms previously reported in the literature, a PubMed search was performed using these search terms: [preterm AND (transcriptome OR transcriptomic)]. The electronic search was performed on March 19th, 2018 with no restrictions to identify all articles related to DE genes in all gestational tissues. The results were analyzed based on 5 inclusion criteria: published in English, original research, human chorioamnion tissue samples, DE between term and PTB and candidate gene list assembled. All articles were cross-checked with database search results to find any additional available transcriptome data.

Other statistical analysis and data visualization

Maternal demographics, clinical characteristics of the term and severe preterm study groups and qPCR real-time data were compared using Student's t-test for between-group comparisons of continuous data. For comparisons of categorical data, the Chi-square test was used. All statistical analyses were performed using the R environment for statistical computing version 3.4.3 [19]. After multiple test correction, a *p*-value < 0.05 was used to consider statistical

significance. The ggplot2 R package was used to create plots [48].

Results

Demographic and clinical characteristics of the term and severe preterm groups are presented in Table 1. There were significant differences between cases and controls for gestational age, birth weight and the number of prenatal medical appointments, while the remaining assessed variables did not present significant differences between groups (Table 1). We selected 8 of the 24 chorioamnion membranes to pursue RNA-Seq analysis (Additional file 2). The 8 tissues were selected and paired to match cases and controls based on the mother's demographics, the newborn's sex and socioeconomic status (see Additional file 2).

Exploration and evaluation of overall gene expression

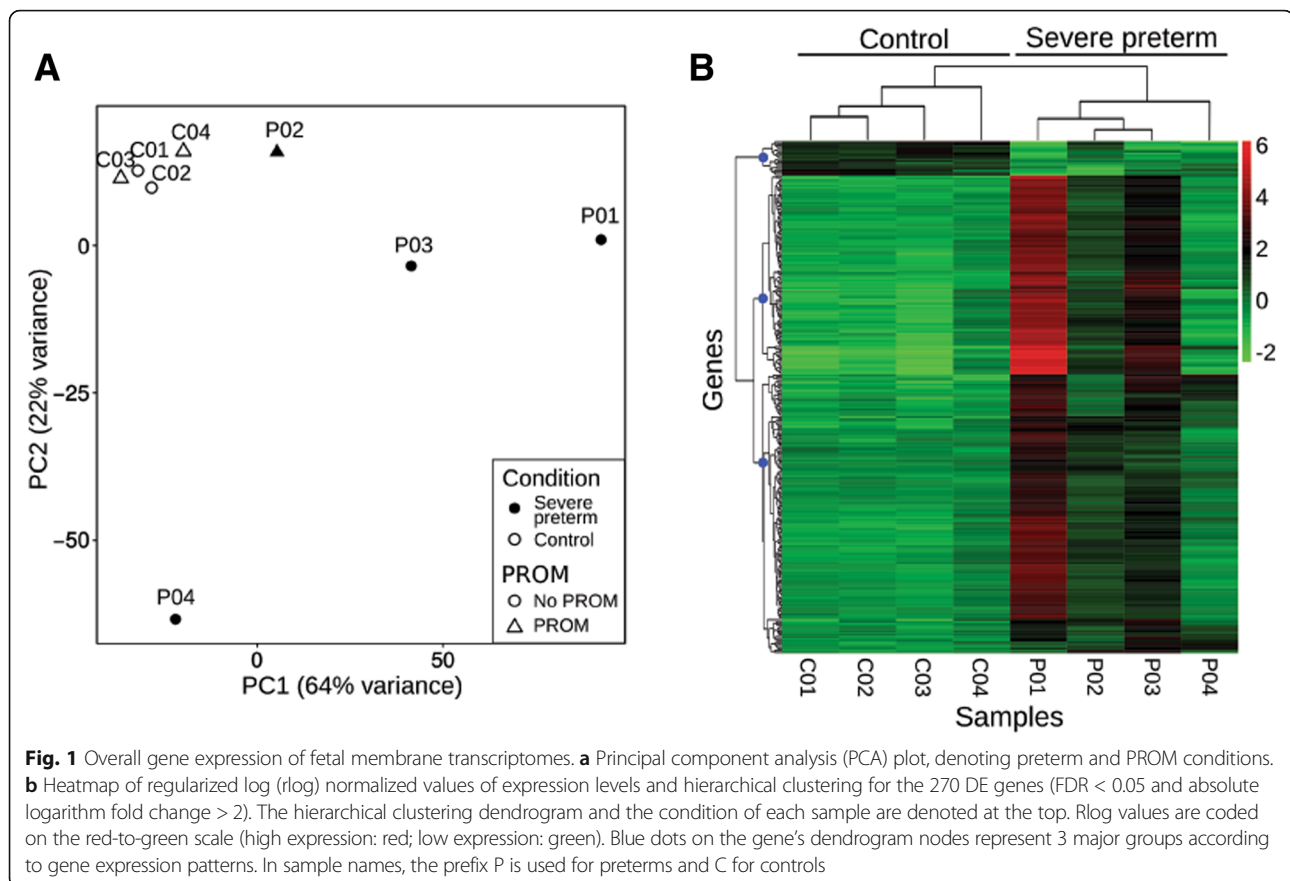
Genome-wide gene expression in chorioamnion tissue samples was evaluated using RNA-Seq technology for 4 term and 4 severe preterm samples. A summary of the sequencing read alignments to the reference genome is presented in Additional file 3. Illumina sequencing effectively produced large numbers of high-quality reads from all samples. On average, 93.7% of the total reads were mapped successfully. Among the aligned reads, 91.5% were mapped to unique genomic regions.

A PCA was performed on the gene expression rlog transformed values (Fig. 1a). A high portion of the overall

Table 1 Demographic and clinical characteristics of term and severe preterm patients

	Term birth	Severe preterm birth	<i>p</i> value ^a
	Total (<i>n</i> = 15)	Total (<i>n</i> = 9)	
Mothers age, years ± SE	24.1 ± 1.6	20.6 ± 1.0	0.146
Gestational age, weeks ± SE	39.3 ± 0.3	30.7 ± 0.9	1.9 × 10 ⁻¹⁰
Education: elementary school or less, <i>n</i> (%)	9 (60.0)	3 (33.3)	0.399
Marital status, % single mother	13.3	22.2	0.472
PROM (%)	4 (26.7)	2 (22.2)	1
Preeclampsia (%)	0	1 (11.1)	0.238
IUGR (%)	0	1 (11.1)	0.238
Parity, <i>n</i> (%)			0.335
Nulliparous	5 (33.3)	4 (44.4)	
≥2	5 (33.3)	2 (22.2)	
Birth weight (g), mean ± SE	3492 ± 120	1696 ± 205	1.6 × 10 ⁻⁰⁷
Birth sex, % female	47	55	0.472
Prenatal appointments ± SE	8.8 ± 0.8	4.3 ± 0.9	0.002

^aSignificance was assessed with Student's t-test or Chi-square test, as applicable
PROM: Preterm rupture of membranes, IUGR: Intrauterine restriction growth, SE: Standard error



variance was accounted for by the first two principal components of this model (PC1, 64%; PC2, 22%). The PCA showed term samples as a clearly packed group, while severe preterm samples showed greater variability.

A total of 270 genes were DE simultaneously according to the 3 methods employed (FDR < 0.05 and absolute logarithm fold change ≥ 2) (Additional file 4). Among the DE genes, 252 were upregulated and 18 were downregulated in severe preterm births relative to term births. The full list of DE genes detected (FDR < 0.05 and absolute log fold change > 2) can be found in Additional file 5a. Additional file 5 also includes references to previous studies that associated PTB with the DE genes found in this work (Additional file 5d). Approximately 50% of DE genes were previously reported to be associated with preterm labor.

When visualizing the rlog transformed read counts of the 270 significantly DE genes, control samples had a homogenous expression pattern (Fig. 1b). Notably, several of the DE genes were associated with immune-related functions such as interleukin 24, CXC motif chemokine ligands and tumor necrosis factor signaling pathways. A hierarchical clustering analysis clustered genes in 3 groups. Remarkably, one of the clusters was a group of 18 genes that had a notably discordant expression pattern for

term and severe preterm samples. This group consisted of the aforementioned 18 downregulated genes (Fig. 2), which potentially characterize a severe preterm transcriptome pattern and therefore are candidate genes for understanding the syndrome. These genes included potassium voltage-gated channel subfamily members, genes involved in cell adhesion functions, placental alkaline phosphatase and a Wnt family member. This gene cluster was related to GO terms involved in the nervous system, organ morphogenesis, and ion channel complexes, but no term remained as enriched with statistical significance in the group after BH *p*-value correction.

The remaining two groups consisted of 147 and 105 genes. Mainly, these genes were related to chemokines, TNF family members, coagulation factors, and transmembrane proteins. A GO term enrichment search revealed 508 and 301 terms, respectively that were statistically significant (BH corrected *p*-value < 0.05). These terms were mostly related to immune and inflammatory responses and the immune system.

We found some overlap between the 270 DE genes and genes related to labor that were previously identified by RNA microarrays [49–52]. In total, 58 of the 270 DE genes found in this study were identified as DE between labor and no labor births (Additional file 5, b).

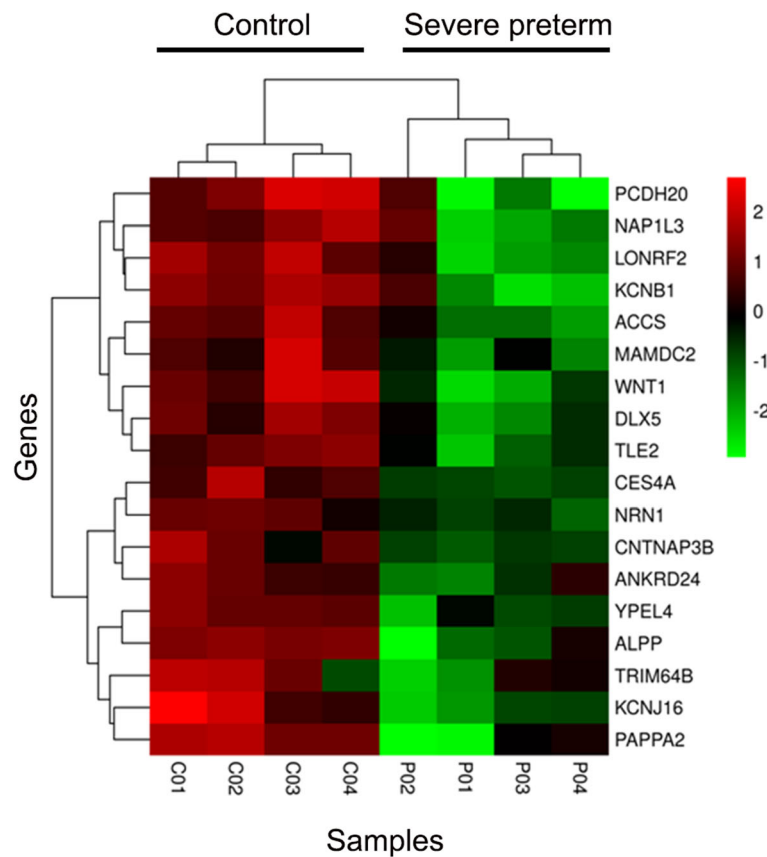


Fig. 2 Gene expression levels of the 18 downregulated genes. Heatmap of regularized log (rlog) normalized values of expression levels and hierarchical clustering for the 18 downregulated DE genes. Gene symbols are indicated to the right. Rlog values are coded on the red-to-green scale (high expression: red; low expression: green). In sample names, the suffix P is used for preterms and C for controls

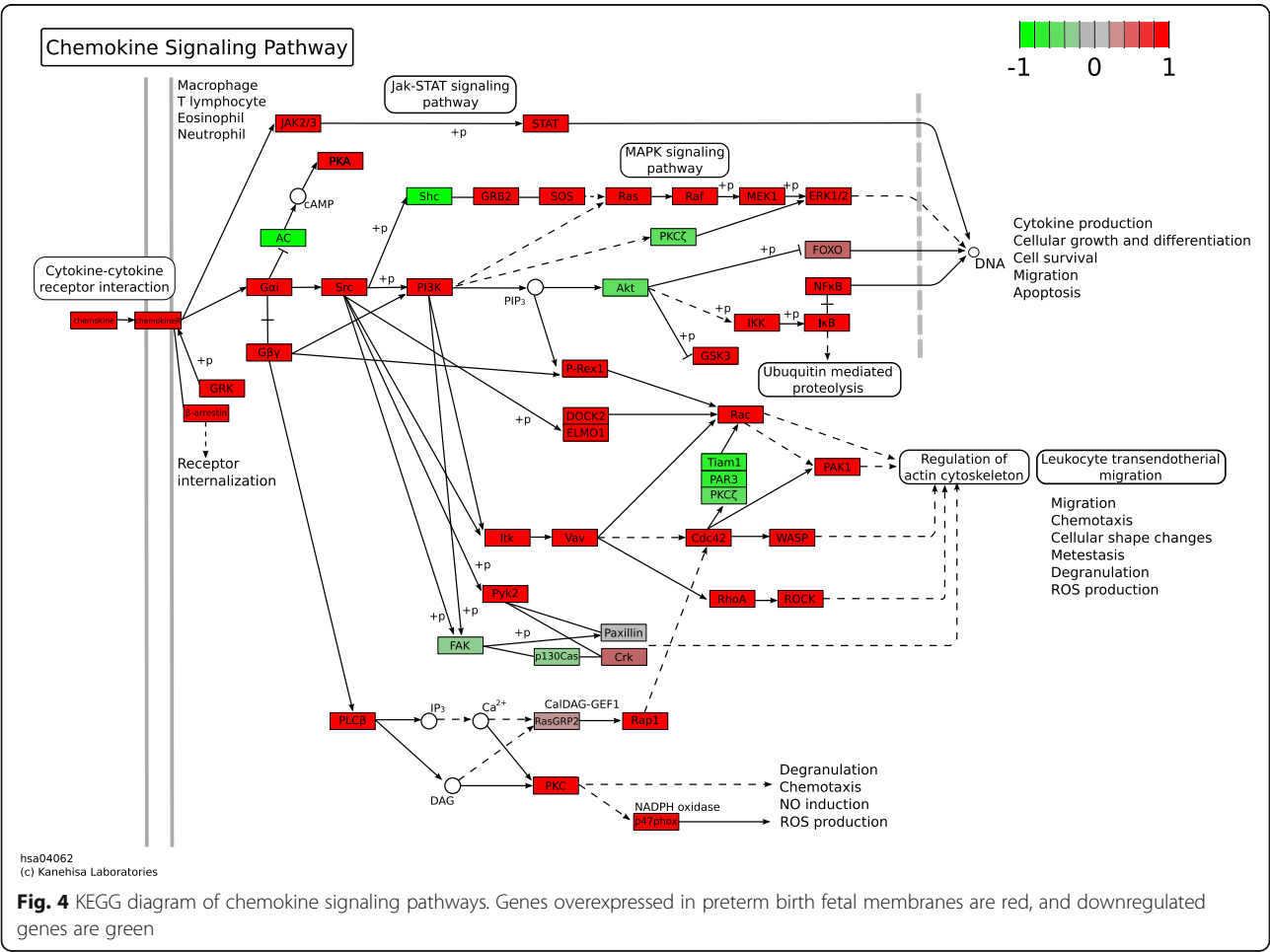
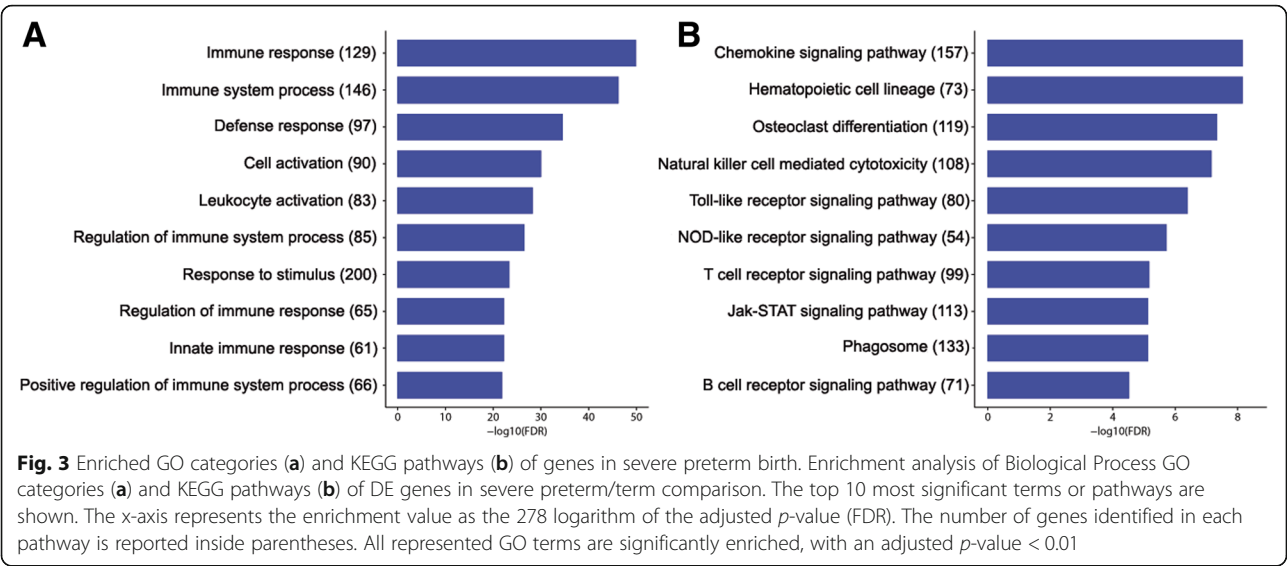
When analyzing whether PROM has an effect on gene expression, we found only 14 genes differentially expressed between groups with and without PROM (FDR < 0.05 and absolute log fold change > 2) (Additional file 5, c). Only 2 of those genes were also detected as DE with respect to preterm condition: KCNJ16 and TRPV6.

Gene set enrichment analysis using gene ontology

In an effort to identify processes and pathways that could be regulated differentially between the severe preterm and term groups, we performed a Gene Set Enrichment Analysis implemented in the Goseq R package. Genes that showed an FDR < 0.05 were tested against the background set of all expressed genes with ENSEMBL annotations. In this sense, for the enrichment analysis comparing gene expression between severe preterm and term births, 270 significant genes were tested against 18,675 background genes. The significant enrichment of GO terms was tested using the Wallenius approximation and by applying multiple hypothesis testing correction (BH). Only terms with 3 or more statistically significant genes were considered.

A total of 1886 GO terms were enriched among the DE genes (FDR < 0.05), corresponding to Biological Processes (562 terms), Cellular Component (49) and Molecular Function (35) (Fig. 3a). Within the Biological Processes category, the terms *immune response*, *immune system process* and *defense response* were the most significantly enriched for DE genes. Likewise, the Cellular Component category subdivided annotated sequences into *plasma membrane*, *cell periphery*, and *cytoplasmic vesicle part*, which were the most highly represented terms. Within the category Molecular Function, the three principal groups were *signaling receptor activity*, *receptor activity*, and *transducer activity*.

When analyzing KEGG pathways, the most significant upregulated pathways were Chemokine signaling pathway and Hematopoietic cell lineage (both with FDR = 7.038×10^{-09} ; Fig. 4). Several relevant processes for PTB were upregulated, such as those involved in intracellular signaling: Toll-like, NOD-like, T cell and Jak-STAT signaling pathway (Fig. 3b). No statistically significant downregulated pathways were found.



Validation of gene expression by qRT-PCR

To validate genes that were significant in the RNA-Seq analysis, six DE genes were selected—*IL1B*, *LCN2*, *MARCO*, *CASP5*, *SERPINA1* and *TNFSF15*—and their expression was assessed by qPCR. The RNA-Seq analysis showed that all genes were significantly upregulated in PTB. Additional file 6 displays the logarithm fold differences in gene expression measured by both RNA-Seq and qRT-PCR. The six genes displayed similar patterns of mRNA abundance with both methods. Five genes were significantly upregulated between severe preterm and controls when analyzing the qPCR expression data: *LCN2*, *MARCO* and *TNFSF15* (p value < 0.001) and *CASP5* and *IL1B* (p value < 0.05); *SERPINA1* was also upregulated, although the differences between groups did not reach statistical significance (Additional file 6).

Validation of RNA seq results by comparison with previous preterm transcriptomic studies

The results presented in this study were compared with those previously reported in the literature. A PubMed search using the terms [preterm AND (transcriptome OR transcriptomic)] yielded 101 results. We collected all 101 abstracts and, after analyzing them, found that most studies focused on preeclampsia, placental villi tissue, and animal models or maternal or cord blood. Six studies used RNA-Seq: two focused on cord blood [53, 54], two on animal models [55, 56] and two on the transcriptome of normal labor [14, 57].

Only one study met all inclusion criteria mentioned above and included the transcriptomes of the chorion and amnion membranes; this study was included for comparison with the results obtained in this work [43]. The study used microarray technology and found 50 genes that identified the preterm labor phenotype, seven of which were DE in the current study (Table 2). The direction of gene expression changes matched between studies in all cases [43].

Discussion

Although PTB is responsible for most neonatal deaths and morbidity [58], there has been a shortage of studies

that permit the creation of a distinctive expression footprint to describe its pathology and to consequently achieve improvements in its prevention, diagnosis, and treatment. To overcome this gap in knowledge, we analyzed the gene expression of the chorioamnion membranes from severe PTB and term newborns by RNA-seq.

In this study, we present a distinct transcriptome pattern of severe PTB. Our study revealed 270 DE genes between both conditions. A number of this magnitude may partially explain the lack of previous molecular signatures for PTB as well as the difficulty in identifying biomarkers and potential therapeutic interventions [4, 39]. Several pathways and genes are involved in labor, and most of them are either turned on or off during pathological conditions [4, 16, 37, 40].

It has been proposed that pregnancy is maintained by the downregulation of chemokines at the maternal-fetal interface [43]. Removal of this downregulation results in term birth, while in the case of PTB, the activation of multiple pathways of the immune system overrides this downregulation [3, 43, 55, 59–62]. In concordance with this idea, we found that upregulated genes were mostly related to the immune response (e.g., *IL1B*, interleukin 24, CXC motif chemokine ligands or *TNF*). Moreover, we found that severe preterm labor was significantly enriched for immune processes in comparison to term labor. Gene ontology and KEGG pathway enrichment analyses indicated that inflammatory and immune pathways were enriched in severe preterm samples (Fig. 3). These findings are in agreement with previously mentioned studies, demonstrating that preterm labor is associated with extensive immune response activation in several gestational tissues. In fact, the incidence of PTB in association with inflammation is higher earlier in pregnancy [63], so the overexpression of inflammatory and immune pathways in severe PTB is consistent with the results reported in the literature.

IL1B was upregulated in severe preterm births in our data. *IL1B* is a central mediator in the pathological process of inflammation since it can stimulate the expression and release of other labor mediators [21, 23, 64–66]. Recent evidence indicates that *IL1B* may be involved in labor through activation of the inflammasome [22]. The inflammasome is

Table 2 DE genes simultaneously found in Bukowski et al. [43] and this study

Gene Symbol	Gene	HGNC ID	Change in direction
BCL2A1	BCL2 related protein A1	HGNC:991	upregulated in PTB
CCRL2	C-C motif chemokine receptor-like 2	HGNC:1612	upregulated in PTB
CHI3L1	chitinase 3 like 1	HGNC:1932	upregulated in PTB
CXCL2	C-X-C motif chemokine ligand 2	HGNC:4603	upregulated in PTB
GPR84	G protein-coupled receptor 84	HGNC:4535	upregulated in PTB
SAMSN1	SAM domain, SH3 domain and nuclear localization signals 1	HGNC:10528	upregulated in PTB
SLC16A10	solute carrier family 16 member 10	HGNC:17027	upregulated in PTB

a multi-protein complex located in the cytoplasm of cells whose activation induces the transformation of pro-IL1B (immature form) into the active forms of IL1B via the actions of active caspase-1 (CASP1) [67–70]. NLR family pyrin domain-containing protein 3 (NLRP3) is an inflammasome recruiting pattern recognition receptor [68, 71]. The amniotic fluid concentrations of some inflammasome mediators (CASP1, IL1B, NLRP3, and IL18) are greater in women who have spontaneous preterm or term labor with intraamniotic infection/inflammation than in those without this clinical condition [22, 72–77]. We found that some inflammasome genes, including CARD domain containing 4 (*NLRC4*), caspase 5 (*CASP5*), absent in melanoma 2 (*AIM2*) and the already mentioned *NLRP3*, were upregulated in our data. Moreover, the inflammasome complex GO term (GO:0061702) was significantly enriched in this group. Taking all data together, inflammasome activation may be implicated in the induction of preterm labor. If the hypothesis that inflammasome activation contributes to PTB is true, it may have medical implications. Components of the inflammasome may be the targets of drugs that could potentially abort PTB once it is clinically suspected and before the full immunological response begins.

As can be observed in Additional file 5, several genes identified as DE in this work were previously associated with PTB. Interestingly, two of the overexpressed genes found in this study were previously proposed as biomarkers of PTB because their concentration in amniotic fluid was higher in PTB than in term labor: C-X-C motif chemokine ligand 10 (*CXCL10*) gene and matrix metalloproteinase 8 (*MMP8*) [46, 78–80]. Consequently, these biomarkers can be analyzed in amniotic fluid and provide information to prevent PTB. *CXCL10* is related to maternal anti-fetal rejection in the same way that it is related to allograft rejection [46]. In this context, *CXCL10* overexpression in PTB may lead to premature labor by triggering a fetal inflammatory response. *MMP8* is known to decrease the integrity of the cervical and fetal membrane extracellular matrix [81]. *MMP-8* may alter the membrane extracellular matrix and permit the passage of bacteria into the endometrium, leading to bacteria-induced PTB [82].

A small group of 18 genes clustered together discriminated between term and severe preterm cases (Fig. 2). All of the genes were downregulated. These genes potentially characterize a severe preterm transcriptome pattern and therefore are candidate genes for understanding the syndrome. None of these downregulated genes were previously proposed in the literature as preterm birth biomarkers. However, the placental alkaline phosphatase (*ALPP*) and pappalysin 1 (*PAPPA*) genes were included in a group of nine placental genes that together could predict GA in a recent pilot study [45]. *ALPP* and Pappalysin 2 (*PAPPA2*), a closely related gene to *PAPPA*, were among the 18 downregulated genes found here. Additionally,

PAPPA2, a type of metalloproteinase with typically high gene expression in the placenta [83], was found to be DE in placenta tissue between spontaneous PTB and term birth [84].

The low number of downregulated genes made it difficult to establish statistically significant pathways: only two of the 18 clustered genes were included in the same KEGG pathway. The Wnt family member 1 (*WNT1*) and distal-less homeobox 5 (*DLX5*) genes belong to the pathway: “Signaling pathways regulating pluripotency of stem cells” (hsa04550). In term births, expression of this pathway increases at the end of pregnancy [85]. The *WNT1* gene belongs to the WNT signaling pathway, which is involved in cell proliferation, development, and tumor progression. WNT signaling pathway inhibition could contribute to miscarriages and PTB by decreasing trophoblast proliferation and invasion [86]. Notably, the *WNT2* gene, which is typically overexpressed in placenta, was not DE in our data. *DLX5* is a member of a homeobox transcription factor gene family. Homeobox genes regulate embryonic as well as placental development [87]. They are widely expressed in the human placenta. A potential role for *DLX5* may be related to neurobehavioral development and to the regulation of stem cell function, both of which are important for normal placental development [87]. Additionally, *DLX5* is upregulated in preeclamptic placentas, presumably due to altered methylation at the *DLX5* locus in preeclampsia that results in loss of imprinting [88]. A decrease in the expression of this gene later in gestation was found in human placenta as a response to an increase in gene body methylation over gestational age [87]. Our results indicate that *WNT1* and *DLX5* expression are lower in severe PTB fetal membranes than in term membranes. In this context, we speculate that downregulated genes may act as maturation genes that set the gestation timing.

These putative maturation genes may be very difficult to discriminate from those leading to preterm birth. A major issue in the study of PTB transcriptomes in humans is the inability to collect healthy control tissue at the same gestational age (GA) to compare with pathologic preterm tissue. Consequently, DE genes between PTB and term fetuses reflect differences in both GA and the pathological events of preterm labor. We are aware that, as a consequence of the design of our study, genes responsible for PTB and GA overlap. By comparing human sPTB and term transcriptomes with GA-matched control transcriptomes from a closely related species (macaques), Eidem et al. [89] identified a low number of distinctive, promising sPTB-specific candidate genes or genes potentially related to GA effects. The authors found similar functions for PTB and GA genes, so they speculated that the effects of sPTB and GA do not correspond to biologically distinct processes [89].

In an attempt to narrow the gene signature of PTB, DE genes were compared with labor-related genes described in previous works [49–52]. Fifty-eight of the DE genes found in this study were identified as labor-related genes, making them candidates for labor independent of GA. In the remaining group, it is probable that PTB genes overlap with those related to GA. However, whether PTB is labor that occurs too soon or represents a pathological event with a new group of expressed genes is still an open question.

The present transcriptome of chorioamnion membranes does not resolve the controversy, but it adds information that will help to complete the PTB expression scenario. On one hand, our results confirmed that immune response pathways are upregulated in PTB. On the other hand, this RNA-Seq approach allowed the identification of upregulated genes and pathways that were not previously reported to be associated with PTB (hematopoietic cell lineage pathway, osteoclast differentiation). Additionally, we found that some of the downregulated genes are involved in the nervous system, morphogenesis (*WNT1*, *DLX5*, *PAPPA2*) and ion channel complexes (potassium voltage-gated channel subfamily J member 16 (*KCNJ16*), potassium voltage-gated channel subfamily B member 1 (*KCNB1*)). It is difficult to speculate on the importance of these genes and pathways in PTB; for example, ion channels are involved in a wide spectrum of different functions that eventually lead to labor or PTB, including implantation [90], Ca^{++} signaling [91], smooth muscle contraction [92] and inflammation [93]. Regarding the possibility that the hematopoietic cell lineage pathway is connected to PTB, it is known that the placenta produces hematopoietic stem cells (HSCs) [94, 95]. The placenta can generate HSCs and cause their expansion but not their differentiation. The placental HSC pool first appears early in gestation and then decreases while the HSC reservoir develops in the liver [95, 96]. Subsequent studies demonstrated that mid-gestation placenta also harbors a large pool of pluripotent hematopoietic stem cells that have self-renewal capacity [94, 97]. Moreover, hematopoietic progenitors were found in full-term placentas within a low percentage of lineage-committed cells [98]. Since the hematopoietic monocyte-macrophage lineage produces osteoclasts, the hematopoietic cell lineage and osteoclast differentiation pathways are connected. Although both pathways appear far removed from term labor or PTB a priori, they may represent interesting unexplored paths to preterm labor. Nevertheless, all of the mentioned genes and biological processes open new lines of research, making them good candidates as biomarkers of PTB.

A literature search to validate these genes revealed that, surprisingly, there are few high-throughput studies analyzing chorioamniotic tissues and none using RNA-Seq technology. To our knowledge, this is the first RNA-Seq study of chorioamniotic tissues in preterm and term births.

When comparing the results from Bukowski et al. [43], only 7 genes are shared with our study. All of these genes may be related to PTB in one way or another: C-C motif chemokine receptor-like 2 (*CCRL2*) and C-X-C motif chemokine ligand 2 (*CXCL2*) belong to inflammatory pathways [99, 100], and G protein-coupled receptor 84 (*GPR84*) is a receptor of fatty acids that is also related to inflammation [101]. BCL2 related protein A1 (*BCL2A1*) and chitinase 3 like 1 (*CHI3L1*) were recently associated with chorioamnionitis in monkeys and humans [102]; solute carrier family 16 member 10 (*SLC16A10*) is upregulated in the mid-gestational placenta [103]; and SAM domain, SH3 domain and nuclear localization signals 1 (*SAMSNI*) is a poorly characterized gene that may be involved in the differentiation of lymphocytes B [104]. The fact that these subsets of genes overlap in both studies, which used different experimental techniques, stresses that they should be targets of further research into PTB.

One interesting finding is the homogeneity for gene expression in the control samples. After a linear reduction of the variability for more than 30,000 expressed genes, the control samples cluster together in a small group. The dispersion of the samples in the PCA analysis is lower compared to other tissue expression profiles using the same exploratory analysis [105]. In contrast, PTB chorioamniotic expression analysis revealed a more dispersed group, further stressing the heterogeneity of this condition. This may be a consequence of the complexity of the syndrome characterization, as it has been previously reported in the expression of other tissues related to PTB like the myometrium [106]. In concordance, Ngo et al. [45] developed a gene expression model based on GA that predicted time until delivery for full-term but that failed for preterm deliveries. This suggests that to predict PTB the classifiers must incorporate at least part of the several outlier physiological events that may lead to PTB [45], e.g. PROM.

Premature and preterm rupture of membranes are frequent complications of birth and, more importantly, preterm birth [107]. The transcriptome profile of the chorioamnion may differ with and without PROM. To overcome this issue, we performed statistical analyses including PROM as a co-variable.

Moreover, when analyzing the differential gene expression of PROM, we did not find substantial differences between samples from newborns with and without PROM. We only detected 14 genes with DE among groups. Additionally, the PCA did not show any clustering based on this condition (Fig. 1). This low number of DE genes related to PROM is likely the consequence of our experimental design, which focused on sampling groups discriminated by the severe preterm condition while matching for the remaining variables. Additionally, when analyzing gene expression between

PROM and no PROM groups, the group that presents rupture of membranes is limited in the number of samples, meaning that the results must be confirmed with a larger sample size.

Finally, this study was performed in an admixed population where PTB is frequent (approx. 10%), indicating that it is a serious health public problem. Population-specific studies are relevant, as genetic background may differently impact different populations, e.g., it is known that African American ancestry increases the risk of PTB [108–110]. Generally, there is an information gap regarding specific populations, except those coming from North America and Europe. With our literature search, we recovered only three out of 101 transcriptomic analyses based on populations from Latin American countries [46, 111, 112], stressing the fact that most studies do not include susceptible populations from this part of the continent.

All of the Latin American studies were performed with microarray analysis, meaning that the data shown in this work are the first RNA-Seq data available to contrast against future studies. Specifically, the Uruguayan population is admixed, with a predominant European contribution and Native American and African contributions of 10.4 and 5.6%, respectively [113, 114]. Common genetic variation affects expression variability [115], as well as the admixed genome structure [116, 117], stressing the necessity of population-specific transcriptome analyses or, alternatively, including diversity in high-throughput population analyses.

Conclusions

The use of state-of-the-art RNA-Seq analytics identified previously reported genes and pathways as well as novel candidates involved in labor and PTB. The combination of up- and downregulated genes permitted the development of a multi-gene disease classifier. These markers were generated in an admixed population in which PTB has a high incidence. Further studies including the testing of these classifiers using a second and larger independent sample are needed to validate the results. Moreover, these markers should be compared to other population studies to obtain a global landscape of severe PTB.

Additional files

Additional file 1: Table S1. Primers used for quantitative PCR analysis. Sequences of the primers used for qRT-PCR. (XLSX 4 kb)

Additional file 2: Table S2. Clinical and demographics characteristics of 4 controls and 4 patients analyzed by RNA-seq. Controls are indicated as C01, C02, C03 and C04 and severe preterms as P01, P02, P03 and P04. (XLSX 12 kb)

Additional file 3: Table S3. Alignment statistics obtained by tophat2. Number of sequenced reads, total mapped reads and uniquely mapped reads for each of the 8 sequenced samples. (XLSX 4 kb)

Additional file 4: Figure S1. Differentially expressed genes. Venn diagram showing the number of differentially expressed genes identified

by each of the three methods employed, when considering FDR < 0.05. (TIF 906 kb)

Additional file 5: Table S4. a. Full list of differentially expressed genes (FDR < 0.05 and absolute logarithm log fold change > 2) detected in the severe preterm vs. term pairwise comparison. b. Subset of DE genes that were also DE between labor and no labor conditions in the literature. c. Full list of differentially expressed genes (FDR < 0.05 and absolute logarithm fold change > 2) detected in the PROM vs. no PROM comparison. d. List of severe preterm DE genes already identified as associated with PTB in available databases. References are indicated in the manuscript. (XLSX 60 kb)

Additional file 6: Figure S2. Validation of overall gene expression. Log2-fold changes of six DE genes measured by RNA-Seq (red) vs. qRT-PCR (blue). *IL1B*: interleukin 1 beta, *LCN2*: lipocalin 2, *MARCO*: macrophage receptor with collagenous structure, *CASP5*: caspase 5, *SERPINA1*: serpin family A member 1, and *TNFSF15*: TNF superfamily member 15. **p* < 0.05; ***p* < 0.01; ****p* < 0.001. (EPS 1243 kb)

Abbreviations

AIM2: Absent in melanoma 2; ALPP: placental alkaline phosphatase; BCL2A1: BCL2 related protein A1; BH: Benjamini and Hochberg's approach; CASP5: Caspase 5; CCRL2: C-C motif chemokine receptor-like 2; CHI3L1: Chitinase 3 like 1; CLAP: Latin American Center for Perinatology; CXCL10: C-X-C motif chemokine ligand 10; CXCL2: C-X-C motif chemokine ligand 2; DE: Differentially expressed; DLX5: Distal-less homeobox 5; FDR: False discovery rate; GA: Gestational age; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; GO: Gene Ontology; GPR84: G protein-coupled receptor 84; HSCs: Hematopoietic stem cells; IL1B: Interleukin 1 beta; IUGR: Intrauterine growth restriction; KCNB1: Potassium voltage-gated channel subfamily B member 1; KCNJ16: Potassium voltage-gated channel subfamily J member 16; LCN2: Lipocalin 2; MARCO: Macrophage receptor with collagenous structure; MMP-8: Matrix metalloproteinase 8; mRNA: Messenger-RNA; NLR4: NLR Family, caspase recruitment domain-containing 4; NLRP3: NLR Family, pyrin domain-containing 3; PAPP: Pappalysin 1; PAPP A2: Pappalysin 2; PCA: Principal Components Analysis; PROM: Premature rupture of membranes; PTB: Preterm birth; qRT-PCR: Quantitative real-time PCR; rlog: Regularized logarithm; RNA-Seq: RNA sequencing; SAMS1: SAM domain, SH3 domain and nuclear localization signals 1; SERPINA1: Serpin family A member 1; SLC16A10: Solute carrier family 16 member 10; sPTB: Spontaneous PTB; TBP: TATA-box protein; TNFSF15: TNF superfamily member 15; TRPV6: Transient receptor potential cation channel subfamily V member 6; WNT1: Wnt family member 1

Acknowledgements

We acknowledge Drs. Ariel Díaz, Natalia Cabrera and Pamela Grimaldi (Pereira Rossell Hospital Center) for their invaluable support in patient and control subject recruitment.

Funding

SP, BB and RS are supported by Sistema Nacional de Investigadores, Agencia Nacional de Investigación e Innovación (ANII, Uruguay). This work was funded by Comisión Sectorial de Investigación Científica (CSIC) and Programa de Desarrollo de las Ciencias Básicas- PEDECIBA (Uruguay) as full project. The funding bodies provided financial support but had no other role in the design of the study, data collection, analysis, and interpretation of data, decision to publish, or preparation of the manuscript.

Availability of data and materials

The data generated and/or analyzed during the current study are available in the Sequence Read Archive (NCBI) repository and are accessible through SRA Series accession number SRP139931. Data are publicly available in: <https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA448842>

Authors' contributions

SP, BB and RS conceived and designed the experiments. CS assisted in sample collection. SP performed the experiments and analyzed the data. All authors discussed results. SP, BB and RS wrote the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The School of Medicine's Ethics Committee of the University of the Republic (Consejo de la Facultad de Medicina, Exp. N° 071140-000907-11) and the Pereira Rossell Hospital Center Ethics Committee (Uruguay) both approved the study protocol. Written informed consent was obtained from mothers prior to collection of biological material.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Departamento de Genética, Facultad de Medicina, Universidad de la República, Av. General Flores 2125, C.P. 11800 Montevideo, Uruguay. ²Clínica Ginecologica "C", Centro Hospitalario Pereira Rossell, Facultad de Medicina, Universidad de la República, Bvar. General Artigas 1590, C.P.11600, Montevideo, Uruguay. ³Departamento de Histología y Embriología, Facultad de Medicina, Universidad de la República, Av. General Flores 2125, C.P. 11800 Montevideo, Uruguay.

Received: 5 September 2018 Accepted: 10 March 2019

Published online: 01 April 2019

References

- Behrman RE, Stith BA. Institute of Medicine Committee on understanding premature birth and assuring healthy outcomes board on health sciences outcomes: preterm birth: causes, consequences, and prevention. Washington, DC: Preterm birth: causes, consequences, and prevention, National Academies Press; 2007.
- Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet*. 2012;379:2151–61.
- Romero R, Dey SK, Fisher SJ. Preterm Labor: One Syndrome, Many Causes. *Science*. 2014;345:760–5.
- Myatt L, Eschenbach DA, Lye SJ, Mesiano S, Murtha AP, Williams SM, et al. A Standardized Template for Clinical Studies in Preterm Birth. *Reprod Sci*. 2012;19:474–82.
- Eidem HR, Ackerman WE, McGary KL, Abbot P, Rokas A. Gestational tissue transcriptomics in term and preterm human pregnancies: a systematic review and meta-analysis. *BMC Med Genomics*. 2015;8:27.
- Vogel JP, Chawanpaiboon S, Moller A-B, Watananirun K, Bonet M, Lumbiganon P. The global epidemiology of preterm birth. *Best Pract Res Clin Obstet Gynaecol*. 2018. <https://doi.org/10.1016/j.bpobgyn.2018.04.003>.
- Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet*. 2008;371. [https://doi.org/10.1016/S0140-6736\(08\)60074-4](https://doi.org/10.1016/S0140-6736(08)60074-4).
- Moutquin J. Classification and heterogeneity of preterm birth. *BJOG*. 2003; 110:30–3.
- Henderson JJ, McWilliam OA, Newnham JP, Pennell CE. Preterm birth aetiology 2004–2008. Maternal factors associated with three phenotypes: spontaneous preterm labour, preterm pre-labour rupture of membranes and medically indicated preterm birth. *J Matern Fetal Neonatal Med*. 2012; 25:642–7.
- Iams JD. Preterm birth categories—labels with consequences. *Am J Obstet Gynecol*. 2017;210:97–8.
- Monangi NK, Brockway HM, House M, Zhang G, Muglia LJ. The genetics of preterm birth: Progress and promise. *Semin Perinatol*. 2015;39:574–83.
- Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc*. 2015;2015:951–69.
- Söber S, Reiman M, Kikas T, Rull K, Inno R, Vaas P, et al. Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes. *Sci Rep*. 2015;5:13336.
- Kim J, Zhao K, Jiang P, Lu Z-X, Wang J, Murray JC, et al. Transcriptome landscape of the human placenta. *BMC Genomics*. 2012;13:115.
- Mikheev AM, Nabekura T, Kaddoumi A, Bammler TK, Govindarajan R, Hebert MF, et al. Profiling gene expression in human placenta of different gestational ages: an OPRU Network and UW SCOR Study. *Reprod Sci*. 2008; 15:866–77.
- Buckberry S, Bianco-Miotto T, Bent SJ, Dekker GA, Roberts CT. Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal-maternal interface. *Mol Hum Reprod*. 2014;20:810–9.
- Rey G, Skowronek F, Alciaturi J, Alonso J, Bertoni B, Sapiro R. Toll receptor 4 Asp299Gly polymorphism and its association with preterm birth and premature rupture of membranes in a South American population. *Mol Hum Reprod*. 2008;14:555–9. <https://doi.org/10.1093/molehr/gan049>.
- Pereyra S, Velazquez T, Bertoni B, Sapiro R. Rapid multiplex high resolution melting method to analyze inflammatory related SNPs in preterm birth. *BMC Res Notes*. 2012;5:69.
- Rey G, Pereyra S, Velazquez T, Grasso D, Alonso J, Bertoni B, et al. The effect of inflammation on preterm birth. In: *Preterm Birth-Mother and Child*. InTech; 2012.
- Pereyra S, Bertoni B, Sapiro R. Interactions between environmental factors and maternal–fetal genetic variations: strategies to elucidate risks of preterm birth. *Eur J Obstet Gynecol Reprod Biol*. 2016;202:20–5.
- Romero R, Grivel J-C, Tarca AL, Chaemsaitong P, Xu Z, Fitzgerald W, et al. Evidence of perturbations of the cytokine network in preterm labor. *Am J Obstet Gynecol*. 2015;213:836.e1–836.e18.
- Strauss JF, Romero R, Gomez-Lopez N, Haymond-Thornburg H, Modi BP, Teves ME, et al. Spontaneous preterm birth: advances toward the discovery of genetic predisposition. *Am J Obstet Gynecol*. 2018;218:294–314.e2.
- Romero R, Espinoza J, Gonçalves LF, Kusanovic JP, Friel L, Hassan S. The role of inflammation and infection in preterm birth. *Semin Reprod Med*. 2007;25:21–39.
- Nhan-Chang C-L, Romero R, Tarca AL, Mittal P, Kusanovic JP, Erez O, et al. Characterization of the transcriptome of chorioamniotic membranes at the site of rupture in spontaneous labor at term. *Am J Obstet Gynecol*. 2010; 202:462.e1–41.
- Challis JR, Lockwood CJ, Myatt L, Norman JE, Strauss JF, Petraglia F. Inflammation and pregnancy. In: *Reproductive Sciences*. 2009. p. 206–15.
- Vrachnis N, Vitoratos N, Iliodromiti Z, Sifakis S, Deligeorgoglou E, Creasas G. Intrauterine inflammation and preterm delivery. *Ann N Y Acad Sci*. 2010; 1205:118–22.
- De Mucio B, Abalos E, Cuesta C, Carroli G, Serruya S, Giordano D, et al. Maternal near miss and predictive ability of potentially life-threatening conditions at selected maternity hospitals in Latin America. *Reprod Health*. 2016;13:1–10.
- Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Evan Johnson W, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet*. 2010;11:733–9.
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.
- Andrews S. FastQC: A quality control tool for high throughput sequence data. <http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/>. 2010. doi: citeulike-article-id:11583827.
- Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*. 2013;14. <https://doi.org/10.1186/gb-2013-14-4-r36>.
- Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. *Database*. 2016;2016. <https://doi.org/10.1093/database/baw093>.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15. <https://doi.org/10.1186/s13059-014-0550-8>.
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11. <https://doi.org/10.1186/gb-2010-11-10-r106>.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
- Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol*. 2013;31:46–53.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57:289–300. <https://doi.org/10.2307/2346101>.
- Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ. GAGE: Generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics*. 2009;10. <https://doi.org/10.1186/1471-2105-10-161>.

39. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 2010;11. <https://doi.org/10.1186/gb-2010-11-2-r14>.
40. Durinck S, Spellman PT, Birney E, Huber W. Mapping Identifiers for the Integration of Genomic Datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009;4:1184–91.
41. Uzun A, Laliberte A, Parker J, Andrew C, Winterrowd E, Sharma S, et al. dbPTB: a database for preterm birth. *Database.* 2012;2012:bar069.
42. Vora B, Wang A, Kostli I, Huang H, Paranjpe I, Woodruff TJ, et al. Meta-Analysis of Maternal and Fetal Transcriptomic Data Elucidates the Role of Adaptive and Innate Immunity in Preterm Birth. *Front Immunol.* 2018;9:993.
43. Bukowski R, Sadovsky Y, Goodarzi H, Zhang H, Biggio JR, Varner M, et al. Onset of human preterm and term birth is related to unique inflammatory transcriptome profiles at the maternal fetal interface. *PeerJ.* 2017;5:e3685.
44. Heng YJ, Pennell CE, Chua HN, Perkins JE, Lye SJ. Whole blood gene expression profile associated with spontaneous preterm birth in women with threatened preterm labor. *PLoS One.* 2014;9:e96901.
45. Ngo TTM, Moufarrej MN, Rasmussen M-LH, Camunas-Soler J, Pan W, Okamoto J, et al. Noninvasive blood tests for fetal development predict gestational age and preterm delivery. *Science.* 2018;360:1133–6.
46. Lee J, Romero R, Chaiworapongsa T, Dong Z, Tarca AL, Xu Y, et al. Characterization of the fetal blood transcriptome and proteome in maternal anti-fetal rejection: evidence of a distinct and novel type of human fetal systemic inflammatory response. *Am J Reprod Immunol.* 2013;70:265–84.
47. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta CT$ method. *Methods.* 2001;25:402–8.
48. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd edn. New York: Springer; 2016. https://doi.org/10.1007/978-3-319-24277-4_5.
49. Haddad R, Tromp G, Kuivaniemi H, Chaiworapongsa T, Kim YM, Mazor M, et al. Human spontaneous labor without histologic chorioamnionitis is characterized by an acute inflammation gene expression signature. *Am J Obstet Gynecol.* 2006;195:394.e1–24.
50. Stephen GL, Lui S, Hamilton SA, Tower CL, Harris LK, Stevens A, et al. Transcriptomic Profiling of Human Choriondecidua During Term Labor: Inflammation as a Key Driver of Labor. *Am J Reprod Immunol.* 2015. <https://doi.org/10.1111/aji.12328>.
51. El-Azzamy H, Balogh A, Romero R, Xu Y, LaJeunesse C, Plazyo O, et al. Characteristic Changes in Decidual Gene Expression Signature in Spontaneous Term Parturition. *J Pathol Transl Med.* 2017;51:264–83.
52. Hamilton SA, Tower CL, Jones RL. Identification of chemokines associated with the recruitment of decidual leukocytes in human labour: potential novel targets for preterm labour. *PLoS One.* 2013;8:e56946.
53. de Jong E, Hancock DG, Wells C, Richmond P, Simmer K, Burgner D, et al. Exposure to chorioamnionitis alters the monocyte transcriptional response to the neonatal pathogen *Staphylococcus epidermidis*. *Immunol Cell Biol.* 2018. <https://doi.org/10.1111/imcb.12037>.
54. Chim S, Wong K, Chung C, Lam S, Kwok J, Lai C-Y, et al. Systematic Selection of Reference Genes for the Normalization of Circulating RNA Transcripts in Pregnant Women Based on RNA-Seq Data. *Int J Mol Sci.* 2017;18:1709.
55. Migale R, MacIntyre DA, Cacciatore S, Lee YS, Hagberg H, Herbert BR, et al. Modeling hormonal and inflammatory contributions to preterm and term labor using uterine temporal transcriptomics. *BMC Med.* 2016;14:86.
56. Willcockson AR, Nandu T, Liu C-L, Nallasamy S, Kraus WL, Mahendroo M. Transcriptome signature identifies distinct cervical pathways induced in lipopolysaccharide-mediated preterm birth. *Biol Reprod.* 2018;98:408–21.
57. Saben J, Zhong Y, McKelvey S, Dajani NK, Andres A, Badger TM, et al. A comprehensive analysis of the human placenta transcriptome. *Placenta.* 2014;35:125–31.
58. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller A-B, Narwal R, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet.* 2012;379:2162–72.
59. Challis JR, Lockwood CJ, Myatt L, Norman JE, Strauss JF 3rd, Petraglia F. Inflammation and pregnancy. *Reprod Sci.* 2009;16:206–15.
60. Behrman RE, Butler AS. *Preterm birth: Causes, consequences, and prevention.* Washington DC: National Academies Press; 2007.
61. Vitoratos N, Papadias K, Makrakis E, Christodoulakos G, Panoulis K, Creatas G. Association between serum tumor necrosis factor- α and corticotropin-releasing hormone levels in women with preterm labor. *J Obstet Gynaecol Res.* 2006;32:497–501.
62. Flood K, Malone FD. Prevention of preterm birth. *Semin Fetal Neonatal Med.* 2012;17:58–63.
63. Korebrits C, Ramirez MM, Watson L, Brinkman E, Bocking AD, Challis JR. Maternal corticotropin-releasing hormone is increased with impending preterm birth. *J Clin Endocrinol Metab.* 1998;83:1585–91.
64. Watari M, Watari H, DiSanto ME, Chacko S, Shi G-P, Strauss JF III. Pro-Inflammatory Cytokines Induce Expression of Matrix-Metabolizing Enzymes in Human Cervical Smooth Muscle Cells. *Am J Pathol.* 1999;154:1755–62.
65. Heng YJ, Liong S, Permezel M, Rice GE, Di Quinzio MKW, Georgiou HM. The interplay of the interleukin 1 system in pregnancy and labor. *Reprod Sci.* 2014;21:122–30.
66. Watari M, Watari H, Strauss J. Lipopolysaccharide and pro-inflammatory cytokines induce expression of matrix metabolizing enzymes in human cervical smooth muscle cells-implication in the mechanism of cervical ripening. *Int J Gynaecol Obstet.* 2000;70:B56.
67. Franchi L, Eigenbrod T, Muñoz-Planillo R, Nuñez G. The inflammasome: a caspase-1 activation platform that regulates immune responses and disease pathogenesis. *Nat Immunol.* 2009;10:241–7.
68. Shao B-Z, Xu Z-Q, Han B-Z, Su D-F, Liu C. NLRP3 inflammasome and its inhibitors: a review. *Front Pharmacol.* 2015;6:1061.
69. van de Veerdonk FL, Netea MG, Dinarello CA, Joosten LAB. Inflammasome activation and IL-1 β and IL-18 processing during infection. *Trends Immunol.* 2011;32:110–16.
70. Petrilli V, Papin S, Tschopp J. The inflammasome. *Curr Biol.* 2005;15:R581.
71. Jin C, Flavell RA. Molecular mechanism of NLRP3 inflammasome activation. *J Clin Immunol.* 2010;30:628–31.
72. Gotsch F, Romero R, Chaiworapongsa T, Erez O, Vaisbuch E, Espinoza J, et al. Evidence of the involvement of caspase-1 under physiologic and pathologic cellular stress during human pregnancy: a link between the inflammasome and parturition. *J Matern Fetal Neonatal Med.* 2008;21:605–16.
73. Gomez-Lopez N, Romero R, Xu Y, Plazyo O, Unkel R, Leng Y, et al. A Role for the Inflammasome in Spontaneous Preterm Labor With Acute Histologic Chorioamnionitis. *Reprod Sci.* 2017;24:1382–401.
74. Plazyo O, Romero R, Unkel R, Balancio A, Mial TN, Xu Y, et al. HMGB1 Induces an Inflammatory Response in the Chorioamniotic Membranes That Is Partially Mediated by the Inflammasome. *Biol Reprod.* 2016;95:130.
75. Romero R, Xu Y, Plazyo O, Chaemsathong P, Chaiworapongsa T, Unkel R, et al. A Role for the Inflammasome in Spontaneous Labor at Term. *Am J Reprod Immunol.* 2018;79:e12440.
76. Panaitescu B, Romero R, Gomez-Lopez N, Xu Y, Leng Y, Maymon E, et al. In vivo evidence of inflammasome activation during spontaneous labor at term. *J Matern Neonatal Med.* 2019;32:1978–91. <https://doi.org/10.1080/14767058.2017.1422714>.
77. Pacora P, Romero R, Maymon E, Gervasi MT, Gomez R, Edwin SS, et al. Participation of the novel cytokine interleukin 18 in the host response to intra-amniotic infection. *Am J Obstet Gynecol.* 2000;183:1138–43.
78. Holt R, Timmons BC, Akgul Y, Akins ML, Mahendroo M. The molecular mechanisms of cervical ripening differ between term and preterm birth. *Endocrinology.* 2011;152:1036–46.
79. Maymon E, Romero R, Pacora P, Gomez R, Athayde N, Edwin S, et al. Human neutrophil collagenase (matrix metalloproteinase 8) in parturition, premature rupture of the membranes, and intrauterine infection. *Am J Obstet Gynecol.* 2000;183:94–9.
80. Kim SM, Romero R, Lee J, Chaemsathong P, Lee M-W, Chaiyasit N, et al. About one-half of early spontaneous preterm deliveries can be identified by a rapid matrix metalloproteinase-8 (MMP-8) bedside test at the time of mid-trimester genetic amniocentesis. *J Matern Fetal Neonatal Med.* 2016;29:2414–22.
81. Arechavala-Velasco F, Marciano D, Díaz-Cueto L, Parry S. Matrix metalloproteinase-8 is expressed in human chorion during labor. *Am J Obstet Gynecol.* 2004;190:843–50.
82. Witkin SS. The vaginal microbiome, vaginal anti-microbial defence mechanisms and the clinical challenge of reducing infection-related preterm birth. *BJOG.* 2015;122:213–8.
83. Farr M, Strübe J, Geppert HG, Kocourek A, Mahne M, Tschesche H. Pregnancy-associated plasma protein-E (PAPP-E). *Biochim Biophys Acta.* 2000;1493:356–62.
84. Chim SSC, Lee WS, Ting YH, Chan OK, Lee SWY, Leung TY. Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma. *PLoS One.* 2012;7:e34328.
85. Winn VD, Haimov-Kochman R, Paquet AC, Yang YJ, Madhusudhan MS, Gormley M, et al. Gene expression profiling of the human maternal-fetal

- interface reveals dramatic changes between midgestation and term. *Endocrinology*. 2007;148:1059–79.
86. Partl JZ, Karin V, Skrtic A, Nikuseva-Martic T, Serman A, Curkovic-Perica M, et al. Negative regulators of Wnt signaling pathway SFRP1 and SFRP3 expression in preterm and term pathologic placentas. *J Matern Neonatal Med*. 2018;31:2971–9. <https://doi.org/10.1080/14767058.2017.1359830>.
87. Novakovic B, Fournier T, Harris LK, James J, Roberts CT, Yong HEJ, et al. Increased methylation and decreased expression of homeobox genes TLX1, HOXA10 and DLX5 in human placenta are associated with trophoblast differentiation. *Sci Rep*. 2017;7:4523.
88. Zadora J, Singh M, Herse F, Przybyl L, Haase N, Golic M, et al. Disturbed Placental Imprinting in Preeclampsia Leads to Altered Expression of DLX5, a Human-Specific Early Trophoblast Marker. *Circulation*. 2017;136:1824–39.
89. Eidem HR, Rinker DC, Ackerman WE 4th, Buhimschi IA, Buhimschi CS, Dunn-Fletcher C, et al. Comparing human and macaque placental transcriptomes to disentangle preterm birth pathology from gestational age effects. *Placenta*. 2016;41:74–82.
90. Davidson LM, Coward K. Molecular mechanisms of membrane interaction at implantation. *Birth Defects Res C Embryo Today*. 2016;108:19–32.
91. Wray S, Burduga T, Noble D, Noble K, Borysova L, Arrowsmith S. Progress in understanding electro-mechanical signalling in the myometrium. *Acta Physiol*. 2015;213:417–31.
92. Buxton ILO, Heyman N, Wu Y-Y, Barnett S, Ulrich C. A role of stretch-activated potassium currents in the regulation of uterine smooth muscle contraction. *Acta Pharmacol Sin*. 2011;32:758–64.
93. Vaeth M, Feske S. Ion channelopathies of the immune system. *Curr Opin Immunol*. 2018;52:39–50.
94. Gekas C, Dieterlen-Lièvre F, Orkin SH, Mikkola HKA. The placenta is a niche for hematopoietic stem cells. *Dev Cell*. 2005;8:365–75.
95. Khodadi E, Shahrabi S, Shahjehani M, Azandeh S, Saki N. Role of stem cell factor in the placental niche. *Cell Tissue Res*. 2016;366:523–31.
96. Alvarez-Silva M, Belo-Diabangouaya P, Salaün J, Dieterlen-Lièvre F. Mouse placenta is a major hematopoietic organ. *Development*. 2003;130:5437–44.
97. Ottersbach K, Dzierzak E. Analysis of the mouse placenta as a hematopoietic stem cell niche. *Methods Mol Biol*. 2009;538:335–46.
98. Kuchma MD, Kyryk VM, Svitina HM, Shablii YM, Lukash LL, Lobyntseva GS, et al. Comparative Analysis of the Hematopoietic Progenitor Cells from Placenta, Cord Blood, and Fetal Liver, Based on Their Immunophenotype. *Biomed Res Int*. 2015;2015:418752.
99. Mattern A, Zellmann T, Beck-Sickinger AG. Processing, signaling, and physiological function of chemerin. *IUBMB Life*. 2014;66:19–26.
100. Kobayashi Y. The role of chemokines in neutrophil biology. *Front Biosci*. 2008;13:2400–7.
101. Mahmud ZA, Jenkins L, Ulven T, Labéguère F, Gosmini R, De Vos S, et al. Three classes of ligands each bind to distinct sites on the orphan G protein-coupled receptor GPR84. *Sci Rep*. 2017;7:17953.
102. Presicce P, Park C-W, Sentharamakannan P, Bhattacharyya S, Jackson C, Kong F, et al. IL-1 signaling mediates intrauterine inflammation and chorio-decidua neutrophil recruitment and activation. *JCI Insight*. 2018;3. <https://doi.org/10.1172/jci.insight.98306>.
103. Uusküla L, Männik J, Rull K, Minajeva A, Kõks S, Vaas P, et al. Mid-gestational gene expression profile in placenta and link to pregnancy complications. *PLoS One*. 2012;7:e49248.
104. Zhu YX, Benn S, Li ZH, Wei E, Masih-Khan E, Trieu Y, et al. The SH3-SAM adaptor HACS1 is up-regulated in B cell activation signaling cascades. *J Exp Med*. 2004;200:737–47.
105. Rahmioglu N, Drong AW, Lockstone H, Tapmeier T, Hellner K, Saare M, et al. Variability of genome-wide DNA methylation and mRNA expression profiles in reproductive and endocrine disease related tissues. *Epigenetics*. 2017;12: 897–908.
106. Ackerman WE 4th, Buhimschi IA, Brubaker D, Maxwell S, Rood KM, Chance MR, et al. Integrated microRNA and mRNA network analysis of the human myometrial transcriptome in the transition from quiescence to labor. *Biol Reprod*. 2018. <https://doi.org/10.1093/biolre/roy040>.
107. Menon R, Richardson LS. Preterm prelabor rupture of the membranes: A disease of the fetal membranes. *Semin Perinatol*. 2017;41:409–19.
108. Plunkett J, Muglia LJ. Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. *Ann Med*. 2008;40:167–95.
109. Plunkett J, Doniger S, Orabona G, Morgan T, Haataja R, Hallman M, et al. An Evolutionary Genomic Approach to Identify Genes Involved in Human Birth Timing. *PLoS Genet*. 2011;7:e1001365.
110. Crider KS, Whitehead N, Buus RM. Genetic variation associated with preterm birth: A HuGE review. *Genet Med*. 2005;7:593–604.
111. Romero R, Mazaki-Tovi S, Vaisbuch E, Kusanovic JP, Chaiworapongsa T, Gomez R, et al. Metabolomics in premature labor: a novel approach to identify patients at risk for preterm delivery. *J Matern Fetal Neonatal Med*. 2010;23:1344–59.
112. Kim J. Identification of genes contributing to preterm birth: insights from genetic, transcriptomic, and epigenetic analyses. PhD thesis, University of Iowa, 2012. <https://doi.org/10.17077/etd.p3zuvpa3>.
113. Sans M, Salzano FM, Chakraborty R. Historical genetics in Uruguay: estimates of biological origins and their problems. *Hum Biol*. 1997;69:161–70.
114. Hidalgo PC, Bengochea M, Abilleira D, Cabrera A, Alvarez I. Genetic Admixture Estimate in the Uruguayan Population Based on the Loci LDLR, GYPA, HBGG, GC and D7S8. *Int J Hum Genet*. 2005;5:217–22.
115. Yang E, Wang G, Yang J, Zhou B, Tian Y, Cai JJ. Epistasis and destabilizing mutations shape gene expression variability in humans via distinct modes of action. *Hum Mol Genet*. 2016;25:4911–9.
116. Lappalainen T, Sammeth M, Friedländer MR, PAC 't h, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501:506–11.
117. Cappetta M, Berdasco M, Hochmann J, Bonilla C, Sans M, Hidalgo PC, et al. Effect of genetic ancestry on leukocyte global DNA methylation in cancer patients. *BMC Cancer*. 2015;15:434.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

