



UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA



**Análisis bioinformático del transcriptoma de
progenitores eritroides con delección bialélica del
gen *Was***

Horacio Javier Sosa Galeano

Tutora: Dra. María José Fernández de Nestosa

Cotutora: Dra. Dora B. Krimer

Tesis presentada para obtener el Título de Máster en Ciencias de la Computación

San Lorenzo - Paraguay

Agosto, 2021

Sosa Galeano, Horacio Javier.

Análisis bioinformático del transcriptoma de progenitores eritroides con delección bialélica del gen *Was* / Horacio Javier Sosa Galeano. -- San Lorenzo : Facultad Politécnica, UNA, 2021. CXLIII; 30 cm.

Tesis (Maestría en Ciencias de la Computación). -- Universidad Nacional de Asunción, Facultad Politécnica.

1. Gen Was. 2. Síndrome de Wiskott-Aldrich. 3. Células MEL. 4. Eritropoyesis. 5. Diferenciación eritroide. 6. Biosíntesis. 7. Enfermedades inmunológicas. 8. Inmunodeficiencia primaria. 9. Bioinformática. 10. Tesis y Disertaciones Académicas -- Paraguay. I. Título.

CDD PY 616.042

**ANÁLISIS BIOINFORMÁTICO DEL TRANSCRIPTOMA DE PROGENITORES
ERITROIDES CON DELECCIÓN BIALÉLICA DEL GEN *WAS***

HORACIO JAVIER SOSA GALEANO

Aprobado en fecha 24 de Agosto de 2021,

Tribunal Examinador:

- | | |
|------------------------|---|
| - José Colbes | FP-UNA |
| - Magaly Martínez | IICS-UNA |
| - Diego Sánchez | Instituto de Patología e Investigación, IPI |
| - María José Fernández | FP-UNA, Tutora |
| - Dora B.Krimer | CIB-CSIC, Co-tutora |

Dra. María José Fernández-Nestosa
Tutora

Dr. Horacio Legal Ayala
Director del Postgrado
Postgrado en Ciencias de la
Computación Facultad Politécnica
Universidad Nacional de Asunción

AGRADECIMIENTOS

A mi familia, en especial a mi madre Graciela, a mi abuela Yolanda, y a mis tías, quiénes me acompañaron y apoyaron incondicionalmente en estos 3 años, dándome aliento, contención y mucho cariño. También a mis amigos, Rodrigo, Anita, Eduardo y Santiago por siempre sacarme una risa, brindarme un hombro y escucharme en momentos difíciles. También quiero agradecer a Santiago por su gran ayuda con el diseño de los gráficos de esta tesis.

A la Dra. María José y a Diego Sánchez, por su inmensa paciencia, sus enseñanzas y su calidez humana durante estos casi tres años.

Por último, a mi yo del 2018, por empezar esta maestría como un desafío en una época oscura, y poder reencontrarme de nuevo conmigo mismo y con las cosas que me gustan, muchas gracias por persistir.

ANÁLISIS BIOINFORMÁTICO DEL TRANSCRIPTOMA DE PROGENITORES ERITROIDES CON DELECCIÓN BIALÉLICA DEL GEN *WAS*

Autor: Horacio Javier Sosa Galeano

Tutora: María José Fernández-Nestosa

Co-tutora: Dora B. Krimer

RESUMEN

La línea celular MEL (*Murine Erytroleukemia cell line*) deriva de proeritroblastos inmortalizados con el SFFV (*Spleen Focus Forming Virus*), el cual se integra en el locus del Sfpi-1/PU.1, corriente arriba de la región codificante del factor de transcripción de PU.1, lo que resulta en una proliferación descontrolada y el bloqueo de la diferenciación eritroide. A partir de la exposición continua de células MEL al agente inductor HMBA, se estableció la línea celular MEL-R, cuya principal característica es la resistencia a los agentes inductores. En estudios previos, se llevaron a cabo análisis transcriptómicos entre MEL y MEL-R logrando identificar genes diferencialmente expresados, entre los que se encontró un grupo de genes relacionados con la organización del citoesqueleto de actina. Dentro de este grupo destacaron los genes que codifican para la proteína tirosina quinasa de Bruton (BTK) y la proteína del síndrome de *Wiskott-Aldrich* (WASp), ambas específicas del linaje eritroide y relacionadas a enfermedades hematológicas severas. La pérdida de función de WASp ocasiona marcados defectos en el citoesqueleto de actina y los pacientes con síndrome de *Wiskott-Aldrich* (SWA) presentan un rango variado de rasgos clínicos que incluyen problemas inflamatorios, inmunodeficiencia, autoinmunidad y, en algunos casos, anemia hemolítica. Si bien el rol de *Was* está ampliamente estudiado en el linaje linfóide y mieloide, su papel dentro del linaje eritroide todavía se mantiene poco claro. En la presente tesis se llevó a cabo un análisis bioinformático del transcriptoma de células MEL con delección bialélica de *Was*, con el fin de estudiar su función en el contexto eritroide y su posible rol en la reprogramación de la diferenciación en células eritroleucémicas.

Los resultados de la presente tesis confirmaron que *Was* es prescindible en la maduración final del linaje eritroide. Además, la ausencia de *Was* altera la expresión de genes relacionados con la biosíntesis del grupo hemo, apoptosis, metabolismo celular y dianas de oncogenes del linaje eritroide. Por último, nuestros datos también sugieren que la expresión de BTK es necesaria para el correcto funcionamiento de WASp, lo cual podría tener implicancias en el tratamiento de pacientes con SWA.

BIOINFORMATIC ANALYSIS OF ERYTHROID PROGENITOR TRANSCRIPTOME WITH BIALLELIC DELETION OF THE WAS GENE

Author: Horacio Javier Sosa Galeano
Mentor: María José Fernández-Nestosa
Co-mentor: Dora B. Krimer

ABSTRACT

The MEL cell line (*Murine Erytroleukemia cell line*) is derived from proerythroblasts immortalized with SFFV (*Spleen Focus Forming Virus*), which integrates into the PU.1 locus, upstream of the coding region, resulting in uncontrolled proliferation and blocking erythroid differentiation. The MEL-R cell line was established from continuous exposure of MEL cells to the differentiation-inducing agent HMBA, and its main characteristic is the resistance to inducing agents. In previous studies, transcriptomic analyses were carried out between MEL and MEL-R, identifying differentially expressed genes, including a group of genes related to the organization of the actin cytoskeleton. Within this group, genes that encode for Bruton's protein tyrosine kinase (BTK) and Wiskott-Aldrich syndrome protein (WASp) stood out because both are specific to the erythroid lineage and are related to severe hematological diseases. Loss of WASp function causes remarkable defects in the actin cytoskeleton, and patients with Wiskott-Aldrich syndrome (WAS) present a wide range of clinical features, including inflammatory problems, immunodeficiency, autoimmunity, predisposition to cancer, and in some cases, hemolytic anemia. While the role of *Was* has been widely studied in the lymphoid and myeloid lineage, its role within the erythroid lineage remains unclear. In the present thesis, a bioinformatic analysis of the transcriptome of MEL cells with biallelic deletion of the gene *Was* was performed, in order to study its function in the erythroid context and its possible role in reprogramming the differentiation of erythroleukemic cells.

The results of the present thesis confirmed that *Was* is dispensable during the final maturation of the erythroid lineage. Furthermore, the absence of *Was* changes the

expression of genes related to the biosynthesis of the heme group, apoptosis, cellular metabolism and targets of oncogenes of the erythroid lineage. Finally, our data also suggest that the expression of BTK is necessary for the correct function of WASp, which could have implications in treatments for WAS patients.

TABLA DE CONTENIDOS

	Pág
1 INTRODUCCIÓN	1
1.1 EXPANSIÓN Y DIFERENCIACIÓN DE LA SERIE ERITROPOYÉTICA	2
1.2 ERITROPOYESIS <i>IN VIVO</i>	2
1.2.1 Niveles de regulación de la eritropoyesis.....	5
1.3 ERITROPOYESIS <i>IN VITRO</i>	8
1.3.1 Las células MEL como modelo <i>in vitro</i> de la diferenciación eritroide	9
1.3.2 Caracterización de líneas celulares eritroleucémicas resistentes a la acción de agentes inductores de la diferenciación	12
1.4 IDENTIFICACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS EN CÉLULAS ERITROLEUCÉMICAS RESISTENTES A LA ACCIÓN DE AGENTES INDUCTORES DE LA DIFERENCIACIÓN.....	13
1.5 FUNCIÓN DE WAS Y BTK EN LA ORGANIZACIÓN DEL CITOESQUELETO DE ACTINA	14
1.5.1 Características y funciones de Was.....	14
1.5.2 Características y funciones de Btk	21
1.6 ANÁLISIS TRANSCRIPTÓMICOS: <i>RNA-SEQ</i>	23
1.6.1 Análisis bioinformático del transcriptoma.....	25
1.6.2 Análisis de calidad de las lecturas.....	26
1.6.3 Pre-procesamiento de lecturas	26
1.6.4 Alineamiento de lecturas contra un genoma de referencia.....	27
1.6.5 Estimación de niveles de expresión.....	27
1.6.6 Comparación de expresión génica entre grupos.....	28
1.6.7 Anotación funcional de genes.....	29
1.7 RELEVANCIA DEL ESTUDIO DE WASP EN EL CONTEXTO DE LA DIFERENCIACIÓN ERITROIDE.....	29
2 OBJETIVOS.....	31
2.1 OBJETIVO GENERAL:.....	31
2.2 OBJETIVOS ESPECÍFICOS:	31
3 MATERIALES Y MÉTODOS	33
3.1 LÍNEAS CELULARES	33
3.2 TRANSCRIPTOMAS ANALIZADOS.....	34
3.2.1 Eritroblastos murinos en estadios finales de la diferenciación eritropoyética.....	34
3.2.2 Macrófagos de pacientes con el síndrome de Wiskott-Aldrich	34
3.3 MÉTODOS EXPERIMENTALES.....	35
3.3.1 Validación de la delección bialélica de los genes Was en células MEL	35
3.3.2 Aislamiento y purificación de RNA para la preparación de librerías.....	36
3.4 ANÁLISIS DE <i>RNA-SEQ</i>	38
3.4.1 Alineamiento de lecturas	38
3.4.2 Cuantificación de expresión génica.....	39
3.4.3 Análisis de expresión diferencial	40
3.4.4 Métodos para el análisis funcional	41
4 RESULTADOS	43
4.1 ANÁLISIS BIOINFORMÁTICO DE SECUENCIAS TRANSCRIPTÓMICAS OBTENIDAS MEDIANTE <i>RNA-SEQ</i>	43
4.2 VALIDACIÓN DE LA DELECCIÓN BIALÉLICA DE LOS GENES WAS Y BTK EN CÉLULAS MEL.....	44
4.3 DESCRIPCIÓN Y ANÁLISIS DE CALIDAD DE LAS SECUENCIAS OBTENIDAS.....	46

4.4	ANÁLISIS DE CORRELACIÓN ENTRE MUESTRAS ESTUDIADAS.....	47
4.5	IDENTIFICACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS	48
4.6	BÚSQUEDA DE LA FUNCIONALIDAD DE LOS GENES IDENTIFICADOS MEDIANTE EXPRESIÓN DIFERENCIAL.....	49
4.6.1	<i>Anotación funcional de genes diferencialmente expresados mediante ORA usando la base de datos KEGG</i>	50
4.6.2	<i>Anotación funcional de genes diferencialmente expresados empleando el método de GSEA</i>	52
4.7	COMPARACIÓN DE LOS GRUPOS FUNCIONALES IDENTIFICADOS MEDIANTE MÉTODOS DISTINTOS	55
4.8	COMPARACIÓN DE LOS GRUPOS FUNCIONALES IDENTIFICADOS EN MEL $W_{AS}^{-/-}$ CON CÉLULAS DE PACIENTES CON SÍNDROME DE WISKOTT-ALDRICH	55
4.9	ANÁLISIS DE PATRONES DE EXPRESIÓN DE GENES PERTENECIENTES A LAS PRINCIPALES CATEGORÍAS FUNCIONALES IDENTIFICADAS USANDO TRANSCRIPTOMAS DE ESTADIOS FINALES DE LA DIFERENCIACIÓN ERITROIDE Y MODELOS DE REPROGRAMACIÓN DE LA DIFERENCIACIÓN EN CÉLULAS ERITROLEUCÉMICAS.....	57
4.9.1	<i>Análisis de expresión de genes relacionados con apoptosis</i>	57
4.9.2	<i>Análisis de expresión de genes pertenecientes a la vía de señalización de HIF</i>	59
4.9.3	<i>Análisis de expresión de genes diana de c-Myc</i>	61
4.9.4	<i>Análisis de la expresión de genes que codifican para proteínas que regulan la biosíntesis del grupo hemo.</i>	62
4.10	COMPARACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS EN MEL $W_{AS}^{-/-}$ Y MEL-R CON RESPECTO A MEL	64
4.11	ANÁLISIS DE LA EXPRESIÓN DE GENES QUE CODIFICAN PARA MIEMBROS DE LA FAMILIA WASP	66
4.12	AMPLIACIÓN DE LA ANOTACIÓN FUNCIONAL USANDO LA BASE DATOS GO.....	70
4.13	ANÁLISIS DE LA EXPRESIÓN DIFERENCIAL DE TRANSCRITOS ENTRE MEL $W_{AS}^{-/-}$ Y MEL	71
5	DISCUSIÓN.....	75
5.1	CONVERGENCIA DE ANOTACIONES FUNCIONALES	76
5.2	ROL DE W_{AS} EN LA REGULACIÓN DE PROCESOS RELACIONADOS CON LA DIFERENCIACIÓN ERITROPOYÉTICA ...	77
5.3	ROL DE W_{AS} EN LA REGULACIÓN DE LA EXPRESIÓN DE DIANAS ONCOGENES EN PROGENITORES ERITROIDES...	79
5.4	ROL NUCLEAR DE W_{AS} : REGULACIÓN DEL <i>SPLICING</i> Y EL <i>SPLICEOSOMA</i>	81
5.5	WASP Y BTK: INTERACCIÓN Y REGULACIÓN DE FUNCIONAL.	83
5.6	TRABAJOS FUTUROS.....	84
6	CONCLUSIONES	86
7	REFERENCIAS BIBLIOGRAFICAS	88
8	ANEXOS.....	105
8.1	COMUNICACIONES A CONGRESOS NACIONALES E INTERNACIONALES DERIVADAS DE ESTA TESIS DE MAESTRÍA	129

TABLAS DE FIGURAS

FIGURA 1. DIFERENCIACIÓN DE LA SERIE ERITROPOYÉTICA <i>IN VIVO</i> .	5
FIGURA 2. FASES DE LA TRANSFORMACIÓN DE PROERITROBLASTOS MEDIANTE EL COMPLEJO VÍRICO FRIEND.	10
FIGURA 4. DOMINIOS Y CAMBIOS DE CONFORMACIÓN DE LA PROTEÍNA WASP.	16
FIGURA 6. ANÁLISIS DE CLONES CON DELECCIONES BIALÉLICAS.	36
FIGURA 7. VALIDACIÓN EXPERIMENTAL DE LA DELECCIÓN BIALÉLICA DE LOS GENES <i>Was</i> Y <i>Btk</i> .	44
FIGURA 8. DESCRIPCIÓN GENERAL DEL ANÁLISIS DE LOS TRANSCRIPTOMAS ESTUDIADOS.	47
FIGURA 9. GENES CUYA EXPRESIÓN DIFIERE SIGNIFICATIVAMENTE ENTRE MEL Y MEL <i>Was</i> ^{-/-} .	49
FIGURA 10. CLASIFICACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS EN CATEGORÍAS FUNCIONALES DE KEGG.	51
FIGURA 12. COMPARACIÓN DE LAS ANOTACIONES FUNCIONALES USANDO DIFERENTES MÉTODOS.	55
FIGURA 13. RESULTADOS DE LA COMPARACIÓN DE LAS ANOTACIONES FUNCIONALES USANDO DIFERENTES MUESTRAS DE PACIENTES CON SWA.	56
FIGURA 14. PERFIL DE EXPRESIÓN DE GENES INVOLUCRADOS EN APOPTOSIS.	58
FIGURA 15. PERFIL DE EXPRESIÓN DE GENES INVOLUCRADOS EN HIPOXIA (VÍA DE SEÑALIZACIÓN DE HIF-1).	60
FIGURA 16. PERFIL DE EXPRESIÓN DE DIANAS DE <i>c-Myc</i> .	62
FIGURA 17. PERFIL DE EXPRESIÓN DE GENES INVOLUCRADOS EN EL METABOLISMO DEL GRUPO HEMO.	63
FIGURA 19. PERFIL DE EXPRESIÓN DE GENES DE LA FAMILIA WASP.	67
FIGURA 20. PERFIL DE EXPRESIÓN DE GRUPOS FUNCIONALES RELACIONADOS A FUNCIONES YA IDENTIFICADAS DE LA FAMILIA WASP.	69
FIGURA 21. AMPLIACIÓN DE LA ANOTACIÓN FUNCIONAL DE GENES CUYA EXPRESIÓN DISMINUYE EN MEL <i>Was</i> ^{-/-} .	70
FIGURA 22. TRANSCRITOS DIFERENCIALMENTE EXPRESADOS ENTRE MEL Y MEL <i>Was</i> ^{-/-} .	72

TABLAS

TABLA 1. DISTRIBUCIÓN DE LAS LECTURAS POR MUESTRA TENIENDO EN CUENTA LA LONGITUD Y NÚMERO DE LECTURAS	46
TABLA 2. ISOFORMAS DE mRNA DIFERENCIALMENTE EXPRESADAS ENTRE MEL <i>Was</i> ^{-/-} Y MEL	74

ABREVIATURAS

ARN	Ácido ribonucleico
ARP2/3	del inglés <i>Actin-Related Proteins-2/3</i>
BTK	del inglés <i>Bruton Tyrosine Kinase</i>
CDC42	del inglés <i>Cell Division Control Protein 42 homolog</i>
EPO	Eritropoyetina
EPO-R	Receptor de la eritropoyetina
DMSO	Dimetilsulfóxido
GATA-1	del inglés <i>GATA-binding protein 1</i>
GATA-2	del inglés <i>GATA-binding protein 2</i>
HIF-1	del inglés <i>Hypoxia-inducible factor 1</i>
HMBA	N,N'- Hexameten-bis-acetamida
JAK	del inglés <i>Janus Kinase 2</i>
STAT	del inglés <i>Signal Transducer and activator of transcription</i>
MAPK	del inglés <i>Mitogen-Activated Protein Kinase</i>
MEL	del inglés <i>Murine erythroleukemia cell line</i>
MEL-R	del inglés <i>Murine erythroleukemia cell line resistant</i>
MYC	del inglés <i>Myelocytomatosis oncogene</i>
PI3K	Fosfoinositol 3-quinasa
RNA-seq	del inglés <i>RNA-sequencing</i>
SWA	Síndrome de Wiskott-Aldrich
TNFA	del inglés <i>Tumour necrosis factor Alpha</i>
WASp	del inglés <i>Wiskott-Aldrich syndrome protein</i>
XLN	del inglés <i>X-linked neutropenia</i>
XLT	del inglés <i>X-linked thrombocytopenia</i>
E2F	del inglés <i>Eukaryotic 2 Transcription factor</i>

1 INTRODUCCIÓN

La eritroleucemia es un tipo de leucemia mieloide aguda (AML) en la que no se forman eritrocitos maduros debido a un bloqueo en el proceso de diferenciación. Uno de los sistemas mejor caracterizados de reprogramación de la diferenciación en células eritroleucémicas, que permite el estudio de los mecanismos de regulación de la diferenciación eritroide, son las células MEL (*Murine Erythroleukemia cell line*), las cuales proceden de precursores eritroides transformados con el complejo vírico Friend. En estudios anteriores, se utilizó la técnica *RNA-seq* para comparar el transcriptoma de las células MEL con el de una línea celular derivada de MEL con resistencia a los inductores de la diferenciación celular (MEL-R) (Fernández-Nestosa et al., 2008). Entre los genes diferencialmente expresados (GDEs), destacó un grupo de genes que corresponde a proteínas relacionadas con la organización del citoesqueleto y cuya expresión es mayor en MEL con respecto a MEL-R. Dentro de este grupo resaltan *Was* (*Wiskott-Aldrich syndrome*) y *Btk* (*Bruton's tyrosine kinase*) por dos razones: son genes exclusivos del linaje hematopoyético y están relacionados a enfermedades hematológicas severas (Fernández-Calleja et al., 2017).

En la presente tesis, se utilizó las líneas celulares MEL *Was*^{-/-}, obtenida a partir de células MEL usando el sistema CRISPR/Cas9, como modelo para estudiar la función molecular de *Was* en progenitores eritroides y su papel en la diferenciación eritropoyética, mediante el análisis bioinformático de datos transcriptómicos obtenidos mediante la técnica de *RNA-seq*.

1.1 Expansión y diferenciación de la serie eritropoyética

Los eritrocitos o glóbulos rojos son un tipo de célula exclusiva de los vertebrados (mamíferos, anfibios, reptiles y aves) (Anderson et al. 2018). Además de esto, son el componente principal y más abundante de la sangre por lo que se estima que un adulto posee entre 20 a 30 trillones de eritrocitos (Bianconi et al., 2013). En mamíferos, los eritrocitos se caracterizan por la ausencia de organelos al madurar, entre los cuales destacan el núcleo y las mitocondrias (Anderson et al., 2018). Debido a la ausencia de organelos, los eritrocitos no son capaces de sintetizar aminoácidos ni ácidos grasos lo que limita su tiempo de vida. Otras características importantes de los eritrocitos son: la forma aplanada y biconcava, la flexibilidad y la estructura de la membrana y el citoesqueleto, las cuales permiten a los eritrocitos pasar a través de capilares sin sufrir daños y así llegar a todas las partes de cuerpo.

Las fuerzas evolutivas han convertido a los eritrocitos en células especializadas en la provisión y suministro de oxígeno. De hecho, el color rojo de la sangre se debe a la presencia de hemoglobina, la cual constituye alrededor de 95% de todos los componentes de los eritrocitos. El principal objetivo y función de la hemoglobina es el transporte de oxígeno desde zonas con alta concentración de oxígeno (pulmones) a zonas de baja concentración (tejidos). Además de esto, también posee la capacidad de transportar dióxido de carbono desde tejidos a los pulmones, en donde será removido y reemplazado por oxígeno. El transporte y suministro de oxígeno no son las únicas funciones de los eritrocitos. Trabajos anteriores demostraron que los eritrocitos también están involucrados en la promoción de la activación y el mantenimiento de la quiescencia inmunológica (Anderson et al, 2018).

1.2 Eritropoyesis *in vivo*

Los glóbulos rojos son generados por el cuerpo mediante un proceso complejo conocido como eritropoyesis. En estudios anteriores del flujo sanguíneo se logró identificar dos grupos de células eritroides (Maximow, 1909). El primer grupo se

caracteriza por un mayor tamaño y la presencia del núcleo, mientras que el segundo grupo presenta un tamaño reducido y la ausencia del núcleo. La presencia temporal del primer grupo durante la gestación temprana se denomina eritropoyesis primitiva. Por el contrario, la eritropoyesis definitiva hace referencia al proceso que ocurre no sólo en etapas gestacionales, sino que continua a lo largo de la vida posgestacional (Palis, 2014).

La eritropoyesis definitiva tiene lugar en el hígado y, principalmente, en la médula ósea, en una unidad anatómica llamada isla eritroblástica, en donde los precursores eritroides proliferan, se diferencian y enuclean (Demirci et al., 2018). En este microambiente, la eritropoyesis es regulada mediante interacciones entre célula-célula, estímulos extrínsecos y también componentes de la matriz extracelular como fibronectinas y lamininas, que afectan la eritropoyesis en estadios terminales de diferenciación.

El primer paso de la eritropoyesis ocurre cuando las células madre hematopoyéticas (HSC) se diferencian y en sucesivos pasos dan lugar a células pertenecientes al compartimiento de progenitores eritroides (Fig. 1). Los progenitores eritroides se diferencian de los progenitores mieloides por la expresión del receptor de eritropoyetina (EPO-R) y su capacidad de responder a la eritropoyetina (EPO).

El compartimiento de progenitores eritroides está formado por BFU-E (unidades formadoras de brotes eritroides) y por CFU-E (unidades formadoras de colonias eritroides). Las BFU-E se caracterizan por un alto grado de proliferación, lo cual les permite generar brotes de colonias de más de mil células en aproximadamente 14 días. Además, las BFU-E presentan un nivel bajo de expresión de EPO-R (Lodish et al., 2010). Por el contrario, las CFU-E se caracterizan por formar colonias más pequeñas (6 a 12 células) y mostrar la expresión de genes relacionados a la biosíntesis del grupo hemo. Además, las CFU-E son sensibles a la presencia de EPO y dependen de esta citoquina para su supervivencia (Mide et al, 2001).

Los progenitores eritroides dan lugar a precursores eritroides. Este compartimiento consiste en células que todavía conservan el núcleo y cuya caracterización está basada en su morfología. La maduración de precursores eritroides transita los siguientes estadios de: proeritroblasto (Pro-E), eritroblasto basofílico (Baso), eritroblasto policromatofílico (Poly), eritroblasto ortocromático (Ortho), reticulocito (Ret) y eritrocito. A medida que estas células proliferan y diferencian, la producción de proteínas características de los eritrocitos aumenta rápidamente, particularmente la hemoglobina.

La maduración de precursores eritroides se caracteriza por una serie de cambios, entre los que se encuentran: (1) la acumulación de hemoglobina, (2) la disminución del tamaño celular, (3) la condensación de la cromatina, (4) la remoción de organelos (incluidos las mitocondrias y ribosomas), (5) la enucleación y (6) la transición estructural a una forma menos rígida y bicóncava (Myneni et al., 2021). Este proceso de maduración tiene lugar en las islas eritroblásticas de la médula ósea. Este nicho especializado está compuesto por uno o dos macrófagos ubicados en el centro, rodeados de eritroblastos en distintos estadios madurativos (Ovchinnikova et al., 2018). La eliminación de organelos ocurre mediante dos procesos importantes: la autofagocitosis y la exocitosis (Grosso et al., 2017). Los macrófagos, entre otras funciones, juegan un rol fundamental en la enucleación ya que son responsables de la fagocitosis y degradación del núcleo liberado por los eritroblastos ortocromáticos.

En la etapa final, luego de liberar el núcleo, los reticulocitos inmaduros entran en circulación mediante la migración a través de poros citoplasmáticos de células endoteliales (Rai et al., 2021). Por tanto, el último compartimiento eritroide está formado por eritrocitos maduros que son liberados al flujo sanguíneo y persisten alrededor de 120 días. Los eritrocitos que cumplieron su periodo de vida son removidos nuevamente por los macrófagos mediante el reconocimiento de antígenos característicos de la senescencia.

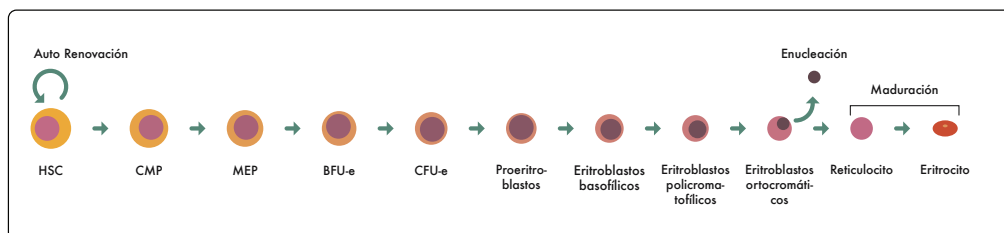


FIGURA 1. DIFERENCIACIÓN DE LA SERIE ERITROPOYÉTICA IN VIVO. Modelo del proceso de diferenciación y maduración de células eritroides. Se muestran las células madre pluripotenciales (HSC) con capacidad de renovación a largo y corto plazo (con capacidad de auto renovación) que generan células progenitoras multipotenciales (sin capacidad de auto renovación), como el progenitor común mieloide (CMP), el cual, a su vez, da lugar al progenitor megacariocítico-eritrocítico (MEP), del cual derivan las células progenitoras comprometidas BFU-E y CFU-E, que proliferan principalmente bajo el control de la EPO, y finalmente se diferencian hacia células precursoras: proeritroblastos, eritroblastos basofílicos, eritroblastos policromatófilos, eritroblastos ortocromáticos. Estas células precursoras proliferan y maduran en reticulocitos y eritrocitos.

1.2.1 Niveles de regulación de la eritropoyesis

El proceso de eritropoyesis es complejo por lo que está regulado por la combinación de efectos microambientales y factores de transcripción, los cuales controlan la supervivencia, proliferación y diferenciación de los progenitores eritroides. Además, ciertos factores nucleares también regulan la transcripción de genes importantes con la maduración y el correcto funcionamiento de los eritrocitos. El estudio de los diferentes niveles de regulación de este complejo sistema brinda nuevas oportunidades para la creación de tratamientos más eficaces contra enfermedades hematológicas, enfermedades degenerativas y el cáncer (Vandekerckhove et al., 2009).

1.2.1.1 Regulación dependiente de eritropoyetina

La eritropoyesis como todo proceso complejo está regulada a lo largo de varios niveles y por varios factores. Uno de los factores más estudiados es la citoquina EPO, la cual se une a su receptor de membrana (EPO-R) y activa la tirosina quinasa JAK2 (*Janus kinase 2*), la cual tiene como efectores a genes de las vías de PI3K/AKT (*Phosphoinositide 3-kinase/ Protein kinase B*), RAS/RAF-2/MAPK (*Rat sarcoma*

virus/ Raf-2 Proto-Oncogene, Serine/Threonine Kinase/ Mitogen-activated Protein Kinase) y STAT5 (*Signal Transducer and Activator of Transcription 5*), todas involucradas en la expansión y maduración de células del linaje eritroide (Karayel et al., 2020). Es importante destacar que la expresión del gen que codifica para la EPO depende de los factores de transcripción HIF-1 α y HIF-2 α , los cuales se activan en condiciones de hipoxia (Haase, 2013).

Las células BFU-E son las células progenitoras más precoces comprometidas con el linaje eritroide y son sensibles a la presencia de EPO. Sin embargo, estudios previos mostraron que el desarrollo normal de las colonias BFU-E es independiente de la presencia de EPO y estas presentan un bajo nivel de expresión de EPO-R (Anagnostou et al., 1994). Por otro lado, la presencia de EPO es fundamental para la formación y supervivencia de colonias CFU-E, las cuales muestran un alto nivel de expresión de EPO-R (Singh et al., 2018). De esta forma, EPO es capaz de controlar la producción de eritrocitos como respuesta a fenómenos externos e internos (sangrados, hipoxia, apoptosis y senescencia).

1.2.1.2 Funciones diferenciales de GATA-1 y GATA-2 en la regulación de la proliferación y diferenciación de progenitores eritroides

Existen varios factores de transcripción que juegan un papel fundamental en la diferenciación del linaje eritroide. Entre ellos se encuentra GATA-1, el cual forma parte de la familia GATA y es un factor esencial para la determinación y diferenciación del linaje eritroide. En etapas tempranas de la diferenciación, así como durante cuadros clínicos de anemia, GATA-1 es responsable de la activación de EPO-R, promoviendo así, la proliferación de células en el estadio CFU-E. Además, en estadios terminales de la diferenciación eritroide, GATA-1 promueve la expresión de genes esenciales para la maduración de eritrocitos, entre los que se incluyen enzimas responsables de la biosíntesis del grupo hemo, globinas alfa y beta, y proteínas características de la membrana eritroide (Gutiérrez et al., 2019).

Otro factor de transcripción importante en la diferenciación eritroide es GATA-2, también perteneciente a la familia GATA. Tanto las células madre hematopoyéticas como los progenitores pluripotentes presentan un alto nivel de expresión de GATA-2 (Moriguchi et al., 2014). Este patrón de expresión es contrario al perfil de expresión de GATA-1, cuya expresión empieza durante la determinación al linaje eritroide y se incrementa considerablemente durante la diferenciación de los proeritroblastos (Suzuki et al., 2013). Estudios anteriores demostraron que GATA-1 es capaz de promover su propia expresión y reprimir la expresión GATA-2, mientras que GATA-2 es capaz de activar la expresión tanto de GATA-2 como de GATA-1 (Suzuki, 2009). Durante la diferenciación eritroide se observa un cambio en la proporción entre estos genes, favoreciendo la expresión de GATA-1 y, por tanto, la represión de GATA-2 (Kaneko et al., 2010).

La expresión de miembros de la familia GATA regula de forma negativa a HIF-2 α que a su vez controla la eritropoyesis mediante su relación con EPO. La expresión de miembros de la familia GATA ocurre como una respuesta citoquinas proinflamatorias como IL-1, IL-6 o TNF α . Entre los miembros de la familia GATA, resalta GATA-2 por su potencial rol en la regulación de la síntesis de EPO. Por tanto, inhibidores de miembros de la familia GATA han mostrado ser capaces de estabilizar al factor de transcripción HIF-2 α , estimulando la síntesis de EPO y, por ende, la eritropoyesis (López-Gómez et al., 2016).

1.2.1.3 Factores que regulan la enucleación de los eritroblastos ortocromáticos

La producción de eritrocitos es el resultado final de una transformación ordenada de un proeritroblasto con un gran núcleo a un eritrocito sin núcleo. El paso decisivo para la transición de células del compartimiento de precursores eritroides a eritrocitos maduros y funcionales, es la enucleación. Los proeritroblastos se dividen y diferencian sometándose a numerosos cambios substanciales, como la reducción en tamaño, condensación de la cromatina y hemoglobinización, previa a su enucleación, dando como resultado reticulocitos. Existen varias histonas, proteínas modificadoras de histonas y factores de cromatina que son responsables del proceso de enucleación. En

estudios previos se demostró que la disminución de la expresión de la histona-desacetilasa HDAC2 es suficiente para el bloqueo de la condensación y la extrusión nuclear (Wang et al., 2020). Otros estudios revelaron que la disminución de la expresión del oncogén *c-Myc* en progenitores eritroides es necesaria para que se produzca la condensación nuclear y la desacetilación de histonas (Jayapal et al., 2010).

Trabajos previos también demostraron que el proceso de enucleación está relacionado al proceso de citocinesis, es decir también depende de la membrana y de las proteínas del citoesqueleto (Konstantinidis et al., 2012). Entre las proteínas involucradas en la enucleación que están relacionadas al citoesqueleto, se encuentran las Rac GTPasas, las cuales activan a sus dianas, Formina y mDIA2, facilitando la formación del anillo contráctil de actina. La desregulación de las Rac GTPasas inhibe la formación del anillo contráctil de actina en estadios finales de la eritropoyesis, bloqueando la proliferación y diferenciación de eritroblastos (Ji P et al., 2008). Durante los estadios finales de la diferenciación eritroide, varias otras proteínas del citoesqueleto sufren una reorganización. Este es el caso de la proteína 4.1, anquirina, tubulina, actina y espectrina, todas ellas presentes en el futuro reticulocito y ausentes en el pirenocito (núcleo extruído).

1.3 Eritropoyesis *in vitro*

En las últimas décadas se han realizado grandes esfuerzos para optimizar la producción *ex vivo* de eritrocitos como una alternativa para la transfusión (Rousseau et al., 2013). Tanto los avances en cultivos celulares como el avance del conocimiento entorno a los procesos de regulación de eritropoyesis son claves para el desarrollo de procedimientos experimentales que permitan la producción de reticulocitos funcionales a partir de células madre hematopoyéticas. Numerosos grupos de investigación han centrado sus esfuerzos en mejorar los procesos como la enucleación y profundizar en la caracterización molecular del proceso de maduración eritroide.

1.3.1 Las células MEL como modelo in vitro de la diferenciación eritroide

La eritroleucemia aguda es una forma rara de leucemia mieloide aguda, caracterizada por defectos en la proliferación y diferenciación eritroide (Boddu et al., 2018). Debido a la naturaleza compleja y fundamental de la eritroleucemia, diversas condiciones de cultivo fueron desarrolladas para analizar los eventos implicados en la reversión de la eritroleucemia y la reprogramación de la diferenciación eritroide (Latif et al., 2008).

Una de las líneas celulares establecidas son las células Friend o células MEL, las cuales han sido ampliamente utilizadas para estudiar el mecanismo de diferenciación eritroide. Las células MEL derivan de proeritroblastos inmortalizados debido a la infección por el complejo vírico Friend (Friend et al. 1966). Esta línea celular crece en cultivos en suspensión y bajo ciertas condiciones puede reanudar el proceso de diferenciación hasta el estadio de reticulocito (Tsiftoglou et al., 2003). En trabajos anteriores se demostró que las células MEL también se pueden diferenciar hacia el linaje de megacariocitos (Hyaman et al., 2001).

El complejo responsable de la transformación de las células MEL está compuesto dos componentes virales: el virus SFFV (*Spleen Focus-Forming Virus*), el cual es defectuoso en su replicación y el virus F-MuLV (*Friend Murine Leukemia Virus*), el cual es competente en la replicación (Spiro et al. 1988; Ruscetti 1995). En el proceso de transformación, el F-MuLV actúa como un soporte para la replicación del SFFV, mientras que SFFV es el responsable de la inmortalización de los progenitores eritroides (Moreau-Gachelin, 2008).

El proceso de transformación de precursores eritroides con el SFFV consta de dos fases bien caracterizadas, según se observa en la figura 2. Durante la primera fase, el EPO-R es activado mediante la glicoproteína gp55 codificada por el gen *env* del SFFV, activando las mismas rutas de señalización de la EPO. Además, se demostró que gp55 recluta una forma truncada del receptor STK (stem-cell kinase receptor) conocida como sf-STK (Zhang y col., 2006). Con la activación de los efectores de EPO, EPO-R y sf-STK, consecutivamente se activan también las vías de señalización JAK2-

STAT, RAS/RAF-1/MAPK y PI3K (Muszynski y col., 1998; Nishigaki y col., 2000). Sin embargo, a diferencia de la respuesta a EPO, la infección con el SFFV provoca una activación constitutiva de estas vías de señalización, lo cual resulta en la expansión excesiva del compartimiento de progenitores eritroides (Nishigaki et al. 2000, Li et al., 2000).

La segunda fase, ocurre aproximadamente de 6 a 12 semanas después de la infección. En esta etapa, aparecen clones de células malignas que han perdido la capacidad de respuesta a la diferenciación inducida por EPO. En la gran mayoría de las eritroleucemias, el virus SFFV se integra corriente arriba del promotor del gen *Sfpi/PU.1* (*SFFV Proviral Integration site-1/ Purine Rich*), que codifica para el factor de transcripción PU.1, perteneciente a la familia ETS, el cual actúa como un oncogén en el linaje eritroide (Cmarik et al., 2010). La sobreexpresión de *PU.1*, junto con una inactivación de *p53*, ya sea debido a un silenciamiento del gen que codifica para el supresor tumoral o a la expresión de una proteína mutada (Mowat y col., 1985; Munroe y col., 1990), resulta en una proliferación descontrolada y un bloqueo de la diferenciación eritroide (Ruscetti et al., 1999).

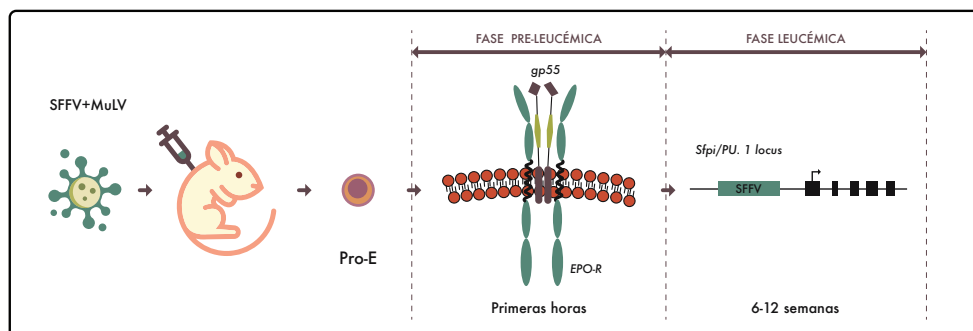


FIGURA 2. FASES DE LA TRANSFORMACIÓN DE PROERITROBLASTOS MEDIANTE EL COMPLEJO VÍRICO FRIEND. La transformación de precursores eritroides mediada por el SFFV en ratones infectados se produce en dos fases. La primera fase consiste en la sobre activación de EPO-R y sf-STK, mediante la glicoproteína gp55 del SFFV. La segunda fase ocurre gracias a la integración del virus al genoma murino, específicamente en el promotor de *Sfpi-1/PU.1*, lo cual causa la sobreexpresión de este oncogén y la continua proliferación de proeritroblastos que son incapaces de diferenciarse.

Las líneas celulares eritroleucémicas presentan características propias del estadio de proeritoblastos y pueden proliferar indefinidamente *in vitro*. La gran utilidad de las células MEL radica en su capacidad para retomar la diferenciación eritroide en presencia de agentes químicos como el HMBA (N,N'-Hexameten-bis-acetamida) o el DMSO (dimetilsulfóxido). Según resultados previos de nuestro y otros laboratorios, a partir de las 48 horas de tratamiento con 5 mM HMBA el porcentaje de células diferenciadas aumenta gradualmente hasta alcanzar alrededor del 90% a las 120 horas (Fig. 3). Como resultado a la exposición a agentes inductores, las células MEL pueden recapitular todos procesos importantes para la maduración eritroide, tales como la acumulación de hemoglobina, la salida del ciclo celular, la condensación de cromatina y la disminución del tamaño celular, gracias a la disminución de la expresión de *PU.1* y otros oncogenes como *c-Myc* y *c-Myb* (Marks et al. 1994, Papetti et al., 2007) junto con un incremento en la expresión de factores de transcripción específicos del linaje eritropoyético, tales como ELKF y GATA-1, e inhibidores del ciclo celular, p21, p27y Mad1. La culminación de la diferenciación se caracteriza por una parada de la proliferación celular, como también por una acumulación de células en la fase G1 del ciclo celular (Kiyokawa et al. 1993; Vanegas et al. 2003; Fernández-Nestosa et al. 2008).

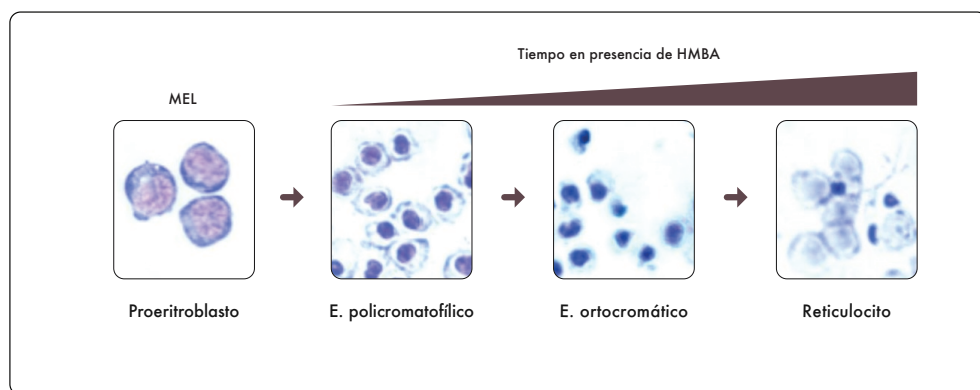


FIGURA 3. REPROGRAMACIÓN DE LA DIFERENCIACIÓN DE DIFERENCIACIÓN DE CÉLULAS ERITROLEUCÉMICAS MEL CULTIVADAS EN PRESENCIA DE HMBA. El cultivo de células MEL en presencia del agente inductor HMBA causa una reactivación del programa de diferenciación eritroide, permitiéndoles recapitular los estadios de eritroblasto policromatófilo, eritroblasto ortocromático hasta la fase de reticulocito.

En otros estudios se demostró que la sobreexpresión de *PU.1* en células MEL en presencia de agentes inductores es capaz de inhibir la diferenciación eritroide, mientras que su silenciamiento es suficiente para inducir la diferenciación incluso en ausencia de agentes químicos (Atar et al., 2005, Papetti et al., 2007). Entre los genes reclutados por *PU.1*, se encuentra la quinasa dependiente de ciclina *Cdk6*, la cual es capaz de inhibir la diferenciación de células MEL (Choe et al., 2010). *Cdk6*, asociada a las ciclinas D, regula el ciclo celular en la fase G1, y juntamente con PU.1 se expresan constitutivamente en las células proliferativas MEL (Matushansky et al., 2003). La inducción mediante agentes inductores disminuye la expresión de *PU.1* y en consecuencia los niveles de *Cdk6*, inhibiendo la proliferación y promoviendo la diferenciación. Es importante destacar que PU.1 actúa como antagonista de GATA-1, factor de transcripción crucial para la diferenciación eritroide. Ambos factores de transcripción se inhiben mutuamente y su interacción es crucial para la determinación de células CMP (*Common Myeloid Progenitor*) hacia el linaje mieloide (PU.1) o el linaje eritroide (GATA-1) (Burda et al., 2010).

1.3.2 Caracterización de líneas celulares eritroleucémicas resistentes a la acción de agentes inductores de la diferenciación

Un obstáculo para las terapias contra el cáncer es la adquisición de resistencia a los medicamentos que se desarrollan en respuesta a terapias repetidas, que finalmente conduce a una recaída en la mayoría de los pacientes (Rebucci et al., 2013). En este contexto, se han establecido modelos de diferenciación *in vitro* que han demostrado ser extremadamente útiles para estudiar los eventos moleculares asociados con el bloqueo de la diferenciación celular exhibido por algunas células tumorales y los requisitos para el reingreso al programa de diferenciación celular. Según se describió en el apartado anterior, aproximadamente un 90% de las células MEL es capaz de retomar la diferenciación en presencia de agentes inductores. Sin embargo, en todos los casos, se observa un porcentaje reducido, cercano al 10%, de células cultivadas en presencia de HMBA, que no responde al agente inductor y permanece en el estadio de proeritroblasto.

La línea celular MEL-R se estableció al aislar el 10% de células MEL que no responden a la presencia continua de HMBA (Fernández-Nestosa et al., 2008). Al ser una línea derivada de MEL, las células MEL-R mantienen la mayor parte de las características fenotípicas observadas en su línea parental. Sin embargo, un estudio previo demostró que en las células MEL-R, la expresión del oncogén *PU.1* está inhibida, incluso en ausencia de agentes inductores. Además, se observó que las células MEL-R presentan: altos niveles de expresión de genes responsables de la proliferación celular como *c-Myc* y *c-Myb*, junto con un silenciamiento de genes que antagonizan con los recién mencionados y genes específicos del linaje eritroide, como GATA-1 y EKLF (Callejas-Fernández, 2018).

Resultados de estudios anteriores demostraron que la sobreexpresión de genes que codifican para las proteínas GATA-1 o MAD1, un gen maestro para el linaje eritroide y un antagonista del oncogén *c-Myc*, respectivamente, es suficiente para revertir la resistencia en MEL-R para retomar el proceso de diferenciación en presencia de agentes inductores (Fernández-Nestosa et al., 2008). Estos resultados sugieren que la activación de la diferenciación eritropoyética depende de una compleja red de factores, entre los que se encuentran factores de transcripción específicos y factores constitutivos, que funcionan en un delicado equilibrio para activar o reprimir el programa de diferenciación eritroide (Dore et al., 2011).

1.4 Identificación de genes diferencialmente expresados en células eritroleucémicas resistentes a la acción de agentes inductores de la diferenciación

Análisis transcriptómicos entre MEL y MEL-R, revelaron que el 81.5% de genes están sobreexpresados en MEL con respecto a MEL-R. Además, mediante el análisis bioinformático se identificaron posibles factores involucrados en la resistencia a los agentes inductores observada en MEL-R (Fernández-Callejas et al., 2017). Entre los genes diferencialmente expresados, destacaron dos grupos de genes, uno relacionado a histonas variantes y otro asociado a la organización del citoesqueleto de actina. El

grupo de histonas variantes mostró una sobreexpresión en MEL-R, mientras que los genes del grupo responsable de la organización del citoesqueleto presentaron una disminución de expresión en MEL-R con respecto a MEL.

1.5 Función de Was y Btk en la organización del citoesqueleto de actina

En el análisis de expresión diferencial en células eritroleucémicas resistentes a la acción de agentes inductores de la diferenciación fueron identificados genes involucrados en la organización del citoesqueleto de actina, entre los cuales resaltan, *Was* y *Btk*, debido a que son específicos del linaje eritroide y están relacionados a enfermedades hematológicas severas. Ambos genes fueron seleccionados para profundizar el estudio de su rol dentro del contexto eritroide y eritroleucémico, así como una posible relación con la resistencia a agentes inductores de la diferenciación.

1.5.1 Características y funciones de Was

El gen *Was* ubicado en el cromosoma X (región Xp 11.22-Xp 11.3.41), codifica para la proteína WASp, una proteína con múltiples dominios y compuesta por 502 aminoácidos. *Was* se expresa de forma única en el linaje hematopoyético (Rivers et al., 2017), donde su función principal es interconectar las señales de la membrana con la polimerización de actina, mediante su interacción con el complejo ARP2/3 (Massaad et al., 2013). Anteriormente, el rol de *Was* estaba delimitado al citoplasma, promoviendo procesos como migración, citocinesis, transporte de membrana y comunicación celular (Snapper et al., 1998). Sin embargo, estudios previos demostraron que *Was* en los linajes mieloide y linfoide, también cumple funciones nucleares relacionadas a remodelación de la cromatina, procesamiento de RNA, estabilidad nuclear y transcripción (Yeng Looi et al., 2014; Sarkar et al., 2016).

Es importante resaltar que WASp cuenta con los siguientes dominios: WH1 (Wasp-Homology Domain 1), una región básica (B), GBD (Guanosine Triphosphatase-Binding Domain), una región extensa rica en prolinas (PRD) y el dominio VCA (Verprolin-Cofilin-Acidic) (Fig. 4 A). El primer dominio N-terminal WH1 es el sitio

de unión de la proteína WIP (*Wasp-Interacting Protein*), la cual enmascara el dominio WH1, evitando su degradación proteosomal (de la Fuente et al., 2007). En células T, también se observó que miembros de la familia *Src*, *Lck* y *Fyn*, son capaces de unirse al dominio WH1. Además, de su rol protector, WIP también ha sido identificado como uno de los factores responsables de la localización de WASp en áreas de polimerización de actina (Ramesh et al., 1997).

El siguiente dominio es el dominio B, el cual es crucial en la regulación conformacional de WASp, ya que es capaz de unirse a PIP2 (fosfatidil inositol 4,5 bifosfato) y junto con la Rho GTPasa CDC42 (*Cell Division Cyclin 42*) promueve la forma activa de WASp, rompiendo su conformación auto inhibida (Thrasher et al., 2010).

Como se observa en la figura 4 B, la proteína WASp permanece en una conformación auto inhibida gracias a una interacción entre los dominios VCA y GBD (Kim et al., 2000). La interacción de WASp con otras proteínas, tales como la Rho CDC42 y la fosforilación de la tirosina Y291 del dominio GBD mediada por miembros de la familia *Src* quinasas, es capaz de cambiar la conformación de WAS a la forma activa. Este cambio conformacional resulta en la exposición del dominio VCA, facilitando la unión de WAS con el complejo ARP2/3, lo cual induce la polimerización de actina para la formación de filamentos ramificados (Campellone et al., 2010). En cuanto respecta a la degradación de WASp, la fosforilación de la tirosina Y291 del dominio GDB también juega un papel importante, ya que es capaz de actuar como un marcador de proteínas destinadas a la degradación proteosomal (Chou et al., 2006).

Estudios previos reportaron que CDC42 en su forma activa recluta a WASp a la membrana citoplasmática en donde es fosforilada por proteínas del citoplasma pertenecientes a la familia *Src* y *Tec*. Numerosas proteínas como NCK y otras pertenecientes a las familias *Src* y *Tec*, entre ellas BTK, promueven la activación de WASp mediante su unión al dominio PRD, regulando así, la organización del

citoesqueleto de actina en el linaje hematopoyético (Guinamard et al., 1998; Blundell et al., 2010).

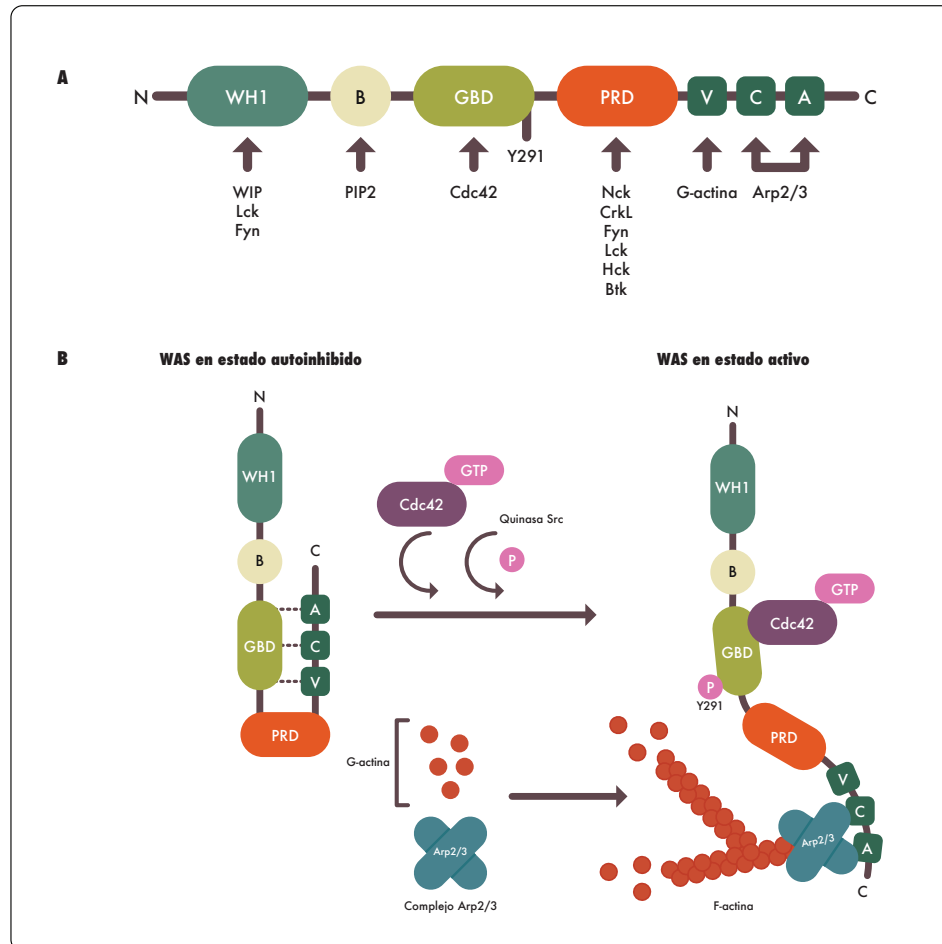


FIGURA 4. DOMINIOS Y CAMBIOS DE CONFORMACIÓN DE LA PROTEÍNA WASp. Esquema genómico de los dominios funcionales de la proteína WASp (A). Cambios conformacionales de WASp. (B) En su conformación inactiva, WASp presenta un estado auto inhibido debido a la interacción del dominio VCA con la región GBD y a la estabilización mediada por WIP. La activación de WAS tiene lugar mediante la interacción con diferentes moléculas, entre las que se encuentran CDC42 y miembros de la familia *Src*, responsables de la fosforilación de la tirosina Y291 del dominio GDB, lo que resulta en el cambio de conformación y la exposición del dominio VCA para su interacción con el complejo ARP2/3. El complejo ARP2/3 junto con WAS, promueven cambios en la organización del citoesqueleto mediante la formación de filamentos de actina.

1.5.1.1 La familia WASP: promotores de la nucleación de actina

WASp es miembro de una familia de proteínas que participan en la transducción de señales desde la superficie celular al citoesqueleto de actina. Los miembros de la familia WASP interactúan con proteínas de la membrana para promover la polimerización de actina a través del complejo ARP2/3 (Ngoenkam et al., 2021). Entre los miembros de la familia WASP se encuentran: su homólogo con expresión prevalente en el sistema nervioso, *N-Wasp* (*Neural-Wasp*), el supresor de proteínas homólogas a verprolina de la familia cAR/WASp (SCAR/WAVE1-3), la proteína WAS y homólogo de SCAR (WASH), el homólogo de WAS asociado a la actina, membrana y microtúbulos (WHAMM), y la proteína reguladora y mediada por uniones (JMY) (Massaad et al., 2013).

Todos los miembros de la familia WASP comparten el dominio C-terminal, pero varían en el dominio N-terminal, es decir, todos los miembros poseen el dominio VCA, responsable de la polimerización y formación de filamentos de actina (Massaad et al., 2013). En particular, N-WASp muestra un alto nivel de conservación en su homología, compartiendo más del 80% de conservación con WASp. Si bien WASp se expresa de forma única en el linaje hematopoyético (Rivers et al., 2017), sus miembros más cercanos, N-WASp y las proteínas SCAR/WAVE1-3, se expresan de forma ubicua (Pollitt et al., 2009; Takenawa et al., 2007). También, es sabido que miembros de la familia WASP son capaces de compensar la ausencia de otros miembros para la correcta polimerización de actina. En el citoplasma, estudios previos en macrófagos han demostrado que N-WASp, es capaz de compensar la pérdida de WASp y restaurar la integridad del citoesqueleto de actina (Isaac et al., 2010). Por otra parte, en el núcleo de timocitos, la ausencia de WASp estimula un incremento en la expresión de *WASF2*, gen que codifica para la proteína WAVE-2 (Kuznetsov et al., 2017).

1.5.1.2 Síndrome de Wiskott-Aldrich

El síndrome de Wiskott-Aldrich (SWA) es una inmunodeficiencia primaria con un patrón de transmisión hereditaria ligada al cromosoma X, que se caracteriza por la

tríada clásica de eccema, inmunodeficiencia severa y micro trombocitopenia. Este trastorno, causado por mutaciones (o alteraciones) en el gen *Was* que producen una alteración molecular en la proteína WASp, fue descrito por primera vez en 1937 por el Dr. Alfred Wiskott. Diecisiete años después, el Dr. Robert Aldrich fue capaz de aclarar que la enfermedad se transmitía de generación en generación de manera recesiva ligada al cromosoma X. Esto significa que los niños varones desarrollan la enfermedad, pero sus madres y sus hermanas, las cuales pueden tener una copia del gen de la enfermedad, no presentan síntomas.

Este síndrome afecta entre uno y diez de cada millón de recién nacidos vivos, y sin tratamiento, les otorga una esperanza de vida de 15 años. Se estima que aproximadamente un tercio de los pacientes diagnosticados con SWA no tienen historia familiar identificable y que, en su lugar, son el resultado de nuevas mutaciones genéticas que ocurrieron al momento de la concepción. Hasta el momento, se han identificado más de 300 mutaciones que ocurren dentro del gen *Was*, las cuales corresponden a mutaciones sin sentido, mutaciones que cambian el marco de lectura, deleciones, inserciones y mutaciones complejas (Ochs et al., 2006).

Según su efecto en la funcionalidad de la proteína WASp, las mutaciones han sido relacionadas con 3 fenotipos clínicos bien caracterizados. Por un lado, se encuentran las mutaciones que desencadenan en una pérdida de función y están asociadas con: trombocitopenia ligada al cromosoma X, XLT (*X-linked thrombocytopenia*), la cual se caracteriza por defectos en plaquetas y se traduce en sangrados, o el SWA, en donde además de la trombocitopenia, predominan la autoinmunidad, inmunodeficiencia y el cáncer, particularmente linfoma o leucemia (Imai et al., 2004). Por otro lado, están las mutaciones que generan una ganancia de función y desencadenan una forma de neutropenia llamada XLN (*X-linked neutropenia*), caracterizada por defectos en glóbulos blancos, lo cual predispone a pacientes a infecciones recurrentes (Sarkar et al., 2016). Las mutaciones XLN-WASP (L270P, S272P y I294T) destruyen la conformación auto inhibida de WASP y generan una proteína desplegada con mayor actividad en la polimerización de actina (Devriendt et al., 2001; Beel et al., 2009)

La severidad de SWA se evalúa mediante un sistema de puntuación en la escala del uno al cinco, el cual busca diferenciar a pacientes con SWA de los pacientes con XLT (Ochs et al., 2009). Los pacientes con XLT reciben un puntaje de uno a dos, mientras que los pacientes con SWA quienes presentan formas más graves, reciben puntajes de tres al cinco. A pesar de que el síndrome Wiskott-Aldrich es considerado como una enfermedad rara, en los últimos años se han observado importantes avances en el diagnóstico y el cuidado clínico de pacientes. Sin embargo, el pronóstico de pacientes con formas severas de SWA sigue todavía siendo negativo. Los tratamientos dependen de la severidad clínica de los pacientes y se clasifican en cuatro: tratamientos convencionales, trasplante de células madre hematopoyéticas, la terapia y la edición génica. Los tratamientos convencionales consisten en: la extirpación quirúrgica del bazo para incrementar el conteo de plaquetas, el suministro de antibióticos e inmunoglobulina intravenosa para reducir el riesgo de infección en pacientes esplenectomizados (Litzman et al., 1996). Sin embargo, los tratamientos convencionales no son capaces de reducir los riesgos de autoinmunidad y tampoco los desórdenes proliferativos de células T y B, por lo que los trasplantes de células madre hematopoyéticas son siempre más recomendados, ya que permiten la reconstitución inmune y corregir los defectos hematológicos de pacientes con SWA (Moratto et al., 2011). Tratamientos más recientes incluyen la terapia génica, la cual consiste en infundir en pacientes con SWA, un vector viral que expresa *Was*. Sin embargo, la integración viral del vector ocurre de forma casi aleatoria lo cual incrementa el riesgo de genotoxicidad en pacientes. Con el objetivo de mejorar los efectos adversos de la terapia génica, trabajos recientes reportaron el uso de la plataforma de edición génica CRISPR/Cas9 para la corrección precisa de la expresión de *Was* en células madre y progenitoras hematopoyéticas humanas (Rai et al., 2020), posicionando a las plataformas de edición génica como posible futuro tratamiento de enfermedades genéticas.

1.5.1.3 El rol de WASp en diferentes tipos de células del sistema hematopoyético

La ausencia completa de WASp ocasiona marcados defectos en el citoesqueleto de actina, lo cual resulta en efectos deletéreos en la funcionalidad de varios linajes hematopoyéticos, entre los que se encuentran los linajes: linfoide, mieloide y megacariocítico (He et al., 2017).

En el linaje linfoide, la ausencia de WASp produce inmunodeficiencia y linfopenia, desencadenando la producción anormal de células T, causada por defectos en el citoesqueleto de actina (Zhou et al., 2017). En células B, a pesar de que el desarrollo y diferenciación no se encuentran drásticamente afectados, ratones con deleciones bialélicas de *Was* presentan un menor número de conteo de células B. Además, pacientes con SWA presentan múltiples manifestaciones autoinmunes, causadas por la producción de auto anticuerpos (Castiello et al., 2014).

Otra característica común entre pacientes con formas severas de SWA es la trombocitopenia. La ausencia de WASp en plaquetas, resulta en la producción de plaquetas con estructuras aberrantes, lo cual aumenta el riesgo de sufrir hemorragias (Rivers et al., 2019). Además, se observó que la citotoxicidad y capacidad de migración de las células NK (*Natural Killers*) de pacientes con SWA se ven severamente reducidas en comparación a células que expresan WASp de forma regular, a pesar de presentar un alto conteo en el torrente sanguíneo (Mace et al., 2018).

En monocitos y macrófagos de pacientes con SWA también se observó un menor número de filopodios, además de presentar defectos graves en la formación de podosomas, polarización celular y migración, lo cual compromete su capacidad de migrar a la zona infectada o inflamada, para reconocer y destruir las células objetivo, favoreciendo las apariciones de tumores (Ishihara et al., 2012).

Por último, es importante destacar que hasta el momento el rol de WASp en el linaje eritroide es casi inexplorado. La única conexión entre defectos relacionados a WASp

y el linaje eritroide, se da en pacientes de SWA que presentan anemia hemolítica (Catucci et al., 2012). Si bien la anemia hemolítica ha sido relacionada con desórdenes autoinmunes propios de SWA, también se observan casos no autoinmunes (Sokolic et al., 2014), lo cual representa un hecho importante y atractivo para profundizar el estudio del rol de WASp en el contexto eritroide.

1.5.2 Características y funciones de *Btk*

La tirosina quinasa de Bruton y su gen fueron descubiertos en el año 1993 por dos grupos en paralelo (Tsukada et al. 1993; Vetrie et al. 1993). Las mutaciones en el gen *Btk* son causantes de la agammaglobulinemia ligada al cromosoma X, XLA (*X-linked agammaglobulinemia*), que se caracteriza por defectos en el desarrollo de células B y la producción de inmunoglobulinas (Suri et al., 2016).

El gen *Btk* pertenece a la familia de genes *Tec* quinasas, la segunda familia más grande de proteínas tirosinas quinasas que no son receptores de membrana (Berg et al., 2005). La proteína BTK consta de 659 aminoácidos y, como se muestra en la figura 4, presenta cinco dominios: el dominio PH (*Pleckstrin homology*) característico de las quinasas, el dominio TH (*Tec homology*), el dominio SH3 (*Src homology 3*), SH2 (*Src homology 2*) y el dominio quinasa SH1 (*Src homology 1*). En humanos, los genes pertenecientes a la familia *Tec* se expresan de forma específica en el linaje hematopoyético.

En la actualidad, los mecanismos de activación de BTK todavía no son comprendidos en su totalidad. BTK se encuentra principalmente en el citoplasma y es reclutada en la membrana celular de forma transitoria mediante su dominio PH, el cual interactúa con PIP₃ (*Phosphatidylinositol-3 kinase*) (Fig. 5). De hecho, la activación de BTK ocurre en dos pasos sucesivos luego de su reclutamiento en la membrana. Primeramente, quinasas miembros de la familia Src, presumiblemente LYN, fosforilan a BTK en la posición Y551, lo cual promueve la actividad catalítica de BTK y su autofosforilación en la posición Y223 del dominio SH3. La función de la auto fosforilación de la

posición Y223 todavía es poco clara, pero se cree que tendría como función la estabilización de la conformación activa de la proteína, lo cual favorecería su actividad quinasa (Pal Singh et al., 2018). En estudios previos se demostró que la activación de BTK es dependiente de PU.1 (Himmelman et al., 1996). En su estado activo, BTK es capaz de promover varias vías de señalización, incluyendo PI3K/AKT, fosfolipasa C y NF-Kb, las cuales están involucradas en proliferación, supervivencia, motilidad, metabolismo, angiogénesis y evasión de la respuesta inmune antitumoral (Lindvall et al., 2005).

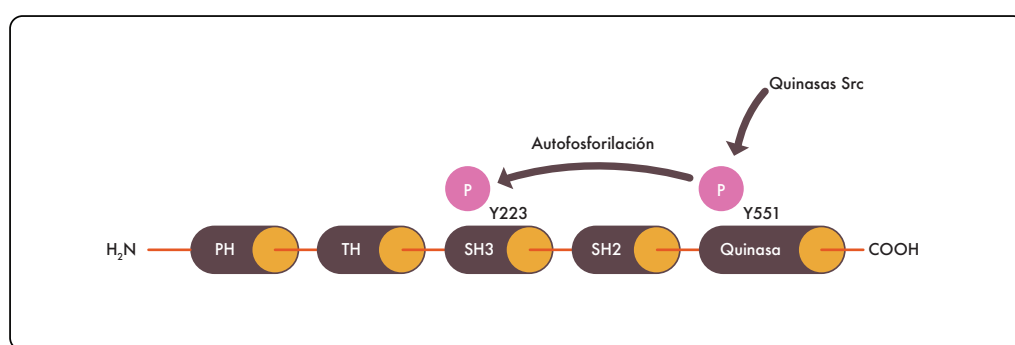


FIGURA 5. DOMINIOS DE LA PROTEÍNA BTK. Se muestran los dominios PH (*Pleckstrin homology domain*), TH (*Tec homology domain*), SH2, SH3 y Src1. BTK se encuentra mayormente en el citoplasma y su activación ocurre mediante una interacción temporal con la membrana celular, lugar donde ocurre la fosforilación de la tirosina Y551 del dominio SH1, mediada por integrantes de la familia de las quinasas *Src*. La fosforilación de BTK en la tirosina Y551 favorece la auto fosforilación de la tirosina Y223 en el dominio SH3.

Estudios *in vitro* también demostraron que el dominio PH es responsable de la promoción de la formación de filamentos de actina. Además, los miembros de la familia *Tec* son capaces de colocalizar con el citoesqueleto de actina, en donde actúan sobre las proteínas Rho GTPasas para así reorganizar el citoesqueleto en respuesta a estímulos (Roman-Garcia et al., 2018).

Trabajos previos en el linaje mieloide han reportado la interacción física entre WASp y BTK. Esta interacción ocurre entre el dominio SH3 de BTK y el dominio N-terminal de WASp y juega un papel importante en la activación de vías señalización relacionadas a la respuesta inflamatoria. La inhibición de la interacción de WASp y

BTK en macrófagos provoca alteración en la fosforilación del inhibidor de la quinasas IKK $\alpha\beta$ y el factor nuclear NF- κ B con una reducción en la expresión de transcritos de citosinas inflamatorias como IL-6, IL-1 β y TNF- α (Sakuma et al., 2015). Sin embargo, el rol de esta interacción en el contexto de la inflamación observada en pacientes con el síndrome Wiskott-Aldrich y su efecto en otros linajes, se mantienen como temas poco explorados.

Btk también se expresa en células del linaje eritroide. En este linaje, BTK es capaz de interactuar con jugadores claves de la eritropoyesis, tales como EPO-R y JAK2. En progenitores eritroide, defectos en la expresión de BTK resultan en una disminución de la fosforilación de EPO-R, JAK2, STAT5A/B y PLC γ 1, así como también causa una reducción en la respuesta de células a EPO. Además, en ausencia de BTK, los progenitores eritroides se vuelven más susceptibles a la apoptosis inducida mediante la activación de los receptores TNF α y TRAIL (*TNF-related apoptosis-inducing ligand*) (Schmidt et al., 2004).

Por último, estudios anteriores identificaron una relación entre la activación mediante BTK de vías de señalización relacionadas a la proliferación y la supervivencia de neoplasias malignas de células B. Además, BTK también se expresa en células mieloides, en donde promueve la activación de la vía NLRP3, la cual activa una forma inflamatoria de muerte celular en macrófagos, que resulta finalmente en la promoción de la progresión de células cancerosas (Ito et al., 2015). Por tanto, la inhibición de BTK mediante agentes químicos ha demostrado ser un atractivo candidato como terapia anticancerígena para neoplasias malignas de células B e incluso en tumores sólidos (Pal Singh et al., 2018).

1.6 Análisis transcriptómicos: *RNA-seq*

El transcriptoma consiste en el conjunto de transcritos cuantificados de células bajo ciertas condiciones de estudios para la caracterización de perfiles de expresión. El estudio del transcriptoma en diferentes contextos se lleva a cabo desde hace décadas. Inicialmente se usaban técnicas de bajo rendimiento, como *Northern-blot*, hibridación

in situ y la reacción en cadena de la polimerasa en transcripción inversa (RT-PCR), que sólo permiten el análisis cualitativo o semicuantitativo de un gen candidato. Posteriormente, el desarrollo de la técnica cuantitativa RT-qPCR aumentó el rendimiento; varios transcritos podían ser analizados al mismo tiempo, aunque esto todavía estaba lejos de una cobertura a gran escala de todo el transcriptoma.

No fue hasta la aparición de los *microarrays* que se logró la caracterización de perfiles de expresión de miles de transcritos simultáneamente. Los perfiles de expresión reflejan en cierta medida los cambios en el comportamiento a nivel molecular que pueden estar relacionados a cambios en el fenotipo de células expuestas a diferentes condiciones. Por tanto, el conocimiento del transcriptoma ha demostrado ser esencial en la identificación de elementos funcionales del genoma y jugadores moleculares involucrados en las condiciones estudiadas (Wang et al., 2019). Los resultados de este tipo de estudios son usados en diversas áreas como en investigaciones biomédicas, diagnóstico de enfermedades, descubrimiento de biomarcadores, estudios de riesgo y seguridad de nuevas drogas (Hrdlickova et al., 2017).

La aparición de la nueva tecnología llamada secuenciación de nueva generación (NGS) ha revolucionado el mundo de la investigación y gracias a su gran versatilidad hizo posible la secuenciación masiva de todas las moléculas de ARN presentes en una muestra y, por tanto, del transcriptoma completo, mediante la llamada *RNA-seq* (*RNA-sequencing*). Debido al gran número de aplicaciones del *RNA-seq*, el flujo de trabajo a ser usado depende del diseño experimental y de los objetivos de la investigación. Sin embargo, un flujo de trabajo completo consiste en la preparación del diseño experimental, el aislamiento del RNA, la preparación de librerías para cada muestra, la secuenciación química de esas librerías y los análisis bioinformáticos posteriores (Conesa et al., 2016).

El análisis transcriptómico mediante *RNA-seq* presenta varias ventajas frente a las estrategias de hibridación utilizadas con más frecuencia en el pasado. Entre sus ventajas se encuentran una mayor sensibilidad para identificar genes con un bajo o

alto nivel expresión, un rango más amplio y dinámico de niveles de expresión para detectar transcritos, una menor variación técnica y altos niveles de reproducibilidad (Han et al., 2015). Otra ventaja es que la técnica de *RNA-seq* no está supeditada al conocimiento del genoma del organismo en estudio, pudiendo aplicarse en especies con genomas todavía no secuenciados (Haas et al., 2013). Además, el *RNA-seq* no presenta el sesgo introducido por la hibridación usada por los *microarrays* (Swindell et al., 2014). Estas ventajas hacen que sea ampliamente usado en el descubrimiento de nuevos genes, la expresión diferencial de genes, análisis de variantes, detección de nuevas isoformas e incluso la detección de la fusión de genes (Geraci et al., 2020).

La gran mayoría de los análisis transcriptómicos se basan en la suposición de que las células son homogéneas por lo que estos estudios no tienen en cuenta la variabilidad entre células del mismo tipo. Sin embargo, en los últimos años se demostró que incluso células de un mismo tipo presentan una expresión génica heterogénea (Shalek et al., 2014), por lo que los estudios del transcriptoma de células individuales o *sc-RNA-seq* (*single-cell RNA-seq*) surgen como una respuesta para dilucidar funciones celulares y cómo la expresión génica se traduce en beneficios y detrimentos para el sistema de estudio (Hwang et al., 2018).

1.6.1 Análisis bioinformático del transcriptoma

Como fue descrito anteriormente, la técnica *RNA-seq* es una poderosa tecnología con múltiples aplicaciones que varían desde el descubrimiento de genes y variantes de *splicing* hasta análisis de expresión diferencial y detección de fusión de genes, variantes y edición de ARN. Debido a sus múltiples aplicaciones, existen una variedad de protocolos disponibles, los cuales se adaptan a las necesidades y preguntas científicas del investigador. Además, el área de análisis de datos mediante *RNA-seq* está en constante evolución, por lo que constantemente se desarrollan nuevas herramientas y técnicas, ofreciendo numerosas alternativas para cada paso del análisis.

A continuación, se presenta una descripción de los pasos principales para el análisis de expresión diferencial génica, el cual constituye una de las aplicaciones más comunes del *RNA-seq*.

1.6.2 Análisis de calidad de las lecturas

El análisis de secuencias empieza con la obtención de lecturas crudas, generalmente en formato FASTQ. El análisis de calidad consiste en evaluar de forma general la calidad de los millones de lecturas obtenidas. Las lecturas son analizadas en busca de bases con baja fiabilidad, composición de nucleótidos sesgada, duplicaciones, adaptadores, etc. El resultado de este análisis provee estadísticos como el número de lecturas e información sobre la calidad, los cuales permiten el pre-procesamiento de lecturas y garantizan la fiabilidad de los datos para la posterior generación de hipótesis (Li et al., 2014). Comúnmente, el software FASTQC es usado para el análisis de calidad. FASTQC está disponible como programa Java independiente con una interfaz gráfica de usuario y también en su versión de línea de comando. La mayor ventaja de FASTQC radica en su rapidez para analizar millones de secuencias en minutos y la visualización de sus resultados sobre la calidad de base y contenido, longitud de la lectura, contenido *k-mer*, presencia de bases ambiguas, secuencias sobrerrepresentadas y duplicados (Andrews, 2010).

1.6.3 Preprocesamiento de lecturas

En este paso, el principal objetivo es remover las secuencias con baja calidad y artefactos como adaptadores, de lecturas individuales que, de no ser removidas, podrían interferir con el análisis en pasos posteriores. La forma más simple consiste en remover las secuencias de acuerdo con una cierta longitud o remover un cierto número de bases de cualquiera de los extremos (Pereira et al., 2017). Una de las herramientas más usadas para el preprocesamiento de lecturas es Trimmomatic, la cual está basada en Java y puede usarse vía línea de comando o con una interfaz gráfica de

usuario. Su principal ventaja radica en su rapidez y facilidad de uso, ya que varios procesos pueden ser realizados en un solo comando (Bolger et al., 2014).

1.6.4 Alineamiento de lecturas contra un genoma de referencia

El principal objetivo de este proceso es identificar el punto de origen de cada una de las lecturas obtenidas dentro del genoma de referencia. En el caso de no contar con un genoma de referencia disponible, las lecturas pueden ser alineadas contra un transcriptoma, el cual puede ser creado *de novo* a partir de las lecturas. Para realizar el alineamiento es necesario contar con secuencias de referencia, así como con las secuencias preprocesadas. Cuando una lectura es alineada a un sitio dentro del genoma de referencia, se crea una secuencia de alineamiento.

El proceso de alineamiento es computacionalmente costoso, ya que implica encontrar sitios en el extenso genoma que coincidan con los millones de secuencias obtenidas, a pesar de la presencia de errores de secuenciación, variación genética y elementos repetitivos. Por tanto, varios programas de alineamiento utilizan diversas estrategias que permiten acelerar el proceso y mantener la fiabilidad del alineamiento. Entre las estrategias se encuentran la conversión y compactación del genoma a índices, así como el uso del enfoque de *seed-and-extend*, para acelerar el proceso de alineamiento. Entre los programas de alineamientos más usados se encuentra STAR (*Spliced Transcripts Alignment to a Reference*) el cual destaca por su rapidez. Sin embargo, en contrapartida STAR requiere considerablemente más memoria que otros alineadores como TopHat. Además, STAR es capaz de realizar búsquedas sin sesgos de *splice junctions* porque no necesita información previa sobre sus localizaciones o longitud de intrón.

1.6.5 Estimación de niveles de expresión

Un proceso clave dentro del análisis de *RNA-seq* es la cuantificación de lecturas por genes y transcritos. En este paso, las lecturas son asociadas individualmente a un gen

en específico, basándose en la localización del alineamiento. La estimación de la abundancia es reportada en conteos crudos o en unidades normalizadas. Las normalizaciones permiten comparaciones no solo entre muestras sino también entre genes. Entre las unidades normalizadas se encuentran RPKM (*reads per thousand nucleotides in transcript per million mapped reads*), FPKM (*fragments per thousand nucleotides per million mapped reads*) o TPM (*transcripts per kilobase million*). Tanto FPKM como RPKM son usados para corregir los recuentos en base al tamaño de la librería y la longitud del transcrito, mientras que TPM toma en cuenta la distribución de la longitud del transcrito en la muestra y, por ende, genera estimaciones de abundancia más consistentes entre muestras. En la última década, una de las herramientas más utilizadas en la estimación de la expresión es featureCounts, por ser fácil de usar, rápida y precisa. Otra característica de featureCounts es que ignora a las lecturas alineadas a varias regiones genómicas, ya sea por homología o repetición de secuencias (Liao et al., 2014).

1.6.6 Comparación de expresión génica entre grupos

Los objetivos fundamentales de un análisis de expresión diferencial son: (1) estimar la magnitud de la expresión diferencial entre dos o más condiciones, basadas en la cuantificación de la expresión en las muestras, y (2) estimar el grado de significancia de la diferencia encontrada. Las herramientas mejores categorizadas por su funcionamiento son edgeR, DESeq y lima-voom (Anders et al., 2010). Si bien DESeq y lima-voom tienden a ser más conservativas que edgeR (mejor control de falsos positivos), esta última es más recomendada para trabajar con experimentos de menos de 12 replicados (Schurch et al., 2015). Estas herramientas básicamente usan la misma estrategia, es decir, estiman la diferencia en la expresión diferencial usando modelos basados en regresión lineal, seguidos de una prueba estadística cuya hipótesis nula sostiene que la diferencia de expresión es cercana a cero, lo cual significa que las diferencias no pueden ser explicadas por las condiciones. El resultado de este paso es una lista de genes diferencialmente expresados en una condición determinada con respecto a otra.

1.6.7 Anotación funcional de genes

La anotación funcional de genes se realiza con el objetivo de obtener una perspectiva más biológica de los resultados de los análisis bioinformáticos del transcriptoma. Los análisis más comunes de anotación funcional incluyen el análisis de sobrerrepresentación ORA (*Over Representation Analysis*) y el análisis de enriquecimiento de genes de tipo GSEA (*Gene Set Enrichment Analysis*). El análisis de sobrerrepresentación básicamente categoriza genes en grupos funcionales en base a una función común, a una vía común o su presencia en una localización celular específica. Para determinar si los grupos están sobrerrepresentados, es necesario determinar la probabilidad de tener la proporción observada de genes asociados a un grupo funcional en base a la proporción de genes asociados con la misma categoría, pero en un conjunto de genes (pudiendo ser todos los genes). Una de las herramientas más usadas para el análisis de sobrerrepresentación es clusterProfiler. Esta herramienta toma como entrada una lista de genes diferencialmente expresados y realiza análisis estadísticos usando la prueba hipergeométrica.

Otro método ampliamente usado en los últimos años corresponde a las herramientas de puntuación de clases funcionales, como es el caso de GSEA, el cual usa estadísticos a nivel de gen ($\log_2FC$ entre otros) obtenidas del análisis diferencial para determinar si los conjuntos de genes pertenecientes a vías moleculares o procesos biológicos están más relacionados a una de las condiciones estudiadas.

1.7 Relevancia del estudio de WASp en el contexto de la diferenciación eritroide

Tanto *Was* como *Btk* juegan un papel fundamental en el linaje hematopoyético (Xia et al., 2020). Las mutaciones identificadas en estos genes han sido relacionadas con patologías hematopoyéticas severas, las cuales afectan principalmente la funcionalidad del linaje linfoide. La importancia del estudio de *Was* radica en su relación con el Síndrome de Wiskott-Aldrich, enfermedad que afecta a 10 de cada millón de recién nacidos. Además, algunos pacientes con SWA presentan anemia hemolítica no autoinmunes entre sus síntomas, lo cual compromete al linaje eritroide.

En los últimos años, han sido descubiertas numerosas nuevas funciones de *Was*, entre las cuales se encuentran: la supresión tumoral, la regulación de la transcripción y la organización nuclear y mitocondrial, la mayoría de ellas caracterizadas en el linaje linfoide, lo cual muestra el complejo rol de *Was* dentro de la célula. La presente tesis tiene como objetivo dilucidar nuevas funciones moleculares de *Was* que puedan contribuir al conocimiento de posibles efectos del SWA sobre el contexto eritroide. Es por este motivo que, en la presente tesis, se analizaron los transcriptomas de las líneas celulares MEL y MEL *Was*^{-/-} mediante *RNA-seq*. Además, mediante el análisis bioinformático de transcriptoma se comparó el perfil de expresión génica de la delección bialélica de *Was* con los perfiles de expresión de pacientes con SWA. Por último, debido a que en trabajos previos se sugiere un posible rol de *Was* en la regulación de la transcripción (Wu et al., 2004), en la presente tesis se analizó la relación entre la ausencia de *Was* y la expresión diferencial de isoformas entre las condiciones MEL y MEL *Was*^{-/-} con el objetivo de identificar un posible rol de *Was* en la regulación de la expresión génica y el funcionamiento de la maquinaria de *splicing*

2 OBJETIVOS

2.1 Objetivo general:

Analizar la función molecular del gen *Was* en progenitores eritroides y su papel en la diferenciación eritroide.

2.2 Objetivos específicos:

1. Realizar la secuenciación masiva de los transcriptomas de MEL, MEL-R, MEL *Btk*^{-/-} y MEL *Was*^{-/-}.
2. Identificar genes cuya expresión diferencial sea significativa entre MEL y MEL *Was*^{-/-}.
3. Realizar la anotación funcional de genes diferencialmente expresados entre MEL y MEL *Was*^{-/-} y analizar su expresión en MEL, MEL-R, MEL *Btk*^{-/-} y MEL *Was*^{-/-}.
4. Comparar los grupos (o categorías) funcionales identificados en MEL *Was*^{-/-} con los identificados en pacientes con el síndrome de Wiskott-Aldrich.
5. Comparar los perfiles de expresión de los grupos (o categorías) funcionales de genes identificados con datos transcriptómicos correspondientes a eritroblastos de los estadios finales de la diferenciación eritroide.
6. Comparar los perfiles de expresión de los grupos (o categorías) funcionales de genes identificados con datos transcriptómicos correspondientes a células

MEL cultivadas en presencia o ausencia de un agente inductor de la diferenciación.

3 MATERIALES Y MÉTODOS

3.1 Líneas celulares

MEL-DS19 (*Murine Erythroleukemia cell line*). Línea celular cedida por el laboratorio de la Dra. Dora Krimer (España). Deriva de una de las líneas originales establecidas por Charlotte Friend a partir de tumores subcutáneos de ratones infectados con el complejo vírico Friend (Friend y col., 1966; Marks y Rifkind, 1978) formado por el SFFV (*Spleen Focus Forming Virus*) y el MuLV (*Murine Leukemia Virus*).

MEL-R. Línea celular derivada de MEL DS-19 establecida mediante pases sucesivos en presencia de N,N'-Hexametileno-bis-acetamida (HMBA) (Sigma) 5 mM durante 6-8 semanas (Fernandez-Nestosa et al. 2008)

MEL *Was*^{-/-}. Línea celular derivada de MEL DS-19 establecida mediante la deleción bialélica de una región del genoma de 9,5 kb que comprende todos los exones de *Was* localizado en el cromosoma X:7658591-7667617 de ratón (desde el segmento -497 a +9.006) utilizando el sistema CRISPR/Cas9 (Fernández-Calleja et al., 2019).

MEL *Btk*^{-/-}. Línea celular derivada de MEL DS-19 establecida mediante la deleción de 5,8 Kb de una región del genoma que incluye los tres primeros exones de *Btk* localizado en el cromosoma X:131076879-131117679 de ratón (desde el segmento -300 a +5.549) empleando la estrategia CRISPR/Cas9.

3.2 Transcriptomas analizados

- Librerías de secuencias de **MEL-DS19**, **MEL-R**, **MEL *Was*^{-/-}** y **MEL *Btk*^{-/-}** obtenidas luego de la secuenciación masiva de las muestras descritas en el apartado 3.1.
- Librerías de secuencias de células MEL cultivadas en presencia (**MEL *induced***) o ausencia (**MEL *uninduced***) de 1,5% de dimetilsulfóxido (DMSO) durante 96 horas. Los datos crudos generados a través de *RNA-seq*, fueron obtenidos de la base de datos *Gene Expression Omnibus (GEO)* con código de acceso GSM3906016, GSM3906017, GSM3906018, GSM3906019 y corresponden a datos no publicados.

3.2.1 Eritroblastos murinos en estadios finales de la diferenciación eritropoyética

Las muestras pertenecientes a este grupo corresponden a células extraídas de la médula ósea y clasificadas usando marcadores de la diferenciación TER119 y CD44, en: **proeritroblastos** (Pro-e), **eritroblastos basofílicos** (Baso), **policromatofílicos** (Poly) y **ortocromáticos** (Orto). Los datos crudos generados a través de *RNA-seq*, fueron obtenidos de la base de datos *Sequence Read Archive (SRA)* con código de acceso SRR1106099, SRR1106100, SRR1106101, SRR1106102, SRR1106103, SRR1106104, SRR1106105, SRR1106106, SRR1106107, SRR1106108, SRR1106109, SRR1106100 y corresponden a los datos presentados por An et al. (2014).

3.2.2 Macrófagos de pacientes con el síndrome de Wiskott-Aldrich

Las muestras analizadas corresponden a macrófagos derivados de células madre pluripotentes inducidas (*iPSCs*) a partir de fibroblastos de dos pacientes con el síndrome de Wiskott-Aldrich y células isogénicas con rescate del gen *Was*.

Paciente 1: presentó una mutación sin sentido (c.107_108del, p.F36*) lo que se traduce en un truncamiento después del dominio WH1. Puntaje de severidad clínica 4 (del 1 al 5, siendo 5 el más severo)

Paciente 2: presentó una mutación una mutación no reportada anteriormente (c.1271dupG, p.L425fs) que causa la delección del dominio VCA. Actualmente se desconoce su puntaje de severidad clínica.

Los datos crudos generados a través de *RNA-seq*, fueron obtenidos de la base de datos *Sequence Read Archive (SRA)* con código de acceso SRR3113677, SRR3113678, SRR3113679, SRR3113680, SRR3113681 y corresponden a datos no publicados de Li et al. (2018).

3.3 Métodos experimentales

3.3.1 Validación de la delección bialélica de los genes *Was* en células MEL

La delección genómica de los genes *Was* y *Btk* fue validada mediante *PCR (Polimerase Chain Reaction)*, usando cebadores de delección, que anillan en secuencias localizadas corriente arriba y corriente abajo de los sitios de unión de los RNA guías (sgRNAs), en los extremos 3' y 5' del gen; y cebadores de no delección, los cuales amplifican una región interna de la zona escindida (Fig. 6), siguiendo el protocolo descrito por Fernández-Calleja et al. (Fernández-Calleja et al., 2019). La lista de oligonucleótidos empleados se observa en la tabla S1.

La ausencia de expresión se confirmó mediante la técnica de *Western blot* a partir de extractos proteicos totales de $5-10 \times 10^6$ células, empleando anticuerpos específicos para cada gen (Tabla S2). La inmunodetección se llevó a cabo según las condiciones descritas en Fernandez-Nestosa et al. (2008).

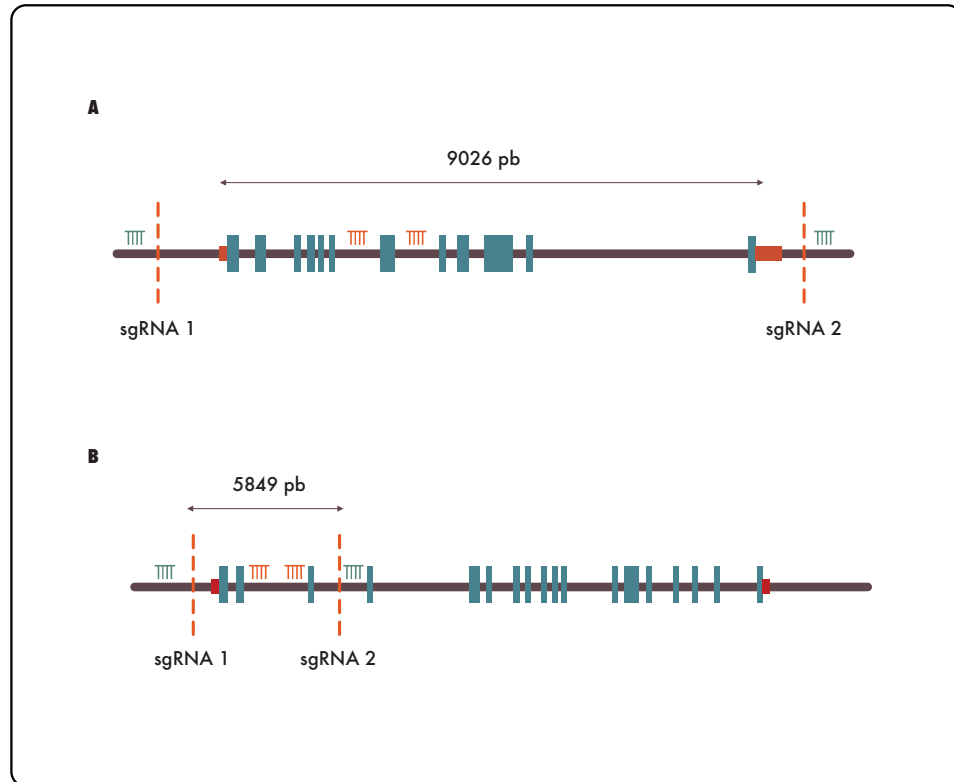


FIGURA 6. ANÁLISIS DE CLONES CON DELECCIONES BIALÉLICAS. Mapa genómico de genes deletados mediante CRISPR/Cas9. *Was* ubicado en cromosoma X:7658591-7667617 de ratón (desde el segmento -497 a +9.006) (A) y *Btk* localizado en el cromosoma X:131076879-131117679 de ratón (desde el segmento -300 a +5.549) (B). Los exones están representados por rectángulos de color azul y el 5' y 3' UTR con rectángulos de color rojo. Los sitios de corte para ambos genes están señalados por líneas discontinuas de color rojo. Los fragmentos verdes y rojos corresponden a las parejas de oligonucleótidos de deleción (verde) y no deleción (rojo), respectivamente, que se emplearon en las ampliaciones por PCR.

3.3.2 Aislamiento y purificación de RNA para la preparación de librerías

El ARN total fue aislado a partir de 1×10^7 células usando el kit *RNeasy* (Qiagen). En el proceso se incluyó la digestión con *DNase I* para evitar cualquier contaminación con DNA genómico.

La cuantificación del ARN aislado fue realizada en un espectrofotómetro de volumen múltiple (*Multiskan GO*, ThermoFisher Scientific) a partir de 2 μ l de cada muestra, los cuales fueron depositados en una placa μ DropTM (ThermoFisher Scientific). Se realizó la lectura de la densidad óptica (OD) a 260 y 280 nm. Los datos fueron procesados mediante el software *SkatIt* (ThermoFisher Scientific). El resultado de las mediciones puede ser observado en la tabla S3.

La integridad de las muestras de ARN se analizó mediante electroforesis en geles de agarosa en condiciones denaturalizantes. Las muestras se prepararon a partir de una mezcla de 2-5 μ g de ARN, formamida al 50%, tampón de carga al 2%, formaldehído 37% (5,5%) y bromuro de etidio a una concentración de 1 mg/ml. Los geles se prepararon con agarosa en MOPS 1x y formaldehído 37%. La electroforesis se llevó a cabo durante 1 hora a 100 V. El gel se visualizó en un transiluminador de luz ultravioleta a 250 nm de longitud de onda.

3.3.3 Cuantificación de la expresión génica relativa mediante PCR a tiempo real

La determinación de la expresión relativa de genes que presentaron una expresión diferencial en MEL *Was*^{-/-} y MEL se realizó utilizando RT-PCR a tiempo real. Para cada muestra se empleó una mezcla de 2x FastStart SYBR Green Master (Roche), 0.2 μ g de cDNA y 0,3 μ M de cada oligonucleótido. Los oligonucleótidos empleados, representados en la Tabla S4, fueron extraídos del sitio web PrimerBank (<https://pga.mgh.harvard.edu/primerbank/>) y sintetizados por Sigma. La reacción se llevó a cabo en el LightCycler® 96 (Roche) siguiendo las condiciones que se detallan: una pre-denaturación del ADN a 95°C durante 5 minutos, 40 ciclos de 95°C durante 10 segundos y 60°C durante 30 segundos, con un paso final a 95°C durante 10 segundos, 60°C durante 1min y un segundo a 97°C. El análisis de la expresión relativa a partir de las Cts obtenidas se efectuó mediante el método $2^{-\Delta\Delta C_t}$ (Schmittgen and Livak 2008) empleando el software LightCycler y Microsoft Office Excel. Previamente, se comprobó la eficiencia de cada pareja de oligonucleótidos mediante una curva estándar de calibrado.

3.4 Análisis de RNA-Seq

Las librerías de *RNA-seq* se prepararon a partir de 1 µg de ARN total, en un volumen máximo de 50 µl de H₂O, basado en aislamiento de polyA+ (*TruSeq RNA Sample Preparation Kit*, Illumina). Las librerías con una longitud de 337-367 nucleótidos se cuantificaron por PCR (Kapa Biosystems) empleando una librería previa estándar. Las librerías se cargaron en un equipo *Cluster Station* (Illumina) y se secuenciaron en un *Illumina GAIIx* (Parque Científico de Madrid).

3.4.1 Control de calidad

Las lecturas obtenidas fueron filtradas mediante estándares de calidad requeridos con el objetivo de evitar errores que puedan afectar el alineamiento. Para el análisis de calidad se usó el software *FastQC* versión 3 (Leggett, Ramirez-Gonzalez et al. 2013), el cual brinda información referente a parámetros de calidad: calidad de secuencia por base, contenido de GC por base, contenido de nucleótidos no identificados (N) por base, distribución de longitud de la secuencia, niveles de duplicación y secuencias sobrerrepresentadas.

En el caso de que los resultados del análisis de calidad de lecturas sean óptimos, no es necesario llevar a cabo la eliminación de pares de bases de baja calidad para asegurar la calidad de los datos. El software *Trimmomatic* versión 0.36 (Bolger, Lohse et al. 2014) se usó para encontrar y remover adaptadores (*Truseq*, Illumina), de forma parcial o total en los extremos 3' de las secuencias, los cuales podían interferir en el alineamiento.

3.4.2 Alineamiento de lecturas

Para el análisis de expresión diferencial tanto el genoma de referencia (en formato *FASTA*) como su anotación (en formato *gtf*), se descargaron de la página web *GENCODE* (Harrow, Frankish et al. 2012). En este caso en particular, el genoma de referencia usado corresponde a la versión de ensamble GRCm38 y al lanzamiento M24 de *GENCODE*.

Los índices del genoma requeridos para el alineamiento, se realizaron usando el genoma de referencia y su anotación, mediante el *software* STAR (*Spliced Transcript Alignment to a Reference*) versión 2.7 (Dobin, Davis et al. 2013) especificando la longitud de las lecturas en 75 bp. Las lecturas filtradas, obtenidas luego de la remoción de adaptadores, fueron alineadas contra el genoma de referencia usando los índices generados. En el alineamiento *splice-aware* se utilizó el filtro por defecto de *STAR*, que permite un máximo de hasta 10 alineamientos de una lectura, y en el caso de sobrepasarlo, la lectura es descartada de entre los resultados. La longitud mínima de semilla utilizada fue la predeterminada en *STAR* y corresponde a 5 nucleótidos. Debido a que los parámetros predeterminados de *STAR* ya están optimizados para genomas de mamíferos, no se hicieron otras modificaciones significativas.

Para el análisis de uso diferencial de transcritos, se usó el transcriptoma descargado de *GENCODE*, correspondiente al lanzamiento M24. Para cumplir con los requerimientos del *software* *Salmon*, versión 0.10.0 (Patro, Duggal et al. 2017) se concatenó el genoma al final del transcriptoma (*decoy*) para la generar índices que fueron usados en el alineamiento. La creación de índices fue hecha usando parámetros predeterminados citados en Robert et al. (2013).

3.4.3 Cuantificación de expresión génica

La cuantificación de genes expresados se hizo usando el algoritmo de conteo *featureCounts* versión 2.0.1 (Liao, Smyth et al. 2014) en *R*. El conteo fue hecho a nivel de genes, considerando sólo aquellos que localizan a un único sitio del genoma de referencia.

Para el análisis de expresión diferencial de transcritos, la abundancia de transcritos de nuestro conjunto de lecturas fue realizada mediante el paquete, previamente descrito, *Salmon*, frente a los índices realizados previamente. Los parámetros usados fueron los predeterminados mencionados en Shi et al. (2019).

3.4.4 Análisis de expresión diferencial

El análisis de expresión génica diferencial fue realizado a partir de los conteos generados en el paso anterior, mediante el paquete de *R/Bioconductor*, edgeR versión 3.0 (Robinson, McCarthy et al. 2010).

Previo al análisis de expresión diferencial, los genes con menor expresión (conteo) fueron removidos del conjunto de datos, usando el filtro de edgeR, para tener en cuenta sólo los genes transcripcionalmente activos. Se considerados genes transcripcionalmente activos aquellos cuyos conteos fueron iguales o mayores a 10 en un número mínimo de muestras. Las muestras de diferentes grupos fueron normalizadas por el método *TMM* (*Trimmed mean of $\underline{M}$ values*) para remover el efecto *batch* entre muestras de diferentes grupos.

Los genes diferencialmente expresados (DEGs) se identificaron usando el modelo binomial negativo (glmfit) de edgeR y una comparación por pareja entre MEL *Was*^{-/-} y MEL, usando MEL como condición basal. Fueron considerados genes diferencialmente expresados aquellos que cumplen con el criterio de punto de corte: *fold change* mayor o igual a 2 y un valor ajustado de *p* (*quasi-likelihood F-test* y *FDR: método Benjamini-Hochberg*) menor o igual a 0.05.

Para el análisis del uso de isoformas y *splicing* alternativo entre MEL y MEL *Was*^{-/-}, las cuantificaciones a nivel de transcritos generadas por *Salmon* se importaron al ambiente de *R/Bioconductor* mediante el paquete *tximport* (Soneson, Love et al. 2015). La matriz de conteo fue generada mediante *tximport* escalando cada muestra (*scaledTPM*) por la estimación de la abundancia de transcritos por millón (TPM). El análisis estadístico del uso diferencial de transcritos fue realizado a través del paquete *DRIMSeq* (Nowicka, Robinson 2016), el cual asume el modelo multinomial de Dirichlet (MMD) para cada gen. Antes de aplicar el modelo usado por *DRIMSeq*, los transcritos fueron filtrados siguiendo los siguientes parámetros: una proporción de abundancia (proporción de la expresión génica total que proviene de cada uno de los transcritos de los genes entre diferentes condiciones) relativa de al menos 0.1 y un conteo total correspondiente a cada gen de al menos 10 en las tres muestras. Luego de

ser aplicado el modelo, los transcritos diferencialmente usados fueron identificados usando el valor de p , el cual determina si las proporciones de uso de ese transcrito cambian dentro de cada gen.

Luego del análisis a nivel de transcritos, la confirmación de transcritos de genes que muestran evidencia de expresión diferencial se hizo usando el paquete *StageR* versión 3.12 (Van den Berge, Soneson et al. 2017). El punto de corte para la identificación de genes que muestran uso diferencial de transcritos fue valor ajustado de p (FDR , método *Benjamini Hochberg*) igual o menor a 0.05.

3.4.5 Métodos para el análisis Funcional

3.4.5.1 ORA (*Over Representation Analysis*)

El paquete *clusterProfiler* versión 3.12 (Yu, Han et al. 2012) se utilizó para realizar el análisis de sobrerrepresentación de genes diferencialmente expresados usando los *pathways* pertenecientes al modelo *Mus musculus* de la base datos *KEGG*, mediante la estimación de proporciones usando la prueba hipergeométrica. Para la estimación de grupos funcionales sobrerrepresentados en *KEGG*, los DEGs fueron divididos en dos grupos en base al siguiente criterio: A) genes con sobreexpresión en MEL *Was*^{-/-}: aquellos con *fold-change* positivo y B) genes con disminución de expresión en MEL *Was*^{-/-}: aquellos con *fold-change* negativo. Posterior a la conformación de estos grupos, el análisis *ORA* fue realizado de forma individual para ambos grupos. Los grupos funcionales resultantes para ambos grupos (sobreexpresados y con disminución de expresión) fueron seleccionados usando como punto de corte un valor ajustado de p menor o igual a 0.05 (FDR , método *Benjamini Hochberg*).

3.4.5.2 GSEA (*Gene Set Enrichment Analysis*)

Para el análisis de enriquecimiento se utilizó GSEA versión 2.2.4 (Subramanian, Tamayo et al. 2005) con el conjunto de datos perteneciente a los transcriptomas de MEL y MEL *Was*^{-/-}, previamente filtrado para eliminar genes duplicados y con baja expresión (según fue descrito en 4.4). Los genes del conjunto de datos

transcriptómicos, provenientes de la nomenclatura MGI (*Mouse Genome Informatics*), fueron convertidos a símbolos de la nomenclatura HUGO (*Human Genome Organization*), con el objetivo de obtener mayor compatibilidad al ser comparados contra los grupos de genes de GSEA. Los genes fueron ordenados según: el estadístico F , y el signo capturado por el *fold-change* el cual indica sobreexpresión, en el caso de ser positivo, o disminución de la expresión, en caso de ser negativo. Estos mismos estadísticos fueron los usados para la identificación de grupos de genes enriquecidos pertenecientes al conjunto de firmas genéticas de GSEA (*Hallmarks*) (<http://www.gsea-msigdb.org/gsea/msigdb>), el cual contiene 50 grupos de genes relacionados con una de las dos condiciones comparadas. Los grupos enriquecidos para una de las dos condiciones estudiadas fueron aquellos con un valor ajustado de p (FDR , método *Benjamini Hochberg*) menor o igual a 0.25, de acuerdo con el criterio establecido para GSEA.

3.4.5.3 Visualización de resultados de la anotación funcional

Los resultados provenientes del análisis de *ORA* y *GSEA* fueron visualizados mediante el paquete *ggplot2* de *R/Bioconductor*. Además, los resultados de ambos métodos fueron comparados y visualizados mediante el paquete *Venn.diagram* de *R*.

Los grupos funcionales explorados en detalle en la sección de resultados se escogieron en base a su importancia en el proceso de diferenciación eritroide. Los mapas de calor se generaron usando valores *logCPM* de genes pertenecientes a las categorías funcionales de interés y visualizados mediante el paquete *Complex.Heatmap* de *R*. Por último, los gráficos de barra horizontales fueron generados usando la lista de genes de cada grupo funcional de interés y sus respectivos *logFC* mediante el paquete *ggplot2*.

4 RESULTADOS

4.1 Análisis bioinformático de secuencias transcriptómicas obtenidas mediante *RNA-seq*

El presente estudio transcriptómico fue realizado con objetivo de estudiar la función molecular del gen *Was* dentro del contexto eritroide, comparando las líneas celulares: MEL, MEL-R, MEL *Was*^{-/-}, y MEL *Btk*^{-/-}.

La línea celular eritroleucémica MEL deriva de progenitores eritroides cuya diferenciación ha sido interrumpida o bloqueada por el complejo vírico Friend (Friend et al., 1966; Marks y Rifkind, 1978). La integración de este complejo vírico, formado por el SFFV y el F-MuLV, bloquea el proceso de diferenciación en el estadio de proeritroblasto. Algunos compuestos químicos como el HMBA o el DMSO actúan de un modo aún desconocido provocando el reinicio del programa de diferenciación. Estas características han convertido a las células MEL en un sistema idóneo para el estudio de las bases moleculares de la diferenciación eritroide, así como del mecanismo de acción de algunos agentes inductores del proceso, que podrían eventualmente utilizarse como agentes terapéuticos en procesos tumorales. La línea celular MEL-R, derivada de MEL, es incapaz de retomar el proceso de diferenciación en presencia de agentes inductores. En tanto que las líneas celulares MEL *Was*^{-/-} y MEL *Btk*^{-/-}, también derivadas de la línea parental MEL, poseen delecciones bialélicas de los genes que codifican para las proteínas WASp y BTK, respectivamente. Dichas delecciones fueron llevadas a cabo mediante la estrategia CRISPR/Cas9 y consistieron en la eliminación total o parcial de los genes *Was* y *Btk*, respectivamente.

En este primer apartado se describe el proceso de validación de las deleciones de estas líneas celulares y se analizan algunos parámetros relacionados con la calidad y correlación de las muestras.

4.2 Validación de la deleción bialélica de los genes *Was* y *Btk* en células MEL

La validación de deleciones bialélicas de los genes *Was* y *Btk* en las muestras de MEL *Was*^{-/-} y MEL *Btk*^{-/-} se realizó mediante PCR y los amplicones se analizaron empleando la técnica de electroforesis en geles de agarosa según las condiciones descritas en *Materiales y Métodos 3.3.1*.

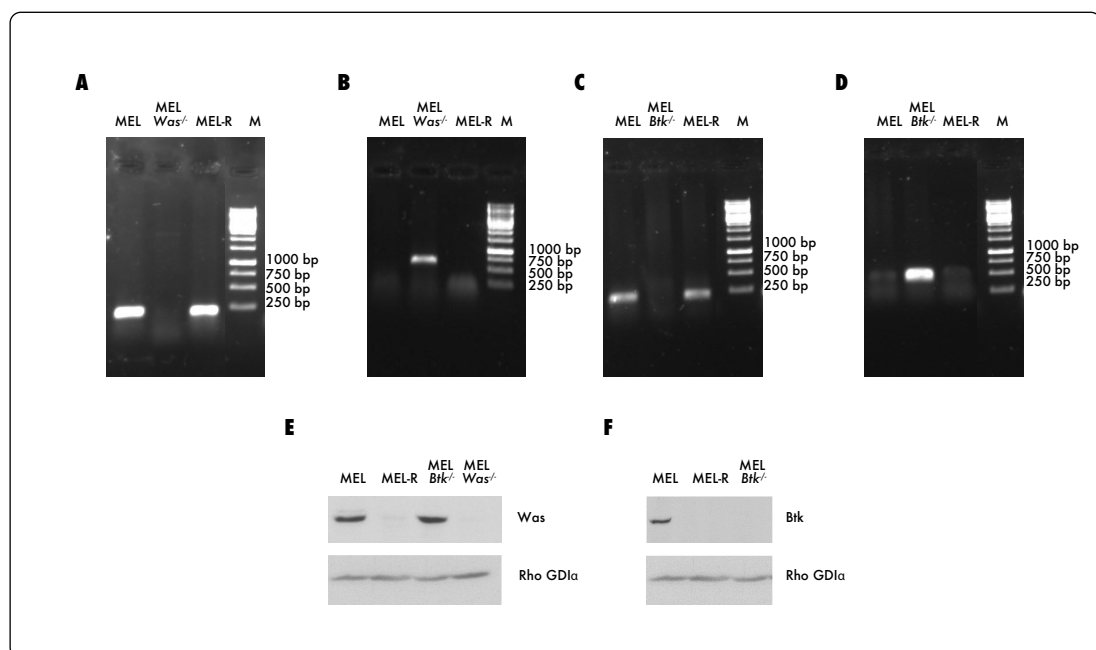


FIGURA 7. VALIDACIÓN EXPERIMENTAL DE LA DELECCIÓN BIALÉLICA DE LOS GENES *Was* Y *Btk*. Electroforesis en geles de agarosa de los productos de PCR con cebadores de no deleción (A) y deleción (B) de *Was* y cebadores de no deleción (C) y deleción (D) de *Btk*. Inmunodetección de WASp (E) y BTK (F) en extractos proteicos totales de MEL, MEL-R, MEL *Was*^{-/-} y MEL *Btk*^{-/-} (E). Se utilizó Rho GDIα como control de carga en ambas inmunodetecciones.

En la figura 7 A se observan bandas tanto en MEL como en MEL-R, las cuales corresponden a productos de la amplificación con los cebadores de la no delección de *Was*, es decir oligonucleótidos que amplifican regiones internas con respecto a los sitios de unión de los *sgRNAs*. El tamaño esperado de la banda correspondiente al producto de la amplificación con cebadores de la no delección es de 191 pb. Por el contrario, en MEL *Was*^{-/-} no se observaron bandas, indicando la imposibilidad de los cebadores de anillar en las regiones delecionadas. En la figura 7 B se muestran los productos de la amplificación con cebadores de la delección de *Was*. La delección del gen *Was* permite amplificar fragmentos de 605 pb y su presencia confirma la delección de *Was*. La banda observada en MEL *Was*^{-/-} corresponde a la amplificación de fragmentos que contienen regiones situadas corriente arriba y corriente abajo de los sitios de escisión de las Caspasa 9, mientras que en MEL y MEL-R no se observan bandas correspondientes a productos de amplificación con los cebadores de la delección. En la figura 7 C, se muestran los productos de la amplificación con cebadores de la delección de *Btk*. El tamaño esperado de los productos amplificados con cebadores de la no delección de *Btk* es de alrededor de 203 pb. Tanto en MEL como en MEL-R se observaron bandas correspondientes a amplificaciones del gen *Btk* con cebadores de la no delección. Sin embargo, no se observaron bandas en la muestra MEL *Btk*^{-/-}. En la figura 7 D se muestran los productos de las amplificaciones con cebadores de la delección de *Btk*. Mientras que en MEL y MEL-R no se observaron bandas, en MEL *Btk*^{-/-} se observó una banda de aproximadamente 672 pb, la cual corresponde a regiones corriente arriba y corriente de los sitios de escisión de las Caspasa 9 en *Btk*. La presencia de esta banda confirmó la delección de *Btk*.

Con el fin de confirmar la ausencia de expresión a nivel de proteína, se realizó la inmunodetección de las proteínas codificadas por *Was* y *Btk*. En la figura 7 E y F se muestran los resultados de la inmunodetección. La ausencia de bandas en MEL *Was*^{-/-} y MEL *Btk*^{-/-} confirma la ausencia de ambas proteínas, y por tanto la delección exitosa de los genes de *Was* y *Btk*.

Tomados en conjunto, nuestros resultados confirman la delección, y correspondiente ausencia de expresión, tanto a nivel mRNA como de proteína, de los genes *Was* y *Btk* mediante la estrategia CRISPR/Cas9.

4.3 Descripción y análisis de calidad de las secuencias obtenidas

El resultado de las mediciones para cuantificar y determinar la pureza de las muestras se observa en la tabla S3. Allí se muestran las lecturas de la densidad óptica (OD) a 260 y 280 nm, la razón de pureza (OD 260/280 y 260/230) y las concentraciones de RNA obtenidas. La integridad de las muestras se analizó mediante electroforesis en geles de agarosa, tal como se muestra en la figura S1.

En la tabla 1 se muestra la descripción de los resultados de la secuenciación del transcriptoma de las muestras de MEL *Btk*^{-/-}, MEL *Was*^{-/-}, MEL-R y MEL. En todas las muestras se obtuvo un número de lecturas entre 19 y 25 millones. La longitud promedio de lecturas fue de 75 pb, la cual es la recomendada para análisis transcriptómicos.

TABLA 1. DISTRIBUCIÓN DE LAS LECTURAS POR MUESTRA TENIENDO EN CUENTA LA LONGITUD Y NÚMERO DE LECTURAS

Muestra	Número de lecturas	Longitud (pb)	Fecha de secuenciación
MEL <i>Btk</i> ^{-/-} 2	22.180.613	75	11-02-2020
MEL <i>Was</i> ^{-/-} 1	20.922.350	75	11-02-2020
MEL <i>Was</i> ^{-/-} 2	19.783.414	75	11-02-2020
MEL-R	16.715.855	75	28-04-2012
MEL	25.197.684	75	28-04-2012

4.4 Análisis de correlación entre muestras estudiadas

Con el objetivo de comprender mejor la relación entre muestras, se realizó un análisis global de los perfiles de las librerías de *RNA-seq* normalizadas. La distancia representada por correlación de Pearson en la matriz determina la variabilidad entre grupos y entre muestras del mismo grupo, tomando en cuenta la expresión génica de todos los genes presentes en el transcriptoma. Una mayor distancia entre muestras puede ser interpretada como cambios importantes en el *fold change* en un conjunto de genes que probablemente sean los responsables de clasificar a las muestras en distintos grupos.

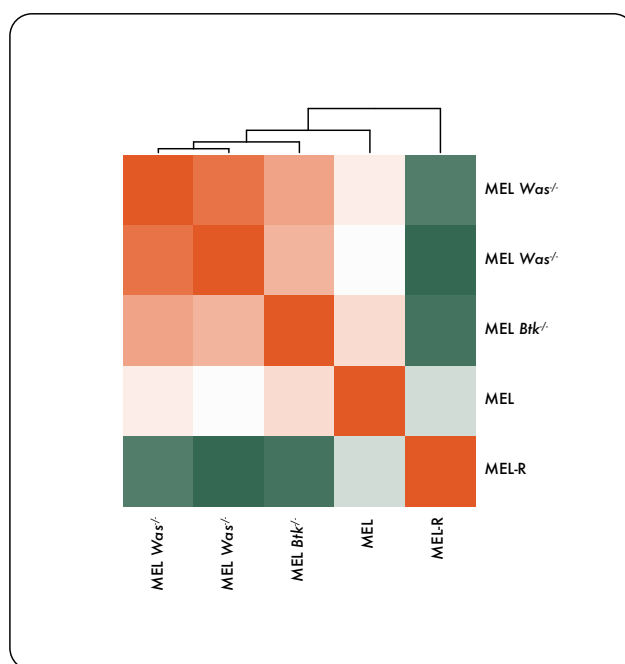


FIGURA 8. ANÁLISIS DE CORRELACIÓN DE LOS TRANSCRIPTOMAS ESTUDIADOS. Clasificación jerárquica de conjunto de datos de expresión génica de MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R.

En la figura 8 se observa un alto grado de correlación en las muestras de MEL *Was*^{-/-}, lo cual indica el alto grado de similitud de sus perfiles de expresión génica, confirmando que ambas muestras pertenecen al mismo grupo. Además, la agrupación de las muestras estudiadas refleja en gran parte el diseño experimental, mostrando en

un nivel, la línea parental MEL, de la cual se desprenden las líneas celulares derivadas de MEL: MEL *Was*^{-/-} y MEL *Btk*^{-/-}. Contrariamente a lo esperado, teniendo en cuenta que los genes *Btk* y *Was* están inactivos en las células MEL-R (Fernández-Callejas et al., 2019), se observó que las líneas MEL *Was*^{-/-} y MEL *Btk*^{-/-} no se agruparon con MEL-R, lo que sugiere un bajo nivel de similitud en términos de expresión génica entre estas muestras.

También es importante destacar la gran similitud a nivel de expresión génica global entre los grupos MEL *Was*^{-/-} y MEL *Btk*^{-/-}, lo que refuerza la idea de una posible interacción reguladora entre estos dos genes (Callejas, 2017).

4.5 Identificación de genes diferencialmente expresados

Debido a la diferencia de tamaño de las librerías y al hecho de que las muestras procesadas pertenecen a corridas de secuenciación diferentes, las muestras fueron normalizadas por el método *TMM*. El objetivo de esta normalización es remover el *batch effect*, según lo descrito en el apartado 3.4.4 de *Materiales y métodos* y en el código anexado a la presente tesis en el apartado de Anexos. Posterior a la normalización, se removieron genes con baja o nula expresión, quedando un total de 13.330 genes que se consideraron para el análisis de expresión diferencial.

Con la finalidad de identificar genes diferencialmente expresados que puedan elucidar nuevas funciones de *Was* en el contexto eritroide, se realizó la comparación de grupos por pares en base a expresión génica de MEL *Was*^{-/-} y MEL, según lo descrito en *Materiales y métodos 3.4.4*, tomando a MEL como condición de referencia.

Se identificaron un total de 2.457 genes diferencialmente expresados entre MEL y MEL *Was*^{-/-}. El criterio para la identificación de genes diferencialmente expresados fue un valor ajustado de *p* por *Benjamini-Hochberg* < 0.05 y un *fold change* ≥ |2|. Como se muestra en la figura 9 A, del total de 2.457 GDEs, 1.430 correspondieron a genes cuya expresión aumenta en MEL *Was*^{-/-} y 1.027 a genes con una disminución de

expresión en MEL *Was*^{-/-} con respecto a MEL. Además, con el fin de validar los resultados del *RNA-seq* se realizó un análisis cuantitativo utilizando RT-qPCR a tiempo real (Fig. 9 B). Los resultados de la validación confirmaron el cambio y la tendencia en la expresión de los genes analizados.

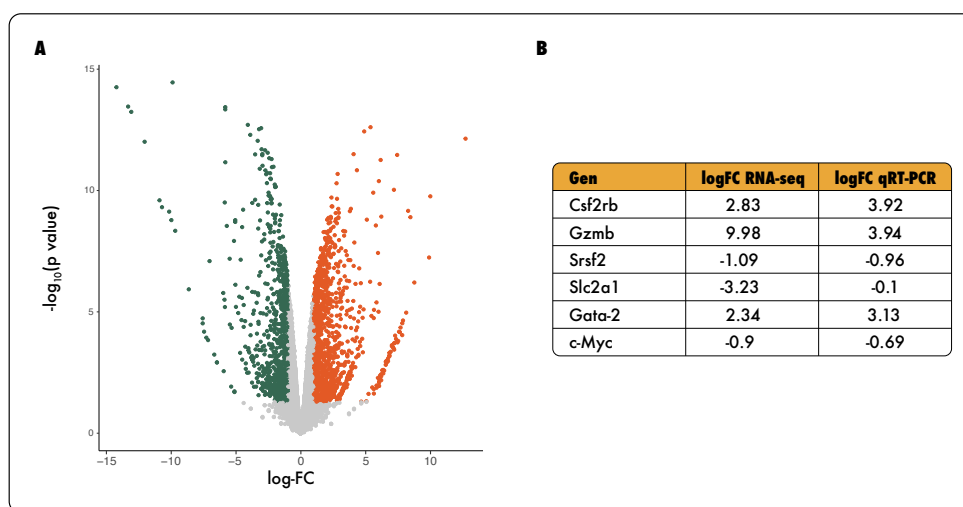


FIGURA 9. GENES CUYA EXPRESIÓN DIFIERE SIGNIFICATIVAMENTE ENTRE MEL Y MEL *Was*^{-/-}. Predicción de genes cuya expresión difiere significativamente entre MEL y MEL *Was*^{-/-} (A) se muestra la significancia estadística con respecto a la magnitud de la diferencia en la expresión de cada gen perteneciente a los transcriptomas MEL y MEL *Was*^{-/-}. Los puntos en color naranja corresponden a genes sobreexpresados en MEL *Was*^{-/-}, mientras que los puntos en color verde corresponden a genes cuya expresión aumenta en MEL. Los puntos en color gris hacen referencia a genes que no presentan cambios significativos. El punto de corte para la identificación de GDEs fue: *fold change* mayor o igual a dos y valor de *p* menor a 0.05. Validación mediante RT-qPCR ciertos genes obtenidos mediante el análisis de expresión diferencial (B).

4.6 Búsqueda de la funcionalidad de los genes identificados mediante expresión diferencial

Con el objetivo de identificar funciones biológicas y moleculares de *Was* en estadios finales de la eritropoyesis, los genes diferencialmente expresados fueron

categorizados en grupos funcionales usando diferentes métodos. En primer lugar, los genes diferencialmente expresados fueron analizados por separado en dos grupos según su expresión en MEL *Was*^{-/-}: genes sobreexpresados y genes cuya expresión disminuye significativamente. Esta categorización previa facilitó la identificación de grupos funcionales y vías metabólicas involucradas en la respuesta transcripcional de las células eritroleucémicas a la ausencia de *Was*.

4.6.1 Anotación funcional de genes diferencialmente expresados mediante *ORA* usando la base de datos KEGG

La utilización de la base de datos KEGG permitió realizar la anotación funcional de los 2.457 genes diferencialmente expresados en MEL *Was*^{-/-} con respecto a MEL. En la figura 10 se muestran los grupos funcionales identificados (test de Fisher, valor ajustado de *p* Benjamini-Hochberg < 0.05) en KEGG, los cuales están asociados con genes sobreexpresados (Fig. 10 A) y genes con disminución de la expresión en MEL *Was*^{-/-} (Fig. 10 B).

Dentro de los grupos funcionales sobreexpresados en MEL *Was*^{-/-}, se observaron algunos relacionados a la diferenciación eritroide, tales como apoptosis, degradación lisosomal, vía de señalización MAPK, los cuales contribuyen al proceso de maduración a través de la enucleación en estadios tardíos (Keerthivasan et al., 2011). También se identificó la vía de señalización *Notch*, la cual está asociada a la inducción de muerte celular en la diferenciación eritroide (Roberto-Moreno et al., 2007).

En cuanto a los grupos funcionales cuya expresión disminuye, se observaron grupos mayormente relacionados con metabolismo energético, tales como metabolismo del carbono, glicólisis y gluconeogénesis, síntesis de aminoácidos, metabolismo del piruvato, entre otros, así como otros procesos importantes para el contexto eritroide, entre los que destacan la vía de señalización HIF-1 y p53.

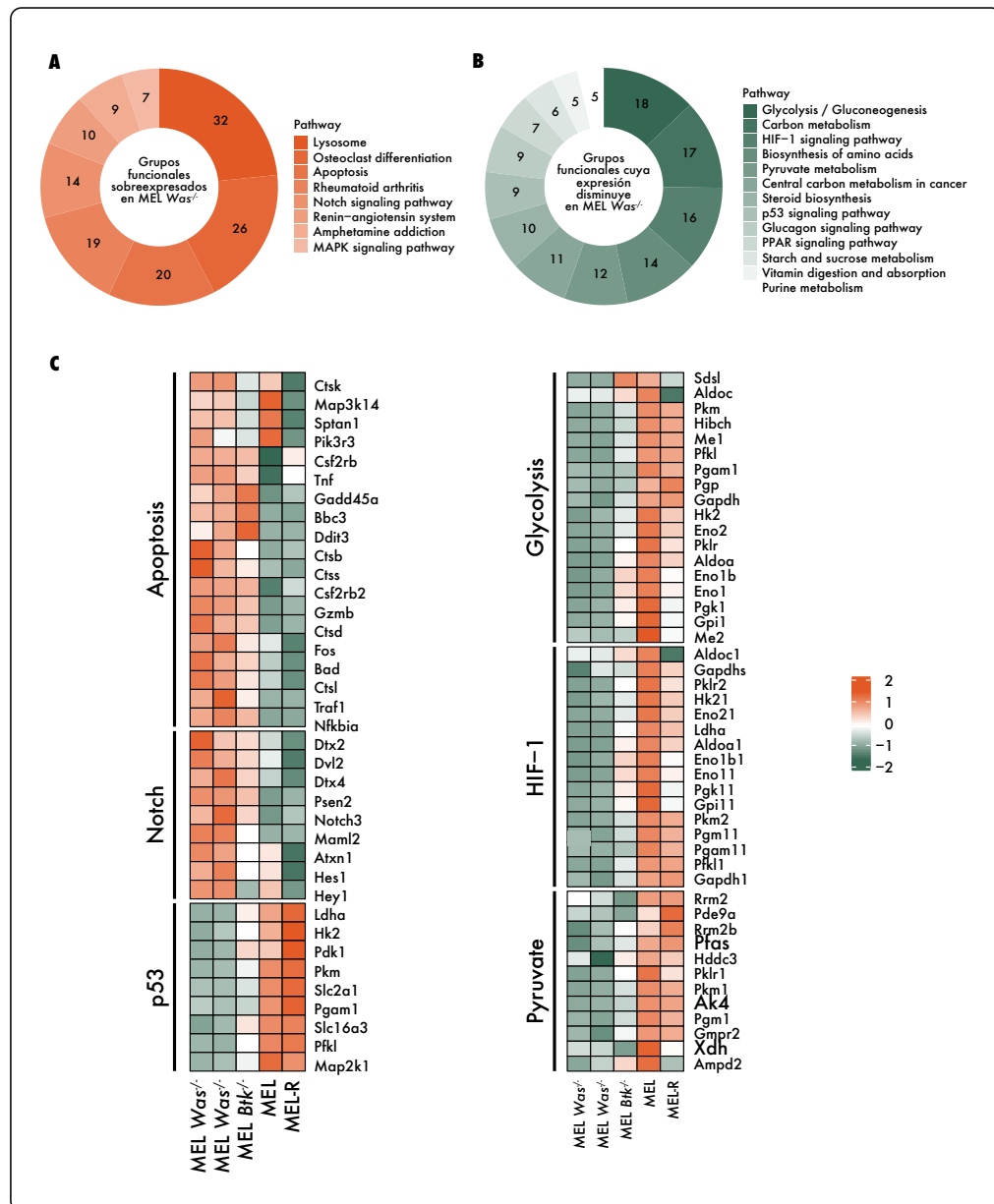


FIGURA 10. CLASIFICACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS EN CATEGORÍAS FUNCIONALES DE KEGG. Los genes diferencialmente expresados entre MEL *Was*^{-/-} y MEL fueron comparados contra un total de 543 categorías funcionales de KEGG. Representación de grupos funcionales cuya expresión génica aumenta (A) o disminuye en MEL *Was*^{-/-} con respecto a MEL (B). Los números dentro de cada gráfico representan el número exacto de genes presente en cada categoría. Los mapas de calor muestran la expresión de GDEs en todas las muestras secuenciadas, MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R

(C). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representada por $\log\text{CPM}$.

En la figura 10 C se muestran los perfiles de expresión para todos los grupos secuenciados: MEL, MEL-R, MEL *Was*^{-/-} y MEL *Btk*^{-/-}, en donde se confirmó el aumento de la expresión de ciertos grupos funcionales como apoptosis y *Notch*; así como la disminución de la expresión génica en otros grupos como la vía de señalización de p53, metabolismo del piruvato, glicolisis y HIF-1 en MEL *Was*^{-/-} y MEL *Btk*^{-/-}. Es importante destacar que a lo largo de los diferentes grupos funcionales analizados se observó un patrón de expresión bastante similar entre los grupos MEL *Was*^{-/-} y MEL *Btk*^{-/-} por un lado, y por el otro MEL y MEL-R.

En conjunto, nuestros resultados sugieren que la ausencia de *Was* estaría asociada a un aumento en la expresión génica de ciertas vías de señalización como la de apoptosis y *Notch*; y una caída general de la expresión de vías relacionadas con el metabolismo celular. Además, se comprobó que los perfiles de expresión de células MEL *Was*^{-/-} y MEL *Btk*^{-/-} presentan un alto grado de similitud.

4.6.2 Anotación funcional de genes diferencialmente expresados empleando el método de GSEA

Con el fin de aumentar la fiabilidad del estudio transcriptómico, se analizaron los datos mediante GSEA, otra técnica de análisis funcional ampliamente utilizada. La principal ventaja de GSEA es que no realiza el análisis en base a genes diferencialmente expresados, sino que toma en cuenta todos los genes dentro del conjunto de datos de manera individual y los asocia con grupos funcionales según una puntuación de enriquecimiento.

Nuestro conjunto de datos fue comparado contra la colección de firmas genéticas (50 conjuntos de genes) de GSEA. En las figuras 11 A y B se muestran las 22 categorías funcionales identificadas en la colección de firmas genéticas (valor ajustado de p

Benjamini-Hochberg < 0.25, según lo recomendado por GSEA). Entre los grupos funcionales enriquecidos positivamente (Fig. 11 A) se encontraron algunos relacionados a muerte celular: apoptosis, vía de señalización TNFA, interferón alfa y gamma, y las vías de señalización *Notch*, IL6 y JAK-STAT

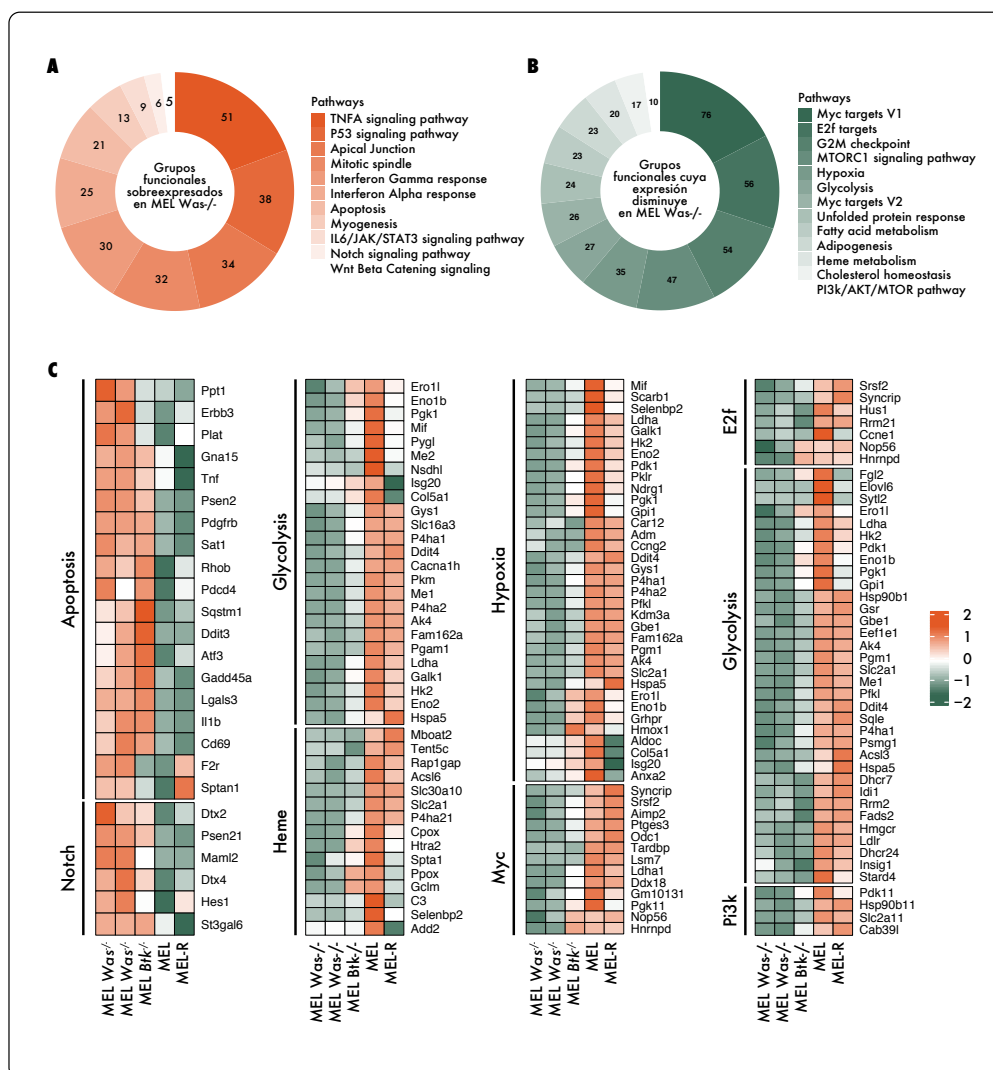


FIGURA 11. CLASIFICACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS USANDO LA COLECCIÓN DE FIRMAS GENÉTICAS DE GSEA. Los transcriptomas de MEL y MEL *Was*^{-/-} fueron analizados mediante el *Gene Set Enrichment Analysis* contra un total de 50 firmas genéticas. Se muestran los grupos de genes con sobreexpresión (A) y aquellos cuya expresión disminuye significativamente en MEL *Was*^{-/-} con respecto a MEL (B). Los números dentro de cada gráfico representan el número exacto de genes presente en cada categoría. Los mapas de calor muestran la expresión de genes anotados

funcionalmente en MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R (C). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representada por *logCPM*.

Contrariamente a lo encontrado en el análisis de *ORA* usando *KEGG*, se observó que la vía de señalización de p53 está sobreactivada en las células MEL *Was*^{-/-}, sugiriendo que los subconjuntos de genes relacionados a p53 están regulados en formas opuestas (regulación positiva y negativa) por el gen *Was*.

La figura 11 B muestra grupos funcionales enriquecidos en genes cuya expresión disminuye en las células MEL *Was*^{-/-}, entre los cuales resaltan aquellos relacionados a funciones metabólicas, tales como glicólisis, hipoxia, metabolismo de ácidos grasos, homeostasis de colesterol y biosíntesis del grupo hemo, lo cual sugiere nuevamente un rol de *Was* en la regulación del metabolismo celular de los proeritroblastos.

En la figura 11 C se observa en detalle la expresión de cada uno de los genes pertenecientes a los grupos funcionales identificados en la colección de firmas genéticas usando *GSEA*, incluyendo además MEL *Btk*^{-/-} y MEL-R. Es importante resaltar que los perfiles de expresión génica de los grupos MEL *Was*^{-/-} y MEL *Btk*^{-/-} presentan un alto grado de similitud a lo largo de las colecciones de firmas genéticas analizadas.

Por otro lado, es importante destacar que en ausencia de *Was* se observó un patrón de disminución en la expresión de las dianas de *c-Myc* y *E2f*. Esto podría sugerir una posible regulación positiva entre *Was* y estos oncogenes dentro del contexto eritroide. Por último, entre los grupos con disminución de expresión génica se identificó el grupo de genes relacionados de biosíntesis del grupo hemo (Fig. 11 C), cuyos niveles disminuyen significativamente tanto en MEL *Was*^{-/-} y MEL *Btk*^{-/-}.

Tomados en conjunto los datos sugieren que la expresión de *Was* en progenitores eritroides estaría relacionada con la regulación de vías metabólicas, la proliferación celular y la regulación de la biosíntesis del grupo hemo.

4.7 Comparación de los grupos funcionales identificados mediante métodos distintos

Con el objetivo de aumentar la fiabilidad de la anotación funcional, los resultados obtenidos usando ambos métodos fueron comparados para identificar grupos funcionales comunes. Como se observa en la figura 12, al intersectar grupos funcionales obtenidos mediante *GSEA*, usando sus firmas genéticas, y *ORA* usando *KEGG*, se encontraron cuatro grupos funcionales comunes: glicólisis, vía de señalización *Notch*, apoptosis e hipoxia (HIF-1). Estos resultados refuerzan la fiabilidad de los grupos funcionales identificados. Debido a su significancia en el contexto eritroide, los perfiles de expresión de genes pertenecientes a algunos de estos grupos, juntamente con otros grupos funcionales importantes para la diferenciación eritropoyética, fueron analizados con mayor profundidad.

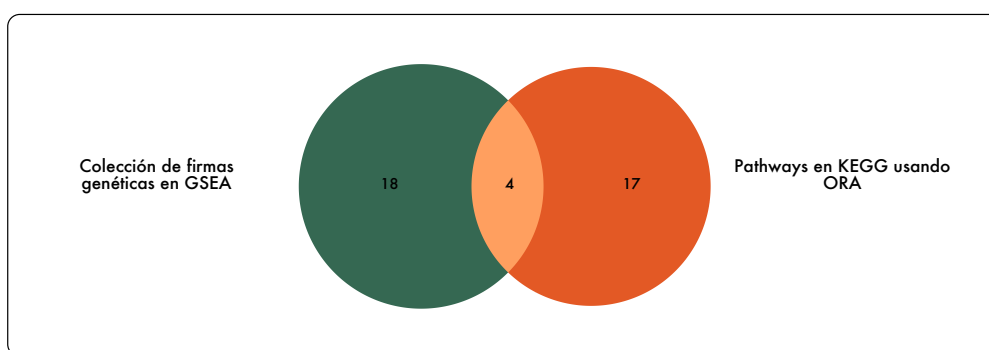


FIGURA 12. COMPARACIÓN DE LAS ANOTACIONES FUNCIONALES USANDO DIFERENTES MÉTODOS. El diagrama de *Venn* muestra grupos funcionales comunes entre las anotaciones funcionales para KEGG y GSEA, cuyos valores de significancia son inferiores a 0.05 y 0.025, respectivamente

4.8 Comparación de los grupos funcionales identificados en MEL *Was*^{-/-} con células de pacientes con síndrome de Wiskott-Aldrich

Dada la relación de *Was* con el síndrome al cual debe su nombre, también fueron comparados los resultados de GSEA obtenidos a partir de la comparación de MEL

Was^{-/-} y MEL, con resultados obtenidos de la comparación de muestras de macrófagos de pacientes con SWA y células isogénicas con un rescate de *Was* (*Materiales y métodos* 3.2.2).

Como se muestra en la figura 13, se identificaron un total de ocho grupos funcionales comunes, los cuales corresponden a las firmas genéticas de: las dianas de *c-Myc*, hipoxia, dianas de *E2f*, respuesta al interferón alfa y respuestas al interferón gamma, huso mitótico, vía de señalización TNFA y G2M *checkpoint*. Si bien estos grupos funcionales son comunes entre pacientes SWA y células eritroleucémicas con delección de *Was*, en algunos casos, la tendencia en la expresión fue distinta. Concretamente, en MEL *Was*^{-/-} los grupos funcionales relacionados con las dianas de *c-Myc*, hipoxia, dianas de *E2f*, respuesta al interferón alfa y respuestas al interferón gamma, huso mitótico, vía de señalización TNFA y G2M *checkpoint* se observó un aumento en la expresión y una disminución en las dianas *c-Myc*, *E2f*, hipoxia y el G2M *checkpoint*. Este hallazgo sugiere que, si bien WASp estaría involucrado en la regulación de estos grupos funcionales, su rol como activador o inhibidor sería dependiente del contexto.

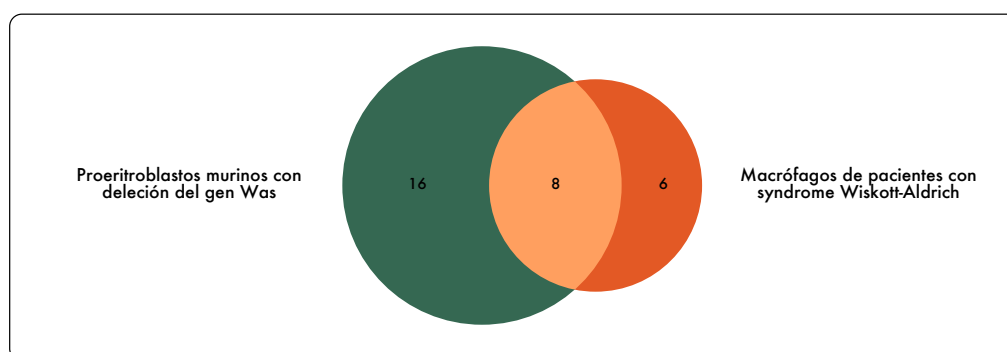


FIGURA 13. COMPARACIÓN DE LAS ANOTACIONES FUNCIONALES CON MUESTRAS DE PACIENTES CON SWA. El diagrama de *Venn* muestra grupos funcionales comunes entre los resultados de análisis de enriquecimiento de genes realizados comparando transcriptomas de MEL *Was*^{-/-} con respecto a MEL (en verde) y transcriptomas de macrófagos de pacientes con el síndrome Wiskott-Aldrich con respecto a células isogénicas con rescate del gen *Was* (en naranja). Fueron considerados como grupos significativos aquellos con un valor de *p* ajustado (método BH) inferior a 0.25.

Estos resultados demuestran que nuestras muestras, en las cuales no hay expresión de la proteína WASp, presentan un alto grado de solapamiento, en términos de grupos funcionales identificados, con muestras de pacientes con SWA. Los resultados obtenidos sugieren que las categorías funcionales identificadas comunes entre los modelos humano y murino, podrían ser consideradas como claves para el estudio del síndrome a nivel molecular.

4.9 Análisis de patrones de expresión de genes pertenecientes a las principales categorías funcionales identificadas usando transcriptomas de estadios finales de la diferenciación eritroide y modelos de reprogramación de la diferenciación en células eritroleucémicas.

Con el objetivo de contextualizar las posibles funciones de *Was* en los estadios finales de la eritropoyesis y la reprogramación de células eritroleucémicas, se compararon los perfiles transcriptómicos de los grupos funcionales identificados con transcriptomas correspondientes a estadios finales de la eritropoyesis y datos transcriptómicos obtenidos al comparar células MEL cultivadas en presencia o ausencia del agente inductor DMSO.

4.9.1 Análisis de expresión de genes relacionados con apoptosis

Uno de los principales y más estudiados procesos celulares es la apoptosis. Este proceso no sólo es importante para el balance en la producción de eritrocitos, sino también es fundamental para la maduración de éstos. En la figura 14 A se puede observar un estricto patrón de sobreexpresión de genes relacionados con apoptosis tanto en MEL *Was*^{-/-} como en MEL *Btk*^{-/-} con respecto a MEL y MEL-R, lo cual sugiere un rol de *Was* y *Btk* como reguladores negativos de la apoptosis. Al analizar la expresión de los mismos genes anotados funcionalmente en estadios finales de eritropoyesis (Fig. 14 B), se observó un incremento de la expresión en la mayoría genes analizados a lo largo de la diferenciación, lo cual coincide con estudios previos

que demostraron un aumento de la expresión de genes relacionados apoptosis en estadios finales de la diferenciación (Keerthivasan, et al. 2011).

Al comparar el mismo grupo de genes en muestras de reprogramación de células eritroleucémicas se encontró que, de manera similar a lo que observado en MEL *Was*^{-/-} (Fig. 14 C), en células MEL cultivadas en presencia del agente inductor DMSO (*MEL induced*) existe una tendencia hacia la activación de los genes analizados, con excepción de algunos genes cuya expresión disminuye (Fig. 14 D).

Tomados en conjunto, nuestros resultados demuestran que el patrón de expresión de genes relacionados apoptosis observado en MEL *Was*^{-/-} es similar al observado en los estadios finales dela diferenciación eritroide, lo cual sugiere un rol regulador de *Was* sobre la apoptosis en el contexto eritroide.

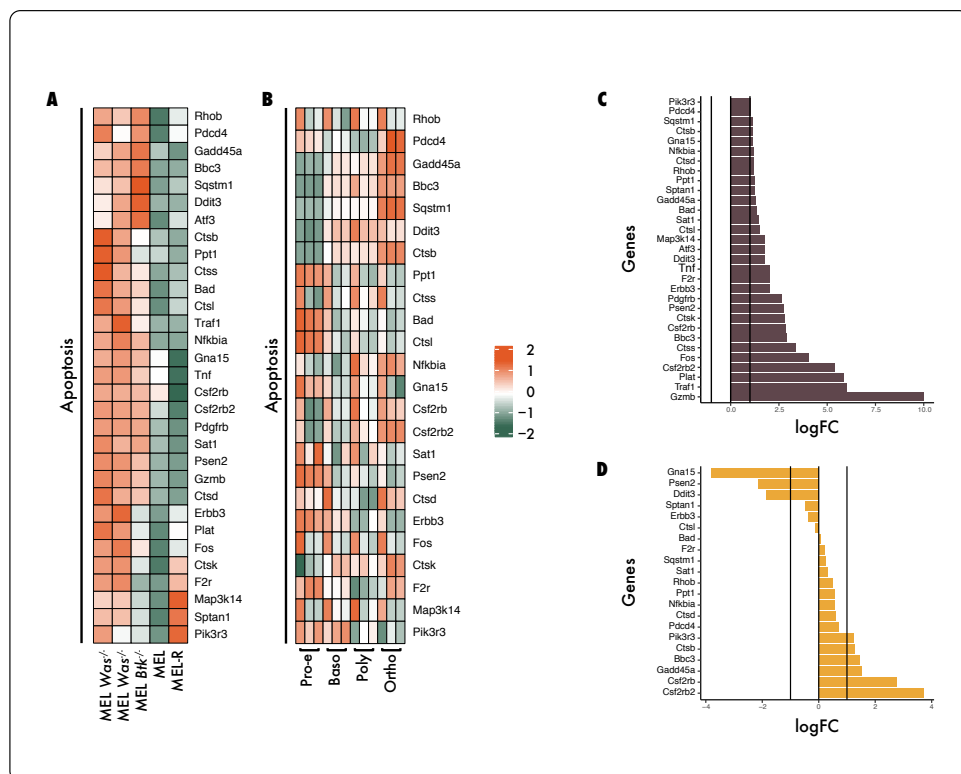


FIGURA 14. PERFIL DE EXPRESIÓN DE GENES INVOLUCRADOS EN APOPTOSIS. Los mapas de calor muestran la expresión de genes involucrados en apoptosis comparando muestras de MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R (A) y diferentes estadios de la diferenciación eritroide: proeritroblasto (Pro-e),

eritroblasto basofílico (Baso), policromatófilicos (Poly) y ortocromático (Orto) (B). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*. Gráficos de barras representado la expresión de genes involucrados en apoptosis en función al *logFC* comparando MEL *Was*^{-/-} con respecto a MEL (C) y células MEL cultivadas en ausencia (*MEL uninduced*) o presencia del agente inductor *DMSO* durante 96 horas (*MEL induced*) (D).

4.9.2 Análisis de expresión de genes pertenecientes a la vía de señalización de HIF

La hipoxia es una condición caracterizada por bajos niveles de oxígeno (Tang et al. 2013). Esta condición es el resultado de cualquier tipo de desperfecto en el suministro de oxígeno que comprometa los requerimientos biológicos. Uno de esos desperfectos es el cáncer, en dónde células tumorales incrementan de forma excesiva los requerimientos de oxígeno, causando modificaciones en la expresión de genes para alcanzar los niveles requeridos de oxígeno.

El factor HIF-1 inducido por la hipoxia (HIF, por *Hypoxia Inducible Factor*) es un complejo proteico que incrementa la expresión de genes específicos en presencia de bajas concentraciones de oxígeno. Con el objetivo de dilucidar el patrón de expresión de genes involucrados en la vía de señalización de HIF dentro del contexto eritroide, se procedió a la comparación del cambio de expresión de los genes identificados entre las células eritroleucémicas de nuestro modelo (MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R) y en los diferentes estadios de la maduración final de la eritropoyesis (proeritroblasto, eritroblasto basofílico, policromatófilicos y ortocromático).

Como se muestra en la figura 15 A, los genes involucrados en la señalización HIF, cuya expresión disminuye significativamente en MEL *Was*^{-/-} y MEL *Btk*^{-/-}, presentan un patrón de expresión opuesto en MEL y MEL-R. En el contexto de diferenciación eritroide, nuestros resultados muestran un silenciamiento paulatino en la mayoría de los genes relacionados a HIF a lo largo de los estadios finales (Fig. 15 B). Estos resultados demuestran que el perfil de expresión de genes relacionados con HIF de las

células MEL *Was*^{-/-} y MEL *Btk*^{-/-} es similar al observado en los estadios terminales de la eritropoyesis.

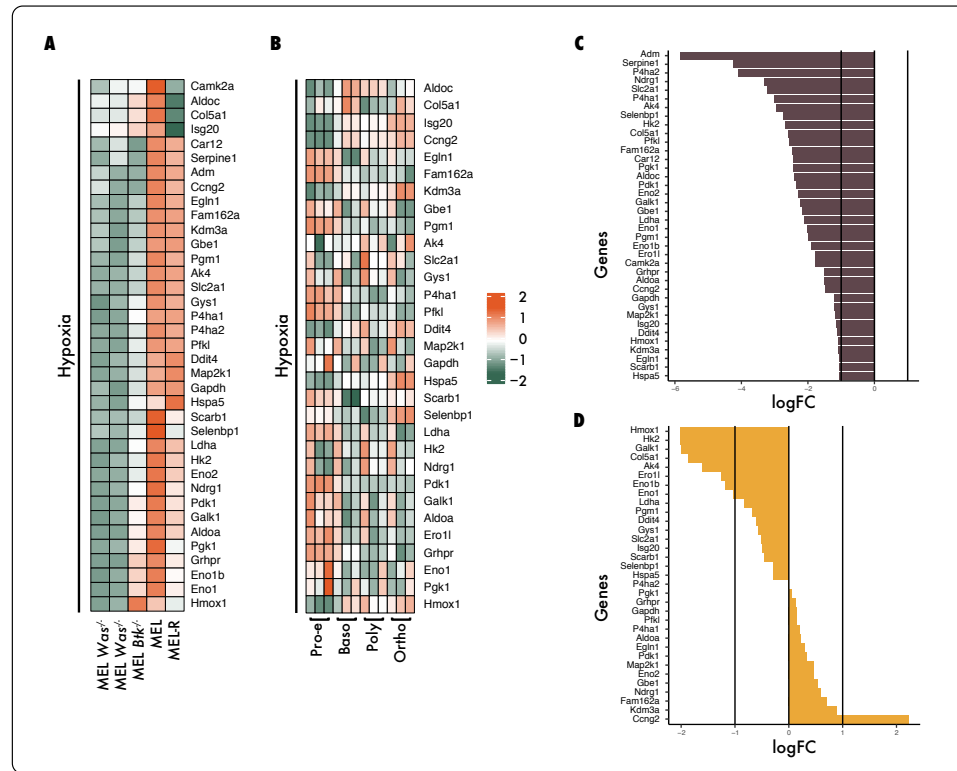


FIGURA 15. PERFIL DE EXPRESIÓN DE GENES INVOLUCRADOS EN HIPOXIA (VÍA DE SEÑALIZACIÓN DE HIF-1). Los mapas de calor muestran la expresión de genes involucrados en apoptosis comparando muestras de MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R (A) y diferentes estadios de la diferenciación eritroide: proeritroblasto (Pro-e), eritroblasto basofílico (Baso), policromatofílicos (Poly) y ortocromático (Orto) (B). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*. Gráficos de barras representado la expresión de genes involucrados en apoptosis en función al *logFC* comparando MEL *Was*^{-/-} con respecto a MEL (C) y células MEL cultivadas en ausencia (*MEL uninduced*) o presencia del agente inductor *DMSO* durante 96 horas (*MEL induced*) (D).

Por último, también se compararon los perfiles de expresión de la vía de señalización de HIF con muestras del modelo de reprogramación eritroleucémica. En la figura 15 C se muestra una clara desregulación del perfil de expresión, con una tendencia a la disminución de genes de la vía de señalización HIF en la ausencia de *Was*, lo cual

sugiere una relación de regulación positiva entre *Was* y HIF, mientras que en células MEL inducidas a la diferenciación, no se observó una tendencia clara con respecto a la expresión de genes relacionados a hipoxia.

Nuestros resultados demuestran que el patrón de expresión de genes de la vía de señalización HIF-1 observado en MEL *Was*^{-/-} es similar al observado en los estadios finales de la diferenciación eritroide.

4.9.3 Análisis de expresión de genes diana de *c-Myc*

La familia de genes MYC está compuesta por *c-Myc*, *n-Myc* y *l-Myc*. Los integrantes de MYC son factores de transcripción que juegan un rol importante en la regulación de la diferenciación y proliferación celular. Las anomalías en la expresión de miembros de MYC resultan en la promoción y mantenimiento de células tumorales en numerosos tipos de cáncer, incluidas la leucemia linfóide y mieloide (Xu et al. 2017).

En la figura 16 A, se observa la disminución de la expresión de dianas de *c-Myc* en células MEL con deleciones bialélicas de *Was* y *Btk*, con respecto a MEL y MEL-R. En lo que respecta a los estadios finales de la diferenciación eritroide, se observó una tendencia a la disminución de la expresión de las dianas de *c-Myc* conforme al avance hacia estadios finales, tal como se muestra en la figura 16 B. En ese sentido, nuestros resultados sugieren que el perfil de expresión de las dianas de *c-Myc* en células MEL *Was*^{-/-} se asemeja a los estadios finales del proceso de diferenciación eritroide.

De manera similar a lo observado en MEL *Was*^{-/-} (Fig. 16 C), la expresión de dianas de *c-Myc* sufre una notable disminución de expresión en células MEL diferenciadas (Fig. 16 D). Estos resultados demuestran que el patrón de expresión de genes dianas de *c-Myc* observado en MEL *Was*^{-/-} es similar al observado en los estadios finales de la diferenciación eritroide y sugiriendo un rol de *Was* como regulador positivo de *c-Myc* dentro del contexto eritroide.

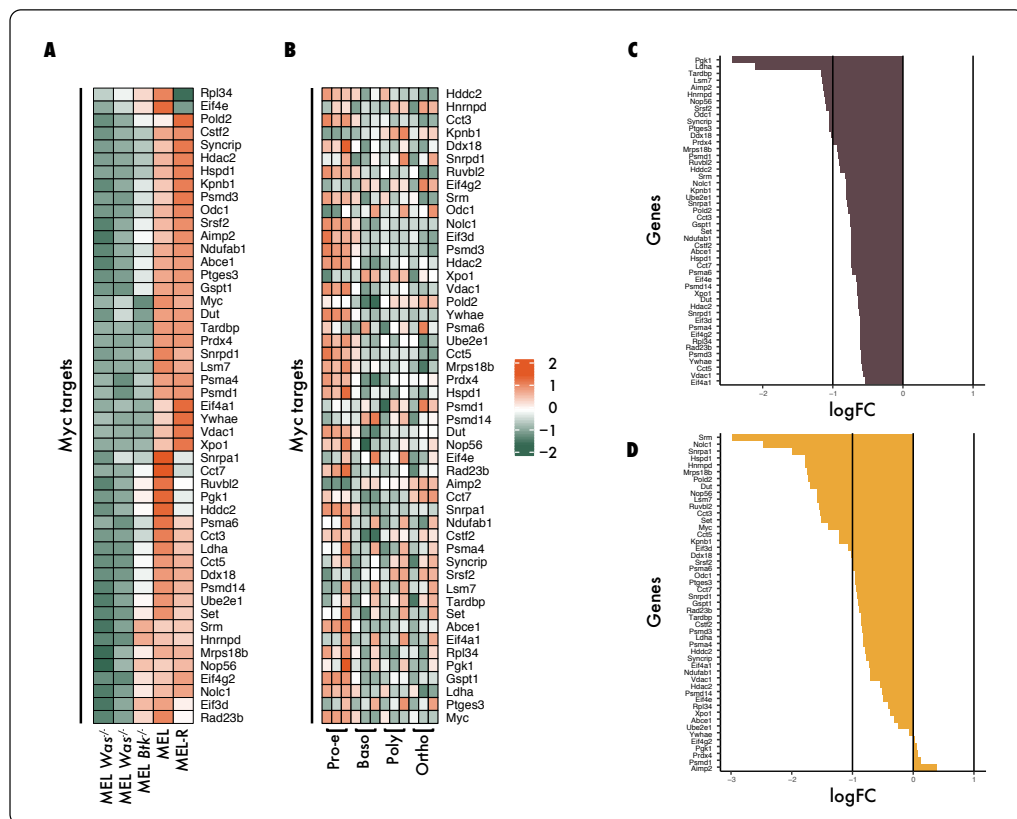


FIGURA 16. PERFIL DE EXPRESIÓN DE DIANAS DE *c-MYC*. Los mapas de calor muestran la expresión de dianas de *c-Myc* comparando muestras de MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R (A) y diferentes estadios de la diferenciación eritroide: proeritroblasto (Pro-e), eritroblastos basofílico (Baso), policromatofílicos (Poly) y ortocromático (Orto) (B). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*. Los gráficos de barras muestran la expresión de dianas de *c-Myc* comparando: MEL *Was*^{-/-} con respecto a MEL (C) y células MEL cultivadas en ausencia (MEL *uninduced*) o presencia del agente inductor *DMSO* durante 96 horas (MEL *induced*) (D).

4.9.4 Análisis de la expresión de genes que codifican para proteínas que regulan la biosíntesis del grupo hemo.

El grupo hemo es un grupo no aminoácido que contiene hierro y puede ser encontrado en una variedad de proteínas. Su síntesis consiste en un proceso enzimático de ocho pasos, cuatro de ellos mitocondriales y cuatro citosólicos (Ajioka, Phillips et

al. 2006). Entre las múltiples funciones del grupo hemo, se encuentra la de ser el componente encargado del transporte de oxígeno en la hemoglobina, proteína característica del linaje eritroide. Las disrupciones en cada uno de los ocho procesos enzimáticos están asociadas con enfermedades cuyos síntomas están bien caracterizados. En estadios finales de la diferenciación eritroide, se incrementa la demanda de hierro y del grupo hemo. Por este motivo, la síntesis del grupo hemo está ligada a la maduración final de eritroblastos (Chung, Chen et al. 2012).

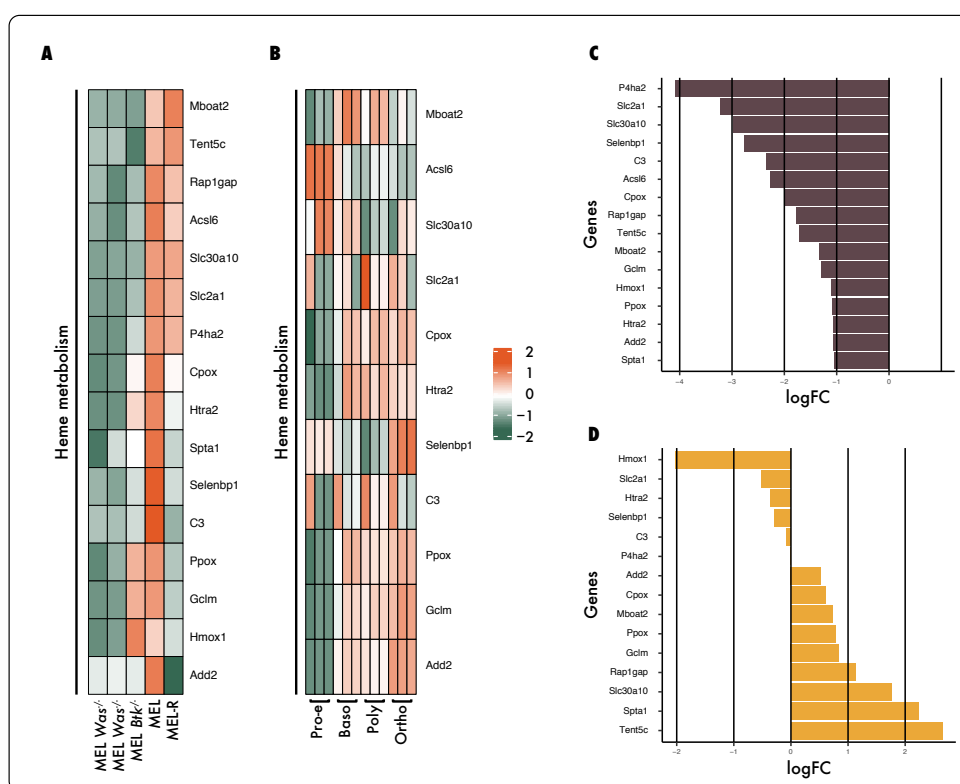


FIGURA 17. PERFIL DE EXPRESIÓN DE GENES INVOLUCRADOS EN EL METABOLISMO DEL GRUPO HEMO. Los mapas de calor muestran la expresión de genes responsables del metabolismo del grupo hemo, comparando muestras de MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R (A) y diferentes estadios de la diferenciación eritroide: proeritroblasto (Pro-e), eritroblasto basofílico (Baso), policromatofílicos (Poly) y ortocromático (Ortho) (B). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*. Los gráficos de barras muestran la expresión de genes del metabolismo del grupo hemo comparando MEL *Was*^{-/-} con respecto a MEL (C) y células MEL cultivadas en ausencia (*MEL uninduced*) o presencia del agente inductor *DMSO* durante 96 horas (*MEL induced*) (D).

Al comparar la expresión de genes involucrados en la biosíntesis del grupo hemo se observó una disminución general de la expresión en los grupos *MEL Was*^{-/-} y *MEL Btk*^{-/-}, que fue parcial en el caso de MEL-R, con respecto a la línea parental MEL (Fig. 17 A). Tal como era de esperar, se constató un incremento en la expresión de genes relacionados a la biosíntesis del grupo hemo a lo largo de la diferenciación eritroide (Fig. 17 B).

Al contrario de lo observado en ausencia de *Was* (Fig. 17 C), en células MEL no inducidas e inducidas a la diferenciación con DMSO, se observó una tendencia hacia el aumento de expresión de genes involucrados en la biosíntesis del grupo hemo (Fig. 17 D), en concordancia con lo observado en estadios terminales de la diferenciación eritroide. Nuestros datos demuestran una tendencia opuesta en el patrón de expresión de genes involucrados en la biosíntesis de grupo hemo en *MEL Was*^{-/-} y en los estadios finales de la diferenciación eritroide, lo cual sugiere que *Was* estaría relacionado con la síntesis del grupo hemo.

4.10 Comparación de genes diferencialmente expresados en *MEL Was*^{-/-} y MEL-R con respecto a MEL

Un estudio previo de *RNA-Seq* identificó a *Was* como uno de los genes cuya expresión disminuye en las células MEL-R con respecto a la línea parental MEL (Fernández-Calleja, 2017). Paralelamente, se comprobó que el promotor del gen codifica para la proteína WASp está metilado y, por lo tanto, inactivo en MEL-R. En un trabajo posterior se llevó a cabo la delección de *Was* en células MEL y se demostró que, a diferencia de las células MEL-R, las células *MEL Was*^{-/-} son capaces de retomar el programa de diferenciación eritroide en presencia del agente inductor HMBA (Fernández-Calleja, 2019). Estos resultados sugieren que la ausencia de *Was* no estaría directamente relacionada con la resistencia a los agentes inductores de las células MEL-R.

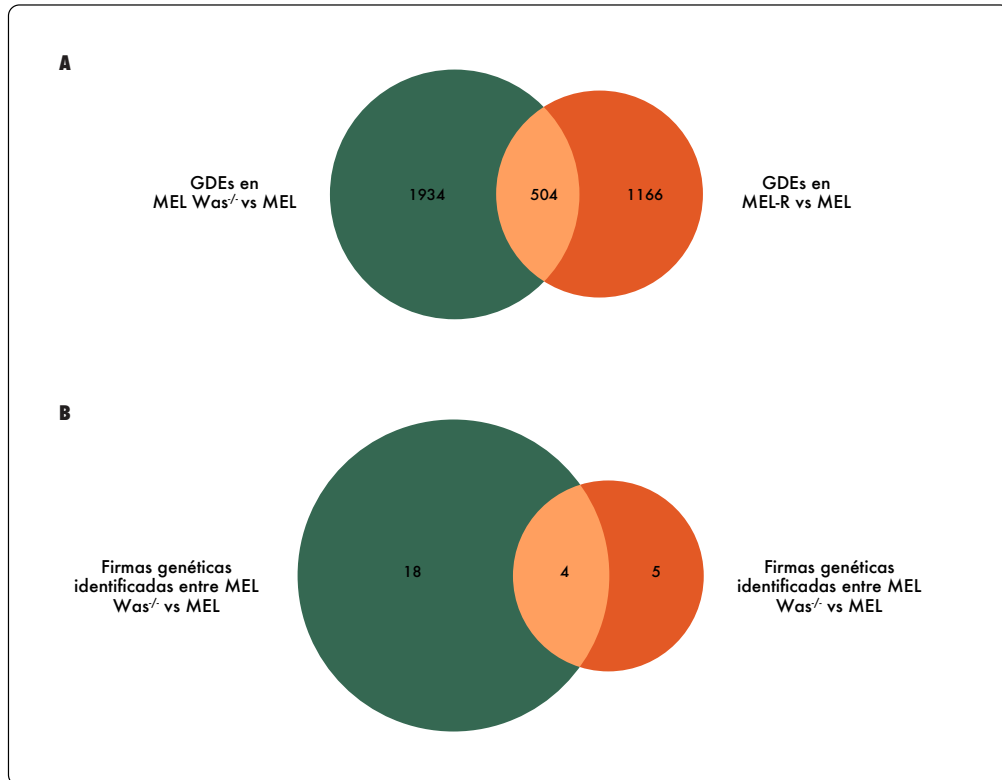


FIGURA 18. COMPARACIÓN DE GENES DIFERENCIALMENTE EXPRESADOS Y GRUPOS FUNCIONALES ENTRE MEL *Was*^{-/-}, Y MEL-R. El diagrama de *Venn* muestra el solapamiento de 504 genes diferencialmente expresados tanto entre MEL *Was*^{-/-} como en MEL-R con respecto a MEL, dejando 1934 genes exclusivos para MEL *Was*^{-/-} y 1166 exclusivos para MEL-R (A). En el diagrama de *Venn* se muestra el solapamiento de 4 grupos funcionales, obtenidos mediante GSEA (firmas genéticas), comparando MEL *Was*^{-/-} y MEL-R con respecto a MEL (B).

La ausencia de expresión de *Was*, a nivel de mRNA y proteína, es una característica compartida entre MEL *Was*^{-/-} y MEL-R. Debido a este punto en común, se procedió a comparar estos transcriptomas con respecto a la línea parental MEL, con el objetivo de conocer el grado de solapamiento entre ambas líneas celulares en términos de genes diferencialmente expresados y su anotación funcional. Como se observa en la figura 18 A, fueron identificados 504 genes comunes entre ambas comparaciones, número que representa al 26% y 30% de los GDEs identificados en MEL-R y MEL *Was*^{-/-} con respecto a MEL, respectivamente.

En cuanto a los resultados de las anotaciones funcionales, tal como se muestra en la figura 18 B, existe solapamiento de 4 firmas genéticas: unión apical, respuesta al interferón alfa, respuesta al interferón gamma, hipoxia, al comparar los perfiles de expresión entre MEL-R y MEL *Was*^{-/-} con respecto a la línea parental MEL. Sin embargo, si bien se observaron vías comunes, estos grupos funcionales están sobreactivados en MEL *Was*^{-/-} con respecto a MEL, mientras que su expresión es significativamente menor en MEL-R con respecto a MEL.

En conjunto, nuestros resultados muestran que si bien existen grupos funcionales comunes entre MEL-R y MEL *Was*^{-/-}, se observa una tendencia opuesta en los patrones de expresión de los GDEs, lo cual plantea la hipótesis de que la inactivación de *Was* en MEL-R podría estar compensada parcialmente por otros miembros de la familia WASP.

4.11 Análisis de la expresión de genes que codifican para miembros de la familia WASP

WASp es el miembro fundador de una familia de factores promotores de la nucleación de actina que incluye al menos cinco subfamilias, (1) WASp y neuronal-WASp (N-WASp; también conocido como WASL), (2) las tres isoformas homólogas de verprolina de la familia WASP (WAVE1 – WAVE3; también conocidas como SCAR1 – SCAR3 y WASF1 – WASF3), (3) el homólogo de WASp asociado con actina, membranas y microtúbulos (WHAMM), (4) el homólogo WASp y SCAR (WASH; también conocido como WASHC1), y (5) proteína reguladora mediadora de la unión (JMY) (Alekhina et al., 2017). A pesar de que todos contienen dominios únicos, todos ellos conservan el dominio VCA requerido para la polimerización de actina. Debido al alto grado de conservación, han demostrado que los miembros N-WASp y WAVE2, pueden compensar la pérdida de WASp, asistiendo en la restauración del citoesqueleto de actina (Isaac et al., 2010; Kuznetsov et al., 2017).

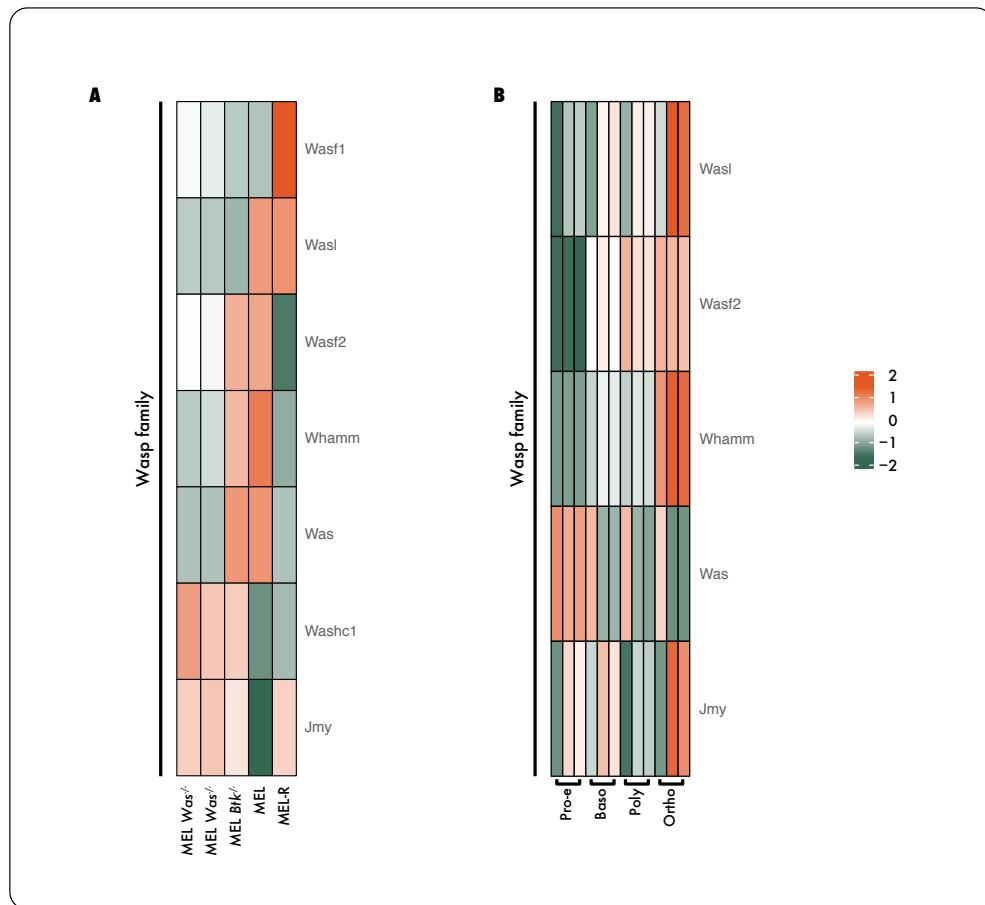


FIGURA 19. PERFIL DE EXPRESIÓN DE GENES DE LA FAMILIA WASP. El mapa de calor de la expresión de genes pertenecientes a la familia WASP en: MEL, MEL-R, *MEL Btk*^{-/-}, *MEL Was*^{-/-} (A). y diferentes estadios de la diferenciación eritroide: proeritroblasto (Pro-e), eritroblasto basofílico (Baso), policromatofílico (Poly), ortocromático (Orto) (B). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*.

Con el objetivo de identificar posibles compensaciones que tienen lugar en ausencia de *Was*, se analizó la expresión de otros integrantes de la familia WASp. En la figura 19 A, se observa el perfil de expresión en las muestras secuenciadas. En MEL *Was*^{-/-} se observa un incremento en la expresión de *Washc1* (Complejo WASH) y *Jmy* con respecto a MEL. En MEL *Btk*^{-/-} se observa que la mayoría de los miembros de la familia WASP se expresan, con excepción de *Wasf1* (WAVE1) y *Wasl* (N-WASp),

los cuales, junto con *Jmy*, muestran un aumento de expresión en MEL-R, con respecto a la línea parental. El análisis transcriptómico también demostró que existen diferencias en el patrón de expresión de genes de la familia WASP entre MEL *Was*^{-/-} y MEL-R, lo cual podría explicar las diferencias observadas en el comportamiento de los grupos funcionales identificados entre ambas muestras. Otro hallazgo importante por destacar es que los mutantes MEL *Btk*^{-/-} expresan *Was*, a niveles similares a los observados en la línea parental MEL.

En cuanto a la expresión de los miembros de familia WASP dentro del contexto de la diferenciación eritroide según se observa en la figura 19 B, existe una disminución en la expresión de *Was* en estadios finales. Esta observación estaría en línea con los resultados de los apartados previos en los cuales los perfiles de expresión de MEL *Was*^{-/-} mostraron una alta similitud con los perfiles de los estadios finales de la diferenciación, más concretamente con el estadio de eritroblasto ortocromatofílico, donde *Was* está ausente. Sin embargo, se observó un aumento en la expresión de otros miembros de la familia, como *Wasl* (N-WASp), *Wasf2* (WAVE2), *Whamm* y *Jmy*.

Luego de haber comparado el perfil de expresión de los distintos miembros la familia WASP, se analizaron funciones bien estudiadas que están asociadas a la familia WASP. Particularmente, aquellos procesos que también son importantes para la maduración de progenitores eritroides, entre ellos la autofagia y la degradación lisosomal, ambos relacionados con el proceso de enucleación. Tal como se muestra en las figuras 20 A y B, los genes relacionados a estos procesos se encuentran sobreexpresados en MEL *Was*^{-/-} con respecto a MEL, mientras que el patrón de expresión en MEL-R es similar al de la línea parental. Nuevamente se observó que las células con la delección de *Btk*, pueden recapitular el patrón de expresión génica de MEL *Was*^{-/-}.

Además, teniendo en cuenta el rol de supresor tumoral de *Was* identificado en otros linajes (Menotti, et al. 2018), se analizaron los grupos funcionales que corresponden: genes relacionados a la anafase y metafase mitótica, y genes relacionados al G2M

checkpoint. En las figuras 20 C y D se muestra que para ambos grupos funcionales se observaron disminuciones significativas tanto en MEL *Was*^{-/-} como en MEL *Btk*^{-/-} con respecto a MEL y MEL-R.

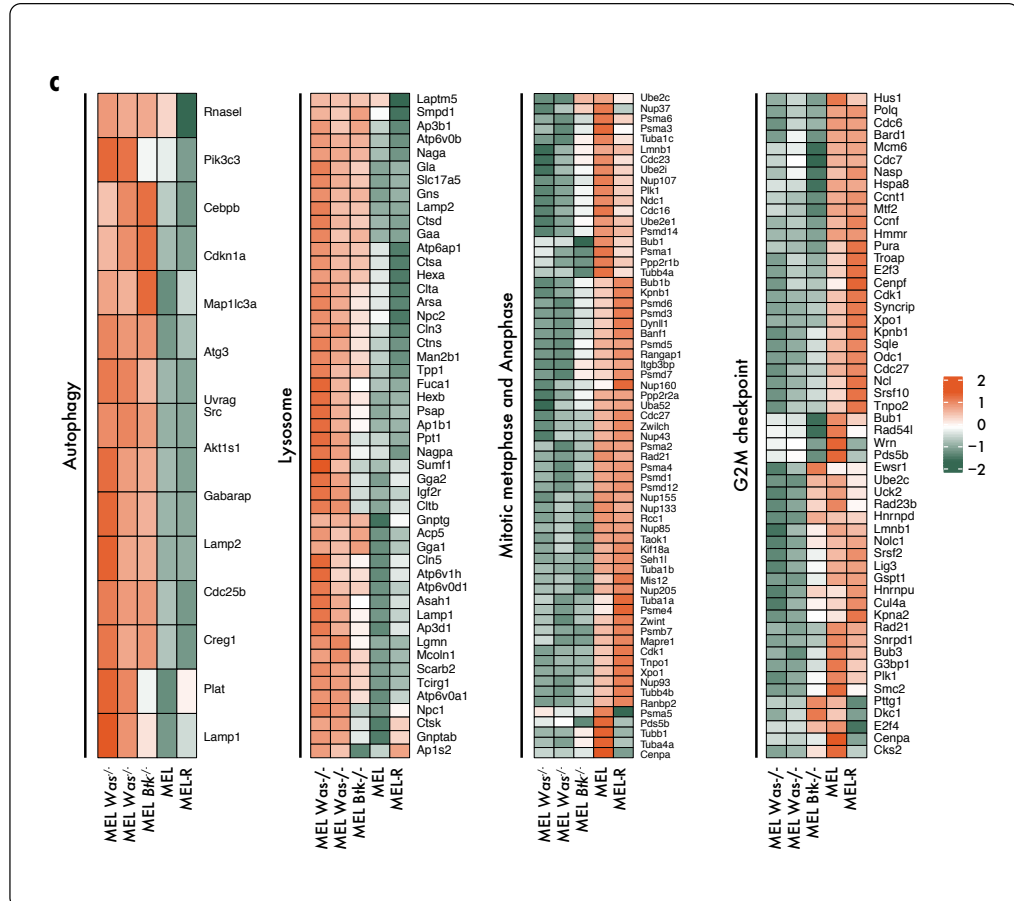


FIGURA 20. PERFIL DE EXPRESIÓN DE GRUPOS FUNCIONALES RELACIONADOS A FUNCIONES YA IDENTIFICADAS DE LA FAMILIA WASP. El mapa de calor de la expresión de genes involucrados en autofagia (A), lisosomas (B), anafase y mitofase (C) y G2M checkpoint (D), en las muestras de MEL, MEL-R, MEL *Btk*^{-/-}, MEL *Was*^{-/-}. La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*.

En conjunto, nuestros resultados sugieren que la inactivación de *Was* en MEL-R podría estar parcialmente compensada por la activación de otros miembros de la familia WASP, en particular por genes que codifican para las proteínas N-WASp, WAVE1 y JMY.

4.12 Ampliación de la anotación funcional usando la base datos GO

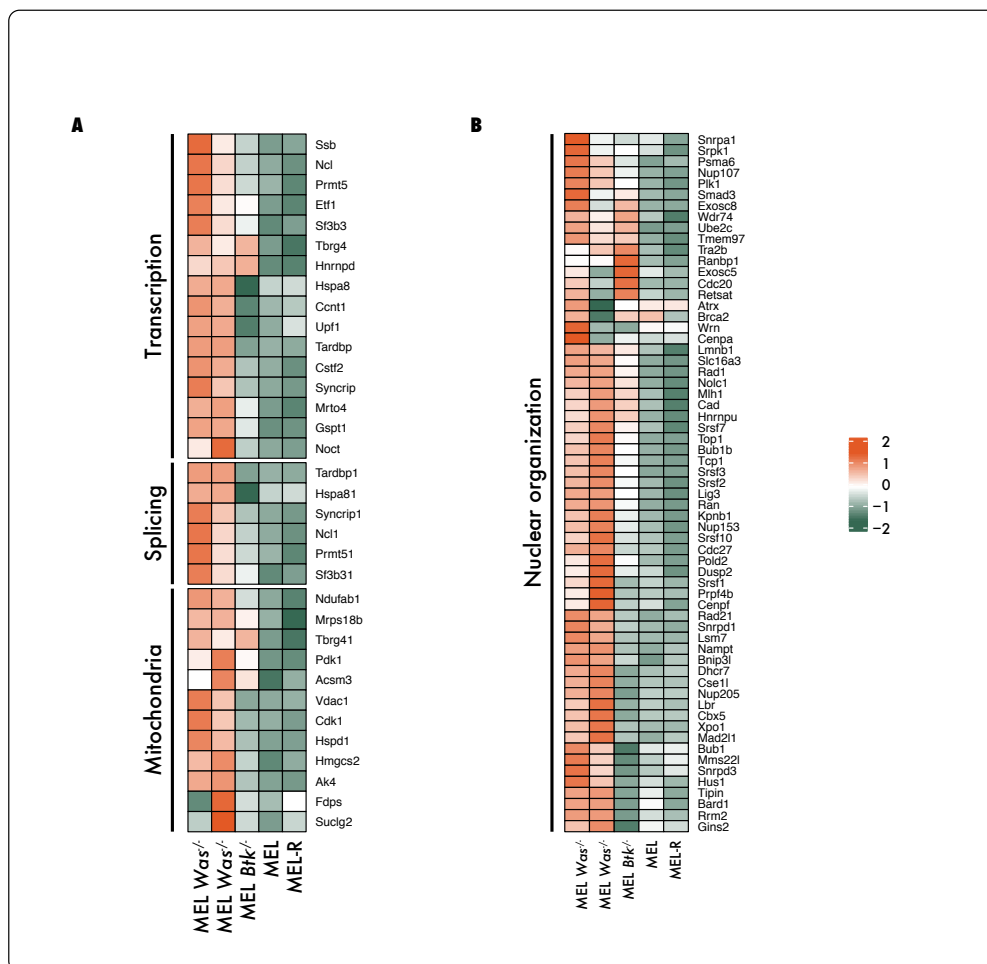


FIGURA 21. AMPLIACIÓN DE LA ANOTACIÓN FUNCIONAL DE GENES CUYA EXPRESIÓN DISMINUYE EN MEL *Was*^{-/-}. Los mapas de calor resumen la expresión de genes anotados funcionalmente por GSEA y reanotados usando la base de datos *Gene Ontology* (GO). La intensidad de los colores naranja y verde indica un mayor o menor nivel de expresión, respectivamente, representado por *logCPM*.

Con el objetivo de obtener más información acerca de la anotación funcional de genes cuya expresión disminuye significativamente en ausencia de *Was*, se seleccionaron los genes identificados mediante GSEA como más enriquecidos para la condición MEL para ser nuevamente categorizados mediante el método ORA, usando la base datos GO. Tal como se observa en la figura 21, la ampliación funcional muestra como

resultado procesos que en su mayoría tienen lugar en el núcleo, entre los que se encuentran la transcripción, *splicing*, organización nuclear, así como en la mitocondria.

En lo que respecta al grupo funcional de la transcripción se observa un subconjunto minoritario de genes que presentan un aumento de expresión en MEL *Was*^{-/-} y MEL *Btk*^{-/-}, mientras que los genes restantes, que representan la mayoría de los analizados, tienden a una caída en la expresión tanto en MEL *Was*^{-/-} y MEL *Btk*^{-/-} con respecto a MEL y MEL-R. En el resto de los grupos funcionales, correspondientes a *splicing* y organización nuclear, se observó una disminución generalizada de la expresión génica en MEL *Was*^{-/-} y MEL *Btk*^{-/-} con respecto a MEL y MEL-R. En cuanto a los genes relacionados con mitocondria, se observó un patrón de inactivación de todos los genes analizados tanto en MEL *Was*^{-/-} como en MEL *Btk*^{-/-}, con respecto a MEL y MEL-R.

Estos resultados sugieren que, tanto en lo referente a la función nuclear de WASP como su relación con la estructura y funcionalidad de la mitocondria, MEL *Was*^{-/-} y MEL *Btk*^{-/-} presentan un patrón de expresión similar para todos los grupos funcionales analizados, que en todos los casos fue opuesto al observado en MEL-R. Por otro lado, es importante destacar el posible rol de *Was* como regulador de la expresión de genes relacionados a transcripción, *splicing* y organización nuclear en progenitores eritroides.

4.13 Análisis de la expresión diferencial de transcritos entre MEL *Was*^{-/-} y MEL

Los análisis realizados previamente detectaron diferencias en expresión génica total entre las condiciones MEL y MEL *Was*^{-/-}. Estos análisis toman en cuenta todas las isoformas conocidas de un gen. Sin embargo, un mismo gen puede dar lugar a varias isoformas, generadas a través de empalme o *splicing* alternativo del mRNA, pudiendo tener efectos funcionales diferentes. Estos cambios en la expresión de isoformas han sido reportados en varias enfermedades, entre ellos el cáncer y desórdenes neurológicos (Zhang, 2013). Por otra parte, en el apartado anterior se reportó una

expresión diferencial de genes relacionados con transcripción y *splicing*, lo cual sugiere un posible rol de *Was* en la regulación de ambos procesos.

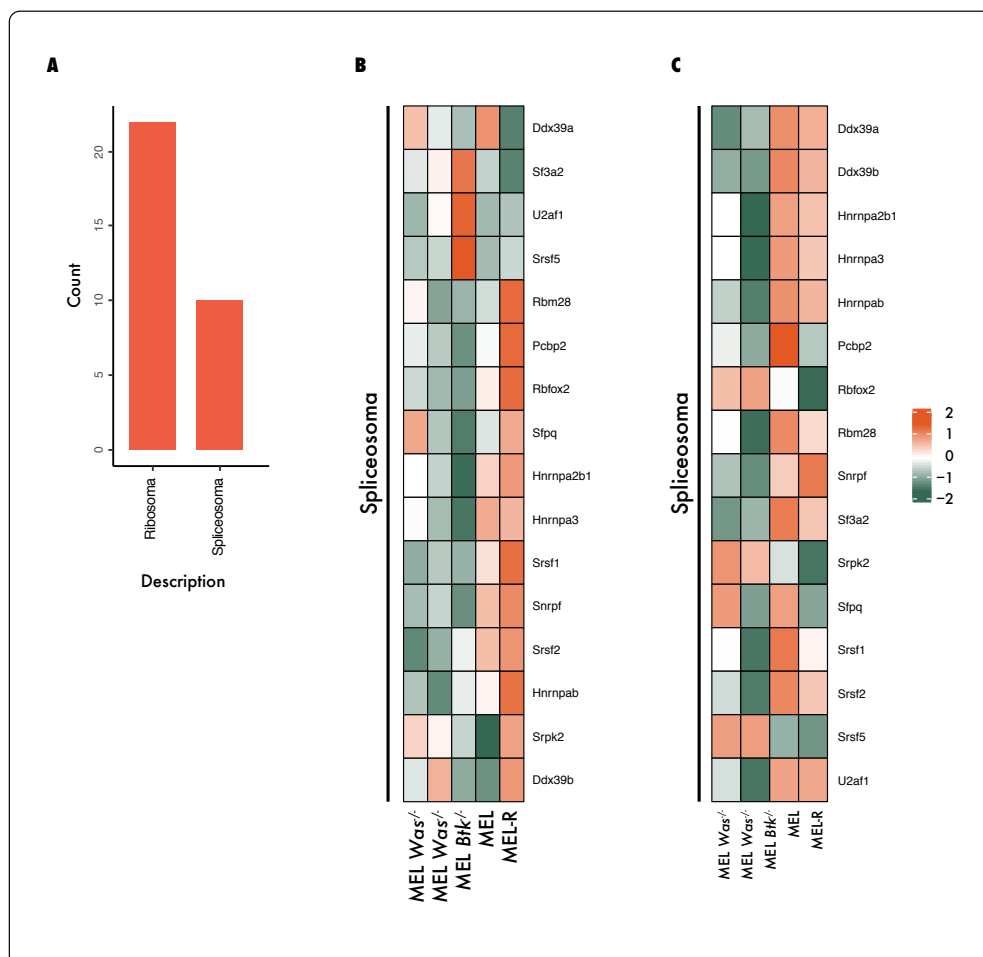


FIGURA 22. TRANSCRITOS DIFERENCIALMENTE EXPRESADOS ENTRE MEL Y MEL *Was*^{-/-}. Categorías funcionales identificadas en función al número de genes (A). Mapa de calor de la expresión de genes involucrados la maquinaria de *splicing* en MEL *Was*^{-/-}, MEL *Btk*^{-/-}, MEL y MEL-R (B) y en células MEL cultivadas en ausencia (MEL *uninduced*) o presencia del agente inductor *DMSO* (MEL *induced*) durante 96 horas (C).

El análisis de expresión diferencial de isoformas entre las condiciones MEL y MEL *Was*^{-/-}, identificó 492 transcritos diferencialmente expresados que corresponden a 254 genes. En la figura 22 A se muestra la anotación funcional de estos genes usando *ORA*,

que asocia los transcritos diferencialmente expresados con dos categorías funcionales de la base de datos *KEGG*, los cuales corresponden al spliceosoma y a ribosomas. Estos resultados sugieren que *Was* podría estar jugando un papel importante en la regulación del *splicing* alternativo, lo cual estaría en concordancia con los resultados obtenidos en el apartado anterior (Fig. 21).

En la figura 22 B se muestra el perfil de expresión de genes involucrados en la maquinaria de *splicing*. Tanto en *MEL Btk^{-/-}* como en *MEL Was^{-/-}* se constató una disminución en la expresión de genes pertenecientes a esta categoría funcional con respecto a MEL. Contrariamente a lo esperado, se observó un aumento en la expresión de los genes de esta categoría en MEL-R, con respecto a MEL, desmarcando a este grupo de todos los demás y sugiriendo una sobreactivación del *splicing* en MEL-R.

En cuanto al perfil de expresión del spliceosoma en la reprogramación de la diferenciación eritroide mediada por agentes inductores, se observó una tendencia hacia la caída de expresión en células MEL cultivadas en presencia de DMSO (Fig. 22 C). Este hecho concuerda con la inactivación de genes relacionados al procesamiento de mRNA y proteínas observada en los estadios finales de la diferenciación tanto murina como humana (An, et al. 2014).

En la tabla 2 se muestran las isoformas identificadas como diferencialmente expresadas para genes pertenecientes a la categoría de spliceosoma. En todos los transcritos analizados se observó un cambio en la relación entre la isoforma funcional, mayoritaria en MEL, y las isoformas no funcionales, predominantes en MEL *Was^{-/-}*.

Tomados en conjunto, nuestros datos sugieren que la ausencia de *Was* en células eritroleucémicas promueve la disminución de la expresión de genes relacionados a la maquinaria de *splicing*, al mismo tiempo que favorece la expresión de isoformas no funcionales de genes que codifican para componentes del splicesoma, lo cual podría comprometer el correcto funcionamiento del spliceosoma.

TABLA 2. ISOFORMAS DE mRNA DIFERENCIALMENTE EXPRESADAS ENTRE MEL WAS^{-/-} Y MEL

Gen	Función biológica	Proporción	Isoformas	Proteína funcional
Zmat2	Componente del spliceosoma	MEL	ENSMUST00000235399.1 ENSMUST00000001419.1	No Sí
		MEL Was ^{-/-}		
Phf5a	Componente del complejo SF3B (splicing factor)	MEL	ENSMUST00000230134.1 ENSMUST00000023117.9 ENSMUST00000160415.1	No Sí No
		MEL Was ^{-/-}		
Sf3b6	Componente del spliceosoma	MEL	ENSMUST00000218650.1 ENSMUST000000046207.8	No Sí
		MEL Was ^{-/-}		
Snrpg	Componente principal del complejo SMN-Sm implicado en el ensamblamiento de la maquinaria snRNP.	MEL	ENSMUST000000089558.6 ENSMUST000000203782.2 ENSMUST00000204768.1	Sí No No
		MEL Was ^{-/-}		
Snu13	Componente del spliceosoma	MEL	ENSMUST000000080622.8 ENSMUST000000072904.4 ENSMUST00000230953.1	Sí No No
		MEL Was ^{-/-}		

5 DISCUSIÓN

En trabajos previos se compararon los transcriptomas de la línea celular MEL-R, la cual es resistente a agentes inductores de la diferenciación, y su línea parental MEL. Dentro de los genes diferencialmente expresados, destacaron un grupo de genes relacionados a la organización del citoesqueleto de actina, cuya expresión disminuye significativamente en MEL-R con respecto a MEL (Fernández-Callejas et al., 2017). Como candidatos para estudiar posibles implicaciones en la diferenciación eritroide se escogieron los genes *Was* y *Btk*, debido a su relación con la organización y polimerización del citoesqueleto de actina, además de estar asociados con enfermedades hematológicas que varían de leves a severas (Catucci et al. 2012.; Hendriks et al., 2014).

Estudios posteriores demostraron que la delección de *Was* es insuficiente para bloquear la respuesta a los agentes inductores (Fernández-Callejas et al., 2019). Sin embargo, aunque los pacientes con SWA presentan anemia hemolítica dentro del cuadro clínico, el rol de *Was* en el contexto eritroide sigue siendo inexplorado. Teniendo en cuenta este aspecto y el hecho de que *Was* presenta una expresión diferencial en células eritroleucémicas resistentes a los agentes inductores de la diferenciación, en el presente estudio se investigaron las posibles funciones moleculares de *Was* dentro del linaje eritroide, más específicamente en proeritroblastos. Nuestros resultados sugieren un rol importante de *Was* en la regulación del proceso de apoptosis, el metabolismo celular, la biosíntesis del grupo hemo y el procesamiento alternativo de pre-mRNA, procesos implicados en la diferenciación y maduración de progenitores eritroides (Liu et al., 2017).

5.1 Convergencia de anotaciones funcionales

En células hematopoyéticas, el gen *Was* es uno de los principales reguladores de la organización del citoesqueleto de actina en el citoplasma (Thrasher et al., 2010). Además de esta función, *Was*, siempre a través de su relación con la polimerización de actina, ha sido asociado con otros importantes procesos celulares tales como: la organización nuclear, el procesamiento de pre-mRNA, la apoptosis, la transcripción y la diferenciación de linajes hematopoyéticos (Kuznetsov et al., 2017). Es importante destacar que la mayoría de estas funciones han sido caracterizadas en el linaje linfóide y, en menor medida, en el mieloide, debido a su relación directa con la inmunodeficiencia y trombocitopenia observada en pacientes con SWA. Sin embargo, aunque no se detectó expresión de *Was* en células en estadios finales de la diferenciación eritroide (Rivers et al., 2017), aún sabemos poco acerca de su rol en durante la eritropoyesis.

Los resultados de las anotaciones funcionales mediante diferentes métodos en progenitores eritroides con delección del gen *Was*, mostraron el solapamiento de cuatro grupos funcionales entre los cuales se encuentran: la glicólisis, la vía de señalización Notch, la apoptosis y la hipoxia. Estudios anteriores han relacionado algunos de estos grupos funcionales con *Was*. Un trabajo previo ha demostrado que la disminución de uno de los genes más conocidos implicados en la glicólisis, la *Aldolasa*, es capaz de inhibir funciones relacionadas a WASp (Ritterson et al., 2012). Además, mutaciones conocidas de *Was* impiden el correcto reclutamiento de proteínas pertenecientes al complejo SWI/SNF, afectando la remodelación de la cromatina, y desregulando la expresión de genes de la vía de señalización Notch (Sarkar et al., 2014). Por otro lado, estudios recientes indicaron que la disminución de la expresión de *Was* es suficiente para desregular la estructura y funcionamiento mitocondrial (Rivers et al., 2020), ambos procesos regulados por genes relacionados a hipoxia (Fuhrmann et al., 2017). Por último, deficiencias en la expresión de *Was* también han sido relacionadas con un incremento de apoptosis dependiente de mitocondria (Obydennyi et al., 2020), lo cual refuerza el posible rol de *Was* en la regulación del funcionamiento de la mitocondria y la apoptosis.

La comparación entre los resultados de la anotación funcional de proeritroblastos y macrófagos de pacientes con SWA, confirmó que la ausencia de *Was*, en especies y linajes hematopoyéticos distintos, resulta en una desregulación en los grupos funcionales de dianas de *c-Myc*, hipoxia, interferón alfa y gamma, la vía de señalización TNFA, dianas de E2F, G2M *checkpoint* y el huso mitótico. Varios de estos grupos funcionales ya han sido asociados y caracterizados en pacientes con SWA. Entre ellos se encuentran los interferones alfa y gamma, así como TNFA, cuya sobreexpresión juega un rol crucial en el desarrollo de enfermedades autoinmunes (Steinman et al., 2007).

En conjunto, nuestros datos muestran que, incluso en contextos diferentes, la ausencia de WASp provoca una desregulación de los grupos funcionales comunes, los cuales representan candidatos fiables a ser profundizados experimentalmente para el estudio de las bases moleculares del SWA en distintos linajes hematopoyéticos.

5.2 Rol de *Was* en la regulación de procesos relacionados con la diferenciación eritropoyética

En respecta al rol de *Was* en la diferenciación eritroide, los resultados del presente trabajo mostraron que desde el estadio de proeritroblasto hasta el de eritroblasto ortocromático, hay una disminución gradual en la expresión de *Was*. Además, es importante destacar que los perfiles de expresión de apoptosis, HIF-1 y *c-Myc* en MEL *Was*^{-/-}, presentaron una gran similitud con los patrones de expresión de células MEL inducidas a la diferenciación con DMSO, así como con los perfiles de expresión de estadios terminales de la diferenciación eritroide. Estos resultados sugieren que la disminución de expresión de *Was* sería necesaria para la maduración hasta el estadio de eritroblasto ortocromático, así como para la reprogramación de la diferenciación celular en proeritroblastos mediante agentes inductores (DMSO). En caso de ser confirmados experimentalmente, estos hallazgos podrían ser profundizados en muestras de pacientes SWA con mutaciones causantes de la sobreactivación de WASp,

clínicamente denominada como XLN (*X-linked neutropenia*), lo que permitiría confirmar la función de *Was* y el efecto de su sobreactivación en el contexto eritroide.

Por otro lado, nuestros datos sugieren que la ausencia de *Was* ocasiona una disminución en la expresión de un grupo de genes relacionado con la biosíntesis del grupo hemo, grupo cuya correcta expresión es requerida en estadios tardíos terminales de la diferenciación. Contrariamente a lo observado en MEL *Was*^{-/-}, en estadios finales de la diferenciación eritroide y en células MEL diferenciadas en presencia DMSO se observó una tendencia al aumento de expresión de este grupo de genes, lo cual es de esperar teniendo en cuenta el rol fundamental del grupo hemo en la funcionalidad de los eritrocitos. Por tanto, la caída en la expresión de genes responsables de la biosíntesis del grupo hemo observada en MEL *Was*^{-/-} comprometería la funcionalidad de los eritrocitos de pacientes con SWA y podría estar relacionada con formas no autoinmunes de anemia hemolítica reportadas en pacientes (Sokolic et al., 2014).

Trabajos previos también mostraron que MEL *Was*^{-/-} y MEL-R presentan defectos en la organización del citoesqueleto de actina debido a la ausencia de *Was* (Fernández-Callejas et al., 2019). Esta ausencia es una característica compartida por ambas líneas celulares y ocurre tanto nivel de mRNA como de proteína. Además, es sabido que la ausencia de WASp puede ser compensada por ciertos miembros de la familia WASP, debido al alto grado de homología de dominios que presentan. Los resultados del análisis de la expresión de los miembros de esta familia sugieren que en ausencia de *Was* se produce un aumento en la expresión de dos miembros, *Washc1* y *Jmy*, ambos responsables de la polimerización de actina, en lisosomas (*Wash*) y en citoplasma, así como de la regulación de p53 en el núcleo (*Jmy*). Resulta llamativo que dos miembros de la familia WASP, N-WASp y WAVE2, previamente identificados como genes que compensan la ausencia de WASp (Isaac et al., 2010; Kuznetsov et al., 2017), presentaron un aumento de su expresión tanto en MEL-R como en estadios finales de la diferenciación eritroide, pero no así en MEL *Was*^{-/-}. Estos diferentes mecanismos de compensación de la ausencia de WASp, podrían explicar, al menos en parte, las diferencias observadas en cuanto al comportamiento de los grupos funcionales

analizados en MEL *Was*^{-/-} y MEL-R, los cuales difieren en los patrones de expresión de todos los grupos funcionales analizados.

En conjunto, los resultados de la presente tesis sugieren que la expresión de *Was* es prescindible para la maduración morfológica de los progenitores eritroides. Sin embargo, la ausencia de *Was* podría desregular ciertos procesos importantes para el contexto eritroide, como la expresión de genes responsables de la biosíntesis del grupo hemo, lo cual se constituye en una primera evidencia del rol de *Was* dentro del contexto eritroide.

5.3 Rol de *Was* en la regulación de la expresión de dianas oncogenes en progenitores eritroides

En la médula ósea, varios factores de transcripción controlan la expresión de genes cruciales para mantener el balance en los diferentes linajes hematopoyéticos. Los resultados de nuestros análisis mostraron el rol regulador de *Was* sobre dos conocidos oncogenes del linaje eritroide.

El factor de transcripción *E2f* es conocido por jugar un papel fundamental en la proliferación celular y regula la expresión de varios genes que codifican para proteínas involucradas en numerosas funciones biológicas. En la gran mayoría de tumores malignos en humanos, *E2f* se encuentra desregulado de forma contexto dependiente, lo cual sugiere roles distintos, a veces contradictorios, en casi todos los tipos de cáncer (Meng et al., 2014). Trabajos previos mostraron que la desregulación de *E2f* es capaz de bloquear la diferenciación terminal de células del linaje mieloide, provocando una marcada proliferación de células inmaduras (Gibbs et al., 2008). Nuestros resultados mostraron que la ausencia de *Was* es suficiente para disminuir el nivel de la expresión de genes dianas de *E2f*.

Otra conocida familia de factores de transcripción es la familia de oncogenes *Myc*, la cual está compuesta por *c-Myc*, *n-Myc* y *l-Myc*, y es conocida por su rol crítico en la proliferación celular y las desregulaciones en su expresión se ha asociado a varios tipos de cáncer (Chen et al., 2018) Estudios previos reportaron que pacientes con

eritroleucemia presentan una sobreexpresión de *c-Myc* y *n-Myc*, siendo capaz de inducir el bloqueo la diferenciación eritroide (Liu et al., 2017; Geiler et al., 2014). Nuestros resultados demostraron que en ausencia de *Was* existe una marcada disminución de dianas de *c-Myc*, lo cual sugiere una posible acción reguladora a nivel de expresión entre *Was* y *c-Myc*. Un estudio anterior demostró que *Wash*, otro miembro de la familia WASP con un alto grado de homología de dominio con *WASP*, es capaz de suprimir tanto la activación transcripcional de *c-Myc* como la de sus dianas (Xia et al., 2014). Además, en el mismo estudio también se demostró que la ausencia de *Wash* en células hematopoiéticas es suficiente para producir defectos en la eritropoyesis causando citopenia y anemia severa. Por otro lado, *c-Myc* debido a su rol de oncogén, también regula la expresión de genes involucrados en el metabolismo celular y la síntesis de ribosomas, procesos necesarios para el estado proliferativo (Dang, 2012). Llamativamente, varios de los procesos metabólicos influenciados por *c-Myc*, tales como el metabolismo de la glucosa, metabolismo de ácidos grasos e hipoxia, también mostraron una disminución en su expresión génica en ausencia de *Was*, lo cual refuerza el posible papel regulador de *Was* sobre la expresión de dianas del oncogén *c-Myc*.

En relación con el contexto eritroide, es importante recordar que, en estadios finales de la diferenciación, *c-Myc* sufre una disminución en su expresión, hecho que coincide con la caída de procesos relacionados a proliferación, transcripción y translación (An et al., 2014, Jayapal et al., 2010). Además, trabajos previos realizados en células eritroleucémicas, demostraron que la expresión ectópica de *Mad1*, antagonista de *c-Myc*, el cual es capaz de inducir la diferenciación de MEL-R en presencia de HMBA (Fernández-Nestosa et al., 2008). Nuestros resultados confirman que tanto en estadios terminales de la diferenciación eritroide como en células MEL cultivadas en presencia de un agente inductor de la diferenciación (DMSO), se produce una caída en la expresión génica de las dianas de *c-Myc*, lo cual coincide con la caída de la expresión de *Was* y el patrón de expresión observado en las células MEL *Was*^{-/-}.

La relación reguladora observada entre *Was* y los oncogenes *c-Myc* y *E2f* parece no ser exclusiva del contexto eritroide. Nuestros resultados en macrófagos de pacientes

con SWA, identificaron a las dianas de *c-Myc* y *E2f* entre los grupos resultantes de la anotación funcional. Sin embargo, esta relación fue opuesta a la observada en MEL *Was*^{-/-}, sugiriendo que la relación *Was* con estos oncogenes es contexto dependiente.

Tomados en conjunto, nuestros datos evidencian que *Was* actuaría como un regulador positivo de la expresión de las dianas de *c-Myc* y *E2f* en proeritroblastos. En el contexto de la diferenciación eritroide, estos datos confirmarían la importancia de la inactivación de *Was* y su posible efecto sobre la inactivación de *c-Myc* y *E2f* en los estadios finales de la diferenciación eritroide. Por otro lado, nuestros resultados sugieren que el rol de supresor tumoral atribuido a *Was* en el linaje linfoide (Menotti et al., 2019) podría variar en otros linajes hematopoyéticos.

5.4 Rol nuclear de *Was*: regulación del *splicing* y el spliceosoma

El uso alternativo de exones se encuentra en más del 90% de genes humanos. Sin embargo, en los últimos años la desregulación del *splicing* ha sido relacionada a un gran número de enfermedades humanas, incluidas ciertas patologías específicas del linaje eritroide (Reimer et al., 2018). Defectos en la expresión y actividad de genes que codifican para factores de *splicing*, dan lugar a isoformas aberrantes que favorecen el cáncer (Zhang et al., 2013). Un ejemplo son los síndromes mielodisplásicos (MDS) en donde mutaciones en proteínas del centro regulador del spliceosoma, son suficientes para producir una eritropoyesis defectuosa y aumentar la predisposición a leucemias mieloides agudas (Shukla et al., 2017).

Nuestros datos sugieren que la ausencia de *Was* favorece un cambio en el patrón de expresión de los transcritos de genes relacionados con el spliceosoma, favoreciendo específicamente la expresión de isoformas no funcionales. Además, se observó que la ausencia de *Was* también ocasiona una disminución generalizada en la expresión total de genes relacionados al proceso de *splicing*. Este resultado coincide con observaciones de otros grupos, que relacionaron formas truncadas de WASp con una desregulación en la expresión de genes relacionados con la maquinaria de *splicing* (Li et al., 2018), lo cual apoya el posible rol de *Was* en el proceso de *splicing*.

Dentro del contexto de la diferenciación eritroide, un trabajo previo de García-Sacristán y colaboradores demostró que la exposición de células MEL al HMBA favorece la expresión de formas no funcionales de proteínas asociadas al spliceosoma, mientras que en células no diferenciadas prevalece de expresión de isoformas funcionales (García-Sacristán et al., 2005). Es importante destacar que nuestros resultados demostraron que tanto en células MEL cultivadas en presencia de un agente inductor (DMSO) como en MEL-R hay un aumento en la expresión de genes relacionados al spliceosoma (Fig. 22 B), siendo esta la única característica que desmarcó a las células MEL-R de su línea parental.

Dentro del contexto de la diferenciación eritroide, un trabajo previo demostró que la exposición de células MEL a HMBA favorece la expresión de formas no funcionales de proteínas asociadas al spliceosoma, mientras que en células no diferenciadas prevalece la expresión de isoformas funcionales (García-Sacristán et al., 2005). Estos estarían en línea con nuestros resultados, ya que las células MEL-R fueron obtenidas a partir exposición constante de células MEL a agentes inductores, y en su perfil de expresión se observó un aumento de expresión de genes relacionados a la maquinaria de *splicing* (Fig. 22 B), lo cual por primera vez desmarca a las células MEL-R del resto de los grupos de estudio. Esto también coincide con observaciones anteriores de otros grupos, en donde ya se relacionó proteínas truncadas de WASp con una marcada activación de la expresión de genes relacionados a la maquinaria de *splicing*, lo cual apoya el rol de *Was* en el proceso de *splicing*.

Por último, trabajos anteriores sugieren un rol de *Was* en la actividad nuclear en el contexto linfoide (Kuznetsov et al., 2017). En ese sentido, nuestros datos sugieren que existe una disminución general en el nivel de expresión de genes relacionados a la organización nuclear en la ausencia de *Was*. Este hallazgo está en concordancia con estudios previos que demostraron que tanto la disminución de expresión como la pérdida de la función en *Wash* y *Was* causan defectos en la organización estructural de compartimientos nucleares tanto en *Drosophila* como en pacientes con SWA (Verboon et al., 2015; Li et al., 2018).

Tomados en conjunto nuestros datos sugieren que WASp regula la expresión de transcritos de genes relacionados con la maquinaria de *splicing*, aumentando la expresión de isoformas no funcionales, así como también la expresión de genes ligados a la organización nuclear.

5.5 WASp y BTK: interacción y regulación de funcional.

Trabajos previos en macrófagos, demostraron que las proteínas WASp y BTK interactúan físicamente mediante sus dominios N-terminal (WASp) y SH3 (BTK), formando un complejo cuya función es crucial para la activación de la vía de señalización TLRs (*T*oll-*L*ike *R*eceptors), la cual es una de las vías inductoras más potentes de la respuesta inflamatoria (Sakuma et al., 2014). Otros trabajos mostraron que la inhibición de la interacción entre WASp y BTK resulta en una desregulación de la expresión de genes relacionados a la respuesta inflamatoria, entre los que se encuentran TNFA, IL-6, IL-1 e IL-1 β . (Sakuma et al., 2015).

Los resultados del análisis de transcriptomas de progenitores eritroides con deleciones bialélicas de *Btk* y *Was*, realizados en el marco de la presente tesis, pusieron de manifiesto un alto grado de similitud en sus perfiles de expresión a lo largo de todos los grupos funcionales estudiados. Esto resulta muy llamativo, teniendo cuenta que las células MEL *Btk*^{-/-} expresan *Was*, con niveles similares a los observados en la línea parental MEL. En ese sentido, los resultados obtenidos demostraron que la ausencia de *Btk* es capaz de replicar los perfiles de expresión observados en MEL *Was*^{-/-}, sugiriendo la importancia de BTK para la funcionalidad de la proteína WASp. En un estudio previo, se demostró que la inhibición de la interacción WASp-BTK, a través de anticuerpos que se unen específicamente al dominio N-terminal de WASp, es capaz de evitar la fosforilación de WASp. Nuestros resultados, sumados a estos hallazgos, resaltan la importancia de la interacción WASp-BTK y del rol de BTK en el correcto funcionamiento de WASp (Sakuma et al., 2015).

Además de la interacción física entre WASp y BTK, interacciones regulatorias a nivel de expresión también han sido reportadas previamente. En un trabajo anterior de Fernández-Callejas y colaboradores, se demostró que en células MEL-R, en donde el gen *Btk* está metilado e inactivo, la expresión ectópica de *Was* es suficiente para promover la expresión de BTK (Fernández-Callejas et al., 2019).

Tomados en conjunto, nuestros resultados demostraron que la ausencia de *Btk* en precursores eritroides recapitula los perfiles de MEL *Was*^{-/-}, lo cual sugiere que la expresión de BTK sería necesaria para el correcto funcionamiento de WASp. Este hallazgo podría ser tomado en cuenta en el estudio de las bases moleculares del SWA y podría traducirse en el sustento para terapias nuevas y eficientes.

5.6 Trabajos futuros

Como trabajos futuros, se plantea realizar validaciones experimentales de la expresión diferencial de genes pertenecientes a grupos funcionales que podrían aportar información sobre la función de WASp en el contexto del linaje eritroide, tales como los genes relacionados a biosíntesis del grupo hemo, apoptosis, hipoxia y las dianas de *c-Myc*.

Por otro lado, a fin de confirmar que la inactivación de *Was* es necesaria para el avance hacia los estadios terminales de la diferenciación eritroide, se planea emplear células MEL *Was*^{-/-} y MEL transfectadas con un vector de expresión ectópica de *Was* (pcDNA 3.1 *Was*), para luego analizar la respuesta de estas células a los agentes inductores de la diferenciación (HMBA y/o DMSO).

Además, también nos planteamos llevar a cabo la confirmación experimental de la relación entre *Was* y la maquinaria de *splicing*. Esto se llevaría a cabo mediante la validación de la expresión diferencial de isoformas no funcionales de componentes del spliceosoma en ausencia *Was*, así como también analizar la posible co-localización de *Was* con componentes del spliceosoma, entre ellos *Srsf2*.

Por último, resultaría muy interesante explorar la posible interacción y relación funcional entre WASp y BTK, ambas asociadas con inmunodeficiencias primarias ligadas al cromosoma X. Esto podría ser llevado a cabo mediante análisis de *Western Blot* que nos permitirían evaluar la activación y estabilidad de WASp en presencia o ausencia de BTK usando las células MEL y MEL *Btk*^{-/-}. En ese sentido, también cabe destacar que queda pendiente el análisis transcriptómico de la línea celular MEL *Btk*^{-/-}, obtenida mediante el sistema CRISPR/Cas9, con el objeto de identificar genes y categorías funcionales regulados por el *Btk* en el contexto del linaje eritroide.

6 CONCLUSIONES

El análisis bioinformático del transcriptoma de la línea celular MEL *Was*^{-/-} llevado a cabo en la presente tesis constituye el primer precedente del análisis del rol WASp en precursores eritroides. En conjunto, nuestros resultados sugieren que la inactivación *Was* es necesaria para la diferenciación de progenitores eritroides en estadios finales y su ausencia se caracteriza por cambios en la expresión de genes relacionados a la biosíntesis del grupo hemo y dianas de oncogenes.

De acuerdo con los resultados obtenidos nuestras conclusiones son las siguientes:

1. Se ha demostrado que grupos de genes implicados en los procesos de apoptosis y respuesta inflamatoria, entre los que se encuentran TNFA, INF- α e INF- γ , están sobreexpresados en la línea MEL *Was*^{-/-} con respecto a la línea parental MEL.
2. Se ha demostrado que grupos de genes implicados en procesos metabólicos entre los que destacan hipoxia (HIF-1), glicólisis y biosíntesis del grupo hemo presentan una disminución en la expresión en la línea MEL *Was*^{-/-} con respecto a la línea parental MEL.
3. Se ha demostrado que grupos de genes dianas de los oncogenes *c-Myc* y *E2f* presentan una disminución en la expresión en la línea MEL *Was*^{-/-} con respecto a la línea parental MEL.

4. Se han constatado que el 33% de los grupos funcionales identificados en MEL *Was*^{-/-} coinciden con grupos funcionales identificados en macrófagos de pacientes con SWA. Los grupos funcionales comunes correspondieron a la vía de señalización TNFA, respuesta al interferón alfa y gamma, huso mitótico, dianas de *c-Myc*, dianas de *E2f* e hipoxia (HIF-1).
5. Se ha comprobado que, en ausencia de *Was*, los patrones de expresión de los genes pertenecientes a los grupos funcionales de apoptosis, hipoxia, dianas de *c-Myc*, son similares a los observados en eritroblastos de estadios finales de la diferenciación eritroide y en células MEL cultivadas en presencia de DMSO durante 96 horas.
6. Se ha comprobado que, en ausencia de *Was*, los patrones de expresión de los genes pertenecientes a los siguientes grupos funcionales: apoptosis, p53, vía de señalización *Notch*, hipoxia, glicólisis, la vía de señalización IL-6/JAK/STAT, vía de señalización TNFA, la biosíntesis del grupo hemo, dianas de *c-Myc*, dianas de PI3K, dianas de mTOR, dianas de *E2f*, autofagia, lisosomas, metafase y anafase mitótica, G2M *checkpoint*, *splicing*, mitocondrias, transcripción, organización nuclear y spliceosoma, son similares a los observados en la línea celular MEL *Btk*^{-/-}.
7. Se demostró que los grupos de genes implicados en procesamiento de pre-mRNA, organización de núcleo y mitocondria presentan una disminución de expresión en la línea MEL *Was*^{-/-} con respecto a la línea parental MEL.
8. Se ha demostrado de que la ausencia *Was* induce la expresión diferencial de transcritos asociados a ribosomas y spliceosoma, favoreciendo la expresión de isoformas no funcionales de los genes *Zmat2*, *Phf5a*, *Sf3b6*, *Snrpg* y *Snu13*, que codifican para componentes del spliceosoma.

7 REFERENCIAS BIBLIOGRAFICAS

1. AJIOKA, R. S., PHILLIPS, J. D., KUSHNER, J. P. 2006. Biosynthesis of heme in mammals. *Biochimica et biophysica acta*, 1763(7), 723–736. <https://doi.org/10.1016/j.bbamcr.2006.05.005>
2. AN, X., SCHULZ, V. P., LI, J., WU, K., LIU, J., XUE, F., HU, J., MOHANDAS, N., & GALLAGHER, P. G. 2014. Global transcriptome analyses of human and murine terminal erythroid differentiation. *Blood*, 123(22), 3466–3477. <https://doi.org/10.1182/blood-2014-01-548305>
3. ANAGNOSTOU, A., LIU, Z., STEINER, M., CHIN, K., LEE, E. S., KESSIMIAN, N., & NOGUCHI, C. T. 1994. Erythropoietin receptor mRNA expression in human endothelial cells. *Proceedings of the National Academy of Sciences of the United States of America*, 91(9), 3974–3978. <https://doi.org/10.1073/pnas.91.9.3974>
4. ANDERS, S., & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome biology*, 11(10), R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
5. ANDERSON, H. L., BRODSKY, I. E., & MANGALMURTI, N. S. 2018. The Evolving Erythrocyte: Red Blood Cells as Modulators of Innate Immunity. *The Journal of Immunology*, 201(5), 1343-1351. doi:10.4049/jimmunol.1800565
6. ANDREWS S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
7. ASARO, R. J., ZHU, Q., & CABRALES, P. 2018. Erythrocyte Aging, Protection via Vesiculation: An Analysis Methodology via Oscillatory Flow. *Frontiers in Physiology*, 9. doi:10.3389/fphys.2018.01607
8. ASARO, R. J., ZHU, Q., & CABRALES, P. 2018. Erythrocyte Aging, Protection via Vesiculation: An Analysis Methodology via Oscillatory Flow. *Frontiers in Physiology*, 9. doi:10.3389/fphys.2018.01607

9. BEEL, K., COTTER, M. M., BLATNY, J., BOND, J., LUCAS, G., GREEN, F., VANDUPPEN, V., LEUNG, D. W., ROONEY, S., SMITH, O. P., ROSEN, M. K., & VANDENBERGHE, P. 2009. A large kindred with X-linked neutropenia with an I294T mutation of the Wiskott-Aldrich syndrome gene. *British journal of haematology*, 144(1), 120–126. <https://doi.org/10.1111/j.1365-2141.2008.07416.x>
10. BERG, L. J., FINKELSTEIN, L. D., LUCAS, J. A., & SCHWARTZBERG, P. L. 2005. Tec family kinases in T lymphocyte development and function. *Annual review of immunology*, 23, 549–600. <https://doi.org/10.1146/annurev.immunol.22.012703.104743>
11. BIANCONI, E., PIOVESAN, A., FACCHIN, F., BERAUDI, A., CASADEI, R., FRABETTI, F., . . . CANAIDER, S. 2013. An estimation of the number of cells in the human body. *Annals of Human Biology*, 40(6), 463–471. doi:10.3109/03014460.2013.807878
12. BLUNDELL, M. P., WORTH, A., BOUMA, G., & THRASHER, A. J. 2010. The Wiskott-Aldrich syndrome: The actin cytoskeleton and immune cell function. *Disease markers*, 29(3-4), 157–175. <https://doi.org/10.3233/DMA-2010-0735>
13. BOLGER, A. M., LOHSE, M., & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
14. BURDA P, LASLO P, STOPKA T. 2010. The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. *Leukemia* 24: 1249-1257.
15. CAMPELLONE, K. G., & WELCH, M. D. 2010. A nucleator arms race: Cellular control of actin assembly. *Nature Reviews Molecular Cell Biology*, 11(4), 237–251. doi:10.1038/nrm2867
16. CASTIELLO, M. C., BOSTICARDO, M., PALA, F., CATUCCI, M., CHAMBERLAIN, N., VAN ZELM, M. C., DRIESSEN, G. J., PAC, M., BERNATOWSKA, E., SCARAMUZZA, S., AIUTI, A., SAUER, A. V., TRAGGIAI, E., MEFFRE, E., VILLA, A., & VAN DER BURG, M. 2014.

- Wiskott–Aldrich syndrome protein deficiency perturbs the homeostasis of B-cell compartment in humans. *Journal of Autoimmunity*, 50, 42–50. <https://doi.org/10.1016/j.jaut.2013.10.006>
17. CATUCCI, M., CASTIELLO, M. C., PALA, F., BOSTICARDO, M., & VILLA, A. 2012. Autoimmunity in Wiskott–aldrich syndrome: An Unsolved enigma. *Frontiers in Immunology*, 3. <https://doi.org/10.3389/fimmu.2012.00209>
 18. CHEN, H., LIU, H., & QING, G. 2018. Targeting oncogenic Myc as a strategy for cancer treatment. *Signal transduction and targeted therapy*, 3, 5. <https://doi.org/10.1038/s41392-018-0008-7>
 19. CHOE, K. S., RADPARVAR, F., MATUSHANSKY, I., REKHTMAN, N., HAN, X. Y SKOULTCHI, A. I. 2003. Reversal of tumorigenicity and the block to differentiation in erythroleukemia cells by GATA-1. *Cancer Res* 63, 6363-6369.
 20. CHOU, H. C., ANTÓN, I. M., HOLT, M. R., CURCIO, C., LANZARDO, S., WORTH, A., BURNS, S., THRASHER, A. J., JONES, G. E., & CALLE, Y. 2006. WIP regulates the stability and localization of WASP to podosomes in migrating dendritic cells. *Current biology : CB*, 16(23), 2337–2344. <https://doi.org/10.1016/j.cub.2006.10.037>
 21. CMARIK, J. L., MIN, H., HEGAMYER, G., ZHAN, S., KULESZ-MARTIN, M., YOSHINAGA, H., MATSUHASHI, S., & COLBURN, N. H. 1999. Differentially expressed protein Pdc4 inhibits tumor promoter-induced neoplastic transformation. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24), 14037–14042. <https://doi.org/10.1073/pnas.96.24.14037>
 22. CONESA, A., MADRIGAL, P., TARAZONA, S., GOMEZ-CABRERO, D., CERVERA, A., MCPHERSON, A., SZCZEŚNIAK, M. W., GAFFNEY, D. J., ELO, L. L., ZHANG, X., & MORTAZAVI, A. 2016. A survey of best practices for RNA-seq data analysis. *Genome biology*, 17, 13. <https://doi.org/10.1186/s13059-016-0881-8>

23. DANG C. V. 2012. MYC on the path to cancer. *Cell*, 149(1), 22–35. <https://doi.org/10.1016/j.cell.2012.03.003>
24. DE LA FUENTE, M. A., SASAHARA, Y., CALAMITO, M., ANTÓN, I. M., ELKHAL, A., GALLEG0, M. D., SURESH, K., SIMINOVITCH, K., OCHS, H. D., ANDERSON, K. C., ROSEN, F. S., GEHA, R. S., & RAMESH, N. 2007. WIP is a chaperone for Wiskott-Aldrich syndrome protein (WASP). *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), 926–931. <https://doi.org/10.1073/pnas.0610275104>
25. DEVRIENDT, K., KIM, A. S., MATHIJS, G., FRINTS, S. G., SCHWARTZ, M., VAN DEN OORD, J. J., VERHOEF, G. E., BOOGAERTS, M. A., FRYNS, J. P., YOU, D., ROSEN, M. K., & VANDENBERGHE, P. 2001. Constitutively activating mutation in WASP causes X-linked severe congenital neutropenia. *Nature genetics*, 27(3), 313–317. <https://doi.org/10.1038/85886>
26. DORÉ, L. C., & CRISPINO, J. D. 2011. Transcription factor networks in erythroid cell and megakaryocyte development. *Blood*, 118(2), 231-239. doi:10.1182/blood-2011-04-285981
27. FERNÁNDEZ-CALLEJA V, FERNÁNDEZ-NESTOSA M-J, HERNÁNDEZ P, SCHVARTZMAN JB, KRIMER DB. 2019. CRISPR/Cas9- mediated deletion of the Wiskott-Aldrich syndrome locus causes actin cytoskeleton disorganization in murine erythroleukemia cells. *PeerJ* 7:e6284 <http://doi.org/10.7717/peerj.6284>
28. FERNÁNDEZ-CALLEJA V, HERNÁNDEZ P, SCHVARTZMAN JB, GARCÍA DE LAC0BA M, KRIMER DB. 2017. Differential gene expression analysis by RNA-seq reveals the importance of actin cytoskeletal proteins in erythroleukemia cells. *PeerJ* 5:e3432 <https://doi.org/10.7717/peerj.3432>
29. FERNANDEZ-NESTOSA MJ, HERNANDEZ P, SCHVARTZMAN JB, KRIMER DB. 2008. PU.1 is dispensable to block erythroid differentiation in Friend erythroleukemia cells. *Leuk Res* 32: 121-130.
30. FRIEND C, PATULEIA MC, DE HARVEN E. 1966. Erythrocytic maturation in vitro of murine (Friend) virus-induced leukemic cells. *Natl Cancer Inst Monogr* 22: 505-522.

31. FUHRMANN, D. C., & BRÜNE, B. 2017. Mitochondrial composition and function under the control of hypoxia. *Redox biology*, 12, 208–215. <https://doi.org/10.1016/j.redox.2017.02.012>
32. GARCÍA-SACRISTÁN, A., FERNÁNDEZ-NESTOSA, M. J., HERNÁNDEZ, P., SCHVARTZMAN, J. B., & KRIMER, D. B. 2005. Protein kinase clk/STY is differentially regulated during erythroleukemia cell differentiation: a bias toward the skipped splice variant characterizes postcommitment stages. *Cell research*, 15(7), 495–503. <https://doi.org/10.1038/sj.cr.7290319>
33. GEILER, C., ANDRADE, I., & GREENWALD, D. 2014. Exogenous c-myc blocks differentiation and improves expansion of human erythroblasts in vitro. *International Journal of Stem Cells*, 7(2), 153–157. <https://doi.org/10.15283/ijsc.2014.7.2.153>
34. GERACI, F., SAHA, I., & BIANCHINI, M. 2020. Editorial: Rna-seq analysis: Methods, applications and challenges. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00220>
35. GIBBS, J. D., LIEBERMANN, D. A., & HOFFMAN, B. 2007. Egr-1 abrogates the E2F-1 block in terminal myeloid differentiation and suppresses leukemia. *Oncogene*, 27(1), 98-106. doi:10.1038/sj.onc.1210627
36. GROSSO, R., FADER, C. M., & COLOMBO, M. I. 2017. Autophagy: A necessary event during erythropoiesis. *Blood Reviews*, 31(5), 300-305. doi:10.1016/j.blre.2017.04.001
37. GUINAMARD, R., ASPENSTRÖM, P., FOUGEREAU, M., CHAVRIER, P., & GUILLEMOT, J. C. 1998. Tyrosine phosphorylation of the Wiskott-Aldrich syndrome protein by Lyn and Btk is regulated by CDC42. *FEBS letters*, 434(3), 431–436. [https://doi.org/10.1016/s0014-5793\(98\)01016-3](https://doi.org/10.1016/s0014-5793(98)01016-3)
38. GUTIÉRREZ, L., CABALLERO, N., FERNÁNDEZ-CALLEJA, L., KARKOULIA, E., & STROUBOULIS, J. 2019. Regulation of GATA1 levels in erythropoiesis. *IUBMB Life*, 72(1), 89-105. doi:10.1002/iub.2192
39. HAASE, V. H. 2013. Regulation of erythropoiesis by hypoxia-inducible factors. *Blood Reviews*, 27(1), 41-53. doi:10.1016/j.blre.2012.12.003

40. HAN, Y., GAO, S., MUEGGE, K., ZHANG, W., & ZHOU, B. 2015. Advanced Applications of RNA Sequencing and Challenges. *Bioinformatics and biology insights*, 9(Suppl 1), 29–46. <https://doi.org/10.4137/BBI.S28991>
41. HE, X., ZOU, R., ZHANG, B., YOU, Y., YANG, Y., & TIAN, X. 2017. Whole Wiskott-Aldrich syndrome protein gene deletion identified by high throughput sequencing. *Molecular medicine reports*, 16(5), 6526–6531. <https://doi.org/10.3892/mmr.2017.7416>
42. HENDRIKS, R. W., YUVARAJ, S., & KIL, L. P. 2014. Targeting Bruton's tyrosine kinase in B cell malignancies. *Nature reviews. Cancer*, 14(4), 219–232. <https://doi.org/10.1038/nrc3702>
43. HRDLICKOVA, R., TOLOUE, M., & TIAN, B. 2017. RNA-Seq methods for transcriptome analysis. *Wiley interdisciplinary reviews. RNA*, 8(1), 10.1002/wrna.1364. <https://doi.org/10.1002/wrna.1364>
44. HYMAN, T. 2001. Structural characterization of erythroid and megakaryocytic differentiation in Friend erythroleukemia cells. *Experimental Hematology*, 29(5), 563-571. doi:10.1016/s0301-472x(01)00616-6
45. ISAAC, B. M., ISHIHARA, D., NUSBLAT, L. M., GEVREY, J. C., DOVAS, A., CONDEELIS, J., & COX, D. 2010. N-WASP has the ability to compensate for the loss of WASP in macrophage podosome formation and chemotaxis. *Experimental cell research*, 316(20), 3406–3416. <https://doi.org/10.1016/j.yexcr.2010.06.011>
46. ISHIHARA, D., DOVAS, A., PARK, H., ISAAC, B. M., & COX, D. 2012. The Chemotactic defect in wiskott-aldrich syndrome Macrophages is due to the REDUCED persistence of Directional Protrusions. *PLoS ONE*, 7(1). <https://doi.org/10.1371/journal.pone.0030033>
47. ITO, M., SHICHITA, T., OKADA, M., KOMINE, R., NOGUCHI, Y., YOSHIMURA, A., & MORITA, R. 2015. Bruton's tyrosine kinase is essential FOR NLRP3 inflammasome activation and contributes TO ISCHAEMIC brain injury. *Nature Communications*, 6(1). <https://doi.org/10.1038/ncomms8360>

48. JAYAPAL, S. R., LEE, K. L., JI, P., KALDIS, P., LIM, B., & LODISH, H. F. 2010. Down-regulation of Myc is essential for terminal erythroid maturation. *The Journal of biological chemistry*, 285(51), 40252–40265. <https://doi.org/10.1074/jbc.M110.181073>
49. JI, P., JAYAPAL, S. R., & LODISH, H. F. 2008. ENUCLEATION OF CULTURED MOUSE FETAL ERYTHROBLASTS REQUIRES RAC GTPASES AND MDA2. *Nature cell biology*, 10(3), 314–321. <https://doi.org/10.1038/ncb1693>
50. KANEKO, K., FURUYAMA, K., FUJIWARA, T., KOBAYASHI, R., ISHIDA, H., HARIGAE, H., & SHIBAHARA, S. 2014. Identification of a novel erythroid-specific enhancer for the ALAS2 gene and its loss-of-function mutation which is associated with congenital sideroblastic anemia. *Haematologica*, 99(2), 252–261. <https://doi.org/10.3324/haematol.2013.085449>
51. KARAYEL, Ö., XU, P., BLUDAU, I., VELAN BHOOPALAN, S., YAO, Y., ANA RITA, F. C., SANTOS, A., SCHULMAN, B. A., ALPI, A. F., WEISS, M. J., & MANN, M. 2020. Integrative proteomics reveals principles of dynamic phosphosignaling networks in human erythropoiesis. *Molecular systems biology*, 16(12), e9813. <https://doi.org/10.15252/msb.20209813>
52. KEERTHIVASAN, G., WICKREMA, A., & CRISPINO, J. D. 2011. Erythroblast enucleation. *Stem cells international*, 2011, 139851. <https://doi.org/10.4061/2011/139851>
53. KIM, A. S., KAKALIS, L. T., ABDUL-MANAN, N., LIU, G. A., & ROSEN, M. K. 2000. Autoinhibition and activation mechanisms of the Wiskott-Aldrich syndrome protein. *Nature*, 404(6774), 151–158. <https://doi.org/10.1038/35004513>
54. KIYOKAWA H, RICHON VM, VENTA-PEREZ G, RIFKIND RA, MARKS PA. 1993. Hexamethylenebisacetamide-induced erythroleukemia cell differentiation involves modulation of events required for cell cycle progression through G1. *Proceedings of the National Academy of Sciences of the United States of America* 90: 6746-6750.

55. KONSTANTINIDIS, D. G., PUSHKARAN, S., JOHNSON, J. F., CANCELAS, J. A., MANGANARIS, S., HARRIS, C. E., Kalfa, T. A. 2012. Signaling and cytoskeletal requirements in erythroblast enucleation. *Blood*, 119(25), 6118-6127. doi:10.1182/blood-2011-09-379263
56. LANE, D., & LEVINE, A. 2010. P53 Research: The Past Thirty Years and the Next Thirty Years. *Cold Spring Harbor Perspectives in Biology*, 2(12). doi:10.1101/cshperspect.a000893
57. LEGGETT, R. M., RAMIREZ-GONZALEZ, R. H., CLAVIJO, B. J., WAITE, D., & DAVEY, R. P. 2013. Sequencing quality assessment tools to enable data-driven informatics for high throughput genomics. *Frontiers in Genetics*, 4. <https://doi.org/10.3389/fgene.2013.00288>
58. LI, C. J., LI, J., ZHANG, Q. M., LV, L., CHEN, R., LV, C. F., YU, P., & YU, D. M. 2012. Efficacy and safety comparison between liraglutide as add-on therapy to insulin and insulin dose-increase in Chinese subjects with poorly controlled type 2 diabetes and abdominal obesity. *Cardiovascular diabetology*, 11, 142. <https://doi.org/10.1186/1475-2840-11-142>
59. LI, M., SUZUKI, K., ZHOU, X., YAMAUCHI, T., MORESCO, J., DUNN, S., HERNANDEZ- BENITEZ, R., HISHIDA, T., YOUNG, KIM N., ANDIJANI, M., KU M., TAKAHASHI .Y., XU, J., QIU, J., HUANG, L, BENNER, C., AIZAWA. E., QU, J., LIU, G., LI, Z., YI, F., BI, C., YANAL GHOSHEH, CHANGWEI SHAO, MAXIM SHOKHIREV, PATRIZIA COMOLI, FRANCESCO FRASSONI, JOHN R. YATES III, XIANG-DONG FU, RODRIGUEZ, C., IZPISUA J. 2018. Wiskott-Aldrich Syndrome Protein Regulates Nuclear Organization, Alternative Splicing and Cell Proliferation.
***No publicado**
60. LIAO, Y., SMYTH, G. K., & SHI, W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* (Oxford, England), 30(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>

61. LINDVALL, J. M., BLOMBERG, K. E., VÄLIAHO, J., VARGAS, L., HEINONEN, J. E., BERGLÖF, A., MOHAMED, A. J., NORE, B. F., VIHINEN, M., & SMITH, C. I. 2005. Bruton's tyrosine kinase: cell biology, sequence conservation, mutation spectrum, siRNA modifications, and expression profiling. *Immunological reviews*, 203, 200–215. <https://doi.org/10.1111/j.0105-2896.2005.00225.x>
62. LITZMAN, J., JONES, A., HANN, I., CHAPEL, H., STROBEL, S., & MORGAN, G. 1996. Intravenous immunoglobulin, splenectomy, and antibiotic prophylaxis in Wiskott-Aldrich syndrome. *Archives of disease in childhood*, 75(5), 436–439. <https://doi.org/10.1136/adc.75.5.436>
63. LIU, X., ZHANG, Y., NI, M., CAO, H., SIGNER, R., LI, D., LI, M., GU, Z., HU, Z., DICKERSON, K. E., WEINBERG, S. E., CHANDEL, N. S., DEBERARDINIS, R. J., ZHOU, F., SHAO, Z., & XU, J. 2017. Regulation of mitochondrial biogenesis in erythropoiesis by mTORC1-mediated protein translation. *Nature cell biology*, 19(6), 626–638. <https://doi.org/10.1038/ncb3527>
64. LODISH, H., FLYGARE, J., & CHOU, S. 2010. From stem cell to erythroblast: Regulation of red cell production at multiple levels by multiple hormones. *IUBMB Life*, 62(7), 492–496. doi:10.1002/iub.322
65. LOOI, C. Y., SASAHARA, Y., WATANABE, Y., SATOH, M., HAKOZAKI, I., UCHIYAMA, M., KURE, S. 2014. The open conformation of WASP regulates its nuclear localization and gene transcription in myeloid cells. *International Immunology*, 26(6), 341–352. doi:10.1093/intimm/dxt072
66. LÓPEZ-GÓMEZ, J. M., ABAD, S., & VEGA, A. 2016. New expectations in the treatment of anemia in chronic kidney disease. *Nefrología (English Edition)*, 36(3), 232–236. <https://doi.org/10.1016/j.nefro.2016.07.002>
67. MACE, E. M., & ORANGE, J. S. 2018. Discovering the cause of wiskott–aldrich syndrome and laying the foundation for understanding immune cell structuring. *The Journal of Immunology*, 200(11), 3667–3670. <https://doi.org/10.4049/jimmunol.1800518>

68. MARKS PA, RICHON VM, KIYOKAWA H, RIFKIND RA. 1994. Inducing differentiation of transformed cells with hybrid polar compounds: a cell cycle-dependent process. *Proceedings of the National Academy of Sciences of the United States of America* 91: 10251-10254.
69. MASSAAD, M. J., RAMESH, N., & GEHA, R. S. 2013. Wiskott-Aldrich syndrome: a comprehensive review. *Annals of the New York Academy of Sciences*, 1285, 26–43. <https://doi.org/10.1111/nyas.12049>
70. MATUSHANSKY, I., RADPARVAR, F. Y SKOULTCHI, A. I. 2003. CDK6 blocks differentiation: coupling cell proliferation to the block to differentiation in leukemic cells. *Oncogene* 22, 4143-4149.
71. MAXIMOW, A. 1908. Untersuchungen über Blut und Bindegewebe. *Archiv Für Mikroskopische Anatomie*, 73(1), 444-561. doi:10.1007/bf02979896
72. MAXIMOW, A. 1908. Untersuchungen über Blut und Bindegewebe. *Archiv Für Mikroskopische Anatomie*, 73(1), 444-561. doi:10.1007/bf02979896
73. MENG, P., & GHOSH, R. 2014. Transcription addiction: Can WE garner the yin and YANG functions Of e2f1 for cancer therapy? *Cell Death & Disease*, 5(8). <https://doi.org/10.1038/cddis.2014.326>
74. MENOTTI, M., AMBROGIO, C., CHEONG, T. C., PIGHI, C., MOTA, I., CASSEL, S. H., COMPAGNO, M., WANG, Q., DALL'OLIO, R., MINERO, V. G., POGGIO, T., SHARMA, G. G., PATRUCCO, E., MASTINI, C., CHOUDHARI, R., PICH, A., ZAMO, A., PIVA, R., GILIANI, S., MOLOGNI, L., ... CHIARLE, R. 2019. Wiskott-Aldrich syndrome protein (WASP) is a tumor suppressor in T cell lymphoma. *Nature medicine*, 25(1),130–140.<https://doi.org/10.1038/s41591-018-0262-9>
75. MIDE, S. M., HUYGENS, P., BOZZINI, C. E., & FERNANDEZ POL, J. A. 2001. Effects of human recombinant erythropoietin on differentiation and distribution of erythroid progenitor cells on murine medullary and splenic erythropoiesis during hypoxia and post-hypoxia. *In vivo (Athens, Greece)*, 15(2), 125–132.
76. MORATTO, D., GILIANI, S., BONFIM, C., MAZZOLARI, E., FISCHER, A., OCHS, H. D., CANT, A. J., THRASHER, A. J., COWAN, M. J., ALBERT,

- M. H., SMALL, T., PAI, S. Y., HADDAD, E., LISA, A., HAMBLETON, S., SLATTER, M., CAVAZZANA-CALVO, M., MAHLAOU, N., PICARD, C., TORGERSON, T. R., ... NOTARANGELO, L. D. 2011. Long-term outcome and lineage-specific chimerism in 194 patients with Wiskott-Aldrich syndrome treated by hematopoietic cell transplantation in the period 1980-2009: an international collaborative study. *Blood*, 118(6), 1675–1684. <https://doi.org/10.1182/blood-2010-11-319376>
77. MOREAU-GACHELIN, F. 2008. Multi-stage Friend murine erythroleukemia: Molecular insights into oncogenic cooperation. *Retrovirology*, 5(1), 99. doi:10.1186/1742-4690-5-99
 78. MORIGUCHI, T., & YAMAMOTO, M. 2014. A regulatory network governing Gata1 and Gata2 gene transcription orchestrates erythroid lineage differentiation. *International journal of hematology*, 100(5), 417–424. <https://doi.org/10.1007/s12185-014-1568-0>
 79. MUSZYNSKI KW, OHASHI T, HANSON C, RUSCETTI SK. 1998. Both the polycythemia- and anemia-inducing strains of Friend spleen focus-forming virus induce constitutive activation of the Raf-1/mitogen-activated protein kinase signal transduction pathway. *Journal of virology* 72: 919-925.
 80. MYNENI, V. D., SZALAYOVA, I., & MEZEY, E. 2021. Differences in Steady-State Erythropoiesis in Different Mouse Bones and Postnatal Spleen. *Frontiers in Cell and Developmental Biology*, 9. doi:10.3389/fcell.2021.646646
 81. MYNENI, V. D., SZALAYOVA, I., & MEZEY, E. 2021. Differences in Steady-State Erythropoiesis in Different Mouse Bones and Postnatal Spleen. *Frontiers in Cell and Developmental Biology*, 9. doi:10.3389/fcell.2021.646646
 82. NGOENKAM, J., PAENSUWAN, P., WIPA, P., SCHAMEL, W. W., & PONGCHAROEN, S. 2021. Wiskott-Aldrich Syndrome Protein: Roles in Signal Transduction in T Cells. *Frontiers in Cell and Developmental Biology*, 9. doi:10.3389/fcell.2021.674572

83. NISHIGAKI K, HANSON C, OHASHI T, THOMPSON D, MUSZYNSKI K, RUSCETTI S. 2000. Erythroid cells rendered erythropoietin independent by infection with Friend spleen focus-forming virus show constitutive activation of phosphatidylinositol 3-kinase and Akt kinase: involvement of insulin receptor substrate-related adapter proteins. *Journal of virology* 74: 3037- 3045.
84. OBYDENNYI, S. I., ARTEMENKO, E. O., SVESHNIKOVA, A. N., IGNATOVA, A. A., VARLAMOVA, T. V., GAMBARYAN, S., LOMAKINA, G. Y., UGAROVA, N. N., KIREEV, I. I., ATAULLAKHANOV, F. I., NOVICHKOVA, G. A., MASCHAN, A. A., SHCHERBINA, A., & PANTELEEV, M. 2019. Mechanisms of increased mitochondria-dependent necrosis in Wiskott-Aldrich SYNDROME PLATELETS. *Haematologica*, 105(4), 1095–1106. <https://doi.org/10.3324/haematol.2018.214460>
85. OCHS, H. D., & THRASHER, A. J. 2006. The Wiskott-Aldrich syndrome. *The Journal of allergy and clinical immunology*, 117(4), 725–739. <https://doi.org/10.1016/j.jaci.2006.02.005>
86. OVCHYNNIKOVA, E., AGLIALORO, F., LINDERN, M. V., & AKKER, E. V. 2018. The Shape Shifting Story of Reticulocyte Maturation. *Frontiers in Physiology*, 9. doi:10.3389/fphys.2018.00829
87. PAL SINGH, S., DAMMEIJER, F., & HENDRIKS, R. W. 2019. Correction to: Role Of bruton's tyrosine kinase in B cells and malignancies. *Molecular Cancer*, 18(1). <https://doi.org/10.1186/s12943-019-1009-z>
88. PALIS, J. 2014. Primitive and definitive erythropoiesis in mammals. *Frontiers in Physiology*, 5. doi:10.3389/fphys.2014.00003
89. PAPETTI, M., & SKOULTCHI, A. I. 2007. Reprogramming Leukemia Cells to Terminal Differentiation and Growth Arrest by RNA Interference of PU.1. *Molecular Cancer Research*, 5(10), 1053-1062. doi:10.1158/1541-7786.mcr-07-0145
90. POLLITT, A. Y., & INSALL, R. H. 2009. WASP and SCAR/WAVE proteins: the drivers of actin assembly. *Journal of cell science*, 122(Pt 15), 2575–2578. <https://doi.org/10.1242/jcs.023879>

91. Pure Erythroleukemia: A Case Report And Literature Review. 2011. *The Internet Journal of Hematology*, 7(2). doi:10.5580/feb
92. RAI, R., ROMITO, M., RIVERS, E., TURCHIANO, G., BLATTNER, G., VETHAROY, W., LADON, D., ANDRIEUX, G., ZHANG, F., ZINICOLA, M., LEON-RICO, D., SANTILLI, G., THRASHER, A. J., & CAVAZZA, A. 2020. Targeted gene correction of human hematopoietic stem cells for the treatment of Wiskott - Aldrich Syndrome. *Nature communications*, 11(1), 4034. <https://doi.org/10.1038/s41467-020-17626-2>
93. RAMESH, N., I.M. ANTON, J.H. HARTWIG, AND R.S. GEHA. 1997. WIP, a protein associated with wiskott-aldrich syndrome protein, induces actin polymerization and redistribution in lymphoid cells. *Proc. Natl. Acad. Sci. USA*. 94:14671–14676.
94. REBUCCI M, MICHIELS C. 2013. Molecular aspects of cancer cell resistance to chemotherapy. *Biochemical pharmacology* 85: 1219-1226.
95. RIBEIL JA, ZERMATI Y, VANDEKERCKHOVE J ET AL. 2007. Hsp70 regulates erythropoiesis by preventing caspase-3-mediated cleavage of GATA-1. *Nature*; 445: 102-105.
96. RITTERSON LEW, C., & TOLAN, D. R. 2012. Targeting of several glycolytic enzymes using RNA interference reveals aldolase affects cancer cell proliferation through a non-glycolytic mechanism. *The Journal of Bio- logical Chemistry*, 287(51), 42554–42563. <https://doi.org/10.1074/ jbc.M112.405969>
97. RIVERS, E., RAI, R., LÖTSCHER, J., HOLLINSHEAD, M., MARKELJ, G., THAVENTHIRAN, J., THRASHER, A. 2020. Wiskott Aldrich syndrome protein regulates non-selective autophagy and mitochondrial homeostasis in human myeloid cells. doi:10.1101/2020.02.06.937946
98. ROBERT-MORENO, À., ESPINOSA, L., SANCHEZ, M. J., DE LA POMPA, J. L., & BIGAS, A. 2007. The notch pathway positively regulates programmed cell death during erythroid differentiation. *Leukemia*, 21(7), 1496–1503. <https://doi.org/10.1038/sj.leu.2404705>
99. ROMAN-GARCIA, S., MERINO-CORTES, S. V., GARDETA, S. R., DE BRUIJN, M. J., HENDRIKS, R. W., & CARRASCO, Y. R. 2018. Distinct

roles FOR Bruton's tyrosine KINASE in B Cell IMMUNE synapse formation. *Frontiers in Immunology*, 9. <https://doi.org/10.3389/fimmu.2018.02027>

100. DSCHMIDT, R. E., GREEN, K. G., FENG, D., DORSEY, D. A., PARVIN, C. A., LEE, J.-M., XIAO, Q., & BRINES, M. 2008. Erythropoietin and its carbamylated derivative prevent the development of experimental diabetic autonomic neuropathy in stz-induced diabetic nod-scid mice. *Experimental Neurology*, 209(1), 161–170. <https://doi.org/10.1016/j.expneurol.2007.09.018>
101. ROUSSEAU, G. F., GIARRATANA, M. C., & DOUAY, L. 2014. Large-scale production of red blood cells from stem cells: what are the technical challenges ahead?. *Biotechnology journal*, 9(1), 28–38. <https://doi.org/10.1002/biot.201200368>
102. RUSCETTI SK. 1999. Deregulation of erythropoiesis by the Friend spleen focus- forming virus. *Int J Biochem Cell Biol* 31: 1089-1109.
103. SAKUMA, C., SATO, M., TAKENOUCHI, T., CHIBA, J., & KITANI, H. 2012. Critical roles of the WASP N-terminal domain and Btk in LPS-induced inflammatory response in macrophages. *PLoS one*, 7(1), e30351. <https://doi.org/10.1371/journal.pone.0030351>
104. SARKAR, K., SADHUKHAN, S., HAN, S., & VYAS, Y. M. 2014. Disruption of hSWI/SNF complexes in T cells by WAS mutations distinguishes X-linked thrombocytopenia from Wiskott-Aldrich syndrome. *Blood*, 124(23), 3409-3419. doi:10.1182/blood-2014-07-587642
105. SCHMITTGEN, T. D., & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, 3(6), 1101-1108. doi:10.1038/nprot.2008.73
106. SCHURCH, N. J., SCHOFIELD, P., GIERLIŃSKI, M., COLE, C., SHERSTNEV, A., SINGH, V., WROBEL, N., GHARBI, K., SIMPSON, G. G., OWEN-HUGHES, T., BLAXTER, M., & BARTON, G. J. 2016. How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?. *RNA (New York, N.Y.)*, 22(6), 839–851. <https://doi.org/10.1261/rna.053959.115>

107. SHUKLA, G. C., & SINGH, J. 2017. Mutations of RNA splicing factors in hematological malignancies. *Cancer letters*, 409, 1–8. <https://doi.org/10.1016/j.canlet.2017.08.042>
108. SINGH, V. K., SAINI, A., TSUJI, K., SHARMA, P. B., & CHANDRA, R. 2014. Manufacturing blood ex vivo: A futuristic approach to deal with the supply and safety concerns. *Frontiers in Cell and Developmental Biology*, 2. doi:10.3389/fcell.2014.00026
109. SNAPPER, S. B., ROSEN, F. S., MIZOGUCHI, E., COHEN, P., KHAN, W., LIU, C. H., HAGEMANN, T. L., KWAN, S. P., FERRINI, R., DAVIDSON, L., BHAN, A. K., & ALT, F. W. 1998. Wiskott-Aldrich syndrome protein-deficient mice reveal a role for WASP in T but not B cell activation. *Immunity*, 9(1), 81–91. [https://doi.org/10.1016/s1074-7613\(00\)80590-7](https://doi.org/10.1016/s1074-7613(00)80590-7)
110. SOKOLIC, R. A., CHONAT, S., RISINGER, M., ECKHAUS, M., GARABEDIAN, E., CALVO, K. R., MARIC, I., KALFA, T. A., & CANDOTTI, F. 2014. Cation leak in red blood cells of patients with wiskott-aldrich Syndrome leads to Non-Immunologic Hemolysis. *Blood*, 124(21), 1338–1338. <https://doi.org/10.1182/blood.v124.21.1338.1338>
111. STEINMAN, R. M. 2007. Dendritic cells: Understanding immunogenicity. *European Journal of Immunology*, 37(S1). <https://doi.org/10.1002/eji.200737400>
112. SURI, D., RAWAT, A., & SINGH, S. 2016. X-linked Agammaglobulinemia. *Indian journal of pediatrics*, 83(4), 331–337. <https://doi.org/10.1007/s12098-015-2024-8>
113. SUZUKI, M., KOBAYASHI-OSAKI, M., TSUTSUMI, S., PAN, X., OHMORI, S., TAKAI, J., YAMAMOTO, M. 2013. GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes to Cells*, 18(11), 921-933. doi:10.1111/gtc.12086
114. SUZUKI, M., MORIGUCHI, T., OHNEDA, K., & YAMAMOTO, M. 2009. Differential contribution of the Gata1 gene hematopoietic enhancer to

- erythroid differentiation. *Molecular and cellular biology*, 29(5), 1163–1175.
<https://doi.org/10.1128/MCB.01572-08>
115. SWINDELL, W. R., STUART, P. E., SARKAR, M. K., VOORHEES, J. J., ELDER, J. T., JOHNSTON, A., & GUDJONSSON, J. E. 2014. Cellular dissection of psoriasis for transcriptome analyses and the post-gwas era. *BMC Medical Genomics*, 7(1). <https://doi.org/10.1186/1755-8794-7-27>
 116. TAKENAWA, T., & SUETSUGU, S. 2007. The WASP-WAVE protein network: connecting the membrane to the cytoskeleton. *Nature reviews. Molecular cell biology*, 8(1), 37–48. <https://doi.org/10.1038/nrm2069>
 117. TANG, N., WANG, L., ESKO, J., GIORDANO, F. J., HUANG, Y., GERBER, H. P., FERRARA, N., & JOHNSON, R. S. 2004. Loss of HIF-1alpha in endothelial cells disrupts a hypoxia-driven VEGF autocrine loop necessary for tumorigenesis. *Cancer cell*, 6(5), 485–495.
<https://doi.org/10.1016/j.ccr.2004.09.026>
 118. ATAR, O. Y LEVI, B. Z. 2005. PU.1 silencing leads to terminal differentiation of erythroleukemia cells. *Biochem Biophys Res Commun* 329, 1288-1292.
 119. THRASHER, A. J., & BURNS, S. O. 2010. WASP: a key immunological multitasker. *Nature reviews. Immunology*, 10(3), 182–192.
<https://doi.org/10.1038/nri2724>
 120. TSUKADA S, SAFFRAN DC, RAWLINGS DJ, PAROLINI O, ALLEN RC, KLISAK I, SPARKES RS, KUBAGAWA H, MOHANDAS T, QUAN S ET AL. 1993. Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. *Cell* 72: 279-290.
 121. VANEGAS N, GARCIA-SACRISTAN A, LOPEZ-FERNANDEZ LA, PARRAGA M, DEL MAZO J, HERNANDEZ P, SCHVARTZMAN JB, KRIMER DB. 2003. Differential expression of Ran GTPase during HMBA-induced differentiation in murine erythroleukemia cells. *Leuk Res* 27: 607-615.
 122. VASANI, N. 2019. Review of "De novo Transcript sequence reconstruction FROM RNA-seq using the TRINITY platform for REFERENCE generation and analysis.". <https://doi.org/10.14293/s2199-1006.1.sorchem.aehhwr.v1.rjovmn>

123. VERBOON, J. M., NAKAMURA, M., DECKER, J. R., DAVIDSON, K. A., NANDAKUMAR, V., & PARKHURST, S. M. 2019. Wash and the WASH Regulatory Complex function in Nuclear Envelope budding. doi:10.1101/2019.12.18.881763
124. VETRIE D, VORECHOVSKY I, SIDERAS P, HOLLAND J, DAVIES A, FLINTER F, HAMMARSTROM L, KINNON C, LEVINSKY R, BOBROW M ET AL. 1993. The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. *Nature* 361: 226-233.
125. WANG, B., KUMAR, V., OLSON, A., & WARE, D. 2019. Reviving the transcriptome studies: An insight into the emergence OF Single-Molecule TRANSCRIPTOME SEQUENCING. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.00384>
126. WANG, P., WANG, Z., & LIU, J. 2020. Correction to: Role of hdacs in normal and malignant hematopoiesis. *Molecular Cancer*, 19(1). <https://doi.org/10.1186/s12943-020-01182-w>
127. XIA, P., WANG, S., HUANG, G., ZHU, P., LI, M., YE, B., DU, Y., & FAN, Z. 2014. WASH is required for the differentiation commitment of hematopoietic stem cells in a c-Myc-dependent manner. *The Journal of experimental medicine*, 211(10), 2119–2134. <https://doi.org/10.1084/jem.20140169>
128. XU, L. Z., LI, S. S., ZHOU, W., KANG, Z. J., ZHANG, Q. X., KAMRAN, M., XU, J., LIANG, D. P., WANG, C. L., HOU, Z. J., WAN, X. B., WANG, H. J., LAM, E. W., ZHAO, Z. W., & LIU, Q. 2017. p62/SQSTM1 enhances breast cancer stem-like properties by stabilizing MYC mRNA. *Oncogene*, 36(3), 304–317. <https://doi.org/10.1038/onc.2016.202>
129. ZHANG, J., RANDALL, M. S., LOYD, M. R., LI, W., SCHWEERS, R. L., PERSONS, D. A., REHG, J. E., NOGUCHI, C. T., IHLE, J. N. Y NEY, P. A. 2006. Role of erythropoietin receptor signaling in Friend virus-induced erythroblastosis and polycythemia. *Blood* 107, 73-78
130. ZHOU, L., LI, W., ZHANG, X., LIU, D., DING, Y., DAI, R., & ZHAO, X. 2017. Abnormal distribution of distinct lymphocyte subsets in children with

Wiskott-Aldrich syndrome. *Human immunology*, 78(9), 565–573.
<https://doi.org/10.1016/j.humimm.2017.05.004>

8 ANEXOS

Tabla S1. Primers usados para la validación de la delección y no delección de genes.

Nombre del primer	Sequencia (5'-3')
<i>Was</i> ND	Fwd: TGTGACGGTCGACATTCAGA
	Rv: TCAGATGCCTTTCCCACCCAT
<i>Was</i> D	Fwd: TATCATTGGCAACGTCCAAA
	Rv: GACATGGTGGCACTTGTCTG
<i>Btk</i> ND	Fwd: TGGGGCTTATGGTTTCAAAG
	Rv: CCCAGTCGAAGTGTGGTTTT
<i>Btk</i> D	Fwd: GGCTATGTTGTCTGGGCTCTA
	Rv: TGTTCGCTGGGTTACCTTC

Tabla S2. Anticuerpos empleados en el desarrollo de esta tesis

Anticuerpo Primario	Referencia	Dilución
Rho GDI	ABClonal ab133248	1:5000
Was	Santa Cruz sc-13139	1:500
Btk	Santa Cruz sc-28387	1:500

Tabla S3. Cuantificación por espectrofotómetro de ARN total extraído.

Muestras	Concentración (ng/μl)	RNA Total en 20 μl (μg)	Ratio 260/280	Ratio 260/230
	Fórmula: Abs260*40*(10/0.52)			

MEL Btk	1394	27.88	2.088	2.296
MEL Was^{-/-} 1	991	19.82	2.091	2.204
MEL Was^{-/-} 2	1141	22.82	2.093	2.134
MEL	2072	41.44	2.005	2.219
MEL-R	1480	29.6	2.085	2.159

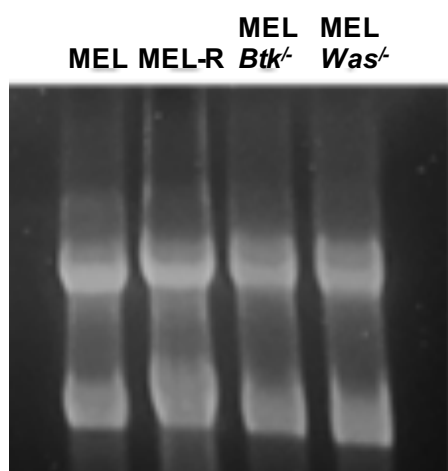


Figura S1. Validación de la integridad del ARN total extraído. Electroforesis en geles de agarosa en condiciones desnaturalizantes de ARN total extraído de las muestras 1) MEL 2) MEL-R 3) MEL *Btk*^{-/-} y 4) MEL *Was*^{-/-}.

Tabla S4. Primers usados para la cuantificación de la expresión génica relativa mediante PCR a tiempo real.

Nombre del primer	Secuencia (5'-3')
<i>Gzmb</i>	Fw: CCACTCTCGACCCTACATGG
	Rv: GGCCCCCAAAGTGACATTTATT
<i>Csf2rb</i>	Fwd: TCCAGCCAGATCGTGACCT
	Rv: AATCCCCAAGAGATACACTCCA
<i>Srsf2</i>	Fw: CGCGCTCCAGATCAACCTC
	Rv: CTTGGACTCTCGCTTCGACAC
<i>Gata2</i>	Fw: CACCCCGCCGTATTGAATG

	Rv: CCTGCGAGTCGAGATGGTTG
<i>c-Myc</i>	Fw: ATGCCCCTCAACGTGAAC TTC
	Rv: CGCAACATAGGATGGAGAGCA

Tabla S5. Equivalentes de genes pertenecientes a la familia *Wasp* en *Mus musculus*

Genes de la familia <i>Wasp</i>	
Miembros	Equivalente en <i>M. musculus</i>
N-Wasp	Wasl
WAVE	Wasf1
	Wasf2
	Wasf3
Wash Complex	Washc1
	Orf19
	Wash
	Wash1
WHAMM	Whamm
	Kiaa1971
	Whdc1

Código de análisis realizados

LINUX COMMAND

###Trimming of adaptors –

####Btk_2 trimming

```
trimmomatic SE -phred33
/home/hsosa/RNAseq/new_data/project_DKR/DKR-
2_S2_R1_001.fastq ./Btk_2.fastq
ILLUMINACLIP:../adapters.fa:2:30:10 SLIDINGWINDOW:4:15
LEADING:20 TRAILING:20 MINLEN:36
```

####Was_1 trimming

```
trimmomatic SE -phred33
/home/hsosa/RNAseq/new_data/project_DKR/DKR-
3_S3_R1_001.fastq ./Was_1.fastq
ILLUMINACLIP:../adapters.fa:2:30:10 SLIDINGWINDOW:4:15
LEADING:20 TRAILING:20 MINLEN:36
```


####Was_2 trimming

```
trimmomatic SE -phred33
/home/hsosa/RNAseq/new_data/project_DKR/DKR-
4_S4_R1_001.fastq ./Was_2.fastq
ILLUMINACLIP:../adapters.fa:2:30:10 SLIDINGWINDOW:4:15
LEADING:20 TRAILING:20 MINLEN:36
```

####Was_1 trimming

```
trimmomatic SE -phred33
/home/hsosa/RNAseq/new_data/project_DKR/DKR-
5_S5_R1_001.fastq ./MEL_1.fastq
ILLUMINACLIP:../adapters.fa:2:30:10 SLIDINGWINDOW:4:15
LEADING:20 TRAILING:20 MINLEN:36
```

####MEL_2 trimming

```
trimmomatic SE -phred33
/home/hsosa/RNAseq/new_data/project_DKR/MEL.fastq ./MEL_
2.fastq ILLUMINACLIP:../adapters.fa:2:30:10
SLIDINGWINDOW:4:15 LEADING:20 TRAILING:20 MINLEN:36
```

####MELR_2 trimming

```
trimmomatic SE -phred33
/home/hsosa/RNAseq/new_data/project_DKR/MELR.fastq ./MEL
R_2.fastq ILLUMINACLIP:../adapters.fa:2:30:10
SLIDINGWINDOW:4:15 LEADING:20 TRAILING:20 MINLEN:36
```

###Genome and annotation were imported from these links
Genome:
[ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/](ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M24/GRCm38.primary_assembly.genome.fa.gz)
[release_M24/GRCm38.primary_assembly.genome.fa.gz](ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M24/GRCm38.primary_assembly.genome.fa.gz)

Reference annotation:
[ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/](ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M24/gencode.vM24.primary_assembly.annotation.gtf.gz)
[release_M24/gencode.vM24.primary_assembly.annotation.gtf](ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M24/gencode.vM24.primary_assembly.annotation.gtf.gz)
.gz

###Index building:

```
STAR --runThreadN 4 --runMode genomeGenerate --genomeDir
/home/hsosa/RNAseq/RNA-seq2/m_musculus --
genomeFastaFiles GRCm38.primary_assembly.genome.fa --
sjdbGTFfile gencode.vM24.primary_assembly.annotation.gtf
--sjdbOverhang 74
```

###Alignment

####Btk_2

```
STAR --genomeDir /home/hsosa/RNAseq/RNA-seq2/m_musculus --
runThreadN 8 --readFilesIn /home/hsosa/RNAseq/RNA-
seq2/trimmed_files/Btk_2.fastq --outFileNamePrefix
/home/hsosa/RNAseq/RNA-seq2/alignment/Btk_2/Btk_2 --outSAMtype
BAM SortedByCoordinate --outSAMunmapped Within --
outSAMattributes Standard
```

####Was_1

```
STAR --genomeDir /home/hsosa/RNAseq/RNA-seq2/m_musculus --
runThreadN 8 --readFilesIn /home/hsosa/RNAseq/RNA-
seq2/trimmed_files/Was_1.fastq --outFileNamePrefix
/home/hsosa/RNAseq/RNA-seq2/alignment/Was_1/Was_1 --outSAMtype
BAM SortedByCoordinate --outSAMunmapped Within --
outSAMattributes Standard
```

####Was_2

```
STAR --genomeDir /home/hsosa/RNAseq/RNA-seq2/m_musculus --
runThreadN 8 --readFilesIn /home/hsosa/RNAseq/RNA-
seq2/trimmed_files/Was_2.fastq --outFileNamePrefix
/home/hsosa/RNAseq/RNA-seq2/alignment/Was_2/Was_2 --outSAMtype
BAM SortedByCoordinate --outSAMunmapped Within --
outSAMattributes Standard
```

####MEL_2

```
STAR --genomeDir /home/hsosa/RNAseq/RNA-seq2/m_musculus --
runThreadN 8 --readFilesIn /home/hsosa/RNAseq/RNA-
seq2/trimmed_files/MEL_2.fastq --outFileNamePrefix
/home/hsosa/RNAseq/RNA-seq2/alignment/MEL_2/MEL_2 --outSAMtype
BAM SortedByCoordinate --outSAMunmapped Within --
outSAMattributes Standard
```

####MELR_2

```
STAR --genomeDir /home/hsosa/RNAseq/RNA-seq2/m_musculus --
runThreadN 8 --readFilesIn /home/hsosa/RNAseq/RNA-
seq2/trimmed_files/MELR_2.fastq --outFileNamePrefix
/home/hsosa/RNAseq/RNA-seq2/alignment/MELR_2/MELR_2 --
outSAMtype BAM SortedByCoordinate --outSAMunmapped Within --
outSAMattributes Standard
```

###Estimation of gene counts - featureCounts

Btk_2

```
featureCounts -a /home/hsosa/RNAseq/RNA-
seq2/gencode.vM25.annotation.gtf -o Btk_2.txt
/home/hsosa/RNAseq/RNA-
seq2/alignment/Btk_2/Btk_2Aligned.sortedByCoord.out.bam
```

```
### Was_1
featureCounts -a /home/hsosa/RNAseq/RNA-
seq2/gencode.vM25.annotation.gtf -o Was_1.txt
/home/hsosa/RNAseq/RNA-
seq2/alignment/Was_1/Was_1Aligned.sortedByCoord.out.bam
```

```
### Was_2
featureCounts -a /home/hsosa/RNAseq/RNA-
seq2/gencode.vM25.annotation.gtf -o Was_2.txt
/home/hsosa/RNAseq/RNA-
seq2/alignment/Was_2/Was_2Aligned.sortedByCoord.out.bam
```

```
### MEL_2
featureCounts -a /home/hsosa/RNAseq/RNA-
seq2/gencode.vM25.annotation.gtf -o MEL_2.txt
/home/hsosa/RNAseq/RNA-
seq2/alignment/MEL_2/MEL_2Aligned.sortedByCoord.out.bam
```

```
### MELR_2
featureCounts -a /home/hsosa/RNAseq/RNA-
seq2/gencode.vM25.annotation.gtf -o MELR_2.txt
/home/hsosa/RNAseq/RNA-
seq2/alignment/MELR_2/MELR_2Aligned.sortedByCoord.out.ba
m
```

R CODE

```
```{r, include = FALSE}
Load libraries
library(edgeR)
library(biomaRt)
library(VennDiagram)
library(RColorBrewer)
library(fgsea)
library(clusterProfiler)
library(reactome.db)
library(statmod)
library(gplots)
library(dplyr)
library(tidyverse)
library(DOSE)
```

```

library(pathview)
library(ReactomePA)
library(grid)
library(lattice)
```

### Read feature counts files
```{r}
fcount_Btk_2 <-
read.table("featureCounts_files/Btk_2.txt", row.names =
NULL, header = TRUE)[, c(1, 7)]

fcount_Was_1 <-
read.table("featureCounts_files/Was_1.txt", row.names =
NULL, header = TRUE)[, c(1, 7)]

fcount_Was_2 <-
read.table("featureCounts_files/Was_2.txt", row.names =
NULL, header = TRUE)[, c(1, 7)]

fcount_MEL_2 <-
read.table("featureCounts_files/MEL_2.txt", row.names =
NULL, header = TRUE)[, c(1, 7)]

fcount_MELR_2 <-
read.table("featureCounts_files/MELR_2.txt", row.names =
NULL, header = TRUE)[, c(1, 7)]

Merge the fcounts_files into a single file
```{r}
fcount <- reduce(function(x, y) merge(x, y, by =
"Geneid"), list(fcount_Btk_2, fcount_Was_1,
fcount_Was_2, fcount_MEL_1, fcount_MEL_2, fcount_MELR_2
```

```{r, include = FALSE}
### mport count matrix done with featureCounts
matrix_counts = read.table(file = "genecounts.txt",
row.names = 1)

### Retrieve the information
matrix_symbols <- getBM(filters = "ensembl_gene_id",
attributes = c("ensembl_gene_id", "external_gene_name",
"entrezgene_id", "description"), values = names_max,
mart = mart)

```

```

#### Format the matrix using gene symbols
matrix_3 <- merge(matrix_2, matrix_symbols, merge =
"ensembl_gene_id")
head(matrix_2)

#### Change the columns names
colnames(matrix_counts) <- c("Btk_2", "Was_1", "Was_2",
"MEL_2", "MELR_2")
head(matrix_counts)

#### Explore the object
class(matrix_counts)
dim(matrix_counts)

#### Creating the grouping factor
group = factor(gsub(pattern = "_[0-9]", "",
colnames(matrix_counts)))

#### Create the DGE object
dge = DGEList(matrix_counts, group = group)

#### Convert DGEList to DESeqDataSet
class(dge)
dim(dge)
```

Filter low-count genes

We attempt to find and filter low counts genes.

```{r, include = FALSE}
#### filter using edgeR
keep <- filterByExpr(dge)

#### Number of genes to keep
sum(keep)
#### Naïve filtering
keep2 <- rowSums(dge$counts) > 1

#### Number of genes to keep2
sum(keep2)

#### Comparing both approaches
sum(keep&keep2) # keep2 includes keep

#### Exploring subset distribution
summary((dge$counts)[keep,])

```

```

#### Subsetting dge. The option keep.lib.sizes=FALSE
causes the library sizes to be recomputed after the
filtering
dge <- dge[keep, , keep.lib.sizes=FALSE]

```

TMM Normalization

Next, we are going to normalize the samples within
samples and within groups. This step allows to eliminate
the batch effect that we might have by using samples
with different time preparation and conditions.

```{r, include = FALSE}
#### edgeR proposed normalization
dge <- calcNormFactors(dge, method = "TMM")

# Exploring results
dge$samples
```

Data Exploration

Before any analysis, we explore the clustering of the
samples using the first and second dimension.

Multidimensional scaling
```{r}
png("MDS.png")
plotMDS(dge)
dev.off()
```

Explore correlation

Next, we explore correlation between samples
```{r}
#### raw gene count correlation
cor(dge$counts, method = "spearman")
```

DGE analysis

In order to make multiple comparison, we need first to
create our data matrix. Here we will compare all groups,
not having an intercept that we will take as a
reference.

```{r, include = FALSE}

```

```

#### Create design matrix
design <- model.matrix(~ 0 + group, data = dge$samples)

#### Assign group to dge columns name
colnames(design) <- levels(dge$samples$group)

#### Estimate dispersion
y <- estimateDisp(dge, design, robust = TRUE)

#### Fit modelling. glmQLFit gives special attention to
handling of zero counts, and in particular to situations
when fitted values of zero provide no useful residual.
fit <- glmQLFit(y, design)
```

```

```

Contrast matrix

```

After creating a design matrix, we also need a contrast matrix that will allow comparison in pairs between groups. The comparison we would like to see are the following: MELRvsMEL = MELR - MEL, BtkvsMEL = Btk - MEL, WasvsMEL = Was - MEL, WasvsBtk = Was - Btk, WasvsMELR = Was - MELR, BtkvsMELR = Btk - MELR.

```

```{r, include = FALSE}
#### Prepare contrast matrix
my.contrast <- makeContrasts(MELRvsMEL = MELR - MEL,
BtkvsMEL = Btk - MEL, WasvsMEL = Was - MEL, WasvsBtk =
Was - Btk, WasvsMELR = Was - MELR, BtkvsMELR = Btk -
MELR, levels = design)
```

```

### ### Differential Expression

Now, we can find the DGE.

```

Explore WasvsMEL comparison
```{r, include = FALSE}
#### Replaces likelihood ratio tests with empirical Bayes
quasi-likelihood F-tests.
qlf_WasvsMEL <- glmQLFTest(fit, contrast= my.contrast[,
"WasvsMEL"])

#### Extract most differentially expressed genes from the
test object
topTags(qlf_WasvsMEL)

#### Extract significant genes from qlf_WasvsMEL (logfc
greater than 2 and p.adjusted less than 0.01)

```

```

de_WasvsMEL <- decideTestsDGE(qlf_WasvsMEL, p = 0.05,
adjust="BH", lfc = 1)

### Extract data of differentially expressed genes
de_WasvsMEL_stat <-
qlf_WasvsMEL$table[as.logical(de_WasvsMEL), ]

myc
### Select keys for inquiry
genes_WasvsMEL <- rownames(de_WasvsMEL_stat)

### Extract gene symbols
gene_symbols_WasvsMEL <- getBM(filters =
"ensembl_gene_id_version", attributes =
c("ensembl_gene_id_version", "external_gene_name",
"entrezgene_id", "description", "chromosome_name",
"start_position", "end_position", "strand"), values =
genes_WasvsMEL, mart = mart)

### Assign rownames to a column to do the merging
de_WasvsMEL_stat$ensembl_gene_id_version <-
rownames(de_WasvsMEL_stat)

### Merge
de_WasvsMEL_stat <- merge(gene_symbols_WasvsMEL,
de_WasvsMEL_stat, by = "ensembl_gene_id_version")

### Explore the object
head(de_WasvsMEL_stat)

### Extract raw counts from dge$table
### Find the positions of de_WasvsMEL_stat genes in
dge$table
WasvsMEL_pos <- which(rownames(dge$counts) %in%
genes_WasvsMEL)

### Extract the values from dge
WasvsMEL_ext <-
as.data.frame(dge$counts[WasvsMEL_pos, ])
WasvsMEL_ext$ensembl_gene_id_version <-
rownames(WasvsMEL_ext)
WasvsMEL_ext <- WasvsMEL_ext[, c(6, 1:5)]
rownames(WasvsMEL_ext) <- NULL

### Compute the mean value for Was observations and
extract the mean and MEL columns row values
raw_WasMEL <- WasvsMEL_ext %>% rowwise() %>%
mutate(Avg_Was=mean(c(Was_1, Was_2))) %>%

```



```

dplyr::select(c(ensembl_gene_id_version, Avg_Was,
MEL_2))
head(raw_WasMEL)

### Add Avg_Was and MEL raw counts to de_WasvsMEL_stat
de_WasvsMEL_stat <- merge(de_WasvsMEL_stat, raw_WasMEL,
by = "ensembl_gene_id_version")

### Explore
head(de_WasvsMEL_stat)

### Save to csv
write.csv(de_WasvsMEL_stat, "WasvsMEL_DGE.csv")
```

See up and down regulated genes
```{r}
summary(de_WasvsMEL)
```

Plot WasvsMEL

Let's observe the DEGs. Positive logFC make reference
genes upregulated in Was while negative logFC point to
genes downregulated in Was or upregulated in MEL.

Smear plot
```{r, include = FALSE}
png("Smear_plot_WasvsMEL.png")
plotMD(qlf_WasvsMEL)
dev.off()
```

Volcano plot
```{r, include = FALSE}
### Format the table with UP-DOWN-NOT SIGNIFICANT

de_diff <- qlf_WasvsMEL$table
head(de_diff)
de_diff$diffexpressed <- "NO"

# if log2Foldchange > 1 and pvalue < 0.05, set as "UP"
de_diff$diffexpressed[de_diff$logFC >= 1 &
de_diff$PValue < 0.05] <- "UP"
# if log2Foldchange < -0.6 and pvalue < 0.05, set as
"DOWN"

```

```

de_diff$diffexpressed[de_diff$logFC <= -1 &
de_diff$PValue < 0.05] <- "DOWN"

### Customize the dots
mycolors <- c("#356852", "#E35925", "grey79")
names(mycolors) <- c("DOWN", "UP", "NO")

### Plot volcano
pdf("Volcano_paper_MELWAS_MEL.pdf")
ggplot(de_diff) + geom_point(aes(x=logFC, y=-
log10(PValue), color=diffexpressed)) +
  xlab("log2 fold change") +
  ylab("-log10 adjusted p-value") +
  #scale_y_continuous(limits = c(0,50)) +
  scale_colour_manual(values = mycolors)+
  theme(legend.position = "none",
        plot.title = element_text(size = rel(1.5),
hjust = 0.5),
        axis.title = element_text(size =
rel(1.25))) + theme(panel.grid.major = element_blank(),
panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line =
element_line(colour = "black"))
dev.off()

### volcano plot
{png("Volcano_plot_WasvsMEL.png")
smoothScatter(de_diff$logFC, -log10(de_diff$PValue),
              ylim=c(-1, 10), xlim = c(-7, 7),
              xlab="log fold-change", ylab="Negative
log10 pvalue",
              main="Volcano plot: Was vs MEL\n", cex =
0.8)
              grid(col="grey75")
              abline(h=0, v=0,
col="grey50", lty=2)
              points(qlf_WasvsME
L$table$logFC[as.logical(de_WasvsMEL)],
              -
log10(qlf_WasvsMEL$table$PValue[as.logical(de_WasvsMEL)]
),
col="salmon", pch=16, cex=1)
dev.off()
}
```


Over-representation analysis of common genes


```

```

Was_up genes
```{r}
Was_up <- WasvsMEL_dge[which(WasvsMEL_dge$logFC > 0), ]

#### Overrepresentation test
mkk <- enrichKEGG(gene = Was_up$entrezgene_id,
                  organism = 'mmu')

#### Convert the result to friendly version
mkk_up_df <- as.data.frame(mkk)
```

Genes downregulated in Was
```{r}
#### Extract downreg genes in Was
Was_down <- WasvsMEL_dge[which(WasvsMEL_dge$logFC <
0), ]

#### Overrepresentation analysis
mkk_down <- enrichKEGG(gene = Was_down$entrezgene_id,
                      organism = 'mmu')

#### Convert results
mkk_down_df <- as.data.frame(mkk_down)
```

####GSEA

Load the file of the pathway using fgsea
gmtPathways()
```{r}
hallmarks <-
gmtPathways("/Users/biososa/EdgeR/h.all.v7.0.symbols.gmt
")
```

Add HUGO genes (GSEA has human notation and we work
with mouse)
```{r}
# Basic function to convert mouse to human gene names
convertMouseGeneList <- function(x){

human = useMart("ensembl", dataset =
"hsapiens_gene_ensembl")
mouse = useMart("ensembl", dataset =
"mmusculus_gene_ensembl")

```

```

genesV2 = getLDS(attributes = c("mgi_symbol"), filters =
c("mgi_symbol"), values = x , mart = mouse, attributesL
= c("hgnc_symbol"), martL = human, uniqueRows=T)
humanx <- unique(genesV2[, 2])

# Print the first 6 genes found to the screen
print(head(humanx))
return(humanx)
}

HUGO <- convertMouseGeneList(DGE$external_gene_name)
HUGO_Was_up_DEG <-
convertMouseGeneList(Was_up_DGE$external_gene_name)
HUGO_Was_down_DEG <-
convertMouseGeneList(Was_down_DGE$external_gene_name)

human = useMart(host="useast.ensembl.org",
biomart="ENSEMBL_MART_ENSEMBL",
dataset="hsapiens_gene_ensembl")

genes_htom <- getBM(filters = "hgnc_symbol", attributes
= c("mmusculus_homolog_associated_gene_name",
"external_gene_name", "mmusculus_homolog_ensembl_gene"),
values = HUGO, mart = human)

genes_htom_Was_up_DGE <- getBM(filters = "hgnc_symbol",
attributes = c("mmusculus_homolog_associated_gene_name",
"external_gene_name", "mmusculus_homolog_ensembl_gene"),
values = HUGO_Was_up_DEG, mart = human)

genes_htom_Was_down_DGE <- getBM(filters =
"hgnc_symbol", attributes =
c("mmusculus_homolog_associated_gene_name",
"external_gene_name", "mmusculus_homolog_ensembl_gene"),
values = HUGO_Was_down_DEG, mart = human)

colnames(genes_htom)[2] <- "HUGO.symbol"
colnames(genes_htom_Was_up_DGE)[2] <- "HUGO.symbol"
colnames(genes_htom_Was_down_DGE)[2] <- "HUGO.symbol"
colnames(genes_htom)[3] <- "ensembl_gene_id"
colnames(genes_htom_Was_up_DGE)[3] <- "ensembl_gene_id"
colnames(genes_htom_Was_down_DGE)[3] <-
"ensembl_gene_id"

### Find duplicates for GSEA table
dup <-
unique(genes_htom$HUGO.symbol[duplicated(genes_htom$HUGO
.symbol)])

```

```

table(genes_htom$HUGO.symbol[genes_htom$HUGO.symbol %in%
dup])

### Find duplicates for Was_up_DGE
dup_was_up <-
unique(genes_htom_Was_up_DGE$HUGO.symbol[duplicated(gene
s_htom_Was_up_DGE$HUGO.symbol)])
table(genes_htom_Was_up_DGE$HUGO.symbol[genes_htom_Was_u
p_DGE$HUGO.symbol %in% dup_was_up])

### Find duplicates for Was_down_DGE
dup_was_down <-
unique(genes_htom_Was_down_DGE$HUGO.symbol[duplicated(gene
s_htom_Was_down_DGE$HUGO.symbol)])
table(genes_htom_Was_down_DGE$HUGO.symbol[genes_htom_Was
_down_DGE$HUGO.symbol %in% dup_was_down])
```



```

### Some cases have more than 1 gene expression, choose
the first occurrence
```{r}
Eliminate duplicates for GSEA TABLES
for (duplicates in dup) {
 x <- which(genes_htom$HUGO.symbol %in% duplicates)
 toDrop <- x[2:length(x)]
 genes_htom <- genes_htom[-toDrop,]
}

Check elimination
table(genes_htom$external_gene_name[genes_htom$external_
gene_name %in% dup])

Eliminate duplicates for Was_up_DGE
for (duplicates_2 in dup_was_up) {
 y <- which(genes_htom_Was_up_DGE$HUGO.symbol %in%
duplicates_2)
 toDrop_2 <- y[2:length(y)]
 genes_htom_Was_up_DGE <- genes_htom_Was_up_DGE[-
toDrop_2,]
}

Check elimination
table(genes_htom_Was_up_DGE$external_gene_name[genes_ho
m_Was_up_DGE$external_gene_name %in% dup_was_up])

Eliminate duplicates for Was_down_DGE
for (duplicates_3 in dup_was_down) {

```


```

```

      z <- which(genes_htom_Was_down_DGE$HUGO.symbol %in%
duplicates_3)
      toDrop_3 <- z[2:length(z)]
      genes_htom_Was_down_DGE <- genes_htom_Was_down_DGE[-
toDrop_3, ]
    }

    ### Check elimination
    table(genes_htom_Was_down_DGE$external_gene_name[genes_h
tom_Was_down_DGE$external_gene_name %in% dup_Was_down])
    ""

    ### Merge HUGO symbols with DGE
    ""{r}
    DGE_2 <- merge(DGE, genes_htom, by = "ensembl_gene_id")
    Was_up_2 <- merge(Was_up_DGE, genes_htom_Was_up_DGE, by
= "ensembl_gene_id")
    Was_down_2 <- merge(Was_down_DGE,
genes_htom_Was_down_DGE, by = "ensembl_gene_id")
    ""

    ### Order FC from top to bottom
    ""{r}
    results.ord <- DGE_2[order(-DGE_2[, "logFC"]), ]
    head(results.ord)

    ### Assign sign to F. Since F captures better the
changes in raw count, we proceed to use it

    prueba <- results.ord[10000: 10005, ]

    F_2 <- data.frame()
    for(i in 1:length(results.ord$logFC)){
      if (results.ord$logFC[i] < 0){
        new <- results.ord$F[i] * -1
        F_2 <- rbind(F_2, new)
      }
      else if (results.ord$logFC[i] > 0 |
results.ord$logFC[i] == 0){
        new <- results.ord$F[i]
        F_2 <- rbind(F_2, new)
      }
    }
  }

  names(F_2) <- "F_2"

  ### Append column to df
  results.ord <- cbind(results.ord, F_2)

```

```

#### Assign ranks
ranks <- results.ord$F_2
names(ranks) <- results.ord$HUGO.symbol
head(ranks)
```

Run GSEA
```{r}
fgseaRes <- fgseaMultilevel(hallmarks, ranks, minSize=5,
maxSize = 500)
class(fgseaRes)
dim(fgseaRes)
head(fgseaRes)
fgseaRes
```

See the results
```{r}
barplot(ranks)
```

Order according significance
```{r}
fgseaResOrdered <- fgseaRes[order(padj), ]
fgseaResOrdered

## How many "interesting" pathways we found (FDR
threshold)
sum(fgseaResOrdered[, padj < 0.29], na.rm = TRUE)

## Significant Pathways
sigpath <- fgseaResOrdered[which(fgseaResOrdered[, padj
< 0.25]),]
sigpath

## Save table into text format
fwrite(sigpath, file="Significant pathways in
WasvsMEL.txt", sep="\t", sep2=c("", " ", ""))
```

GSEA Pathways related to Was
```{r}
#### Pathways more related to Was
Up_Was <- sigpath[which(sigpath$NES > 0), ]
Up_Was
```

GSEA Pathways downregulated in Was (more related to
MEL)
```{r}

```

```
Down_Was <- sigpath[which(sigpath$NES < 0), ]
Down_Was
```

```

### ###DTU

#### ## Procedures

#### ### Alignment and quantification with Salmon

Transcript-level quantification was done with Salmon. Transcriptome, genome and gtf files were downloaded from Genecode (genecode.vM25).  
[https://www.gencodegenes.org/mouse/release\\_M25.html](https://www.gencodegenes.org/mouse/release_M25.html)

Salmon gets the relative abundance of transcripts in TPM, and it is unable to detect new splicing events.

#### ### Prepare salmon files

Salmon indexing requires the names of the genome targets, which is extractable by using the grep command:

```
```{r, include = FALSE}
### Metadata
grep "^>" <(gunzip -c
GRCm38.primary_assembly.genome.fa.gz) | cut -d " " -f 1
> decoys.txt
sed -i.bak -e 's/>//g' decoys.txt
```

Concatenate transcriptome and genome reference file for index

```
cat gencode.vM23.transcripts.fa.gz
GRCm38.primary_assembly.genome.fa.gz > gentrome.fa.gz
```

Indexing

```
salmon index -t gentrome.fa.gz -d decoys.txt -p 12 -i
salmon_index --gencode
```

Quantification

Btk_2

```
salmon quant -i /home/hsosa/RNAseq/RNA-seq2/salmon_index
-l A -r /home/hsosa/RNAseq/data/DKR-2_clean3p_Q20.fastq
-o /home/hsosa/RNAseq/RNA-
seq2/salmon_alignment/Btk_2.salmon --seqBias --useVBOpt
```

Was_1

```
salmon quant -i /home/hsosa/RNAseq/RNA-seq2/salmon_index
-l A -r /home/hsosa/RNAseq/data/DKR-3_clean3p_Q20.fastq
```



```

-o /home/hsosa/RNAseq/RNA-
seq2/salmon_alignment/Was_1.salmon --seqBias --useVBOpt

### Was_2
salmon quant -i /home/hsosa/RNAseq/RNA-seq2/salmon_index
-l A -r /home/hsosa/RNAseq/data/DKR-4_clean3p_Q20.fastq
-o /home/hsosa/RNAseq/RNA-
seq2/salmon_alignment/Was_2.salmon --seqBias --useVBOpt

### MEL_2
salmon quant -i /home/hsosa/RNAseq/RNA-seq2/salmon_index
-l A -r /home/hsosa/RNAseq/data/MEL.fastq -o
/home/hsosa/RNAseq/RNA-
seq2/salmon_alignment/MEL_2.salmon --seqBias --useVBOpt

### MELR_2
salmon quant -i /home/hsosa/RNAseq/RNA-seq2/salmon_index
-l A -r /home/hsosa/RNAseq/data/MELR.fastq -o
/home/hsosa/RNAseq/RNA-
seq2/salmon_alignment/MELR_2.salmon --seqBias --useVBOpt
```

```

### ### Load files with tximport

Now, we import the data obtained in Salmon to R.

```

```{r, include = FALSE}
### Specify files name
dirs <-
list.files("/Volumes/Transcend/salmon_quant/MELvsWas")
### List files
quant_files <- list.files("MELvsWas/", pattern=
"quant.sf", recursive = TRUE, full.names = TRUE)
### Assign names to quant files
names(quant_files) <- dirs

### Inspect salmon output
quants <- read_tsv(quant_files[1])
head(quants)
```

```{r, include = FALSE}

## Creating a trascript database
txdb <-
GenomicFeatures::makeTxDbFromGFF("gencode.vM25.annotatio
n.gtf.gz")

### See keys and columns of the database

```

```

keytypes(txdb)
columns(txdb)

### Get names of all transcripts
txdf <- AnnotationDbi::select(txdb, keys(txdb,
"GENEID"), "TXNAME", "GENEID")

tab <- table(txdf$GENEID)

txdf$ntx <- tab[match(txdf$GENEID, names(tab))]
```

```{r, include = FALSE}
### Import transcript-level counts
txi <- tximport(quant_files, type="salmon", txOut=TRUE,
ignoreTxVersion = TRUE, countsFromAbundance =
"scaledTPM")
```

```{r, include = FALSE}
### Create design matrix
myDesign <- data.frame(
  sample_id = colnames(txi$abundance),
  condition = c("MEL", "Was", "Was")
)
```

```{r}
### Filter low and zero counts
cts <- txi$counts
cts <- cts[rowSums(cts) > 0,]

### Explore object
head(cts)
dim(cts)
```

```{r, include = FALSE}
### Transcript-to-gene mapping through txdb
txdf <- txdf[match(rownames(cts), txdf$TXNAME), ]
all(rownames(cts) == txdf$TXNAME)

counts <- data.frame(gene_id=txdf$GENEID,
                     feature_id=txdf$TXNAME,
                     cts)
```

###DRIMSeq

```

We create a DRIMSeq with counts and matrix design

```
```{r}
d <- dmDSdata(counts=counts, samples= myDesign)
d
```
```

In total we have around 22000 genes. Therefore, it is very important to filter low-count genes.

### ### Filtering

Here we filter the object by the following parameters: `n` is the total number of samples, in our case it is 3. Then we have `n.small` which is the sample size (numbers of replicates in a group), which in our case is 1 (MEL has no replicates). Thus, we filter by asking for isoform counts of 10 at least in 1 sample, a relative abundance portion of 0.1 and total count of corresponding gene is at least 10 in all 3 samples.

```
```{r, include = FALSE}
n <- 3 ### Number of samples
n.small <- 1 ### Smalles number of samples in a group
d <- dmFilter(d,
              min_samps_feature_expr=n.small,
min_feature_expr=5,
              min_samps_feature_prop=n.small,
min_feature_prop=0.1,
              min_samps_gene_expr= 2, min_gene_expr=10)
d
table(table(counts(d)$gene_id))
```
```

After filtering we ended up with ~8000 genes that have more than one isoform.

We create a design matrix, using conditions to make the regression.

```
```{r}
### Matrix design
design_full <- model.matrix(~condition,
data=DRIMSeq::samples(d))
design_full
```
```

Then three functions were used to estimate parameters for DTU: precision, dispersion. Dispersion is inversely related to precision. ( $\text{dispersion} = 1 / 1 + \text{precision}$ ).

We test for the difference between Was and MEL called "conditionWas"

```

```{r, include= FALSE}
###Extract results
set.seed(1)
system.time({
  d <- dmPrecision(d, design=design_full)
  d <- dmFit(d, design=design_full)
  d <- dmTest(d, coef="conditionWas")
})
```

```

We then run the DRIMSeq function result. It generates p-values per gene which test for DTU within genes or transcript (we did both).

```

```{r, include = FALSE}
### DTU at gene level
res <- DRIMSeq::results(d)
res <- res[order(res$pvalue, decreasing = FALSE), ]
top_gene_id <- res$gene_id[1]
head(res)

```

```

### DTU at transcript level
res.txp <- DRIMSeq::results(d, level="feature")
head(res.txp)

```

```

### Remove NA values
no.na <- function(x) ifelse(is.na(x), 1, x)
res$pvalue <- no.na(res$pvalue)
res.txp$pvalue <- no.na(res.txp$pvalue)
```

```

### ### Stager

Any regular DTU analysis finds first what genes are involved in DTU and then find the transcripts of those genes involved in DTU. We perform that process with stager

```

```{r}
### Remove version id from genes and tx
pScreen <- res$pvalue
strp <- function(x) gsub(pattern = "[.][0-9]", "", x)

names(pScreen) <- strp(res$gene_id)

pConfirmation <- matrix(res.txp$pvalue, ncol=1)
rownames(pConfirmation) <- strp(res.txp$feature_id)

### Create a two column dataframe

```

```

tx2gene <- res.txp[,c("feature_id", "gene_id")]
for (i in 1:2) tx2gene[,i] <- strp(tx2gene[,i])

### Test for OFDR for cutoff of 0.05
stageRobj <- stageRTx(pScreen=pScreen,
pConfirmation=pConfirmation,
                        pScreenAdjusted=FALSE,
tx2gene=tx2gene)
stageRobj <- stageWiseAdjustment(stageRobj,
method="dtu", alpha=0.05)
suppressWarnings({
  drim.padj <- getAdjustedPValues(stageRobj,
order=FALSE,

onlySignificantGenes=TRUE)
})

dim(drim.padj)
drim.padj

### Save drim.padj to final
final <- drim.padj
colnames(final)[1] <- "ensembl_gene_id"

#### Add gene symbols to final table
#### Create mart object db
mart = useMart(host="useast.ensembl.org",
biomart="ENSEMBL_MART_ENSEMBL",
dataset="mmusculus_gene_ensembl")

#### Get symbols
symbols <- getBM(filters = "ensembl_gene_id", attributes
= c("ensembl_gene_id", "external_gene_name",
"entrezgene_id","description"), values =
final$ensembl_gene_id, mart = mart)

#### Merge
final <- merge(final, symbols, by = "ensembl_gene_id" )
head(final)

#### The transcripts with values in the column,
transcript, less than 0.05 pass the confirmation stage
on a target 5% overall false discovery rate, or OFDR

significant_tx <- final %>% filter(transcript < 0.05)
dim(significant_tx)
length(significant_tx$external_gene_name)

```

```

write.csv(significant_tx,
"Significant_DTU_WasvsMEL.csv")
head(significant_tx)

```

GO annotation
Now we proceed to annotate the genes for DTUs. We
perform BP (Biological Process), CC (cellular
compartment) and MF (Molecular Function)
```{r, include = FALSE}
### Biological process GO
bp_tdx <- enrichGO(gene = significant_tx$entrezgene_id,
                    OrgDb      = org.Mm.eg.db,
                    ont         = "BP",
                    pAdjustMethod = "BH",
                    pvalueCutoff  = 0.01,
                    qvalueCutoff  = 0.05)

bp_tdx_df <- as.data.frame(bp_tdx)
```

```

## 8.1 Comunicaciones a congresos nacionales e internacionales derivadas de esta tesis de Maestría

1. **5to Simposio Argentino de Jóvenes Investigadores en Bioinformática. (Presentación oral). Agosto 2020.**

Título del trabajo: RNA-seq analysis reveals importance of regulation of actin cytoskeleton in erythroleukemia cells

Link: <https://rsg-argentina.netlify.app/talk/5sajib/>

2. **Women Bioinformatics Network LA (Póster digital). Septiembre 2020.** Título del trabajo: RNA Sequencing analysis revealed the importance of RHO GTPases and integrins in virus-induced erythroleukemia cells

Link: <https://wbds.la/conferences/1WBDSLAC/diseasesVPS.html>

3. **Jornadas de jóvenes universitarios. UNA PY. (Póster digital y presentación oral). Septiembre 2020.**

Título del trabajo: Análisis de la expresión diferencial en células Eritroleucémicas con delección de proteínas específicas del linaje hematopoyético: análisis funcional de Was

Link de la lista de seleccionados: <http://investigacion.una.py/jovenes-investigadores/wp-content/uploads/2020/09/JJI-2020-Seleccionados-Salud.pdf>. [https://www.youtube.com/watch?v=4TU\\_5AaIrtg](https://www.youtube.com/watch?v=4TU_5AaIrtg)

4. **Jornada de Difusión de Proyectos de Investigación de la FP-UNA (Presentación oral). Septiembre 2020.** Título del trabajo: Efecto de la delección de proteínas específicas del linaje hematopoyético sobre la reprogramación de la diferenciación en células eritroleucémicas.

Link: <https://www.facebook.com/pol.una.py/posts/10161730053594988>.

5. **Conferencia virtual EMBL: From Functional Genomics to Systems Biology. Noviembre 2020.** Título del trabajo: RNA-Seq of FV-induced erythroleukemia cells suggests a role of Wasp in apoptosis.

Link: <https://functional-embl.ipostersessions.com/default.aspx?s=41-C2-DA-2F-A9-BD-D0-BB-36-74-DC-C0-0F-BA-CD-E7>