



**Universidad Nacional Autónoma de México**  
**Maestría y Doctorado en Ciencias Bioquímicas**

*“Papel de SIRT1 en el control de los programas transcripcionales que dirigen la diferenciación a nivel de célula única.”*

**Tesis**

**QUE PARA OPTAR POR EL GRADO DE:**

Maestro en ciencias

**PRESENTA:**

Alejandro Alarcón del Carmen

**TUTORA PRINCIPAL:**

Dra. Lorena Aguilar Arnal  
[Instituto de Investigaciones Biomédicas, UNAM](#)

**MIEMBROS DEL COMITÉ TUTOR:**

Dr. Víctor Julián Valdés.

[Instituto de Fisiología Celular, UNAM](#)

Dr. Osbaldo Resendis Antonio.

[Instituto Nacional de Medicina Genómica](#)

CD. MX. 18 de Junio de 2023



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## Agradecimientos

Este proyecto de maestría se realizó bajo la tutoría de la Dra. Lorena Aguilar Arnal en el Departamento de Biología Celular del Instituto de Investigaciones Biomédicas de la UNAM, se agradece también el apoyo brindado por el Dr. Victor Julián Valdés (IFC) y el Dr. Osbaldo Resendis Antonio (INMEGEN) quienes participaron como miembros del comité tutor. Agradezco al Programa de Posgrado de Maestría en Ciencias Bioquímicas de la UNAM, a su personal administrativo, así como a la plantilla de profesores que participaron en mis cursos de formación durante mi maestría. De igual forma agradezco el apoyo económico brindado por CONAHCyT a través de su programa de Becas Nacionales para Estudios de Posgrado 2021.

También agradezco al laboratorio de Dr. Victor Julián Valdés por facilitarnos sus instalaciones para llevar a cabo el protocolo de single cell RNA seq, por brindarnos todas las comodidades y en especial a la técnica de su laboratorio, la Dra. Nallely Cano Domínguez por acompañarnos durante la realización de este experimento en cada paso del protocolo. De igual forma se agradece a Roberto Carvante de Analitek *life* por su apoyo técnico durante la realización del protocolo de single cell RNA seq.

Agradezco también a la Unidad de Cómputo del Instituto de Investigaciones Biomédicas de la UNAM por el apoyo brindado con el servidor, a la M. en C. Marcia Bustamante Zepeda por su apoyo técnico y por compartir su conocimiento durante mi estancia en el laboratorio. También a todos mis compañeros de laboratorio, a Roberto López Valiente, Roman Sitten, Dra. Carolina Cid, Dra. Lucia Mendoza, especialmente a Mirna González, Luis Ricardo Hernández e Iraiz Nuñez por ser, además de increíbles científicos, excelentes amigos y compañeros, de forma particular agradezco a Edgar Sánchez quien estuvo a mi lado como mentor en cada paso de mi formación experimental, agradezco de forma individual a la Dra. Lorena por su paciencia inmensa y la atención a mi desarrollo durante esta maestría. Este trabajo fue realizado con el apoyo y financiamiento del proyecto no. FORDECYT – PRONACES/15758/2020 de CONAHCyT, el Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT-DGAPA de la UNAM, número IN208022), y el Proyecto “CHRomatin dynamics and nuclear METabolism: an intimate interplay uncovered by non-linear optics”, con número RGY0078/2017 de Human Frontiers Science Program. Se extiende un agradecimiento al Programa Institucional Brain-Body-Behavior Interaction, por las discusiones sobre el tema y el apoyo prestado.

# Índice de contenidos

Índice de figuras.....	6
Índice de abreviaturas.....	8
Resumen.....	10
Abstract.....	13
1.- Introducción: .....	15
<b>1.1. Células pluripotenciales y células troncales embrionarias (ESC´s).....</b>	<b>15</b>
<b>1.2. Metabolismo de ESCs.....</b>	<b>16</b>
<b>1.3. NAD<sup>+</sup> y regulación REDOX.....</b>	<b>18</b>
<b>1.4. Enzimas consumidoras de NAD<sup>+</sup> .....</b>	<b>18</b>
<b>1.5. SIRT1 y las ESCs.....</b>	<b>20</b>
<b>1.6. SIRT1 y programas de diferenciación.....</b>	<b>22</b>
<b>1.7. scRNA seq y la heterogeneidad celular.....</b>	<b>24</b>
1.7.1. Aislamiento de células.....	25
1.7.2. Corrección del ruido técnico.....	26
1.7.3. Alta dimensionalidad.....	27
1.7.4. Visualización de la estructura y sus transiciones en datos de alta dimensionalidad (PHATE).....	28
<b>1.8. Obtención de genes diferencialmente expresados (DEG).....</b>	<b>32</b>
<b>1.9. Inferencia de actividad de Factores de transcripción.....</b>	<b>34</b>
<b>1.10. Análisis de trayectoria y pseudotiempo .....</b>	<b>36</b>
2.-Antecedentes específicos.....	42
3. Justificación.....	46
<b>3.1. Hipótesis.....</b>	<b>47</b>
<b>3.2. Objetivo principal.....</b>	<b>47</b>
<b>3.3. Objetivos particulares.....</b>	<b>47</b>
4. Metodología.....	48
<b>4.1. Cultivo celular de células H9, H9 SIRT1 -/- y H9 SIRT1ΔEx4 .....</b>	<b>48</b>
<b>4.2. Protocolo de diferenciación hacia adipocitos.....</b>	<b>48</b>
<b>4.3. Tinción Rojo oleoso (ORO).....</b>	<b>49</b>
<b>4.3. Protocolo de diferenciación hacia NPC.....</b>	<b>49</b>
<b>4.3. Preparación de solución single cell de células H9 y H9 SIRT1-/-.....</b>	<b>51</b>

4.4. Protocolo single cell 10X Chromium .....	51
4.5 Creación de librerías y secuenciación.....	52
4.6. Análisis de datos scRNAseq .....	54
4.7. Importación en R y métodos de análisis: singleCellExperiment y Seurat. ....	55
4.8. singleCellExperiment .....	55
4.9. Seurat.....	58
4.10 PHATE.....	61
4.11.- Pseudotiempo de difusión.....	62
4.12.- Trayectoria Slingshot.....	62
4.13.- <i>tradeSeq</i> y genes oscilantes a lo largo de la trayectoria.....	63
5. Resultados.....	65
5.1. Diferenciación de línea celular H9 y H9 <i>SIRT1</i> <sup>ΔEx4</sup> hacia linaje adipogénico.....	65
5.2. Diferenciación de línea celular H9 y H9 <i>SIRT1</i> <sup>-/-</sup> hacia NPC .....	66
5.3. scRNAseq 10X Chromium .....	68
5.4. Obtención de datos crudos scRNAseq.....	69
5.5.Reportes de calidad Cellranger .....	69
5.6 Análisis utilizando <i>SingleCellExperiment</i> .....	70
5.6.1. Análisis de calidad .....	70
5.6.2. Modelado de varianza. ....	71
5.6.3 Representación tSNE y Anotación de identidad .....	72
5.6.4. Obtención de genes diferencialmente expresados (DEG): Pseudobulk .....	78
5.7. Seurat.....	80
5.7.1. Control de calidad .....	80
5.7.2. Clustering y anotación de identidad. ....	82
5.7.3. Clustering y anotación de identidad día 6 .....	86
5.7.4 Integración de datos: H9 vs H9 <i>SIRT1</i> <sup>-/-</sup> día 0 .....	88
5.7.5 Identificación de Genes Diferencialmente expresados (DEG) .....	89
5.7.6 Integración de datos: H9 vs H9 <i>SIRT1</i> <sup>-/-</sup> día 6.....	93
5.8.1 Inferencia de actividad de Factores de Transcripción (TF): Día 0.....	95
5.8.2 Inferencia de actividad de Factores de Transcripción (TF): Día 6.....	96
5.9 PHATE y pseudotiempo de difusión .....	97
5.10 <i>Slingshot</i> y trayectoria celular .....	101
5.11. Genes asociados al pseudotiempo .....	102
Discusión.....	114

Conclusiones .....	126
Referencias:.....	127
Figuras suplementarias. ....	137

## Índice de figuras

<b>Figura 1:</b> Autorrenovación de ESC's.....	15
<b>Figura 2:</b> Tipos celulares y embriogénesis.....	16
<b>Figura 3:</b> Pluripotencialidad y su relación con el metabolismo. ....	17
<b>Figura 4:</b> Desacetilación catalizada por sirtuinas.....	20
<b>Figura 5:</b> OTX2 y la pluripotencia "naive" .....	22
<b>Figura 6:</b> Comparación de métodos de visualización y reducción de dimensiones en set artificial.....	31
<b>Figura 7:</b> Pseudotiempo y trayectorias. . ....	37
<b>Figura 8:</b> Algoritmos de inferencia de trayectorias (IT) y determinación de GAT.....	40
<b>Figura 9:</b> "Generalized additive model" o GAM.....	41
<b>Figura 10:</b> Validación de la delección de SIRT1 en la línea celular de hESCs H9.....	42
<b>Figura 11:</b> Niveles de abundancia de mRNA de marcadores de pluripotencia en células H9 y las mutantes isogénicas en el gen SIRT1.....	43
<b>Figura 12:</b> Niveles de expresión de mRNA de marcadores mesodérmicos y endodérmicos en las mutantes isogénicas en SIRT1 respecto a la condición control.....	44
<b>Figura 13:</b> Niveles de expresión de mRNA de marcadores ectodérmicos en las mutantes isogénicas con delección de SIRT1 respecto a la condición control. ....	44
<b>Figura 14:</b> Esquema temporal de protocolo de diferenciación de células troncales pluripotentes a adipocitos... 48	
<b>Figura 15:</b> Esquema temporal de protocolo de diferenciación de células troncales pluripotentes aun linaje dopaminérgico.....	50
<b>Figura 16:</b> Esquema del protocolo Chromium Next GEM Single Cell 3' Reagent Kits; primera parte. ....	52
<b>Figura 17:</b> Esquema del protocolo Chromium Next GEM Single Cell 3' Reagent Kits; segunda parte.....	53
<b>Figura 18:</b> Fotografías representativas (20X) de las células inducidas con medio de diferenciación adipogénico en el día 5 y 10 del protocolo de diferenciación a adipocitos en las células H9 WT y la Clona 8 SIRT1 <sup>-/-</sup> .....	66
<b>Figura 19:</b> Fotografías representativas (10X) de la morfología celular observada en los días 1,3,4,5,12 y 14 del protocolo de diferenciación a neuronas dopaminérgicas en las células H9 control y las H9 SIRT1 <sup>-/-</sup> .....	67
<b>Figura 20:</b> Electroferogramas de cDNA obtenido de cada muestra tras la RT de los mRNA capturados, generados con Qsep1 Lite .....	68
<b>Figura 21:</b> Electroferogramas de las librerías obtenido de cada muestra. ....	69
<b>Figura 22:</b> Gráfica de control de calidad Single cell experiment.....	71
<b>Figura 23:</b> Modelado de varianza.....	72
<b>Figura 24:</b> Representación tSNE de células H9 día 0.....	73
<b>Figura 25:</b> Representación tSNE de células H9 SIRT1 <sup>-/-</sup> día 0.....	74
<b>Figura 26:</b> Representación tSNE de células H9 día 6.....	75
<b>Figura 27:</b> Porcentaje de la población asignado a cada identidad del set de referencia de (La Manno et al., 2016) en la condición H9 día 6. ....	75
<b>Figura 28:</b> Representación tSNE de células H9 SIRT1 <sup>-/-</sup> día 6.....	76
<b>Figura 29:</b> Porcentaje de la población asignado a cada identidad del set de referencia de (La Manno et al., 2016) en la condición H9 día 6. ....	77
<b>Figura 30:</b> Barplot de DEG encontrados mediante pseudobulking en día 0.....	79
<b>Figura 31:</b> Gráfico de violín de parámetros de calidad en muestras día 0 .....	80
<b>Figura 32:</b> Gráfico de violín de parámetros de calidad en muestras día 6 .....	81
<b>Figura 33:</b> UMAPs H9 día 0 en ambos criterios de calidad. ....	84

<b>Figura 34:</b> UMAPs H9 SIRT1 <sup>-/-</sup> día 0 en ambos criterios de calidad..	85
<b>Figura 35:</b> UMAP H9 WT día 6..	87
<b>Figura 36:</b> UMAP H9 SIRT1 <sup>-/-</sup> día 6..	87
<b>Figura 37:</b> UMAPs de la integración de los datos de H9 y H9 SIRT1 <sup>-/-</sup> día 0 con ambos criterios de calidad.....	89
<b>Figura 38:</b> Barplots de enriquecimiento DEG día 0 criterios laxos.....	91
<b>Figura 39:</b> Barplots de enriquecimiento DEG día 0 criterios estrictos..	92
<b>Figura 40:</b> UMAP de la integración de los datos de H9 y H9 SIRT1 <sup>-/-</sup> día 6. S.....	93
<b>Figura 41:</b> Barplots de enriquecimiento DEG día 6..	95
<b>Figura 42:</b> Heatmap de actividad de TFs inferida de los datos Dia 0..	96
<b>Figura 43:</b> Heatmap de actividad de TFs inferida de los datos Dia 6..	97
<b>Figura 44:</b> Representación PHATE y pseudotiempo de difusión en eje de diferenciación D0 a D6 WT..	98
<b>Figura 45:</b> Representación PHATE y pseudotiempo de difusión en eje de diferenciación D0 a D6 SIRT1 <sup>-/-</sup> ..	100
<b>Figura 46:</b> Trayectorias trazadas por Slingshot en ejes de diferenciación D0 a D6, condiciones H9 y H9 SIRT1 <sup>-/-</sup> .....	102
<b>Figura 47:</b> Análisis de enriquecimiento con genes diferenciales al final del pseudotiempo..	104
<b>Figura 48:</b> Genes involucrados en los procesos de enriquecimiento “Organización de la estructura extracelular”, “Organización de la matriz extracelular” y “Regulación positiva de la locomoción “..	107
<b>Figura 49:</b> Genes involucrados en los procesos de enriquecimiento “Respuesta al daño del DNA regulada por la transducción de señales de p53“..	108
<b>Figura 50:</b> Genes involucrados en los procesos de enriquecimiento “Ciclo Celular, Mitosis”, “Fase G1 y transición G1/S” y “Transición G1/S”.....	110
<b>Figura 51:</b> Genes involucrados en las identidades celulares “Células troncales pluripotentes” y “Células del epiblasto” ..	112
<b>Figura 52:</b> Genes involucrados en las identidades celulares “Células troncales Neuronales/Precursores neuronales”, “Tanicitos”, y “Glía Radial” ..	113

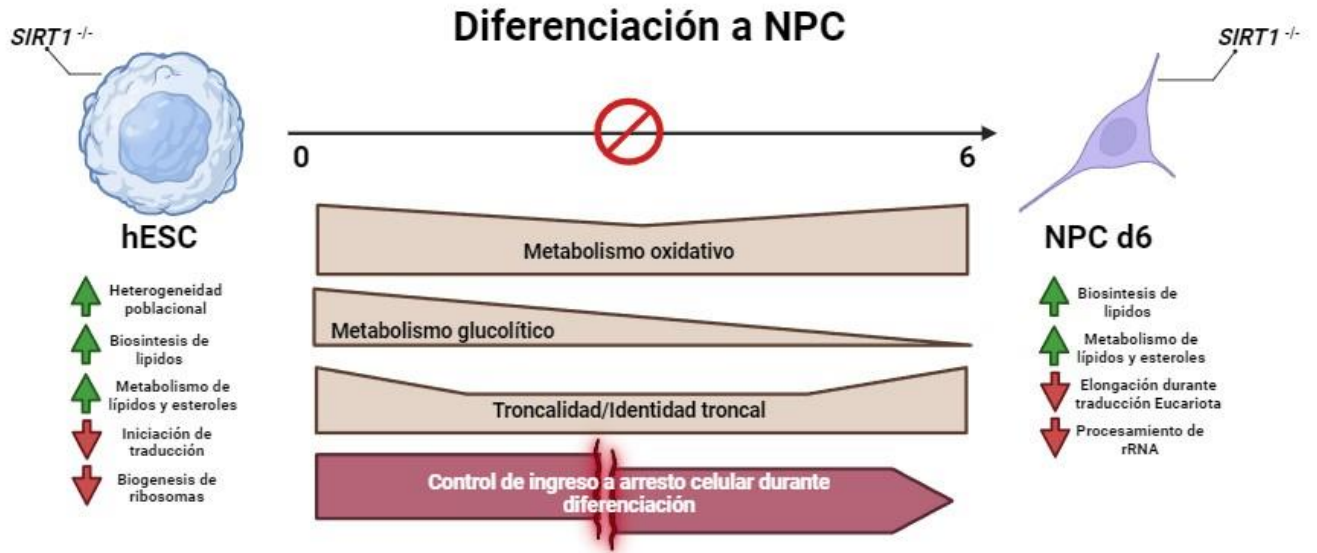


## Índice de abreviaturas

ARTs	ADP-ribosil transferasas
BEAM	Modelado de expresión basado en ramificaciones
bp	Pares de bases
BP	Proceso biológico
DEG	Genes diferencialmente expresados
ERCC	External RNA Controls Consortium
ESCs	Células troncales embrionarias
FACS	Sorting celular activado por fluorescencia
GAM	Modelos generales aditivos
GAT	Genes asociados a la trayectoria
GLM	Modelos lineales generalizados
HDACs	Deacetilasas de histonas
hEndo	Células endoteliales
hESCs	Células troncales embrionarias humanas
hMgl	Microglía
hNbML	Neuroblastos mediolaterales
hNProg	Progenitores neurales
hOMNT	Núcleo oculomotor
hPeric	Pericitos
hProgBP	Progenitores plato basal
hProgFPL	Progenitores piso lateral
hProgFPM	Progenitores piso medio basal
hProgM	Progenitores línea media
hRgl1	Glia-like radial 1
hRgl2	Glia-like radial 2
iPS	Células pluripotentes inducidas
KSR	Knockout Serum Replacement
mESCs	Células troncales embrionarias de ratón
MM	Modelo mixto
mMSCs	Células mesenquimales troncales de ratón
MSC	Células troncales mesenquimales

mst	Mínimum spaning tree
NAD	Nicotinamida adenina dinucleótido
NE	Neuroepitelio
NPC	Progenitores Neurales
NU	Neuronas
ORO	Tinción rojo oleoso
PAR	Poli-ADP-ribosa
PARPs	Poli-ADP-ribosa polimerasas
PCA	Análisis de componentes principales
PCR	Reacción en cadena de la polimerasa
PHATE	Potencial de difusión de calor para embeddings de transición basadas en afinidad
ROBO	Rondabout
ROS	Especie reactiva de oxígeno
RT	Retrotranscripción
scRNAseq	RNA seq de célula única
SIRT1	Sirtuína 1
SIRT6	Sirtuína 6
SIRT7	Sirtuína 7
TF	Factores de transcripción
TSNE	Incrustación de vecinos estocásticos distribuidos en t
UMAP	Proyección y aproximación de variedad uniforme para reducción de dimensiones
UMIS	Identificadores moleculares únicos

# Resumen gráfico



## Resumen

Las células troncales humanas hESCs poseen características únicas como son la pluripotencialidad y la autorrenovación, la manutención de estas capacidades dependen de factores de transcripción como OCT4, SOX2 y NANOG, también del mantenimiento de un nicho hipóxico acompañado de un metabolismo caracterizado por un elevado consumo de glucosa y producción de lactato a través de la glucólisis aeróbica. Estas células tienen la capacidad de dar origen a todos los linajes embrionarios a través de procesos de diferenciación que conlleva a una reprogramación metabólica donde el flujo energético deja de depender de la glucólisis anaerobia y comienza a generarse ATP a través de la fosforilación oxidativa principalmente, este metabolismo oxidativo genera cofactores como el NAD<sup>+</sup> que, además de ser un cofactor esencial en las reacciones metabólicas de óxido-reducción, sirve como sustrato de enzimas dependientes de NAD<sup>+</sup> como son las PARPs o las sirtuínas. Recientemente ha comenzado a reportarse una relación entre los cambios en el metabolismo de óxido-reducción y los procesos de diferenciación en células troncales, donde enzimas como las sirtuínas juegan papeles esenciales no solo en la reprogramación metabólica si no también en la regulación de la pluripotencia y homeostasis de las hESCs. La sirtuína 1 (SIRT1) es una enzima con función de desacetilasa de histonas que ha sido vinculada con procesos de diferenciación celular y su regulación a través de la transición metabólica, sin embargo, aún se desconoce cómo regula este proceso a nivel transcriptómico y de forma dinámica. Para conocer el papel de SIRT1 durante un proceso de diferenciación se utilizó una línea celular de hESCs con una delección total de SIRT1 (H9 *SIRT1*<sup>-/-</sup>) y un modelo de knockdown (H9 *SIRT1*<sup>ΔEx4</sup>) que fueron sometidos a protocolos de diferenciación para obtener neuronas dopaminérgicas y adipocitos (respectivamente), para posteriormente utilizar una técnica de scRNAseq de 10X Genomics para obtener el transcriptoma de estas células. Los resultados de este trabajo demostraron que SIRT1 es esencial para la correcta diferenciación celular de las hESCs, en la condición H9 *SIRT1*<sup>ΔEx4</sup> se obtuvo una menor cantidad de adipocitos maduros respecto a la condición H9 WT, esto corroborado a través de una tinción de rojo oleoso, en contraste, la condición H9 *SIRT1*<sup>-/-</sup>, mostró que SIRT1 es esencial no solo en la correcta diferenciación celular si no que es necesaria para la supervivencia de estas, las células H9 *SIRT1*<sup>-/-</sup> no respondían al estímulo de diferenciación a neuronas dopaminérgicas y finalmente morían al alcanzar el día 12 del protocolo de diferenciación, es por esto que se eligieron las células de esta condición para realizar el scRNAseq en los días 0 (estado basal) y 6 (progenitor neural o NPC). Al analizar los datos de la condición H9 *SIRT1*<sup>-/-</sup> en el día 0 se encontró que la delección de SIRT1 producía heterogeneidad interpoblacional debido a la aparición de identidades relacionadas con mesodermo (*Mesenchymal Stem Cells* o *MSC*) y a que un porcentaje elevado de la población se asociaba con la identidad de *iPScells* y no con *hESCs* como ocurría en la condición WT, esto indica que SIRT1 tiene un papel importante en la conservación del

estado pluripotente, así como las características de troncalidad en hESCs. Posteriormente a través de análisis de enriquecimiento de los genes diferencialmente expresados entre la condición H9 WT y H9 *SIRT1*<sup>-/-</sup> se encontró que los procesos más afectados eran el metabolismo y la traducción, ya que la condición H9 *SIRT1*<sup>-/-</sup> mostraba activación de metabolismo de lípidos mediada por *SREBP* y los genes regulados a la baja en esta misma condición mostraba un decrecimiento en la actividad traduccional y procesamiento de RNA ribosomal, por lo que SIRT1 está involucrada en el control del switch metabólico en las células hESCs así como en el control del nivel de actividad traduccional celular. Finalmente se realizó el análisis de los genes que cambiaban a lo largo del proceso de diferenciación, usando técnicas de pseudotiempo y análisis de trayectoria, y se encontraron 94 genes que tenían comportamientos diferencialmente significativos en la etapa final del protocolo (cercano al día 6), estos genes estaban relacionados con el proceso de remodelación extracelular y correcto establecimiento del linaje neuroectodermico, con la regulación del ciclo celular y la entrada al arresto celular, que es esencial durante las etapas tempranas de diferenciación en hESCs y como última instancia, relacionados con la conservación de marcadores de pluripotencia, la condición H9 *SIRT1*<sup>-/-</sup> mostró un nivel de expresión bajo de los genes encargados de estos tres procesos lo que sugiere que SIRT1 regula el eje de diferenciación de hESCs a NPC a través de un correcto establecimiento de la matriz extracelular y la interacción con ella, de la mano es importante para permitir que las células H9 entren en arresto celular y tengan control sobre las fases G1 y G2/M para diferenciarse correctamente. El caso de los marcadores de pluripotencia, donde en la condición H9 *SIRT1*<sup>-/-</sup> no se observaba el decrecimiento en la expresión en contraste a la condición H9 WT, puede ser indicador de que la SIRT1 es esencial para llevar a cabo el proceso de “escape de la troncalidad” necesario para llevar a cabo el compromiso de linaje y posterior diferenciación.

## Abstract

Human embryonic stem cells (hESCs) possess unique characteristics such as pluripotency and self-renewal, and the maintenance of these abilities depends on transcription factors such as OCT4, SOX2, and NANOG, as well as the maintenance of a hypoxic niche accompanied by a metabolism characterized by high glucose consumption and lactate production through aerobic glycolysis. These cells have the capacity to give rise to all embryonic lineages through differentiation processes that involve metabolic reprogramming, where the energy flow shifts from anaerobic glycolysis to primarily generating ATP through oxidative phosphorylation. This oxidative metabolism generates cofactors such as NAD<sup>+</sup>, which, in addition to being an essential cofactor in redox metabolic reactions, serves as a substrate for NAD<sup>+</sup>-dependent enzymes such as PARPs or sirtuins.

Recently, a relationship has been reported between changes in redox metabolism and stem cell differentiation processes, where enzymes such as sirtuins play essential roles not only in metabolic reprogramming but also in the regulation of pluripotency and homeostasis of hESCs. Sirtuin 1 (SIRT1) is a histone deacetylase enzyme that has been linked to cellular differentiation processes and its regulation through metabolic transition. However, it is still unknown how SIRT1 dynamically regulates this process at the transcriptomic level.

To investigate the role of SIRT1 during differentiation processes, an hESC cell line with a total deletion of SIRT1 (H9 SIRT1<sup>-/-</sup>) and a knockdown model (H9 SIRT1 $\Delta$ Ex4) were used to obtain dopaminergic neurons and adipocytes (respectively) through differentiation protocols, followed by 10X Genomics scRNAseq technique to obtain the transcriptome of these cells. The results of this study demonstrated that SIRT1 is essential for proper cellular differentiation of hESCs. In the H9 SIRT1 $\Delta$ Ex4 condition, a lower number of mature adipocytes were obtained compared to the H9 WT condition, as confirmed through an oil red staining. In contrast, the H9 SIRT1<sup>-/-</sup> condition showed that SIRT1 is not only essential for proper cellular differentiation but is also necessary for the survival of these cells. The H9 SIRT1<sup>-/-</sup> cells did not respond to the dopaminergic neuronal differentiation stimulus and eventually died on day 12 of the differentiation protocol. Therefore, cells from this condition were chosen to perform the scRNAseq on days 0 (basal state) and 6 (neural progenitor or NPC).

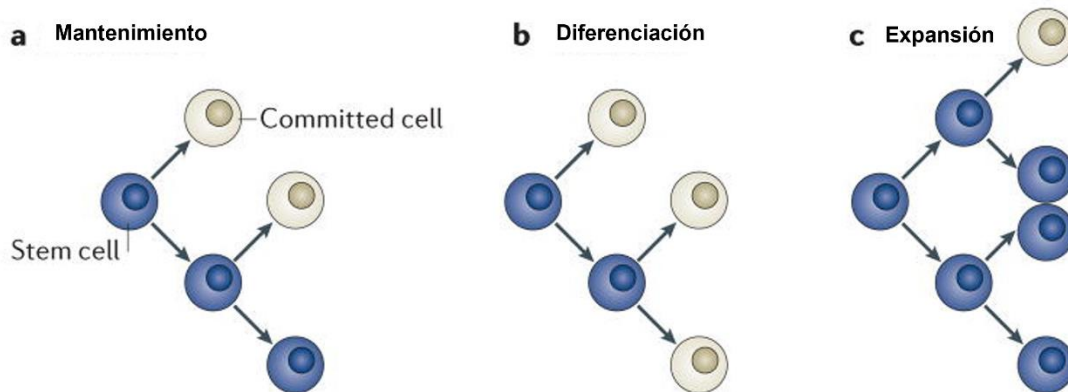
Upon analyzing the data from the H9 SIRT1<sup>-/-</sup> condition on day 0, it was found that SIRT1 deletion produced interpopulation heterogeneity due to the appearance of identities related to mesoderm (Mesenchymal Stem Cells or MSCs) and a high percentage of the population being associated with iPSCs identity rather than hESCs as observed in the WT condition. This indicates that SIRT1 plays an important role in the preservation of pluripotency state and stemness characteristics in hESCs.

Furthermore, the enrichment analysis of differentially expressed genes between the H9 WT and H9 SIRT1<sup>-/-</sup> conditions revealed that the most affected processes were metabolism and translation. The H9 SIRT1<sup>-/-</sup> condition showed activation of lipid metabolism mediated by SREBP, while the downregulated genes in this condition showed decreased translational activity and ribosomal RNA processing, indicating that SIRT1 is involved in the control of the metabolic switch in hESCs and cellular translational activity. Finally, analysis of genes that changed during the differentiation process using pseudotime and trajectory analysis identified 94 genes with differentially significant behavior in the final stage of the protocol (around day 6), which were related to extracellular remodeling, proper establishment of neuroectodermal lineage, cell cycle regulation, and entry into cell arrest, which is essential during early differentiation stages in hESCs. The H9 SIRT1<sup>-/-</sup> condition showed low expression levels of genes responsible for these three processes, suggesting that SIRT1 regulates the hESC-to-NPC differentiation axis through proper establishment and interaction with the extracellular matrix. It is also important for allowing H9 cells to enter cell arrest and have control over the G1 and G2/M phases to differentiate properly. The case of pluripotency markers, where the H9 SIRT1<sup>-/-</sup> condition did not show a decrease in expression in contrast to the H9 WT condition, may indicate that SIRT1 is essential for carrying out the process of "exit from stemness" necessary for lineage commitment and subsequent differentiation.

## 1.- Introducción:

### 1.1. Células pluripotenciales y células troncales embrionarias (ESC's)

Las células troncales embrionarias (ESC's) son esenciales para el desarrollo de los tejidos durante la embriogénesis, así como en el mantenimiento de la homeostasis y regeneración de estos tejidos en condiciones como lesiones o heridas durante las etapas posnatales. Estas células poseen características específicas como la capacidad de dar origen a distintos linajes celulares (totipotencialidad, pluripotencialidad o multipotencialidad) mediante procesos de diferenciación. Otra característica esencial de estas células es la capacidad de autorrenovación (Ito & Suda, 2014) ya que tras una división celular pueden dar origen a nuevas células troncales y a la vez generar células dirigidas a un patrón de diferenciación determinado (Figura 1).



**Figura 1:** Autorrenovación de ESC's. (a) División asimétrica de mantenimiento, se da origen a una ESC y una célula comprometida a un nuevo linaje celular; (b) División asimétrica de diferenciación, se da origen a nuevas células comprometidas y se agota la población ESC's; (c) División de expansión, ocurren una serie de divisiones simétricas que expanden el número de ESC's para posteriormente continuar generando células comprometidas a diferenciación. Modificado de: Ito K. y Suda T. *Nat Rev Mol Cell Biol*, 2014

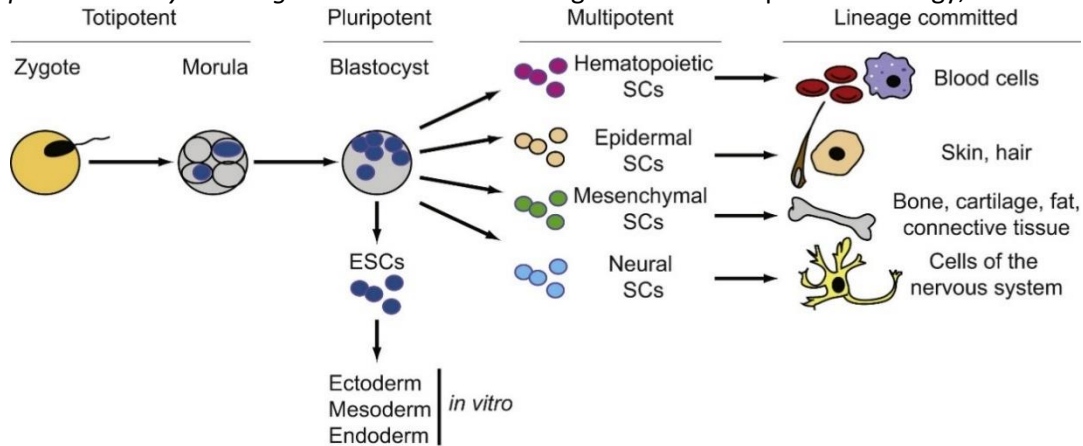
Durante la embriogénesis, la unión de los gametos da origen a un cigoto que posteriormente se convierte en mórula tras procesos de proliferación y división celular, en esta etapa las células son totipotenciales debido a que pueden dar origen a todos los tejidos embrionarios y extraembrionarios. Las células totipotenciales continúan diferenciándose hasta llegar a una etapa conocida como blastocisto donde dan origen a un grupo de células más diferenciadas conocidas como "células pluripotenciales" que darán origen a las tres capas germinales del embrión (endodermo, mesodermo y ectodermo) que se diferenciarán aún más hasta formar todos los tejidos del embrión (Weger et al., 2017).

Las ESC's se derivan de la masa celular interna del blastocisto y poseen todas las características



pluripotenciales así como la capacidad de autorrenovación, pueden ser cultivadas *in vitro* e incluso pueden ser inducidas a diferenciarse en los distintos tipos de tejido (Figura 2).

**Figura 2:** Tipos celulares y embriogénesis. Tomado de: Weger et al. *Developmental Biology*, 2017.



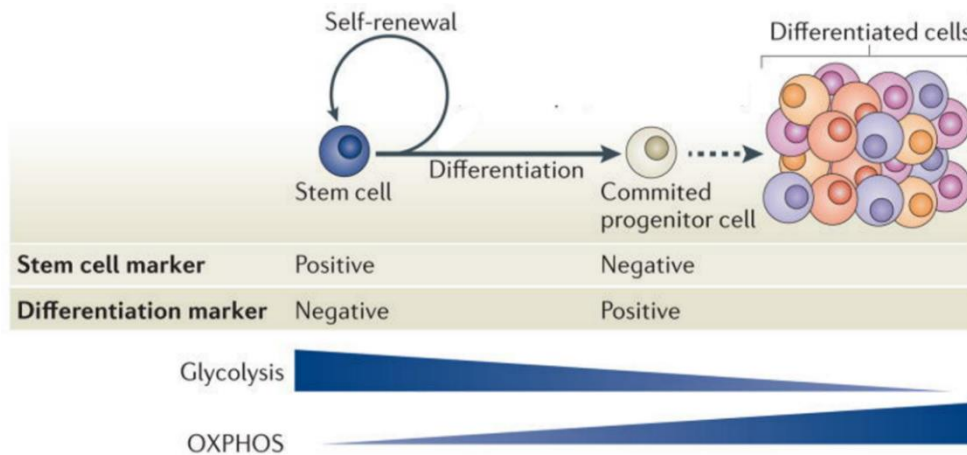
Las ESC´s poseen también algunos factores de transcripción específicos que permiten identificarlas y aislarlas, algunos de estos son los factores de transcripción Oct4 y Nanog que están implicados en la formación y supervivencia de este linaje celular, por otro lado, el factor de transcripción Sox2 es reconocido como esencial para mantener la pluripotencialidad en estas células (Rodrigues et al., 2020).

## 1.2. Metabolismo de ESCs.

Durante la formación de la masa interna del blastocisto y el establecimiento de las ESCs las células se compactan y migran lo que desencadena una nueva fase proliferativa y requerimientos metabólicos específicos, estas condiciones generan que las ESCs aumenten el flujo de glucosa hacia su interior e incrementa la expresión de transportadores de glucosa como GLUT1 y GLUT3, también incrementa la actividad de enzimas de la vía como la hexocinasa y fosfofructocinasa 1 (Shyh-Chang et al., 2013). Por otro lado se ha reportado que en la ESC´s las mitocondrias presentan un bajo potencial en su membrana lo que es indicativo de una baja actividad mitocondrial (Van Blerkom, 2009) lo que indica que las ESC´s dependen fuertemente de la glucólisis en condiciones anaeróbicas para obtener su energía.

La obtención de ATP a través de la fosforilación oxidativa y la cadena respiratoria es característica de células mucho más diferenciadas (Figura 3) debido a que se ha observado que el aumento en la actividad mitocondrial se relaciona directamente con el nivel de diferenciación celular, por otro lado un

metabolismo basado en glucólisis aporta cofactores y subproductos como el  $\text{NAD}^+$  que son cruciales en el desarrollo, proliferación y mantenimiento de la pluripotencialidad en las ESC (Ito & Suda, 2014).



**Figura 3:** Pluripotencialidad y su relación con el metabolismo. Modificado de: Ito K. y Suda T. *Nat Rev Mol Cell Biol*, 2014

La dependencia de la glucólisis anaerobia puede parecer desventajosa ya que solo se obtienen 4 ATP por molécula de glucosa que ingresa a la vía, en comparación la fosforilación oxidativa es capaz generar 36 ATP por molécula de glucosa ingresada. Sin embargo, una cantidad semejante de ATP puede ser obtenida si se logra mantener un flujo constante de glucosa en la vía de la glucólisis. El flujo de glucosa constante se obtiene debido a la alta concentración de la coenzima nicotinamida adenina dinucleótido ( $\text{NAD}^+$ ) que es esencial en la reacción de conversión de gliceraldehido 3-fosfato a 1-3 bífosfoglicerato durante glucólisis, esta alta concentración de  $\text{NAD}^+$  es regenerada a través de la acción de la enzima lactato deshidrogenasa cuando se genera lactato en condiciones anaerobias (Heo et al., 2017).

Las funciones principales del  $\text{NAD}^+$  y el  $\text{NADH}$  se puede clasificar en las siguientes categorías: 1) Reacciones de óxido reducción donde  $\text{NAD}^+$  y  $\text{NADH}$  actúan como coenzimas; 2) Reacciones catalizadas por la  $\text{NADH}$  oxidasa que oxida  $\text{NADH}$  a  $\text{NAD}^+$  generando especies reactivas oxígeno (ROS) como subproducto; 3) El catabolismo de  $\text{NAD}^+$  para generar otras moléculas que contengan nicotinamida a través de enzimas consumidoras de  $\text{NAD}^+$ ; 4) La conversión de  $\text{NAD}^+$  en  $\text{NADP}$  a través de cinasas de  $\text{NAD}^+$  (Ying, 2006).

### 1.3. NAD<sup>+</sup> y regulación REDOX

La producción de NADP así como la conversión de NAD<sup>+</sup> a NADH y viceversa a través de reacciones redox son componente principales del metabolismo, por un lado otorgan la energía química necesaria para el anabolismo de moléculas a través de la vía de las pentosas fosfato que dependen de la presencia de NADPH o durante el catabolismo de moléculas para la obtención de energía (como el flujo de glucosa previamente descrito) sin embargo la producción de ROS consecuencia de esta conversión tiene también un papel fundamental en la mantención de las capacidades de autorrenovación y troncalidad en las ESCs.

En las células troncales embrionarias los cambios en el estado de oxidación, también llamados regulación redox, son responsables de la comunicación entre el núcleo celular y las mitocondrias (Gomes et al., 2013). La comunicación núcleo-mitocondria mediada por la regulación redox es parte importante de la coordinación entre el metabolismo celular y la remodelación de la cromatina, por ende parte importante de la regulación de expresión génica, control del ciclo celular, reparación del DNA y la diferenciación celular.

La regulación del balance redox dentro de las ESCs es esencial para mantener la capacidad de autorrenovación (Ushio-Fukai & Rehman, 2014), las ESCs utilizan enzimas antioxidantes para eliminar el ROS intracelular, algunos de los factores involucrados en esto son la familia de factores de transcripción *forkhead homeobox type O (FOXO)* y el factor *NRF2* (Shi et al., 2012)

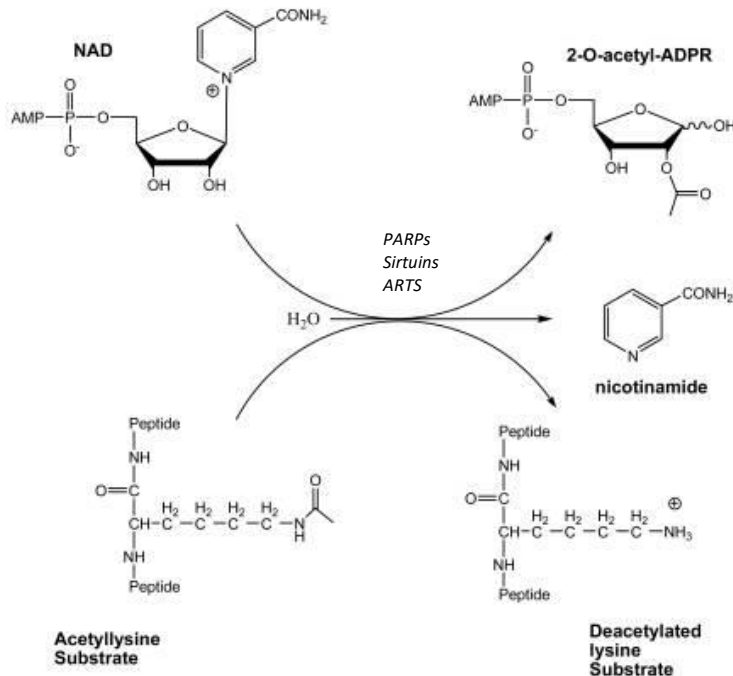
Otra proteína involucrada en el mantenimiento de la capacidad de autorrenovación es la proteína multifuncional *APE1/REF1* debido a su capacidad de controlar la respuesta a estrés oxidativo y la disminución de la acumulación de ROS mediante su unión a factores de transcripción como *HIF1-α*, *NERF2* y *p53* (Angkeow et al., 2002).

### 1.4. Enzimas consumidoras de NAD<sup>+</sup>

El NAD<sup>+</sup> tiene también un papel fundamental como sustrato de algunas enzimas que a partir de su catabolismo producen moléculas con nicotinamida o moléculas que contienen ADP-ribosa, estas reacciones pueden tener varias funciones biológicas importantes. El primer grupo de enzimas y uno de lo más importantes son las Poli-ADP-ribosa polimerasas (PARPs) que consumen NAD<sup>+</sup> para producir nicotinamida y polímeros de poli-ADP-ribosa (PAR), su función es crucial en la reparación del daño al DNA ya que cuando este ocurre por algún agente químico o físico la PARP se une al sitio del daño comenzando a polimerizar moléculas de PAR que genera mediante el catabolismo del NAD<sup>+</sup> disponible, esto genera el reclutamiento de otras proteínas involucradas en la reparación de los sitios de ruptura

en el DNA. En segundo lugar, tenemos a las enzimas mono ADP-ribosil transferasas (ARTs) (Lupi et al., 2000) que de igual forma catabolizan el NAD<sup>+</sup> para producir nicotinamida y monómeros de ADP-ribosa, estos últimos son transferidos a algunas proteínas modificando su función generando un estado permanentemente activo o por otro lado inhibiendo la función de la proteína blanco (Lupi et al., 2000). Las NAD<sup>+</sup> cinasas catalizan la conversión de NAD<sup>+</sup> en NADP, este último es esencial en vías anabólicas donde las ESCs utilizan la energía química producida por esta molécula y su estado reducido para sintetizar nucleótidos a través de la vía de las pentosas fosfato (Ying, 2006). Por último el NAD<sup>+</sup> funciona también como un sustrato de un grupo de enzimas remodeladoras de la cromatina que se encargan de regular los programas epigenéticos y que mantienen la proliferación y pluripotencialidad de las ESCs, estas enzimas son las desacetilasas de histonas (HDACs) dependientes de NAD<sup>+</sup> conocidas como sirtuinas. Estas enzimas llevan a cabo la desacetilación al transferir el grupo acetilo de algún residuo de lisina presente en una histona u otra proteína a una molécula de NAD<sup>+</sup> (Sauve, 2010), en el proceso se libera la nicotinamida y se produce una molécula de 2-O-acetil-ADPR (Figura 4).

Existen 7 sirtuinas en mamíferos, las cuales tienen distintos roles durante el desarrollo embrionario y la diferenciación de la ESCs a distintos linajes celulares maduros, al mismo tiempo desempeñan un rol crucial durante el mantenimiento de las características de troncalidad en las ESCs. En las células embrionarias troncales de ratón (mESCs) la sirtuina 1 (SIRT1) es la sirtuina más expresada, seguida por sirtuina 6 (SIRT6) y sirtuina 7 (SIRT7) (Efroni et al., 2008). El knock out de estas tres sirtuinas en células germinales causa defectos notorios en el desarrollo embrionario enmarcando la importancia que tienen en los procesos de diferenciación y regulación epigenética del desarrollo (Fang et al., 2019).



**Figura 4:** Desacetilación catalizada por sirtuinas. Se puede observar que mediante hidrólisis y el catabolismo de una molécula de NAD se remueve un grupo acetilo de un residuo de lisina en la proteína objetivo, generando una molécula de 2-O-acetil-ADPR, nicotinamida y la proteína objetivo desacetilada. Tomado de: Sauve, A. A. BBA, 2010

### 1.5. SIRT1 y las ESCs

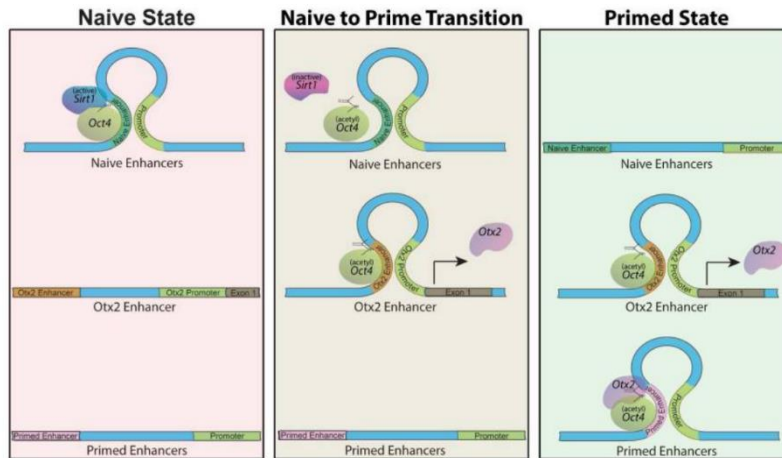
De las sirtuinas anteriormente mencionadas SIRT1 tiene un papel fundamental característico durante el desarrollo y diferenciación en las ESCs. Los niveles de expresión de SIRT1 son muy elevados en los embriones preimplantados de ratón, a partir de la etapa de dos células donde se empieza a expresar el genoma embrionario propio, indicando que la SIRT1 es uno de los primeros genes activados durante la embriogénesis (Tang et al., 2017).

Se ha reportado que la delección de SIRT1 en células troncales embrionarias humanas (hESCs) promueve su diferenciación a células pancreáticas beta capaces de producir insulina con expresión de genes específicos de estos islotes celulares como la proteína homeobox de páncreas/duodenos 1 (PDX1) y Nkx6 (Seo et al., 2019).

También se ha observado que SIRT1 tiene un papel importante en mantener el correcto establecimiento de las líneas celulares durante etapas posteriores del desarrollo en ESCs, un ejemplo de lo anterior se da a través de la regulación de la actividad de la ADN metiltransferasa tipo 3 (Dnmt3l) en una línea celular troncal embrionaria de ratón con delección de SIRT1 (mESCs SIRT1 <sup>-/-</sup>); mediante interacción directa la SIRT1 desacetila esta metiltransferasa generando inestabilidad en la misma y su posterior degradación. También se encontró que la SIRT1 se unía a la región del promotor y el cuerpo

del gen codificante de la *Dnmt3l*, lo anterior nos permite inferir que SIRT1 regula esta metiltransferasa a nivel transcripcional reprimiéndola y regulando la metilación de algunos sets de genes cruciales durante el desarrollo. En ausencia de SIRT1, algunos genes relacionados con el establecimiento de los linajes celulares germinales y genes relacionados con la impronta embrionaria se encontraban hipermetilados en sus regiones reguladoras, por ejemplo *Meg3* (Maternally expressed gene 3), *Dlk1* (delta like non-canonical Notch ligand 1), *Ngn* (Necdin), *Ddx3y* (DEAD-box helicase 3 Y-linked), *Pla2g1b* (phospholipase A2, group IB) y *Stra8* (Stimulated by retinoic acid 8) entre otros (Heo et al., 2017). Los cambios en la actividad de la DNMT3L ocasionados por esta falta de expresión de SIRT1 produjeron a nivel fenotípico un retraso en la neurogénesis y espermatogénesis en modelos murinos de células troncales embrionarias (mESCs) (Heo et al., 2017).

La SIRT1 también es un regulador crucial de la pluripotencialidad en las ESCs mediante la desacetilación de factores de transcripción que modulan de forma directa esta característica. En las ESCs humanas, SIRT1 es encargada de la regulación del factor de transcripción *Oct4* mediante la represión de p53, un sustrato de desacetilación bien descrito de SIRT1 (Williams et al., 2016). Se ha demostrado que la SIRT1 tiene una función importante en el mantenimiento del estado de pluripotencialidad *naive* y la transición a un estado de pluripotencialidad *primed* en mESCs. Los niveles bajos de SIRT1 producen la hiperacetilación del factor de transcripción OCT4, esto induce su traslocación e interacción con un enhancer distal promoviendo así la expresión de otro factor de transcripción llamado OTX2 (Figura 5), este último tiene una afinidad de unión alta a OCT4 en su estado hiperacetilado y promueve su unión a genes relacionados con la transición a pluripotencia *primed* (Figura 5) (Williams et al., 2016).



**Figura 5:** OTX2 y la pluripotencia "naive". Se observa que durante el estado "naive" de pluripotencia SIRT1 desacetila OCT4 y lo mantiene unido a genes relacionados con su mantenimiento, cuando SIRT1 es removida OCT4 se acetila y se desplaza hacia enhancers de genes que posibilitan la transición a un estado de pluripotencia "primed" como OTX2.

Existe interacción reportada de SIRT1 con algunos otros factores de transcripción relacionados con la embriogénesis, desarrollo y pluripotencialidad un ejemplo de esto es su cooperación con los factores de transcripción de la familia Myc (N-myc, L-myc y c-myc) donde en un modelo de mESCs la delección de *Sirt1* genera la hiperacetilación de los factores de transcripción N-Myc y c-Myc, lo que disminuye su estabilidad y reduce su reclutamiento al promotor de la enzima metionil adenosil transferasa 2A (MAT2A). Esto genera un decrecimiento en la actividad principal mediada por esta enzima que es la conversión de metionina a S-adenosilmetionina (SAM) que funge como la principal molécula donadora de grupos metilo en múltiples reacciones de metilación, esto de forma colateral genera una disminución considerable en los niveles de metilación en las histonas y compromete la conservación del programa transcripcional necesario para mantener la pluripotencialidad de las mESCS (Tang et al., 2017).

### 1.6. SIRT1 y programas de diferenciación.

Además de su participación como un modulador importante de la pluripotencialidad en ESCs y su rol durante es establecimiento de las líneas germinales embrionarias, la SIRT1 tiene como una de sus funciones principales la desacetilación y regulación de factores de transcripción maestros en programas de diferenciación.

La activación de SIRT1 ha demostrado favorecer la diferenciación osteogénica en células mesenquimales troncales de ratón (mMSCs) mediante la regulación positiva del factor de transcripción maestro de osteoblastos RUNX2 (Zainabadi et al., 2017). La delección de SIRT1 genera un efecto nulo

sobre la expresión este factor de transcripción; sin embargo, uno de sus genes objetivo río abajo en la cascada de diferenciación en osteoblastos, el factor de transcripción *Osx* (osterix), reduce su expresión considerablemente (Zainabadi et al., 2017). El factor de transcripción OSX (también conocido como SP7) juega un papel crucial en la maduración de los osteoblastos siendo regulador principal de genes marcadores específicos de esta clase celular como son *Bsp* (Bone sialoprotein), *fibromodulina*, *Osteocalcina*, *Dkk1* (Dickkopf WNT Signaling Pathway Inhibitor 1) y *Col1a1* (Collagen type I Alpha 1 chain) (Liu et al., 2020).

Estudios recientes han demostrado que la SIRT1 es capaz de bloquear el camino de diferenciación adipogénica en hMSCs mediante la estimulación de la vía de señalización Wnt/B-catenina lo que resulta en el compromiso de estas células hacia un linaje osteogénico (Hsu et al., 2018), SIRT1 desacetila B-catenina y promueve su acumulación en el núcleo celular lo que conlleva a la transcripción de genes relacionados con la diferenciación osteogénica como el factor de transcripción *Runx2* previamente mencionado (Feng et al., 2016).

Otra de las vías de regulación de diferenciación mediadas por SIRT1 está relacionada con el coactivador 1a de PPAR- $\gamma$ , *PCG-1 $\alpha$* . Este coactivador transcripcional es un regulador fundamental de las vías de biogénesis de mitocondrias en tejidos como los músculos y la grasa (adipocitos). La acetilación de *PCG-1 $\alpha$*  inhibe su actividad y está demostrado que SIRT1 es capaz de desacetilar a esta proteína de manera dependiente del balance celular redox entendido como el ratio de  $NAD^+/NADH$ , lo que define la capacidad de diferenciación hacia un linaje adipogénico (Zhao et al., 2014).

Sin embargo el papel de SIRT1 en la diferenciación adipogénica no se limita a la inhibición de la formación adipocitos. Se ha reportado que SIRT1 es un regulador importante de la vía de diferenciación adipogénica en sus estadios tempranos (preadipocitos) a través de modular la expresión de los receptores de ácido retinoico ( $RAR-\alpha$ ,  $RAR-\beta$  y  $RAR-\gamma$ ). La inhibición de SIRT1 en un modelo de mESCs utilizando el inhibidor EX527 aumentó la acetilación del factor de transcripción NCOR1 inhibiendo su capacidad de unirse a los promotores de los genes codificantes para el receptor  $RAR-\beta$ , quien forma parte crucial de la respuesta a ácido retinoico que promueve la diferenciación de mESCs a un linaje adipogénico (Bost et al., 2002). A su vez, estas células deficientes de SIRT1 mostraban una deficiencia en la expresión de marcadores de preadipocitos como *Pref-1*, *LPL* y *C/EBP- $\beta$*  en comparación con las mESCs wild type (Jung et al., 2020).



SIRT1 también tiene roles importantes en la proliferación y maduración de progenitores neurales (NPC) en adultos a través del eje de regulación *CREB-SIRT1-Hes1* encontrado en mESC. Se reportó que la expresión de *Hes-1* (hairy and enhancer of Split 1) es modulada en NPC a través de los niveles de glucosa extracelular y la acción coordinada de *CREB* (cyclic AMP responsive element binding protein) y SIRT1, un exceso de glucosa reduce la expresión de *Hes-1* mediada por *CREB* y a su vez promueve la represión del promotor de *Hes-1* mediada por SIRT1 a través de la desacetilación de histonas en zonas de cromatina aledañas a este elemento regulador. Esto tiene como consecuencia una disminución en la proliferación celular de los NPC y su capacidad de diferenciación. En contraste, los niveles bajos de glucosa *CREB* desplaza a SIRT1 y se une a una región CRE (cAMP response elements) promoviendo la transcripción de *Hes-1* y la maduración y proliferación de los NPC (Fusco et al., 2016).

Por último, la SIRT1 reprime el proceso de diferenciación en las hESCs. Los niveles altos de expresión de SIRT1 se relacionan con la represión de genes de diferenciación como *DLL4*, *PAX6* y *WNT6* que están relacionados con la morfogénesis de la retina. De forma recíproca, la diferenciación celular reduce los niveles de mRNA y proteína de SIRT1 en hESCs (Calvanese et al., 2010).

Lo anterior demuestra la importancia que tiene la SIRT1 en la ESCs manteniendo la regulación de la pluripotencia mediante interacción directa desacetilando factores de transcripción o favoreciendo la interacción de estos para regular vías de diferenciación, al mismo tiempo SIRT1 es esencial para el establecimiento de los linajes embrionarios (Mesodermo, endodermo, ectodermo) posteriores en modelos como el ratón o el ser humano. Sin embargo, se desconocen los efectos específicos sobre los programas de transcripción que dependen de esta regulación.

### **1.7. scRNA seq y la heterogeneidad celular.**

Las técnicas de análisis transcriptómicos acoplados a tecnologías de secuenciación de nueva generación permiten conocer los valores de expresión del RNA mensajero de un grupo celular determinado con mucha precisión. Además, permiten llevar a cabo estudios de comparación o expresión diferencial entre grupos para distinguir genes o factores de transcripción relevantes en condiciones clínicas o en ámbitos ecológicos y ambientales. Sin embargo, la mayoría de estas solo permiten obtener un promedio general de la expresión de las células muestra (Stark et al., 2019) y no poseen la capacidad de discernir a nivel célula individual las diferencias de expresión que pueden existir entre los tipos celulares presentes. Lo anterior puede volverse problemático ya que a pesar de que los tejidos pueden parecer homogéneos a simple vista, las células poseen programas de expresión génica individuales y altamente específicos incluso dentro de los mismos tipos celulares (Hang, 2009).

Para poder resolver lo anterior se desarrolló la técnica llamada single cell RNA seq (scRNA seq), que permite obtener los niveles de transcripción de células de forma individual, constituyendo una herramienta que permite obtener un set de genes diferencialmente expresados o análisis de enriquecimiento y a la par provee información para distinguir grupos dentro de las muestras analizadas, anotar y caracterizar esos grupos con genes variables de nuestro interés e incluso formular modelos de interacción entre estas subpoblaciones (Hwang et al., 2018).

El scRNAseq no es un método único, sino un conjunto de protocolos variables con fortalezas (Hwang et al., 2018) y limitaciones que se adaptan a distintos sistemas y diferentes preguntas científicas, algunas de sus aplicaciones puede ser la identificación de poblaciones celulares enrarecidas, lo que significa que una enorme cantidad de células debe ser examinada. Un ejemplo de esto es el trabajo de (Campbell et al., 2017) donde se secuenciaron 20,921 células de hipotálamo de ratón para poder identificar subpoblaciones neuronales compuestas hasta por 50 células bien caracterizadas. Otra aplicación de esta técnica es el discernimiento entre poblaciones celulares sumamente parecidas lo que requiere una técnica que brinde mayor precisión en la captación de genes poco expresados y disminución del ruido técnico, lo cual se puede lograr utilizando protocolos de scRNAseq de alta sensibilidad.

De forma resumida, todo experimento de scRNAseq consta de tres etapas: 1.- Aislamiento de células individuales, 2.- Construcción de librerías y 3.- Secuenciación.

### **1.7.1. Aislamiento de células**

La primera etapa del scRNAseq puede ser llevada a cabo de distintas formas, el *sorting* activado por fluorescencia (FACS) es un tipo especial de citometría que permite separar células en pozos individuales mediante el acoplamiento de sondas fluorescentes a las células que son posteriormente detectados por la interacción con anticuerpos específicos (Gross et al., 2015). Sin embargo, esta técnica tiene como requisito una suspensión celular inicial con alto contenido de células, alrededor de 10,000, por lo que muestras que cuentan con un bajo contenido celular o condiciones experimentales que generan la recuperación de un bajo número celular no son compatibles. Otra limitación de esta técnica es que puede afectar la viabilidad celular de la muestra debido a la velocidad a la que son sometidas las células durante el *sorting* (Hu et al., 2016).

Una segunda aproximación para el aislamiento de células individuales es la micromanipulación o selección manual. Esta técnica consta de un microscopio invertido con un par de micropipetas que son controladas a través de motores para mayor precisión. Esta técnica brinda ventajas como la posibilidad

de observar las células para separar aquellas que cumplen con una morfología óptima y permite un manejo preciso y delicado de muestras frágiles como son células de embriones. A pesar de lo anterior, esta técnica está limitada al requerimiento de especialistas para llevarla a cabo así como la cantidad de células que se pueden aislar en cierta cantidad de tiempo (Citri et al., 2011).

En la actualidad una de las metodologías mejor estandarizadas y más ampliamente utilizadas es el aislamiento de células mediante tecnología de *droplets*. Algunos de los protocolos más conocidos son InDrop, Dropseq y 10X Chromium, que se basan en chips de microfluidos donde se mezcla la muestra de células, perlas de gel o *gel beads* que contienen oligos con secuencias capaces de etiquetar y capturar el mRNA de las células, aceite de partición y un mastermix que permite llevar a cabo la reacción de retrotranscripción. Las células son aisladas mediante la formación de una emulsión donde cada *droplet* contiene una *gelbead*, una célula y una fracción del mastermix (Salomon et al., 2019). Una de las ventajas más notorias de este sistema es su eficiencia de captura de muestra así como el acoplamiento de los pasos de retro transcripción (RT) del mRNA de la muestras, amplificación del cDNA obtenido tras la RT y la creación de librerías dentro de un mismo protocolo estandarizado (Andrews & Hemberg, 2018b).

### **1.7.2. Corrección del ruido técnico**

El volumen y complejidad de los datos producidos por un scRNA-seq ha abierto nuevas fronteras en diversos campos de la investigación biomédica que ha sido superficialmente explorada en los años recientes (Angerer et al., 2017). Lo anterior tiene como consecuencia la aparición de mucho ruido técnico que podría ser contraproducente en el posterior análisis. Para lidiar con esto, se utilizan identificadores moleculares únicos (UMIS) o RNA exógeno (ERCC Spike-ins) que ayudan a reconocer las señales biológicas dentro de la muestra (Andrews & Hemberg, 2018b).

Los UMIS son secuencias o índices de 4 a 10 bp que se agregan en el extremo 3' o 5' de los transcritos capturados durante la creación de librerías. Estos permiten etiquetar los mRNA capturados y ayudan a eliminar el ruido técnico causado por el evento de amplificación del DNA, asignando a cada uno de estos un evento de retrotranscripción individual, estimando así el conteo inicial de moléculas durante la captura del mRNA de cada célula y eliminando duplicados durante la reacción en cadena de la polimerasa (PCR) (Zheng et al., 2017).

El uso de UMIS como controles para el ruido técnico son recomendados en tres escenarios: 1.- Muestra de inicio baja o poca cantidad de mRNA por célula, 2.- Secuenciación profunda de librerías de scRNA-seq (>80 millones de lecturas por muestra) y 3.- La detección de mutaciones específicas de muy baja frecuencia durante la secuenciación del DNA (Phipson, 2017).

Por otro lado los ERCC Spike-ins son 176 secuencias de DNA que son sintetizadas como DNA creadas en el año 2003 por el External RNA Controls Consortium (ERCC), estas secuencias son agregadas a los experimentos durante el diseño experimental y la finalidad de esto era otorgar secuencias en concentraciones iniciales conocidas y estandarizadas, que permitieran distinguir el ruido de fondo generado por duplicados de PCR o extrema variación en la detección de ciertos genes a través de una comparación con el valor de los ERCC spike-ins ((Cronin et al., 2004).

Sin embargo los ERCC spike-ins han mostrado una variabilidad técnica muy alta dentro de los experimentos en comparación con los genes endógenos a comparar ("A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium," 2014). También se ha demostrado que estas secuencias son influenciadas por efectos biológicos como la baja disponibilidad de mRNA y no permitían la captura eficiente del transcriptoma en muestras con estas características (Tung et al., 2017) siendo una medida de control de ruido técnico insuficiente para discernir entre una muestras de baja o alta calidad.

### **1.7.3. Alta dimensionalidad.**

Una vez superada la problemática causada por la distinción entre ruido técnico y señal biológica tras la secuenciación de las librerías obtenidas durante un protocolo de scRNAseq, debemos solucionar la existencia de un espacio con múltiples dimensiones y su representación. Un experimento de scRNA seq nos permite obtener información de todos los genes que están siendo transcritos en todas las células de nuestra muestra (o de un número determinado durante el diseño experimental), pero la aparición de miles de genes en miles de células supone un desafío estadístico. Este fenómeno es conocido como "Maldición de la dimensionalidad" y se refieren a que en espacios estadísticos con tantas dimensiones se torna complicado distinguir entre la variación Inter poblacional y las verdaderas diferencias que distinguen a cada población una de otra (Altman & Krzywinski, 2018). Para lidiar con esta problemática se puede optar por proyectar los datos en espacios de menor dimensión en un procedimiento conocido como "Reducción de dimensiones". Estos espacios son definidos a través de algoritmos que buscan capturar la variabilidad entre las células/muestras así como algunas características específicas. El uso de estas técnicas conlleva a la pérdida de información por lo que la selección de qué algoritmo se aplica depende de que sets de características son de interés para el experimento (Andrews & Hemberg, 2018b).

El primero de los algoritmos es el análisis de componentes principales (PCA) que permite capturar la variabilidad entre las muestras a través de la creación de nuevas variables conocidas como "componentes" que buscan contener la mayor cantidad de esta variabilidad en ellas. Posteriormente, las muestras son proyectadas en un espacio compuesto por dos dimensiones donde el eje x representa el componente 1 (que alberga el mayor porcentaje de variabilidad del experimento) y el eje

y el componente 2 (el segundo componente con mayor variabilidad dentro del experimento). En este tipo de datos, las muestras con mayor similitud entre ellas se agrupan y a la vez se alejan de aquellas muestras con mayor disimilitud. Una de las mayores consideraciones a tomar con el uso del PCA es que este algoritmo asume que los datos se encuentran en una distribución normal, lo que no es característico de los *sets* de datos obtenidos en experimentos de scRNAseq (Andrews & Hemberg, 2018a).

Por otro lado, la incrustación de vecinos estocásticos distribuidos en t (tSNE) es un algoritmo estocástico diseñado específicamente para la representación de datos de alta dimensionalidad preservando las interacciones locales entre las células, refiriéndose a los vecinos más cercanos de cada células, lo que permite distinguir con precisión a través de las poblaciones celulares generando *clusters* bien definidos entre ellas, sin embargo mucha de la información global (clústers que separan poblaciones celulares o estados) se pierde al llevar los datos a un espacio métrico euclidiano en 2 dimensiones. (Andrews & Hemberg, 2018b).

Una de las desventajas más notorias de este algoritmo es que al ser estocástico su aplicación múltiples veces sobre un set de datos puede generar resultados visualmente distintos, así como diferencias en los agrupamientos. Una buena práctica al utilizar este algoritmo es la aplicación múltiple sobre los datos para verificar la robustez de los *clusters* observados (Andrews & Hemberg, 2018b).

Los perfiles de expresión obtenidos mediante scRNA seq han sido utilizados para interpretar la relación de los distintos linajes celulares durante el desarrollo (Petropoulos et al., 2016), inferir las trayectorias de diferenciación celular a partir de perfiles de expresión determinados (Miragaia et al., 2019), explorar el repositorio inmune ante enfermedades infecciosas (Yao et al., 2019), entender la heterogeneidad de los tumores (Wagner et al., 2019) y dar nuevos panoramas en el entendimiento de la resistencia a fármacos en el carcinoma (Shaffer et al., 2017). Lo anterior convierte al scRNA seq en la técnica ideal para descifrar la heterogeneidad en las clonas de H9 con delección de SIRT1 en su estado pluripotente y durante un proceso de diferenciación a algún promotor celular temprano (NPC en este caso particular), no solo por su capacidad de discernir diferencias mínimas entre células individuales, sino también por la cantidad de información que puede obtenerse de los datos generados por la técnica.

#### **1.7.4. Visualización de la estructura y sus transiciones en datos de alta dimensionalidad (PHATE)**

Como se mencionó anteriormente, hoy en día contamos con una amplia variedad de métodos y algoritmos de reducción de dimensiones como el PCA, tSNE y UMAP que se encuentran optimizados y son ampliamente aplicados en datos generados por scRNAseq con la finalidad de poder visualizar los

clúster y poblaciones formadas en las muestras procesadas, sin embargo en ocasiones pueden no ser la mejor decisión al momento de buscar información más detallada en la estructura global de los datos debido a su alta sensibilidad al ruido dentro de los datos.

Las técnicas anteriores sin embargo poseen una gran cantidad de limitaciones y problemas al momento de buscar reducir un espacio multidimensional (el universo de células y la expresión de sus miles de genes) a un espacio de bajas dimensiones como puede ser un plano en 2 dimensiones, por ejemplo, se ha demostrado que el tSNE es altamente sensible al número de vecinos cercanos elegidos al implementarlo (parámetro conocido como “*perplexity*”) generando visualizaciones erróneas de la estructura de los datos (Kobak & Berens, 2019), por su lado UMAP tiene problemas similares debido a que se ha reportado que en distancias locales posee una distancia de Jaccard (un coeficiente de disimilitud entre dos sets de datos que va de 0 a 1 siendo 1 el valor de disimilitud más alto) promedio de 0.7 respecto al espacio multidimensional (definido como la matriz de expresión de genes normalizada y escalada); lo mismo se observa con las interacciones globales (aquellas ocurridas entre distintos tipos celulares o condiciones) donde el orden de cercanía en el espacio métrico euclidiano de las identidades celulares tanto en UMAP como en tSNE tenía una correlación muy baja respecto a lo observado en el espacio multidimensional (Chari & Pachter, 2023). Para resolver estos problemas se han presentado algunas nuevas formas de visualización que buscan conservar la máxima cantidad de información global y local como es el caso del Potencial de difusión de calor para embeddings de transición basadas en afinidad (PHATE).

PHATE genera una visualización en dos dimensiones que muestra una representación sin ruido de ambas tipos de relaciones en los datos (global y local), esta es una de sus ventajas más fuertes cuando se trata de obtener trayectorias a diferencia de otras herramientas para este mismo fin como Monocle 2 (Moon et al., 2019) que asume como verdad base que los datos se pueden ajustar a la estructura de un árbol con ramificaciones. La forma en que PHATE logra la reducción de dimensiones y visualización de los datos se puede resumir de la siguiente forma:

1.- *Matriz de distancias y conversión a afinidades.* – En primer lugar, se crea una matriz de distancia entre todos los puntos en los datos (en nuestro caso cada punto representa una célula), posteriormente estas distancias se convierten en los *inputs* de un proceso de normalización de filas y conversión de distancia a afinidades (o similitud) mediante un *kernel* Gaussiano.

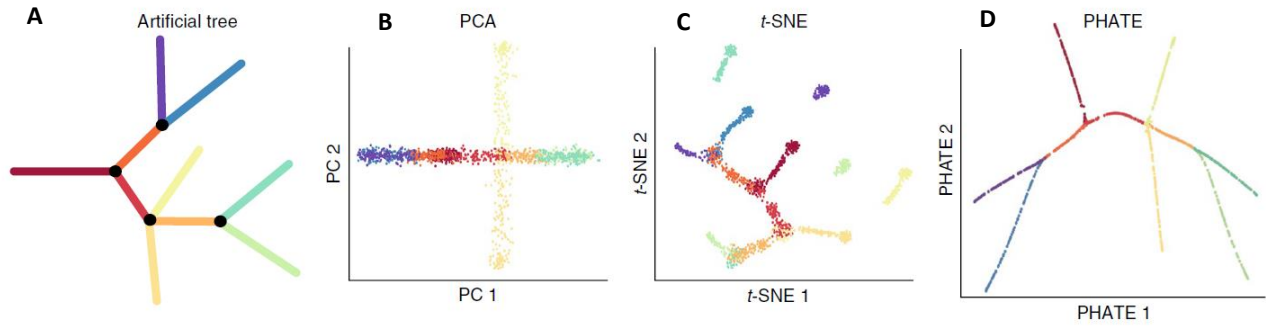
2.- *Codificación de la estructura global de los datos.* – Una vez que se estableció la relación entre cada punto y aquellos inmediatamente adyacentes se debe evaluar la distancia y relación

de cada punto con todos los puntos restantes en el set de datos, para esto PHATE implementa un método conocido como “difusión” que consiste en evaluar la probabilidad de que se haga una caminata al azar desde un punto hacia otro todos los demás presentes en nuestros datos. Se toman en cuenta todas rutas posibles de este caminando azaroso y se elimina el ruido dentro de la estructura de los datos mediante la eliminación de caminos pocos probables generados por el ruido dentro del set de datos. Finalmente se hace una comparación usando el contexto global de difusión de cada punto con todos los restantes del set conservando así las relaciones entre los puntos adyacentes y aquellos más distantes. Esto nos permite representar cada una de las células como el *set* de probabilidades de difusión hacía todas las demás dentro de los datos

3.- *Calculo de divergencia M entre células.* – Debido a que cada punto de nuestros datos se representa como un conjunto de probabilidades se dibujan una distribución de probabilidades para cada célula, la distancia entre cada una de las células se calcula usando la divergencia  $M$  que es una medida de distancia estadística y se puede definir como la separación de una distribución de probabilidades  $P$  a una distribución de probabilidades  $Q$  (una forma ideal de obtener distancias entre nuestras células en este punto del proceso).

4.- *Transformación de potencial de difusión a representación visual.* – Finalmente se utilizan las divergencias previamente calculadas y se aplica la Métrica de Escalamiento Multidimensional o MDS que permite crear un gráfico de 2 dimensiones permitiendo conservar tanto las distancias globales como locales de nuestros datos.

Una forma de apreciar la capacidad de este método de visualización creado por el grupo de (Moon et al., 2019) es mediante la comparativa de PHATE con otros métodos (Figura 6), ya que ellos llevaron a cabo la simulación de un set de datos de scRNAseq con una estructura artificial ramificada y previamente conocida (Figura 6 A) que posteriormente buscaban recapitular mediante el uso de tres métodos de reducción de dimensiones y representación visual: PCA, tSNE y PHATE (Figura 6 B,C y D respectivamente).



**Figura 6:** Comparación de métodos de visualización y reducción de dimensiones en set artificial. (A) Set de datos artificial utilizando como verdad inicial, (B) Implementación de PCA sobre set artificial, (C) Implementación de tSNE sobre set artificial de datos, (D) Implementación de PHATE sobre set de datos artificiales. Tomado y modificado de Moon et al., 2019.

El set de datos artificial (Figura 6 A) muestra una serie de clústeres de distintos colores que se conectan en patrones locales específicos (por ejemplo, clúster naranja conectado con clústeres azul y morado) y patrones globales (la específica secuencia de los colores de cada clúster), en el caso del PCA (Figura 6 B) se observa que se conserva en mayor parte la estructura global de los datos ya que la secuencia de los colores de cada clúster es muy similar a la del set original, sin embargo se puede observar que a nivel local muchos de los clúster se superponen no permitiendo identificar la estructura local de cada rama del set de datos (por ejemplo los clúster morado y azul se ven mezclados).

En el tSNE (Figura 6 C) se puede apreciar que los contactos locales entre cada clúster es mejor conservada que en el caso del PCA, un claro ejemplo de esto es la similitud que hay en las ramas dentro de este gráfico en comparación al original, pese a lo anterior también podemos notar que esta metodología de reducción de dimensiones fragmenta los clúster y la estructura general de estos datos mostrando clústeres dispersos y sin conectividad, esto es muestra de su debilidad conservar la estructura global de los datos.

Por último, PHATE (Figura 6 D) muestra la mayor conservación de estructura tanto a nivel global como local ya que la secuencia de colores de los clústeres así como las ramas del árbol generado con esta metodología es idéntica a la del set artificial creado, lo anterior es una clara muestra de las ventajas que otorga PHATE en sets de datos con transiciones que no solo dependen de las interacciones de poblaciones adyacentes sino también de aquellas poblaciones o clústeres más alejados.

Por lo anterior podemos afirmar que PHATE permite conservar la estructura de los datos especialmente en aquellos que poseen una estructura de transición o continua, un ejemplo claro de esto son los sets de datos relacionados con procesos de diferenciación o aquellos compuestos de muestras que son sometidas a tratamientos y posteriormente sus datos son obtenidos en puntos



temporales específicos. Esta metodología se vuelve entonces un candidato ideal para probar las diferencias existentes en el proceso de diferenciación a progenitores neurales de nuestras muestras con delección de SIRT1 y WT tanto en su estructura como en los genes que favorecen dichas estructuras.

### **1.8. Obtención de genes diferencialmente expresados (DEG)**

Otro de los objetivos centrales al realizar un scRNAseq es la obtención de los genes que se encuentran expresados en nuestras muestras al momento de realizar nuestro experimentos, este perfilado detallado es el que nos permite clasificar a la células en clústeres ya que aquellas células más similares expresarán un programa transcriptómico similar, sin embargo en muchos casos también se dese conocer la diferencia puntual entro dos grupos (en nuestro caso WT vs *SIRT1*<sup>-/-</sup>) para poder tener una idea más acertada de que cambios en la expresión dentro de las células podrían impulsar las diferencias fenotípicas observadas.

Al proceso de obtención de estos genes se le conoce como análisis diferencial y tiene múltiples funciones como puede ser encontrar los genes marcadores que definen nuevas poblaciones celulares o encontrar aquellos genes que promueven heterogeneidad celular dentro de una población, también estos DEG funcionan como inputs para análisis posteriores como pueden ser análisis de enriquecimiento de vías o análisis de alteración de vías canónicas (Schubert et al., 2018).

En un inicio para llevar a cabo este análisis se utilizaban las herramientas desarrolladas para esta tarea en bulk-RNAseq, sin embargo éstas no estaban diseñadas para el manejo de las características de un scRNAseq como son la alta dimensionalidad generada por tener grandes cantidades de células con un orden de cientos a miles de genes secuenciados y expresados, sumado a lo anterior también podemos encontrar que estas herramientas desarrolladas en bulk-RNAseq no estaban estadísticamente diseñadas para manejar la gran cantidad de ceros presentes en las matrices de cuentas de los datos de scRNAseq ((Dal Molin et al., 2017).

Ante esta problemática se han desarrollado una amplia variedad de nuevos algoritmos y metodología para obtener DEG especialmente diseñadas para scRNAseq, a grandes rasgos podemos resumirlas de la siguiente forma:

1. *Modelos lineales generalizados (GLM)*. - Este tipo de modelos asumen que la expresión de los genes (o counts) siguen un patrón de familia exponencial, que puede expresarse con una función, entre los genes que estamos midiendo y los DEG que buscamos. Existen múltiples metodologías que utilizan este modelo como ZINB-WaVE (Risso et al., 2018), ZingeR (Van den Berge et al., 2018) y SwarnSeq (Das & Rai, 2021).

Algunas de las limitaciones más grandes de este tipo de modelo es el hecho de que asume una distribución probabilista dentro del set de datos, de forma típica asume una distribución de la familia de exponenciales (como una distribución binomial, multinomial o de Poisson). De igual forma, este tipo de modelos son sumamente intensos en términos de recursos computacionales tomando incluso 10 a 12 horas en llevar a cabo el análisis diferencial en un grupo (Das et al., 2021).

2. *Modelos generales aditivos (GAM)*. – Estos modelos son muy similares a los GLM, sin embargo, poseen mayor flexibilidad a la hora de estimar la expresión de un gen en función a las condiciones, grupo o contexto de las células ya que permite que estas estimaciones se den en distribuciones lineales y no lineales, no requiere de una distribución específica de los valores de *counts* ya que utiliza una función suave para ajustar una curva a los datos de expresión, esto con la finalidad de capturar la información y co-varianza en la expresión de todas las células, algunas de las librerías o paqueterías que utilizan este modelo son Monocle (Trapnell et al., 2014), Monocle2 y tradeSeq (Van den Berge et al., 2020).

Dentro de las limitaciones más grandes de estos modelos encontramos la necesidad de contar con un ordenamiento de las células, por ejemplo, sobre un eje de tiempo (pseudotiempo) o sobre una trayectoria (análisis de trayectoria) a través del espacio multidimensional que forman las células (Das et al., 2021).

3. *Modelo mixto (MM)*. – El planteamiento central de este modelo asume que la distribución de los valores de expresión observados en todos los genes está compuesta por una mezcla de otras múltiples distribuciones, cada uno representando una fuente distinta de variación como puede ser ruido técnico, variabilidad biológica o errores técnicos. El objetivo final de estos modelos es encontrar la composición de distribuciones dentro de la muestra y así lograr obtener el valor “real” de expresión de cada gen. Algunas de las herramientas más populares que utilizan este enfoque son SCDE (Kharchenko et al., 2014), BASiCS (Vallejos et al., 2015) y SC2P (Wu et al., 2018).

4. *Aproximaciones basadas en el modelo Hurdle*. – En las aproximaciones basadas en el modelo Hurdle podemos observar que los datos son separados en dos sets, el primero de ellos representando los genes que tienen un valor de expresión igual a cero en todas las células (zero-count) y en el segundo de ellos todos aquellos genes que se expresan en algún nivel (non-zero counts). La primera parte del modelo consiste en ajustar una distribución a los genes en el set zero-counts, posteriormente modela la distribución de los genes non-zero counts, finalmente genera una función de distribución

truncada en cero, y que los niveles de expresión crudos de un gen no pueden ser negativos. Esta clase de modelos destacan debido a su capacidad de manejar la gran cantidad de ceros presentes en las matrices de expresión de los datos de scRNAseq evitando que el ruido generado por estos enmascare la expresión de los genes biológicamente relevantes en la muestra. Una de las herramientas más conocidas dentro de esta categoría es MAST (Finak et al., 2015), sin embargo es importante mencionar que este modelo puede verse limitado si los ceros presentes en la matriz de datos responden al carácter estructural de la población ya que asume que todos los ceros presentes responden a un origen común, el muestreo (Jiang et al., 2022).

5. *Aproximaciones de comparación paramétrica.* – Este tipo de aproximación y modelo evita el uso de funciones complicadas o el ajuste de curvas o funciones no lineales en los datos, este grupo de modelos se basa en parámetros más simples de los datos como el promedio de expresión general de los genes, su dispersión y varianza para posteriormente comparar la expresión general de cada gen con la global al comparar dos grupos, existen muchas herramientas basadas en esta idea dentro de las cuales destacan scDD (Korthauer et al., 2016) y DEsingle (Miao et al., 2018).

La limitación más conocida de este tipo de modelos es que solo puede realizar la comparación de dos grupos en situación uno a uno lo que no es muy común en datos generados por scRNAseq ya que estos suelen estar compuestos por múltiples poblaciones y muchos tipos celulares (Soneson & Robinson, 2018).

La metodología seleccionada para analizar los datos de expresión diferencial dependerá de la naturaleza o la biología de la muestra a analizar, por ejemplo, si es algo esencial una cantidad alta de genes con una expresión de cero no se recomendaría usar un modelo de expresión diferencial basado en modelos Hurdle, en cambio si se tiene un set de datos con mucho ruido generado por la inflación de ceros (proveniente de un muestreo o captura ineficiente de RNA o problema técnicos durante la obtención de los datos) en la muestra esta podría ser la mejor opción a elegir.

### **1.9. Inferencia de actividad de Factores de transcripción**

Dentro del análisis de datos de scRNAseq podemos encontrar un grupo de análisis conocido como análisis accesorios, estos análisis van más allá de los análisis básicos que incluyen expresión diferencial, clustering, visualización, anotación de identidades celulares y enriquecimientos, sin embargo hay una mayor cantidad de información aún disponible en los datos de scRNAseq y que pueden ayudarnos a comprender aún más nuestras muestras.

La regulación de la expresión de los genes es fundamental en procesos celulares como el desarrollo, la diferenciación y en el proceso de conservar la homeostasis en nuestros tejidos, los mecanismos

regulando esta expresión pueden ser de muchos tipos, desde marcas epigenéticas que alteran la estructura tridimensional del genoma y evitan la expresión de sitios genómicos específicas, hasta la interacción directa de proteínas conocidas como factores de transcripción (TF) que regulan la expresión de un set de genes.

Debido a su capacidad interactuar con regiones reguladoras en el DNA los TF se han propuesto como marcadores de la actividad de vías y con el tiempo la capacidad de medir o inferir su actividad ha ganado atención. Una de las formas *in silico* utilizada para inferir la actividad de estos TF es la estimación de la actividad de un determinado TF a través de los niveles de expresión de sus genes objetivo, esto es conocido como reguloma de TFs (Trescher et al., 2017).

Se han desarrollado una gran cantidad de metodologías con la finalidad de definir que combinación de genes y en qué nivel de expresión podrían permitir determinar la actividad de un TF específico, estas estrategias se clasifican de acuerdo con la fuerza y robustez que respalda la interacción entre un TF y sus genes objetivo; una de estas estrategias es la de crear bases de datos con material de publicaciones científicas curada manualmente. Esto permite tener datos de alta calidad debido a que se cuenta con experimentos puntuales que refuerza las interacciones (Lesurf et al., 2016; Orchard et al., 2014); sin embargo, cada base de datos cuenta con sus propios criterios de selección, protocolos de control de calidad y curación lo que puede crear discrepancias entre ellas, lo que finalmente limita la información disponible para contraste (Tripathi et al., 2016).

En respuesta a esta clase de problemas otras bases de datos proponían la curación de datos de técnicas acopladas a tecnologías de secuenciación como el CHIP-seq que permitían determinar los objetivos con los que se unía un TF, sin embargo la limitación más inmediata a esta metodología era que los experimentos no eran realizados bajo las mismas condiciones en muchos de los casos y que se reportaba que algunas de las uniones detectadas en estas metodologías podían no causar ningún cambio en la expresión del gen (Fisher et al., 2012).

Una tercera metodología complementaria surgió para solucionar este problema, esta se basada en la predicción y presencia de “motivos de unión” específicos para el TF en la región donde se detectaba la interacción. Estos motivos eran predichos a través de las preferencias que tienen los sitios de unión al DNA que se han reportado en múltiples TF. A pesar de que este proceso permitía robustecer la información sobre qué genes son objetivos potenciales de un TF específico, se veía truncada por el hecho de que solo se podía predecir aquellos motivos conocidos y validados previamente (Khan et al., 2018; Kulakovskiy et al., 2018).

En la actualidad se han creado metodologías con la capacidad de mezclar estas estrategias mencionadas y acoplarlas a valores de expresión obtenidos de análisis como el scRNAseq. Esta clase de análisis no es dependiente de DEG si no que toma en cuenta aquellos genes que poseen una mayor variación en sus niveles de expresión a lo largo de todas las células (Genes altamente variables) para poder determinar si el patrón de expresión de estos puede dar señal de la actividad de TFs específicos. Una de las herramientas que permite llevar a cabo esto es DOROTHEA (García-Alonso et al., 2019). DOROTHEA permite acceder a un compendio de 1541 TF que fueron curados y validados con tres métodos; curación manual de literatura que valida las interacciones TF-objetivo, 2829 CHIP-seqs con control de calidad y normalización para los TF curados de la literatura y los sitios de unión de TF predichos en humano y ratones presentes en bases de datos como JASPAR y HOMOCOCO.

Esta clase de análisis permiten llevar a cabo una integración profunda de los datos obtenidos tras un scRNAseq ya que utilizan otras propiedades de la señal biológica capturada de forma que aportan mayor entendimiento sobre el problema biológico evaluado.

#### **1.10. Análisis de trayectoria y pseudotiempo**

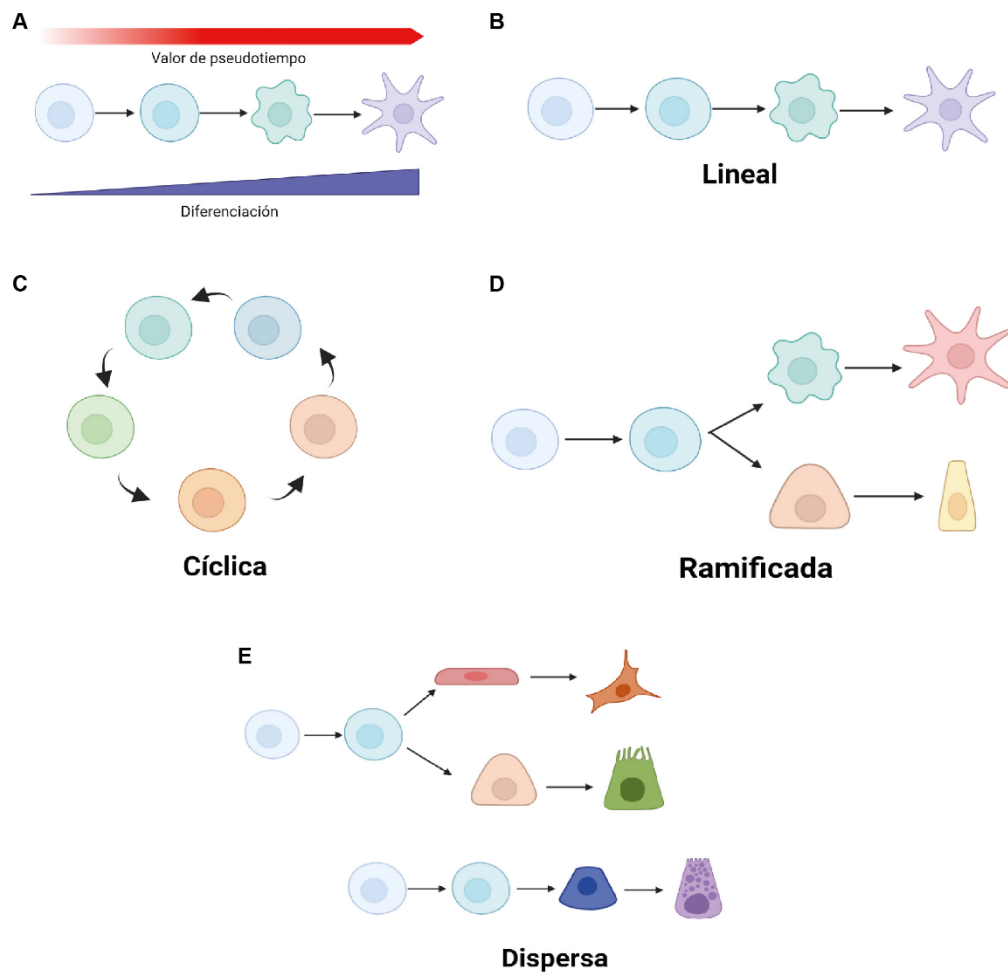
Muchos procesos biológicos pueden manifestarse como un continuo de cambios dinámicos en el estado de la célula, uno de los ejemplos más obvio es la diferenciación celular que parte de un estado troncal o no diferenciado hacia subtipo con una mayor especialización; sin embargo, también pueden incluirse procesos como el ciclo celular o la activación inmune.

Uno de los cambios que pueden observarse durante estos continuos son las variaciones en la expresión de genes específicos o relacionados al fenómeno. Estos cambios pueden ser relacionados con una “trayectoria” que marca la dirección del flujo de los cambios.

Las trayectorias pueden ser sumamente simples siendo una línea o una curva que va del punto A al punto B; sin embargo, también pueden encontrarse trayectorias con topología complejas donde se puedan observar ramificaciones y múltiples nodos de termino.

En segunda instancia tenemos el pseudotiempo que se define como la posición de las células a lo largo de esta trayectoria. Se representa con un valor número que indica que tan adelante o atrás se encuentra una célula en el fenómeno, por ejemplo, en un eje de diferenciación de células troncales embrionarias hacia un linaje como progenitores neurales, las células troncales tendrían el valor de pseudotiempo más pequeño (Figura 7 A).

Existen una gran variedad de métodos para inferir trayectorias dentro de datos de scRNAseq, cada uno de ellos se adapta a distintas topologías que podrían estar presentes en la estructura del proceso estudiado (Figura 7). Los primeros modelos desarrollados para esta función se basaban en modelos lineares (Figura 7 B) donde se proponía ajustar una línea que pasará a través de los diversos clústers dentro de los datos (Bendall et al., 2014), sin embargo, estos métodos no logran siempre capturar la distribución de las células ya que dependen de linealidad dentro de los datos; algunos ejemplos de estos algoritmos son *Wanderlust* (Bendall et al., 2014) y *MATCHER* (Welch et al., 2017).



**Figura 7:** Pseudotiempo y trayectorias. (A) Incremento del valor de pseudotiempo conforme avanza el proceso de diferenciación en una trayectoria lineal; (B,C,D,E) Distintos tipos de trayectorias posibles al realizar un análisis de trayectorias.

Existen otro tipo de algoritmos que pueden ser utilizados en procesos dinámicos cíclicos como el paso de una muestra celular a lo largo del ciclo celular, este tipo de algoritmo produce trayectorias cíclicas (Figura 7 C) algunas de las herramientas especializadas en este tipo de datos son *EIPiGraph* (Albergante et al., 2020) y *reCAT* (Ye et al., 2019).

Es importante tomar en cuenta que no todos los procesos biológicos dinámicos ocurren en trayectorias lineares (punta A hacia punto B), sino que pueden incluir diversas ramificaciones (Figura 7 D) donde las células se comprometen a diversos tipos de linajes; por ejemplo, los diversos linajes producidos durante la hematopoyesis. Algunos de los algoritmos que tienen la capacidad de encontrar este tipo de trayectoria dentro de las células son *Slingshot* (Street et al., 2018) y *fateID* (Herman et al., 2018).

Por último, existen herramientas con la capacidad de no solamente encontrar ramificaciones dentro de las trayectorias sino que además pueden encontrar grupos de células que poseen una trayectoria separada, este fenómeno es conocido como trayectorias dispersas (Figura 7 E). Dentro de las herramientas más destacables en este caso encontramos *PAGA* (Wolf et al., 2019) y *TinGa* (Todorov et al., 2020).

### **1.11. Genes asociados a la trayectoria**

Como se mencionó previamente, hoy en día existen una amplia variedad de metodologías para inferir trayectorias que cruzan a través de las poblaciones celulares presentes en datos obtenidos de un scRNAseq. Este tipo de análisis complementado con el valor de pseudotiempo asignado a cada paso del proceso de diferenciación (Figura 7 A) nos brinda la posibilidad de determinar que genes dentro de la población celular varían sus niveles de expresión en función al valor del pseudotiempo y durante la trayectoria; conocidos como genes asociados a la trayectoria (GAT).

La capacidad de discernir cuales son los GAT depende también del método de inferencia de trayectoria elegido, como se mencionó en la sección anterior, no todos los eventos biológicos estudiados corresponden a una trayectoria simple y lineal por lo que muchas herramientas de búsqueda de GAT basadas en este tipo de trayectoria fallan en su labor.

Esta deficiencia en algunas herramientas ha impulsado la búsqueda de nuevas metodologías para encontrar GAT, como por ejemplo el uso de los DEG dentro de los clústers celulares. Sin embargo, las aproximaciones iniciales basadas en esta idea no logran capturar la información deseada debido a que sus metodologías dependen de la elección por parte del usuario; este debe determinar que clústers se comparan, cuales resultados son importantes para hacer esta comparativa y cuáles de las comparaciones son independientes entre sí.

Algunos métodos posteriores buscan optimizar estas metodologías basadas en DEG mediante el modelado de la expresión de cada gen a través de una función suave a lo largo de los linajes, entre estas metodologías podemos destacar a *Monocle* (Trapnell et al., 2014) que evalúa si la expresión de un gen está relacionada con el pseudotiempo mediante la implementación y ajuste de la expresión del gen a un modelo generativo general en función al cambio en el pseudotiempo (Figura 8 B), esta aproximación también ha sido tomada por metodología como *TSCAN* (Ji & Ji, 2016), este último utiliza un *minimum spanning tree* o *mst* (Figura 8 A) que dibuja a través de los distintos clústeres poblacionales. El *mst* pasa a través de los centroides de cada uno de estos clústeres, y el cambio de los genes es entonces calculado a partir del orden de cada célula en cada clúster a través del *mst*. Otro tipo de métodos son los que por ejemplo implementa *GPfates* (Lönnberg et al., 2017) que utiliza una mezcla de procesos Gaussianos sobrepuestos, donde cada componente mixto del modelo representa un linaje celular, para determinar los genes más importantes se hace la comparativa de similitud de los datos originales con una versión bifurcada, testeando si los genes y su expresión diferencial son esenciales en dichas bifurcaciones. De forma similar, *Monocle2* (Qiu et al., 2017) utiliza una metodología llamada *Branch expression análisis modeling* (BEAM) que permite determinar cuáles son los genes con mayor importancia en los eventos de ramificación de las trayectorias celulares (Figura 8 B).

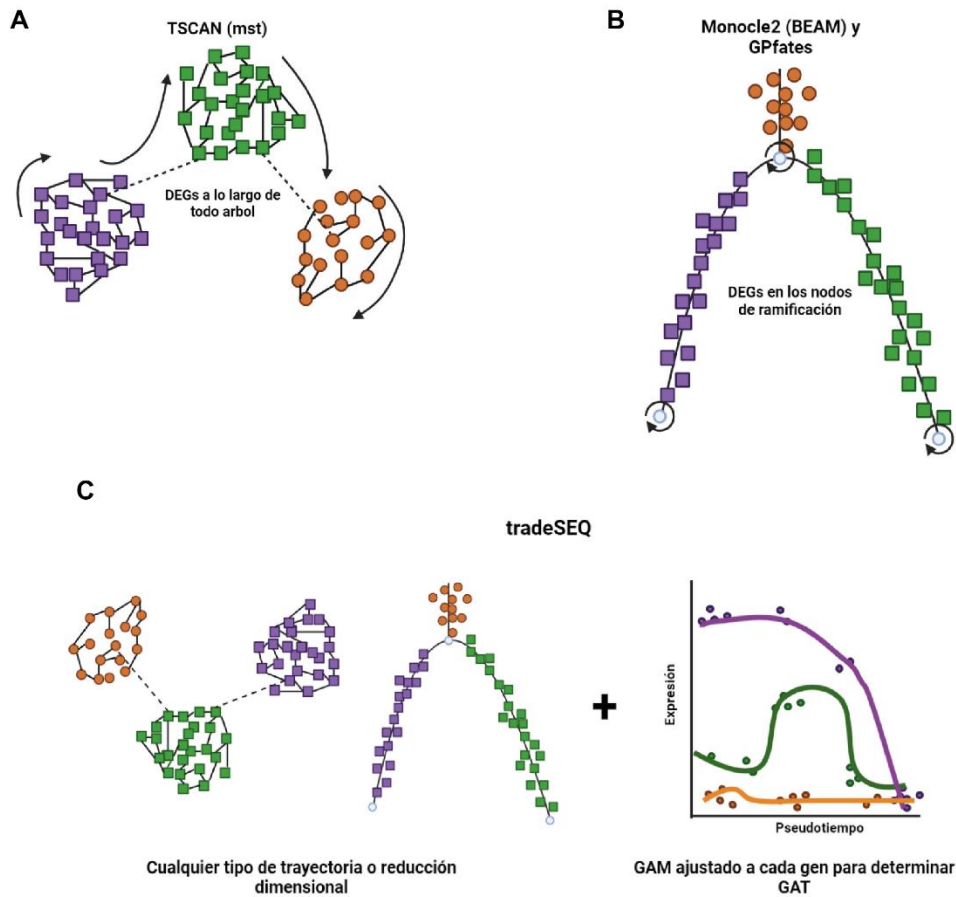
En conjunto, estas metodologías optimizan la inferencia de genes importantes durante los eventos de diferenciación respecto a las versiones basadas en aproximaciones discretas respecto a clústeres celulares debido a su capacidad de capturar información continua de expresión y utilizando una sola evaluación a lo largo de cada linaje, limitando la comparación a linaje contra linaje.

Sin embargo, *GPfates* tiene limitaciones claras como que solo funciona en trayectorias que consisten en máximo una bifurcación, limitando su aplicabilidad a problemas biológicos con máximo dos linajes. Por su parte, BEAM está limitado por la técnica de reducción dimensional disponibles en el software de *Monocle2* como son el análisis de componentes independientes (ICA) (Qiu et al., 2017) y *DDRTree* (Qi et al., 2017).

Para poder resolver las limitaciones de reducción dimensional limitada y trayectorias con una sola bifurcación han surgido alternativas como *tradeSeq* (Van den Berge et al., 2020), que utiliza modelos generales aditivos (GAM) para inferir funciones suavizadoras de expresión de los genes a lo largo del pseudotiempo en cada linaje (Figura 8 C). Posteriormente, realiza pruebas de Wald para poder determinar cuáles genes tienen una expresión que varía de acuerdo con el pseudotiempo. Los valores y estadísticos de estas funciones suavizadoras pueden ser utilizados en una amplia variedad de análisis posteriores que permitirán determinar qué genes son importantes al inicio o al final del



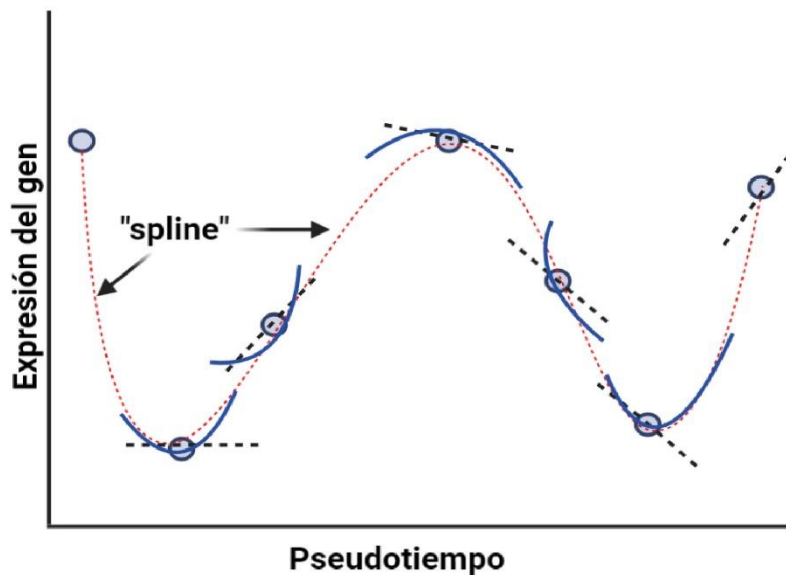
pseudotiempo. Contar con la información de estos suavizadores permite conocer cuáles genes tienen patrones de comportamiento distintos dentro de los linajes y en comparación.



**Figura 8:** Algoritmos de inferencia de trayectorias (IT) y determinación de GAT. (A) Metodología de *mst* aplicada por TSCAN, los nodos representan células donde cada color representa un clúster de células, las aristas representan el MST que conecta cada célula a lo largo de todos los clústeres con mínima distancia, finalmente las flechas representan la dirección hacia donde avanza la trayectoria y la expresión de los GAT; (B) Metodología de BEAM aplicada por Monocle2 y GPfates, los nodos representan células donde cada color representa un clúster de células, las aristas representan la estructura de árbol ramificado asumido como verdad base en estas metodologías, por último las flechas cíclicas indican los puntos donde se calculan los GAT como esenciales para los eventos de bifurcación; (C) Metodología utilizada por tradeSEQ que puede incluir cualquiera de las metodologías A o B agregando además modelos GAM para determinar los GAT de nuestra trayectoria.

El uso de GAMS requiere que la medida de expresión de cada gen y cada célula en el pseudotiempo sean alimentados al modelo como una variable independiente (Figura 9), mientras que el valor de pseudotiempo funge como variable independiente (Van den Berge et al., 2020). Una vez realizado esto, los GAMS utilizan curvas diferenciables en porciones (llamados también “*splines*”) para estimar la relación entre el pseudotiempo y el valor de expresión de los genes (Figura 9). Estas *splines* pueden ser cúbicas, cuadráticas o tener otras formas de función y se eligen a partir de la naturaleza de los

datos (Hastie, 2017). El objetivo final de los GAMs es encontrar el suavizador que describa mejor la relación entre la expresión de los genes y el pseudotiempo. La suavidad de estas funciones se ajusta a través de una validación cruzada entre las múltiples variables dependientes.

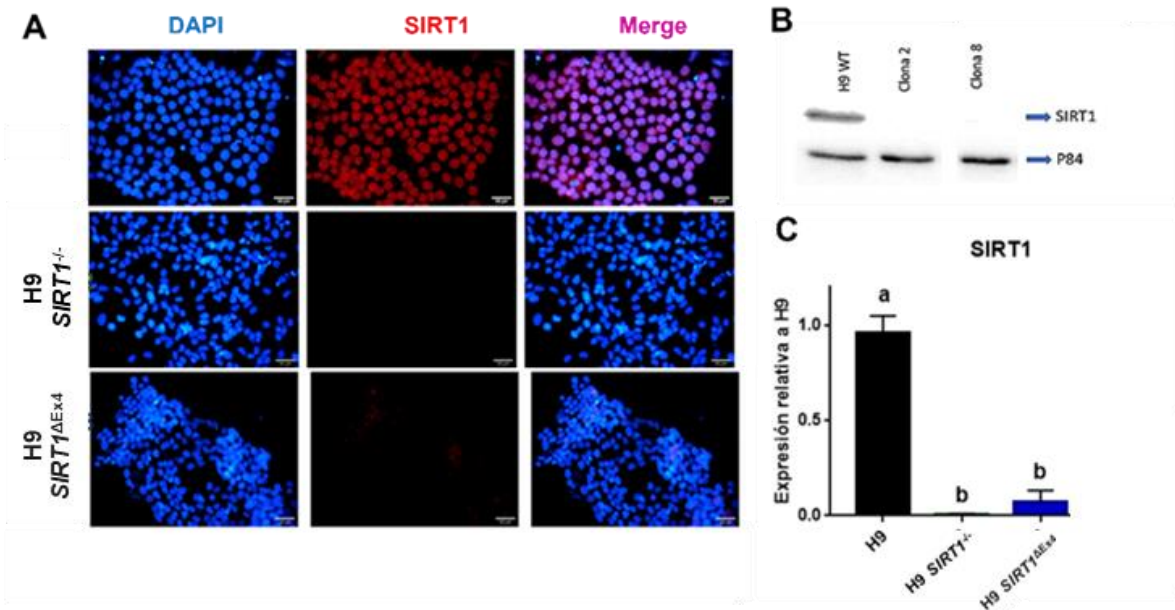


**Figura 9:** "Generalized additive model" o GAM. Se observa el ajuste de suavizadores en cada punto de expresión del gen a lo largo del pseudotiempo que es finalmente conectado por una "spline" o función suavizadora.

Otra ventaja de *tradeSeq* es que puede ser utilizado con cualquier tipo de reducción dimensional (Figura 8 C). De igual forma, puede aceptar el valor de pseudotiempo obtenido de una amplia variedad de metodologías ya que solo requiere que estos sean brindados en forma de una matriz de dos dimensiones. Otra ventaja que posee es la capacidad de analizar trayectorias con múltiples linajes e incluso puede considerar condiciones experimentales dentro de los linajes y llevar a cabo las comparaciones entre estos considerando cada condición para calcular que genes son importantes.

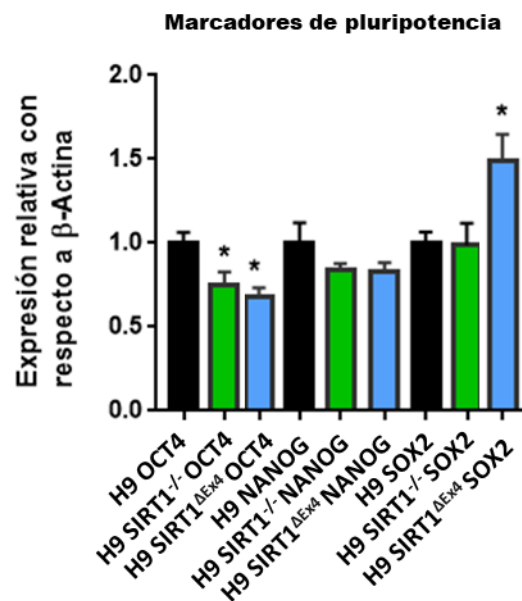
## 2.-Antecedentes específicos:

En el laboratorio ya se ha establecido un modelo de estudio en la línea celular de ESCs humanas H9 (WA09, WiCell Research Institute, Inc.) donde mediante el sistema de edición genética CRISPR-CAS9 se logró hacer una delección de SIRT1 y producir diversas mutantes isogénicas (Figura 10). Los niveles de esta proteína fueron validados a través de Western Blot (Figura 10 B) y de inmunofluorescencia (Figura 10 A) con su respectivo análisis de intensidad relativa de fluorescencia (Figura 10 C).



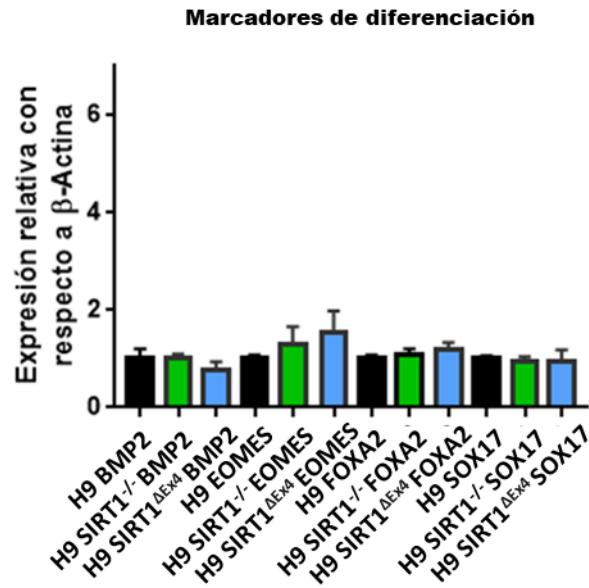
**Figura 10:** Validación de la delección de SIRT1 en la línea celular de hESCs H9. A.- Inmunofluorescencia de la proteína SIRT1; B.- Western Blot representativo de los niveles de proteína SIRT1 en las monoclonas editadas respecto a la condición Wild Type (WT); C.- Análisis de intensidad relativa de fluorescencia de SIRT1 en las monoclonas editadas respecto a la condición H9 WT, estadística se determinó mediante ANOVA de una vía seguida de una prueba post hoc de Tukey's para múltiples comparaciones. Tomado de: Sánchez Ramírez E. Tesis Doctoral, 2021 (Datos no publicados)

Posteriormente se buscó evaluar el efecto causado por la delección de SIRT1 en las mutantes isogénicas (H9 *SIRT1*<sup>-/-</sup> y H9 *SIRT1*<sup>ΔEx4</sup>). Para cumplir esta meta se evaluó presencia y abundancia de los RNA mensajeros (mRNA) de genes marcadores de pluripotencia y de algunos genes de diferenciación temprana de las 3 capas germinales embrionarias, esto se realizó a través de un RT-PCRq. Se encontró una reducción en la abundancia de *OCT4* en todas las mutantes isogénicas respecto a la condición control (Figura 11), en el caso de *NANOG* este fenómeno no fue observado en ninguna de las dos mutantes,; por último *SOX2* muestra un incremento significativo en la condición H9 *SIRT1*<sup>ΔEx4</sup> respecto a todas las condiciones restantes (Figura 11).



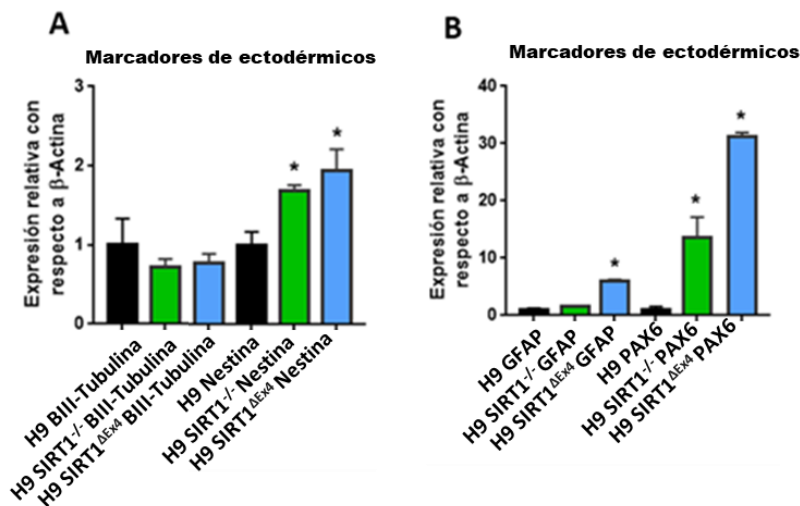
**Figura 11:** Niveles de abundancia de mRNA de marcadores de pluripotencia en células H9 y las mutantes isogénicas en el gen *SIRT1*. Se muestra el promedio +/- la desviación estándar, \* =  $p \leq 0.05$  respecto a condición control. Tomado de: Sánchez Ramírez E. Tesis Doctoral, 2021 (Datos no publicados)

En el caso específico de los marcadores de diferenciación temprana la expresión de los genes *BMP2* y *EOMES*, que son genes de linajes mesodérmicos, no mostraron cambios significativos en ninguna de las mutantes isogénicas *SIRT1*<sup>-/-</sup>, Los genes relacionados con linajes endodérmicos *FOXA2* y *SOX17* no mostrarán ningún cambio significativo respecto a la condición control en ninguna de las mutantes isogénicas editadas (Figura 12).



**Figura 12:** Niveles de expresión de mRNA de marcadores mesodérmicos y endodérmicos en las mutantes isogénicas en *SIRT1* respecto a la condición control. Se muestra el promedio  $\pm$  desviación estándar (SE),  $*= p \leq 0.05$  con respecto a la condición control. Tomado de: Sánchez Ramírez E. Tesis Doctoral, 2021 (Datos no publicados)

Los genes de linajes ectodérmicos como son Nestina, *GFAP* y *PAX6* aumentaron significativamente en las dos mutantes isogénicas respecto a la condición control (Figura 13).



**Figura 13:** Niveles de expresión de mRNA de marcadores ectodérmicos en las mutantes isogénicas con delección de *SIRT1* respecto a la condición control. Se muestra el promedio  $\pm$  desviación estándar (SE),  $*= p \leq 0.05$  con respecto a la condición control. Tomado de: Sánchez Ramírez E. Tesis Doctoral, 2021 (Datos no publicados).

Por otro lado en el laboratorio se había realizado una diferenciación de la línea celular H9 a neuronas de linaje dopaminérgico utilizando el protocolo descrito por (Fedele et al., 2017) buscando evaluar así la capacidad de esta línea celular de convertirse en Neuronas Dopaminérgicas maduras, sin embargo parte del presente trabajo incluyo la exploración de la capacidad de diferenciación de las mutantes con deleción de SIRT1 (específicamente las células H9 SIRT1<sup>-/-</sup>).

Por los resultados anteriormente expuestos se eligió a la H9 SIRT1<sup>-/-</sup> como grupo experimental comparativo en este trabajo, debido a que presenta heterogeneidad en la expresión de marcadores de diferenciación temprana a linajes ectodérmicos, pérdida de expresión del marcador de pluripotencia OCT4 y de las mutantes isogénicas obtenidas es la única con la condición de homocigoto KO de SIRT1 (SIRT<sup>-/-</sup>).

### 3. Justificación:

- La delección de *SIRT1* en la línea celular de ESCs H9 generó cambios en el fenotipo de estas células, de igual forma se presentaron alteraciones en la expresión de algunos marcadores canónicos de pluripotencia y troncalidad (NANOG y SOX2) así como la aparición de otros marcadores de diferenciación a linajes tardíos.

Para poder entender que cambios ocurrieron tras la delección de *SIRT1* se propone realizar la comparación de los programas transcriptómicos de las condiciones H9 WT y H9 *SIRT1*<sup>-/-</sup>, en un estadio basal o de mantenimiento y en un estadio de diferenciación avanzado que será el día 6 de un protocolo de diferenciación a NPCs.

La técnica propuesta para realizar esta comparación es el scRNAseq ya que posee la capacidad de capturar la información transcriptómica de células de forma individual, esto nos permite diferenciar los efectos que tuvo la delección en la población de células H9, también nos brinda la oportunidad de anotar la identidad de las células lo que resulta informativo acerca de cuan troncales continúan siendo las células H9 *SIRT1*<sup>-/-</sup> tras la delección de *SIRT1*.

Los datos obtenidos mediante el scRNAseq posibilitan también el estudio de los cambios en la expresión de los genes a lo largo de un proceso dinámico, como es el protocolo de diferenciación hacia NPCs, mediante análisis de pseudotiempo y trayectoria; lo anterior nos permitirá entender a más profundidad que grupo de genes en las ESCs H9 se ven afectados o dependen de *SIRT1* para llevar a cabo su papel durante la diferenciación celular.

### 3.1. Hipótesis:

- La delección de *SIRT1* genera un cambio significativo en la regulación transcripcional de genes relacionados con pluripotencialidad generando heterogeneidad poblacional intercelular y limitando la capacidad de diferenciación de estas células, estos cambios se verán reflejado en los genes asociados al pseudotiempo y trayectoria en un eje de diferenciación a progenitores neurales.

### 3.2. Objetivo principal:

- Determinar las diferencias en el patrón de expresión de genes dentro de los grupos experimentales y estudiar la heterogeneidad generada tras la delección de *SIRT1* en la línea celular de hESCs H9.

### 3.3. Objetivos particulares:

- Cultivar las líneas celulares de hESCs H9 wild type (H9 WT) y la clona H9 con delección de *SIRT1* (H9 *SIRT1*<sup>-/-</sup>) y realizar la diferenciación celular a linajes mesodérmicos (adipocitos) y ectodérmicos (progenitores neurales, NPC).
- Utilizar la técnica de microfluidos de 10X genomics para aislar las células y generar la biblioteca de cDNA etiquetado para realizar el scRNA seq de las condiciones H9 WT vs H9 *SIRT1*<sup>-/-</sup> antes y tras la diferenciación a NPC.
- Analizar la calidad de los datos obtenidos tras la secuenciación de scRNA-seq, así como el análisis posterior comparando las condiciones H9 WT vs H9 *SIRT1*<sup>-/-</sup> utilizando herramientas bioinformáticas dedicadas.
- Realizar un análisis de enriquecimiento con los genes diferencialmente expresados entre las condiciones WT y *SIRT1*<sup>-/-</sup>.
- Realizar un análisis de trayectoria durante el proceso de diferenciación de ambas condiciones para comparar los genes encontrados durante ambos pseudotiempo.



## 4. Metodología

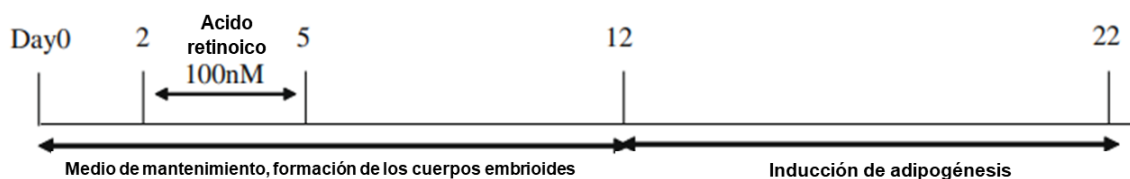
### 4.1. Cultivo celular de células H9, H9 SIRT1 -/- y H9 SIRT1 $\Delta$ Ex4

Se trabajó con las células troncales embrionarias humanas (hESCs) de la línea celular H9 (WA09, WiCell Research Institute Inc.), su cultivo y expansión se realizó sobre de una monocapa de Geltrex™ hESC-Qualified (A15696-01-50 mL, Gibco) que recubría el fondo de la superficie de cultivo, el medio de mantenimiento utilizado es el mTeSR1 (StemCell Technologies, # 05850) suplementado con mTeSR1 5X suplemento (StemCell Technologies #05851).

### 4.2. Protocolo de diferenciación hacia adipocitos.

Las hESCs H9 SIRT1 $\Delta$ Ex4 fueron separadas de la membrana de Geltrex™ utilizando Dispasa (07923, STEMCELL Technologies) y las colonias flotantes fueron recuperadas y resembradas en cajas de 6 pozos de baja adherencia (CORNING), el medio de mantenimiento utilizado se realizó con DMEM/F-12 (Biological Industries) suplementado con 20% de KSR (*Knock Out Serum Replacement*; 10828028, Gibco), 1% de aminoácidos no esenciales (01-340-1B, Biological Industries), 1mM de L-Glutamina (03-020-1B, Biological Industries), 0.1 mM de  $\beta$ -Mercaptoetanol (Sigma-Aldrich) y 1X de penicilina-streptomina (15140122, Gibco).

Del día 2 al día 5 del protocolo (Figura 14) se añadió una concentración de 100nM de Ácido retinoico (R2625, Sigma-Aldrich), tras 12 días de formación de los cuerpos embrioides (Figura 14) estos fueron transferidos a placas de 12 pozos cubiertas con Geltrex™ para lograr su adherencia a la placa, el medio fue sustituido por *MesenCult Adipogenic Differentiation Kit medium* (Stem Cell Technologies) para inducir la diferenciación de los cuerpos embrioides hacia adipocitos, se cambió el medio cada 2 días hasta alcanzar un total de 10 días.



**Figura 14:** Esquema temporal de protocolo de diferenciación de células troncales pluripotentes a adipocitos. Modificado de (Taura et al., 2009).

### 4.3. Tinción Rojo oleoso (ORO).

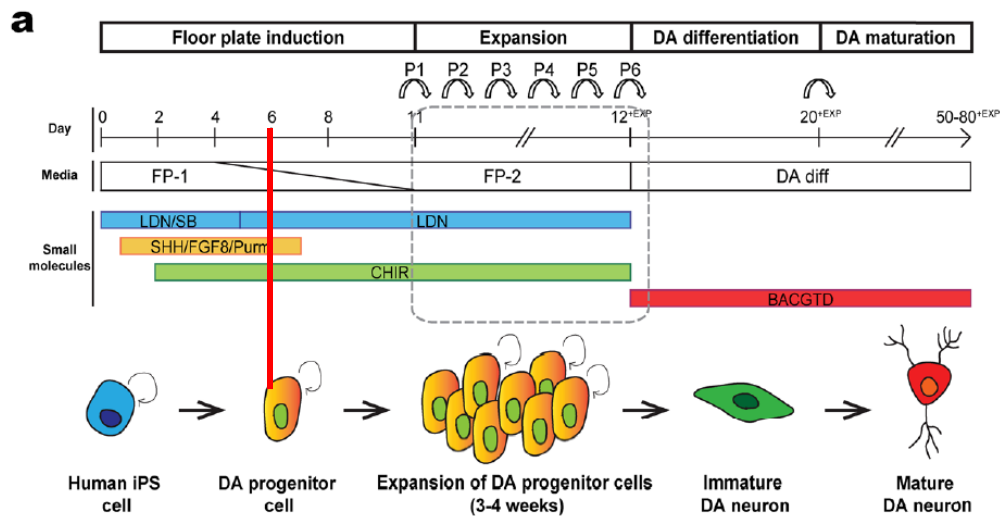
Las células crecidas en placas fueron fijadas y teñidas en el día 5 y 10 posterior a su transferencia a placas de 12 pozos recubiertas con Geltrex™, para la fijación se retiró el medio de los pozos y las células fueron lavadas con PBS 1X, posteriormente se agregaron 500 uL de paraformaldehído (PFA) fresco al 4% durante 45 minutos.

Para realizar la solución de tinción se utilizaron 1.5 partes de solución stock de ORO (01391, Sigma-Aldrich) con 1 parte de agua estéril, la solución fue filtrada utilizando un *cell strainer* de 40 uM (352340, Corning); pasados los 45 minutos se removió el PFA y se realizaron dos lavados con agua estéril, consecutivamente se agregaron 500 uL de isopropanol al 60% a cada pozo y se incubó a temperatura ambiente durante 5 minutos. Una vez terminada la incubación se retiró el isopropanol y se añadieron 500 uL de la solución de tinción ORO a cada pozo, se incubó la placa durante 30 minutos a temperatura ambiente.

Transcurridos los 30 minutos de tinción la solución fue retirada y se realizaron 5 lavados por pozo con agua para remover los precipitados de la solución, las fotografías fueron tomadas con la cámara Axiocam EEc 5s acoplada a un microscopio ZEISS Primovert usando un aumento 20x.

### 4.3. Protocolo de diferenciación hacia NPC

Para la obtención de NPC a partir de hESC se utilizó el protocolo escrito por (Fedele et al., 2017). Este protocolo consiste en tres etapas de diferenciación: fase de inducción o especificación, fase de expansión de los progenitores y una fase de diferenciación a dopaminérgicas (Figura 15).



**Figura 15:** Esquema temporal de protocolo de diferenciación de células troncales pluripotentes aun linaje dopaminérgico. La línea roja indica la etapa hasta la que se llevaron las células de la línea celular H9 y H9 SIRT1<sup>-/-</sup>. Modificado de Fedele S., et al, 2017).

Las células troncales embrionarias H9 y las células de la línea H9 SIRT1<sup>-/-</sup> fueron sembradas sobre Geltrex™ y en medio de cultivo mTeSR1 suplementado. Cuando las células alcanzaron un 70% de confluencia se inició el tratamiento con las moléculas pequeñas. Inicialmente del día 0 al día 5 se suplemento el medio celular con SB431542 (10 µM) junto con dorsomorfina, la molécula SB431542 se retiró en el día 5 mientras que la dorsomorfina se mantuvo en el medio durante los 7 días de diferenciación. En el día 1 del protocolo se suplemento con la molécula SAG 1µm, purmorfamina 2 µM y FGF8 0.1 µg/mL hasta el día 7. Llegado el día 3 del protocolo se agregó al medio la molécula CHIR99021 (3 µM) hasta el día 7. A partir del día 5 el medio de crecimiento fue retirado de forma gradual y reemplazado por medio N2 (17052048, Thermofisher) y 2 mM de Glutamax (Biological Industries) hasa el día 7 (Tabla 1).

09/05/22	10/05/22	11/05/22	12/05/22	13/05/22	14/05/22	15/05/22	16/05/22
<b>Día 0</b>	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Día 7
<b>SB 45</b>	SB 45	SB 45	SB 45	SB 45	SB 45		
<b>DorsoM</b>	DorsoM	DorsoM	DorsoM	DorsoM	DorsoM	DorsoM	DorsoM
			CHIR99	CHIR99	CHIR99	CHIR99	CHIR99
	SAG	SAG	SAG	SAG	SAG	SAG	SAG
	PURMO	PURMO	PURMO	PURMO	PURMO	PURMO	PURMO
	FGF8	FGF8	FGF8	FGF8	FGF8	FGF8	FGF8
<b>Mtesr1 medio</b>	Mtesr1 medio	Mtesr1 medio	Mtesr1 medio	Mtesr1 medio	Mtesr1 medio (75%) y medio N2*(25%)	Mtesr1 medio (75%) y medio N2*(25%)	Mtesr1 medio (50%) y medio N2*(50%)

**Tabla 1:** Protocolo de diferenciación a NPC. Se muestran las fechas de realización del experimento, así como las moléculas que se suplementaron a lo largo del protocolo durante 7 días.

### **4.3. Preparación de solución single cell de células H9 y H9 SIRT1<sup>-/-</sup>**

Las células fueron despegadas utilizando 1 ml de TrypLE™ Express (Cat. No. 12604021, Gibco) a cada uno de los pozos de caja de 6 pozos a despegar, se incubó durante 5 minutos a 37 °C, posteriormente se agregó 1 ml de medio mTeSR 1 y las células de recolectaron en un tubo falcon de 15 ml. Se llevó a cabo una centrifugación a 1000 rpm durante 4 minutos, terminando el proceso anterior se retiró el sobrenadante y se agregó 1 ml de PBS 1x para lavar las células y eliminar la mayor cantidad de residuos celulares posible, se centrifugo la muestra una vez más con los parámetros anteriores. Tras la centrifugación se retiró el sobrenadante y las células fueron resuspendidas en 500 µl de medio mTeSR 1, las muestras fueron entonces filtradas utilizando un tubo Falcon de 5 ml con cell strainer de 35 µm (Cat. No. 352235, Corning).

### **4.4. Protocolo single cell 10X Chromium**

Para obtener las librerías de nuestras muestras se siguió el protocolo *Chromium Next GEM Single Cell 3' Reagent Kits*, Número de documento CG000204 Rev D, 10x Genomics, (2019, Noviembre), que se explican en la figura 16 y 17.

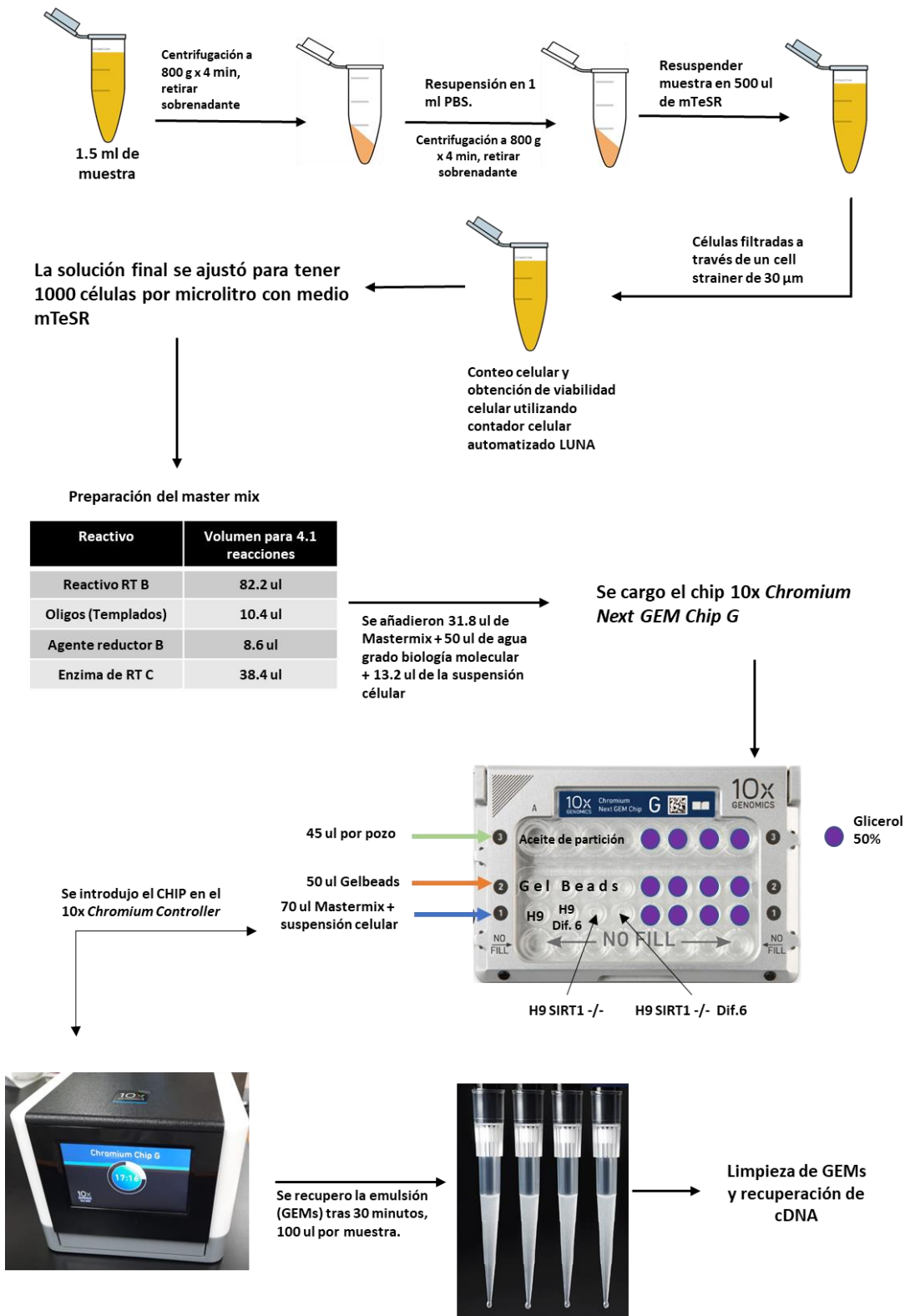


Figura 16: Esquema del protocolo Chromium Next GEM Single Cell 3' Reagent Kits; primera parte.

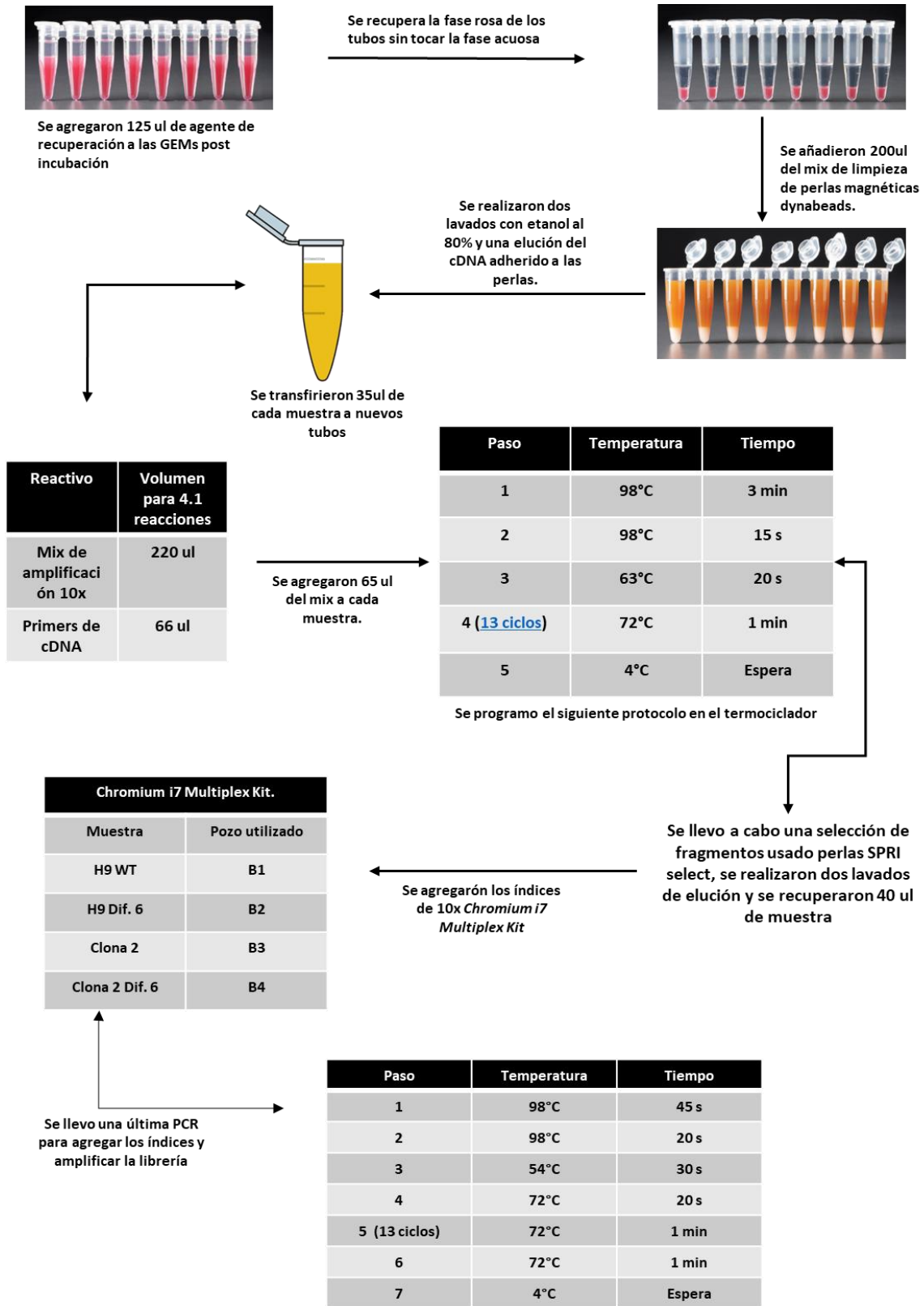


Figura 17: Esquema del protocolo Chromium Next GEM Single Cell 3' Reagent Kits; segunda parte.

#### 4.5 Creación de librerías y secuenciación.

Se realizó la construcción de la librería de cada muestra utilizando los índices *Chromium i7 Multiplex Kit* (PN-120262) de los pozos B1 a B4, asignando cada uno a una de las muestras, se cuantificó una vez más la concentración de DNA en cada muestra (Tabla 2).

Muestra ID Tubo	Pozo del kit (tomado de <i>Chromium i7 Multiplex Kit ; PN-120262</i> )	Secuencias de oligos	Volumen de muestra (ul)	Concentración
H9 SIRT1 <sup>-/-</sup> día 6	B4	ACTTCATA	59 ul	6.62 ng/ul
		GAGATGAC		
		TGCCGTGG		
		CTAGACCT		
H9 día 6	B2	TACTCTTC	59 ul	8.10 ng/ul
		CCTGTGCG		
		GGACACGT		
		ATGAGAAA		
H9 SIRT1 <sup>-/-</sup> día 0	B3	GTGTATTA	59 ul	3.92 ng/ul
		TGTGCGGG		
		ACCATAAC		
		CAACGCCT		
H9 día 0	B1	GTAATCTT	59 ul	4.77 ng/ul
		TCCGGAAG		
		AGTTCGGC		
		CAGCATCA		

**Tabla 2:** Índices asignados por muestra y concentración de las librerías obtenidas de cada condición

Las librerías se enviaron a secuenciar a la compañía Novogene Inc., donde se utilizó el secuenciador Novaseq6000 (Illumina) en modo 150 bp *paired end*, buscando una profundidad de 100GB por muestra.

#### 4.6. Análisis de datos scRNAseq

Los datos crudos fueron inicialmente procesados utilizando el pipeline de 10X Genomics llamado *Cellranger* en su versión 6.1.2. a través de la función *cellranger count*. Este pipeline alinea las lecturas contenidas en los archivos *fastq* crudos con un genoma de referencia elegido, y el output de esta función son archivos compatibles con otros softwares de análisis como los disponibles en R que se utilizaron en este análisis.

Se utilizó el genoma de referencia *GRCh38-2020-A* brindado y recomendado por la página 10X Genomic. La función *cellranger count* nos brinda también tres archivos por muestra que contienen los

datos crudos de las lecturas alineadas, estos archivos son:

**-Matrix.mtx .-** Este archivo es una matriz con el conteo de lecturas crudas de cada células dentro de la muestra

**-Barcodes.tsv.-** Un archivo con los identificadores de cada una de las células (correspondiente a la posición de cada una respecto a las columnas de la matrix.mtx).

**-Features.tsv.-** Contiene los nombres de los genes que fueron detectados dentro de la muestra (correspondiente a la posición de cada gen respecto a las filas de la matrix.mtx).

Estos tres archivos son esenciales para poder cargar los datos en paqueterías de análisis de scRNAseq en R(R Development Core Team, 2010).

#### **4.7. Importación en R y métodos de análisis: singleCellExperiment y Seurat.**

Se utilizaron dos metodologías para llevar a cabo el análisis de los datos producidos tras aplicar *cellranger count* a los datos crudos. La primera, denominada “singleCellExperiment”, consiste en un pipeline que utiliza varias de librerías para cada uno de los pasos del análisis, siendo su fundamento la creación de un tipo de objeto conocido como “singleCellExperiment” (Amezquita et al., 2020), este objeto es altamente interoperable lo que permite mucha libertad en la decisión acerca de qué herramienta elegir a lo largo de las etapas de análisis sin necesidad de convertir nuestro objeto.

Por otro lado Seurat (Satija et al., 2015) otorga un kit de herramientas contenido dentro de una sola librería en R que está diseñado para control de calidad, análisis y exploración de datos de scRNAseq, permite además llevar a cabo algunos análisis acoplados manteniendo un solo formato estándar a través de todo el análisis. Ambas metodologías se realizaron utilizando R 4.2.0 (R Development Core Team, 2010).

#### **4.8. singleCellExperiment**

##### **Importación de datos y control de calidad**

Se utilizó la librería *DropletUtils* (Lun et al., 2019) para importar los datos de las muestras con su función *read10XCounts*, esta función requiere la creación de una variable que contenga la dirección del directorio donde se almacenan los archivos Matrix.mtx, Barcodes.tsv y Features.tsv para cada una de las muestras.



Posteriormente se utilizó la librería *AnnotationHub* (Morgan & Shepherd, 2022) para anotar los genes detectados en nuestras muestras brindándole la información contenida en la librería *EnsDb.Hsapiens.v86* (Rainer, 2017), esta permite la creación de una conexión a una base de datos que contiene las anotaciones de los genes humanos en Ensembl hasta el día 19 de Mayo del 2017. Una vez anotados los genes en nuestras muestras se realizó la eliminación de las células detectadas que corresponden a ruido técnico causado principalmente por las droplets que no lograron capturar una célula, para esto se utilizó la función *emptyDrops* de la librería *DropletUtils*. El análisis de control de calidad se llevó a cabo con la función *perCellQCMetrics* contenida en la librería *scater* (McCarthy et al., 2017) que realiza el cálculo de controles de calidad de las células dentro de nuestra muestra, como el contenido de genes mitocondriales o células con baja variabilidad. Posteriormente, estas células fueron eliminadas de los datos a procesar con la función *isOutlier* de esta misma librería.

### **Normalización**

Se utilizó la librería *scran* (Lun, 2016) para realizar una normalización por deconvolución. Primero se utilizó la función *quickCluster* de esta librería para agrupar las células de acuerdo con su similitud en expresión y varianza en los valores de expresión, posteriormente se utiliza la función *computeSumFactors* que realiza una normalización a través de todos los clústeres previamente creados y finalmente se utiliza la función *logNormCounts* de la librería *scatter* para realizar una transformación logarítmica de los valores previamente normalizados en la matriz de cuentas de nuestros datos.

### **Modelado de varianza**

Para poder llevar a cabo la comparación entre las células dentro de las muestras es importante conservar aquellos genes que nos aportan la mayor cantidad de información y eliminar aquellos que contienen ruido aleatorio. Para cumplir lo anterior se buscaron aquellos genes que tienen la expresión más variable a través de todas las células con la función *modelGeneVarByPoisson* de la librería *scran*, esta función calcula la varianza y promedio de los valores de expresión normalizados y transformados logarítmicamente de cada gen, posteriormente ajusta una línea de tendencia a la varianza calculada de los genes contra el promedio de expresión y variación de una serie de cuentas simuladas utilizando una distribución de Poisson.

Tras realizar este modelado de la expresión y varianza de los genes de nuestros datos se utilizó la función *getTopHVGs* de la misma librería para obtener la lista de genes que cumplían con la característica de ser altamente variables.

## Reducción de dimensiones

Para poder llevar a cabo la reducción de dimensiones se utilizó la función *denoisePCA* de la librería *scrn* que realiza un análisis de componentes principales (PCA) y automatiza el proceso de eliminación de ruido técnico mediante la eliminación de los componentes calculados que dejan de aportar información importante sobre la estructura de los datos dentro de la muestra. Lo anterior basado en el concepto fundamental de un PCA que nos indica que los primeros componentes (casi siempre componente 1 y 2) capturan la mayor cantidad de información de nuestras muestras.

Se llevó a cabo también una reducción de dimensiones no lineal, en el caso de *singleCellExperiment* se utiliza la proyección llamada *TSNE* utilizando la función *runTSNE* de la librería *scater* que aplica el algoritmo *t-stochastic neighbour embedding* o *tSNE*, que busca una representación no lineal de los datos usando la mínima cantidad de dimensiones que preserven las distancias entre cada células y sus células vecinas en el espacio multidimensional original

## Clustering

Para determinar los clusters dentro de nuestros datos se utilizó la función *buildSNNGraph* de la librería *scrn* Esta función lleva a cabo la creación de un grafo que conecta las células de nuestra muestra a través del espacio multidimensional, posteriormente se asigna un valor o “peso” a las conexiones entre las distintas células, finalmente se identifican nodos o comunidades donde los pesos de las conexiones entre las células similares marcando los límites entre comunidades a través de determinar pesos más bajos asociados a comunidades alejadas, esta metodología es conocida como *Nearest Neighbors Tree* (Kobak & Berens, 2019).

## Anotación de identidad

Para poder identificar la identidad de las células dentro de nuestros datos se utilizó la librería *SingleR* (Aran et al., 2019) que contiene la función *SingleR*, está última requiere un set de datos previamente anotado para poder realizar una comparación entre nuestros datos y el set de referencia y así asignar una identidad celular a partir de una matriz de correlación entre ambos sets de datos.

El set de datos de referencia utilizado para anotar las muestras H9 y H9 *SIRT1<sup>-/-</sup>* día 0 fue obtenido a través de la función *HumanPrimaryCellAtlas* contenida en la librería *celldex* (Aran et al., 2019). Esta función genera un set de datos que contiene los valores de expresión de 713 microarreglos realizados en células progenitoras y troncales de humano publicada por (Mabbott et al., 2013).

Para anotar las identidades celulares de las muestras H9 y H9 *SIRT1<sup>-/-</sup>* día 6 se utilizó la librería *scRNAseq* (Risso & Cole, 2022) para obtener los datos de (La Manno et al., 2016) que contiene datos de scRNAseq con 25 identidades celulares en un modelo de desarrollo neuronal en humano. Estas

identidades se obtuvieron a través de la evaluación de líneas celulares durante procesos de diferenciación y cerebros humanos embrionarios.

### **Obtención de genes diferencialmente expresados**

Se utilizó la función *pseudoBulkDGE* de la librería *scrn* para llevar a cabo la expresión diferencial de los genes entre nuestros de grupos de muestras: H9 día 0 vs H9 *SIRT1*<sup>-/-</sup> día 0 y H9 día 6 vs H9 *SIRT1*<sup>-/-</sup> Día 6.

Esta función convierte los datos de scRNAseq en un output similar al obtenido al realizar un Bulk RNAseq tradicional, esto nos permite utilizar algoritmos de búsqueda de genes diferencialmente expresados diseñados para Bulk RNA seq, como son *limma* (Ritchie et al., 2015) y DESeq2 (Love et al., 2014).

## **4.9. Seurat**

### **Importación de datos y control de calidad.**

Para importar los datos crudos de nuestras muestras a la herramienta de análisis *Seurat* utilizamos la función *Read10X*, posteriormente utilizamos la función *CreateSeuratObject* dónde aplicamos el primer filtro a los datos crudos estableciendo el mínimo de *features* (Genes en nuestro caso particular) detectados por célula a 200 y un mínimo de expresión en 3 células por cada *feature* para determinar que es un *feature* válido.

Posteriormente utilizamos la función *PercentageFeatureSet* de la librería *Seurat* para poder obtener el porcentaje de genes detectados en cada una de las células de cada muestra que mapearon a mitocondria, posteriormente se utilizó la función *subset* para llevar a cabo la eliminación de células de baja calidad de acuerdo a los siguientes criterios:

**H9 día 0 .-** Se utilizaron dos conjuntos de criterios para determinar el umbral de buena calidad en estas muestras; el primero de ellos consideraba células que expresarán un mínimo de 200 genes y un máximo de 5000 con el límite de genes mapeados a mitocondria del 5%; por otro lado también se consideró como segundo conjunto aquellas células que tuvieran un mínimo de 2000 genes y un máximo de 5500 expresados con un valor de mapeo de mitocondria menor o igual al 5%.

**H9 *SIRT1*<sup>-/-</sup> día 0 .-** De igual forma se consideraron dos conjuntos de criterios para determinar el umbral de buena calidad en esta muestra; el primero consideraba células que expresarán un mínimo de 200 genes y un máximo de 4000 con el límite de genes mapeados a mitocondria del 5%; el segundo conjunto que incluyo células que tuvieran un mínimo de 1500 genes y un máximo de 5000 expresados con un valor de mapeo de genes a mitocondria menor o igual al 5%.

**H9 día 6 .-** Se consideraron aquellas células que tuvieron al menos 1000 y máximo de 5000 genes expresados con un porcentaje menor al 5% de estos mapeado a mitocondria.

**H9 SIRT1<sup>-/-</sup> día 6** .- Se consideraran aquellas células que tuvieran al menos 1500 y un máximo de 5500 genes expresados con un porcentaje menor al 5% de estos mapeados a mitocondria.

Los criterios de calidad anteriores se determinaron a partir de la observación de los gráficos de calidad obtenidos al utilizar la función *Vlnplot* e incluyendo tres criterios a graficar: Número de genes detectados (features) por célula, número de moléculas de RNA detectado por célula (Count RNA) y el porcentaje de genes mapeados a mitocondria.

### **Normalización**

Para poder normalizar los valores de expresión en nuestros datos se utilizó la función *NormalizeData* de la librería *Seurat* utilizando el método de normalización *LogNormalize* y un factor de escalamiento de 10,000. A grandes rasgos esta forma de normalización divide los valores de expresión (*counts*) de cada gen entre el número total de *counts* de la célula, multiplica este valor por 10,000 y aplica una transformación logarítmica al valor obtenido ( $\ln p + 1$ , donde  $p$  es el valor obtenido en *counts*).

### **Detección de genes altamente Variables**

Se utilizó la función *FindVariableFeatures* para poder encontrar aquellos genes cuya expresión variaba más de células a célula, la metodología de selección utilizada fue *vst*, esta última ajusta una línea a el valor de la varianza transformada logarítmicamente y al valor promedio de la expresión transformado logarítmicamente de cada gen usando una regresión polinomial local, posteriormente estandariza la expresión usando el promedio de expresión observado y la varianza esperada (de acuerdo a la línea ajustada), finalmente determina los genes más variables de acuerdo a los estándares calculados.

### **Escalamiento de datos**

Se utilizó la función *scaleData* para poder llevar un escalamiento de la expresión de los genes previo a la realización de la reducción de dimensiones, esto para evitar que los genes con mayor variabilidad resulten intrusivos en la conservación del patrón de genes con menor heterogeneidad en su expresión.

### **Reducción de dimensiones (PCA)**

El PCA se llevó a cabo con la función *RunPCA*, esta efectúa el análisis sobre los datos previamente escalados considerando como inputs los genes altamente variables obtenidos previamente.

### **Clustering**

Para llevar a cabo el clustering de nuestras células en el pipeline de *Seurat* se utilizó primero la función *FindNeighbors* que construye un grafo utilizando *nearest neighbor* calculando la distancia euclideana entre las células dentro de los componentes de nuestro PCA previamente calculado.

Posteriormente se utiliza la función *FindClusters* que utiliza el algoritmo Louviano, este algoritmo es iterativo y diseñado para la obtención de comunidades dentro de grandes redes de datos, se calcula el número de aristas (conexiones) dentro de las comunidades detectadas y se asigna un valor entre -1 a 1, esto se itera hasta obtener el mejor valor posible dentro de toda la red de células.

### **Reducción no-lineal de dimensiones: UMAP**

Se utilizó la función *RunUMAP* para poder visualizar los clústeres de nuestras muestras utilizando la aproximación y proyección.

### **Obtención de genes diferencialmente expresados.**

Para poder encontrar aquellos genes diferencialmente expresados entre la línea células H9 y H9 *SIRT1<sup>-/-</sup>* en las condiciones de día 0 y día 6 se utilizó el procedimiento de integración de datos propuesto en *Seurat*.

El proceso se inició con la adición de una columna indicadora de origen a cada muestra utilizando la función *AddMetadata*, esto nos permitió marcar cada una de las células como *WT* (línea H9) y la deletada (H9 *SIRT1<sup>-/-</sup>*), posteriormente se creó una lista donde se incluyeron ambas muestras con la función *list* de la paquetería base de R. Una vez creado el objeto donde se listaban los objetos que contenían los valores de expresión de las muestras *WT* y la deletada se usó la función *SelectIntegrationFeatures* que permite la selección de aquellos features que varían de forma consistente en ambas muestras, estos feautres son posteriormente introducidos en la función *FindIntegrationAnchors* que junto con la lista de objetos que representan nuestras muestras define punto de anclaje que se asemejan en nuestras muestras para finalmente realizar la integración de ambos objetos con la función *IntegrateData*, como paso final de la integración, se cambia el grupo de datos default a la matriz de cuentas crudas o *RNA assay* usando la función *DefaulAssay*.

Posterior a la creación del objeto integrado donde se incluyen nuestras muestras a comparar se realizan los mismos pasos anteriormente mencionados para la normalización, escalado de datos, búsqueda de genes altamente variables, reducción lineal con PCA, clustering y reducción de dimensiones no lineal UMAP.

Finalmente se utilizó la función *Findarkers* para encontrar los genes diferencialmente expresados, esta función ofrece varios algoritmos de búsqueda, en el caso de este trabajo se utilizó *MAST* (Finak et al., 2015) que es un algoritmo optimizado para la búsqueda de expresión diferencial en datos single cell.

### **Inferencia de actividad de Factores de Transcripción**

Para poder llevar a cabo el análisis de la actividad de factores de transcripción se utilizaron los objetos integrados creados en la sección de obtención de genes diferencialmente expresados (WT día 0 vs KO día 0; WT día 6 vs KO día 6), se obtuvo su matriz de cuentas normalizada y filtrada de ambos utilizando la función *GetAssayData* con el parámetro *slot* ajustado a *'counts'*, posteriormente este objeto fue convertido a la clase matriz utilizando la función *as.matrix* de R base, esta matriz fue almacenada en dos nuevos objetos llamados WTvsKO\_D0 y WTvsKO\_D6.

Posteriormente se utilizó la librería *DOROTHEA* para extraer la información de interacciones de TF almacenada en su base de datos, se creó una variable llamada *net* utilizando la función *get\_dorothea* con el parámetro de *organism* ajustado a *'human'* y extrajimos solo las interacciones con el mayor nivel de confianza ajustando el parámetro *levels* con la concatenación *c('A', 'B')*.

Una vez que se obtuvo la información de DOROTHEA se utilizó la función *runwmean* de la librería *DecoupleR* (Badia-i-Mompel et al., 2022) que requiere la matriz de cuentas de la muestra a analizar y un set de datos que incluya interacciones de TF con sus múltiples objetivos (la variable *net*), el objeto resultante es una lista de factores de transcripción y su nivel de actividad inferido en la muestra, este objeto es posteriormente sub seteado para obtener el marco de datos llamado *acts* que contiene la asignación de la actividad del TF con los respectivos clústeres, esto permite conocer las diferencias de actividad entre las poblaciones conformadas por células KO y WT.

Finalmente se utilizó la función *heatmap* de R base para la creación de los heatmaps que representan de forma visual la diferencia en el nivel de actividad de cada TF en las poblaciones de las muestras WT y KO en las condiciones Día 0 y Día 6.

#### **4.10 PHATE**

Se llevó a cabo la implementación del método de visualización de datos de alta dimensionalidad PHATE en los objetos integrados WTvsKO\_D0 y WTvsKO\_D6. Se utilizó la función *GetAssayData* con el parámetro *slot* ajustado a *'counts'* para obtener la matriz de cuentas de ambos objetos, se utilizó la misma función con el parámetro *slot* ajustado a *'metadata'* para crear un objeto que contuviera las identidades asignadas a cada célula en el paso de anotación de identidad. La metodología de PHATE corre en el lenguaje de programación Python por lo que las matrices de cuentas obtenidas y la metadata de anotación de identidad fueron exportadas a formato .csv utilizando la función *write.csv* de la paquetería R base.

Una vez creados los archivos .csv se importaron a un ambiente desarrollado en Jupyter Notebook (Kluyver et al., 2016) corriendo un kernel de Python 3.10, se utilizó la función *.io.load\_csv* del módulo *scprep* (Gigante et al., 2023) para poder crear una variable que contenga los datos de cuentas, de

igual forma se importaron las identidades celulares almacenadas en formato .csv con la función `.read_csv` del módulo *pandas* (Pandas, 2020), las variable creada en Python para la matriz de cuentas fue entonces procesada con la función `phate.fit_transform` y finalmente se creó un nuevo dataframe que contuviera las coordenadas de cada células en la visualiación de PHATE usando la función `.DataFrame` del módulo *pandas*.

Se plotearon los datos utilizando las coordenadas almacenada previamente y asignando el color de cada célula de acuerdo a su identidad almacenada en la metadata, los scater plot fueron creados utilizando la función `.plot.scatter2d` del módulo *scprep*.

#### **4.11.- Pseudotiempo de difusión.**

Se utilizó la implementación de pseudotiempo de difusión del módulo de Python *scanpy* (Wolf et al., 2018), para poder realizar esto convertimos la matriz de datos de cada muestra a un objeto de la clase *AnnotatedData* utilizando la función `.AnnData` del módulo *anndata* (Virshup et al., 2021), una vez creados los objetos establecimos la célula con los valores de coordenadas de PHATE más bajos como el punto inicial del pseudotiempo, posteriormente se realizó un PCA con la función `.pp.pca`, después se llevó a cabo un cálculo de vecinos cercanos (nearest neighbors, nn) usando la función `.pp.neighbors`, una vez obtenido esto se calculó el pseudotiempo utilizando la función `.tl.diffmap` seguida de la función `.tl.dpt`.

Los gráficos de pseudotiempo se crearon utilizando la función `.plot.scatter2d` y asignando el color de cada célula de acuerdo a su valor de pseudotiempo calculado.

#### **4.12.- Trayectoria Slingshot**

Se calculó la trayectoria dentro de la representación de PHATE utilizando la librería *Slingshot* (Street et al., 2018) en R. Los datos de coordenadas de PHATE fueron exportados al formato .csv usando la función `.to_csv` del módulo *pandas*, de igual forma se exportaron los nn asignados a cada célula usando la misma función.

Los datos fueron posteriormente importados en R utilizando la función `read.csv` de R base, posteriormente se obtuvo el nombre de los genes altamente variables usando los siguientes argumentos de subset `nombredelobjeto@assays$RNA@var.features` para llevar a cabo el análisis de tendencia de expresión. Se utilizarón las coordenadas de PHATE importadas en la función *slingshot* asignando las etiquetas de nn obtenidas en PHATE al parámetro de `cluster_labels`.

Una vez obtenido el objeto *Pseudotimeordering* de *Slingshot* procedimos a utilizar *tradeSeq* para comprobar si la expresión de los genes altamente variables previamente filtrados respondía de alguna

forma a la trayectoria en dirección al pseudotiempo de difusión calculado, se utilizó la función *fitGAM* de esta librería usando los parámetros obtenidos en el objeto de clase *Pseudotimeordering*.

#### **4.13.- *tradeSeq* y genes oscilantes a lo largo de la trayectoria.**

Para poder identificar aquellos genes que oscilan o sufren cambios en su expresión en relación con el avance del proceso de diferenciación se utilizó la librería *tradeSeq* (Van den Berge et al., 2020). Previo a su uso se fusionaron las matrices de cuentas crudas de ambas condiciones obtenidas previamente en la realización del análisis *PHATE* utilizando la función *merge* de *Seurat*. El valor de pseudotiempo se obtuvo mediante la fusión de las matrices de pseudotiempo de difusión obtenidas en la sección 4.11, y se asignaron dos linajes: el linaje 1 correspondiente a la identidad H9 y el linaje 2 a la condición H9 *SIRT1*<sup>-/-</sup>. Finalmente se creó una matriz binaria de dos columnas denominada *Cellweights* donde se asignó cada una de las células de la matriz a su respectivo linaje.

Una vez obtenidos todas las matrices anteriores se procedió a utilizar la función *fitGAM* de la librería *tradeSeq*, se utilizó el argumento *counts* en el que se asignó la matriz de cuentas fusionadas previamente creada, en el argumento *pseudotime* se introdujo la matriz de los pseudotiempo fusionados de ambas condiciones, en el argumento *cellWeights* se introdujo la matriz binaria denominada *Cellweights*, debido a que este proceso es muy exhasutivo y requiere una gran cantidad de recursos computacionales se utilizó computación paralela mediante el argumento *BPPARAM* de esta función, en esta se asignaron un total de 12 '*workers*' (núcleos de procesador). El output de este análisis es un objeto de tipo *singleCellExperiment* que se usa en los análisis posteriores

El objeto *singleCellExperiment* es sometido a una prueba de asociación para saber si el modelo aditivo general calculado para cada gen a través de *fitGAM* se relacionada con el avance del pseudotiempo en ambos linajes. Para esto se usó la función *associationTest* de la librería *tradeSeq*, donde se usó el argumento *lineages* con el identificador lógico *TRUE* para que las comparaciones se hicieran entre ambos linajes. De igual forma se ajustó el valor del argumento *l2fc* en 1 para que se buscara asociación de aquellos genes cuyo cambio a lo largo de la trayectoria fuera el doble o la mitad en alguno de los puntos.

Posteriormente se utilizó la función *patternTest* de la misma librería para poder encontrar los genes que tenían un patrón de expresión distinto a lo largo del pseudotiempo de diferenciación, de igual forma se ajustó el argumento *l2fc* a 1 para poder obtener una lista de los genes con la mayor diferencia en su patrón de expresión entre ambos linajes.



Se realizó también un análisis de aquellos genes en la etapa final del pseudotiempo (genes marcadores tardíos) utilizando la función *diffEndTest*, de forma similar se ajustó el valor del argumento *l2fc* a 1 para obtener sólo aquellos genes que poseyeran un cambio significativo al final de esta etapa.

## 5. Resultados

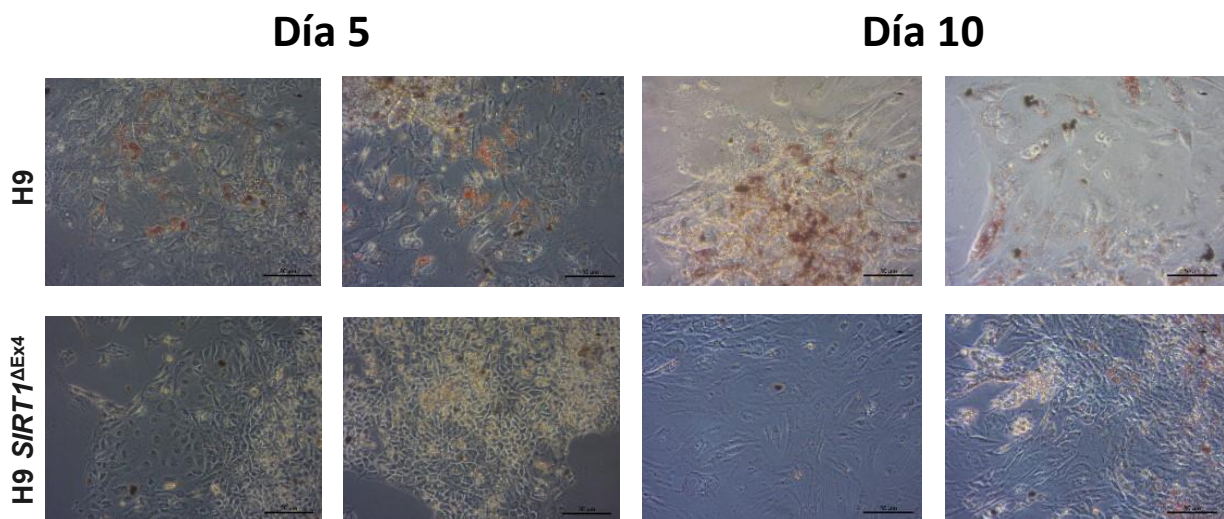
### 5.1. Diferenciación de línea celular H9 y H9 *SIRT1*<sup>ΔEx4</sup> hacia linaje adipogénico

Con la finalidad de evaluar la capacidad de la línea H9 WT y la línea H9 *SIRT1*<sup>ΔEx4</sup> se realizó un protocolo diferenciación hacia adipocitos del linaje mesenquimal como se describe en la sección de métodos. Para poder realizar esta evaluación se utilizó una tinción ORO que tiñe de rojo brillante los lípidos neutros acumulados dentro de las células, que es una característica definitoria del linaje celular de los adipocitos. Las tinciones se realizaron los días 5 y 10 posteriores al proceso de inducción a linaje adipogénico.

En el día 5 se pudo observar que las células WT comienzan a mostrar acumulación de lípidos en las periferias de los cuerpos embrioides inducidos (Figura 18) caso contrario a las células H9 *SIRT1*<sup>ΔEx4</sup> donde las células no muestran ninguna acumulación de lípidos intracelular, lo anterior nos lleva a pensar que en primera instancia la respuesta al medio de inducción no es llevada a cabo en estas células.

Se observó el mismo comportamiento en el día 10 de inducción (Figura 18), las células H9 mostraron una gran cantidad de células teñidas lo que nos permite inferir que se logró diferenciarlas hacia adipocitos, la condición H9 *SIRT1*<sup>ΔEx4</sup> mostró una cantidad baja de células teñidas por lo que aun cuando se completó el protocolo de diferenciación estas no lograron responder al estímulo de inducción a linaje adiposo.

Este primer intento de diferenciación nos mostró que SIRT1 parece esencial durante procesos de diferenciación en las hESCs hacia un linaje mesodérmico como son los adipocitos, la ausencia de esta deacetilasa genera células troncales incapaces de responder a un estímulo de diferenciación y determinación de linaje.



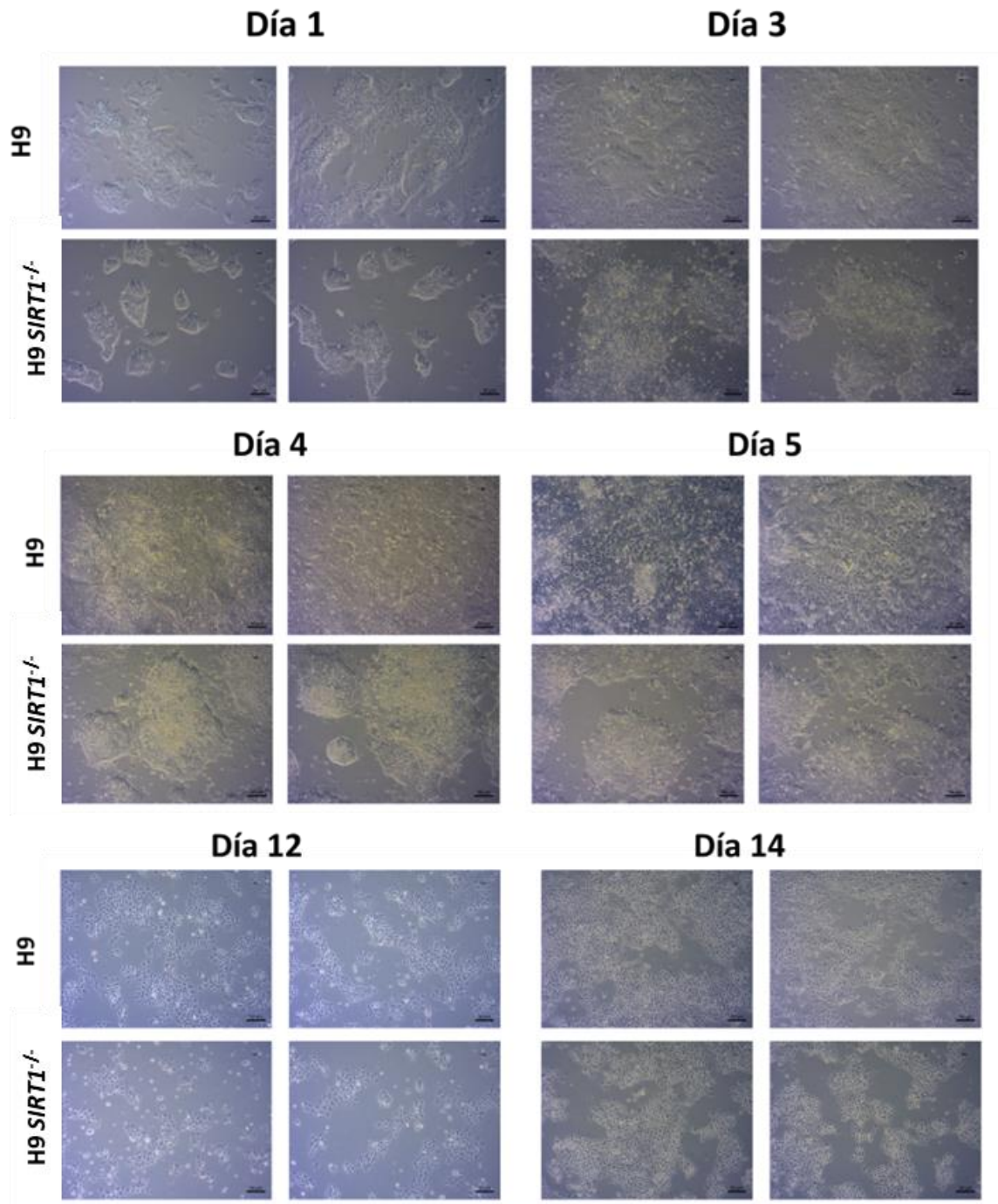
**Figura 18:** Fotografías representativas (20X) de las células inducidas con medio de diferenciación adipogénico en el día 5 y 10 del protocolo de diferenciación a adipocitos en las células H9 WT y la Clona 8 SIRT1<sup>-/-</sup>.

## 5.2. Diferenciación de línea celular H9 y H9 SIRT1<sup>-/-</sup> hacia NPC

Con el propósito de evaluar la capacidad de diferenciación de la línea celular H9 WT y la H9 SIRT1<sup>-/-</sup> en un segundo linaje embrionario (ectodérmico) se realizó la diferenciación de ambas hacia neuronas de linaje dopaminérgico como se indica en la sección de métodos. Se observó que la H9 SIRT1<sup>-/-</sup> continúa presentando una morfología colonial a pesar de las moléculas añadidas al medio en el día 1. Por su parte las células H9 control exhibieron una alta proliferación cubriendo por completo la superficie del pozo en el día 3, la línea H9 SIRT1<sup>-/-</sup> en contraste presenta menor confluencia y aún conserva su morfología colonial (Figura 19).

En los días 4 y 5 del protocolo de diferenciación la línea H9 SIRT1<sup>-/-</sup> continua mostrando una proliferación menor respecto al control de células H9, además de lo anterior las células H9 SIRT1<sup>-/-</sup> muestra morfología colonial en algunas zonas del pozo mientras que la condición control presenta células con una morfología alargada (Figura 19).

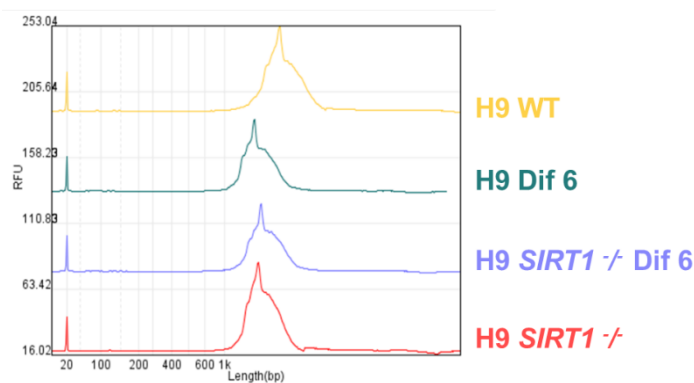
Posteriormente en el día 11 del protocolo de diferenciación se realizó un pase de las células de cada condición para continuarla fase de expansión, al día siguiente la confluencia de la H9 SIRT1<sup>-/-</sup> era menos a las células control H9, sumado a eso las células H9 SIRT1<sup>-/-</sup> presentaron una morfología distinta a la condición control durante esa fase del protocolo y las células del pozo murieron en el día 15 (Figura 19). Lo anterior nos permite inferir que la línea H9 SIRT1<sup>-/-</sup> no tenía la capacidad de sobrevivir más allá de la etapa de progenitores neurales dopaminérgicos (NPC) haciendo este estadio el candidato en comparativa para la realización de nuestro scRNAseq.



**Figura 19:** Fotografías representativas (10X) de la morfología celular observada en los días 1,3,4,5,12 y 14 del protocolo de diferenciación a neuronas dopaminérgicas en las células H9 control y las H9 SIRT1<sup>-/-</sup>.

### 5.3. scRNAseq 10X Chromium

Se realizó el protocolo de scRNAseq de la empresa 10x Chromium para poder obtener los niveles de los genes transcritos en 4 condiciones experimentales: 1.- Células H9 WT (H9 WT); 2.- Células H9 diferenciadas a NPC en el día 7 (H9 Dif. 6); 3.- Células H9 *SIRT1*<sup>-/-</sup>; 4.- Células H9 *SIRT1*<sup>-/-</sup> diferenciadas a NPC en el día 7. Se obtuvo el cDNA de cada condición tras la retrotranscripción (RT) de los mRNA capturados en cada muestra usando el equipo Qsep1 Lite de BioOptic.Inc (Figura 20) y la concentración obtenida a través de Qubit (Tabla 2).

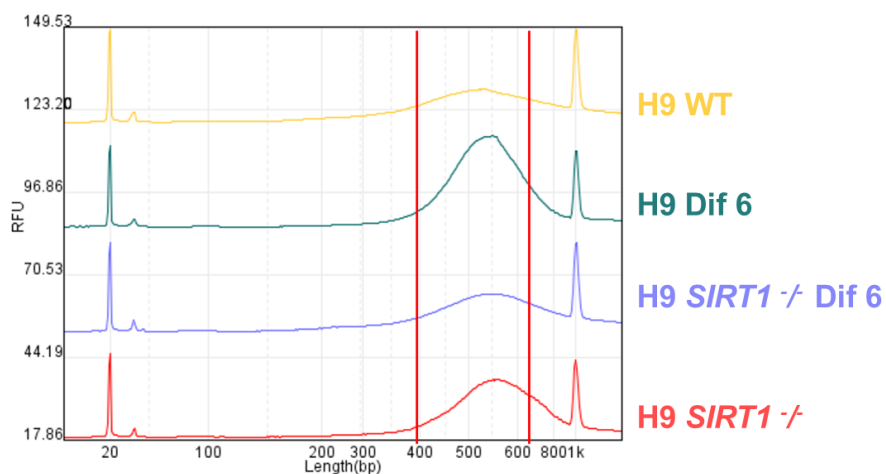


**Figura 20:** Electroferogramas de cDNA obtenido de cada muestra tras la RT de los mRNA capturados, generados con Qsep1 Lite

Muestra	cDNA cuantificado por Qbit (ng/ul)
H9 WT	50
H9 Dif. 6	48.2
Clona 2	48.5
Clona 2 Dif. 6	44.4

**Tabla 2:** Concentración del cDNA obtenido tras la RT de los mRNA capturados de cada muestra

Por último, se obtuvieron los electroferogramas utilizando el equipo Qsep1 Lite para identificar los tamaños de fragmentos en cada una de las librerías construidas, el tamaño óptimo de acuerdo con 10X genomics es de fragmentos que van de los 300 a 1000 pares de bases (bp) con un pico de abundancia entre los 550 y 425 bp. Las librerías obtenidas de todas las muestras cumplieron con lo especificado anteriormente (Figura 21).



**Figura 21:** Electroferogramas de las librerías obtenido de cada muestra. La línea roja muestra el área cubierta por la distribución de abundancia del tamaño de los fragmentos obtenidos. Electroferograma obtenida utilizando Qsep1 Lite.

#### 5.4. Obtención de datos crudos scRNAseq

Se obtuvieron los datos crudos por parte de Novogene de las 4 muestras, todas las muestras pasaron los análisis de control de calidad y la cantidad de lecturas, dimensión en gigabytes y score de calidad se reportan en la siguiente tabla:

Muestra	Lecturas crudas	Dimensión (GB)	Score Q20 (%)	Score Q30 (%)
H9	648555112	97.3	85	77.08
H9 <i>SIRT1</i> <sup>-/-</sup>	585348082	87.8	84.20	73.15
H9 Día 6	480817274	72.1	83.23	75.50
H9 <i>SIRT1</i> <sup>-/-</sup> Día 6	618321236	92.7	84.53	76.54

**Tabla 3:** Información de datos crudos obtenidos tras secuenciación.

#### 5.5. Reportes de calidad Cellranger

Se obtuvieron los siguientes valores tras correr el pipeline automatizado *cellranger* en los fastq de nuestras muestras:

Muestra	No. De células	Reads promedio por células	Genes promedio por célula	Porcentaje de alineamiento al genoma
H9	5971	54309	3906	91.6%
H9 <i>SIRT1</i> <sup>-/-</sup>	7911	36996	3022	91,7%
H9 día 6	5874	40928	2604	94.6%
H9 <i>SIRT1</i> <sup>-/-</sup> día 6	5030	66131	3742	95.5%

**Tabla 4:** Reportes de calidad de cellranger 10X. Valores de calidad obtenidos tras realizar el alineamiento utilizando cellranger.

## 5.6 Análisis utilizando *SingleCellExperiment*.

### 5.6.1. Análisis de calidad

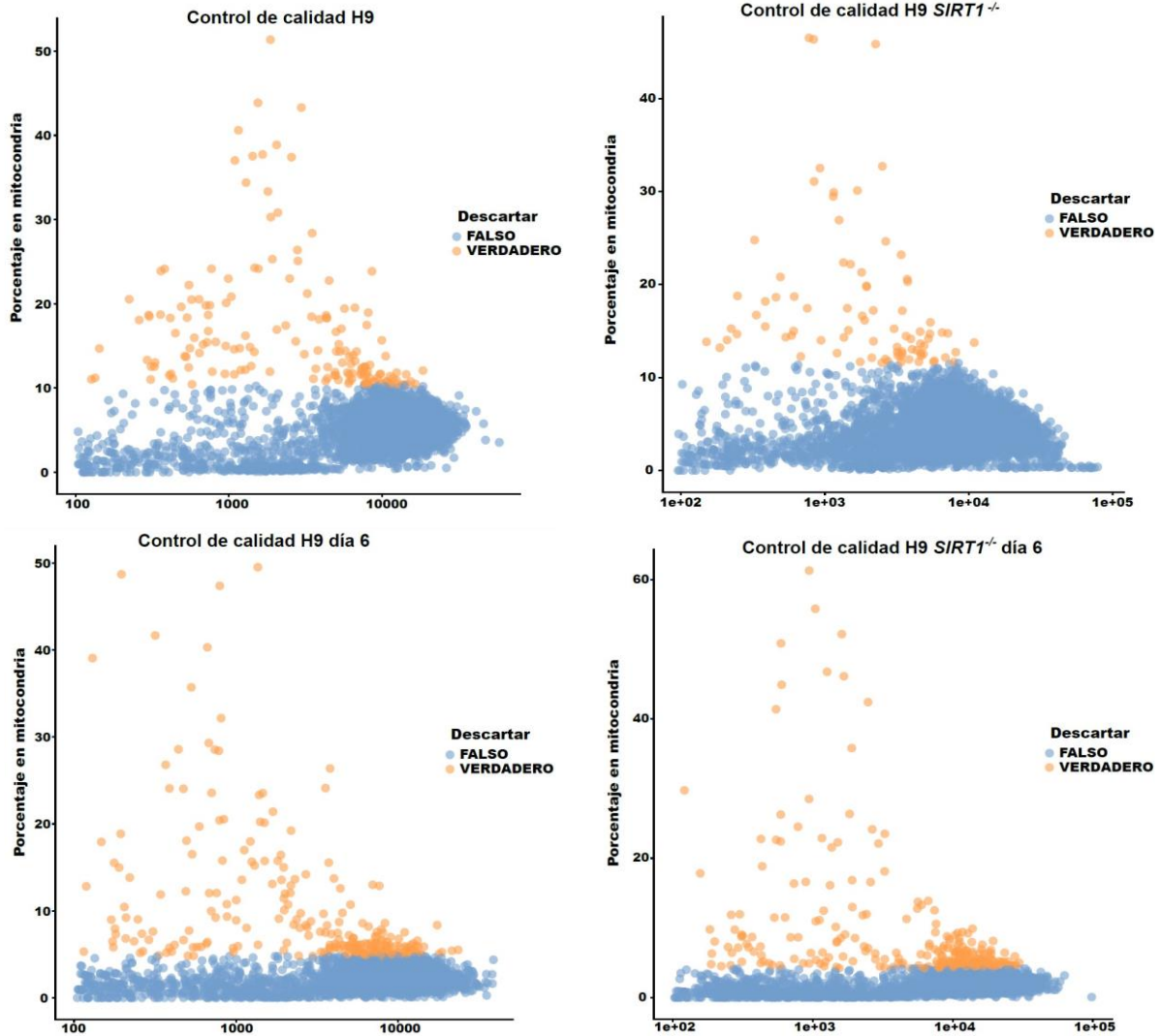
Se realizó el control de calidad de las 4 muestras (tras llevar a cabo la eliminación de las células vacías o que representaban ruido técnico en cada una), para eliminar aquellas células que tuvieran baja calidad se utilizó el porcentaje de genes mapeados a mitocondria, en el caso de la muestra H9 día 0 se obtuvieron 6271 células tras el primer filtro, de las cuales 161 fueron eliminada por su contenido mitocondrial (Figura 22 A, puntos naranjas), quedando con un total de 6110 células con buena calidad para los análisis posteriores.

Para los datos de H9 *SIRT1*<sup>-/-</sup> día 0 se obtuvieron 8432 células tras el primer filtro, se encontraron 83 células con alto contenido mitocondrial (Figura 22 B, puntos naranjas), dejando 8349 células con buena calidad para los análisis posteriores.

En la muestra H9 día 6 se obtuvieron 6255 células tras el primer filtro, posteriormente se eliminaron 259 células por su alto contenido de genes mitocondriales (Figura 22 C, puntos naranjas), dejando 5996 células para los análisis posteriores.

Por último, en la muestra H9 *SIRT1*<sup>-/-</sup> día 6 se obtuvieron 5314 células tras aplicar el primer filtro de las cuales 284 fueron eliminadas por su alto contenido mitocondrial (Figura 22 D, puntos naranjas), dejando 5030 células para los análisis posteriores.



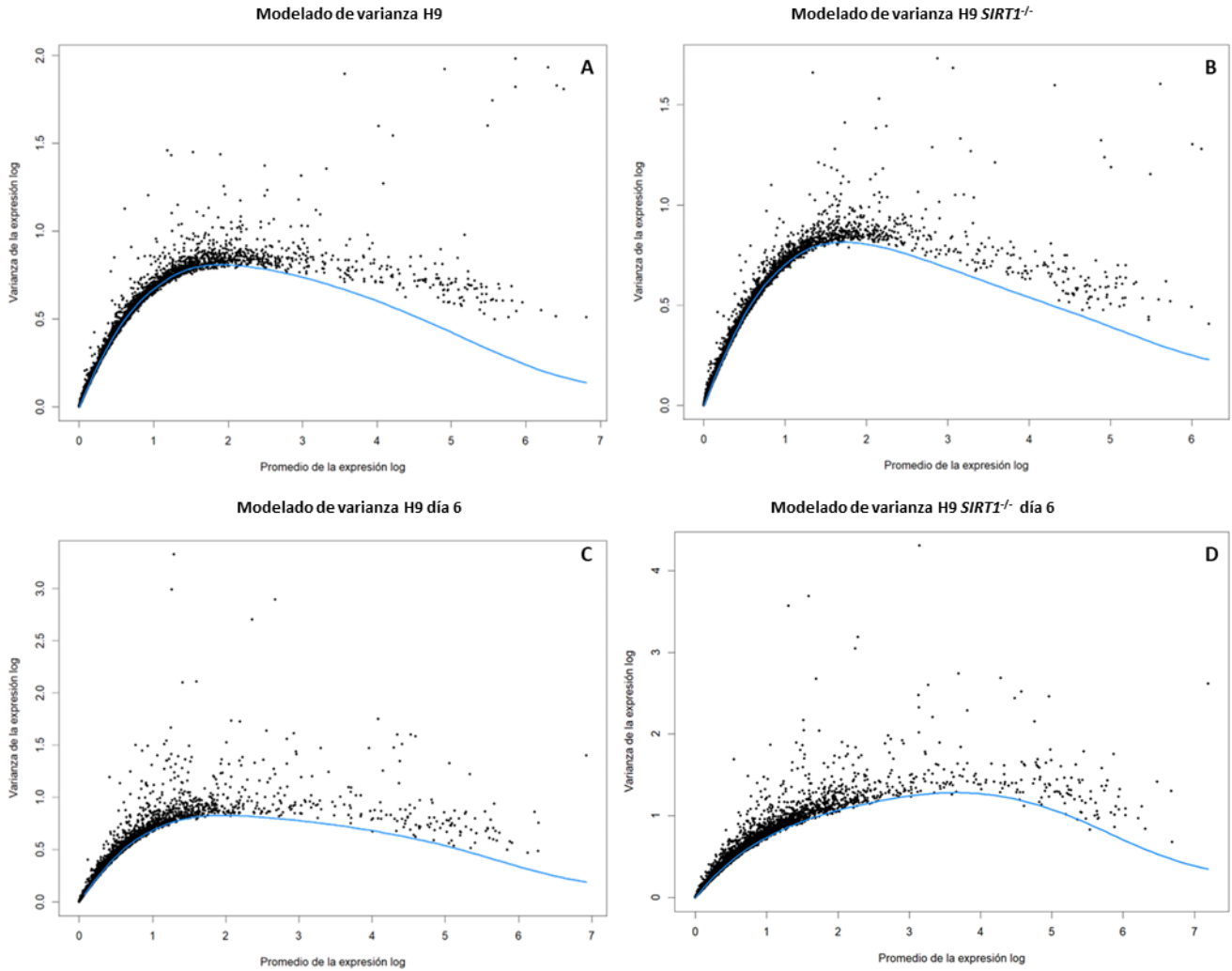


**Figura 22:** Gráfica de control de calidad. Se muestran las 4 muestras analizadas (A) H9 día 0; (B) H9 SIRT1<sup>-/-</sup> día 0; (C) H9 día 6; (D) H9 SIRT1<sup>-/-</sup> día 6, se representan las células que tienen una baja calidad debido a su alto contenido de genes mitocondriales (naranja) y aquellas que pasaron los filtros de calidad y poseen un bajo contenido de genes mitocondriales (azul)

### 5.6.2. Modelado de varianza.

Se utilizó el modelado de varianza por distribución de Poisson en las 4 muestras con la finalidad de eliminar el sesgo causado por los genes altamente variables y por la gran cantidad de valores iguales o cercanos a cero en las cuentas crudas de cada muestra, se muestran las gráficas de punto de la muestra H9 día 0 (Figura 23 A), H9 SIRT1<sup>-/-</sup> día 0 (Figura 23 B), H9 día 6 (Figura 23 C) y H9 SIRT1<sup>-/-</sup> día 6 (Figura 23 D), la línea azul muestra la tendencia ajustada a los valores de expresión y varianza de los genes detectados tras utilizar la distribución de Poisson.





**Figura 23:** Modelado de varianza. Se muestran los análisis y modelado de varianza de las cuatro muestras analizadas (A) H9 día 0; (B) H9 SIRT1<sup>-/-</sup> día 0; (C) H9 día 6; (D) H9 SIRT1<sup>-/-</sup> día 6, cada punto negro representa un gen detectado en las muestras, el eje X representa el promedio de expresión a través de todas las células transformado logarítmicamente, el eje Y representa la varianza de expresión a través de todas las células de la muestra, por último, la línea azul representa el ajuste de varianza obtenido tras aplicar una distribución de Poisson.

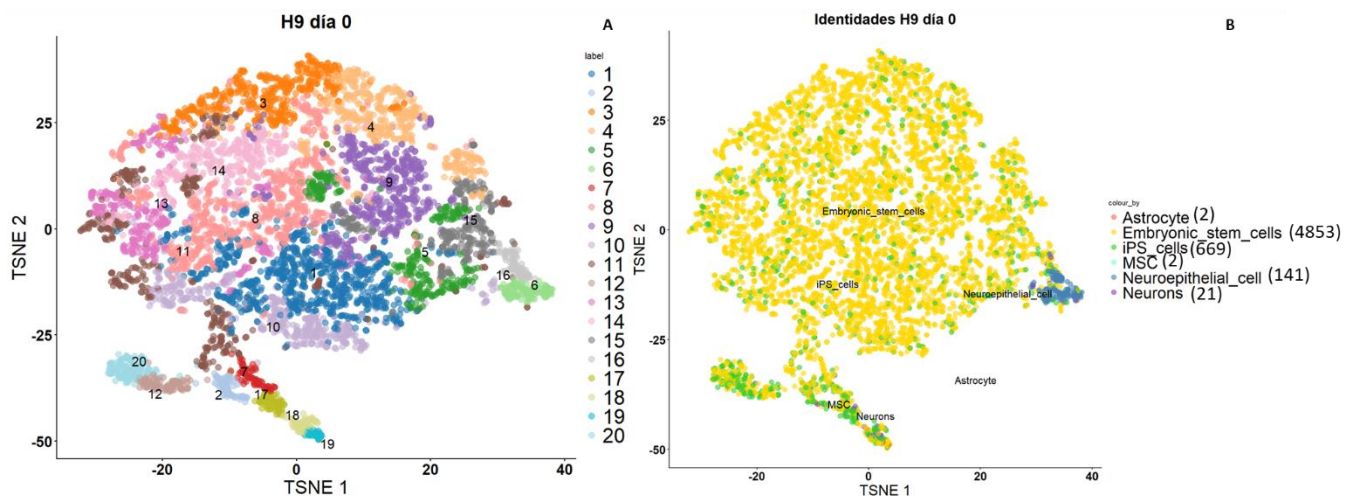
### 5.6.3 Representación tSNE y Anotación de identidad

Se utilizó la representación *t-Distributed Stochastic Neighbor Embedding* (t-SNE) para representar las células de cada muestra y observar los clusters formados tras realizar el clustering basado en grafos de *KNN* (*K Nearest-Neighbors*), posteriormente se utilizó la librería *SingleR* para determinar la identidad celular dentro de las muestras.

En el caso de las células H9 día 0 se obtuvieron un total de 20 clusters (Figura 24 A) con una forma homogénea a lo largo y ancho de la población en general, se observaron algunos clusters no tan

segregados o definidos como los clusters 11, 1 y 4.

Posteriormente se realizó la asignación de identidades celulares utilizando la información del HPCA, se obtuvieron un total de 6 identidades celulares (Figura 24 B), de las cuales *Embryonic\_Stem\_Cells* y *iPS\_cells* fueron las identidades con un mayor número de células con 4853 y 669 respectivamente (Figura 22 B), también se detectaron identidades relacionadas con linajes neuronales como son *Astrocyte*, *Neuroepithelial\_cell* y *Neurons* (Figura 24 B).

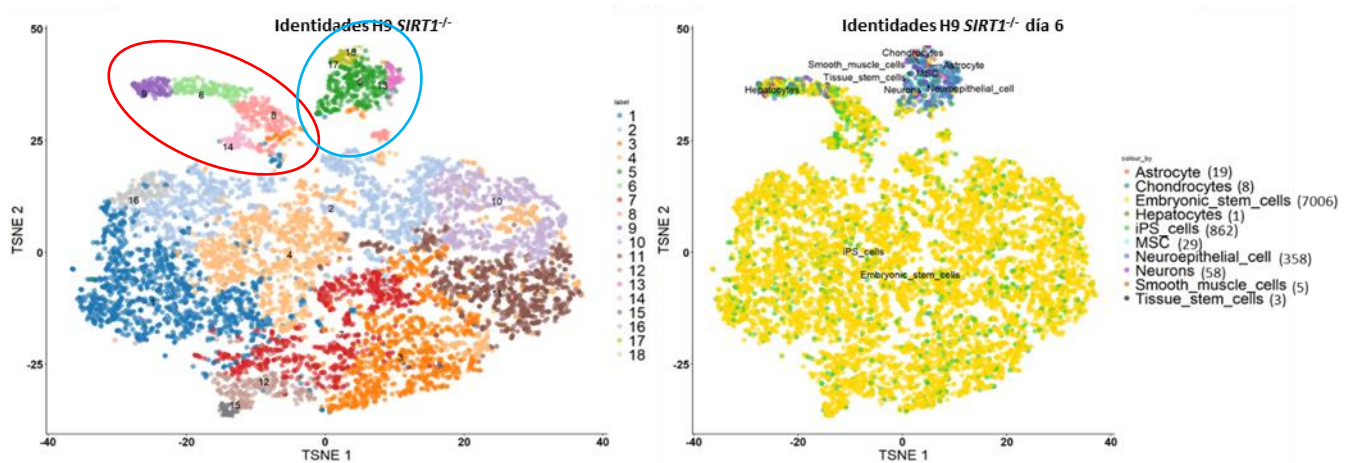


**Figura 24:** Representación tSNE de células H9 día 0. (A) Se muestran los 20 clusters presentes en la muestra H9 día 0 con algunos clusters bien definidos (20, 12, 7, 17, 18, 19) y algunos difusos y menos segregados (11, 1 y 14); (B) Se observan las 6 identidades presentes en las células H9 día 0, Astrocitos (Rojos), ESCs (Amarillo), células iPS (verde), MSC (azul claro), Neuroepitelio (Azul oscuro) y Neuronas (Morado).

En el caso de las células de la H9 *SIRT1* día 0 se obtuvieron un total de 18 clusters (Figura 25 A), se puede observar que los clusters 9, 6, 14 y 8 forman un grupo separado del conjunto de clusters más grande (Figura 25 A, círculo rojo), compuesto por los clusters 15, 12, 7, 3, 11, 16, 4, 2 y 10, de igual forma los clusters 17, 18, 13 y 5 forman otro conjunto de clusters distinto a los dos anteriores (Figura 25 A, círculo azul).

La anotación de identidad en estas células se llevó a cabo con la misma referencia utilizada en la muestra H9 día 0 el HPCA, en contraste con las células H9 se obtuvieron nuevas identidades celulares (Figura 25 B) entre las que podemos observar *Chondrocytes*, *Hepatocytes*, *Smooth\_muscle\_cells* y *Tissue\_stem\_cells*.

La aparición de identidades relacionadas con linaje mesodérmico es un primer indicio que refuerza la hipótesis de que la delección de SIRT1 en H9 *SIRT1*<sup>-/-</sup> genera heterogeneidad poblacional dentro de las células de esta muestra, finalmente se procedió a anotar las identidades celulares en el tSNE de la Clon 2 día 0 (Figura 25 B).



**Figura 25:** Representación tSNE de células H9 *SIRT1*<sup>-/-</sup> día 0. (A) Representación tSNE de células H9 *SIRT1*<sup>-/-</sup> día 0. Se muestran los 18 clusters presentes en la muestra H9 *SIRT1*<sup>-/-</sup> día 0, se señalan dos grupos de clusters separados del conjunto de clusters principales (círculo rojo y azul); (B) Se observan las 10 identidades presentes en las células H9 *SIRT1*<sup>-/-</sup> día 0, Astrocitos (Rojos), Condrocitos (Verde oscuro), ESCs (Amarillo), Hepatocitos (Olive), células iPS (verde), MSC (azul claro), Neuroepitelio (Azul oscuro), Neuronas (Morado), Musculo liso (Marrón) y Células troncales de tejido (Negro).

En el caso de las muestras H9 y H9 *SIRT1*<sup>-/-</sup> día 6 se eligió un set de referencia distinto para asignar identidad, se hizo una revisión de los datos disponibles y se eligió el set de (La Manno et al., 2016), este set de datos contiene etiquetas creadas a partir de scRNAseqs realizados en líneas células troncales diferenciadas a linajes neuronales y en cerebros humanos embrionarios.

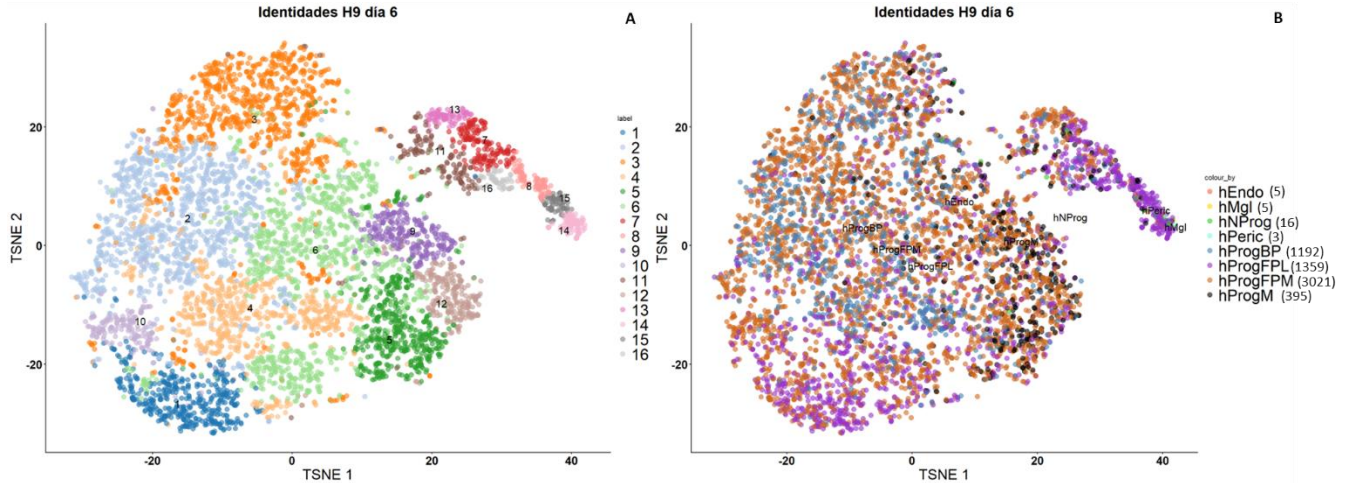
En el caso de las células H9 día 6 se obtuvieron 16 clústeres observables en el tSNE (Figura 26 A), se pueden observar dos grupos de clústeres, uno de ellos compuesto por los clústeres 14, 15, 8, 16, 7, 11 y 13 mientras que los clústeres restantes forman un segundo grupo separado.

Al anotar la identidad de estas células se obtuvieron 8 identidades (Figura 26 B): hEndo (células endoteliales), hMgl (microglía), hNProg (progenitores neurales), hPeric (pericitos), hProgBP (progenitores de plato basal), hProgFPL (progenitores de piso lateral basal), hProgFPM (progenitores de piso medio basal) y hProgM (progenitores de línea media).

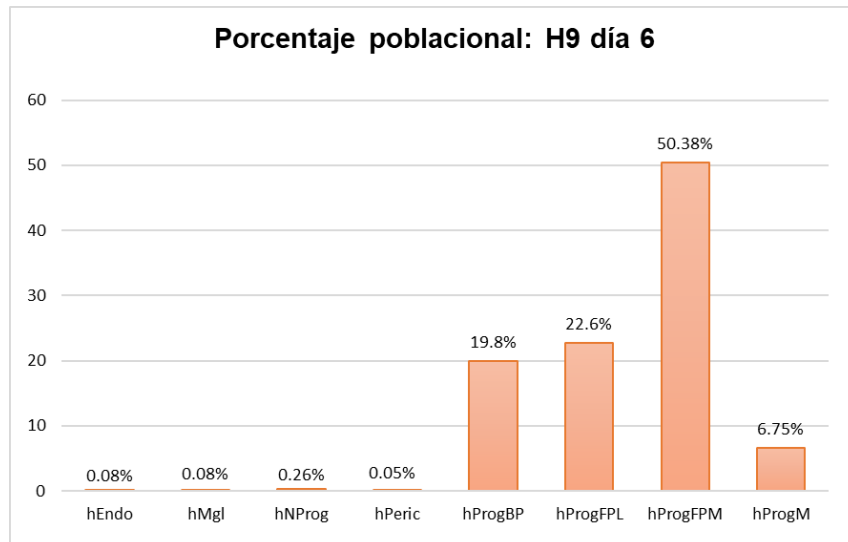
Al analizar la composición poblacional (Figura 27) observamos que los hProgFPM representan el 50.38% del total (3021 células), lo anterior tiene sentido debido a que el protocolo de diferenciación

hacia progenitores neurales de (Fedele et al., 2017) que se utilizó en estas muestras de día 6 compromete a las células a un linaje proveniente del piso medio basal; la segunda identidad celular más presente son los hProgFPL que representan el 22.66% (1359 células) seguido por la identidad hProgBP representando el 19.8% (1192 células).

**Figura 26:** Representación tSNE de células H9 día 6. (A): Representación tSNE de células H9 día 6. Se muestran



los 16 clusters presentes en la muestra H9 día 6.; (B) Se observan las 6 identidades presentes en las células H9 día 6, hEndo (Células endoteliales; Rojos), hMgl (Microglia; Amarillo), hNProg (Progenitores Neurales; verde), hPeric (Pericitos; azul claro), hProgBP (Progenitores plato basal; Azul oscuro), hProgFPL (Progenitores piso lateral; Morado), hProgFPM (Progenitores piso medio basal; Marrón) y hProgM (Progenitores línea media; Negro).



**Figura 27:** Porcentaje de la población asignado a cada identidad del set de referencia de (La Manno et al., 2016) en la condición H9 día 6.



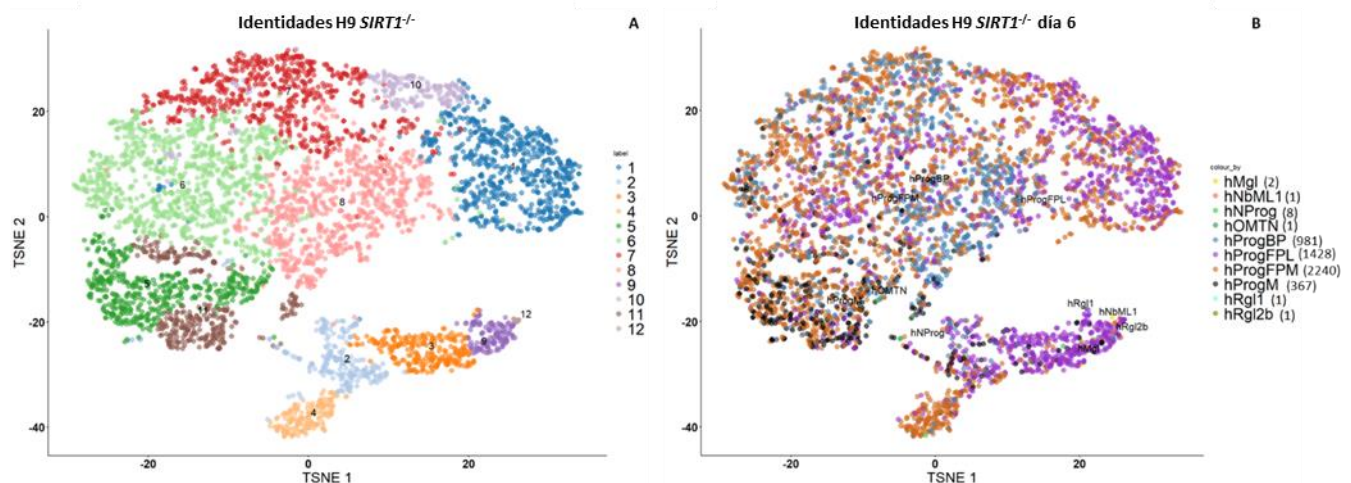
Por último, en las células H9 *SIRT1*<sup>-/-</sup> día 6 encontramos 16 clústeres al realizar la proyección del tSNE, se pueden observar dos grupos de clústeres, el más pequeño de ellos conformado por los clústers 4,2,3,9 y 12, los restantes se agrupan en el conjunto de mayor tamaño (Figura 28 A).

Posteriormente, la anotación de identidad celular encontró 10 identidades en las células H9 *SIRT1*<sup>-/-</sup> día 6 (Figura 28 B), hMgl, hNbML1 (Neuroblastos mediolaterales), hNProg, hOMTN (Núcleo oculomotor y troclear), hProgBP, hProgFPL, hProgFPM, hProgM, hRgl1 (Glia-like radial 1) y hRgl2b (Glia-like radial 2b).

Una vez más H9 *SIRT1*<sup>-/-</sup> muestra una composición poblacional distinta a la observada en las células H9 en el mismo día de diferenciación, la aparición de más células de linaje glial así como de progenitores de neuronas oculomotoras y del nervio ocular refuerza la teoría de heterogeneidad poblacional celular causada por la delección de SIRT1 en esta clona.

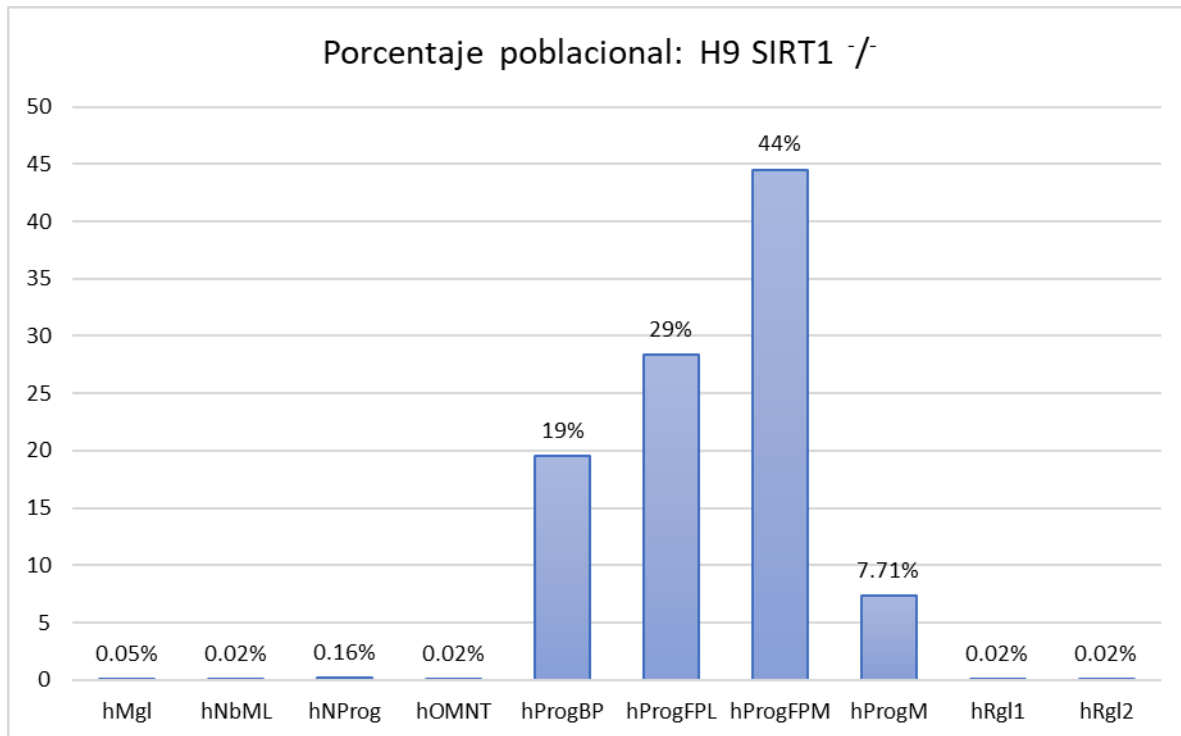
Para continuar estudiando esta heterogeneidad se revisó la composición porcentual de cada identidad en H9 *SIRT1*<sup>-/-</sup> día 6 usando el porcentaje de células asignado a cada identidad (Figura 29).

La identidad celular más abundante son los hProgFPM con el 44% de la población (2240 células), en segundo lugar, tenemos los hProgFPL con el 29% (1428 células) y en tercer lugar a los hProgBP con el 19% (981 células). Los porcentajes anteriores nos muestran que hay una pérdida en la cantidad de células comprometidas a piso medio basal y una ganancia en las células con identidad de piso basal lateral, lo anterior pese al sometimiento al protocolo de diferenciación a progenitores de piso medio basal.



**Figura 28:** Representación tSNE de células H9 *SIRT1*<sup>-/-</sup> día 6. (A) : Se muestran los 18 clusters presentes en la muestra H9 *SIRT1*<sup>-/-</sup> día 6.; (B) Se observan las 10 identidades presentes en las células H9 *SIRT1*<sup>-/-</sup> día 6, hNbML1 (Neuroblastos mediolaterales; Rojos), hOMTN (Núcleo oculomotor; Verde oscuro), hMgl (Microglía; Amarillo), hRgl2b (Glia-like radial 2; Olivo), hNProg (Progenitores Neurales; verde), hRgl1 (Glia-like Radial 1; azul claro),

*hProgBP (Progenitores de plato Basal; Azul oscuro), hProgFPL (Progenitores de piso lateral; Morado), hProgFPM (Progenitores de piso medio basal; Marrón) y hProgM (Progenitores línea media; Negro).*



**Figura 29:** Porcentaje de la población asignado a cada identidad del set de referencia de (La Manno et al., 2016) en la condición H9 día 6.

Tras realizar observaciones en los resultados obtenidos con esta metodología podemos señalar que a pesar de que las muestras H9 día 0 y día 6 poseen una mayor cantidad de clústeres (20 y 16 respectivamente) en comparación a los clústeres detectados en las muestras de H9 *SIRT1*<sup>-/-</sup> día 0 y día 6 (18 y 12 respectivamente) la asignación de identidades celulares muestra una mayor heterogeneidad celular presente en las células de H9 *SIRT1*<sup>-/-</sup>.

La presencia de identidades neuronales fue compartida entre las células H9 día 0 (Figura 24) y las células H9 *SIRT1*<sup>-/-</sup> día 0 (Figura 25), teniendo en común las identidades de astrocitos, neuroepitelio y neuronas, sin embargo, las identidades exclusivas de H9 *SIRT1*<sup>-/-</sup> día 0 (Tabla 7) muestran linajes embrionarios distintos a los observados en el día 0 de la muestra H9, siendo principalmente linajes correspondientes a la capa mesodérmica del embrión; estas identidades incluyen condrocitos, hepatocitos y células de músculo liso (Evseenko et al., 2010).

La aparición de estos linajes nos muestra que la delección de *SIRT1* genera heterogeneidad dentro de la población aún en condiciones de mantenimiento basal, aún en presencia de un medio especializado

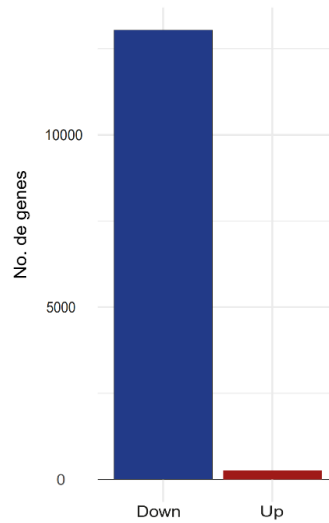
para ESCs. Lo mismo puede ser observado al analizar y comparar las muestras H9 día 6 y H9 *SIRT1* / día 6 utilizando un set de datos especializado en desarrollo neuronal en humano (La Manno et al., 2016), de forma similar ambas muestras mostraron identidades en común, específicamente aquellas relacionadas con la formación del plato basal (hProgBP, hProgFPL y hProgFPM) durante la formación del tubo neural en el embrión, sin embargo, las células H9 *SIRT1* / día 6 mostraron la presencia de tipos celulares neuronales más avanzados y especializados como los progenitores de neuronas oculomotoras (OMTN) que aparecen en puntos de desarrollo mucho más avanzados de acuerdo a lo publicado por (La Manno et al., 2016), de igual forma la aparición de identidades relacionadas con Glia (Tabla 7, identidades hRgl1 y hRGl2b) puede ser asociada con la iniciación del programa de gliogenesis (Kriegstein & Alvarez-Buylla, 2009).

A pesar de la presencia de heterogeneidad al realizar la asignación de identidad celular en las muestras y aun cuando la composición poblacional se veía alterada, los clusters formados por la reducción dimensional y clustering de *SingleCell* así como la detección de pocas células en algunas identidades nos llevó a evaluar el uso de otro software de análisis o pipeline disponible para datos de scRNAseq, la elección fue utilizar *Seurat* y comparar los datos obtenidos para entender mejor el significado biológico dentro de ellos.

#### **5.6.4. Obtención de genes diferencialmente expresados (DEG): Pseudobulk**

Otro de los análisis que se realizaron fue la obtención de DEG en objetos compuestos por ambas muestras a través de la metodología de PseudoBulking presente en la paquetería *scran*, estos genes son de interés ya que reflejan los cambios en expresión causados directamente por el silenciamiento de *SIRT1*, además permiten identificar mediante análisis de enriquecimiento los procesos biológicos afectados por el cambio de expresión en estos genes.

De acuerdo a este análisis se obtuvieron 2430 genes a la baja y 1668 a la alta en las células H9 *SIRT1* / respecto a las células H9 en el día 0; en el caso del día 6 de ambas muestras se encontraron 4395 genes regulados a baja y 3885 genes a la alta en las células H9 *SIRT1* / respecto a las células H9. Los resultados obtenidos tras este análisis se evaluaron y se encontró un total de 13300 genes diferencialmente expresados (Figura 30), de estos se estimaba que 259 estaban a la alta en el modelo con delección de *SIRT1* mientras que los 13042 restantes estaban expresados a la baja.



**Figura 30:** Barplot de DEG encontrados mediante pseudobulking en día 0. Se muestran los 259 genes a la alta (rojo) y los 13042 genes a la baja (azul) obtenidos tras realizar pseudobulking en los datos de día 0 en la librería SingleCellExperiment.

Por la anterior se llegó a la conclusión de que estos números eran excesivos y existía una alta probabilidad de que la gran mayoría de ellos fueran falsos positivos o sobre estimaciones de expresión diferencial, esto último en conjunto con lo mencionado en la sección de clustering y asignación de identidad celular nos hizo optar por evaluar *Seurat* al notar que ofrecía soluciones mucho más flexibles y robustas a la hora de realizar la obtención de DEG.

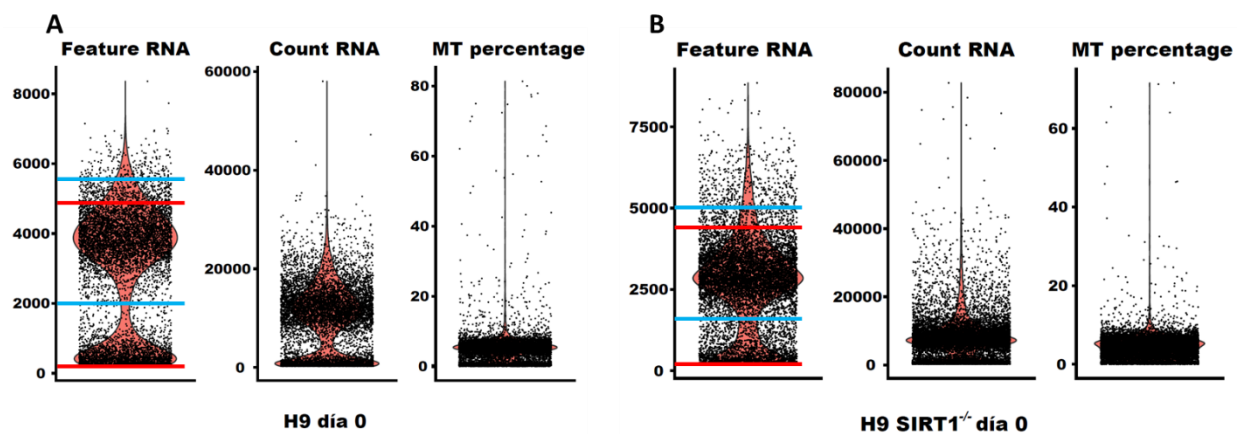


## 5.7. Seurat

### 5.7.1. Control de calidad

El primer paso tras crear los objetos en *Seurat* que almacenan nuestras muestras es la eliminación de las células de mala calidad., De acuerdo a lo recomendado por (Satija et al., 2015), un mínimo de 200 genes (que llamaremos *features*) por células y un máximo de 5% de genes mapeados a mitocondria es suficiente para asegurar obtener solo células con buena calidad, sin embargo se recomienda consultar los gráficos de violín para cada muestra y ajustar este filtro.

Se realizaron los gráficos de violín de calidad para los valores de *features* por célula (*nFeature\_RNA*), número de cuentas o moléculas de RNA por célula (*nCount\_RNA*) y porcentaje de genes mapeados a mitocondria (*percent.mt*) para cada una de las muestras, y se eligieron parámetros de calidad a partir de lo observado en ellos. En los gráficos de las células H9 día 0 podemos observar que una gran cantidad de células se encuentra en la parte inferior del violín (Figura 31 A) con menos de 2000 *features*, y un segundo grupo de células se observa que contiene entre los 2000 y hasta 5500 *features*. Debido a esto, y ante la posibilidad de que estas diferencias en el conteo de *features* sesgara los análisis, se determinó que se evaluarían dos conjuntos de datos para la muestra H9 día 0, el primero de ellos con los criterios de mínimo 200 *features*, máximo 5000 *features* y con un máximo del 5% de estos mapeados a mitocondria (Figura 31 A, líneas rojas); el segundo conjunto de criterios abarcaría las células con un mínimo de 2000 *features* detectados, un máximo de 5500 y con un límite de 5% de estos mapeados a mitocondria (Figura 31 A, líneas azules). Con el primer conjunto de criterios se obtuvieron 4036 células de buena calidad mientras que con el segundo conjunto de criterios se obtuvieron 1959 células.



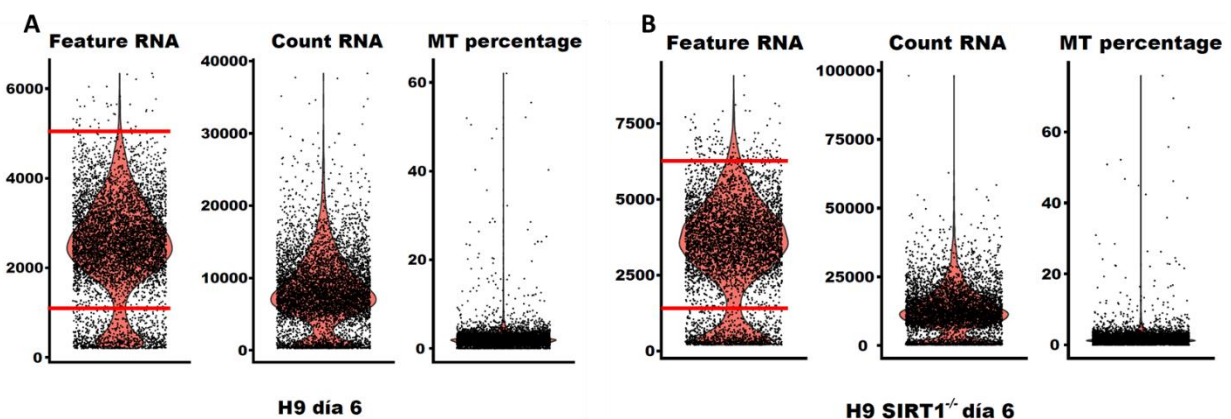
**Figura 31:** Gráfico de violín de parámetros de calidad en muestras día 0. Se observan el número de *features* detectados por células (*nFeature\_RNA*), el número de moléculas de RNA por célula (*nCount\_RNA*) y el porcentaje de genes mapeados a mitocondria (*percent.mt*) en la muestra H9 día 0 (A) y H9 SIRT1<sup>-/-</sup> día 0 (B)

En el caso de la muestra H9 *SIRT1*<sup>-/-</sup> día 0 encontramos algo similar, ya que una cantidad notable de células se encontraba por debajo de los 1000 *features* detectados (Figura 31 B), de forma similar a la muestra H9 día 0, elegimos dos conjuntos de criterios para determinar el umbral de buena calidad, el primer conjunto considerado un mínimo de 200 *features* y un máximo de 4000 *features* con menos del 5% de estos mapeados a mitocondria (Figura 29 B, líneas rojas); el segundo grupo de criterios abarcando un mínimo de 1500 *features*, un máximo de 5000 y con un límite de 5% de estos mapeados a mitocondria (Figura 31 B, líneas azules).

El primer grupo de criterios resultó en 5008 células que pasaron el filtro de buena calidad, mientras que en el segundo este número se redujo a 3690.

Para la muestra H9 día 6 se observó una distribución más uniforme de las células a través de los tres criterios de calidad, a partir de la observación de los gráficos de violín de calidad se decidió que las células con un mínimo de 1000 *features*, un máximo de 5000 y un porcentaje menor al 5% de estos mapeados a mitocondria serían consideradas de buena calidad (Figura 32 A, líneas rojas). Este filtro de calidad detectó 5431 células que cumplían con los criterios en la muestra H9 día 6.

Por último, se realizaron los gráficos de violín de los tres parámetros de calidad de H9 *SIRT1*<sup>-/-</sup> día 6 (Figura 32 B), se observó que la gran mayoría de las células se encontraban agrupadas entre 1500 y 6000 *features* detectados, por lo que se eligieron como el mínimo y máximo respectivamente en los criterios de calidad, se agregó además la condicional de que el límite de ellos mapeados a mitocondria podía ser del 5% (Figura 32 B, líneas rojas). Tras aplicar los criterios de calidad se obtuvieron 4140 células para esta muestra.



**Figura 32:** Gráfico de violín de parámetros de calidad en muestras día 6. Se observan el número de *features* detectados por células (*nFeature\_RNA*), el número de moléculas de RNA por célula (*nCount\_RNA*) y el porcentaje de genes mapeados a mitocondria (*percent.mt*) en la muestra H9 (A) y H9 *SIRT1*<sup>-/-</sup> (B).

Se realizó una tabla para resumir los criterios de calidad definidos en ambas condiciones (Tabla 5), esta tabla se usará en futuras referencias para identificar el número de células y los criterios de calidad que dieron origen a las mismas.

<b>Criterios de calidad</b>	<b>Mínimo de genes</b>	<b>Máximo de genes</b>	<b>% mapeado a mitocondria</b>	<b>No. De células</b>
<b>H9 día 0 laxos</b>	200	5000	5%	4036
<b>H9 día 0 estrictos</b>	2000	5500	5%	1959
<b>H9 <i>SIRT1</i> / día 0 laxos</b>	200	4000	5%	5008
<b>H9 <i>SIRT1</i> / día 0 estrictos</b>	1500	5000	5%	3690
<b>H9 día 6</b>	1000	5000	5%	5431
<b>H9 <i>SIRT1</i> / día 6</b>	1500	6000	5%	4140

**Tabla 5:** Criterios de calidad Seurat. Se muestran los 6 criterios de calidad de las 4 muestras de las células H9 día 0, H9 *SIRT1* / día 0, H9 día 6 y H9 *SIRT1* / día 6.

### 5.7.2. Clustering y anotación de identidad.

Una vez que se definieron los criterios de calidad para filtrar nuestras muestras se procedió a realizar la normalización, escalamiento de datos, detección de genes variables, reducción de dimensiones lineal (PCA), clustering y reducción de dimensiones utilizando UMAP en la muestra H9 día 0 con los criterios laxos (ver Tabla 5) se encontraron un total de 8 clústeres agrupados en dos conjuntos de clústeres, el primero de ellos (Figura 33 A, círculo rojo) conformado por los clústeres 5,2 y 0, mientras que el segundo conjunto (Figura 33 A, círculo azul) conformado por los clúster 6,7,1,3 y 4.

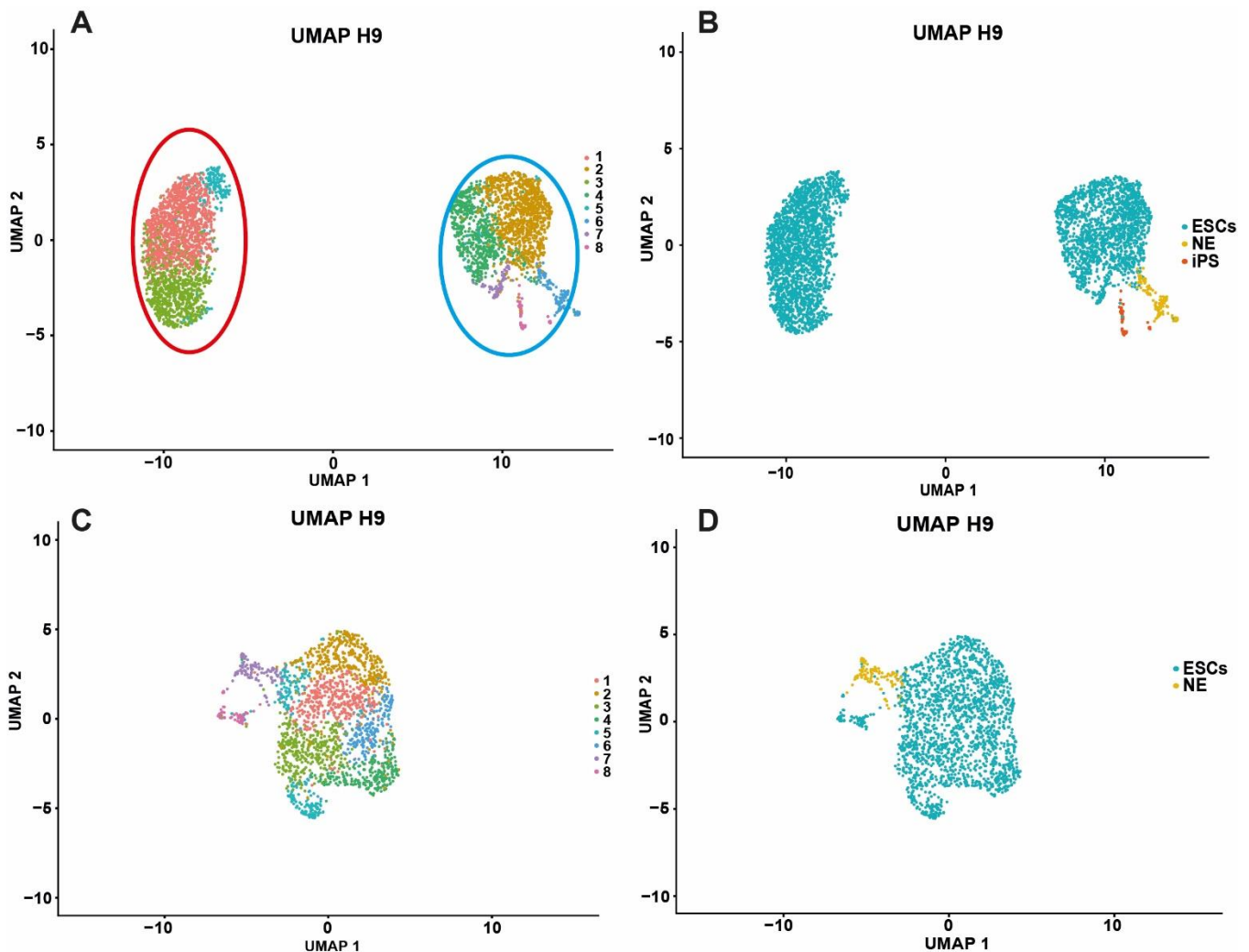
Una vez obtenido el UMAP se procedió a anotar la identidad de este conjunto de células utilizando los datos contenidos en el HPCA, la librería de anotación fue la misma utilizada en la metodología de *SingleCellExperiment* sin embargo *Seurat* utiliza el etiquetado por cluster y no por célula individual como en *SingleCellExperiment*.

La anotación de la muestra H9 día 0 con criterios de calidad laxos identificó 3 tipos celulares dentro de la población (Figura 33 B): *Embryonic\_stem\_cells* (ESCs), *Neuroepithelial\_cell* (NE) e *iPS\_cells* (iPSc).

Se repitió el mismo proceso con el set de datos generados utilizando los criterios de calidad estrictos (Ver tabla 5), en la representación UMAP se observaron 8 clústeres con una distribución bastante homogénea (Figura 33 C).

Se realizó la anotación de la identidad celular de la muestra utilizando la referencia del HPCA, en el caso de los criterios de calidad estrictos se encontraron solo dos identidades celulares asignadas a la muestra, *ESCs* y *NE* (Figura 33 D) siendo la primera de éstas la más abundante con 1847 células (Figura 31 D).

Los UMAPS de ambos sets de datos obtenidos de la muestra H9 día 0 nos permiten inferir que los criterios de calidad estrictos eliminan el grupo de células observadas en la Figura 33 A círculo rojo, así como a un grupo de células cuya identidad era la de células iPS, lo anterior podría ser debido a que esta población celular contenía un bajo conteo de features detectados colocándolos cerca del nivel más bajo observado en el gráfico de violín de Features en la Figura 31. Se decidió conservar ambos sets de datos para los análisis posteriores.



**Figura 33:** UMAPs H9 día 0 en ambos criterios de calidad. Se muestran los clústers de la muestra H9 día 0 en ambos criterios de calidad; laxos (panel A) y estrictos (panel C). También se muestran las identidades celulares asignadas a cada clúster en ambos criterios de calidad; laxos (panel B) y estrictos (panel D). Las abreviaturas corresponden a su identidad de la siguiente manera: ESCs (Células troncales embrionarias; azul), NE (Neuroepitelio; amarillo) e iPS (Células iPS; naranja).

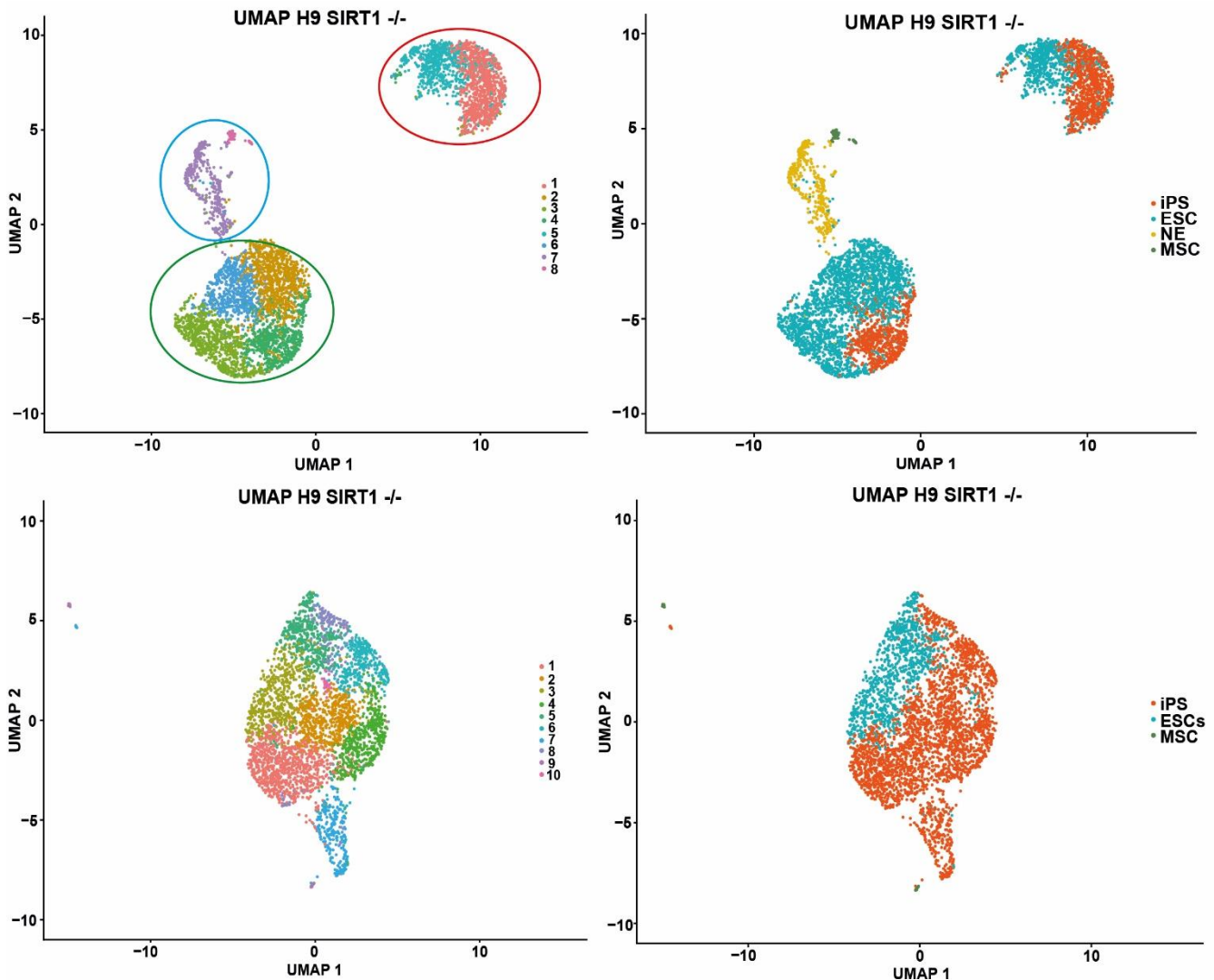
En el caso de H9 *SIRT1*  $\uparrow$  día 0 el primer conjunto de criterios (criterios de calidad laxos) generó un total de 8 clústeres que a su vez forman tres grupos dentro del UMAP (Figura 34 A), el primer grupo está conformado por los clústeres 1 y 3 (Figura 34 A, círculo rojo), en el segundo grupo se observan dos clústeres dispersos que son el 6 y 7 (círculo azul); finalmente el tercer grupo se compone por los cluster 2,4,5 y 0 (círculo verde)

La anotación de identidad de esta muestra se realizó con la misma referencia (HPCA) y se obtuvieron 4 identidades: ESCs, NE, iPS e MSC (Figura 34 B). La aparición de una población con identidad de células troncales mesenquimales parece ser persistente en ambas metodologías de análisis de datos (*SingleCellExperiment* y *Seurat*).

Otro cambio a resaltar en comparación a los sets de la muestra H9 día 0 es la disminución de la proporción de células asignadas a la etiqueta de células troncales embrionarias (ESCs) y el aumento de las células asignadas como iPS (Figura 33 B).

En el caso de los criterios de calidad estrictos en la H9 *SIRT1*<sup>-/-</sup> día 0 se detectaron un total de 10 clusters (Figura 34 C) con una distribución bastante homogénea.

Tras realizar la anotación de identidades celulares utilizando la referencia del HPCA (Figura 34 D) encontramos que las células asignadas a la identidad de neuroepitelio se perdían y la cantidad de células iPS aumentaba considerablemente superando incluso la cantidad de células asignadas a ESCs (Figura 34 D).



**Figura 34:** UMAPs H9 *SIRT1*<sup>-/-</sup> día 0 en ambos criterios de calidad. Se muestran los clústers de la muestra H9 *SIRT1*<sup>-/-</sup> día 0 en ambos criterios de calidad; laxos (panel A) y estrictos (panel C). También se muestran las identidades celulares asignadas a cada clúster en ambos criterios de calidad; laxos (panel B) y estrictos (panel D).

*Las abreviaturas corresponden a su identidad de la siguiente manera: ESCs (Células troncales embrionarias; azul), NE (Neuroepitelio; amarillo), iPS (Células iPS; naranja) y MSC (Células troncales mesenquimales; verde).*

Tras analizar ambos sets de datos nos percatamos de que H9 *SIRT1*<sup>-/-</sup> presenta un linaje celular distintivo en comparación de la muestra H9 día 0, la presencia de un linaje mesodérmico parece ser efecto de la delección de *SIRT1* en esta clona, se analizará en un futuro los genes marcadores de esta población. Por otro lado es importante destacar que una diferencia muy clara entre la muestra WT y *SIRT1*<sup>-/-</sup> es la aparición de una gran cantidad de células con identidad iPS, recordemos que este tipo celular posee marcadores de pluripotencia, sin embargo, son producto de una reprogramación celular por lo que no son idénticas a una células troncal embrionario, un objetivo futuro será evaluar que genes se ven alterados en estas poblaciones para conocer los efectos producidos sobre ellas al realizar la delección de *SIRT1*.

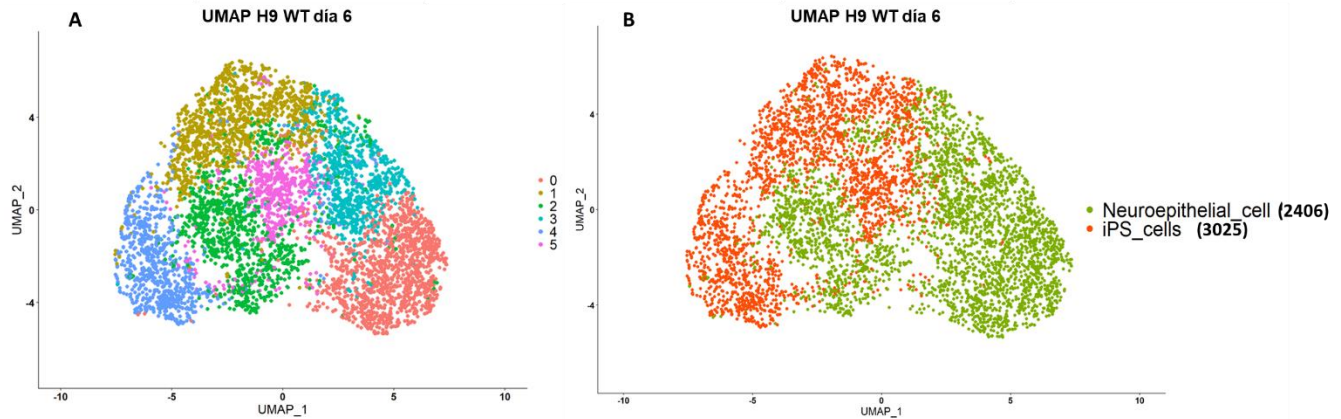
De forma más cuantitativa, en el caso de los criterios laxos y estrictos de calidad de la muestra H9 *SIRT1*<sup>-/-</sup> día 0, las ESCs representan el 58 y 34% respectivamente mientras que en la muestra H9 día 0 se observaba un 95% (criterios de calidad laxos) y 94% (criterios de calidad estrictos).

### **5.7.3. Clustering y anotación de identidad día 6.**

Se llevo a cabo la representación gráfica usando la técnica de reducción multidimensional de las muestras H9 WT y H9 *SIRT1*<sup>-/-</sup> correspondientes al día 6 del protocolo de diferenciación a NPC utilizando los criterios de calidad elegidos para ambas condiciones (Tabla 5).

En el caso de la condición H9 WT, se recuperaron un total de 5431 células que se agruparon en 6 clústers (Figura 35 A), al anotar la identidad de estos pudimos encontrar solo dos identidades celulares: NE y iPSc (Figura 35 B).





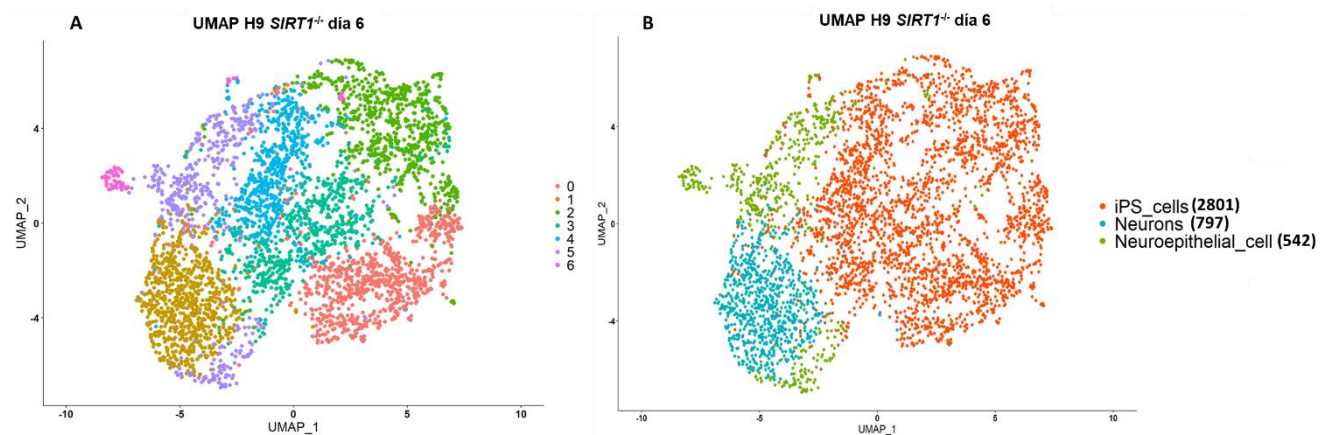
**Figura 35:** UMAP H9 WT día 6. (A) Se observan los clústers detectados por Seurat en la muestra H9 WT día 6; (B) Se observan las identidades asignadas a cada clúster en la muestra H9 WT día 6, Neuroepitelio (Verde) y células iPS (Naranja).

Al analizar la cantidad de células de cada identidad nos damos cuenta de que las iPSc son las más abundante y en segundo lugar las células NE.

Se realizo el mismo procedimiento con las células de la condición H9 SIRT1 <sup>-/-</sup> y se encontraron 7 clústers que al ser anotados se asociaron a tres identidades celulares distintas (Figura 36 A y B respectivamente): iPSc, NE y Neurons (NU).

**Figura 36:** UMAP H9 SIRT1 <sup>-/-</sup> día 6. (A) Se observan los clusters detectados por Seurat en la muestra H9 SIRT1 <sup>-/-</sup> día 6; (B) Se observan las identidades asignadas a cada cluster en la muestra H9 SIRT1 <sup>-/-</sup> día 6, Neuroepitelio (Verde), células iPS (Naranja) y Neuronas (azul).

En el caso de la condición H9 SIRT1 <sup>-/-</sup> se conserva la similitud en la porporción de iPSc en la población siendo estas las más abundantes, en segundo lugar, las NU y en tercer lugar de abundancia las células NE.





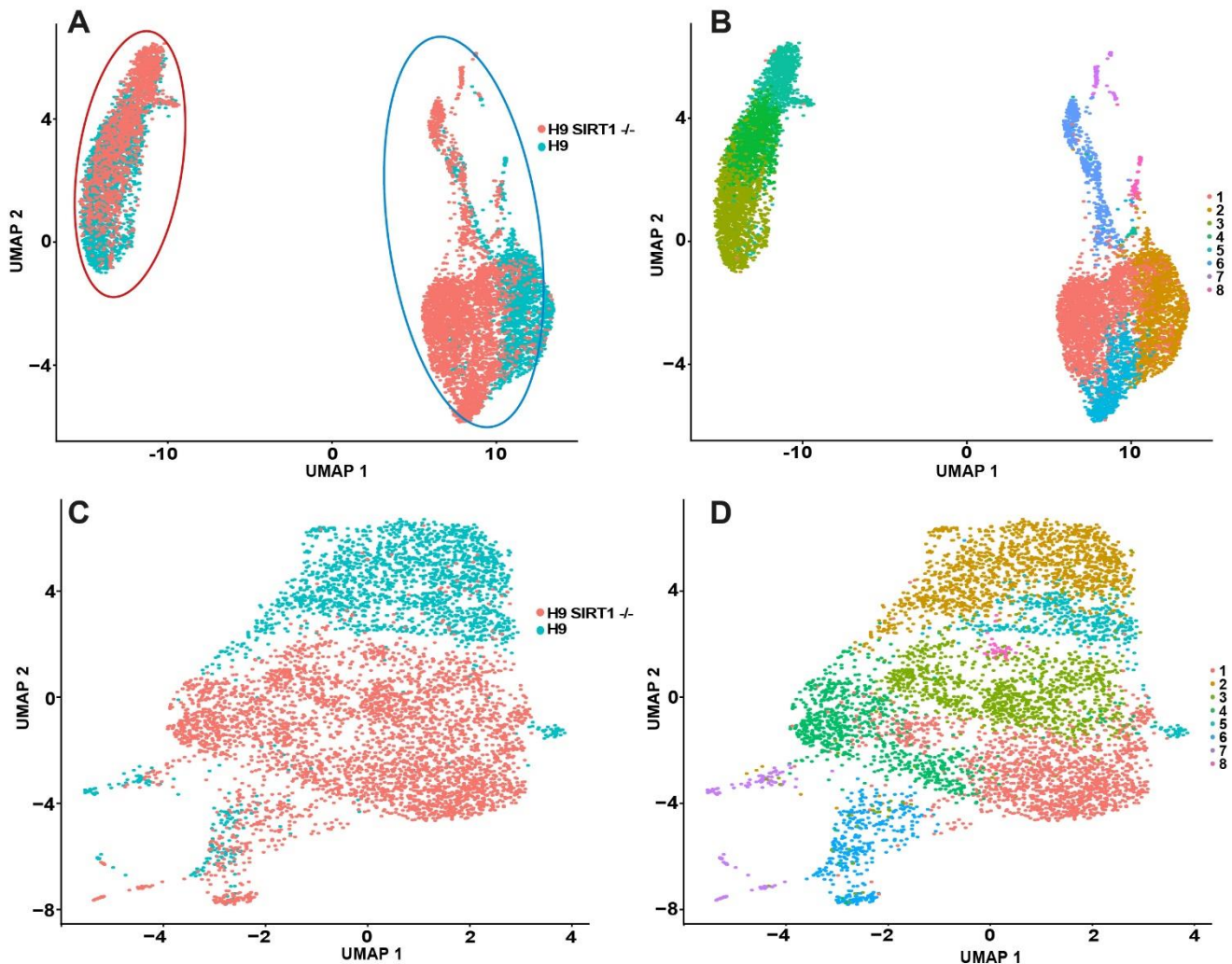
#### 5.7.4 Integración de datos: H9 vs H9 *SIRT1*<sup>-/-</sup> día 0

Se utilizó la metodología de integración de *Seurat* sobre las muestras H9 y H9 *SIRT1*<sup>-/-</sup> día 0 para poder obtener los genes diferencialmente expresados entre la condición WT (H9) y H9 *SIRT1*<sup>-/-</sup>. Para realizar la integración se utilizaron los dos sets producidos al aplicar criterios distintos de calidad (laxos y estrictos); se realizó un análisis rápido de estos genes mediante un enriquecimiento utilizando la herramienta *Enrichr* (Kuleshov et al., 2016).

La primera de las integraciones se realizó utilizando los sets de datos producidos con los criterios de calidad laxos, en total el objeto integrado estaba conformado por 9044 células, su distribución en un UMAP formaba dos grupos de clúster (Figura 37 A).

Este objeto integrado muestra dos grupos de clúster, el primero de ellos (Figura 37 A; círculo rojo) compuesto por tres clúster; 2,3 y 4 (Figura 37 B), que muestra mucha homogeneidad, esto nos lleva a pensar que no existe una diferencia tan marcada entre las células de cada condición, sin embargo, el segundo grupo (Figura 37 A; círculo azul) muestra que el clúster 1 y 8 (Figura 37 B) están compuestos principalmente por células de la H9 *SIRT1*<sup>-/-</sup> día 0, generando una fuente de comparación con los clústeres restantes de ese grupo compuestos por células H9 día 0.

Posteriormente se realizó la integración del segundo *set* de datos producidos por los criterios de calidad estrictos utilizando el mismo procedimiento que la integración anterior. Se obtuvo un solo conjunto de clústeres (Figura 37 D), se observó que cada una de las muestras se separaba a partir de su condición (WT o *SIRT1*<sup>-/-</sup>), en el caso de las células *SIRT1*<sup>-/-</sup> se observa que los clústeres 1 y 4 están compuestos casi exclusivamente por células de este origen (Figura 37 C), los clústeres 2, 3, 0, 7 y 5 estaban compuestos principalmente por células de origen WT (Figura 37 C).



**Figura 37:** UMAPs de la integración de los datos de H9 y H9 SIRT1<sup>-/-</sup> día 0 con ambos criterios de calidad. Se observa la formación de dos grupos de clústers en la condición de criterios laxos (Panel A: círculos rojo y azul) así como la ausencia de estos en los criterios estrictos (Panel C). Se muestra el número de clústers que compone los criterios de calidad laxos (Panel B) y estrictos (Panel D)

### 5.7.5 Identificación de Genes Diferencialmente expresados (DEG)

Se utilizó el algoritmo de búsqueda de genes diferencialmente expresado *MAST* sobre las muestras integradas, se encontraron 270 genes expresados al alta y 186 genes expresados a la baja en las células H9 SIRT1<sup>-/-</sup> en los criterios de calidad laxos.

Estos genes se introdujeron en la herramienta web *Enrichr* y se evaluaron lo genes a la alta y a la baja por separado, los genes a la alta mostraron tener un impacto en el tipo de metabolismo en las células H9 SIRT1<sup>-/-</sup> día 0 (Figura 38 A) de acuerdo a la base de datos de *Reactome 2022* (Gillespie et al., 2022), es importante remarcar que una de las características más esenciales de una ESC es la

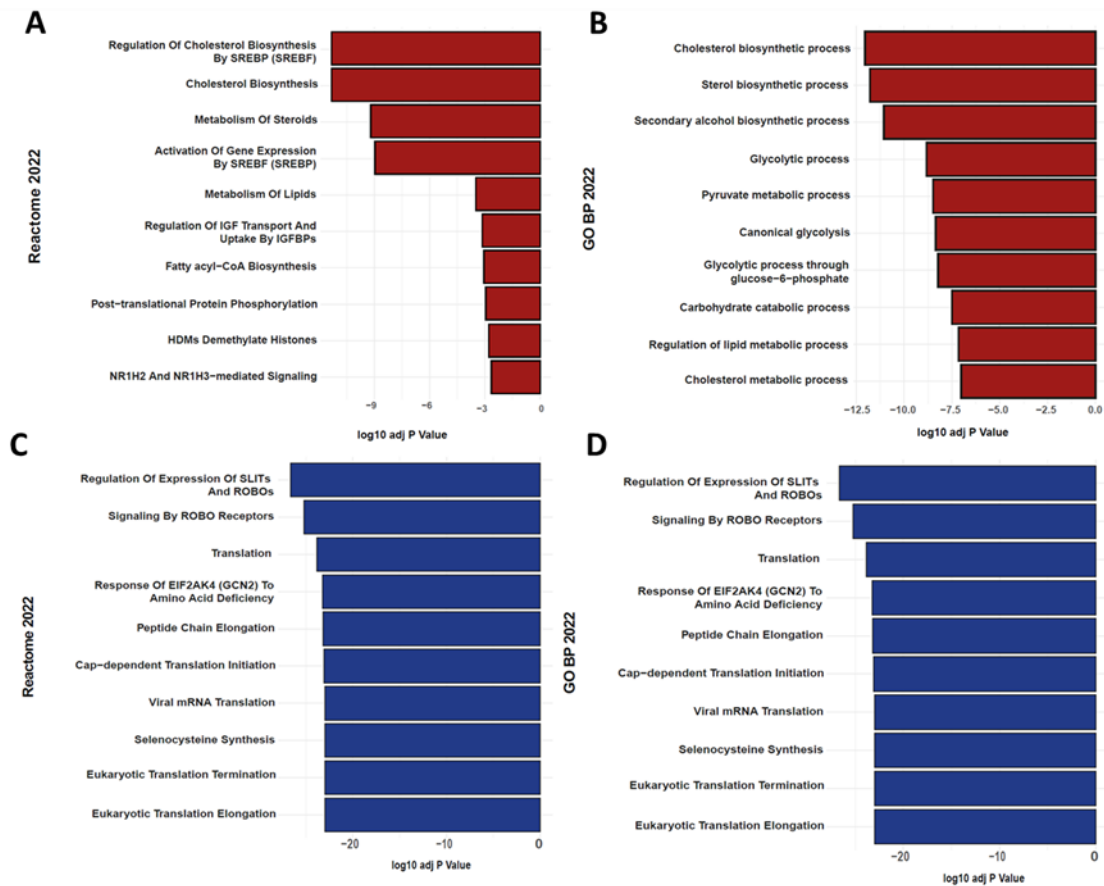
dependencia de un metabolismo basado en la glucólisis, el cambio en el tipo de metabolismo se ha relacionado previamente con la pérdida del fenotipo troncal (Ito & Suda, 2014).

De igual forma al consultar la base de datos de *Gene Ontology* ("The Gene Ontology resource: enriching a GOld mine," 2021) en el apartado de "*Biological Process*" o procesos biológicos (BP) nos percatamos que los procesos más enriquecidos se relacionaban con la ganancia de un metabolismo basado en lípidos, sin perder de igual forma el enriquecimiento en procesos biológicos relacionados con metabolismo de glucosa (Figura 38 B).

El proceso se repitió utilizando los genes a la baja y consultando la misma base de datos (*Reactome 2022*) para obtener los procesos más involucrados con estos genes (Figura 38 C).

Los dos primeros procesos encontrados al consultar *Reactome 2022* están relacionados con la expresión y regulación de *ROBO* y los *SLITS*. *ROBO* (Rondabout) está relacionado profundamente con el desarrollo del sistema neuronal, la migración células y la guía de los axones, sin embargo una de las funciones más interesantes de *ROBO* es la mediación de la transición entre progenitores neurales primarios a intermedios, específicamente durante la etapa de expansión, migración y establecimiento de estos durante el desarrollo neuronal mediante la activación del efector de Notch, *HES1* (Borrell et al., 2012). La deficiencia en la señalización y activación de las vías de desarrollo controladas por *SLIT* y *ROBO* podrían tener relación con el fenotipo observado en las células H9 *SIRT1*<sup>-/-</sup> debido a que no pueden sobrevivir la etapa de expansión y transición de neurales primarios a intermedios durante el protocolo de diferenciación al que se les sometió.

Los procesos enriquecidos restantes estaban relacionados con la traducción de proteínas y la elongación de la cadena de péptidos en estas células *SIRT1*<sup>-/-</sup>. En el caso de los procesos biológicos enriquecidos en la base de datos *Gene Ontology* con los genes a la baja en H9 *SIRT1*<sup>-/-</sup> (Figura 38 D) vemos que se relacionan con la traducción y la síntesis de péptidos durante esta, de igual forma observamos la presencia de procesos como la degradación de mRNA mensajero y la biogénesis de ribosomas, esto en conjunto nos permite inferir que la traducción de los mRNA así como la síntesis de proteínas en H9 *SIRT1*<sup>-/-</sup> se encuentran afectados tras la delección de *SIRT1*, esto podría sugerir un papel de esta última en la regulación de la traducción de mRNA en la célula.



**Figura 38:** Barplots de enriquecimiento. Se muestran los barplots producidos con los datos de enriquecimiento obtenidos de *Enrichr* con los genes diferencialmente expresados en *H9 SIRT1*<sup>-/-</sup> día 0 respecto a *H9* día 0 (criterios de calidad laxos). (A) Enriquecimiento genes al alta en la base de datos *Reactome 2022*; (B) Enriquecimiento de genes ala alta en la base de datos *Gene Ontology*, en el apartado *Biological Process*; (C) Enriquecimiento genes a la baja en la base de datos *Reactome 2022*; (B) Enriquecimiento de genes a la baja en la base de datos *Gene Ontology*, en el apartado *Biological Process*.

En el caso de los datos integrados con criterios de calidad estrictos; al utilizar el algoritmo *MAST*, se encontraron 206 genes diferencialmente expresados, de los cuales 105 se encontraban expresados a la alta en *H9 SIRT1*<sup>-/-</sup> y 101 a la baja, se utilizaron estos genes para llevar a cabo un análisis de enriquecimiento usando de nuevo *Enrichr* y consultando la base de datos de *Reactome* y *Gene Ontology BP 2022*.

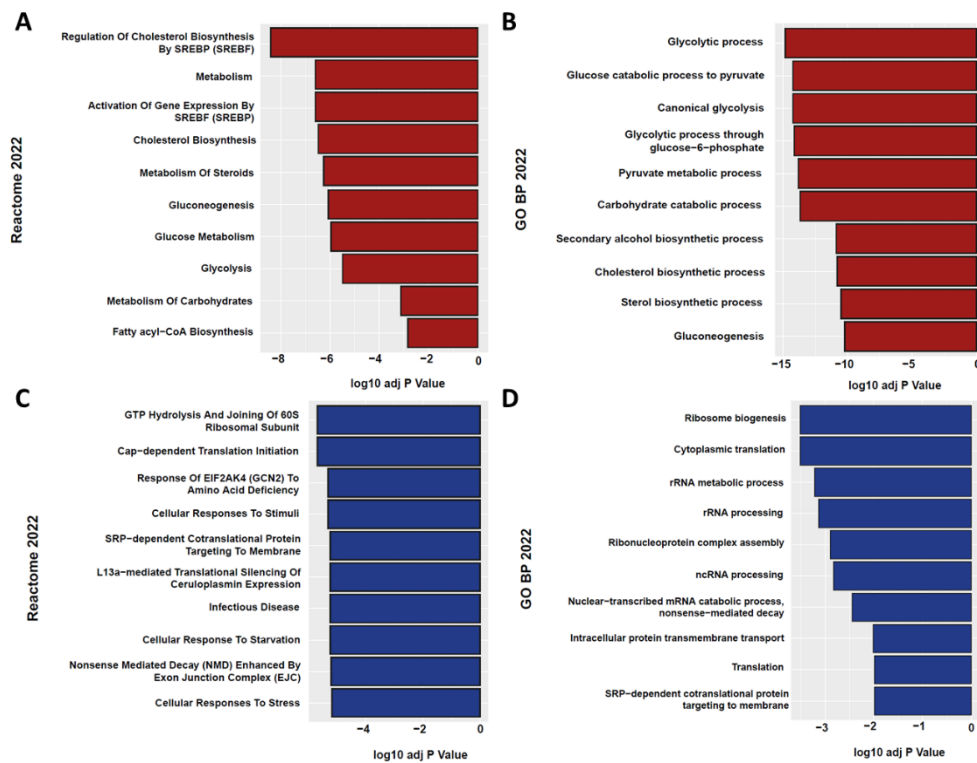
Los procesos obtenidos al introducir los genes a la alta en *H9 SIRT1*<sup>-/-</sup> son los mismos obtenidos en la integración de datos con criterios laxos, esto permite reforzar la idea de que un cambio en el metabolismo de las células tras la delección de *SIRT1* es un resultado consistente a explorar más adelante (Figura 39 A).

De igual forma los *BP* enriquecidos en estos genes a la alta se relacionan principalmente con metabolismo, sin embargo en el top de procesos aparecen algunos relacionados con metabolismo de

glucosa mientras que aquellos relacionados con metabolismo de lípidos aparecen más abajo en la lista (Figura 39 B).

Sin embargo, al introducir los genes expresados a la baja en la condición *SIRT1*<sup>-/-</sup> los procesos enriquecidos en *Reactome* cambiaron respecto a los obtenidos en el set integrado de criterios laxos (Figura 39 C).

La mayoría de los procesos presentes en este enriquecimiento están relacionados con el establecimiento del ribosoma, el inicio de la transcripción y la regulación de la transcripción y regulación de los micro RNAs. Por último, los BP enriquecidos en la base de datos *Gene Ontology* se centraban en regulación de la traducción, pero también se observan procesos como el procesamiento de RNAs ribosomales y el procesamiento de RNAs no condificantes (Figura 39 D).

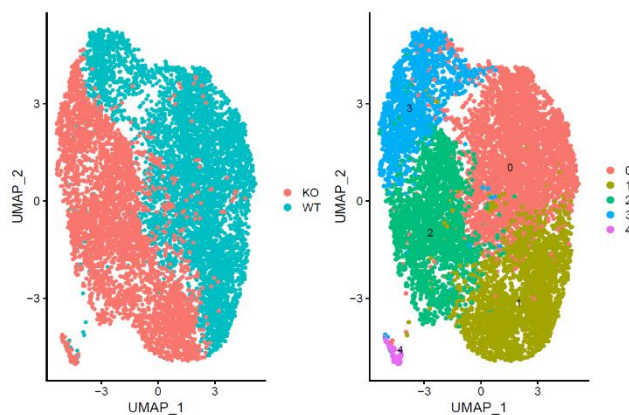


**Figura 39:** Barplots de enriquecimiento. Se muestran los barplots producidos con los datos de enriquecimiento obtenidos de Enrichr con los genes diferencialmente expresados en la H9 *SIRT1*<sup>-/-</sup> día 0 respecto a H9 día 0 (criterios de calidad estrictos). (A) Enriquecimiento genes al alta en la base de datos Reactome 2022; (B) Enriquecimiento de genes ala alta en la base de datos Gene Ontology, en el apartado Biological Process; (C) Enriquecimiento genes a la baja en la base de datos Reactome 2022; (B) Enriquecimiento de genes a la baja en la base de datos Gene Ontology, en el apartado Biological Process.

### 5.7.6 Integración de datos: H9 vs H9 *SIRT1*<sup>-/-</sup> día 6

Se utilizó la metodología de integración de datos de *Seurat* para poder unir los datos de las células H9 y H9 *SIRT1*<sup>-/-</sup> en el día 6 del protocolo de diferenciación hacia NPC, esto con la finalidad de obtener los genes diferencialmente expresados en estas condiciones.

En el caso de los datos integrados del día 6 podemos observar que se forman un total de 5 clústers (Figura 40 B) de los cuales el clúster 2 y 4 son exclusivos prácticamente de las células *SIRT1*<sup>-/-</sup> y el clúster 0 de las células WT, los clústeres restantes (3 y 1) se comparten en una proporción del 50% a cada una de las condiciones (Figura 40 A).



**Figura 40:** UMAP de la integración de los datos de H9 y H9 *SIRT1*<sup>-/-</sup> día 6. Se observa que cada condición (WT y *SIRT1*<sup>-/-</sup>) posee clústers exclusivos; WT clusters 0; KO clusters 2 y 4, el resto de los clústers se comparten entre ambas condiciones.

Posteriormente se procedió a utilizar el algoritmo MAST para poder encontrar los DEG entre la condición *SIRT1*<sup>-/-</sup> y la WT, se encontraron un total de 552 genes genes diferencialmente expresados en la condición *SIRT1*<sup>-/-</sup> respecto a la WT, de los cuales 335 se encontraban expresados al alta y 217 a la baja.

Los DEG encontrados fueron introducidos en *Enrichr* para poder llevar a cabo el enriquecimiento, se consultaron las bases de datos *Reactome* y *Gene Ontology BP 2022*.

Al observar los datos de enriquecimiento podemos notar una consistencia con los resultados obtenidos en enriquecimientos del día cero, esto debido a que muchos de los process que se encontraron enriquecidos a partir de los DGE coinciden, en el caso de la base de datos *Reactome 2022* (Figura 41 A) con los genes regulados al alta en el fenotipo *SIRT1*<sup>-/-</sup> no encontramos una vez más con la regulación de la biosíntesis de colesterol mediada por SREBP, es importante destacar que mucho de los genes objetivo mediados por SREBP están relacionados con la lipogénesis, de igual forma

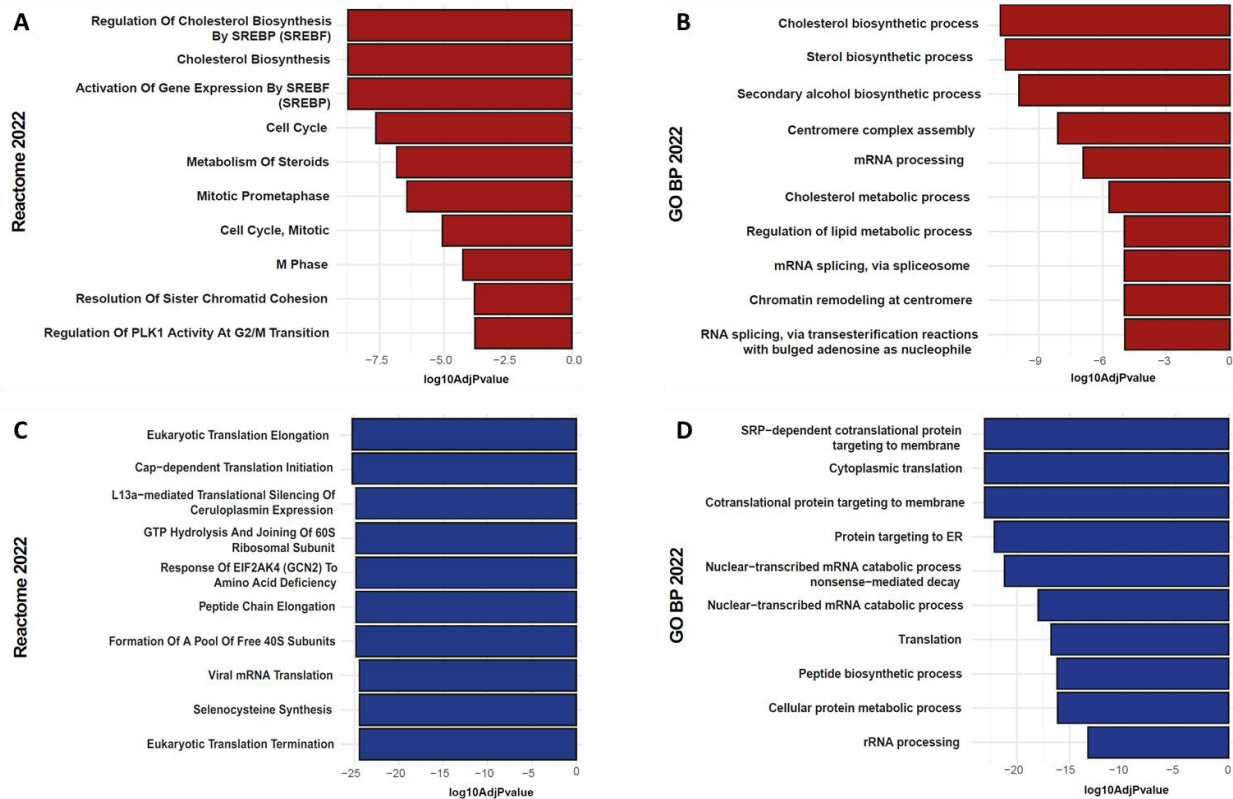
encontramos la biosíntesis de colesterol y la activación de la expresión de los genes mediados por SREBF lo que nos permite reforzar que el cambio en el tipo de metabolismo dentro de las células *SIRT1<sup>-/-</sup>* está presente aun cuando son sometidas a un proceso de diferenciación y de alguna forma podrían estar limitando este último. Otro de los procesos celulares más afectados tras la delección de SIRT1 parece ser el ciclo celular que, de igual forma que en el día 0, muestra alteraciones durante la prometafase mitótica y la mitosis.

En el caso de los procesos biológicos enriquecidos en la base de datos *GO BP 2022* volvemos a encontrar la biosíntesis de colesterol y de esteroides dentro de los principales *BP* enriquecidos, mientras que más abajo podemos encontrar procesos relacionados con el *splicing* y procesamiento del RNA (Figura 41 B).

El proceso se repitió introduciendo los genes regulados a la baja en las células *SIRT1<sup>-/-</sup>* en ambas bases de datos, en el caso de *Reactome 2022* (Figura 41 C) encontramos que los 3 procesos con mayor significancia estadística de acuerdo a nuestro enriquecimiento se encuentran relacionados con la traducción de mRNA hacia péptidos, incluyendo las etapas de iniciación de la traducción y elongación de la cadena polipeptídica, también podemos encontrar en tercera posición un proceso de silenciamiento de la traducción mediado por la subunidad L13a de la subunidad ribosomal 60s, los procesos enriquecidos restantes están relacionados con el ensamble de las unidades ribosomales, la acumulación de unidades 40s sin unirse y la terminación del proceso de traducción.

Por último, introduciendo los DGE a la baja de nuestras células *SIRT1<sup>-/-</sup>* en la base de datos *GO BO 2022* (Figura 41 D) encontramos que los *BP* con más significancia estadística están relacionados con transporte de proteínas a la membrana, catabolismo de mRNA nuclear, biosíntesis de péptidos, procesamiento de rRNA y traducción celular. Los resultados encontrados en el día 6 entonces nos sugieren que el cambio en el tipo de metabolismo en tras la delección de SIRT1 permanece aún después de una inducción a un proceso de diferenciación celular, además podemos observar otras consecuencias notorias tras la delección de SIRT1 donde se afecta principalmente su capacidad de tener una transición a través de las diversas fases del ciclo celular, viéndose más afectadas (de acuerdo a los enriquecimientos) las fases M y G2, por último también es importante destacar que las células *SIRT1<sup>-/-</sup>* presentan fallas en el proceso de traducción de sus mRNA hacia proteínas e incluso se sugiere que se ve afectado el proceso de *splicing* y maduración de sus mRNA durante la transcripción.





**Figura 41:** Barplots de enriquecimiento. Se muestran los barplots producidos con los datos de enriquecimiento obtenidos de Enrichr con los genes diferencialmente expresados en H9 *SIRT1*<sup>-/-</sup> día 6 respecto a H9 día 6. (A) Enriquecimiento genes al alta en la base de datos Reactome 2022; (B) Enriquecimiento de genes ala alta en la base de datos Gene Ontology, en el apartado Biological Process; (C) Enriquecimiento genes a la baja en la base de datos Reactome 2022; (B) Enriquecimiento de genes a la baja en la base de datos Gene Ontology, en el apartado Biological Process.

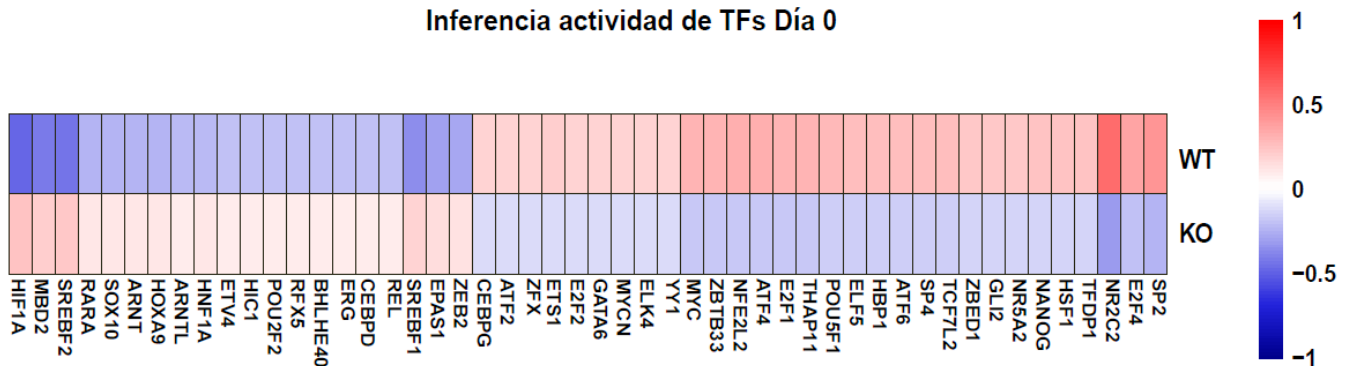
### 5.8.1 Inferencia de actividad de Factores de Transcripción (TF): Día 0

Se utilizó la librería DecoupleR para inferir los TF más activos dentro de nuestros datos, esta inferencia se realizó en los sets de datos integrados de ambas condiciones con las identidades anotadas de acuerdo con el HPCA, de igual forma se utilizó una anotación donde se compararon todas las células WT contra todas las células *SIRT1*<sup>-/-</sup>, sin importar su identidad celular.

En el caso de las células del día 0, para los análisis posteriores se decidió trabajar solamente con los datos producidos por los criterios de calidad estrictos, esto último con la finalidad de disminuir el ruido producido por las células con baja captura de *features* que se observan en los criterios de calidad laxos (Figura 35, círculo rojo); por su parte los datos producidos por los criterios de calidad estrictos forman un solo grupo de clústers diferenciado de la condición H9 *SIRT1*<sup>-/-</sup> (Figura 37, panel C).



La inferencia con *DecoupleR* encontró un total de 183 factores de transcripción activos en nuestros datos de día 0 (Figura suplementaria 1), con la finalidad de poder observar el top de TF más variables en actividad en nuestras muestras creamos un heatmap con el top 50 de estos (Figura 42) donde encontramos que TFs como *NR2C2*, *SP2*, *SP4*, *MYC* y *E2F4* con la mayor pérdida de actividad en el modelo *SIRT1*<sup>-/-</sup> respecto a la condición WT. De forma contraria, podemos observar que *HIF1A*, *SREBF2*, *EPAS1* y *SREBF1* son TFs que aumentan su actividad en la condición *SIRT1*<sup>-/-</sup>.

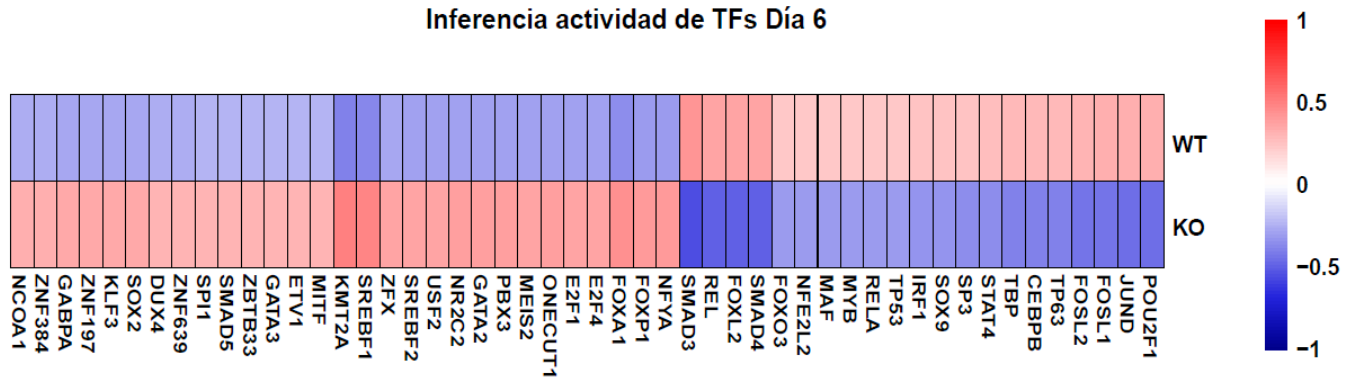


**Figura 42:** Heatmap de actividad de TFs inferida de los datos Día 0. Se muestran el top 50 de TFs más variables comparando las células WT contras las *SIRT1*<sup>-/-</sup> en los datos integrados del día 0.

### 5.8.2 Inferencia de actividad de Factores de Transcripción (TF): Día 6

En el caso de la inferencia de actividad de TFs en el día 6 se encontraron un total de 164 estadísticamente significativos (Figura suplementaria 2)., Con la finalidad de evaluar algunos de estos TF se creó un heatmap con el top 50 de TF con mayor variabilidad de actividad dentro de los datos del día 6 (Figura 43).

Algunos de los TFs con una menor actividad en las células *SIRT1*<sup>-/-</sup> en contraste con las WT son *SMAD3*, *SMAD4*, *FOXL2* y *REL* (Figura 43). Los TFs *FOXA1*, *SREBF1* y *KMT2* mostraron un aumento en su actividad en el modelo *SIRT1*<sup>-/-</sup> respecto a las células WT.



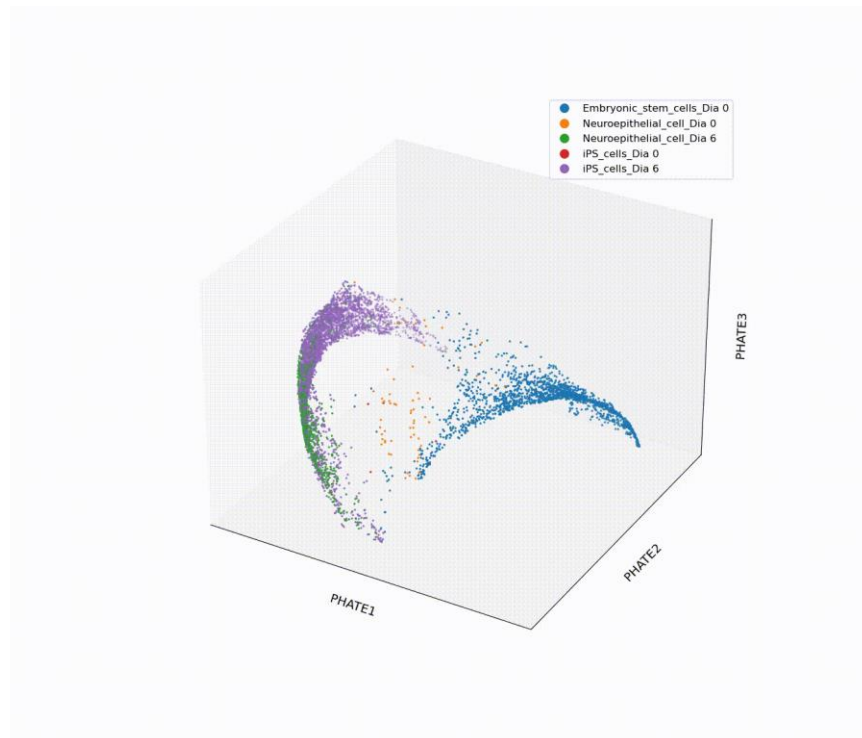
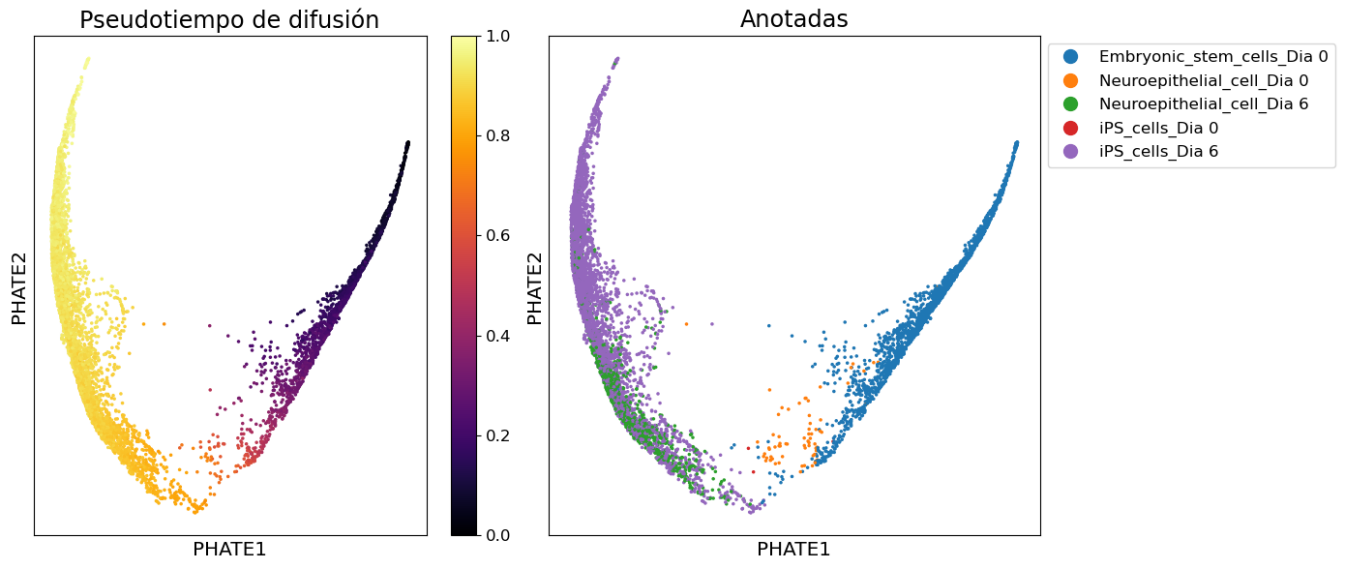
**Figura 43:** Heatmap de actividad de TFs inferida de los datos Día 6. Se muestran el top 50 de TFs más variables comparando las células WT contra las *SIRT1*<sup>-/-</sup> en los datos integrados del día 6.

### 5.9 PHATE y pseudotiempo de difusión

Se llevó a cabo la representación multidimensional PHATE para poder representar los datos integrados del día 0 y 6 de las condiciones WT y *SIRT1*<sup>-/-</sup>. En el caso de ambas condiciones se utilizaron los datos día 0 generados con los criterios estrictos de calidad.

El pseudotiempo de los ejes de diferenciación DO a D6 de ambas condiciones se calculó y proyectó sobre la representación PHATE, y la célula con el menor valor de pseudotiempo se eligió usando la célula con el menos valor de PHATE.

En el caso del eje de diferenciación WT (Figura 44) podemos observar que las células del día 0 se agrupan en una estructura curva muy similar a una letra U, observado el pseudotiempo (Figura 44, Pseudotiempo de difusión) nos damos cuenta que las células con el menor pseudotiempo (color morado oscuro) coinciden con las ESCs del día 0 (Figura 44, anotadas), de igual forma podemos observar que el extremo con el mayor pseudotiempo (color amarillo) coinciden con las células anotadas como iPSc del día 6, en conjunto esto nos permite concluir que nuestras células WT cumplen con un eje de diferenciación lineal como el esperado, donde las células del día 0 conforman el inicio del pseudotiempo y conforme se avanza a través de este recorreremos identidades más diferenciadas como neuroepitelio y células iPSc, tanto de día 0 como de día 6. De igual forma se puede apreciar en un video la estructura en tres dimensiones de su trayectoria con PHATE (Figura 44; video).



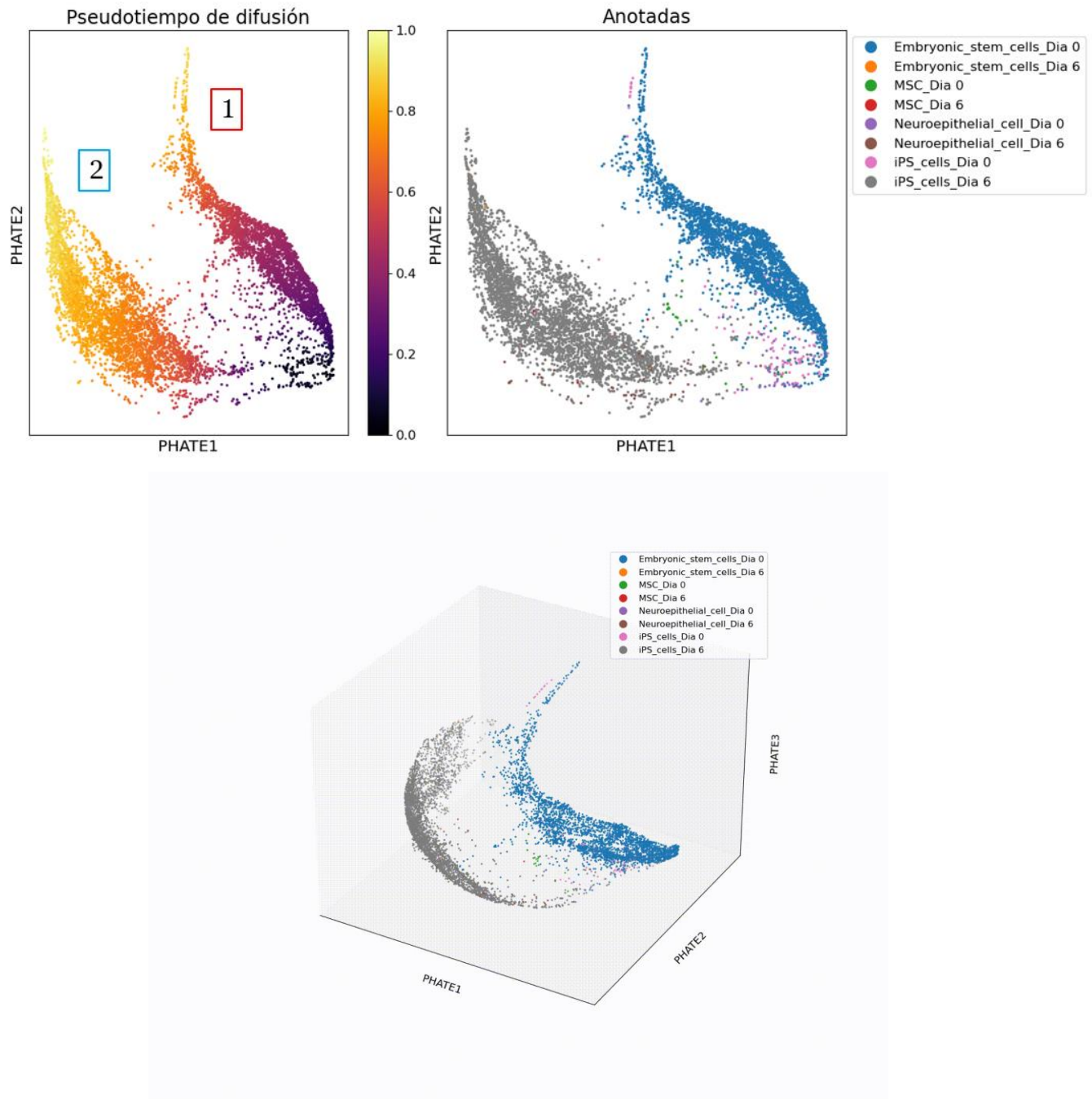
**Figura 44:** Representación PHATE y pseudotiempo de difusión en eje de diferenciación D0 a D6 WT. Se observa la representación generada por el algoritmo PHATE, del lado derecho observamos las identidades anotadas sobre la representación; del lado izquierdo observamos la proyección del pseudotiempo de difusión sobre el mapa PHATE, se observa también la escala de valor de este donde morado representa el valor mínimo (cero) y el amarillo el valor máximo (uno).

Al observar la representación PHATE del eje de diferenciación generado por los datos integrados día 0 y día 6 de las células *SIRT1*<sup>-/-</sup> (Figura 45) se aprecia una estructura similar a la de las células WT, sin embargo, la estructura parece menos definida y las células se ven más dispersas dentro de los clústers de identidad.

Lo anterior se ve reforzado al observar el pseudotiempo encontrado en estos datos (Figura 45, Pseudotiempo de difusión) donde observamos que las células con el menor valor de pseudotiempo se encuentran en el centro de la estructura y a partir de este surgen dos rutas alternativas (Figura 45 recuadro rojo y azul) con extremos correspondientes al valor máximo de pseudotiempo.

Para dar sentido a lo encontrado en el pseudotiempo podemos observar a qué identidades celulares corresponden estos pseudotiempo (Figura 45, Anotadas), el nodo central con el valor mínimo de pseudotiempo corresponde a un grupo pequeño de células pertenecientes a la identidad de ESCs del día 0, a partir de este podemos observar que la ruta 1 va a través de las células ESCs del D0 y que también podemos encontrar algunas células con identidad de Neuroepitelio Día 0 y iPSc día 0, la segunda ruta inicia en las ESCs día 0 hasta y recorre un gradiente en aumento del pseudotiempo a través de una clúster formado por las iPSc del día 6.

Las dos rutas encontradas en los datos *SIRT1*<sup>-/-</sup> nos sugieren que a pesar de estar sometidas al mismo tipo de estímulos y moléculas que conforman el protocolo de diferenciación a NPCs parte de las ESCs más troncales del día 0 no responden a este estímulo y que además la delección de *SIRT1* genera una heterogeneidad notoria dentro de las ESCs incluso en condiciones de mantenimiento, sin ningún estímulo de diferenciación o compromiso a linaje (Día 0). De igual forma se puede apreciar la estructura tridimensional de a proyección PHATE de la condición H9 *SIRT1*<sup>-/-</sup> en el video de la Figura 43, se aprecia que los dos brazos de la trayectoria de esta condición se orientan de forma horizontal en lugar de vertical como se aprecia en la condición H9 (Figura 45).



**Figura 45:** Representación PHATE y pseudotiempo de difusión en eje de diferenciación D0 a D6  $SIRT1^{-/-}$ . Se observa la representación generada por el algoritmo PHATE, del lado derecho observamos las identidades anotadas sobre la representación; del lado izquierdo observamos la proyección del pseudotiempo de difusión sobre el mapa PHATE, se observa también la escala de valor de este donde morado representa el valor mínimo (cero) y el amarillo el valor máximo (uno). Se remarcan dos rutas encontradas al proyectar el pseudotiempo de difusión representadas por el recuadro rojo (1) y el recuadro azul (2).

## 5.10 *Slingshot* y trayectoria celular

Posteriormente al cálculo del pseudotiempo con PHATE se dibujaron las trayectorias celulares a través de este usando *Slingshot*, con la intención de poder encontrar aquellos genes que oscilaban su expresión a lo largo de la trayectoria y en dirección al pseudotiempo.

Para poder utilizar *Slingshot* en nuestros datos es necesario darle el número de clústers presentes, así como el clúster de inicio de la trayectoria, se utilizó el mismo número de clústers calculados durante la obtención de la reducción multidimensional PHATE, en el caso de las células H9 se encontraron 5 clústers a lo largo de su distribución (Figura 46, A), para la condición H9 *SIRT1*<sup>-/-</sup> también se encontraron 5 clústers en toda su distribución (Figura 46, B).

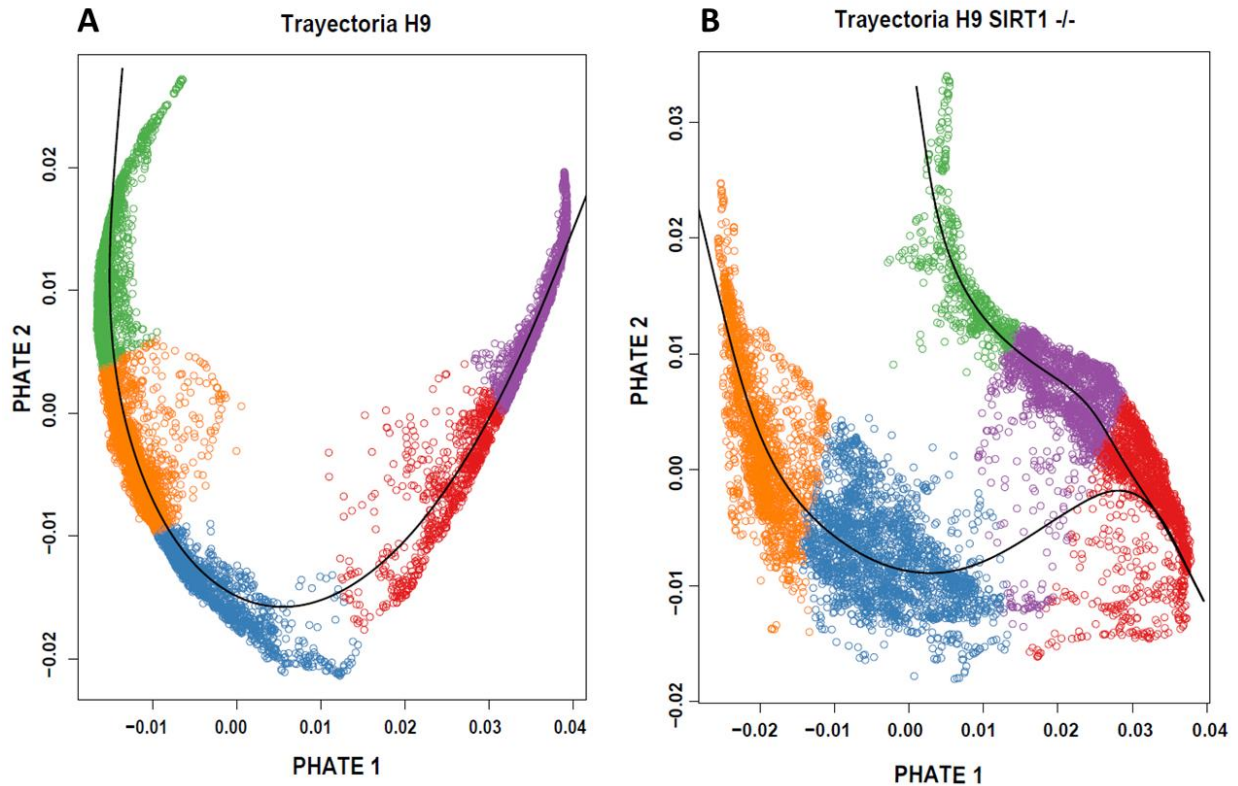
El clúster de inicio se definió cotejando la identidad celular y el clúster que contenía el valor de pseudotiempo más pequeño, en el caso de las células H9 el valor de pseudotiempo menor (Figura 44) se ubica en el clúster morado (Figura 46, A) y coincide con las células identificadas como ESCs del día 0 (Figura 42). En el caso de las células H9 *SIRT1*<sup>-/-</sup> podemos observar que el menor valor de pseudotiempo (Figura 45) se ubica en el cluster de color rojo (Figura 46, B), este último corresponde a algunas células identificadas como ESCs.

La trayectoria dibujada por *Slingshot* corresponde de igual forma con lo encontrado en el pseudotiempo, la condición H9 mostró una sola trayectoria que recorre los cinco clústers (Figura 46, A), este tipo de trayectoria es la esperada en nuestro modelo WT debido a que las células se están sometiendo a un estímulo de diferenciación dirigido, este protocolo no solamente las induce a un linaje específico sino que además se acompaña de una serie de inhibidores que las mantienen dentro del linaje esperado.

Por su parte la condición H9 *SIRT1*<sup>-/-</sup> muestra dos trayectorias (Figura 46, B) por un lado tenemos una trayectoria que parte del clúster rojo, hacia el morado y termina en el clúster verde, esta primera trayectoria corresponde al valor de pseudotiempo que recorre la identidad de ESCs día 0 en esta condición (Figura 44, B), esta primera trayectoria nos indica que la condición H9 *SIRT1*<sup>-/-</sup> muestra una heterogeneidad intracelular que es detectada como un proceso de diferenciación tanto por el pseudotiempo de difusión como por nuestra trayectoria.

La segunda trayectoria encontrada en nuestros análisis tiene su inicio en el clúster rojo y se desplaza hacia el clúster azul y naranja sucesivamente (Figura 46, B), esta trayectoria corresponde a las identidades de ESCs día 0 y su desplazamiento hacia la identidad de iPSc del día 6, es importante

mencionar que en esta condición no existe una clara agrupación de células neuroepiteliales del día 6 como lo que se observa en las de la condición H9.



**Figura 46:** Trayectorias trazadas por Slingshot en ejes de diferenciación D0 a D6, condiciones H9 y H9 *SIRT1*<sup>-/-</sup>. Se observan las trayectorias trazadas a través de la representación multidimensional PHATE en ambas condiciones: (A) Se muestra la trayectoria única trazada en la condición H9; (B) Se muestran las dos trayectorias presentes en la condición H9 *SIRT1*<sup>-/-</sup>.

Este análisis nos permite reforzar la existencia de una heterogeneidad en la línea H9 *SIRT1*<sup>-/-</sup> detectable a través de ambos análisis (pseudotiempo y trayectoria), además nos brinda la capacidad de detectar qué genes responden al avance del pseudotiempo a lo largo de la trayectoria. Para los análisis posteriores se utilizará la trayectoria y pseudotiempo de la línea H9 WT, sin embargo, en la condición H9 *SIRT1*<sup>-/-</sup> se decidió utilizar la trayectoria de diferenciación que va de rojo, a azul y finalmente a naranja, esto debido a que para realizar comparaciones adecuadas entre ejes de diferenciación esta es equivalente a la presente en H9.

### 5.11. Genes asociados al pseudotiempo

Se realizó el análisis de genes asociados al pseudotiempo para poder conocer aquellos genes que tenían diferencias estadísticas significativas en su expresión al final de la trayectoria de pseudotiempo entre las condiciones H9 WT y H9 *SIRT1*<sup>-/-</sup>, en segunda instancia se buscó reconocer aquellos genes

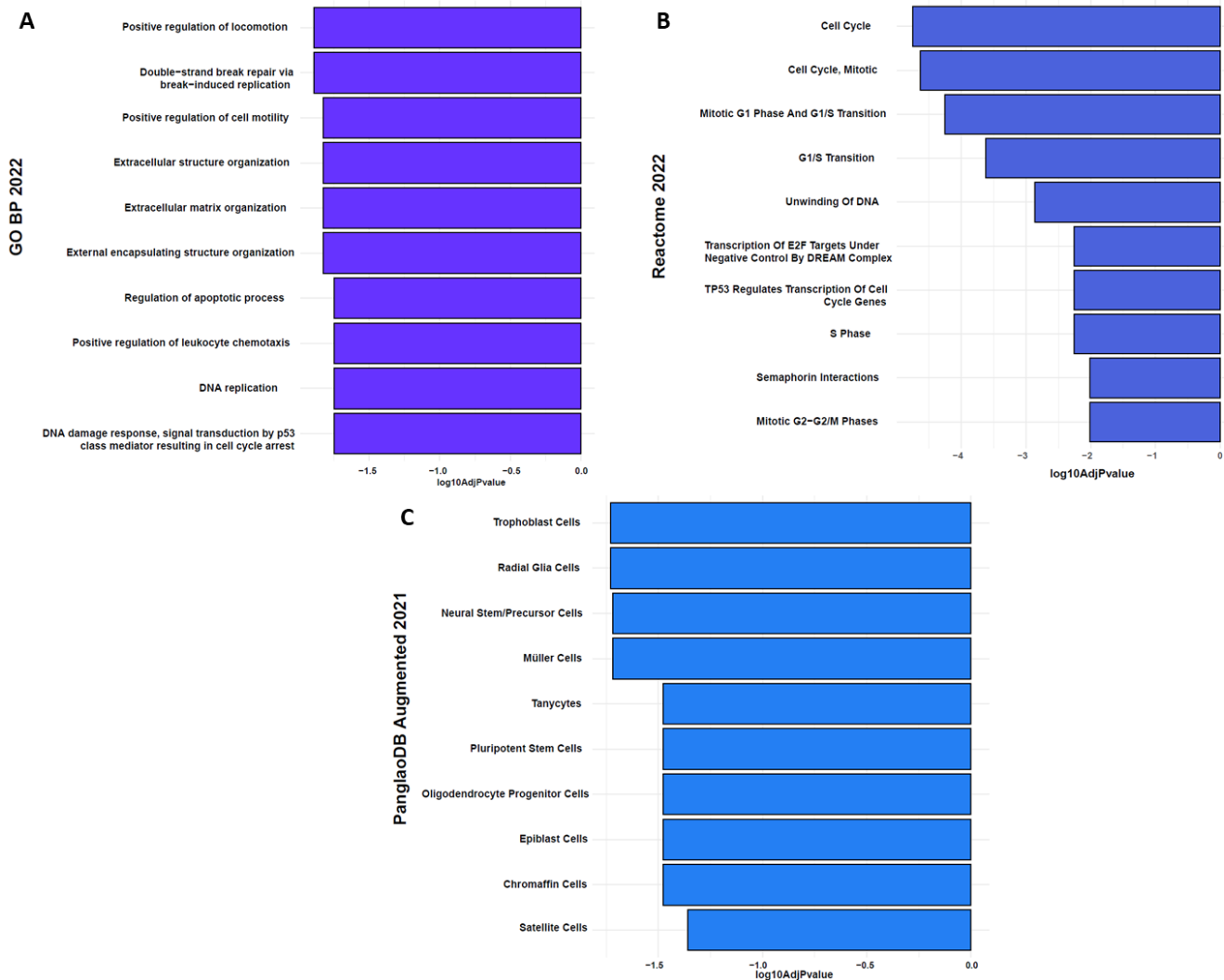
que tuvieron algún cambio significativo en el patrón de expresión a lo largo de todo el eje de diferenciación, en el caso número uno se encontraron un total de 92 genes con cambios significativos al final de la trayectoria (Anexo 1). En el caso de los genes que cambiaron en su patrón de expresión en general se encontraron 838 genes con una expresión diferencial a lo largo de su patrón de expresión durante el pseudotiempo (Anexo 2).

Se llevaron a cabo enriquecimientos usando los 92 genes con cambios diferenciales significativos al final de la trayectoria de diferenciación entre las condiciones H9 WT y H9 *SIRT1*<sup>-/-</sup> (Figura 47), se incluyeron 3 bases de datos: *Gene Ontology Biological Process (GO BP)*, *Reactome* y *PanglaoDB Augmented*. Estas tres bases de datos nos permiten conocer los procesos alterados por los genes encontrados (en el caso de GO BP y Reactome) y por su parte PanglaoDB Augmented otorga información sobre qué tipo celular se reconoce con el patrón de expresión de esos genes, esta última base de datos se integró al análisis debido a que fue diseñada y construida con datos provenientes de scRNAseq de múltiples tipos celulares humanos (Franzén et al., 2019).

En el caso de la base de datos *GO BP* (Figura 47 A) podemos observar que los procesos más enriquecidos se relacionan con organización celular, por ejemplo, los procesos de regulación de la locomoción celular, organización de la estructura extracelular y la organización de la estructura encapsulante externa, también podemos observar representados procesos relacionados con la replicación del DNA, así como con la ruptura de la doble hebra en dos condiciones, la primera causada por la inducción de la misma replicación y el segundo caso como respuesta al daño del DNA regulada por la transducción de señales de p53 resultando en arrestos del ciclo celular.

Por su parte los procesos enriquecidos encontrados en la base de datos de *Reactome* (Figura 47 B) están muy enfocados hacia el ciclo celular incluyendo la mitosis, la fase S, la transición entre fase G1 y S, las fases G2 y G2M así como la regulación de este proceso mediado por TP53. También podemos encontrar el proceso de desenrollamiento del DNA y las interacciones de semaforinas.





**Figura 47:** Análisis de enriquecimiento con genes diferenciales al final del pseudotiempo. Se muestran tres bases de datos con los resultados de enriquecimiento obtenidos tras incluir los 92 genes diferencialmente expresados al final de la trayectoria comparando las condiciones H9 WT y H9 SIRT1<sup>-/-</sup>: (A) Enriquecimientos obtenidos al consultar la base de datos Gene Ontology en el apartado de Biological Porcess; (B) Enriquecimientos obtenidos al utilizar la base de datos Reactome 2022; (C) Enriquecimientos de tipo celular obtenidos al consultar la base de datos Panglao DB.

Por último, al analizar las identidades celulares encontradas de *PanglaoDB* (Figura 47 C) nos damos cuenta que aparecen identidades de naturaleza troncal como son células del trofoblasto, células troncales pluripotentes y células del epiblasto, sin embargo también podemos encontrar identidades

relacionadas con la identidad hacia las que están siendo diferenciadas nuestras células H9 WT y H9 *SIRT1*<sup>-/-</sup> como son los precursores neurales, tanicitos y oligodendrocitos, debido a que son pertenecientes a linajes neurales humanos, es importante recordar que en el día 6 nuestras células poseen una identidad cercada a un progenitor neural pero que aún están lejos de tener una identidad neuronal madura definida, por lo que su programa de transcripción puede ligarse a diversos linajes terminales de este tipo celular, tomando cuenta lo anterior, estos grupos pueden mostrarnos diferencias causadas por nuestra delección de *SIRT1* y cómo afectaría la aparición de estos linajes.

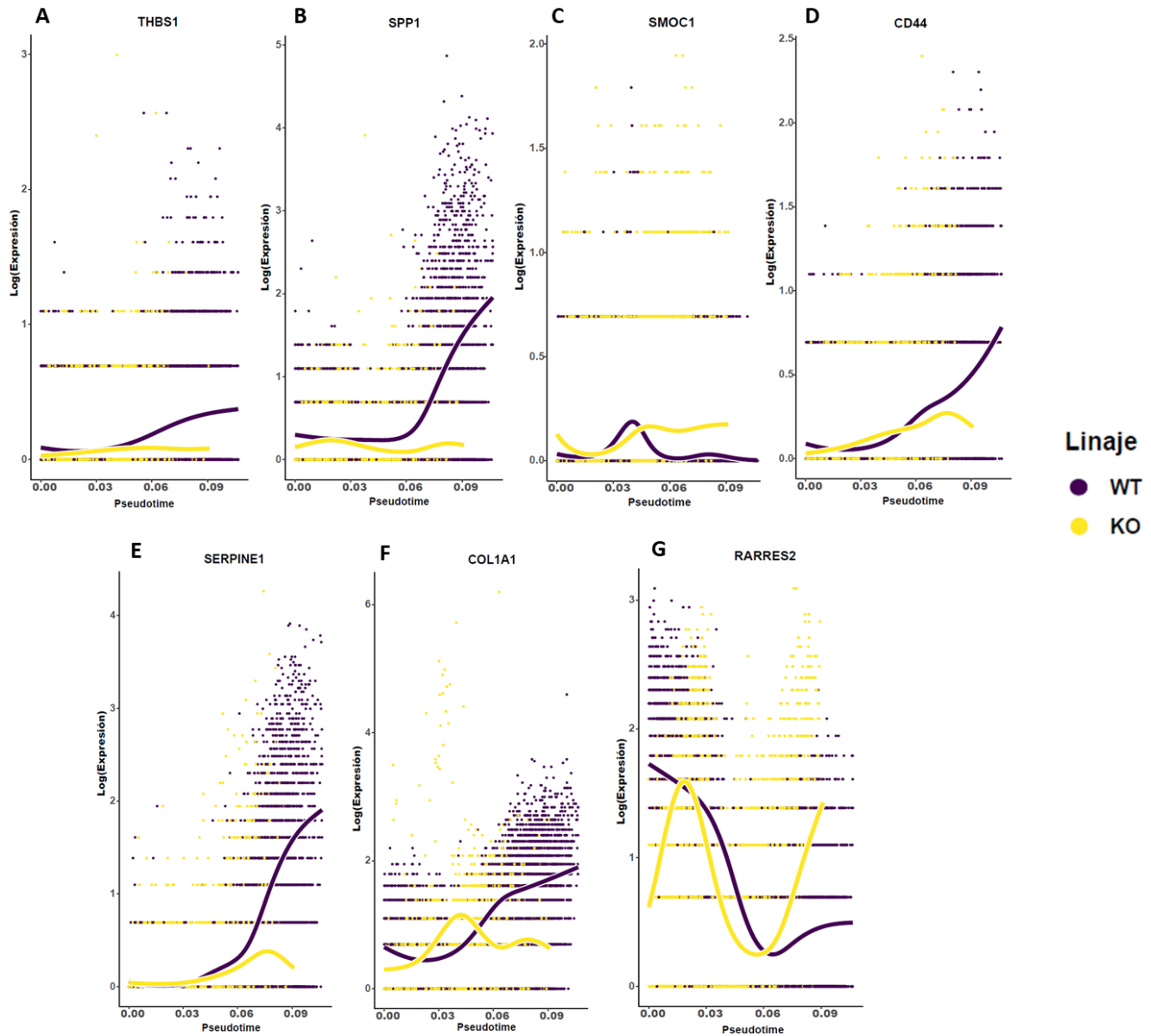
Una vez identificados los procesos enriquecidos, y para entender mejor los efectos de la delección de *SIRT1* en nuestro modelo, realizamos análisis de las diferencias de expresión a lo largo del pseudotiempo de los genes asociados a los diversos procesos e identidades celulares.

En el caso de los procesos enriquecidos en *GO BP* observamos que el proceso de “*Organización de la estructura extracelular*” contiene 5 genes, *THBS1*, *SPP1*, *SMOC1*, *CD44* y *SERPINE1* (Figura 48 A,B,C,D y E respectivamente), estos se comparten con el proceso de “*Organización de la matriz extracelular*” que contiene un séptimo gene siendo este *COL1A1* (Figura 48 F), de igual forma en el caso del proceso “*Regulación positiva de la locomoción*” se comparte el gen *THBS1* y se agrega el gen *RARRES2* (Figura 48 G).

De los genes anteriormente mencionados *THBS1*, *SPP1*, *CD44*, *SERPINE1* y *COL1A1* (Figura 44 A,B,D,E y F respectivamente) muestran una expresión que va en aumento conforme avanza el eje de diferenciación en la línea H9 WT, el comportamiento de estos genes en la línea celular H9 *SIRT1*<sup>-/-</sup> muestra la incapacidad de elevar la expresión de los genes manteniéndose en la mayoría debajo de lo observado en la línea H9 WT, en el caso de *CD44*, *SERPINE1* y *COL1A1* se puede observar que en la línea H9 *SIRT1*<sup>-/-</sup> la expresión incrementa alrededor del valor de pseudotiempo 0.06 sin embargo esta desciende llegando al final del pseudotiempo, en el valor 0.09, no logrando igualar la expresión observada en la condición H9 WT.

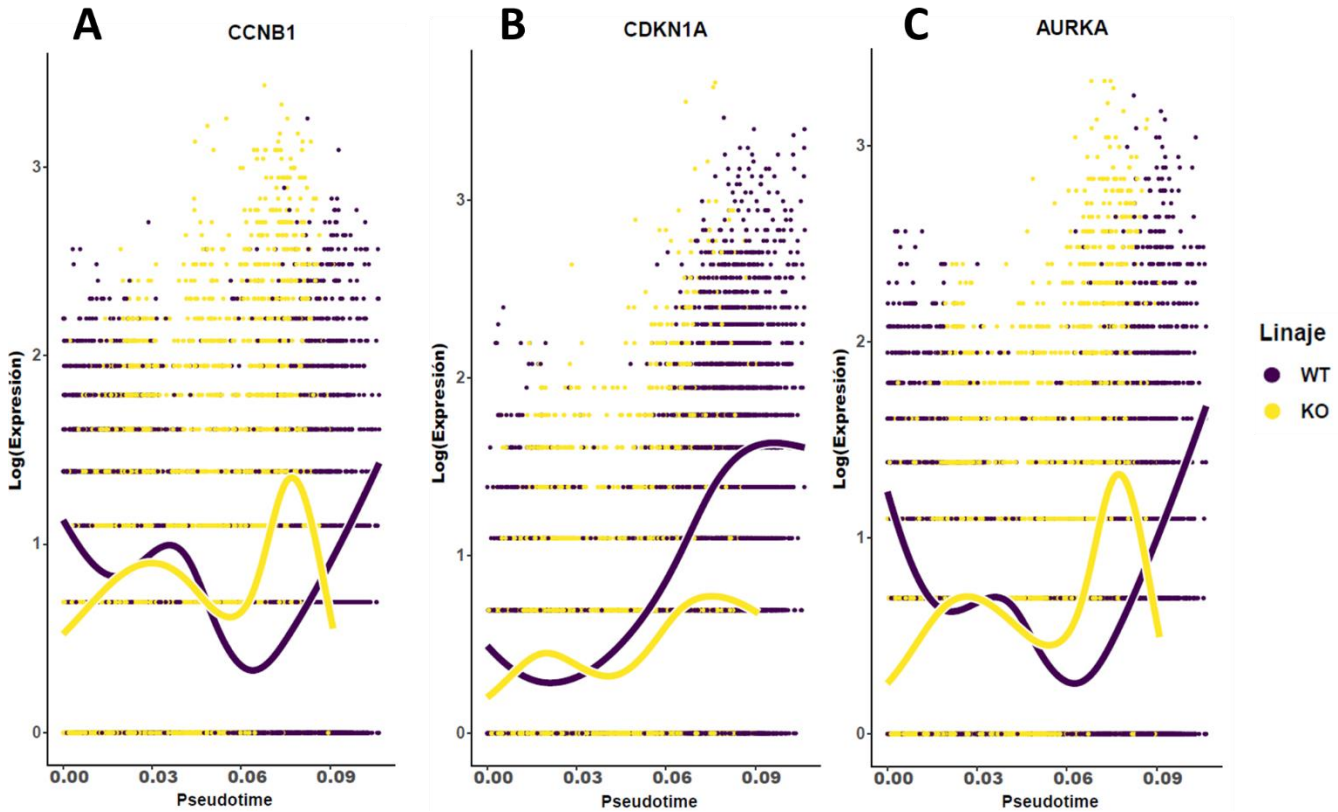
En el caso del gen *SMOC1* y *RARRES2* (Figura 48 C y G respectivamente) la condición H9 *SIRT1*<sup>-/-</sup> muestra un mayor nivel de expresión en comparación a la condición H9 WT al final del eje de diferenciación, el gen *SMOC1* muestra un aumento aproximadamente en el valor de pseudotiempo 0.03 en la línea H9 WT que posteriormente desciende en el valor 0.06 de pseudotiempo, sin embargo la línea H9 *SIRT1*<sup>-/-</sup> no regresa al estado basal de expresión observado en H9 WT, el gen *RARRES2* muestra un comportamiento similar ya que su expresión desciende casi a la par en ambas condiciones aproximadamente en el valor 0.06 de pseudotiempo, por su parte la condición H9 WT se mantiene en

un valor de expresión menor a 1 y cercano a 0.5 pero la línea H9 *SIRT1* muestra un aumento drástico en la expresión de este gen alcanzando un valor de 1.5 de expresión.



**Figura 48:** Genes involucrados en los procesos de enriquecimiento “Organización de la estructura extracelular”, “Organización de la matriz extracelular” y “Regulación positiva de la locomoción”. Expresión transformada logarítmicamente de los genes A) THBS1; B) SPP1; C) SMOC1; D) CD44; E) SERPINE1; F) COL1A1 y G) RARRES2. La línea morada representa la expresión en la condición H9 WT, la línea amarilla representa la expresión en la condición H9 *SIRT1*<sup>-/-</sup>.

Otro de los procesos de interés encontrados en el enriquecimiento usando la base *GO BP* fue la “Respuesta al daño del DNA regulada por la transducción de señales de p53 resultando en arrestos del ciclo celular” donde se agrupan tres genes: *CCNB1*, *CDKN1A* y *AURKA* (Figura 49 A, B y C respectivamente).



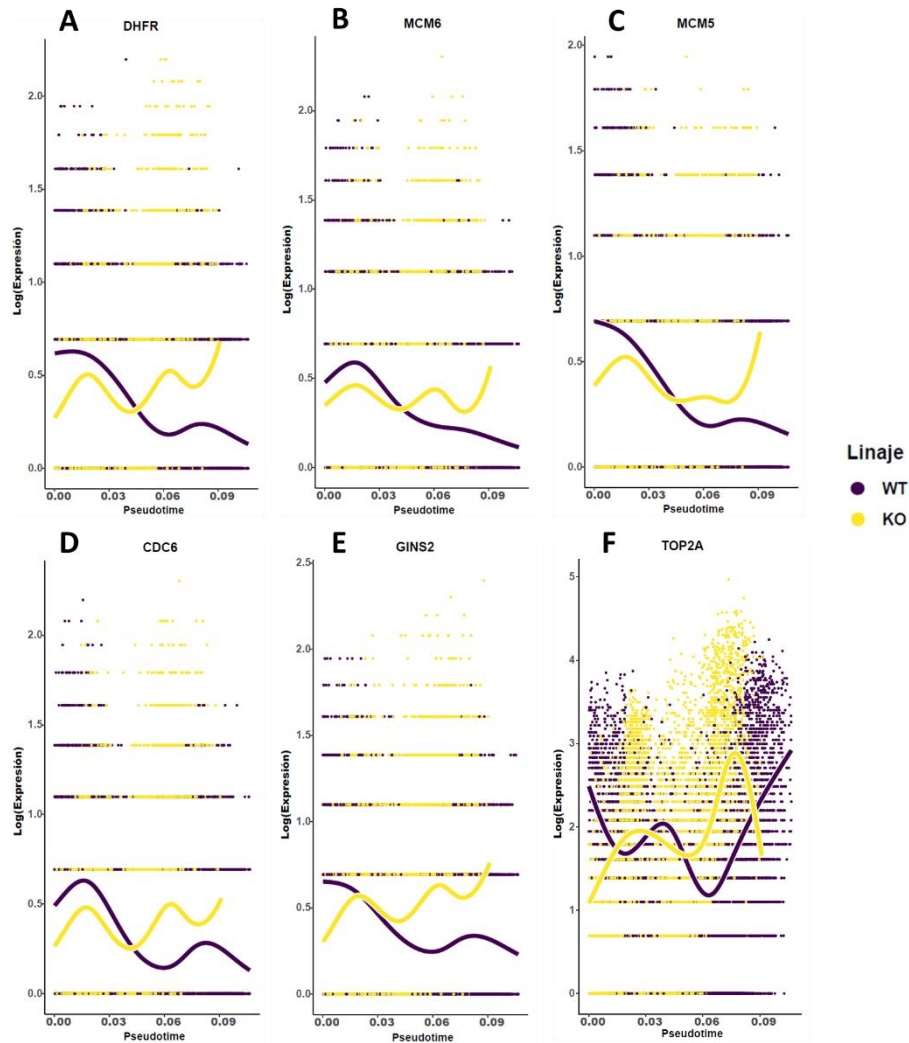
**Figura 49:** Genes involucrados en los procesos de enriquecimiento “Respuesta al daño del DNA regulada por la transducción de señales de p53”. Expresión transformada logarítmicamente de los genes A) *CCNB1*; B) *CDKN1A* y C) *AURKA*. La línea morada representa la expresión en la condición H9 WT, la línea amarilla representa la expresión en la condición H9 *SIRT1*<sup>-/-</sup>.

La expresión de los genes *CCNB1* y *AURKA* (Figura 49 A y C respectivamente) muestran un comportamiento de subida y bajada durante el eje de diferenciación en la condición H9 WT, ambos genes decrecen su expresión en el valor de pseudotiempo 0.03 hasta un valor de expresión cercano a 0.5, su expresión incrementa de nuevo al alcanzar el valor de pseudotiempo 0.06 superando el nivel de expresión inicial, la condición H9 *SIRT1*<sup>-/-</sup> muestra un comportamiento similar durante el eje de diferenciación pero ninguno de los dos genes reduce su expresión al mismo nivel que la condición H9 WT, además de lo anterior, ambos genes sufren una caída drástica en su expresión al llegar al pseudotiempo 0.07 quedando cerca del valor de expresión de 0.5.

El gen *CDKN1A* (Figura 49 B) muestra un aumento general y progresivo durante el eje de diferenciación en la condición H9 WT, el mismo comportamiento se observa en la condición H9 *SIRT1*<sup>-/-</sup>, sin embargo, su nivel de expresión no alcanza al observado en H9 WT.

En el caso de los procesos enriquecidos en la base de datos *Reactome* encontramos que los procesos relacionados con el ciclo celular se comparten con los encontrados en *GO BP* sin embargo los procesos “Ciclo celular, Mitosis”, “Fase G1 y transición G1/S durante la mitosis” y “Transición G1/S” contenían los genes *DHFR*, *MCM6*, *MCM5*, *CDC6*, *GINS2* y *TOP2A* (Figura 48 A, B, C, D, E y F respectivamente).

En el caso de estos genes podemos observar que los genes *DHFR*, *MCM6*, *MCM5*, *CDC6* y *GINS2* (Figura 50 A, B, C, D y E respectivamente) muestran una expresión que decrece a lo largo del pseudotiempo del eje de diferenciación en la condición H9 WT, en la condición H9 *SIRT1*<sup>-/-</sup> estos genes muestran una oscilación donde decrecen y posteriormente aumenta su expresión superando significativamente la observada en H9 WT, en el caso del gen *TOP2A* se observó que la condición H9 WT decrece su expresión a partir del valor de pseudotiempo 0.03 y esta comienza a aumentar aproximadamente en el valor de pseudotiempo 0.06, las células de la condición H9 *SIRT1*<sup>-/-</sup> muestran un comportamiento similar, sin embargo, su expresión deja de aumentar y comienza a decrecer de forma rápida cerca del final del pseudotiempo de diferenciación (valor 0.09).



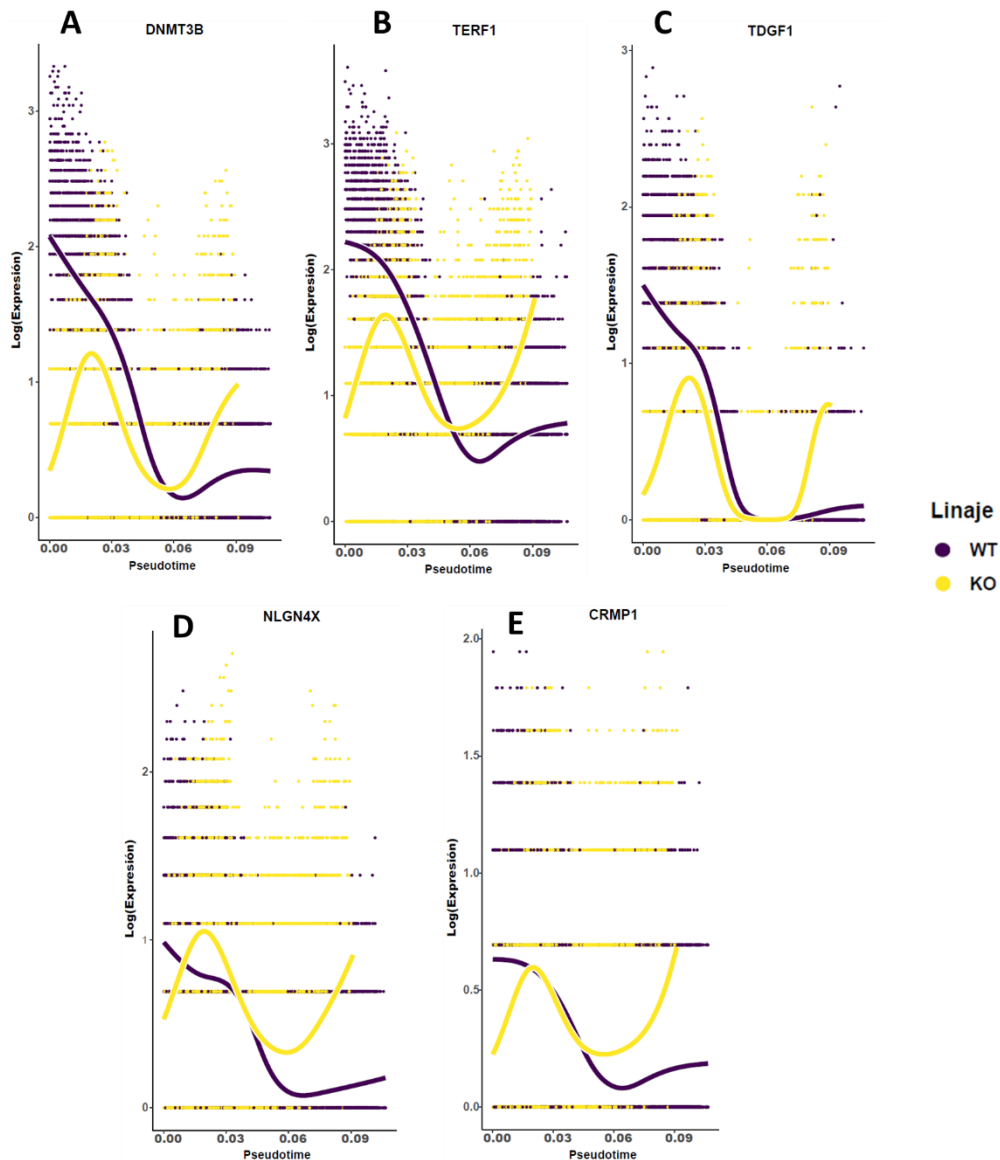
**Figura 50:** Genes involucrados en los procesos de enriquecimiento “Ciclo Celular, Mitosis”, “Fase G1 y transición G1/S” y “Transición G1/S”. Expresión transformada logarítmicamente de los genes A) DHFR; B) MCM6; C) MCM5; D) CDC6; E) GINS2 y D) TOP2A. La línea amarilla representa la expresión en la condición H9 y la línea morada la expresión en la condición H9 SIRT1<sup>-/-</sup>.

Por último, se analizó la expresión de genes presentes en las identidades celulares encontradas en el enriquecimiento utilizando la base de datos *PangloaDB Augmented*, las identidades de interés para este trabajo pueden ser definidas en dos categorías: Identidades de pluripotencia y troncalidad (“Células troncales pluripotentes” y “Células del epiblasto”) e identidades neuronales (“Células troncales Neuronales/Precursores neuronales”, “Tanicitos”, y “glía Radial”).

Los genes encontrados relacionados con la identidad de “Células troncales pluripotentes” son *DNMT3B*, *TERF1* y *TDGF1* (Figura 51 A, B y C respectivamente), de igual forma se encontró asociado a esta identidad el gen *CDC6* (Figura 50 D) previamente asociado con el ciclo celular y la transición de

la fase G1/S; en el caso de la identidad “*Células del epiblasto*” se encontró que los genes *DNMT3B* y *TDGF1* se compartían con la identidad de “*Células troncales pluripotentes*”, sin embargo, también se encontró que los genes *NLGN4X* y *CRMP1* se asociaban a una identidad celular del epiblasto (Figura 51 D y E respectivamente).

Los 5 genes nuevos encontrados en estas dos identidades asociadas a troncalidad y pluripotencia muestran un comportamiento de decrecimiento en la condición H9 WT, mientras que en las células de la condición H9 *SIRT1* /- muestran una disminución en su expresión previo al valor de pseudotiempo 0.06, sin embargo, los 5 genes muestran un aumento significativo hacía el final del eje de diferenciación (pseudotiempo 0.08 y 0.09).

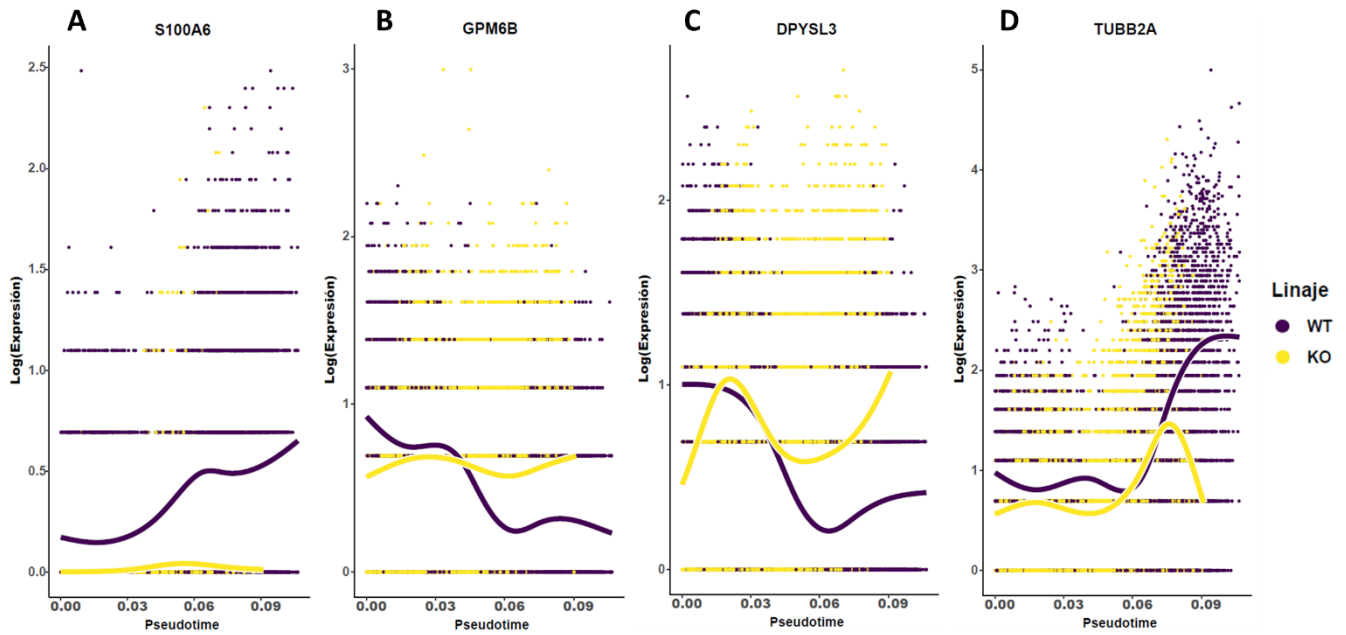




**Figura 51:** Genes involucrados en las identidades celulares “Células troncales pluripotentes” y “Células del epiblasto”. Expresión transformada logarítmicamente de los genes A) DNMT3B; B) TERF1; C) TDGF1; D) NLGN4X y E) CRMP. La línea morada representa la expresión en la condición H9 WT, línea amarilla representa la expresión en la condición H9 SIRT1  $\text{-/-}$ .

En el caso de las identidades asociadas a linajes neurales encontramos que las tres “Células troncales Neuronales/Precursores neuronales”, “Tanicitos”, y “Glía Radial” comparten los genes marcadores S100A6, GPM6B, DPYSL3 y TUBB2A (Figura 50 A, B, C y D respectivamente).

En el caso de los genes TUBB2A y S100A6 podemos observar un aumento gradual en la expresión de ambos genes a lo largo del eje de diferenciación de la condición H9 WT, mientras que la condición H9 SIRT1  $\text{-/-}$  el gen S100A6 nunca logra elevar su expresión a lo largo del eje de diferenciación, por su parte TUBB2A muestra una expresión similar a la observada en H9 WT hasta el valor de pseudotiempo 0.07 donde comienza a disminuir significativamente. Los genes GPM6B y DPYSL3 muestran un comportamiento contrario siendo que en la condición H9 WT ambos genes decrecen su expresión conforme avanza el eje de diferenciación mientras que en la condición H9 SIRT1  $\text{-/-}$  se aprecia que el gen DPYSL3 disminuye su expresión aproximadamente en el valor de pseudotiempo 0.04 pero su expresión vuelve a aumentar a partir del valor de pseudotiempo 0.05, quedando caso al mismo nivel de inicio del pseudotiempo, por su parte el gen GPM6B no muestra cambios drásticos en su expresión durante el eje de diferenciación, manteniendo siempre un valor cercano al observado en el inicio del pseudotiempo.



**Figura 52:** Genes involucrados en las identidades celulares “Células troncales Neuronales/Precursores neuronales”, “Tanicitos”, y “Glía Radial”. Expresión transformada logarítmicamente de los genes A) S100A6; B) GPM6B; C) DPYSL3 y D) TUBB2A. La línea morada representa la expresión en la condición H9 WT, la línea amarilla representa la expresión en la condición H9 SIRT1<sup>-/-</sup>.

## Discusión

Este trabajo tenía como finalidad el estudio de los cambios generados en una línea celular de hESCs tras la delección de la deacetilasa de histonas SIRT1, evaluando estos cambios en su estado basal y durante ejes de diferenciación definidos aquí como procesos dinámicos, utilizando como técnica principal el scRNA-seq y los análisis bioinformáticos que se realizaron a partir de los datos obtenidos de este experimento.

En primer lugar, se evaluó la importancia de SIRT1 en dos procesos diferenciación celular, esto derivado de los antecedentes específicos de este proyecto donde se encontró que las células H9 *SIRT1*<sup>-/-</sup> mostraban heterogeneidad en la expresión de marcadores de troncalidad como OCT4 así como la presencia de marcadores de otros linajes embrionarios en etapas de diferenciación más avanzadas (por ejemplo linajes ectodérmicos); ya que previamente ha sido reportado que los cambios en la expresión y regulación de OCT4 son suficientes para la pérdida de la troncalidad en hESCs (Shi & Jin, 2010).

En el caso del protocolo de diferenciación adipogénica se tenían reportes previos que mostraban que SIRT1 inhibía la adipogénesis en células troncales mesenquimales (MSCs) (Li et al., 2011; Puri et al., 2012; Zhou et al., 2015); sin embargo la mayoría de estos estudios parten de un estado de diferenciación multipotente, como el caso de las MSCs, esto derivó en la idea de evaluar mediante tinciones de ORO la capacidad de nuestra línea H9 *SIRT1*<sup>ΔEx4</sup> la eficiencia de diferenciación hacia adipocitos comparado con las células H9 WT, los resultados mostraron que las células H9 *SIRT1*<sup>ΔEx4</sup> eran capaces de formar adipocitos sin embargo no con la misma velocidad ni cantidad que la condición WT, ya que estos eran apenas visibles durante el día 10 de la diferenciación, en contraste con la condición WT donde se apreciaban células teñidas abundantes desde el día 5 del protocolo. La primera razón causante de este fenómeno que llega a nuestra mente es que las células H9 *SIRT1*<sup>ΔEx4</sup> han visto disminuida su capacidad de controlar el *switch* metabólico necesario durante la adquisición de un fenotipo más diferenciado, donde las células troncales cambian de un estado glucolítico a uno oxidativo conforme adquieren un linaje más específico (Folmes et al., 2012), otra razón evidente causante de este fenómeno es la alteración de la expresión de genes relacionados con la adquisición de linaje adipogénico en estas células, esto se logró comprobar por (Sánchez-Ramírez et al.) quienes llevaron a cabo un análisis de expresión utilizando bulk-RNAseq en la misma línea H9 *SIRT1*<sup>ΔEx4</sup> y encontraron ~5000 DGE durante el día 8 de diferenciación, de los cuales <60 se encontraban alterados significativamente, estos genes permitieron definir que aún con niveles bajos de SIRT1 las células podían avanzar durante el protocolo de diferenciación pero tenían dificultades en la adquisición de un linaje de adipocito maduro (Sánchez-Ramírez et al., 2022).

En el caso del segundo protocolo de diferenciación se eligió utilizar la línea celular H9 *SIRT1*<sup>-/-</sup> para comprobar si la ausencia total de SIRT1 tenía un efecto similar en un eje de diferenciación, lo observado fue contrastante debido a que las células H9 *SIRT1*<sup>-/-</sup> no solo no podían diferenciarse si no que estas morían (se despegaban de la matriz de Geltrex de la caja de cultivo) alrededor del día 14 a 16, esto coincidía con la etapa de expansión del protocolo de (Fedele et al., 2017) donde las células aún conservaban una identidad de progenitor neural, además algo que se pudo observar en nuestras fotografías tomadas durante el protocolo fue que las células H9 *SIRT1*<sup>-/-</sup> mostraban una respuesta nula o muy poco notable al medio enriquecido con moléculas pequeñas para su diferenciación, un ejemplo claro de esto es el crecimiento en colonias que se observa constantemente en los pozos de cultivo de la línea H9 *SIRT1*<sup>-/-</sup> incluso en etapas tan tardías como el día 5, es importante recalcar que este crecimiento colonial es una característica principal de las hESCs cultivadas in vitro (Wadkin et al., 2019); por el contrario en las células H9 WT se puede notar que las células rompen la estructura de las colonias y se expanden cubriendo como monocapa toda la placa de crecimiento.

Por lo tanto, podemos proponer que una cantidad pequeña de SIRT1 (como la presente en la línea H9 *SIRT1*<sup>ΔEx4</sup>) es suficiente para que las células sean viables a lo largo de un proceso de diferenciación pero puede afectar su capacidad de maduración hacia linajes definidos como los adipocitos (Sánchez-Ramírez et al., 2022), por otro lado, la ausencia de SIRT1 vuelve a las células no viables, generando su muerte cuando se les somete a un proceso de diferenciación, además es notorio que la ausencia de SIRT1 genera una falta de respuesta a estímulos de diferenciación, manteniendo comportamientos similares a los presentes en una hESCs, sin embargo se requiere llevar a cabo más experimentos para poder comprobar esta idea con una aproximación más experimental y menos bioinformática.

Otro de los aportes de este trabajo es que la delección de SIRT1 puede causar por sí misma heterogeneidad en una población células de hESCs, en primera instancia este es el fundamento central del trabajo y la razón por la que se decidió utilizar un scRNA-seq como experimento central. El hallazgo de identidades como MSCs durante el proceso de anotación de los datos de las células H9 *SIRT1*<sup>-/-</sup> en el día 0 es prueba de lo anterior, es importante recordar que este tipo celular no se encuentra siquiera en el mismo nivel de potencial de diferenciación de una hESCs, que están sesgadas hacia un linaje mesodérmico y que además son células multipotentes (Tonk et al., 2020), la presencia de esta población era algo consistente ya que aun cuando se decidió ajustar los criterios de calidad para ser más estrictos la población de MSCs permanecía presente y detectable tras el proceso de anotación de identidad.

Otra diferencia asociable a el silenciamiento de SIRT1 es la distribución de identidades celulares dentro de nuestros datos; para la condición H9 WT pudimos observar que el porcentaje de células

asignadas a la identidad ESCs es del 95% en los criterios de datos laxos y del 94.2% en los criterios de calidad estrictos, sin embargo, estos porcentajes son considerablemente menores en ambos criterios de calidad de la condición H9 *SIRT1*<sup>-/-</sup> donde se observó que solo el 51% de la población en los criterios de calidad laxos se asociaban con la identidad ESCs y el solo 27% en el caso de los criterios de calidad estrictos.

Por otro lado, en las células H9 *SIRT1*<sup>-/-</sup> la segunda población más abundante en los criterios de calidad laxos fueron las *iPSc* con el 33% de la población, mientras que en los criterios de calidad estrictos se encontró que el 73% de la población se asignaba a la identidad *iPSc*. En contraste en la condición H9 WT los criterios de calidad laxos poseen menos del 1% de las células asignadas a la identidad *iPSc*, mientras que los criterios de calidad estrictos no poseen ninguna célula asignada a esta identidad.

Todo lo anterior ayuda a reafirmar nuestra propuesta de que la delección de *SIRT1* genera por sí mismo heterogeneidad dentro de la población H9 *SIRT1*<sup>-/-</sup>, aun cuando estas se encuentran en su estado basal de mantenimiento (Día 0 con medio de cultivo de mantenimiento), es importante remarcar que aun cuando las células *iPSc* poseen capacidades similares como la pluripotencialidad de una hESC, éstas no son células troncales humanas. Las células pluripotenciales inducidas, o *iPSc*, como dice su nombre provienen de células somáticas que son “reprogramadas” mediante la inducción de la expresión de marcadores troncales como son *Oct4*, *Sox2*, *Klf4* y *c-Myc*, originalmente esta reprogramación fue desarrollada por Yamanaka y sus colegas (Takahashi & Yamanaka, 2006), la expresión de estos genes producen células con capacidad de dar origen a las capas germinales del embrión (pluripotencialidad) y una capacidad de autorrenovación muy similar al de las hESCs. Aun cuando las células *iPSCs* son muy similares a una hESC se ha reportado que estas poseen características como la memoria somática, esta última a causa de la retención de marcas epigenéticas conservadas de su linaje somático, estas marcas pueden desembocar en la limitación de la capacidad de diferenciación de este tipo celular, generando un sesgo de preferencia hacia diferenciarse al linaje de origen previo a la reprogramación (Hu et al., 2010).

Otra de las características diferenciadoras encontradas reside en la programación metabólica ya que las células somáticas maduras dependen fuertemente de la fosforilación oxidativa como fuente de obtención de ATP principal, por su parte las células troncales poseen un metabolismo basado principalmente en la glucólisis como fuente de ATP, tomando en cuenta lo anterior se ha reportado que las *iPSc* no sufren una reprogramación completa a nivel metabólico teniendo fenotipos mitocondriales mixtos, mezclando mitocondrias en estado de inmadurez y mitocondrias con fenotipo de maduración terminal, mientras que en la hESC estas se encuentran completamente inmaduras (Varum et al.,

2011), otra de las diferencias encontradas y que sugieren que las *iPSc* conservan rasgos de su linaje somático de origen se reportó cuando se comparó el perfil metabólico de células hESCs con el obtenido de células *iPSc* producidas a partir de queratinocitos y fibroblastos humanos (Panopoulos et al., 2012).

Dejando de lado lo anterior es importante también mencionar que el proceso de anotación de identidad celular en datos de scRNA-seq es un proceso con múltiples opciones y que en el caso particular de este trabajo se decidió utilizar la metodología de *SingleR* (Aran et al., 2019) que está basada en una metodología de correlación (Pasquini et al., 2021) para llevar a cabo este proceso, la identificación se hace a nivel de cluster y depende además del set de genes sobre el que se anota (por ejemplo todos los genes de la célula, un set de genes random o los genes altamente variables); una limitación enorme de la anotación por correlación es la necesidad de una referencia de tipos celulares o atlas de identidades celulares que permitan obtener los coeficientes de correlación, si no existe una referencia semejante, este tipo de metodologías no puede ser utilizado. Conocer las limitaciones y características de esta metodología previamente mencionada nos lleva a proponer la exploración de otros métodos de anotación como pueden ser los basados en clasificación supervisada (Pasquini et al., 2021).

El siguiente paso en este trabajo fue entonces buscar los cambios de expresión en genes que podían estar dirigiendo y causando estas diferencias, para hacerlo de forma funcional se decidió utilizar el análisis de enriquecimiento usando los DEG.

En el caso del día 0 ambos criterios de calidad mostraron procesos enriquecidos muy similares, lo que favorece la idea de que los cambios son robustos y replicables aun cuando los criterios se vuelven más astringentes. Uno de los cambios más importantes generados tras la delección de *SIRT1* en el día 0 se ve reflejado en el metabolismo de las células H9 *SIRT1*<sup>-/-</sup>, ya que se enriquece el metabolismo de lípidos, biosíntesis de colesterol, metabolismo de esteroides y la activación de la expresión de genes mediada por *SREBF*, también a la par aparecen enriquecidos procesos relacionados con el metabolismo de lípidos, glucólisis canónica y catabolismo de carbohidratos.

Como se mencionó en la introducción, el metabolismo celular es una de las características esenciales de la hESCs ya que se ha demostrado en diversos estudios que un ambiente glucolítico es necesario para mantener las características de troncalidad, autorrenovación y pluripotencia en este tipo celular, la reducción de esta vía metabólica mediante la inhibición del transportador de lactato (Gu et al., 2016) o de la enzima hexoquinasa (Moussaieff et al., 2015) resultan en diferenciación espontánea de las hESCs. La aparición de *SREBP* y procesos relacionados con su expresión justifican la presencia de otros procesos como la síntesis de lípidos y colesterol, ya que este factor de transcripción es el

regulador central de las vías de síntesis de colesterol y ácidos grasos insaturados (DeBose-Boyd & Ye, 2018), además se ha reportado previamente que *SIRT1* tiene un papel en la inhibición de la actividad de elementos reguladores de *SREBP*, por ejemplo se sabe que SIRT1 inhibe de forma directa la síntesis de lípidos mediada por *SREBP-1c* acetilando este factor de transcripción en células hepáticas (Ponugoti et al., 2010); lo anterior nos lleva a pensar que la línea celular H9 *SIRT1*<sup>-/-</sup> se ve afectada a nivel de regulación de la actividad de los miembros de *SREBP* generando así la activación de vías de síntesis de ácidos grasos y lípidos.

Por su parte las consecuencias del inicio de una síntesis de ácidos grasos de forma anormal puede aportar a la pérdida de troncalidad, la pérdida de autorrenovación e incluso la muerte celular (Luchi et al., 2019); se ha reportado en modelos de células troncales como hMSCs que la acumulación de ácidos grasos como el butirato afectan la heterogeneidad funcional de estas células además de suprimir su capacidad de autorrenovación y potencial de diferenciación hacia adipocitos y osteocitos mediante un incremento en la apoptosis durante el proceso (Xie et al., 2021); lo reportado por estos autores coincide con lo observado en nuestro modelo, donde hay afecciones en la eficiencia de diferenciación (en el caso de la línea H9 *SIRT1*<sup>ΔEx4</sup>) o definitivamente se observa muerte celular en ausencia de SIRT1 durante la diferenciación (en el caso de la línea H9 *SIRT1*<sup>-/-</sup>).

Posteriormente, tras analizar los procesos enriquecidos usando los genes regulados a la baja en la condición H9 *SIRT1*<sup>-/-</sup>, encontramos que la mayoría de los procesos se relacionan con la traducción de mRNA, la biosíntesis de péptidos, la elongación de la cadena peptídica, procesamiento de RNA ribosomal y el inicio de la transcripción. La traducción es uno de los procesos más regulados en las hESCs ya que es esencial en el mantenimiento de la pluripotencialidad, se ha reportado que las hESCs poseen una baja actividad global de síntesis de proteínas en comparación con células en procesos de diferenciación y células diferenciadas (Saba et al., 2021), estudios en modelos murinos utilizando células troncales embrionarias (mESCs) demostraron, mediante la medición de los cambios en la síntesis de proteínas nacientes utilizando etiquetado con [<sup>35</sup>S] metionina a lo largo de 5 días de diferenciación, demostraron un aumento de aproximadamente el doble entre la condición inicial mESCs y los cuerpos embrioides formados tras 5 días de diferenciación, formando un eje de incremento global de la traducción celular conforme se adquiere un nivel avanzado de diferenciación (Sampath et al., 2008).

El control del nivel global de traducción se da nivel de abundancia de ribosomas en las células, así como el procesamiento de los rRNA, se ha asociado que durante las etapas tempranas de diferenciación ocurre un aumento en la biogénesis de ribosomas que posteriormente decrece cuando se alcanzan estadios de maduración avanzados (Ingolia et al., 2011).

Se ha reportado que SIRT1 juega un papel importante en la modulación de la biogénesis ribosomal, la vía más conocida de esto es a través del control de las vías de estrés nutricional donde su cofactor *NAD+* incrementa, de igual forma la *SIRT1* puede desacetilar factores de transcripción como son TAF<sub>68</sub> reduciendo significativamente la transcripción del rDNA y suprimiendo la biogénesis ribosomal (Muth et al., 2001), sin embargo nuestra delección de SIRT1 parece generar el mismo efecto en la biogénesis ribosomal, decreciendo su actividad y llevando los niveles de traducción células aún más comparado con la condición H9 WT.

El análisis de enriquecimiento de ambas condiciones en el día 6 arrojó resultados similares a los encontrados en el día 0, recapitulando, los DEG a la alta en la condición H9 *SIRT1*<sup>-/-</sup> generaban enriquecimientos con procesos relacionados a la biosíntesis de lípidos, la activación de vías relacionadas con *SREBPF* y la expresión de genes objetivos de estos TF. Por su parte los DEG a la baja en la misma condición se relacionaban con procesos biológicos como elongación durante la traducción, traducción en general, procesamiento de rRNA e iniciación de la traducción, reforzando que dos de los efectos de la delección de *SIRT1* más grandes fueron la afección del control de la síntesis de lípidos y los niveles de traducción dentro de la célula, sin embargo, se requiere explorar aún más estas propuestas para brindar un mecanismo más preciso.

Lo anterior nos permite afirmar que la delección de SIRT1 además de causar pérdida de control en la regulación de las vías de síntesis de lípidos genero también un decrecimiento en la actividad de la traducción celular y los ribosomas, esto último más evidente en el día 6 de diferenciación donde de acuerdo a lo reportado, se esperaría un aumento en la actividad global de traducción y actividad ribosomal, sin embargo se observa que estos procesos continúan regulados a la baja en la condición H9 *SIRT1*<sup>-/-</sup> lo que podría explicar su incapacidad de reaccionar al estímulo de diferenciación y posterior muerte, observado durante el protocolo de diferenciación a linajes mesodérmicos (adipocitos) y ectodérmicos (progenitores neurales) a los que fueron sometidos durante este trabajo.

En el caso de los TFs al alta en la condición *SIRT1*<sup>-/-</sup> en el día 0 volvemos a encontrar miembros de la familia de *SREBP* como son *SREBF2* y *SREBF1*, sin embargo algo llamativo encontrado en estos datos es el incremento en la actividad de TF como *HIF1A* ya que ambos se relacionan con la respuesta a los niveles bajos de oxígeno o los procesos de hipoxia celular, en el caso particular de *HIF1A* se ha reportado que es responsable de activar la expresión de genes relacionados con transportadores de glucosa, enzimas involucrada en la glucólisis (Lum et al., 2007).

El caso particular de estos TFs refleja una vez más el efecto sobre el metabolismo celular producido por la delección de SIRT1, recordemos que en las células troncales es esencial el mantenimiento de un



metabolismo basado en glucólisis y condiciones de hipoxia celular para favorecer la dependencia metabólica de estas vías, sin embargo se observa que la delección genera la activación de algunos de los TFs maestros en la regulación de la lipogénesis y biosíntesis de colesterol, en contraste, el aumento en los factores de hipoxia podría ser un mecanismo de respuesta por parte de la célula en una búsqueda de mantener la hipoxia celular y por ende la troncalidad.

Pasando ahora a los TFs que pierden actividad en nuestro modelo *SIRT1*<sup>-/-</sup> encontramos el factor *NR2C2*, que controla procesos como son la homeostasis celular, la diferenciación y el desarrollo, también es esencial en la supresión de la señalización promovida por ácido retinoico, su silenciamiento aumenta la actividad de esta vía e inhibe la diferenciación de células troncales hacia linajes mesenquimales, específicamente osteoblastos y adipocitos (Green et al., 2017), también se ha reportado que su actividad es esencial en el desarrollo del cerebelo (Chen et al., 2007).

Por último, el factor *E2F4* es un factor de transcripción que reprime la expresión de genes relacionados con el ciclo celular, específicamente durante la etapa G0 y G1, esta represión la hace en conjunto con proteínas de la familia *RB* (Verona et al., 1997). A pesar de lo anterior mencionado, se ha encontrado que en otros contextos celulares *E2F4* puede actuar como un activador de genes relacionados con proliferación, por ejemplo, se comprobó que en durante el desarrollo del corazón de ratón los cardiomiocitos presentaban un decrecimiento en la expresión de *E2F4*, al inducir la proliferación postnatal de estas células los niveles de *E2F4* aumentaban de forma recíproca (van Amerongen et al., 2009).

En el caso de los TFs con menor actividad inferida encontrados al comparar las condiciones H9 WT y H9 *SIRT1*<sup>-/-</sup> en el día 6 del protocolo de diferenciación a NPC, encontramos a los TFs *SMAD4* y *SMAD3* que son componentes esenciales en la transmisión de la cascada de señalización de *TGF-β* hacia el núcleo celular, *SMAD4* es un factor de transcripción esencial en diversos procesos de desarrollo y diferenciación celular, un ejemplo de esto es su papel como efector crucial en la respuesta al estímulo de *NODAL* que a su vez es esencial en la especificación del mesodermo y endodermo embrionario, así como en el establecimiento del eje derecho/izquierdo (Robertson, 2014); por otra parte *SMAD4* también es esencial en la respuesta al estímulo por BMP para el establecimiento del linaje ectodérmico del embrión y el establecimiento del eje dorsoventral (Bier & De Robertis, 2015). Por su parte *SMAD3* es también un efector esencial en la transmisión de la cascada de *TGF-β* que permite responder a estímulos de diferenciación y proliferación celular, además de lo anterior se ha reportado que puede formar un complejo *SMAD3/SMAD4/JUN/FOS* para regular la transcripción mediata por *TGF-β* (Zhang et al., 1998).

En el caso de los factores de transcripción con un incremento de actividad en nuestro modelo *SIRT1*<sup>-/-</sup> día 6 volvemos a encontrar a *SREBF1*, la constante presencia al alta de este factor nos muestra que el cambio en el metabolismo de las células *SIRT1*<sup>-/-</sup> permanece aun cuando las células son estimuladas con diversas moléculas hacía un linaje neuronal.

Por último tenemos el TF *FOXA2* que es miembro de la familia de factores de transcripción *FOXA* tiene un papel importante en el establecimiento de las capas germinales, ya que se ha reportado que modelos de ratón con una mutación en el locus de *FOXA2* mueren en el día 10 de desarrollo mostrando anomalías en la formación de las tres capas germinales (Weinstein et al., 1994); también se ha observado que los ratones heterocigotos para *FOXA2* logran sobrevivir y desarrollarse hasta su vejez donde comienzan a presentar fenotipos similares al Parkinson (Kittappa et al., 2007).

Por último, este trabajo se enfocó en tratar de comprender toda la información previamente descrita durante un fenómeno dinámico, es decir, el proceso de diferenciación hacía progenitores neurales llevado a cabo en ambas líneas celulares, H9 WT y H9 *SIRT1*<sup>-/-</sup>. El primer paso de este proceso consistió en la búsqueda de una metodología de colapso de dimensiones más eficiente que el UMAP por lo que se eligió *PHATE*, este paso es todo menos trivial debido a que se ha reportado de forma amplia que la multidimensionalidad de los datos generados por omics de célula sencilla es un desafío difícil de superar, especialmente por la siempre presente “maldición de la dimensionalidad” (Altman & Krzywinski, 2018), en general muchas de las técnicas de reducción de dimensiones generan una pérdida masiva de información y por sí mismas toman supuestos con poco sentido como la posibilidad de generar representaciones en dos dimensiones a partir de métricas como la distancia Euclidiana en espacios multidimensionales donde es imposible dicha premisa (Chari & Pachter, 2022), es por lo anterior que la exploración de múltiples métodos de reducción de dimensiones fue esencial para poder llegar a aquella que nos brindara la mayor cantidad de información biológica contenida en nuestros datos.

En el caso de nuestro protocolo de diferenciación se esperaba obtener un solo linaje que partiera de la identidad *ESCs* del día 0 hacía las células *NE* del día 6, este comportamiento fue observado en la condición H9 WT donde a través de *PHATE* encontramos a las células distribuidas a lo largo del pseudotiempo en una estructura similar a un arco invertido. La condición H9 *SIRT1*<sup>-/-</sup> mostro un comportamiento completamente distinto debido a que se encontraron dos linajes dentro de la representación *PHATE*, uno de ellos correspondiente a las células *ESCs* del día 0 y su heterogeneidad interpoblacional, mientras que el segundo linaje partía desde un grupo de células *ESCs* día 0 en el centro de la estructura hacía las células *iPSc* del día 6.

Al profundizar en los procesos enriquecidos por los genes diferencialmente expresados al final de la trayectoria de diferenciación pudimos encontrar que en general se asociaban a los siguientes procesos: cambios en la estructura extracelular y organización de la matriz extracelular, ciclo celular y su regulación, respuesta a daño mediada por p53 y la incapacidad de la pérdida de marcadores de pluripotencia.

En el caso de los procesos relacionados con la organización de la matriz extracelular podemos comentar que ha sido ampliamente reportado que esta tiene una fuerte influencia en el comportamiento celular durante la diferenciación de células troncales (Noh et al., 2023), un ejemplo más específico de esto es que la formación de una matriz extracelular “suave” lleva a las MSCs hacia un linaje adipogénico mientras que una matriz más “rígida” favorece la diferenciación hacia linajes osteogénicos (Xie et al., 2018); en el caso de nuestro modelo podemos resaltar algunos GAT interesantes como el gen *THBS1* (también conocido como *TSP1*) codifica a la proteína *Thrombospondin 1*, una glucoproteína adhesiva que media las interacciones células-matriz extracelular, se ha reportado que modelos murinos deficientes en la expresión de este gen muestran una proliferación reducida de progenitores neurales *in vivo* mientras que *in vitro* un modelo *TSP1*<sup>-/-</sup> de NPC de ratón mostraba disminución en su capacidad de diferenciación temprana y proliferación (Lu & Kipnis, 2010). Otro de estos genes de interés es CD44 que codifica una glicoproteína con funciones en las interacciones célula a célula, la adhesión celular y su migración, se ha encontrado que en la neurogénesis adulta esta glicoproteína juega un papel importante en el mantenimiento de la autorrenovación de las células troncales neurales (NSC) además de que les permite mantener su capacidad de diferenciación hacia neuronas, astrocitos y oligodendrocitos, de forma contraria un modelo murino de NSC con depleción de CD44 mostraban una diferenciación neuronal lenta y retardada (Su et al., 2017), en el caso de la diferenciación neuronal durante el desarrollo embrionario se ha reportado en modelos murinos que durante el desarrollo temprano del cerebro CD44 aumenta su expresión en precursores de astrocitos, NCS y células precursoras de oligodendrocitos (Naruse et al., 2013). El último GAT interesante asociado a estos procesos es *RARRES2*, también conocido como *Chemerin*, reportado como un regulador de muchos procesos de diferenciación de linajes mesodérmicos con la capacidad de favorecer la proliferación y diferenciación de mioblastos a través de la vía de *ERK1/2* y la vía de mTOR (Yang et al., 2012), otros estudios han reportado en líneas celulares que la delección de *RARRES2* en la línea celular 3T3-L1 inhibía la diferenciación hacia un linaje adipogénico (Goralski et al., 2007); el decremento en la expresión de este gen en la condición H9 WT concuerda con la inhibición de linajes ajenos al neuroectodermo durante el protocolo de diferenciación, sin embargo, la condición H9 *SIRT1*<sup>-/-</sup> muestra un aumento de este gen al final del

pseudotiempo analizado, pudiendo ser uno de los factores que genera heterogeneidad celular y la presencia de la identidad MSC en esta condición.

Los tres genes identificados en respuesta al daño mediado por p53, *CCNB1*, *CDKN1A* y *AURKA* son genes relacionados con el ciclo celular, el gen *CCNB1* codifica la ciclina B1 que es un regulador esencial de la progresión de la fase G2 a la fase M del ciclo celular, durante la fase G2 del ciclo celular esta ciclina se acumula en el citoplasma, conforme se acerca la transición G2/M *CCNB1* forma un complejo con la cinasa dependiente de ciclinas 1 (*CDK1*) que permite el paso hacia la mitosis (Jackson & Bartek, 2009), por su parte el gen *CDKN1A* es un inhibidor de quinasas dependiente de ciclinas, esta proteína promueve el arresto del ciclo celular mediante su unión y posterior inhibición de la actividad de la ciclina *CDK2*, *CDK1* y *CDK4* (Karimian et al., 2016). Por último el gen *AURKA* codifica una cinasa dependiente del ciclo celular relacionada con la maduración del centrómero durante la mitosis (Liu & Ruderman, 2006).

En la condición H9 WT, el gen *CCNB1* muestra un comportamiento de disminución durante el inicio de la diferenciación que posteriormente aumenta hacia el final de la misma, este aumento final no es tan significativo respecto al inicio de la diferenciación, la condición H9 *SIRT1*<sup>-/-</sup> muestra niveles similares de este gen durante el inicio e intermedio de la diferenciación, sin embargo, al final de esta el nivel desciende en comparación con las células H9 WT, esto podría indicar que al ser sometidas al protocolo de diferenciación las células con delección de *SIRT1* pierden control del checkpoint de transición G2/M, por su parte, el gen *CDKN1A* muestra un incremento gradual en la condición H9 WT lo que nos indica que estas células entran en arresto del ciclo celular durante el protocolo de diferenciación, este arresto celular es un proceso esencial para permitir la correcta diferenciación de las células (Ruijtenberg & van den Heuvel, 2016), la delección de *SIRT1* en la condición H9 *SIRT1*<sup>-/-</sup> genera la pérdida de la dinámica de este inhibidor a lo largo del pseudotiempo de diferenciación lo que podría sugerir que su incapacidad de generar NPC normales podría deberse a un mal desarrollo de estos inducido por la incapacidad de salir del ciclo celular.

Los efectos de la delección de *SIRT1* y sus efectos sobre la regulación del ciclo celular se muestran una vez más al observar la dinámica de los genes identificados como participantes del ciclo celular y su regulación, la inhibición de la capacidad de ingresar al arresto celular se refuerza al observar que genes como *MCM5* y *MCM6* pierden su expresión gradual a lo largo del pseudotiempo de diferenciación en la condición de H9 WT, estos dos genes codifican proteínas que regulan la transición de la fase G0 a G1 durante el ciclo celular y la regulación a la baja de esos produce la entrada al arresto del ciclo celular (Jones et al., 2021), el gen *CDC6* tiene una función parecida ya que la inhibición o alteración de su función genera arresto celular en las mismas fases del ciclo celular

(Borlado & Méndez, 2007), similar a lo observado previamente, la condición H9 *SIRT1*<sup>-/-</sup> no muestra la baja en la expresión de estos genes sugiriendo una pérdida del control de ingreso a ciclo celular arrestado, durante la diferenciación, tras la delección de *SIRT1*.

El efecto de *SIRT1* sobre estos genes y la regulación del ciclo celular podría relacionarse con su regulación sobre las funciones del factor de transcripción *p53* (Gonfloni et al., 2014), ya que la inducción de este último puede generar una regulación a la baja de *CCNB1* (Innocente et al., 1999) y de los genes pertenecientes al complejo *MCM* (Agarwal et al., 2007) así como de otros genes reguladores de las fases del ciclo celular, su arresto e incluso la senescencia celular como son la *polo-like kinase* o *PLK1* (Jackson et al., 2005), la ciclina B1 y la ciclina B2 (Krause et al., 2000).

Finalmente en el caso de los GAT asociados a identidades troncales el comportamiento esperado sería una disminución gradual de su expresión conforme avanza el pseudotiempo, sin embargo el comportamiento de estos GAT en la condición H9 *SIRT1*<sup>-/-</sup> señala que tras la pérdida de *SIRT1* existe una incapacidad de perder estos genes marcadores y por lo tanto una incapacidad para salir del estado pluripotente y troncal dos de los GAT más importantes de esta lista son *DNMT3B* que se ha reportado como un gene sensible a los niveles de oxígeno, esta DNA metiltransferasa reduce su expresión cuando las condiciones de hipoxia celulares se anulan e incrementan los niveles de oxígeno en células troncales mesenquimales (MSC) que afecta la capacidad de diferenciación de estas células (Dogan et al., 2021), sumado a lo anterior se ha reportado el papel que tiene la reducción de la expresión de este gen durante la decisión temprana de linaje ya que en células troncales de epiblasto de ratón su incremento favorecía la diferenciación hacia células mesodermales, por el contrario la depleción de su expresión en etapas tempranas favorecía la generación de neuroectodermo (Lauria et al., 2023), este comportamiento previamente reportado robustece el observado en la dinámica de este gen en la condición H9 WT mostrando un decrecimiento gradual conforme se llega hacia la identidad de progenitor neural.

El segundo gen de interés es *TERF1* que codifica una proteína de interacción con los telómeros, también es miembro del complejo protector o telosoma, este gen está fuertemente asociado con la inducción y mantenimiento de la pluripotencia siendo un blanco regulado por el factor de transcripción *OCT4* (Schneider et al., 2013).

Al realizar un ejercicio de contraste entre las dinámicas de expresión de genes asociados a pluripotencia en ambas condiciones podemos afirmar que la delección de *SIRT1* afecta la capacidad celular de la línea H9 para realizar el escape del estado de pluripotencia a nivel de regulación transcripcional. En refuerzo a lo observado en este trabajo se encontró reportado un modelo de células

*iPSc* murinas sometidas a una diferenciación neuronal que los genes *DNMT3B*, *TERF1* y *TDGF1* formaban parte del panel de genes propuestos como marcadores de troncalidad y, además, muestran un decrecimiento considerable al día 6 de su protocolo de diferenciación (Novak et al., 2022).

Al terminar se encontraron 4 genes asociados con la identidad de células troncales neuronales (NSCs) sin embargo no se encontró información contundente sobre el rol de estos en los procesos de diferenciación temprana en linajes neuroectodermicos sugiriendo la necesidad de analizar la identidad neural en días posteriores al 6 pero previos a la muerte celular observada en la condición H9 *SIRT1*  $\neg$  en el día 12.

## Conclusiones

- SIRT1 modula la transición metabólica durante la diferenciación de hESCs hacia NPC, probablemente mediante el control de la activación de vías relacionadas con la biosíntesis y metabolismo de esteroides y ácidos grasos
- SIRT1 es necesaria para mantener un estado pluripotente basal y correcto a nivel transcripcional debido a que su delección afecta el eje dinámico de incremento de la actividad transcripcional durante la diferenciación a NPC, generando un panorama similar al observado en células troncales senescentes.
- La delección de SIRT1 afecta la dinámica de genes asociados a múltiples procesos dentro de los que destaca la regulación del ciclo celular, la salida de la pluripotencia (pérdida de la identidad troncal) y la reorganización celular durante un proceso de diferenciación.
- Los genes de ciclo celular afectados por la delección de SIRT1 muestran una dinámica de expresión durante la diferenciación que sugiere que esta deacetilasa es esencial en el control de la entrada a arresto del ciclo celular, proceso necesario para llevar a cabo una correcta diferenciación en múltiples modelos y más específicamente en este protocolo de diferenciación hacia NPC.
- SIRT1 es necesaria para poder llevar a cabo el escape del estado pluripotente mediante la regulación de la expresión de genes marcadores, de mantenimiento y fomento de la pluripotencia.

## Referencias:

- A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. (2014). *Nat Biotechnol*, 32(9), 903-914. <https://doi.org/10.1038/nbt.2957>
- Agarwal, M. K., Ruhul Amin, A. R., & Agarwal, M. L. (2007). DNA replication licensing factor minichromosome maintenance deficient 5 rescues p53-mediated growth arrest. *Cancer Res*, 67(1), 116-121. <https://doi.org/10.1158/0008-5472.can-06-2835>
- Albergante, L., Mirkes, E., Bac, J., Chen, H., Martin, A., Faure, L., . . . Zinovyev, A. (2020). Robust and Scalable Learning of Complex Intrinsic Dataset Geometry via ELPiGraph. *Entropy (Basel)*, 22(3). <https://doi.org/10.3390/e22030296>
- Altman, N., & Krzywinski, M. (2018). The curse(s) of dimensionality. *Nature Methods*, 15(6), 399-400. <https://doi.org/10.1038/s41592-018-0019-x>
- Amezquita, R. A., Lun, A. T. L., Becht, E., Carey, V. J., Carpp, L. N., Geistlinger, L., . . . Hicks, S. C. (2020). Orchestrating single-cell analysis with Bioconductor. *Nat Methods*, 17(2), 137-145. <https://doi.org/10.1038/s41592-019-0654-x>
- Andrews, T. S., & Hemberg, M. (2018a). Dropout-based feature selection for scRNASeq. *bioRxiv*, 065094. <https://doi.org/10.1101/065094>
- Andrews, T. S., & Hemberg, M. (2018b). Identifying cell populations with scRNASeq. *Molecular Aspects of Medicine*, 59, 114-122. <https://doi.org/https://doi.org/10.1016/j.mam.2017.07.002>
- Angerer, P., Simon, L., Tritschler, S., Wolf, F. A., Fischer, D., & Theis, F. J. (2017). Single cells make big data: New challenges and opportunities in transcriptomics. *Current Opinion in Systems Biology*, 4, 85-91. <https://doi.org/https://doi.org/10.1016/j.coisb.2017.07.004>
- Angkeow, P., Deshpande, S. S., Qi, B., Liu, Y. X., Park, Y. C., Jeon, B. H., . . . Irani, K. (2002). Redox factor-1: an extra-nuclear role in the regulation of endothelial oxidative stress and apoptosis. *Cell Death Differ*, 9(7), 717-725. <https://doi.org/10.1038/sj.cdd.4401025>
- Aran, D., Looney, A. P., Liu, L., Wu, E., Fong, V., Hsu, A., . . . Bhattacharya, M. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology*, 20(2), 163-172. <https://doi.org/10.1038/s41590-018-0276-y>
- Badia-i-Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., . . . Saez-Rodriguez, J. (2022). decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, 2(1). <https://doi.org/10.1093/bioadv/vbac016>
- Bendall, S. C., Davis, K. L., Amir el, A. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., . . . Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell*, 157(3), 714-725. <https://doi.org/10.1016/j.cell.2014.04.005>
- Bier, E., & De Robertis, E. M. (2015). EMBRYO DEVELOPMENT. BMP gradients: A paradigm for morphogen-mediated developmental patterning. *Science*, 348(6242), aaa5838. <https://doi.org/10.1126/science.aaa5838>
- Borlado, L. R., & Méndez, J. (2007). CDC6: from DNA replication to cell cycle checkpoints and oncogenesis. *Carcinogenesis*, 29(2), 237-243. <https://doi.org/10.1093/carcin/bgm268>
- Borrell, V., Cárdenas, A., Ciceri, G., Galcerán, J., Flames, N., Pla, R., . . . Marín, O. (2012). Slit/Robo signaling modulates the proliferation of central nervous system progenitors. *Neuron*, 76(2), 338-352. <https://doi.org/10.1016/j.neuron.2012.08.003>
- Bost, F., Caron, L., Marchetti, I., Dani, C., Le Marchand-Brustel, Y., & Binétruy, B. (2002). Retinoic acid activation of the ERK pathway is required for embryonic stem cell commitment into the adipocyte lineage. *The Biochemical journal*, 361(Pt 3), 621-627. <https://doi.org/10.1042/0264-6021:3610621>



- Calvanese, V., Lara, E., Suárez-Alvarez, B., Abu Dawud, R., Vázquez-Chantada, M., Martínez-Chantar, M. L., . . . Fraga, M. F. (2010). Sirtuin 1 regulation of developmental genes during differentiation of stem cells. *Proc Natl Acad Sci U S A*, *107*(31), 13736-13741. <https://doi.org/10.1073/pnas.1001399107>
- Campbell, J. N., Macosko, E. Z., Fenselau, H., Pers, T. H., Lyubetskaya, A., Tenen, D., . . . Tsai, L. T. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. *Nature Neuroscience*, *20*(3), 484-496. <https://doi.org/10.1038/nn.4495>
- Chari, T., & Pachter, L. (2022). The Specious Art of Single-Cell Genomics. *bioRxiv*, 2021.2008.2025.457696. <https://doi.org/10.1101/2021.08.25.457696>
- Chari, T., & Pachter, L. (2023). The specious art of single-cell genomics. *PLoS Comput Biol*, *19*(8), e1011288. <https://doi.org/10.1371/journal.pcbi.1011288>
- Chen, Y. T., Collins, L. L., Uno, H., Chou, S. M., Meshul, C. K., Chang, S. S., & Chang, C. (2007). Abnormal cerebellar cytoarchitecture and impaired inhibitory signaling in adult mice lacking TR4 orphan nuclear receptor. *Brain Res*, *1168*, 72-82. <https://doi.org/10.1016/j.brainres.2007.06.069>
- Citri, A., Pang, Z. P., Südhof, T. C., Wernig, M., & Malenka, R. C. (2011). Comprehensive qPCR profiling of gene expression in single neuronal cells. *Nat Protoc*, *7*(1), 118-127. <https://doi.org/10.1038/nprot.2011.430>
- Cronin, M., Ghosh, K., Sistare, F., Quackenbush, J., Vilker, V., & O'Connell, C. (2004). Universal RNA Reference Materials for Gene Expression. *Clinical Chemistry*, *50*(8), 1464-1471. <https://doi.org/10.1373/clinchem.2004.035675>
- Dal Molin, A., Baruzzo, G., & Di Camillo, B. (2017). Single-Cell RNA-Sequencing: Assessment of Differential Expression Analysis Methods. *Front Genet*, *8*, 62. <https://doi.org/10.3389/fgene.2017.00062>
- Das, S., Rai, A., Merchant, M. L., Cave, M. C., & Rai, S. N. (2021). A Comprehensive Survey of Statistical Approaches for Differential Expression Analysis in Single-Cell RNA Sequencing Studies. *Genes (Basel)*, *12*(12). <https://doi.org/10.3390/genes12121947>
- Das, S., & Rai, S. N. (2021). SwarnSeq: An improved statistical approach for differential expression analysis of single-cell RNA-seq data. *Genomics*, *113*(3), 1308-1324. <https://doi.org/10.1016/j.ygeno.2021.02.014>
- DeBose-Boyd, R. A., & Ye, J. (2018). SREBPs in Lipid Metabolism, Insulin Signaling, and Beyond. *Trends Biochem Sci*, *43*(5), 358-368. <https://doi.org/10.1016/j.tibs.2018.01.005>
- Dogan, F., Aljumaily, R. M. K., Kitchen, M., & Forsyth, N. R. (2021). DNMT3B Is an Oxygen-Sensitive De Novo Methylase in Human Mesenchymal Stem Cells. *Cells*, *10*(5). <https://doi.org/10.3390/cells10051032>
- Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoepfner, D. J., Dash, C., . . . Meshorer, E. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell*, *2*(5), 437-447. <https://doi.org/10.1016/j.stem.2008.03.021>
- Evseenko, D., Zhu, Y., Schenke-Layland, K., Kuo, J., Latour, B., Ge, S., . . . Crooks, G. M. (2010). Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *Proc Natl Acad Sci U S A*, *107*(31), 13742-13747. <https://doi.org/10.1073/pnas.1002077107>
- Fang, Y., Tang, S., & Li, X. (2019). Sirtuins in Metabolic and Epigenetic Regulation of Stem Cells. *Trends Endocrinol Metab*, *30*(3), 177-188. <https://doi.org/10.1016/j.tem.2018.12.002>
- Fedele, S., Collo, G., Behr, K., Bischofberger, J., Müller, S., Kunath, T., . . . Taylor, V. (2017). Expansion of human midbrain floor plate progenitors from induced pluripotent stem cells increases dopaminergic neuron differentiation potential. *Scientific Reports*, *7*(1), 6036. <https://doi.org/10.1038/s41598-017-05633-1>
- Feng, G., Zheng, K., Song, D., Xu, K., Huang, D., Zhang, Y., . . . Zhang, D. (2016). SIRT1 was involved in TNF- $\alpha$ -promoted osteogenic differentiation of human DPSCs through Wnt/ $\beta$ -catenin signal. *In Vitro Cellular & Developmental Biology - Animal*, *52*(10), 1001-1011. <https://doi.org/10.1007/s11626-016-0070-9>
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., . . . Gottardo, R. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology*, *16*(1), 278. <https://doi.org/10.1186/s13059-015-0844-5>
- Fisher, W. W., Li, J. J., Hammonds, A. S., Brown, J. B., Pfeiffer, B. D., Weiszmann, R., . . . Celniker, S. E. (2012). DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene

- expression in *Drosophila*. *Proc Natl Acad Sci U S A*, 109(52), 21330-21335. <https://doi.org/10.1073/pnas.1209589110>
- Folmes, C. D., Dzeja, P. P., Nelson, T. J., & Terzic, A. (2012). Metabolic plasticity in stem cell homeostasis and differentiation. *Cell Stem Cell*, 11(5), 596-606. <https://doi.org/10.1016/j.stem.2012.10.002>
- Franzén, O., Gan, L.-M., & Björkegren, J. L. M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*, 2019. <https://doi.org/10.1093/database/baz046>
- Fusco, S., Leone, L., Barbati, Saviana A., Samengo, D., Piacentini, R., Maulucci, G., . . . Grassi, C. (2016). A CREB-Sirt1-Hes1 Circuitry Mediates Neural Stem Cell Response to Glucose Availability. *Cell Reports*, 14(5), 1195-1205. <https://doi.org/10.1016/j.celrep.2015.12.092>
- Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., & Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res*, 29(8), 1363-1375. <https://doi.org/10.1101/gr.240663.118>
- Gigante, S., Burkhardt, D., & Tong, A. (2023). *scprep: A collection of scripts and tools for loading, processing, and handling single cell data*. In (Version 1.2.2)
- Gillespie, M., Jassal, B., Stephan, R., Milacic, M., Rothfels, K., Senff-Ribeiro, A., . . . D'Eustachio, P. (2022). The reactome pathway knowledgebase 2022. *Nucleic acids research*, 50(D1), D687-D692. <https://doi.org/10.1093/nar/gkab1028>
- Gomes, A. P., Price, N. L., Ling, A. J., Moslehi, J. J., Montgomery, M. K., Rajman, L., . . . Sinclair, D. A. (2013). Declining NAD(+) induces a pseudohypoxic state disrupting nuclear-mitochondrial communication during aging. *Cell*, 155(7), 1624-1638. <https://doi.org/10.1016/j.cell.2013.11.037>
- Gonfloni, S., Iannizzotto, V., Maiani, E., Bellusci, G., Ciccone, S., & Diederich, M. (2014). P53 and Sirt1: routes of metabolism and genome stability. *Biochem Pharmacol*, 92(1), 149-156. <https://doi.org/10.1016/j.bcp.2014.08.034>
- Goralski, K. B., McCarthy, T. C., Hanniman, E. A., Zabel, B. A., Butcher, E. C., Parlee, S. D., . . . Sinal, C. J. (2007). Chemerin, a novel adipokine that regulates adipogenesis and adipocyte metabolism. *J Biol Chem*, 282(38), 28175-28188. <https://doi.org/10.1074/jbc.M700793200>
- Green, A. C., Kocovski, P., Jovic, T., Walia, M. K., Chandraratna, R. A. S., Martin, T. J., . . . Purton, L. E. (2017). Retinoic acid receptor signalling directly regulates osteoblast and adipocyte differentiation from mesenchymal progenitor cells. *Exp Cell Res*, 350(1), 284-297. <https://doi.org/10.1016/j.yexcr.2016.12.007>
- Gross, A., Schoendube, J., Zimmermann, S., Steeb, M., Zengerle, R., & Koltay, P. (2015). Technologies for Single-Cell Isolation. *Int J Mol Sci*, 16(8), 16897-16919. <https://doi.org/10.3390/ijms160816897>
- Gu, W., Gaeta, X., Sahakyan, A., Chan, A. B., Hong, C. S., Kim, R., . . . Christofk, H. R. (2016). Glycolytic Metabolism Plays a Functional Role in Regulating Human Pluripotent Stem Cell State. *Cell Stem Cell*, 19(4), 476-490. <https://doi.org/10.1016/j.stem.2016.08.008>
- Hastie, T. J. (2017). Generalized additive models. In *Statistical models in S* (pp. 249-307). Routledge.
- Heo, J., Lim, J., Lee, S., Jeong, J., Kang, H., Kim, Y., . . . Shin, D. M. (2017). Sirt1 Regulates DNA Methylation and Differentiation Potential of Embryonic Stem Cells by Antagonizing Dnmt3l. *Cell Rep*, 18(8), 1930-1945. <https://doi.org/10.1016/j.celrep.2017.01.074>
- Herman, J. S., Sagar, & Grün, D. (2018). FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat Methods*, 15(5), 379-386. <https://doi.org/10.1038/nmeth.4662>
- Hsu, Y. C., Wu, Y. T., Tsai, C. L., & Wei, Y. H. (2018). Current understanding and future perspectives of the roles of sirtuins in the reprogramming and differentiation of pluripotent stem cells. *Exp Biol Med (Maywood)*, 243(6), 563-575. <https://doi.org/10.1177/1535370218759636>
- Hu, P., Zhang, W., Xin, H., & Deng, G. (2016). Single Cell Isolation and Analysis [Review]. *Frontiers in Cell and Developmental Biology*, 4. <https://doi.org/10.3389/fcell.2016.00116>

- Hu, Q., Friedrich, A. M., Johnson, L. V., & Clegg, D. O. (2010). Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation. *Stem Cells*, 28(11), 1981-1991. <https://doi.org/10.1002/stem.531>
- Hwang, B., Lee, J. H., & Bang, D. (2018). Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med*, 50(8), 1-14. <https://doi.org/10.1038/s12276-018-0071-8>
- Ingolia, N. T., Lareau, L. F., & Weissman, J. S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, 147(4), 789-802. <https://doi.org/10.1016/j.cell.2011.10.002>
- Innocente, S. A., Abrahamson, J. L., Cogswell, J. P., & Lee, J. M. (1999). p53 regulates a G2 checkpoint through cyclin B1. *Proc Natl Acad Sci U S A*, 96(5), 2147-2152. <https://doi.org/10.1073/pnas.96.5.2147>
- Ito, K., & Suda, T. (2014). Metabolic requirements for the maintenance of self-renewing stem cells. *Nat Rev Mol Cell Biol*, 15(4), 243-256. <https://doi.org/10.1038/nrm3772>
- Iuchi, K., Ema, M., Suzuki, M., Yokoyama, C., & Hisatomi, H. (2019). Oxidized unsaturated fatty acids induce apoptotic cell death in cultured cells. *Mol Med Rep*, 19(4), 2767-2773. <https://doi.org/10.3892/mmr.2019.9940>
- Jackson, M. W., Agarwal, M. K., Yang, J., Bruss, P., Uchiumi, T., Agarwal, M. L., . . . Taylor, W. R. (2005). p130/p107/p105Rb-dependent transcriptional repression during DNA-damage-induced cell-cycle exit at G2. *J Cell Sci*, 118(Pt 9), 1821-1832. <https://doi.org/10.1242/jcs.02307>
- Jackson, S. P., & Bartek, J. (2009). The DNA-damage response in human biology and disease. *Nature*, 461(7267), 1071-1078. <https://doi.org/10.1038/nature08467>
- Ji, Z., & Ji, H. (2016). TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res*, 44(13), e117. <https://doi.org/10.1093/nar/gkw430>
- Jiang, R., Sun, T., Song, D., & Li, J. J. (2022). Statistics or biology: the zero-inflation controversy about scRNA-seq data. *Genome Biol*, 23(1), 31. <https://doi.org/10.1186/s13059-022-02601-5>
- Jones, M. L., Baris, Y., Taylor, M. R. G., & Yeeles, J. T. P. (2021). Structure of a human replisome shows the organisation and interactions of a DNA replication machine. *Embo j*, 40(23), e108819. <https://doi.org/10.15252/emboj.2021108819>
- Jung, Y. J., Park, W., Noh, J. M., Kang, K. P., Nguyen-Thanh, T., Han, M. K., & Kim, W. (2020). SIRT1 induces the adipogenic differentiation of mouse embryonic stem cells by regulating RA-induced RAR expression via NCOR1 acetylation. *Stem Cell Research*, 44, 101771. <https://doi.org/https://doi.org/10.1016/j.scr.2020.101771>
- Karimian, A., Ahmadi, Y., & Yousefi, B. (2016). Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair (Amst)*, 42, 63-71. <https://doi.org/10.1016/j.dnarep.2016.04.008>
- Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., . . . Mathelier, A. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*, 46(D1), D260-d266. <https://doi.org/10.1093/nar/gkx1126>
- Kharchenko, P. V., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat Methods*, 11(7), 740-742. <https://doi.org/10.1038/nmeth.2967>
- Kittappa, R., Chang, W. W., Awatramani, R. B., & McKay, R. D. (2007). The foxa2 gene controls the birth and spontaneous degeneration of dopamine neurons in old age. *PLoS Biol*, 5(12), e325. <https://doi.org/10.1371/journal.pbio.0050325>
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., . . . Corlay, S. (2016). *Jupyter Notebooks-a publishing format for reproducible computational workflows* (Vol. 2016).
- Kobak, D., & Berens, P. (2019). The art of using t-SNE for single-cell transcriptomics. *Nat Commun*, 10(1), 5416. <https://doi.org/10.1038/s41467-019-13056-x>

- Korthauer, K. D., Chu, L. F., Newton, M. A., Li, Y., Thomson, J., Stewart, R., & Kendzierski, C. (2016). A statistical approach for identifying differential distributions in single-cell RNA-seq experiments. *Genome Biol*, 17(1), 222. <https://doi.org/10.1186/s13059-016-1077-y>
- Krause, K., Wasner, M., Reinhard, W., Haugwitz, U., Dohna, C. L., Mössner, J., & Engeland, K. (2000). The tumour suppressor protein p53 can repress transcription of cyclin B. *Nucleic Acids Res*, 28(22), 4410-4418. <https://doi.org/10.1093/nar/28.22.4410>
- Kriegstein, A., & Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. *Annu Rev Neurosci*, 32, 149-184. <https://doi.org/10.1146/annurev.neuro.051508.135600>
- Kulakovskiy, I. V., Vorontsov, I. E., Yevshin, I. S., Sharipov, R. N., Fedorova, A. D., Rumynskiy, E. I., . . . Makeev, V. J. (2018). HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res*, 46(D1), D252-d259. <https://doi.org/10.1093/nar/gkx1106>
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., . . . Ma'ayan, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*, 44(W1), W90-97. <https://doi.org/10.1093/nar/gkw377>
- La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., . . . Linnarsson, S. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell*, 167(2), 566-580.e519. <https://doi.org/10.1016/j.cell.2016.09.027>
- Lauria, A., Meng, G., Proserpio, V., Rapelli, S., Maldotti, M., Polignano, I. L., . . . Oliviero, S. (2023). DNMT3B supports meso-endoderm differentiation from mouse embryonic stem cells. *Nat Commun*, 14(1), 367. <https://doi.org/10.1038/s41467-023-35938-x>
- Lesurf, R., Cotto, K. C., Wang, G., Griffith, M., Kasaian, K., Jones, S. J., . . . Griffith, O. L. (2016). ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res*, 44(D1), D126-132. <https://doi.org/10.1093/nar/gkv1203>
- Li, Y., He, X., He, J., Anderstam, B., Andersson, G., & Lindgren, U. (2011). Nicotinamide phosphoribosyltransferase (Nampt) affects the lineage fate determination of mesenchymal stem cells: a possible cause for reduced osteogenesis and increased adipogenesis in older individuals. *J Bone Miner Res*, 26(11), 2656-2664. <https://doi.org/10.1002/jbmr.480>
- Liu, Q., Li, M., Wang, S., Xiao, Z., Xiong, Y., & Wang, G. (2020). Recent Advances of Osterix Transcription Factor in Osteoblast Differentiation and Bone Formation [Review]. *Frontiers in Cell and Developmental Biology*, 8. <https://doi.org/10.3389/fcell.2020.601224>
- Liu, Q., & Ruderman, J. V. (2006). Aurora A, mitotic entry, and spindle bipolarity. *Proc Natl Acad Sci U S A*, 103(15), 5811-5816. <https://doi.org/10.1073/pnas.0601425103>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- Lu, Z., & Kipnis, J. (2010). Thrombospondin 1--a key astrocyte-derived neurogenic factor. *Faseb j*, 24(6), 1925-1934. <https://doi.org/10.1096/fj.09-150573>
- Lum, J. J., Bui, T., Gruber, M., Gordan, J. D., DeBerardinis, R. J., Covelto, K. L., . . . Thompson, C. B. (2007). The transcription factor HIF-1alpha plays a critical role in the growth factor-dependent regulation of both aerobic and anaerobic glycolysis. *Genes Dev*, 21(9), 1037-1049. <https://doi.org/10.1101/gad.1529107>
- Lun, A. a. M., DJ and Marioni, J. C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor [version 2; peer review: 3 approved, 2 approved with reservations]. *F1000Research*, 5(2122). <https://doi.org/10.12688/f1000research.9501.2>
- Lun, A. T. L., Riesenfeld, S., Andrews, T., Dao, T. P., Gomes, T., Marioni, J. C., & participants in the 1st Human Cell Atlas, J. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biology*, 20(1), 63. <https://doi.org/10.1186/s13059-019-1662-y>



- Lupi, R., Corda, D., & Di Girolamo, M. (2000). Endogenous ADP-ribosylation of the G protein beta subunit prevents the inhibition of type 1 adenylyl cyclase. *J Biol Chem*, 275(13), 9418-9424. <https://doi.org/10.1074/jbc.275.13.9418>
- Lönnerberg, T., Svensson, V., James, K. R., Fernandez-Ruiz, D., Sebina, I., Montandon, R., . . . Teichmann, S. A. (2017). Single-cell RNA-seq and computational analysis using temporal mixture modelling resolves Th1/Tfh fate bifurcation in malaria. *Sci Immunol*, 2(9). <https://doi.org/10.1126/sciimmunol.aal2192>
- Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C., & Hume, D. A. (2013). An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics*, 14(1), 632. <https://doi.org/10.1186/1471-2164-14-632>
- McCarthy, D. J., Campbell, K. R., Lun, A. T. L., & Wills, Q. F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*, 33(8), 1179-1186. <https://doi.org/10.1093/bioinformatics/btw777>
- Miao, Z., Deng, K., Wang, X., & Zhang, X. (2018). DEsingle for detecting three types of differential expression in single-cell RNA-seq data. *Bioinformatics*, 34(18), 3223-3224. <https://doi.org/10.1093/bioinformatics/bty332>
- Miragaia, R. J., Gomes, T., Chomka, A., Jardine, L., Riedel, A., Hegazy, A. N., . . . Teichmann, S. A. (2019). Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. *Immunity*, 50(2), 493-504.e497. <https://doi.org/10.1016/j.immuni.2019.01.001>
- Moon, K. R., van Dijk, D., Wang, Z., Gigante, S., Burkhardt, D. B., Chen, W. S., . . . Krishnaswamy, S. (2019). Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol*, 37(12), 1482-1492. <https://doi.org/10.1038/s41587-019-0336-3>
- Morgan, M., & Shepherd, L. (2022). *AnnotationHub: Client to access AnnotationHub resources*. In (Version R package version 3.4.0.)
- Moussaieff, A., Rouleau, M., Kitsberg, D., Cohen, M., Levy, G., Barasch, D., . . . Nahmias, Y. (2015). Glycolysis-mediated changes in acetyl-CoA and histone acetylation control the early differentiation of embryonic stem cells. *Cell Metab*, 21(3), 392-402. <https://doi.org/10.1016/j.cmet.2015.02.002>
- Muth, V., Nadaud, S., Grummt, I., & Voit, R. (2001). Acetylation of TAF(I)68, a subunit of TIF-IB/SL1, activates RNA polymerase I transcription. *Embo j*, 20(6), 1353-1362. <https://doi.org/10.1093/emboj/20.6.1353>
- Naruse, M., Shibasaki, K., Yokoyama, S., Kurachi, M., & Ishizaki, Y. (2013). Dynamic changes of CD44 expression from progenitors to subpopulations of astrocytes and neurons in developing cerebellum. *PLoS One*, 8(1), e53109. <https://doi.org/10.1371/journal.pone.0053109>
- Noh, K. M., Park, S. J., Moon, S. H., & Jung, S. Y. (2023). Extracellular matrix cues regulate the differentiation of pluripotent stem cell-derived endothelial cells. *Front Cardiovasc Med*, 10, 1169331. <https://doi.org/10.3389/fcvm.2023.1169331>
- Novak, G., Kyriakis, D., Grzyb, K., Bernini, M., Rodius, S., Dittmar, G., . . . Skupin, A. (2022). Single-cell transcriptomics of human iPSC differentiation dynamics reveal a core molecular network of Parkinson's disease. *Commun Biol*, 5(1), 49. <https://doi.org/10.1038/s42003-021-02973-7>
- Orchard, S., Ammari, M., Aranda, B., Breuza, L., Briganti, L., Broackes-Carter, F., . . . Hermjakob, H. (2014). The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*, 42(Database issue), D358-363. <https://doi.org/10.1093/nar/gkt1115>
- Pandas, D. T. (2020). *Pandas*. In Zenodo.
- Panopoulos, A. D., Yanes, O., Ruiz, S., Kida, Y. S., Diep, D., Tautenhahn, R., . . . Izpisua Belmonte, J. C. (2012). The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res*, 22(1), 168-177. <https://doi.org/10.1038/cr.2011.177>
- Pasquini, G., Rojo Arias, J. E., Schäfer, P., & Busskamp, V. (2021). Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J*, 19, 961-969. <https://doi.org/10.1016/j.csbj.2021.01.015>

- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., . . . Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. *Cell*, 167(1), 285. <https://doi.org/10.1016/j.cell.2016.08.009>
- Phipson, B. a. Z., L and Oshlack, A. (2017). Gene length and detection bias in single cell RNA sequencing protocols [version 1; peer review: 4 approved]. *F1000Research*, 6(595). <https://doi.org/10.12688/f1000research.11290.1>
- Ponugoti, B., Kim, D. H., Xiao, Z., Smith, Z., Miao, J., Zang, M., . . . Kemper, J. K. (2010). SIRT1 deacetylates and inhibits SREBP-1C activity in regulation of hepatic lipid metabolism. *J Biol Chem*, 285(44), 33959-33970. <https://doi.org/10.1074/jbc.M110.122978>
- Puri, N., Sodhi, K., Haarstad, M., Kim, D. H., Bohinc, S., Foglio, E., . . . Abraham, N. G. (2012). Heme induced oxidative stress attenuates sirtuin1 and enhances adipogenesis in mesenchymal stem cells and mouse pre-adipocytes. *J Cell Biochem*, 113(6), 1926-1935. <https://doi.org/10.1002/jcb.24061>
- Qi, M., Li, W., Tsang, I. W., & Yijun, S. (2017). Principal Graph and Structure Learning Based on Reversed Graph Embedding. *IEEE Trans Pattern Anal Mach Intell*, 39(11), 2227-2241. <https://doi.org/10.1109/tpami.2016.2635657>
- Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H. A., & Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*, 14(10), 979-982. <https://doi.org/10.1038/nmeth.4402>
- R Development Core Team. (2010). *R: A language and environment for statistical computing*. In (Version 4.1.2)
- Rainer, J. (2017). *EnsDb.Hsapiens.v86: Ensembl based annotation package*. In (Version 2.99.0)
- Risso, D., & Cole, M. (2022). *scRNAseq: Collection of Public Single-Cell RNA-Seq Datasets*. In (Version R package version 2.10.0.)
- Risso, D., Perraudeau, F., Gribkova, S., Dudoit, S., & Vert, J. P. (2018). A general and flexible method for signal extraction from single-cell RNA-seq data. *Nat Commun*, 9(1), 284. <https://doi.org/10.1038/s41467-017-02554-5>
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7), e47-e47. <https://doi.org/10.1093/nar/gkv007>
- Robertson, E. J. (2014). Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Semin Cell Dev Biol*, 32, 73-79. <https://doi.org/10.1016/j.semcdb.2014.03.028>
- Rodrigues, A. S., Pereira, S. L., & Ramalho-Santos, J. (2020). Stem metabolism: Insights from oncometabolism and vice versa. *Biochim Biophys Acta Mol Basis Dis*, 1866(7), 165760. <https://doi.org/10.1016/j.bbadis.2020.165760>
- Ruijtenberg, S., & van den Heuvel, S. (2016). Coordinating cell proliferation and differentiation: Antagonism between cell cycle regulators and cell type-specific gene expression. *Cell Cycle*, 15(2), 196-212. <https://doi.org/10.1080/15384101.2015.1120925>
- Saba, J. A., Liakath-Ali, K., Green, R., & Watt, F. M. (2021). Translational control of stem cell function. *Nature Reviews Molecular Cell Biology*, 22(10), 671-690. <https://doi.org/10.1038/s41580-021-00386-2>
- Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R. E., Neild, A., Farbehi, N., . . . Gallego-Ortega, D. (2019). Droplet-based single cell RNAseq tools: a practical guide. *Lab Chip*, 19(10), 1706-1727. <https://doi.org/10.1039/c8lc01239c>
- Sampath, P., Pritchard, D. K., Pabon, L., Reinecke, H., Schwartz, S. M., Morris, D. R., & Murry, C. E. (2008). A hierarchical network controls protein translation during murine embryonic stem cell self-renewal and differentiation. *Cell Stem Cell*, 2(5), 448-460. <https://doi.org/10.1016/j.stem.2008.03.013>
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., & Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology*, 33(5), 495-502. <https://doi.org/10.1038/nbt.3192>
- Sauve, A. A. (2010). Sirtuin chemical mechanisms. *Biochim Biophys Acta*, 1804(8), 1591-1603. <https://doi.org/10.1016/j.bbapap.2010.01.021>

- Schneider, R. P., Garrobo, I., Foronda, M., Palacios, J. A., Marión, R. M., Flores, I., . . . Blasco, M. A. (2013). TRF1 is a stem cell marker and is essential for the generation of induced pluripotent stem cells. *Nat Commun*, *4*, 1946. <https://doi.org/10.1038/ncomms2946>
- Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., . . . Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nat Commun*, *9*(1), 20. <https://doi.org/10.1038/s41467-017-02391-6>
- Seo, N.-H., Song, H.-R., & Han, M.-K. (2019). SIRT1 Knockdown Enhances the Differentiation of Human Embryonic Stem Cells into Pancreatic  $\beta$  Cells. *Development & Reproduction*, *23*(4), 391-399. <https://doi.org/10.12717/DR.2019.23.4.391>
- Shaffer, S. M., Dunagin, M. C., Torborg, S. R., Torre, E. A., Emert, B., Krepler, C., . . . Raj, A. (2017). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature*, *546*(7658), 431-435. <https://doi.org/10.1038/nature22794>
- Shi, G., & Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res Ther*, *1*(5), 39. <https://doi.org/10.1186/scrt39>
- Shi, X., Zhang, Y., Zheng, J., & Pan, J. (2012). Reactive oxygen species in cancer stem cells. *Antioxid Redox Signal*, *16*(11), 1215-1228. <https://doi.org/10.1089/ars.2012.4529>
- Shyh-Chang, N., Daley, G. Q., & Cantley, L. C. (2013). Stem cell metabolism in tissue development and aging. *Development*, *140*(12), 2535-2547. <https://doi.org/10.1242/dev.091777>
- Soneson, C., & Robinson, M. D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. *Nat Methods*, *15*(4), 255-261. <https://doi.org/10.1038/nmeth.4612>
- Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nat Rev Genet*, *20*(11), 631-656. <https://doi.org/10.1038/s41576-019-0150-2>
- Street, K., Risso, D., Fletcher, R. B., Das, D., Ngai, J., Yosef, N., . . . Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*, *19*(1), 477. <https://doi.org/10.1186/s12864-018-4772-0>
- Su, W., Foster, S. C., Xing, R., Feistel, K., Olsen, R. H., Acevedo, S. F., . . . Sherman, L. S. (2017). CD44 Transmembrane Receptor and Hyaluronan Regulate Adult Hippocampal Neural Stem Cell Quiescence and Differentiation. *J Biol Chem*, *292*(11), 4434-4445. <https://doi.org/10.1074/jbc.M116.774109>
- Sánchez-Ramírez, E., Ung, T. P. L., Alarcón del Carmen, A., del Toro-Ríos, X., Fajardo-Orduña, G. R., Noriega, L. G., . . . Aguilar-Arnal, L. (2022). Coordinated metabolic transitions and gene expression by NAD<sup>+</sup> during adipogenesis. *Journal of Cell Biology*, *221*(12). <https://doi.org/10.1083/jcb.202111137>
- Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663-676. <https://doi.org/10.1016/j.cell.2006.07.024>
- Tang, S., Fang, Y., Huang, G., Xu, X., Padilla-Banks, E., Fan, W., . . . Li, X. (2017). Methionine metabolism is essential for SIRT1-regulated mouse embryonic stem cell maintenance and embryonic development. *EMBO J*, *36*(21), 3175-3193. <https://doi.org/10.15252/embj.201796708>
- Taura, D., Noguchi, M., Sone, M., Hosoda, K., Mori, E., Okada, Y., . . . Nakao, K. (2009). Adipogenic differentiation of human induced pluripotent stem cells: comparison with that of human embryonic stem cells. *FEBS Lett*, *583*(6), 1029-1033. <https://doi.org/10.1016/j.febslet.2009.02.031>
- The Gene Ontology resource: enriching a GOld mine. (2021). *Nucleic Acids Res*, *49*(D1), D325-d334. <https://doi.org/10.1093/nar/gkaa1113>
- Todorov, H., Cannoodt, R., Saelens, W., & Saeys, Y. (2020). TinGa: fast and flexible trajectory inference with Growing Neural Gas. *Bioinformatics*, *36*(Suppl\_1), i66-i74. <https://doi.org/10.1093/bioinformatics/btaa463>
- Tonk, C. H., Witzler, M., Schulze, M., & Tobiasch, E. (2020). Mesenchymal Stem Cells. In B. Brand-Saberi (Ed.), *Essential Current Concepts in Stem Cell Biology* (pp. 21-39). Springer International Publishing. [https://doi.org/10.1007/978-3-030-33923-4\\_2](https://doi.org/10.1007/978-3-030-33923-4_2)

- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., . . . Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol*, 32(4), 381-386. <https://doi.org/10.1038/nbt.2859>
- Trescher, S., Münchmeyer, J., & Leser, U. (2017). Estimating genome-wide regulatory activity from multi-omics data sets using mathematical optimization. *BMC Syst Biol*, 11(1), 41. <https://doi.org/10.1186/s12918-017-0419-z>
- Tripathi, S., Vercruyse, S., Chawla, K., Christie, K. R., Blake, J. A., Huntley, R. P., . . . Kuiper, M. (2016). Gene regulation knowledge commons: community action takes care of DNA binding transcription factors. *Database (Oxford)*, 2016. <https://doi.org/10.1093/database/baw088>
- Tung, P.-Y., Blischak, J. D., Hsiao, C. J., Knowles, D. A., Burnett, J. E., Pritchard, J. K., & Gilad, Y. (2017). Batch effects and the effective design of single-cell gene expression studies. *Scientific reports*, 7, 39921-39921. <https://doi.org/10.1038/srep39921>
- Ushio-Fukai, M., & Rehman, J. (2014). Redox and metabolic regulation of stem/progenitor cells and their niche. In *Antioxid Redox Signal* (Vol. 21, pp. 1587-1590). <https://doi.org/10.1089/ars.2014.5931>
- Vallejos, C. A., Marioni, J. C., & Richardson, S. (2015). BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLoS Comput Biol*, 11(6), e1004333. <https://doi.org/10.1371/journal.pcbi.1004333>
- van Amerongen, M. J., Diehl, F., Novoyatleva, T., Patra, C., & Engel, F. B. (2009). E2F4 is required for cardiomyocyte proliferation. *Cardiovascular Research*, 86(1), 92-102. <https://doi.org/10.1093/cvr/cvp383>
- Van Blerkom, J. (2009). Mitochondria in early mammalian development. *Seminars in Cell & Developmental Biology*, 20(3), 354-364. <https://doi.org/https://doi.org/10.1016/j.semcdb.2008.12.005>
- Van den Berge, K., Perraudeau, F., Soneson, C., Love, M. I., Risso, D., Vert, J. P., . . . Clement, L. (2018). Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. *Genome Biol*, 19(1), 24. <https://doi.org/10.1186/s13059-018-1406-4>
- Van den Berge, K., Roux de Bézieux, H., Street, K., Saelens, W., Cannoodt, R., Saeys, Y., . . . Clement, L. (2020). Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun*, 11(1), 1201. <https://doi.org/10.1038/s41467-020-14766-3>
- Varum, S., Rodrigues, A. S., Moura, M. B., Momcilovic, O., Easley, C. A. t., Ramalho-Santos, J., . . . Schatten, G. (2011). Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One*, 6(6), e20914. <https://doi.org/10.1371/journal.pone.0020914>
- Verona, R., Moberg, K., Estes, S., Starz, M., Vernon, J. P., & Lees, J. A. (1997). E2F activity is regulated by cell cycle-dependent changes in subcellular localization. *Molecular and Cellular Biology*, 17(12), 7268-7282. <https://doi.org/doi:10.1128/MCB.17.12.7268>
- Virshup, I., Rybakov, S., Theis, F. J., Angerer, P., & Wolf, F. A. (2021). anndata: Annotated data. *bioRxiv*, 2021.2012.2016.473007. <https://doi.org/10.1101/2021.12.16.473007>
- Wadkin, L. E., Orozco-Fuentes, S., Neganova, I., Bojic, S., Laude, A., Lako, M., . . . Shukurov, A. (2019). Seeding hESCs to achieve optimal colony clonality. *Sci Rep*, 9(1), 15299. <https://doi.org/10.1038/s41598-019-51897-0>
- Wagner, J., Rapsomaniki, M. A., Chevrier, S., Anzeneder, T., Langwieder, C., Dykgers, A., . . . Bodenmiller, B. (2019). A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell*, 177(5), 1330-1345.e1318. <https://doi.org/10.1016/j.cell.2019.03.005>
- Weger, M., Diotel, N., Dorsemans, A.-C., Dickmeis, T., & Weger, B. D. (2017). Stem cells and the circadian clock. *Developmental Biology*, 431(2), 111-123. <https://doi.org/https://doi.org/10.1016/j.ydbio.2017.09.012>
- Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M., & Darnell, J. E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell*, 78(4), 575-588. [https://doi.org/10.1016/0092-8674\(94\)90523-1](https://doi.org/10.1016/0092-8674(94)90523-1)



- Welch, J. D., Hartemink, A. J., & Prins, J. F. (2017). MATCHER: manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics. *Genome Biol*, *18*(1), 138. <https://doi.org/10.1186/s13059-017-1269-0>
- Williams, E. O., Taylor, A. K., Bell, E. L., Lim, R., Kim, D. M., & Guarente, L. (2016). Sirtuin 1 Promotes Deacetylation of Oct4 and Maintenance of Naive Pluripotency. *Cell Rep*, *17*(3), 809-820. <https://doi.org/10.1016/j.celrep.2016.09.046>
- Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol*, *19*(1), 15. <https://doi.org/10.1186/s13059-017-1382-0>
- Wolf, F. A., Hamey, F. K., Plass, M., Solana, J., Dahlin, J. S., Göttgens, B., . . . Theis, F. J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol*, *20*(1), 59. <https://doi.org/10.1186/s13059-019-1663-x>
- Wu, Z., Zhang, Y., Stitzel, M. L., & Wu, H. (2018). Two-phase differential expression analysis for single cell RNA-seq. *Bioinformatics*, *34*(19), 3340-3348. <https://doi.org/10.1093/bioinformatics/bty329>
- Xie, J., Lou, Q., Zeng, Y., Liang, Y., Xie, S., Xu, Q., . . . Zhao, M. (2021). Single-Cell Atlas Reveals Fatty Acid Metabolites Regulate the Functional Heterogeneity of Mesenchymal Stem Cells. *Front Cell Dev Biol*, *9*, 653308. <https://doi.org/10.3389/fcell.2021.653308>
- Xie, J., Zhang, D., Zhou, C., Yuan, Q., Ye, L., & Zhou, X. (2018). Substrate elasticity regulates adipose-derived stromal cell differentiation towards osteogenesis and adipogenesis through  $\beta$ -catenin transduction. *Acta Biomater*, *79*, 83-95. <https://doi.org/10.1016/j.actbio.2018.08.018>
- Yang, H., Li, F., Kong, X., Yuan, X., Wang, W., Huang, R., . . . Yin, Y. (2012). Chemerin regulates proliferation and differentiation of myoblast cells via ERK1/2 and mTOR signaling pathways. *Cytokine*, *60*(3), 646-652. <https://doi.org/10.1016/j.cyto.2012.07.033>
- Yao, C., Sun, H. W., Lacey, N. E., Ji, Y., Moseman, E. A., Shih, H. Y., . . . Wu, T. (2019). Single-cell RNA-seq reveals TOX as a key regulator of CD8. *Nat Immunol*, *20*(7), 890-901. <https://doi.org/10.1038/s41590-019-0403-4>
- Ye, Y., Gao, L., & Zhang, S. (2019). Circular Trajectory Reconstruction Uncovers Cell-Cycle Progression and Regulatory Dynamics from Single-Cell Hi-C Maps. *Adv Sci (Weinh)*, *6*(23), 1900986. <https://doi.org/10.1002/advs.201900986>
- Ying, W. (2006). NAD<sup>+</sup> and NADH in cellular functions and cell death. *Front Biosci*, *11*, 3129-3148. <https://doi.org/10.2741/2038>
- Zainabadi, K., Liu, C. J., & Guarente, L. (2017). SIRT1 is a positive regulator of the master osteoblast transcription factor, RUNX2. *PLOS ONE*, *12*(5), e0178520. <https://doi.org/10.1371/journal.pone.0178520>
- Zhang, Y., Feng, X. H., & Derynck, R. (1998). Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. *Nature*, *394*(6696), 909-913. <https://doi.org/10.1038/29814>
- Zhao, Y., Ling, F., Griffin, T. M., He, T., Towner, R., Ruan, H., & Sun, X. H. (2014). Up-regulation of the Sirtuin 1 (Sirt1) and peroxisome proliferator-activated receptor  $\gamma$  coactivator-1 $\alpha$  (PGC-1 $\alpha$ ) genes in white adipose tissue of Id1 protein-deficient mice: implications in the protection against diet and age-induced glucose intolerance. *J Biol Chem*, *289*(42), 29112-29122. <https://doi.org/10.1074/jbc.M114.571679>
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., . . . Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*(1), 14049. <https://doi.org/10.1038/ncomms14049>
- Zhou, Y., Zhou, Z., Zhang, W., Hu, X., Wei, H., Peng, J., & Jiang, S. (2015). SIRT1 inhibits adipogenesis and promotes myogenic differentiation in C3H10T1/2 pluripotent cells by regulating Wnt signaling. *Cell Biosci*, *5*, 61. <https://doi.org/10.1186/s13578-015-0055-5>

Figuras suplementarias.

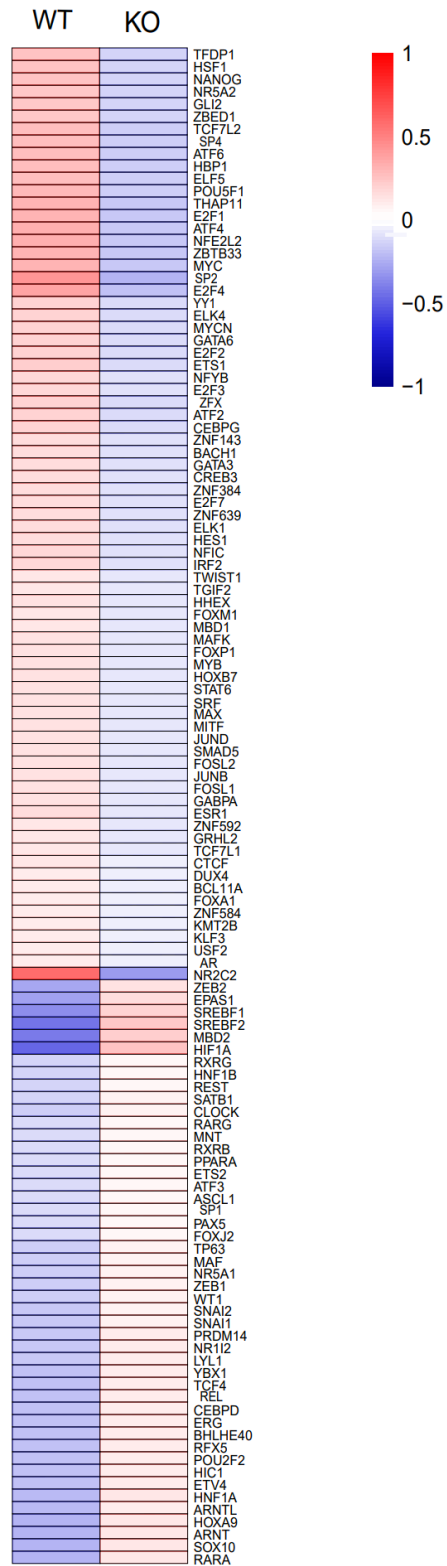


Figura suplementaria 1: Se muestran los 122 TFs estadísticamente significativos en los datos del día 0 obtenidos durante la inferencia de actividad de TFs

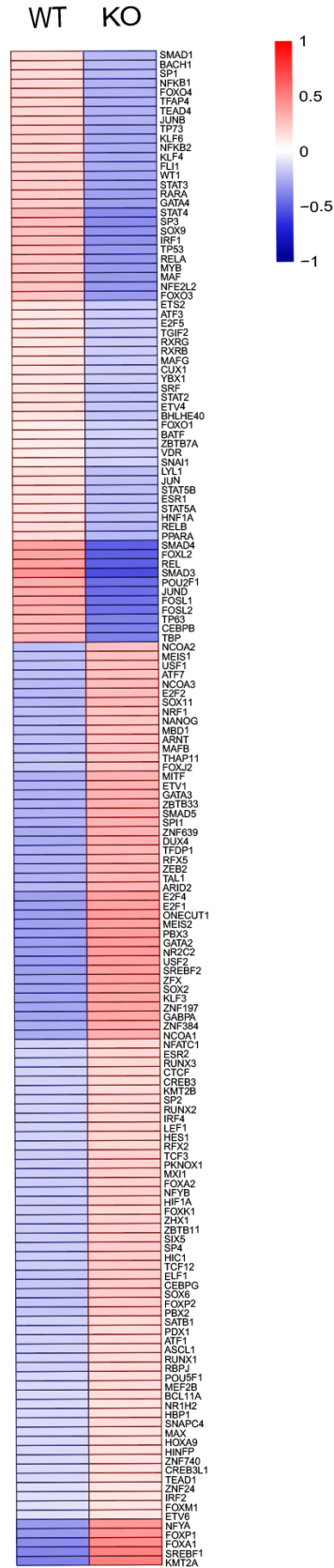


Figura suplementaria 2: Se muestran los 164 TFs estadísticamente significativos en los datos del día 6 obtenidos durante la inferencia de actividad de TFs