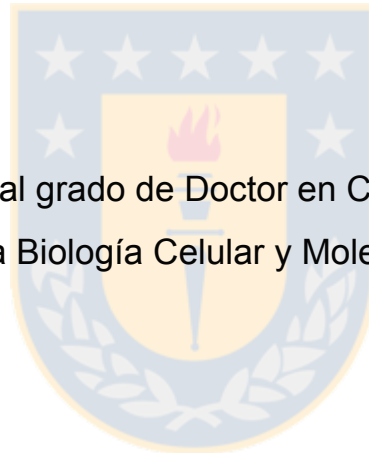




Universidad de Concepción  
Dirección de Postgrado  
Facultad de Ciencias Biológicas -Programa de Doctorado en Ciencias  
Biológicas  
Área Biología Celular y Molecular

## **IDENTIFICACIÓN DE BLANCOS TRANSCRIPCIONALES DE SALL2**



Tesis para optar al grado de Doctor en Ciencias Biológicas  
Área Biología Celular y Molecular

**CARLOS ALEJANDRO FARKAS POOL**  
**CONCEPCIÓN-CHILE**  
**2018**

Profesor Guía: Dra. Roxana Jacqueline Pincheira Barrera  
Dpto. de Bioquímica y Biología Molecular, Facultad de Ciencias Biológicas  
Universidad de Concepción

Esta tesis ha sido realizada en el Departamento de Bioquímica y Biología Molecular de la Facultad de Ciencias Biológicas, Universidad de Concepción.

Profesores integrantes Comisión Evaluadora:

---

Dra. Roxana Pincheira Barrera  
Profesor Guía de Tesis  
Facultad de Ciencias Biológicas

---

Dr. Leonardo Gutiérrez Contreras  
Facultad de Ciencias Biológicas



---

Dr. Sylvain Marcellini Liotaud  
Facultad de Ciencias Biológicas

---

Dr. Mario Galindo Díaz  
Profesor Evaluador Externo  
Universidad de Chile

---

Dr. Juan Pablo Henríquez H.  
Director  
Programa Doctorado en Ciencias  
Biológicas.  
Área Biología Celular y Molecular

## RESUMEN

En esta tesis llevamos a cabo estudios de transcriptoma en un modelo “knockout” (KO) para *Sall2* para comprender la respuesta transcripcional dependiente de *Sall2* y diseccionar los correspondientes genes dependientes de este factor de transcripción. En el modelo utilizado se identificó una cohorte de genes altamente dependientes del fondo genético en donde se mantiene el modelo KO. En base a lo anterior y a la literatura existente, propusimos el uso de datos RNA-Seq como un método para determinar la introgresión genética en genes codificantes, incluyendo estrategias para descubrir variantes vinculadas al genotipo KO y la identificación de potenciales genes modificadores del fenotipo en los organismos genéticamente modificados (GEMs). Empleamos esta metodología en datos de ARN-Seq públicamente disponibles desde cinco modelos KO congénicos, incluyendo los datos experimentales de ARN-Seq de nuestro modelo en estudio *Sall2* KO. Los métodos aplicados demostraron la introgresión del genoma de la célula madre embrionaria (ESC) derivada de la cepa endogámica “129” en gráficos de genoma completo, la detección del segmento embrionario que flanquea a la mutación en estudio (también conocido como huella congénica) y otras estrategias para la detección de loci de carácter cuantitativo de expresión (eQTL) como posibles genes modificadores de fenotipos. Como prueba de concepto, identificamos genes dependientes de *SALL2* desde el transcriptoma derivado del KO para *Sall2* y posteriormente validamos estos genes mediante PCR cuantitativa en tiempo real (qPCR), incluyendo la validación cruzada con otros estudios. Demostramos la existencia de la huella congénica en el modelo e identificamos angiogenina (*Ang*) como un gen modificador de varios fenotipos que se producen en el KO para *Sall2*. Sin embargo, también identificamos a *ANG* como un blanco transcripcional de *SALL2*, conservado en células de ratón y humanas. En resumen, las estrategias presentadas son herramientas convenientes para evaluar el fondo genético, realizar la validación cruzada de genes diferencialmente expresados, identificar potenciales genes modificadores de fenotipos

e interpretar los fenotipos derivados de estudios que empleen modelos genéticamente modificados.



## SUMMARY

In this thesis we aimed to unravel the transcriptional targets of the *Sall2* transcription factor by using a knockout (KO) mice model. For this purpose, we conducted RNA-Seq studies to understand the *Sall2*-dependent transcriptional response by using a *Sall2* knockout model, dissecting a cohort of genes highly dependent on the genetic background where the *Sall2* KO has been maintained. Based on the latter and supported by the available literature, we proposed a pipeline to determine genetic introgression in coding genes by using RNA-Seq data; including strategies to uncover variants linked to the targeted locus and potential modifier genes in genetically modified organisms (GEMs). The above methodology was implemented analyzing publicly available RNA-Seq data from five congenic KO models plus experimental RNA-Seq data from our *Sall2* KO model. The applied methods demonstrated the extent of the 129-derived ESC genome introgression in genome-wide graphs, the detection of the flanking embryonic stem cell (ESC)-derived chromosome (also known as the congenic footprint) and strategies for the detection of expression quantitative trait loci (eQTLs), including potential modifier genes. By applying this computational approach, we dissected *Sall2* responsive genes from the transcriptome of the *Sall2* KO; validating these genes by quantitative real time PCR (qPCR) including cross-validation with other studies. As a proof of concept, we demonstrated the existence of the congenic footprint and the identification of angiogenin (*Ang*) as a modifier gene of several phenotypes occurring in the *Sall2* KO model. Nevertheless, we also identified *ANG* as a conserved transcriptional target of *SALL2* in mouse and human cells. In summary, our novel strategy is a convenient tool to assess genetic background, cross-validation of responsive genes, identification of putative modifier genes and the interpretation of phenotypic outcomes in GEM models.

## TABLA DE CONTENIDOS

**RESUMEN (iii)**

**SUMMARY (v)**

**INDICE DE FIGURAS (ix)**

**INDICE DE TABLAS (xii)**

**ABREVIATURAS (xiii)**

### **I. INTRODUCCIÓN (1)**

Uso de modelos murinos en transgénesis (1)

Consideraciones genéticas en la generación de ratones KO mediante recombinación homóloga (6)

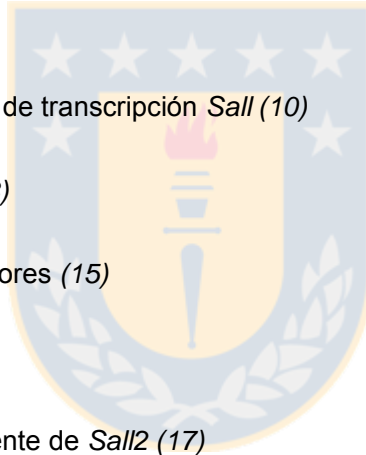
*Sall2* y la familia de factores de transcripción *Sall* (10)

Función normal de *Sall2* (13)

*Sall2* como supresor de tumores (15)

*Sall2* como oncogén (16)

Regulación génica dependiente de *Sall2* (17)



### **II. HIPÓTESIS DE TRABAJO Y OBJETIVOS (19)**

### **III) MATERIALES Y MÉTODOS (20)**

## MATERIALES (20)

Plásmidos (20)

Partidores (21)

Anticuerpos y fluoróforos (23)

## MÉTODOS DE LABORATORIO (25)

Aislamiento de fibroblastos primarios de ratón (MEFs) y cultivo celular (25)

Congelamiento de fibroblastos embrionarios de ratón (25)

Genotipificación (26)

Cultivo celular (27)

Estrés genotóxico o deprivación de suero (27)

Inmunocitoquímica (28)

Detección de proteínas (28)

Construcciones de ADN y digestión preparativa (30)

Transformación en bacterias competentes por calcio (30)

Minipreparación de ADN (31)

Purificación de ADN a gran escala (32)

Electroforesis en geles de agarosa (33)

Transfección transiente en células HEK293 (33)

Transducción de shRNA para SALL2 en células MEF inmortales (iMEFs) (33)

Electroporación de células de mamífero (34)

Generación de silenciamiento y KO mediante el sistema CRISPR-Cas9 (34)

Purificación de ARN total con reactivo TRIzol™ (35)

Preparación de muestras para ARN-seq y purificación de ARN mediante kit (36)

Transcripción inversa acoplada a PCR (37) PCR

cuantitativa en tiempo real (qPCR) (37)

Inmunoprecipitación de cromatina (38)

## MÉTODOS COMPUTACIONALES (39)

Secuenciación de muestras para ARN-seq (39)

Estrategia de alineamiento de las lecturas de ARN-seq al genoma de referencia (40)

Ordenamiento de los archivos en formato BAM y cuantificación de transcritos (42)

Análisis de la expresión diferencial de genes (44)

Llamado de variantes desde los datos de ARN-seq (53)

Visualización de variantes en R (54)

## IV. RESULTADOS (56)

**OBJETIVO 1: Determinar respuestas transcripcionales dependientes de *Sall2* (56)**

**OBJETIVO 1.1: Definir blancos transcripcionales específicos de *SALL2* *in vitro* e *in vivo* (56)**

Experimento de ARN-seq en células MEF *Sall2* silvestres y KO (56)

Detección de variantes a nivel de genoma completo desde datos de ARN-seq (59)

Disección de variantes ligadas al genotipo KO: la huella congénica (88)

Ploidía de huella congénica (95)

La huella congénica influye en la expresión génica de las células MEF KO para *Sall2* (106)

Cribado de loci de carácter cuantitativo de expresión (eQTL) en la región congénica del modelo KO para *Sall2* (120)

Angiogenina como un potencial gen modificador de fenotipos en el modelo KO para *Sall2* (130)

**OBJETIVO 1.2: Investigar el efecto de SALL2 en regiones promotoras de genes blanco** (146)

Búsqueda de sitios de unión de SALL2 en genes diferencialmente expresados (146)

Angiogenina como blanco transcripcional de SALL2 (151)

**OBJETIVO 2: Seleccionar blanco(s) transcripcional(es) de SALL2 asociados a fenotipo canceroso y realizar ensayos funcionales de estos genes y su dependencia de SALL2** (158)

**OBJETIVO 2.1: Analizar el efecto de pérdida/ganancia de la función de genes asociados con SALL2 provenientes del objetivo 1 y corroborar efectos celulares** (158)

Interferencia genética de *Cdkn1a* por angiogenina, un blanco canónico de SALL2 (158)

Otros efectos fenotípicos en MEF *Sall2* KO inherentes a su expresión génica (162)

**V. DISCUSIÓN** (169)

**VI. CONCLUSIONES GENERALES** (179)

**VII. AGRADECIMIENTOS** (181)

**VIII. BIBLIOGRAFÍA** (182)

**IX. ANEXOS** (189)



## ÍNDICE DE FIGURAS

**Figura 1:** Distribución de las subespecies que dieron origen a las cepas endogámicas clásicas. (2)

**Figura 2:** Visión general de las variantes desde 17 genomas de ratón en relación con el genoma de referencia de ratón (C57BL/6J). (3)

**Figura 3:** Esquema de la generación de un KO en células madres embrionarias (ESC) mediante recombinación homóloga. (5)

**Figura 4:** Fuentes de variación genética identificadas en ratones congénicos KO. (9)

**Figura 5:** Representación esquemática de los dominios conservados presentes en las proteínas SALL (11)

**Figura 6:** Representación esquemática del gen SALL2 y sus promotores alternativos (12)

**Figura 7:** Modelo de acción de *Sall2* en células neuronales (14)

**Figura 8:** Escalamiento multidimensional del experimento de ARN-seq (58)

**Figura 9:** Protocolo computacional para la detección de variantes introgresadas derivadas de ESC en estudios de secuenciación masiva en ratón (66)

**Figura 10:** Detección de variantes de la célula madre 129P2OlaHsd en el experimento de secuenciación del modelo KO para *Sall2* (68)

**Figura 11:** Detección de la totalidad de las variantes en el experimento de secuenciación del modelo KO para *Sall2* (70)

**Figura 12:** Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Mecp2* (72)

**Figura 13:** Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Gtf2ird1* (74)

**Figura 14:** Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Stc1* (76)

**Figura 15:** Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Itch* (78)

**Figura 16:** Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Hnrnpd* (*AUF-1*) (80)

**Figura 17:** Detección y distribución de variantes a nivel de genoma completo en los modelos KO analizados (85)

**Figura 18:** Detección de la mutación inactivante del gen *Casp4* a través de las réplicas biológicas del estudio de secuenciación del modelo KO para *Hnrnpd* (87)

**Figura 19:** Protocolo computacional para la detección de variantes ligadas al genotipo KO (89)

**Figura 20:** Disección de variantes ligadas al genotipo KO: la huella congénica (91)

**Figura 21:** Comparación de los genes congénicos detectados en el modelo KO para *Sall2* frente a la predicción de genes ligados al genotipo KO con el método Me-PaMuFind-it (95)

**Figura 22.** Ploidía de la huella congénica en el modelo KO para *Sall2* (97)

**Figura 23:** Confirmación genómica de las variantes de ratones KO para *Sall2* mediante el uso del gen congénico *Rpph1* como marcador (102)

**Figura 24.** Ploidía de la huella congénica en el modelo KO para *Stc1* (104)

**Figura 25.** Análisis de los genes diferencialmente expresados en MEFs silvestre versus KO para *Sall2* (108)

**Figura 26.** La huella congénica influye en la expresión génica de las células MEF KO para *Sall2* (110)

**Figura 27.** Detección del (ARNm) *Pnp* en células MEF *Sall2* KO (114)

**Figura 28.** Ontología genética para los 15 genes provenientes de la superposición del experimento de ARN-Seq con el estudio de microarreglos de la inducción de *Sall2* (117)

**Figura 29.** Validación cruzada de los genes diferencialmente expresados sensibles a SALL2 en células humanas HEK293 (118)

**Figura 30:** Protocolo para la detección de loci de carácter cuantitativo de expresión (eQTL) a partir de datos de ARN-Seq (122)

**Figura 31:** Impacto del tratamiento con doxorubicina como una perturbación global de la expresión génica e histograma de las variantes ligadas al genotipo *Sall2* KO con y sin el tratamiento con doxorubicina (125)

**Figura 32:** Detección de loci con rasgos cuantitativos de expresión (eQTL) en el modelo KO para *Sall2* con o sin perturbación con doxorubicina (129)

**Figura 33:** *Ang* congénico es un eQTL en comparación con C57BL/6J (133)

**Figura 34:** Disminución del (ARNm) *Ang* desde cepas “129” (135)

**Figura 35:** La expresión de *RNASE4* depende de la expresión de *ANG* (138)

**Figura 36.** Regulación del gen *Ang/Rnase4* en células KO y silenciadas para *Sall2* (140)

**Figura 37.** Detección del ARNm de *Ang* desde un modelo con la delección del gen murino *Sall2* mediante CRISPR/Cas9 (141)

**Figura 38.** Búsqueda de sitios de unión de SALL2 en los promotores de genes sensibles a SALL2 (148)

**Figura 39.** Sitios de unión de SALL2 identificados en el promotor murino que controla el gen *Ang/Rnase4* (149)

**Figura 40:** Ensayo de luciferasa con el promotor que controla el gen *Ang/Rnase4* murino electroporado en MEF *Sall2* silvestres y KO (151)

**Figura 41.** Sitios de unión de SALL2 identificados en el promotor humano que controla el gen *ANG/RNASE4* (153)

**Figura 42:** Identificación del gen *ANG/RNASE4* como un blanco transcripcional de SALL2 en células K562 (155)

**Figura 43.** Interferencia genética de *Cdkn1a* por angiogenina, un blanco canónico de SALL2 (160)

**Figura 44.** Razón de traducción en el modelo MEF *Sall2* KO versus el modelo de CRISPR/Cas9 para *Sall2* (163)

**Figura 45:** Capacidad de formación de gránulos de estrés (SG) en células *Sall2* KO versus el modelo silenciado para *Sall2* (166)

**Figura Anexo I:** Protocolo para la supervisión genética de modelos genéticamente modificados (GEMs) (192)

## ÍNDICE DE TABLAS

**Tabla 1:** Grado de alineamiento del experimento de secuenciación en MEF silvestres y KO para *Sall2* (57)

**Tabla 2.** Prueba de Cochran-Armitage en los 6 experimentos de secuenciación analizados (82)

**Tabla 3.** Lista de los genes congénicos presentes en el modelo KO para *Sall2* (93)

**Tabla 4.** Diferencias en la expresión génica entre el modelo *Sall2* “knockout” y el modelo de silenciamiento para *Sall2* (113)

**Tabla 5.** Genes diferencialmente expresados que se localizan en la huella congénica del modelo *KO* para *Sall2* (121)

**Tabla 6.** Genes congénicos adicionales expresados en el tratamiento con doxorubicina en las MEF provenientes del modelo *KO* para *Sall2* (125)

**Tabla 7.** Diferencias en los fenotipos caracterizados entre el modelo *Sall2* “knockout” y el modelo de silenciamiento para *Sall2* (168)

#### **ABREVIATURAS:**

**ADN:** Ácido desoxirribonucleico

**ADNc:** Ácido desoxirribonucleico copia

**ARN:** Ácido ribonucleico

**ARN-seq:** secuenciación de ácido ribonucleico

**BAM:** Binary Alignment File

**BSA:** “Bovine serum albumin”

**ChIP:** “Chromatin Immunoprecipitation”

**CRISPR:** “Clustered Regularly Interspaced Short Palindromic Repeats”

**c-Myc:** “V-Myc Avian Myelocytomatosis Viral Oncogene Homolog”

**DEG:** “differentially expressed gene”

**DEPC:** Dietilpirocarbonato

**DMEM:** “Dulbecco’s Modified Eagle Medium”

**DMSO:** Dimetilsulfóxido

**eQTL:** “Expression quantitative trait loci”

**ESC:** “Embryonic Stem Cell”

**FIAU:** “1-(2-deoxy-2-fluoro-1-D-arabinofuranosyl)-5-iodouracil”

**FBS:** “Fetal Bovine Serum”

**G418:** “Geneticin”

**KO:** “Knockout”

**LB:** Luria-Bertani

**LiCl:** “Lithium chloride”

**MEF:** “Mouse Embryonic Fibroblast”

**mARN: Ácido ribonucleico mensajero**

**NaCl: “Sodium Chloride”**

**NP-40: “Nonyl phenoxyethoxyethanol”**

**PBS: “Phosphate-Buffered Saline”**

**PNP: “Purine Nucleoside Phosphorilase”**

**PCR: “Polymerase Chain Reaction”**

**SDS: “Sodium Dodecyl Sulfate”**

**SDS-PAGE: “Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis”**

**shRNA: “Short Hairpin ARN”**

**SNP: “Single Nucleotide Polymorphism”**

**TAE: “Tris base, Acetic acid and EDTA”**

**TE: “Tris + EDTA”**

**Tris-Cl: “Tris base + hydrochloric acid”**

**TBS: “Tris-buffered saline”**

**WT: “Wild Type”**



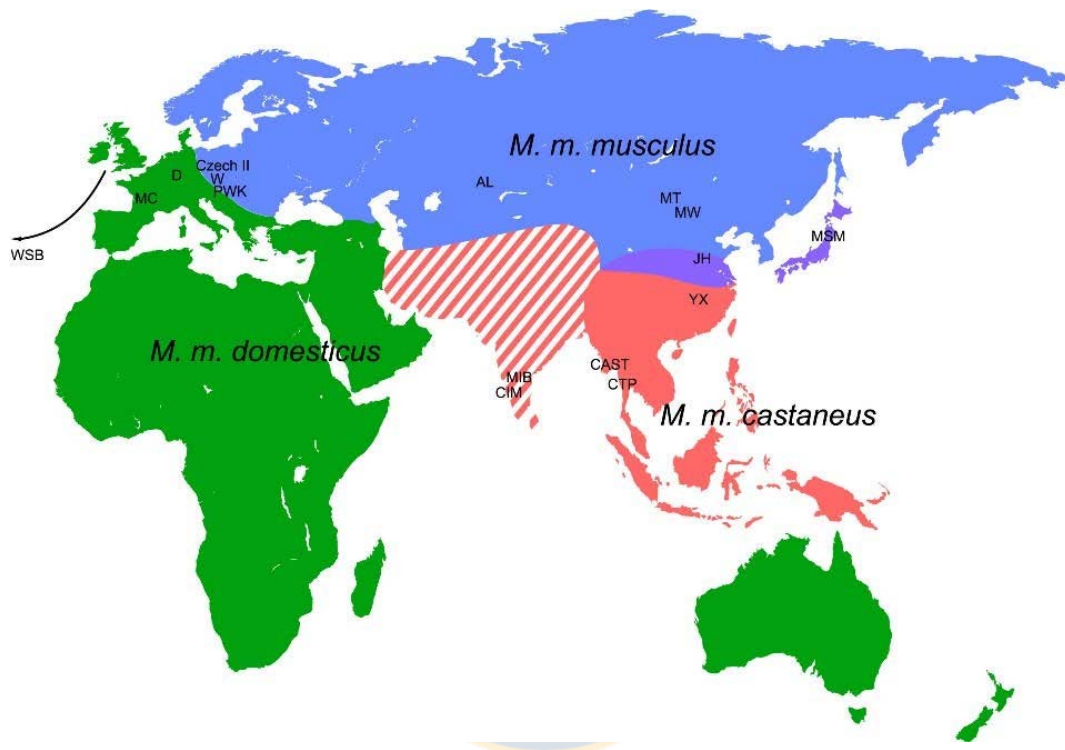
## I. INTRODUCCIÓN

### Uso de modelos murinos en transgénesis

El ratón actual de laboratorio deriva de la especie *Mus musculus*, perteneciente al género *Mus* del orden de los mamíferos. Se ha reportado que la especie *Mus musculus* divergió hace aproximadamente 2 millones de años respecto a las otras subespecies del género *Mus*, incluyendo, *Mus. spretus* y *Mus spicilegus* (Nishioka et al. 1994). La secuenciación genómica de las especies *Homo Sapiens* y *Mus musculus* reveló que *Homo Sapiens* comparte el 99% de los genes con *Mus musculus*, lo cual indica que compartimos la mayoría de los bloques moleculares evolutivos con esta especie (<http://www.informatics.jax.org/>). A lo largo de la historia la especie y subespecies derivadas de *Mus musculus* se han seleccionado como mascotas y/o animales de experimentación debido a su rápido crecimiento, la posibilidad de criarlos bajo consanguinidad y su relativo bajo costo de mantención (Paigen 2003). Estas características han probado ser ventajosas en *Mus musculus* como modelo en ciencias biológicas.

En base a lo anterior, el uso de los modelos murinos en ciencia ha proporcionado un conocimiento exhaustivo en el estudio de la función génica, el modelaje de las enfermedades humanas para su posterior caracterización fisiológica *in vivo* y en el análisis genómico comparativo/evolutivo. El ratón de laboratorio moderno es el resultado de una cuidadosa selección de rasgos que comenzó a principios del siglo XX (Atchley and Fitch 1991). Esto condujo consecutivamente a la generación de cepas murinas endogámicas, ratones obtenidos por al menos 20 generaciones consecutivas de apareamiento hermano-hermana que se desarrollaron originalmente para el estudio del cáncer (Bittner et al. 1944). Después de 20 generaciones, los *loci* de los ratones son homocigóticos en un promedio de un 98,6%, lo que hace teóricamente posible construir un perfil genético de la cepa en estudio caracterizando sólo un individuo de la colonia (Sterken and Kiryluk 2010). Se ha caracterizado que las

cepas endogámicas utilizadas en la actualidad se derivan de al menos tres subespecies murinas derivadas de la especie *Mus musculus* repartidas entre África, Asia y el sudeste asiático. Estas subespecies corresponden a *Mus musculus musculus*, *Mus musculus domesticus* y *Mus musculus castaneus*.

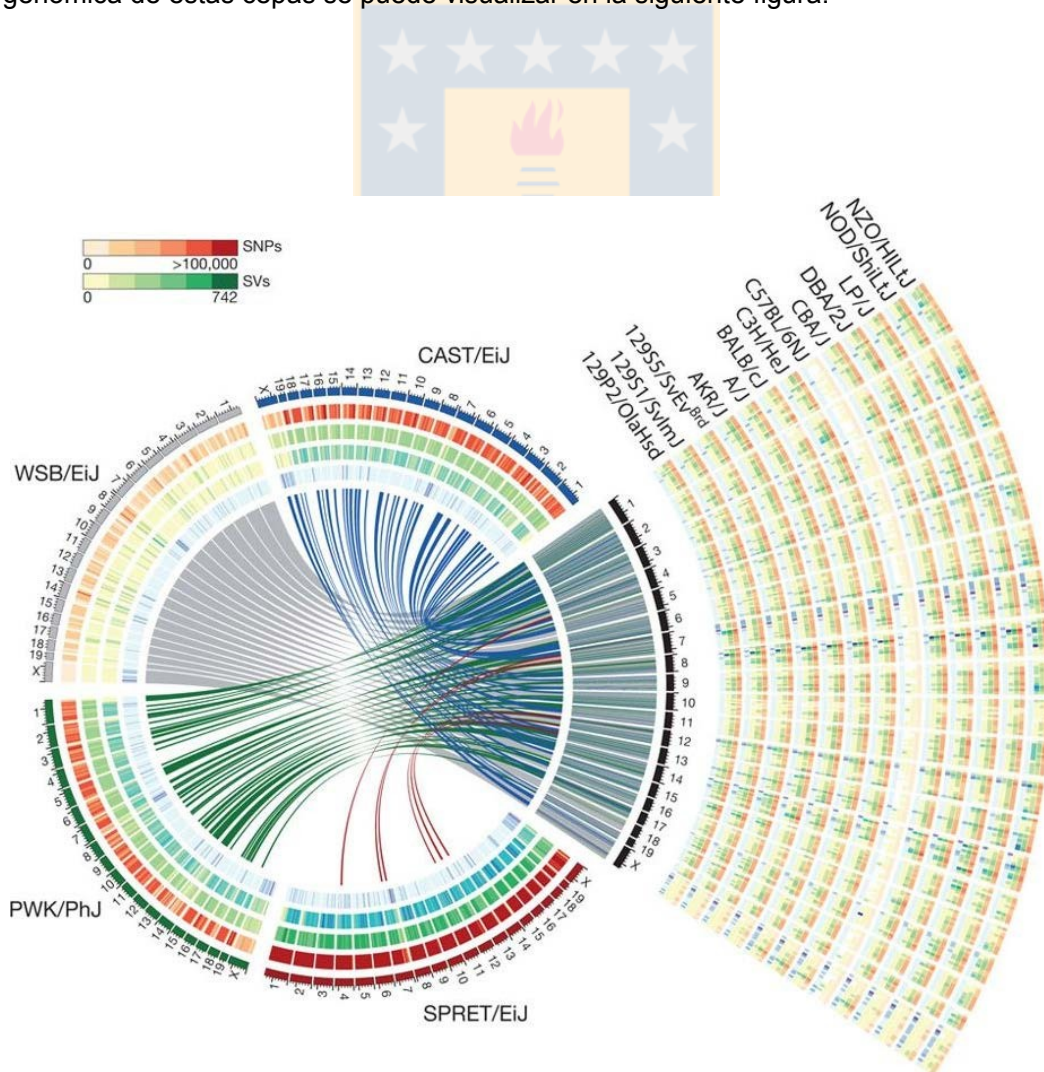


**Figura 1: Distribución de las subespecies que dieron origen a las cepas endogámicas clásicas.** La subespecie *Mus musculus musculus* es originaria de Asia central, abarcando parte de Europa y el sudeste asiático, mientras que la subespecie *Mus musculus domesticus* es originaria de África y *Mus musculus castaneus* del sudeste asiático. Estas subespecies son originarias del género *Mus* y divergieron hace 500000 años. Basado de (Lilue et al. 2013).

Keane et al 2011 reportó la secuenciación de genoma completo de las 17 cepas de ratón más utilizadas e identificó 10 veces más cantidad de variantes respecto a las que se habían reportado previamente en cepas murinas. Se identificó en las



cepas secuenciadas que la gran mayoría de las variantes estructurales incluyendo las variantes polimórficas de un solo nucleótido (derivado del inglés “single nucleotide polymorphisms”) provenían de las subespecies del género *Mus* que dieron origen a estas cepas. En base a lo anterior, este estudio reveló que al menos el 12% de los transcritos expresados muestran un sesgo de expresión significativo dependiendo del tejido, dado por las variaciones genéticas asociadas a las cepas. Adicionalmente se identificaron 718 *loci* de rasgos cuantitativos de expresión (del inglés “Expression Quantitative Trait Loci”) que potencialmente contribuyen a fenotipos en estas cepas. Por tanto, la constitución genética de los modelos genéticamente modificados y otros debe tenerse en cuenta dependiendo de la cepa(s) empleada(s). La caracterización genómica de estas cepas se puede visualizar en la siguiente figura.

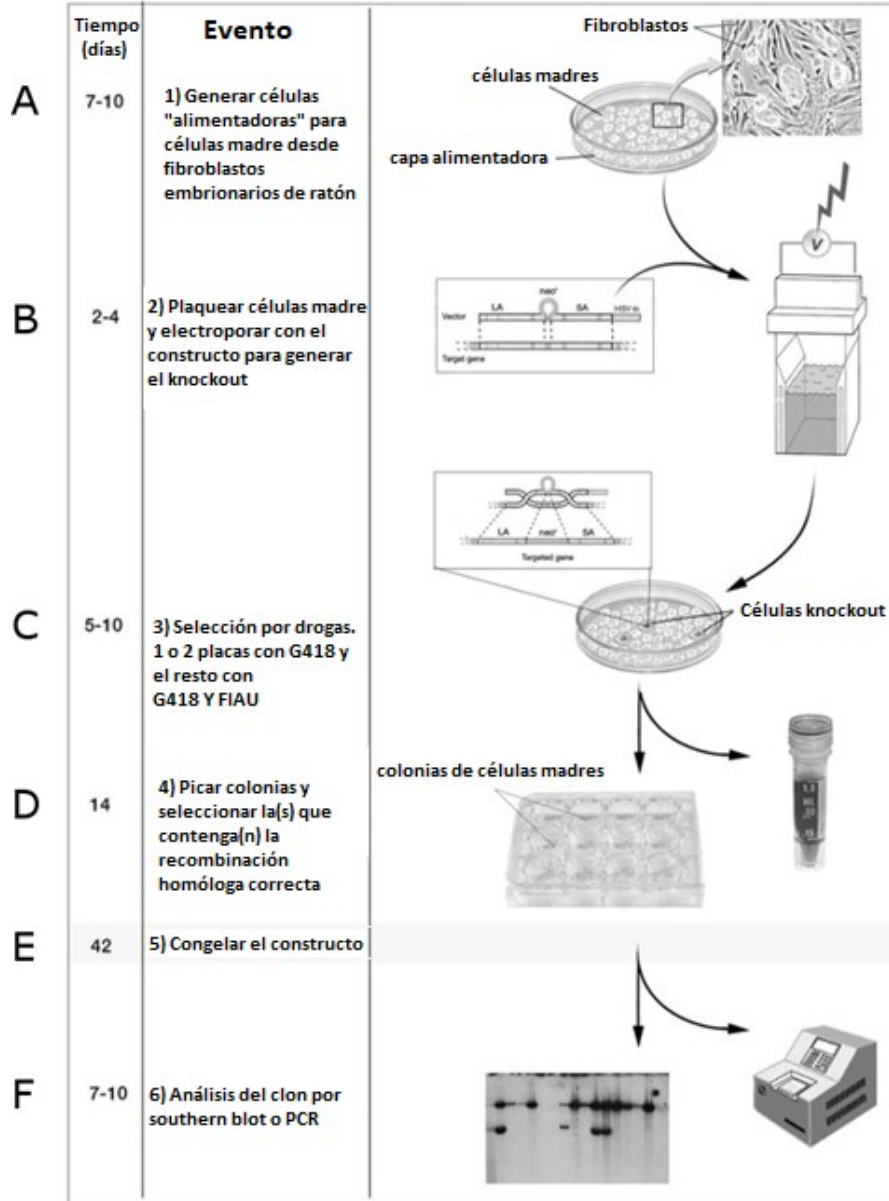


**Figura 2: Visión general de las variantes desde 17 genomas de ratón en relación con el genoma de referencia de ratón (C57BL/6J).** Las cuatro cepas de origen silvestre (CAST/EiJ, WSB/EiJ, PWK/PhJ y SPRET/EiJ) son representativas de las subespecies *Mus musculus castaneus*, *Mus musculus musculus*, *Mus musculus domesticus* y *Mus spretus* e incluyen los progenitores de los cuales se derivaron las cepas clásicas de laboratorio. Estos genomas se muestran en un círculo con líneas que indican la densidad relativa de los polimorfismos de un solo nucleótido (véase “SNPs”) y las variantes estructurales (véase “SV”). A la derecha del círculo, se muestra la contribución de los genomas correspondientes para cada una de las 13 cepas clásicas de laboratorio, por cromosoma. Los enlaces que cruzan el círculo indican regiones contenedoras de 375 kb. Modificado de (Keane et al. 2011).

Algunas de las cepas endogámicas de laboratorio han demostrado tener características ventajosas para la tecnología de la transgenia en células madre embrionarias (ESC) (Taketo et al. 1991). La mayoría de las ESC disponibles para la transgenia han sido derivadas de subcepas provenientes de la cepa "129", generada a partir de cruces de ratones de colores provenientes de criadores ingleses con chinchillas provenientes de la región de Castle (Reino unido), en 1928 (Threadgill et al. 1997). Estas cepas y las subcepas derivadas de ella contienen una mutación denominada “gen de dilución de ojos rosados” (del inglés “pink-eyed dilution” gene) la cual es derivada desde ratones asiáticos del tipo *Mus musculus* (Brilliant et al. 1994). Las ESC derivadas de la cepa “129” han sido particularmente exitosas en la transmisión de la línea germinal y se han usado ampliamente en más de 5000 experimentos de tipo KO (Limaye et al. 2009; Guan et al. 2010). Típicamente, la inactivación de genes en el ratón se realiza en ESC derivadas de 129. Las ESC se electroporan con un constructo que hace recombinación homóloga en el gen a inactivar, el cual se reemplaza con un casete de resistencia al antibiótico Neomicina. Después de seleccionar con Neomicina y explorar un clon ESC que albergue el alelo nulo, este clon se inyecta en los blastocistos de ratón de la cepa pionera C57BL/6

receptora (subcepa N o J), obteniendo un ratón congénito por posterior retrocruzamiento (Seong et al. 2004).

### Esquema de electroporación de células madres y selección



**Figura 3: Esquema de la generación de un KO en células madre embrionarias (ESC) mediante recombinación homóloga.** (A) Las células madre embrionarias (ESC) son crecidas sobre una capa de fibroblastos embrionarios (MEF) previamente irradiados o inactivados en crecimiento como soporte de las ESC. (B) Las ESC pueden incorporar un constructo adecuado para la recombinación homóloga mediante electroporación (C) Distintos clones electroporados pueden ser seleccionados mediante drogas, ya que el constructo contiene genes de resistencia para estas. (D) Las colonias resistentes son crecidas para congelamiento y para obtener ADN genómico. (E) Cuando se alcanza un número suficiente de células, los clones son congelados (F) Técnicas como “Southern Blot” y/o PCR pueden ser empleadas para encontrar el(los) clones que contengan la mutación deseada. Modificado de (Limaye et al. 2009).

### **Consideraciones genéticas en la generación de ratones KO mediante recombinación homóloga**

La estrategia anteriormente mencionada tiene importantes desventajas desde el punto de vista genético. Dado que los ratones resultantes en las generaciones iniciales contienen la introgresión del genoma proveniente de la célula madre y de la cepa utilizada en la retrocruza (usualmente C57BL/6J), los experimentos deben retrasarse al menos diez cruza con C57BL/6J para eliminar la introgresión del genoma de la ESC (Eisener-Dorman et al. 2009). Contrario a lo anterior, si es que se decide estudiar la inactivación de un gen en un fondo genético mixto, el análisis fenotípico puede ser complejo, ya que cada individuo transgénico puede contener un genotipo diferente. Esto último se debe a que los individuos con fondo genético mixto pueden exhibir una segregación diferencial del fondo genético (Gerlai 2001; Gerlai 2016). Por otro lado, la elección de la célula madre embrionaria (ESC) puede ser una fuente importante de contaminación y deriva génica; considerando que estas células pueden contener un fondo genético desconocido o contaminado con el genoma de otras cepas murinas (Threadgill et al. 1997). En la práctica, se ha observado que si se retrocruza los ratones transgénicos de forma extensiva, se obtendrán líneas congénicas; ratones

genéticamente idénticos, a excepción de un locus particular en estudio. Esto es deseable en el estudio de un gen inactivado, ya que se elimina la introgresión de variantes derivadas de la ESC a lo largo del genoma. Sin embargo, es importante destacar que los ratones congénicos siempre contendrán el genoma de la ESC original cerca del gen inactivado, fenómeno genéticamente denominado como "huella congénica" (Schalkwyk et al. 2007). Las características anteriormente mencionadas pueden confundir el análisis de los fenotipos asociados al gen inactivado y se pueden resumir de la siguiente manera:

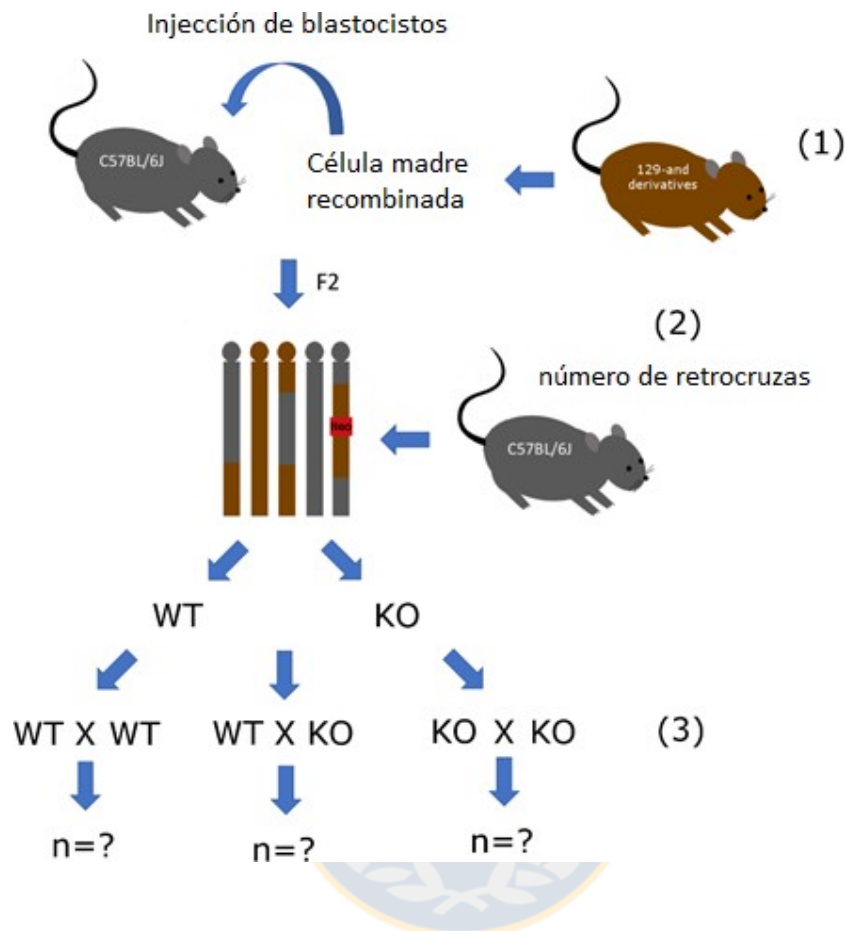
- 1) El fondo genético en donde se mantiene el modelo KO influye profundamente en los fenotipos resultantes, fenómeno bien descrito en la literatura (Clarke et al. 1992; Baribault et al. 1994; Threadgill et al. 1995; Perry et al. 1998; Kallapur et al. 1999). En el caso en que se decida estudiar la mutación en un fondo genético mixto y no en líneas congénicas, los fenotipos asociados al gen inactivado serán estudiados en un fondo genético mixto, con una introgresión genética compleja. Este tipo de estudios han probado tener una baja reproducibilidad debido a la genética única que contienen los modelos en uso. Este tópico fue revisado por (Rivera and Tessarollo 2008; Doetschman 2009).
- 2) La existencia de la huella congénica conduce a una comparación errónea entre los ratones silvestres y los transgénicos (Vanden Berghe et al. 2015). De acuerdo con esto, varios reportes demuestran que los cambios más drásticos en la expresión génica se centran en genes congénicos, sin contar las interacciones genéticas de estos genes con el fondo genético (Schalkwyk et al. 2007; Koks et al. 2009; Noyes et al. 2010; Szabo et al. 2016). Esto supone un sesgo en la disección de un transcriptoma dependiente de la mutante, si está presente, incluyendo la asociación de fenotipos erróneos al gen inactivado (Weissmann and Flechsig 2003; Crusio 2004; Pekarik and Izpisua Belmonte 2008; Vanden Berghe et al. 2015).

- 3) Los errores en la replicación del ADN y la transmisión de la línea germinal crearán naturalmente una variación genética a través de las generaciones de los ratones. Si no se controla el número de generaciones en las cuales se mantiene el gen inactivado en estudio, las mutaciones pasajeras se acumularán, y pueden volverse fijas en la población. Lo anterior puede causar que las cepas deriven genéticamente a lo largo del tiempo, sesgando la asociación de fenotipos al gen inactivado (Casellas and Medrano 2008).

Los sesgos anteriormente referidos deben considerarse cuando se llevan a cabo estudios con animales KO de tipo congénico. Dado lo anterior, es deseable generar métodos rentables para supervisar la genética en el uso de modelos murino.

Respecto a esto, se ha descrito que el genoma de ratón se puede explorar de forma eficiente mediante el uso de paneles de genotipado de polimorfismos de un solo nucleótido (del inglés "SNP genotyping panels") (Petkov et al. 2004), pero como limitantes, estos métodos analizan un número limitado de sondas y no detectan variantes nuevas, fijadas por la deriva génica.

En esta tesis, proponemos el uso de datos RNA seq para controlar la naturaleza, la ploidía y la cantidad de variantes introgresadas en modelos murinos KO, previo a estudiar el(los) fenotipo(s) y transcriptoma dependiente del gen inactivado en estudio. Es deseable además incluir la estimación de la huella congénica y su correspondiente contribución genética en la expresión génica asociada al gen inactivado. Estos procedimientos pueden ayudar a comprender mejor los resultados fenotípicos en los estudios con modelos KO y desentrañar nuevas interacciones genéticas en modelos de ratón. En lo específico, se pretendió identificar blancos transcripcionales del gen *Sal12* mediante el uso de un modelo murino "knockout" (KO) para *Sal12*, disponible en nuestro laboratorio y descrito por Sato y colaboradores (Sato et al. 2003).

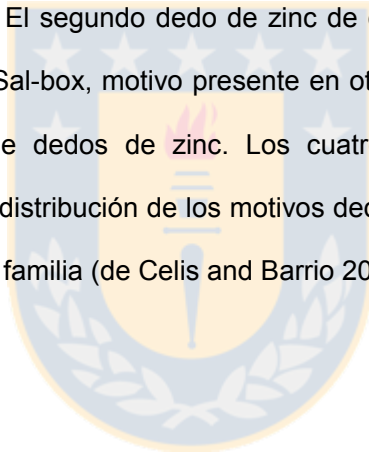


**Figura 4: Fuentes de variación genética identificadas en ratones congénicos KO.**

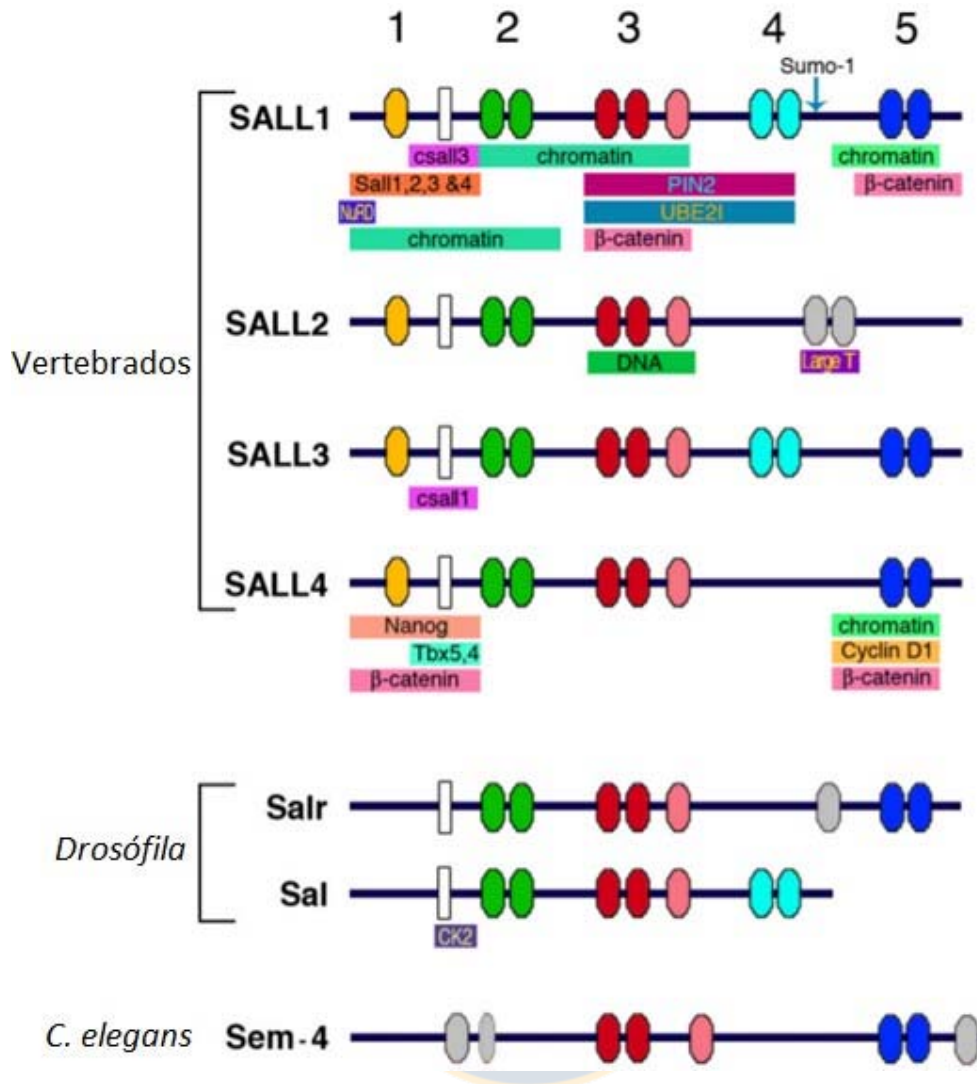
Como se describió en primer lugar, la elección de ESC derivado de los ratones marrones endogámicos es una fuente importante de variación genética y la caracterización genética de estas células podría no estar disponible (1A). En segundo lugar, el número de retrocruzamientos con C57BL/6 (subcepa N o J) dará lugar a una introgresión variable de variantes de las ESC entre los animales de la camada, dependiendo de cuántos retrocruzadas se realicen (2A). Finalmente, la estrategia de mantenimiento del gen inactivado a lo largo de las generaciones determinará la constitución final de variantes en el tipo silvestre y en el mutante, incluida la extensión de la huella congénica (3A). El N de cruces también determinará el número de nuevas variantes fijas en animales, definido por la deriva génica. Fuente: Elaboración propia.

## ***Sall2* y la familia de factores de transcripción *Sal***

Las proteínas SALL (“Sal-like”) están conservadas en vertebrados por su rol en el desarrollo, dado que la actividad de la proteína *Sal* en *Drosophila* es requerida para la especificación de los placodos traqueales durante la embriogénesis tardía, entre otros procesos de diferenciación en tejidos embrionicos y larvales (Kuhnlein and Schuh 1996). Estas proteínas poseen en común múltiples dominios de dedos de zinc distribuidos a lo largo de sus secuencias, lo que los clasifica como factores de transcripción (Kuhnlein et al. 1994; Kohlhase et al. 1996). A continuación, se muestra en la figura 1 que el dominio dedo de zinc 1 corresponde a la clase C2HC y está presente sólo en homólogos vertebrados; el resto de los dominios (2-5) son del tipo C2H2 ordenados en pares. El segundo dedo de zinc de cada par contiene un motivo característico denominado Sal-box, motivo presente en otros factores de transcripción que presentan dominios de dedos de zinc. Los cuatro ortólogos en vertebrados presentan diferencias en la distribución de los motivos dedos de zinc, siendo SALL2 el miembro más distante de la familia (de Celis and Barrio 2009).

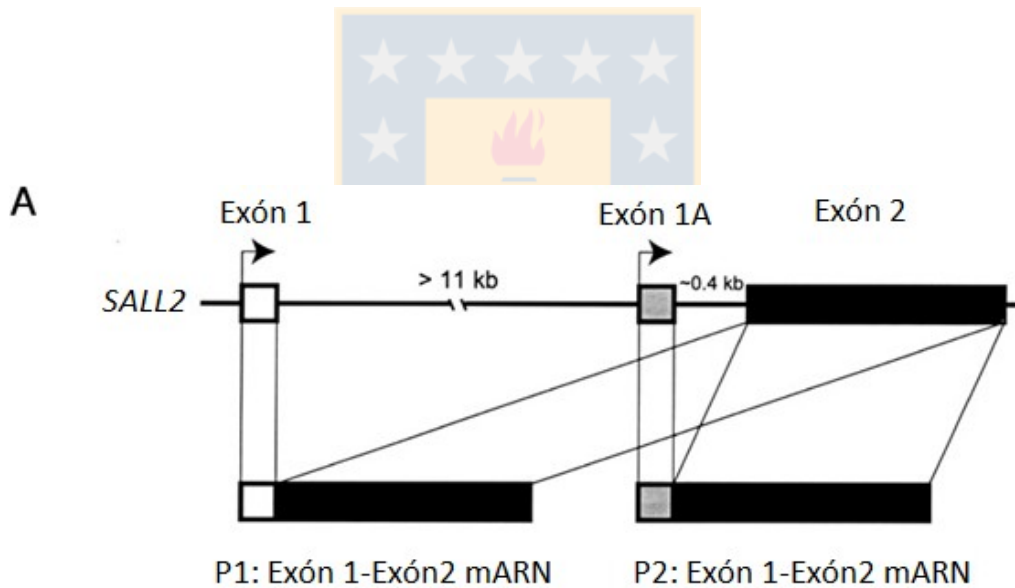






**Figura 5: Representación esquemática de los dominios conservados presentes en las proteínas SALL.** Los óvalos coloreados numerados del 1 al 5 representan los dominios de dedos de zinc conservados en las proteínas homólogas SALL de vertebrados, *Drosophila* y *C. elegans*, respectivamente. Los rectángulos blancos representan las regiones poli Q (repeticiones de glutámico). La flecha indicada en el gen *Sall1* denota un sitio de sumoilación descrito en esta proteína. Las barras coloreadas horizontales debajo de cada proteína indican los dominios de interacción en las proteínas SALL con otras proteínas. Derivado de (de Celis and Barrio 2009).

Se ha demostrado que existen dos isoformas de la proteína SALL2, las cuales se generan debido a la presencia de dos promotores alternativos en el extremo 5' del gen *SALL2* humano (véase exones E1 y E1A) y comparten un exón principal (Exón 2). Cada promotor conduce a la generación de (ARNm) distintos con una expresión diferencial de las isoformas. Se ha descrito que la isoforma E1A tiene expresión ubicua mientras que la isoforma E1 se expresa mayormente en cerebro, riñón y menormente en pulmón, testículo e intestino (Ma et al. 2001). El exón E1 codifica una secuencia aminoacídica de 27 residuos y el exón E1A una secuencia de 25 residuos, por lo que ambas isoformas difieren en composición aminoacídica en el extremo N terminal, como se muestra en la figura 2.



**B**

Exón 1:            MSRRKQRKPQ QLISDCEGPS ASENG

Exón 1A:            MAHESERSSR LGVPCGEPAE LGG

**Figura 6: Representación esquemática del gen SALL2 y sus promotores alternativos.** A) Esquema de la distribución de los exones que dan origen a las isoformas E1 y E1A controladas por los promotores P1 y P2, respectivamente. B) Secuencia aminoacídica de SALL2 humano codificada en exones E1 y E1A respectivamente. Derivado de (Ma et al. 2001).

Nuestro laboratorio se dedica al estudio de las funciones normales y oncogénicas del segundo miembro de la familia SALL en humano y ratón, *Sall2*.

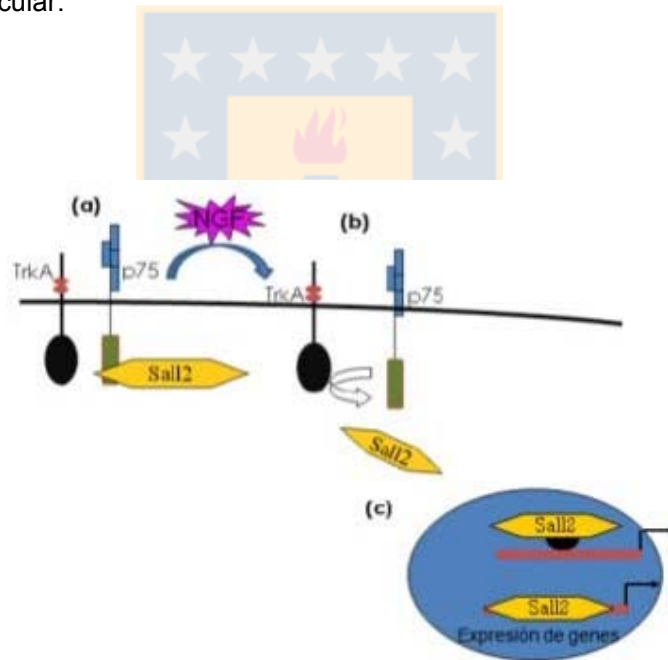
### **Función normal de *Sall2***

En el año 2003, Sato y colaboradores generaron un modelo de ratón deficiente para *Sall2*, en el cual cinco dominios de dedos de zinc situados en el extremo N-terminal del exón 2 fueron deletados mediante la generación de un constructo diseñado para recombinación homóloga (Sato et al. 2003). No obstante, este modelo no presentó un fenotipo anormal, demostrando que la funcionalidad de *Sall2* es dispensable para el desarrollo. Por el contrario, en el año 2008, Böhm y colaboradores generaron un modelo de ratón deficiente para *Sall2* mediante la delección completa del gen, en donde se evidenció una cohorte de animales con anomalías en la formación del tubo neural (Bohm et al. 2008). En base a esta última evidencia se ha propuesto que *Sall2* cumple un rol en el desarrollo del sistema nervioso.

En esta línea, nuestro laboratorio identificó a SALL2 como un interactor del receptor de neurotrofinas p75 (Pincheira et al. 2009). Se demostró que la señalización mediada por NGF (del inglés “nerve growth factor”) disocia un complejo conformado por SALL2 y el receptor de neurotrofinas p75NTR, en donde SALL2 transloca al núcleo e induce la transcripción del gen blanco p21<sup>CIP-1/WAF-1</sup> (*Cdkn1a*). En este mismo estudio, el silenciamiento de *Sall2*, en células PC12 y de hipocampo, demuestra que SALL2 es necesario para la extensión de neuritas en respuesta a NGF y la inducción del gen

*Cdkn1a*, codificante para la proteína p21 y un mediador del arresto del ciclo celular en las fases G1-S. En base a estos estudios, se ha propuesto un modelo de señalización para SALL2 en células neuronales (véase figura 7).

Posteriormente, se identificaron mutaciones en *SALL2* que están ligadas al coloboma ocular recesivo, una enfermedad congénita que se caracteriza por un cierre incompleto de la fisura óptica, generando ceguera en los individuos (Kelberman et al. 2014). Estos defectos fueron corroborados en el modelo deficiente para *SalI2* utilizado por Böhm et al. 2008 y en una familia que presentaba una mutación del gen *SALL2* en el codón correspondiente al residuo glutamina 29, lo cual genera una proteína trunca no funcional. Estos estudios en su conjunto proponen que *SalI2* además cumple un rol en el desarrollo ocular.



**Figura 7: Modelo de acción de *SalI2* en células neuronales.** (A) En ausencia de neurotrofinas, SALL2 se mantiene localizado en el citoplasma unido al receptor p75NTR. (b) La presencia de NGF induce (c) la translocación de SALL2 al núcleo y la inducción de sus genes blanco incluyendo p21<sup>CIP-1/WAF-1</sup>. Derivado de (Pincheira et al. 2009).

## ***Sall2* como supresor de tumores**

Varios estudios apoyan que *SALL2* cumple un rol como supresor de tumores en ciertos tipos de cánceres. En humanos, el gen que codifica para la proteína *SALL2* se encuentra ubicado en el cromosoma 14, en la región 14q11.1–13, anteriormente reportada con pérdida de heterocigocidad en el mesotelioma maligno y cáncer de ovario (Bandera et al. 1997). Una investigación basada en la búsqueda de posibles supresores tumorales que interfirieran en la replicación del virus de polioma de ratón oncogénico, reveló que *SALL2* interactuaba con el antígeno largo de este virus, inhibiendo su replicación (Li et al. 2001). En base a estos estudios iniciales, se propuso a *Sall2* como un supresor tumoral, probablemente en cánceres donde esté ausente respecto a su expresión normal. Estudios posteriores determinaron que la reintroducción de *Sall2* en células de carcinoma ovárico humano (SKOV-3) inhibía la tasa de crecimiento y la capacidad para formar colonias tumorigénicas en agar blando. Además, ocurre una disminución substancial de la capacidad tumoral de estas células en ratones atímicos, acompañado con una inhibición de la síntesis del ADN y un aumento en la apoptosis celular (Li et al. 2004). En concordancia con estos estudios, se ha demostrado recientemente que el silenciamiento de *SALL2* en la línea celular de cáncer de ovario A2780 aumenta la proliferación celular, migración e invasión, inhibiendo de forma conjunta la apoptosis celular (Miao et al. 2017).

En cuanto a los mecanismos moleculares que explican la acción supresora de *SALL2* en estos cánceres figura la inducción transcripcional de  $p21^{CIP-1/WAF-1}$ , revisado en (Li et al. 2004) y la inhibición transcripcional del oncogén *c-myc*, el cual a su vez inhibe la apoptosis celular (Sung et al. 2012). De acuerdo con la función supresora de tumores de *SALL2* en cáncer de ovario, se ha reportado que, en una parte significativa de los carcinomas ováricos, *SALL2* se encuentra silenciado por metilación del promotor que controla la expresión de la isoforma E1A (Sung et al. 2013), fenómeno previamente descrito en los procesos de inactivación de los genes supresores

tumorales (Esteller 2002). Estas mismas observaciones fueron recientemente reportadas en el carcinoma de células esofágicas, en donde los niveles de (ARNm) *SALL2* estaban disminuidos 20 veces en un clon radioresistente de la línea celular esofágica TE-1 versus la línea control radiosensible (Luo et al. 2017). En este último estudio también se reportó la metilación del promotor de *SALL2* como mecanismo de silenciamiento en la línea radioresistente. Consistente con los estudios en cáncer de ovario, la reintroducción de *SALL2* en clones radioresistente de las líneas celulares de cáncer esofágico TE-1 y Eca-109, aumentó la apoptosis y la sensibilidad al quimioterapéutico cisplatino en dichas células. En resumen, estos estudios en su conjunto apoyan a *SALL2* como supresor de tumores en cáncer de ovario y en carcinoma de esófago.

### ***Sall2* como oncogén**

En relación con los estudios anteriores, otros reportes refutan la idea de que *SALL2* es un supresor tumoral. Un ejemplo de ello es el aumento de la expresión de *SALL2* en un 24,5% de casos de carcinoma oral de células escamosas (SCC) (Estilo et al. 2009). En este estudio, se propone que la expresión de *SALL2* actuaría como un potencial predictor de cánceres de SCC de cabeza y cuello. La expresión de *SALL2* también ha sido relacionada con la metástasis a ganglios linfáticos en cáncer de mama (Huang et al. 2003) y a través de microarreglos realizados en tejidos tumorales, en donde se determinó que la expresión nuclear de *SALL2* es elevada en un 43% de los casos de sarcoma sinovial (Nielsen et al. 2003). Otros estudios sugieren una asociación similar en la expresión de *SALL2* con tumorigénesis, como es en el caso del cáncer testicular (Alagaratnam et al. 2011) y tumor de Wilms en el riñón (Li et al. 2002). En el caso particular de los tumores en glioblastoma, se ha reportado que *SALL2* en conjunto con los factores de transcripción *POU3F2*, *SOX2* y *OLIG2* son necesarios para mantener la pluripotencia de las células madre de glioblastoma (Suva et al. 2014). Estos resultados contrastan con los datos que sugieren que *SALL2* es un gen supresor

tumoral. Lo anterior podría deberse a que este gen presentaría funciones contexto-específicas que al ser desreguladas en cáncer lo clasificarían como supresor tumoral y/o oncogén.

Si bien existen avances en relación con la función de *Sall2* en ciertos contextos celulares, no se conocen a cabalidad los blancos transcripcionales de *Sall2* que mediarían su función celular y su relación con cáncer.

### **Regulación génica dependiente de *Sall2***

Varios estudios han caracterizado blancos transcripcionales del gen *SALL2* tanto en células humanas como en modelos murinos. Un blanco bien caracterizado como se ha mencionado anteriormente es el gen *CDKN1A* (gen que codifica para la proteína p21<sup>CIP-1/WAF-1</sup>). Los estudios sobre la actividad supresora tumoral de *SALL2* en cáncer ovárico, indicaron que *SALL2* estimulaba la transcripción de *CDKN1A* de forma independiente de p53 (Li et al. 2004). En este estudio *SALL2* fue silenciado en células normales de la superficie epitelial de ovario humano (HOSE) y se observó una disminución de p21 de forma concomitante con el aumento de la síntesis de ADN. En un estudio posterior, se determinó que *SALL2*, al igual que el factor de transcripción específico 1 (Sp1), tiene preferencia por unirse a secuencias ricas en GC y se definió que su secuencia consenso es GGG(C|T)GG, mediante un ensayo SELEX *in vitro* (Gu et al. 2011). Dicha secuencia efectivamente se encuentra en el promotor de *CDKN1A*, y en este estudio se demuestra que tanto p21 como la proteína pro-apoptótica BAX son blancos transcripcionales regulados positivamente por *SALL2*.

Posterior a estos estudios, se identificó que la proteína supresora de tumores p16 (p16<sup>INK4A</sup>) es otro blanco transcripcional regulado positivamente por *SALL2*, regulando la proliferación celular. En este estudio se identificaron 1616 genes dependientes de la expresión de *SALL2* mediante un estudio de microarreglos de expresión génica y se procedió a validar a *CDKN2A* (gen que codifica para la proteína

p16<sup>INK4A</sup>) como blanco transcripcional, en acuerdo con la actividad supresora tumoral de SALL2. Adicionalmente, se han identificado otros blancos transcripcionales de SALL2 bajo estrés genotóxico. La mencionada proteína proapoptótica BAX (Gu et al. 2011) al igual que la proteína proapoptótica NOXA (Escobar et al. 2015) son blancos regulados positivamente por SALL2 bajo el tratamiento con quimioterapéuticos, mientras que el oncogén *c-myc* es reprimido por SALL2 bajo este contexto (Sung et al. 2012). En un estudio más reciente, se han identificado 104 blancos transcripcionales dependientes de *Sall2* en células embrionarias pluripotentes (ESC) en el fondo genético 129S6/SvEvTac (Yamamizu et al. 2016). En este estudio se aplicó la herramienta de genética ontológica (del inglés “Gene Ontology”) encontrándose que SALL2 regularía una cohorte de genes que codifican para canales de iones controlados por voltaje. Si bien estos estudios aportan de forma significativa en explicar las distintas funciones en las cuales *Sall2* estaría involucrado, la mayoría de sus blancos transcripcionales en otros tipos celulares o contextos no son conocidos y aún no se puede establecer una participación clara de este factor en el cáncer.



## II. HIPÓTESIS DE TRABAJO Y OBJETIVOS

**SALL2 regula positivamente una cohorte de genes asociados a un fenotipo canceroso.**

Para responder a esta hipótesis se propone como objetivo general:

### **Definir blancos transcripcionales de SALL2**

Con el fin de poder lograr el objetivo general hemos establecido dos objetivos específicos:

#### **Objetivo 1: Determinar respuestas transcripcionales dependientes de *Sall2***

1.1 Definir blancos transcripcionales específicos de SALL2 *in vitro* e *in vivo*.

1.2 Investigar el efecto de SALL2 en regiones promotoras de genes blanco.

#### **Objetivo 2: Seleccionar blanco(s) transcripcional(es) de SALL2 asociados a fenotipo canceroso y realizar ensayos funcionales de estos genes y su dependencia de SALL2.**

2.1 Analizar el efecto de pérdida/ganancia de la función de genes asociados con SALL2 provenientes del objetivo 1 y corroborar efectos celulares.

### III) MATERIALES Y MÉTODOS

#### MATERIALES:

##### Plásmidos:

pCMV-dR8.2 dvpr (Plásmido Addgene #8455, (Stewart et al. 2003))

pCMV-VSVG (Plásmido Addgene #8454, (Stewart et al. 2003))

pLKO.1-scramble (Plásmido Addgene #8453, (Sarbasov et al. 2005))

pLKO.1-shSall2 (Plásmido Addgene #8453, (Sarbasov et al. 2005))

pGL3-mAng-promoter (1231 pares de bases de promotor de angiogenina murina)

pmCherry-C1

pmCherry-C1-ANG (proteína madura)

pD1321-AP

pD1321-AP-SALL2

pD1321-AP-Jun\_SALL2

pD1321-AP-mSall2

pD1321-AP-mSall2\_down



#### **Secuencias clonadas en el vector pLKO.1 scramble (shCtrl) y para silenciar Sall2**

shCtrl:

5'-CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGG-3'

shSall2 (**loop** y **término**):

5'- CCGGAAGTCATGGATACAGAAGCACACT**CGAG**TGTGCTCTGTATCCATGAC

**TTTTTTT**G -3'

**Secuencias clonadas en el vector pD1321-AP para la generación de células KO mediante el sistema CRISPR/Cas9**

GGCTCCTTAGGCCAGACGGT (SALL2)

CACCCCAACTAGCGGTTAC (Jun\_SALL2)

GGTGAGCGAGGAATTCGGTC (mSall2)

TAGTCTAGGTGCTCCGGTAC (mSall2\_down)

**Partidores:**

**(1) Partidores utilizados para PCR cuantitativa en tiempo real (Ratón)**

*Ms4a6d\_F* 5'-TTCCATCCTGTTAGAATCTGGCT-3'

*Ms4a6d\_R* 5'-GTGAACCAAAGGCTTAGTCATCT-3'

*Mmp3\_F* 5'-ACATGGAGACTTTGTCCCTTTTG-3'

*Mmp3\_R* 5'-TTGGCTGAGTGGTAGAGTCCC-3'

*Rmrp\_F* 5'-GCTCTGAAGGCCTGTTTCCT-3'

*Rmrp\_R* 5'-GACTTGCGGGGAAGTCTAC-3'

*Anxa8\_F* 5'-GAGCTGCACGATGCCATGA-3'

*Anxa8\_R* 5'-CGGTTCCATAATCTTCTCACC-3'

*Rpph1\_F* 5'-GCTTGGAACAGACTCACGG-3'

*Rpph1\_R* 5'-CCATTTGTTCCAGGCACTG-3'

*Pnp\_F* 5'-TCGTCTGCTAAAGATGCTGG-3'

*Pnp\_R* 5'-TGAGAAACCAAAGACACGGAG-3'

*Cdkn1a\_F* 5'-CTTGCACTCTGGTGTCTGAG-3'

*Cdkn1a\_R* 5'-GCACTTCAGGGTTTTCTCTTG-3'

*Meox1\_F* 5'-GAAACCCCACTCAGAAGATAGC-3'

*Meox1\_R* 5'-TCGTTGAAGATTCGCTCAGTC-3'

*Rn45s\_F* 5'-CTCGAGAGACTCATTGCTTTCC-3'

*Rn45s\_R* 5'-GTTCTCTTCCAAGGGCATTCT-3'  
*Pop1\_F* 5'-GCCCAGTAACGTGACTCTGTC-3'  
*Pop1\_R* 5'-GGCTTCCCCTTTCCAAAATGAT-3'  
*Ang\_F* 5'-CCAGGCCCGTTGTTCTTGAT-3'  
*Ang\_R* 5'-GGAAGGGAGACTTGCTCATTTC-3'  
*Gatm\_F* 5'-GCTTCCTCCCGAAATTCCTGT-3'  
*Gatm\_R* 5'-CCTCTAAAGGGTCCCATTTCGT-3'  
*Nmral1\_F* 5'-GGCCGGTGCATTGCTAGAA-3'  
*Nmral1\_R* 5'-CTCGCACTACCTCAGCACC-3'  
*Isg15\_F* 5'-GGTGTCCGTGACTAACTCCAT-3'  
*Isg15\_R* 5'-TGGAAGGGTAAGACCGTCCT-3'  
*Gch1\_F* 5'-ACTTCACCAAGGGATACCAGG-3'  
*Gch1\_R* 5'-CTTGCTTGTTAGGAAGATAGCCA-3'  
*Cd36\_F* 5'-ATGGGCTGTGATCGGAACTG-3'  
*Cd36\_R* 5'-GTCTTCCAATAAGCATGTCTCC-3'  
*RNApol\_F* 5'-ACTGTGCGGAACTCCATCAA-3'  
*RNApol\_R* 5'-AGCCAGGTTCTGGAACCTCAA-3'

**(2) Partidores utilizados para PCR cuantitativa en tiempo real (Humano)**

*ANG\_F* 5'-CATCTTTGCGTTTTCTACCGG-3'  
*ANG\_R* 5'-CCCATCACCATCTCTTCCAAC-3'  
*ANXA8\_F* 5'-ATGAAGGGCTTAGGAACCAAG-3'  
*ANXA8\_R* 5'-ACTTGTGTCTGCTTGGATGTC-3'  
*PNP\_F* 5'-ACTACGGTCAAATCCCCAAC-3'  
*PNP\_R* 5'-TCAGGAACCCAAACACCAG-3'  
*RPPH1\_F* 5'-GCTTGGAACAGACTCACGG-3'  
*RPPH1\_R* 5'-TCTGAATTGGGTTATGAGGTCC-3'

### (3) Partidores utilizados para secuenciar mediante PCR

*Rpph1\_seq\_F* 5'-CAGCCATTGAACTCGCACTG-3'  
*Rpph1\_seq\_R* 5'-GGGGAGGTGAGTTCCCAGAG-3'  
*Del\_mS2\_F* (ratón) 5'-AGAGCCTCCGTTATCCAACA-3'  
*Del\_mS2\_R* (ratón) 5'-TGCTGAGCCACTGTCTGCTA-3'  
*Del\_S2\_F* (humano) 5'-GGGTTGTCTCCTGTCCACAC-3'  
*Del\_S2\_R* (humano) 5'-CATTTGTGCCTTCCACCAG-3'

### (4) Partidores utilizados para PCR cuantitativa en tiempo real en experimentos de inmunoprecipitación de cromatina (Humano)

*CAC\_F* 5'-CTCCGGCGTTCACAGTT-3'  
*CAC\_R* 5'-CGTGGATGCGTGTGTTTG-3'  
*NO\_RELACIONADO\_F* 5'-AAATTAACGAATGCCTGCTCTG-3'  
*NO\_RELACIONADO\_R* 5'-CCTGCTGACTCAAGACTTC-3'

#### **Anticuerpos y fluoróforos:**

<b>Anticuerpo</b>	<b>Dilución / uso</b>
<b>Anti-angiogenin</b> (ab10600, Abcam)	Dilución 1:500 / "Western Blot"
<b>Anti-p53</b> (PAb240, Abcam)	Dilución 1:500 / "Western Blot"
<b>Anti- p21</b> (sc-6246, Santa Cruz Biotechnology)	Dilución 1:500 / "Western Blot"
<b>Anti-β-Actin</b> (C4, Santa Cruz Biotechnology)	Dilución 1:10000 / "Western Blot"
<b>Anti-SALL2</b> (HPA004162, SIGMA)	Dilución 1:1000 / "Western Blot"
<b>Anti-SALL2</b> (A303-208A, Bethyl laboratories)	4 microgramos por punto de Inmunoprecipitación
<b>Anti-puromycin</b> (12D10, Millipore)	Dilución 1:2000 / "Western Blot" e

	Immunocitoquímica
<b>Anti-TIA-1</b> (C20, Santa Cruz Biotechnology)	Dilución 1:2000 / inmunocitoquímica
<b>Anti-EIF3H</b> (AB2, SIGMA)	Dilución 1:2000 / inmunocitoquímica
<b>Anti-mouse Alexa fluor 488</b> (Life Technologies)	Dilución 1:200 / inmunocitoquímica
<b>Anti-mouse Alexa fluor 546</b> (Life Technologies)	Dilución 1:200 / inmunocitoquímica
<b>Anti-rabbit Alexa fluor 488</b> (Life Technologies)	Dilución 1:200 / inmunocitoquímica
<b>Anti-rabbit Alexa fluor 546</b> (Life Technologies)	Dilución 1:200 / inmunocitoquímica
<b>Anti-rabbit Alexa fluor 647</b> , Jackson Immunoresearch	Dilución 1:200 / inmunocitoquímica
<b>IgG normal (mouse)</b> sc-2025, Santa Cruz.	4 µg por punto de inmunoprecipitación
<b>IgG normal (rabbit)</b> sc-2027, Santa Cruz.	4 µg por punto de inmunoprecipitación
<b>Hoechst</b> , Thermo Fisher Scientific. (1:500-1:1000)	Dilución 1:500-1:1000

## **MÉTODOS DE LABORATORIO**

### **Aislamiento de fibroblastos primarios de ratón (MEFs) y cultivo celular**

El modelo murino *Sal12* KO fue obtenido por intermedio de un acuerdo de transferencia de material desde el Dr. Ryuichi Nishinakamura (Kumamoto University, Kumamoto, Japan) a la Dra. Roxana Pincheira (Universidad de Concepción, Chile) en el año 2010. Los fibroblastos embrionarios de ratón se prepararon desde embriones decapitados 13.5 días *post coitum*. Los órganos rojos fueron retirados, la cabeza congelada a -80°C para posterior genotipificación y el cuerpo de los embriones fue disgregado en pedazos utilizando una cuchilla de afeitar estéril en una placa de cultivo de 10 cm con 5 ml de tripsina/EDTA al 0.25% (Hyclone). Luego de esto, los embriones disgregados fueron incubados en la tripsina durante 15 min a 37 ° C seguido de una dilución en 10 ml de DMEM (Hyclone; del inglés Dulbecco's Modified Eagle Medium con alta glucosa; al 10% FBS del inglés Fetal Bovine Serum (Gibco); 2mM L-glutamina (Hyclone); penicilina/estreptomicina (Gibco)) pipeteando hacia arriba y hacia abajo cada 10 minutos 3 veces. Las células posteriormente fueron centrifugadas y sembradas en placas de cultivo de 100 mm por embrión, utilizando 6 mL de DMEM precalentado a 37°C (pasaje 0). Posteriormente las MEFs fueron obtenidas en estas placas y cultivadas de forma rutinaria hasta el siguiente pasaje en una estufa de cultivo celular (Forma Scientific) a 37°C con un 5% de CO<sub>2</sub>. La obtención de las MEFs se consideró de embriones independientes, los cuales fueron previamente genotipados a partir del ADN genómico proveniente de las cabezas respectivas.

### **Congelamiento de fibroblastos embrionarios de ratón**

Se procedió a criopreservar los cultivos celulares primarios cercanos al 90% de confluencia de la placa en pasaje uno (p1). Para esto, se eliminó el medio cultivo de las placas de cultivo y las células se lavaron con 5 mL de PBS (PBS: del inglés Phosphate Buffered Saline; stock 10X: NaCl 137 mM; KCl 2.7 mM; Na<sub>2</sub>HPO<sub>4</sub> x 2 H<sub>2</sub>O

10 mM;  $\text{KH}_2\text{PO}_4$  2.0 mM; agua ultrapura; pH 7.4) seguido de la adición de tripsina/EDTA al 0.25% por 2-3 minutos a 37°C. Para inhibir el efecto de la tripsina, se adicionó DMEM completo y se centrifugaron los tubos a 1000 rpm por 5 minutos. El pellet celular se conservó y resuspendió adicionando el medio de congelamiento (90% FBS y 10% DMSO: Dimetilsulfóxido; Sigma). Luego, se adicionó 1 ml de medio de congelamiento por cada vial de criogenia (Nalgene) marcados con los datos del embrión, fecha, pasaje y genotipo. Los tubos se almacenaron a -80°C durante 2-3 días como máximo y fueron llevados a nitrógeno líquido (-196°C) para un almacenamiento a largo plazo.

### **Genotipificación**

La genotipificación se llevó a cabo mediante el aislamiento de ADN genómico desde pedazos de cola de los ratones o cabezas de los embriones. Ambos tejidos se incubaron a 55°C por 16 horas en 300  $\mu\text{L}$  de buffer que contenía 100mM de Tris, 5mM de EDTA, 0.2% de SDS, 200mM de NaCl y 60  $\mu\text{g}$  de proteinasa K. La PCR de genotipificación usó un microlitro de ADN por reacción utilizando los siguientes partidores: sentido 5'-CACATTTTCGTGGGCTACAAG-3', antisentido 5'-CTCAGAGCTGTTTTCTGGG-3' y Neo, 5'-GCGTTGGCTACCCGTGATAT-3' los cuales generaron productos de PCR de 188 pares de bases para los embriones silvestres (wild-type) y de 380 pares de bases para los embriones mutantes (KO), respectivamente. El programa de reacción en cadena de la polimerasa (PCR) utilizado fue el siguiente: 94°C por 2,5 minutos para la denaturación inicial; 32 ciclos de 94°C por 30 segundos de denaturación, 58°C por 30 segundos de alineamiento y 72°C por 30 segundos de extensión. Finalmente se efectuó una extensión final de 72°C por 7 minutos. La PCR se efectuó con la enzima TaKaRa ex Taq (Clontech) en un termociclador de PCR convencional (Applied Biosystems).



## **Cultivo celular**

MEFs primarias e inmortalizadas tanto silvestres como mutantes para *SalI2* se cultivaron en medio de cultivo DMEM suplementado con 10% de suero fetal bovino inactivado por calor (FBS, GE Healthcare HyClone), 1% de glutamina (Invitrogen) y 0,5% de penicilina/estreptomicina (Invitrogen). Los experimentos con MEFs primarias se realizaron en pasajes tempranos (pasajes 3-4). Las MEFs inmortalizadas silvestres y mutantes para *SalI2* se obtuvieron utilizando el antígeno T grande proveniente del virus de simio SV40 basado en el protocolo modificado de Zhu et al. (Zhu et al. 1991). Para la transfección de MEF primarias, se utilizó Lipofectamina 2000 (Invitrogen) y 2 µg de un vector de expresión del antígeno T grande SV40 (Addgene Plasmid # 9053, (Zalvide and DeCaprio 1995)). El medio de cultivo fue reemplazado por medio DMEM sin suero conteniendo la Lipofectamina 2000 con los 2 µg de plásmido. Después de la transfección, las células fueron incubadas por 5-6 horas a 37°C, para luego reemplazar el medio por DMEM con suero. Posterior a esto, se procedió a seleccionar MEFs que crecieran en baja densidad. Para completar el proceso de inmortalización, se llevaron a cabo 5-6 pases posteriores a la transfección, sembrando en baja densidad por vez. Adicionalmente, se cultivaron células HEK293 epiteliales de riñón embrionario humano (American Type Culture Collection CRL-1573™) en DMEM suplementado con 10% de FBS, 1% de glutamina, y 0,5% de penicilina/estreptomicina y células de leucemia mieloide crónica K562 (American Type Culture Collection CCL-243™) en RPMI suplementado de la misma forma que el medio DMEM.

## **Estrés genotóxico o deprivación de suero**

Para ejercer estrés genotóxico, células MEF fueron tratadas con la droga doxorubicina a una concentración 1 µM por 9 o 16 horas. En el caso de células HEK293, se aumentó la concentración de la droga al doble. La deprivación de suero se efectuó con la adición de DMEM alta glucosa, 2mM L-glutamina y penicilina/estreptomicina a excepción del suero bovino fetal por 9 o 16 horas. En el

caso de la medición de la tasa de traducción celular (ensayo de SunSET), las células fueron incubadas con la droga puromicina (10 µg/mL) por 10 minutos antes de finalizar los tratamientos y ser colectadas. Tanto la droga como la deprivación de suero se efectuó en cultivos asincrónicos.

### **Inmunocitoquímica**

Las células sujetas a inmunocitoquímica fueron crecidas sobre cubreobjetos y posteriormente fijadas en 4% formaldehído/PBS por 30 minutos a temperatura ambiente (18°C). Después, se realizaron tres lavados con PBS por 5 minutos, para luego permeabilizar las células con dos lavados de 0,2% TRITON X 100/PBS por 5 minutos cada uno a 4°C. Posteriormente se realizaron tres lavados con PBS por 5 minutos seguido de la incubación de las células con el anticuerpo primario diluido en BSA al 1% toda la noche a 4°C. Al día siguiente, se realizaron tres lavados con PBS por 5 minutos y se incubó 1 hora con el anticuerpo secundario correspondiente a la especie del anticuerpo primario, incluyendo la tinción de “Hoechst” para teñir los núcleos. Finalmente, se montaron los cubreobjetos en “Fluorescence Mounting Medium” (Dako) para su posterior análisis bajo microscopía confocal. El microscopio confocal empleado fue un microscopio espectral confocal LSM780 NLO Zeiss (objetivos 40X y 60X, CMA Bío-Bío). La intensidad de fluorescencia de los gránulos de estrés fue cuantificada con la herramienta “Threshold” y luego con análisis de partículas de ImageJ (NIH, USA), graficando posteriormente la relación entre las condiciones de experimentación y los controles. Se evaluó la significancia estadística utilizando la prueba T de Student y los gráficos fueron procesados en R (<https://www.r-project.org/>) o Prism 6.

### **Detección de proteínas**

Se realizó inmunodetección de proteínas a través de la técnica de “Western Blot”. Se realizó lisis celular en 50 µL de “buffer” de lisis por condición (0.5% Igepal

CA030, 10mM Hepes pH7.9, 1mM DTT, 100mM NaCl, 0.5mM PMSF) en presencia de inhibidores de proteasas y fosfatasas (PPC1010-1ML, Sigma Aldrich). Se igualó concentraciones de las proteínas lisadas mediante la técnica de Bradford y se separaron las proteínas a través de electroforesis en gel denaturante al 15, 10, 8 y 6% de acrilamida (SDS-PAGE) dependiendo del tamaño de las proteínas en estudio. Se utilizaron geles de 1.5 mm para la detección de proteínas. Las muestras se incubaron por 5 minutos a 95°C con un buffer de carga para proteínas (62.5mM Tris-HCl pH6.8, 2%SDS, 10% glicerol, 0.01% azul de bromofenol) en presencia de 0.1M de DTT. Luego las muestras se sembraron junto al estándar pre-teñido (PageRuler Prestained Protein Ladder, 10 to 180 kDa Thermo) en el gel, para posteriormente correrlas a 80V constantes en solución de corrida (25mM Tris, 250mM glicina, 0.1% SDS). Se cargaron entre 25-100 µg de proteínas totales para la detección de proteínas específicas, las que posteriormente fueron transferidas desde el gel a una membrana de nitrocelulosa en solución de transferencia (25mM Tris, 192mM glicina, 20% metanol) durante 1 hora y 20 minutos a 100 volts o toda la noche a 30 volts si la transferencia involucraba proteínas de alto peso molecular. Para verificar la correcta transferencia, se tñó la membrana de nitrocelulosa con rojo "Ponceau" (P3504-10G, Sigma Aldrich). Luego de esto, se realizaron múltiples lavados con agua destilada, para luego bloquearla con leche descremada al 5% en TBS-Tween (150mM NaCl, 10mM Tris, 0.05%Tween20) durante 1 hora. Después se incubó con el anticuerpo primario y el secundario por 1 hora a temperatura ambiente o toda la noche a 4°C cada uno, realizando tres lavados con TBS-Tween por 10 minutos después de cada incubación de anticuerpo. Finalmente, la membrana se reveló con líquidos quimioluminocentes ("Western Lightning Plus-ECL", Perkin Elmer) y se expuso a una película autoradiográfica (Ultracruz Autoradiography Film Blue, Santa Cruz Biotechnology). Se realizaron análisis densitométricos de los films revelados utilizando el programa ImageJ 1.49q (NIH, USA).

## **Construcciones de ADN y digestión preparativa**

Se realizó una construcción de traspaso, para subclonar 1231 pares de bases del promotor murino de Angiogenina/Rnase4 al vector pGL3-basic, reportero de luciferasa. Para ello, se emplearon las enzimas de restricción XhoI y SmaI, traspasando el promotor (inserto) desde el vector pUC19 al vector pGL3-basic, y así generar a pGL3-basic/mAng-promoter. Se realizaron digestiones preparativas para purificar al vector y al inserto por separado, empleando 10 µg de ADN y 20 unidades de cada enzima de restricción, utilizando al "buffer" de compatibilidad según las especificaciones de la empresa (New England Biolabs) en 15 µL totales de reacción. La incubación de la reacción se llevó a cabo a 37°C por 8 horas, y luego, se purificaron los fragmentos desde gel de agarosa al 1% en 0.5X TAE (40mM Tris-HCl, 1mM EDTA, 1.14mL/L ácido acético glacial, pH 7.8) en presencia de bromuro de etidio con el kit NucleoSpin Gel and PCR Clean-up (MACHEREYNAGEL). Se cuantificó por gel, analizando la banda generada a través de densitometría empleando el programa ImageJ 1.49q (NIH, USA), para luego realizar la ligación. Para la ligación, se utilizó una relación molar 3:1 (inserto: vector) y al vector sólo como control, empleando la enzima T4 DNA Ligase (Thermo Scientific) según las recomendaciones de la empresa, incubando a 16°C toda la noche. Posteriormente cada mezcla de ligación se transformó en bacterias competentes XL-1 Blue (Catalog #200249, Agilent Technologies) usando el método de transformación por calcio.

## **Transformación de bacterias competentes por calcio**

Para efectuar la transformación bacteriana, se incubó la reacción de ligación con 50 µL de bacterias competentes por 30 minutos en hielo. Posterior a esto, se indujo el shock térmico a 42°C por 2.5 minutos y se incubó la mezcla por 5 minutos en hielo. A continuación, las células se recuperaron por una hora a 37°C en medio LB (Medio Luria-Bertani, MO BIO Laboratories, Inc.) sin antibiótico, para posteriormente sembrar las bacterias en placas agar LB/Ampicilina 100 µg/ml; la cuales se incubaron

toda la noche a 37°C. Al día siguiente, las colonias presentes en la placa se inocularon en medio LB con ampicilina (100 µg/ml) para su posterior crecimiento.

### **Minipreparación de ADN**

Se analizaron al menos 10 colonias por separado para verificar la presencia del fragmento clonado en la nueva construcción. Se crecieron las colonias en 3 ml de medio LB con ampicilina por toda la noche a 37°C. Posteriormente, 1.5 ml de bacterias fueron centrifugadas por 10 minutos a 14.000 rpm. Luego se descartó el sobrenadante, para resuspender el pellet con 300 µL de “buffer” TENS (10mM Tris-HCL pH 8.0, 1mM EDTA pH 8.0, 0.1N NaOH, 0.5% SDS) por vórtex. La mezcla anterior se incubó en hielo por 10 minutos y se le agregó 15 µL de Acetato de sodio 3M pH 5.2, agitando nuevamente en vórtex. A continuación, se realizó una centrifugación por 15 minutos y se transfirió el sobrenadante a un nuevo tubo donde se le agregó 800 µL de etanol al 100% frío para obtener la precipitación del ADN. Posterior a esto se realizó una centrifugación por 5 minutos y se descartó el sobrenadante. El pellet de ADN se lavó con 800 µL de etanol al 70% y se realizó otra centrifugación por 10 minutos. Después, el pellet se resuspendió en agua desionizada conteniendo RNasa a una concentración de 20µg/ml. Finalmente, para verificar si los clones elegidos correspondían a la construcción esperada, se realizó una digestión analítica (1 unidad de enzima en 15 µL finales) empleando enzimas de restricción (New England Biolabs) según descripción de la empresa. Se verificó la liberación de fragmentos mediante un gel de agarosa al 1% en 0.5X de tampón TAE. Una vez detectado el clon que contenía la construcción con el inserto, se generó un vial de “glicerol stock” correspondiente (850 µL de caldo inoculado con 150 µL de glicerol puro) para su almacenamiento a largo plazo a -80°C. La colonia se creció en 500 mL de caldo de cultivo para purificar ADN a gran escala.

## **Purificación de ADN a gran escala**

Para generar ADN plasmidial a gran escala el clon seleccionado se creció en 500 ml de LB/ampicilina toda la noche a 37°C. Luego, se utilizó el kit NucleoBond® Xtra Midi/Maxi (MACHEREYNAGEL) siguiendo las instrucciones del manual. Se centrifugó el cultivo bacteriano por 15 minutos a 6000 rpm y se descartó el sobrenadante. Se resuspendió el pellet en 16 mL de “resuspension buffer” y a continuación se agregó 10 mL de “lysis buffer” mezclando suavemente por inversión e incubando posteriormente por 5 minutos a temperatura ambiente. Seguidamente, se equilibró la columna de purificación con 12 mL de “equilibration buffer” y se agregó 12 mL de “neutralization buffer”, mezclando con suavidad hasta que la mezcla se tornó clara y el ADN genómico se separó como un precipitado blanco. Se centrifugó el lisado a 10.000 rpm por 10 min a 4°C y el sobrenadante obtenido de la centrifugación se agregó a la columna de purificación. El lisado filtrado se descartó y se lavó la columna con 15 mL de “equilibration buffer”. El filtro de la columna de purificación se removió y se procedió a lavar la columna nuevamente con 25 mL de “wash buffer”. El ADN plasmidial se eluyó desde la columna con 15 mL de “elution buffer” y se agregó 10.5 mL de isopropanol, mezclando e incubando por 10 minutos a temperatura ambiente. Se recuperó el pellet centrifugando a 10.000 rpm por 30 minutos a 4°C y se descartó el sobrenadante, para lavar el pellet con etanol al 70% y posterior secado a temperatura ambiente con el objeto de remover el etanol. Finalmente, se disolvió el pellet en 300 µL de “buffer” TE (100mM Tris Cl pH 8.0, 1mM EDTA) y se determinó la concentración de ADN a través de la absorbancia obtenida mediante el uso de un NanoDrop (Thermo Scientific NanoDrop 2000/2000c). Se consideraron las absorbancias a 260 nm (ácidos nucleicos) y 280 nm (proteínas) y se obtuvo la razón 260/280 que idealmente alcanza una razón de 1.8-2.0 para muestras de calidad óptima.

## **Electroforesis en geles de agarosa**

La separación de fragmentos de ADN obtenidos fue realizada mediante geles de agarosa al 1%. El tampón empleado para la electroforesis fue TAE (Tris-Ácido Acético EDTA; Tris-HCl 40 mM, ácido acético 30 mM y EDTA 1 mM; pH 7,6). Los geles de agarosa se prepararon con TAE conteniendo bromuro de etidio 0,5 µg/mL (APEX). El marcador de peso molecular utilizado fue de 100 pb DNA ladder (GeneRuler, Thermo Scientific). Para visualizar las diferentes bandas de ADN en el gel de agarosa, el gel fue ubicado en un equipo transiluminador y fotodocumentador (VilberLourmat).

## **Transfección transiente en células HEK293**

Se emplearon células HEK293 cultivadas a un 50% de confluencia en placas de 10 cm con medio HyClone DMEM (GE Lifesciences) suplementado con 10% de FBS, 1% de glutamina, y 0,5% de penicilina/estreptomicina. Posteriormente, el medio de cultivo fue reemplazado por medio DMEM sin suero conteniendo Lipofectamina 2000 (Invitrogen) y 10 µg de los plásmidos de expresión. Las células fueron incubadas por 5-6 horas a 37°C con el medio de transfección y luego se reemplazó por DMEM con suero por 18 horas adicionales.

## **Transducción de shRNA para SALL2 en células MEF inmortales (iMEFs)**

Para efectuar la transducción de los shRNA control y dirigido contra SALL2 en células iMEFs, las partículas lentivirales se empacaron en células HEK293 mediante la co-transfección de las siguientes construcciones: pCMV-dR8.2 dvpr, pCMV-VSVG y pLKO.1 scramble o shSall2 (ver sección plásmidos). El medio se cambió cada 24 horas en un lapso total de 72 horas y fue filtrado usando un filtro de 0,45 µm para posteriormente añadirse a células blanco iMEF de tipo silvestre. Todos los sobrenadantes se añadieron a las células blanco en conjunto con 9 µg/ml de polibreno como adyuvante de la transducción. Posterior a las 72 horas, las células iMEF se

seleccionaron con 5 µg/ml de puromicina y se recuperaron adicionalmente con medio DMEM nuevo.

### **Electroporación de células**

Se procedió a transfectar fibroblastos embrionarios de ratón inmortalizados (iMEFs) y/o K562 (células humanas de leucemia mieloide crónica, ATCC® CCL-243™) de acuerdo con el protocolo de electroporación de Neon® Transfection System (Thermo Fisher Scientific). 1.5 millones de células MEF o K562 fueron resuspendidas en 90 µL de buffer de resuspensión, volumen el cual se completó con 20 µg de ADN plasmidial a un volumen final de 115 µL. La mezcla de electroporación se colocó en un tubo Eppendorf de 2 mL y fue colectado cuidadosamente con la pipeta de electroporación incluida por el fabricante. La pipeta fue posicionada de forma vertical en la cubeta de electroporación previamente rellena con 3 mL de buffer E2 (denominada Neon® Pipette Station). Para electroporar las células iMEF, se programó el monitor de electroporación para un pulso de 1150 voltios con duración de 35 milisegundos y para las células K562, un pulso de 1200 voltios por 35 milisegundos. Después de electroporadas, las células fueron depositadas en una placa de cultivo de 10 cm en el caso de la generación de KO por el sistema de CRISPR/Cas9 o para la sobreexpresión de proteínas. En el caso de los ensayos de luciferasa, las células fueron depositadas homogéneamente en placas de 6 pocillos.

### **Generación de silenciamiento y KO mediante el sistema CRISPR/Cas9**

Se electroporaron células iMEF como se indicó previamente con 15 µg de los vectores pD1321-AP-mSall2 y pD1321-AP-mSall2\_down, diseñados para generar una deleción genómica en el gen *Sall2* murino mediante la expresión de la proteína Cas9 silvestre acoplado a la proteína RFP (acrónimo del inglés “Red Fluorescent Protein”). Se incluyó como control de la transfección el vector vacío pD1321, sin guías para la proteína Cas9. Tanto la proteína Cas9 como RFP se pueden producir de manera



eficiente desde un único constructo, ya que están separadas por un péptido basado en el péptido 2A autoescindible, permitiendo un salto translacional (Provost et al. 2007). Las células electroporadas fueron disgregadas 18 horas post-electroporación mediante tripsinización y sometidas a citometría de flujo acoplado a sorteo celular (BDFACSAria III cell sorting). En cada experimento se seleccionó un pool de 100 células correspondientes al 1% de las células más fluorescentes. Los clones crecieron en placas hasta alcanzar la confluencia, se congelaron viales de 500000 células por vial y se confirmó la delección genómica por PCR convencional previo a la experimentación. Los partidores utilizados para detectar la delección genómica en el gen *Sall2* murino fueron los siguientes: Del\_mS2\_F y Del\_mS2\_R (véase sección partidores) y el protocolo de PCR convencional fue el mismo utilizado para la genotipificación. La generación de células K562 KO para *SALL2* fue la misma, a excepción de los plásmidos utilizados para generar la delección (pD1321-AP-SALL2 y pD1321-AP-Jun\_SALL2) y de los partidores (Del\_S2\_F y Del\_S2\_R) utilizados para detectar la delección genómica.

#### **Purificación de ARN total con reactivo TRIzol™**

El ARN total desde las células en cultivo y/o tejidos provenientes de los ratones fue extraído mediante el uso de TRIzol™ (catálogo 15596026, Thermo Fisher Scientific). Las células y/o tejidos se resuspendieron en 500 µL de solución TRIzol™ y se colectaron en un tubo Eppendorf de 1.5 mL. Posteriormente, las mezclas fueron homogenizadas por vortex y se les agregó 300 µL de cloroformo seguido de agitación fuerte. Se incubaron las muestras a temperatura ambiente por 5 min y se centrifugaron a máxima velocidad por 5 minutos a 4°C. La fase superior de 200-400 µL que contenía el ARN se colectó en un tubo nuevo y las fases inferiores se descartaron (conteniendo membranas, el ADN y las proteínas). Se agregó 300 µL de alcohol isopropílico a la fase superior, mezclando los tubos varias veces por inversión y posteriormente éstos se incubaron por 10 minutos a temperatura ambiente hasta precipitar el ARN.

Posteriormente se centrifugó a máxima potencia por 15 min a 4°C hasta obtener un pellet de ARN. El sobrenadante se descartó y se lavó el pellet con 500 µL de etanol con agua tratada con dietilpírocarbonato al 0,1% (DEPC 75% v/v), seguido de centrifugación a 7500 rpm por 5 min a temperatura ambiente. Se eliminó el sobrenadante y el pellet de ARN se secó a temperatura ambiente por 10 minutos agitando el tubo de forma intermitente. El pellet fue resuspendido mediante pipeteo suave y repetitivo en 30 µL de agua libre de nucleasas hasta disolverse. Finalmente se midió absorbancia en un NanoDrop para determinar concentración del ARN resuspendido a 260 nm. Adicionalmente se determinó la integridad del ARN por medio de un gel de agarosa no denaturante al 1 % preteñido con bromuro de etidio. Para eliminar ADN genómico residual del ARN aislado, se trató las resuspensiones con una enzima desoxirribonucleasa (DNasa) según el protocolo del kit comercial TURBO DNA-free (Invitrogen).

#### **Preparación de muestras para ARN-seq y purificación de ARN mediante kit**

Se seleccionó un embrión silvestre, un heterocigoto y tres embriones *Sall2* KO, de los cuales se obtuvieron células MEF y se expandieron a 4 placas por embrión. Dos placas fueron seleccionadas como control y las otras dos fueron tratadas con doxorubicina 1 µM por 16 horas. Con el objeto de obtener ARN de calidad para secuenciación masiva de nueva generación (ARN-seq) se utilizó el kit RNeasy Plus Mini Kit (Qiagen) para ARN total, siguiendo las instrucciones del fabricante. Cada placa de células MEF silvestres o *Sall2* KO tratadas o no con doxorubicina, fueron colectadas con 300 µL de buffer RLT con la adición previa de 10 µL de β-mercaptoetanol (β-ME) por 1 ml de Buffer RLT. Las réplicas técnicas se mezclaron para obtener un total de 600 µL y dicha mezcla se posicionó en una columna de eliminación de ADN genómico (“gDNA eliminator”) para descartar el ADN genómico del ARN total. El volumen centrifugado se recuperó, descartando la columna. A este volumen se le adicionó 1 volumen de etanol al 70% v/v, se homogenizó y se transfirió

a una columna “RNeasy spin column”, centrifugando por 15 segundos a 10000 rpm. El volumen centrifugado se descartó y se agregó 700  $\mu\text{L}$  de buffer RW1, centrifugando nuevamente como en el paso anterior. Este paso se repitió 2 veces más con 500  $\mu\text{L}$  de buffer RPE y finalmente el ARN se eluyó con 30  $\mu\text{L}$  de agua grado biología molecular.

### **Transcripción inversa acoplada a PCR**

La transcripción inversa del ARN aislado se realizó con el kit comercial “M-MLV Reverse Transcriptase” (Promega). Se preparó la primera mezcla con 2  $\mu\text{g}$  de ARN total; 1  $\mu\text{L}$  de una dilución 1:10 de oligo dT y agua grado biología molecular hasta completar 15  $\mu\text{L}$  de volumen final. El ARN se desnaturalizó durante 5 minutos a 70°C y se enfrió en hielo por 5 minutos. La segunda mezcla se preparó con agua libre de nucleasas, tampón 5X  $\text{MgCl}_2$ , dNTPs y se adicionó la enzima transcriptasa reversa M-MLV, hasta completar 10  $\mu\text{L}$  de volumen final. A esta segunda mezcla se le agregó la primera que contenía el ARN. Una vez unidas las dos mezclas, se realizaron los ciclos de temperatura en el termociclador: 42°C por 60 minutos (extensión) y 70°C por 15 minutos (inactivación de la transcriptasa inversa). Los ADN complementarios (ADNc) obtenidos desde la retrotranscripción se guardaron a -20°C hasta su utilización. El análisis de la expresión de genes se realizó mediante PCR cuantitativa en tiempo real (qPCR) a partir de los ADNc utilizando oligonucleótidos específicos (véase partidores para PCR en tiempo real).

### **PCR cuantitativa en tiempo real (qPCR)**

Se realizaron las reacciones de PCR cuantitativa en tiempo real (qPCR) por triplicado utilizando “KAPA SYBR FAST qPCR Master Mix (2X) Kit” (Kapa Biosciences) con una concentración de primer de 0,4  $\mu\text{M}$  (véase partidores para PCR en tiempo real). Las condiciones de ciclado fueron las siguientes: desnaturalización inicial a 95 °C por 3 min, luego 40 ciclos de 95 °C por 5 segundos (desnaturalización) y 60 °C por 20 segundos (extensión). Para controlar la especificidad del producto amplificado, se llevó

a cabo un análisis de curva de fusión. No se observó amplificación de producto inespecífico con los partidores utilizados. La expresión de cada gen fue relativa a la expresión del gen ARN Pol II y se representó como veces de cambio en comparación con el control, para cada caso. Los partidores utilizados se seleccionaron desde la base de datos PrimerBank (<https://pga.mgh.harvard.edu/primerbank/>) que compila partidores previamente testeados de forma experimental.

### **Inmunoprecipitación de cromatina**

Células K562 en cultivo fueron centrifugadas y tratadas con formaldehído al 1% por 10 minutos a temperatura ambiente, con el objeto de entrecruzar el ADN con las proteínas. La reacción de entrecruzamiento se detuvo con 0,55 ml de glicina 2,5 M a los 10 minutos de reacción. Posteriormente, las células fueron lavadas dos veces con PBS 1X frío mediante resuspensión y centrifugación. El pellet celular se lisó en un buffer de lisis que contenía 0.5 M HEPES pH 7.8, 1 M MgCl<sub>2</sub>, 1 M KCl y 10% de NP-40 en presencia de inhibidores de proteasas y la lisis se incubó en hielo durante 20 minutos. Luego, los lisados fueron homogenizados en un homogenizador "Dounce" (20 veces, utilizando un bastón apretado) y después de 5 minutos de centrifugación a 1000 r.p.m., los núcleos celulares contenidos en el pellet fueron sometidos a sonicación. Los núcleos fueron resuspendidos en 300 µL de buffer de sonicación (50 mM Hepes pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na deoxycholate y 0.1 % SDS) y sonicados en hielo con el objeto de obtener fragmentos de cromatina entre 200 y 1000 pb, utilizando el sonicador Vibracell Sonics VCX 130. El programa de sonicación empleado fue el siguiente: 4,5 minutos de sonicación intercalando 15 segundos en ON y 20 segundos en OFF a 100 volts de potencia. Las inmunoprecipitaciones se llevaron a cabo durante la noche a 4 °C en una mezcla que contenía 4 µg de anticuerpo anti-SALL2 (Bethyl laboratories) o control IgG de conejo (Santa Cruz) más 40 µg de cromatina (cuantificada por absorciometría a 260 nm) en rotación constante. Posterior a esto, a cada mezcla se le agregó 40 µL de proteína A/G

agarosa y se incubó durante 2 horas a 4 °C en rotación constante. Después de esto, la agarosa se centrifugó durante 5 minutos a 1000 r.p.m. y se lavó una vez en 500 µL de buffer de sonicación, dos veces en 500 µL de buffer de lavado de inmunoprecipitación (500 mM LiCl, 100 mM Tris-Cl pH 8.0, 1% NP-40 y 0.1% Na deoxycholate) y una vez en 500 µL de buffer TE pH=8.0. Los complejos ADN-proteína se eluyeron de la agarosa en 100 µL de un buffer que contenía 1% de SDS y 0.1% NaHCO<sub>3</sub> incubándose durante 65 °C por 15 minutos. Los entrecruzamientos entre el ADN y la proteína se revirtieron por la adición de 4.1 µL de NaCl 5 M y posterior incubación a 65 °C toda la noche. La proteinasa K se añadió durante 4 horas a 42 °C y el ADN se recuperó mediante una extracción con fenol/cloroformo y posterior precipitación con etanol incluyendo un vehículo adecuado (análogo de glicógeno). El ADN proveniente de la cromatina de los “input” se diluyó en 100 µL de buffer TE pH=8.0 y el ADN proveniente de las inmunoprecipitaciones en 30 µL de buffer TE pH=8.0. Se analizó la unión de SALL2 a las regiones promotoras del gen *ANG/RNASE4* por PCR cuantitativa en tiempo real utilizando los partidores detallados la sección “partidores”. Las condiciones de ciclado para todas las reacciones de PCR fueron las detalladas en la sección “PCR cuantitativa en tiempo real”.

## **MÉTODOS COMPUTACIONALES:**

### **Secuenciación de muestras para ARN-seq**

La secuenciación de ARN se efectuó en el servicio de secuenciación de la Universidad de Cambridge (Reino Unido). Las muestras de ARN fueron enviadas a la universidad y allí se efectuó la construcción de las librerías de secuenciación por muestra, empleando el kit NextSeq 500/550 v2, illumina diseñado para el secuenciador NextSeq500. El ARN proveniente de células MEF silvestre, heterocigota y *Sall2* KO tratadas o no con doxorubicina 1 µM por 16 horas fue obtenido mediante el kit de

purificación de ARN (descrito en métodos de laboratorio) y se alicuotaron 5 µg de ARN total por condición para su secuenciación. La secuenciación fue efectuada en la máquina NextSeq500 (illumina), con lecturas de 150 pares de bases de largo, de tipo “single end”.

### **Estrategia de alineamiento de las lecturas de ARN-seq al genoma de referencia**

Los archivos de secuenciación fueron ubicados en el clúster de computación de la universidad de Penn State (Penn State, Pennsylvania, USA) y analizados mediante programación en el sistema SLURM (acrónimo del inglés “**S**imple **L**inux **U**tility for **R**esource **M**anagement”). El sistema permitió planificar tareas de programas de código abierto mediante el acceso controlado de los recursos (nodos) del clúster, gestionando una cola de tareas pendientes. La secuenciación en la máquina NextSeq500 generó cuatro archivos de tipo FASTQ (formato que combina la secuencia de las lecturas junto con la calidad obtenida) los cuales fueron unidos mediante el comando “merge” en UNIX. Los archivos de salida del comando fueron alineados al genoma de referencia de *Mus Musculus* (mm10, ensamblaje GRCm38 de diciembre 2011, <https://genome.ucsc.edu/cgi-bin/hgGateway>) mediante el programa TopHat v1.4.0 en la configuración por defecto. El código utilizado en SLURM para estos efectos fue el siguiente:

```

## Script para alinear secuencias en formato FASTQ hacia archivos BAM ##
# Llamado de recursos #
#!/bin/bash
#SBATCH -C new
#SBATCH --exclusive
#SBATCH --ntasks=5
#SBATCH --cpus-per-task=20
#SBATCH -J carlos.maSeq
#SBATCH -t 14-00:00:00
#SBATCH --mail-type=ALL
#SBATCH --mail-user=cfarkas@udec.cl
## Definición de Variables y directorios ##

reference="/galaxy/data/mm10/bowtie_index/mm10"
carlos5="/nfs/brubeck.bx.psu.edu/scratch5/carlos"
readsDir="/nfs/brubeck.bx.psu.edu/scratch4/cfarkas/FASTQ"

inputName=${1}
outputFolderName=`echo ${1}|cut -d "/" -f 1`

reads=${readsDir}/${inputName}
outDir=${carlos5}/${outputFolderName}
mkdir -p ${outDir}

tophat -o ${outDir} -p 20 ${reference} ${reads}
#####

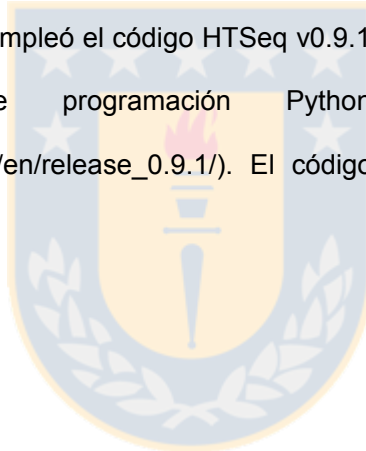
## Ejecución ##
#En BASH, ejemplo para KO-1-Control.:
sbatch tophat-carlos1.sh KO-1-Control/ KO-1-Control.fastq.gz

```

Los archivos de salida para este programa correspondieron a las lecturas alineadas al genoma de referencia en formato BAM (\*.bam, acrónimo del inglés **B**inary **A**lignment **F**ile) que es la versión binaria del formato SAM diseñado para almacenar el alineamiento de lecturas de secuenciación a un genoma de referencia determinado, incluyendo coordenadas genómicas.

### **Ordenamiento de los archivos en formato BAM y cuantificación de transcritos**

Los archivos BAM fueron ordenados en base a las coordenadas cromosómicas del genoma mm10 mediante el script SortSam.jar proveniente desde la colección de programas de la herramienta Picard tools (<https://broadinstitute.github.io/picard/>). Para cuantificar el número de lecturas alineadas en base a coordenadas genómicas desde los archivos BAM/SAM se empleó el código HTSeq v0.9.1 (modo unión) implementado en el lenguaje de programación Python (<https://www.python.org>, [https://htseq.readthedocs.io/en/release\\_0.9.1/](https://htseq.readthedocs.io/en/release_0.9.1/)). El código utilizado en SLURM para este efecto fue el siguiente:





```

## Indexado de archivos BAM por coordenadas genómicas y cuantificación de transcritos
##

## Llamado de recursos ##

#!/bin/bash

#SBATCH -C new

#SBATCH --exclusive

#SBATCH --ntasks=6

#SBATCH --cpus-per-task=10

#SBATCH -J carlos.htseq

#SBATCH --mail-type=ALL

#SBATCH --mail-user=cfarkas@udec.cl

## Definición de variables y directorios ##

BAM=${1}

GFF="/nfs/brubeck.bx.psu.edu/scratch5/carlos/igenomes/Mus_musculus/UCSC/mm10/An
notation/Genes/genes.gtf"

HTCOUNT="/nfs/brubeck.bx.psu.edu/scratch4/cfarkas/HTSeq-0.6.1/build/scripts-
2.7/htseq-count"

java -jar /galaxy/home/boris/boris4/software/bin/picardTools/SortSam.jar ${picard}
l=${BAM} O=sorted.${BAM} SO=coordinate 2>/dev/Null

exitStatus=$?

if [[ ${exitStatus} -eq 0 && -s sorted.${BAM} ]]; then

    ${HTCOUNT} -f bam -r pos -s no -i gene_id -q sorted.${BAM} ${GFF}
>sorted.${BAM}.htseq.report-3 2>>errors-3

else

    "processing ${BAM} failed"

    exit 2

fi

#####

## Ejecución de este script en BASH (por ejemplo, para KO-1-Control) ##

sbatch htseq-carlos.shKO-1-Control-merged.bam

```

### **Análisis de la expresión diferencial de genes**

Los archivos de salida del programa anterior fueron utilizados como archivos de entrada para confeccionar una tabla de conteo de lecturas alineadas por gen (en inglés, “count table”). Se utilizó un script escrito en el lenguaje de programación R (<https://www.r-project.org>) con el fin de construir esta tabla. El código utilizado en R para confeccionar la tabla mencionada se detalla a continuación:



```
## Definición de directorio y lectura de los archivos de salida de HTSEQ ##
setwd("C:/Users/Carlos/Desktop/RNA seq/HTSeq merge-bams/HTSeq output")
data1<-read.table("WT-Control.txt", header=F)
data2<-read.table("Het-Control.txt", header=F)
data3<-read.table("KO-1-Control.txt", header=F)
data4<-read.table("KO-2-Control.txt", header=F)
data5<-read.table("KO-3-Control.txt", header=F)
data6<-read.table("WT-Doxo.txt", header=F)
data7<-read.table("Het-Doxo.txt", header=F)
data8<-read.table("KO-1-Doxo.txt", header=F)
data9<-read.table("KO-2-Doxo.txt", header=F)
data10<-read.table("KO-3-Doxo.txt", header=F)
sum(data1$V2)
sum(data2$V2)
sum(data3$V2)
sum(data4$V2)
sum(data5$V2)
sum(data6$V2)
sum(data7$V2)
sum(data8$V2)
sum(data9$V2)
sum(data10$V2)
table(duplicated(data1$V1))["FALSE"]
table(duplicated(data2$V1))["FALSE"]
table(duplicated(data3$V1))["FALSE"]
table(duplicated(data4$V1))["FALSE"]
table(duplicated(data5$V1))["FALSE"]
table(duplicated(data6$V1))["FALSE"]
table(duplicated(data7$V1))["FALSE"]
table(duplicated(data8$V1))["FALSE"]
table(duplicated(data9$V1))["FALSE"]
table(duplicated(data10$V1))["FALSE"]
```

```

#Renombrando las columnas#
library(plyr)
a<-rename(data1, c("V1"="gene_name", "V2"="WT"))
b<-rename(data2, c("V1"="gene", "V2"="Het"))
c<-rename(data3, c("V1"="gene", "V2"="KO1"))
d<-rename(data4, c("V1"="gene", "V2"="KO2"))
e<-rename(data5, c("V1"="gene", "V2"="KO3"))
f<-rename(data6, c("V1"="gene", "V2"="DW"))
g<-rename(data7, c("V1"="gene", "V2"="DHet"))
h<-rename(data8, c("V1"="gene", "V2"="DKO1"))
i<-rename(data9, c("V1"="gene", "V2"="DKO2"))
j<-rename(data9, c("V1"="gene", "V2"="DKO3"))
data<-cbind(a,b,c,d,e,f,g,h,i,j)
data2<-data[,c("gene_name", "WT", "Het", "KO1", "KO2", "KO3", "DW", "DHet", "DKO1",
"DKO2", "DKO3")]

# Eliminando la línea que contiene: (_no_feature_to_alignment_not_unique) #

data3<-data2[-24067,]
tail(data3)
data4<-data3[-24066,]
tail(data4)
data5<-data4[-24065,]
tail(data5)
data6<-data5[-24064,]
tail(data6)
data7<-data6[-24063,]
tail(data7)
rownames(data7) <- data7[,1]
data7[,1] <- Null
head(data7)

```

```
# Escribiendo la tabla #  
write.table(data7, file = "Count Table.txt")  
#####
```

Al obtener la tabla de conteo de lecturas correspondiente al experimento de secuenciación, se procedió a normalizar los datos mediante el algoritmo RUVSeq implementado en Bioconductor, R (<https://www.bioconductor.org>). Este algoritmo implementa el método de (Risso et al. 2014) y permite remover variación inesperada en datos de ARN-seq de forma empírica. La tabla de conteo de lecturas fue leída en R y se implementó el código RUVSeq para normalizar la tabla de la siguiente forma (no se consideró el embrión heterocigoto):



```

# Lectura de la tabla de conteo de lecturas por gen en R #
setwd("C:/Users/Carlos/Desktop/RNA seq/HTSeq merge-bams/HTSeq analysis")
data<-read.table("Count Table.txt", header=T)
head(data)

# Separando controles y tratamientos #
data_A<-data[,c("WT1","KO1","KO2","KO3")]
write.table(data_A, file = "controls.txt")
data_B<-data[,c("DW1","DKO1","DKO2","DKO3")]
write.table(data_B, file = "treatments.txt")

# Uso de RUVSeq en controles (Normalización empírica) #
library(RUVSeq)
library(edgeR)
data<-read.table("controls.txt", header=T)
head(data)
tail(data)
dim(data)

filter <- apply(data, 1, function(x) length(x[x>5])>=2)
filtered <- data[filter,]
x <- as.factor(rep(c("WT", "KO"), WT=1, KO=3))

set <- newSeqExpressionSet(as.matrix(filtered), phenoData =
data.frame(conditions=c(rep("WT",1),rep("KO",3)), row.names=colnames(filtered)))
head(set)

library(RColorBrewer)
colors <- brewer.pal(3, "Set2")

plotRLE(set, outline=FALSE, ylim=c(-4, 4), col=colors[x])
plotPCA(set, col=colors[x], cex=1.2)

```



```

# Estimando los factores de variación no deseada utilizando residuales#
y <- DGEList(counts=data[,1:4], genes=data[,0:0])
y <- calcNormFactors(y)
y$samples
plotMDS(y)
GENOTYPE = c("WT", "KO", "KO", "KO")
data.frame(Sample=colnames(y),GENOTYPE)
design <- model.matrix(~GENOTYPE)
rownames(design) <- colnames(y)
design
y <- estimateGLMCommonDisp(y, design, verbose=TRUE)
y <- estimateGLMTrendedDisp(y, design)
y <- estimateGLMTagwiseDisp(y, design)
y
fit <- glmFit(y, design)
lrt <- glmLRT(fit, coef=2)
top <- topTags(lrt, n=nrow(set))$table
empirical <- rownames(set)[which(!(rownames(set) %in% rownames(top)[1:5000]))]
set2 <- RUVg(set, empirical, k=1)
pData(set2)
plotRLE(set2, outline=FALSE, ylim=c(-4, 4), col=colors[x])
plotPCA(set2, col=colors[x], cex=1.2)
write.table(normCounts(set2), file = "controls-Norm.txt")

# Uso de RUVSeq en tratamientos (Normalización empírica)#
library(RUVSeq)
library(edgeR)
data<-read.table("treatments.txt", header=T)
head(data)
tail(data)
dim(data)

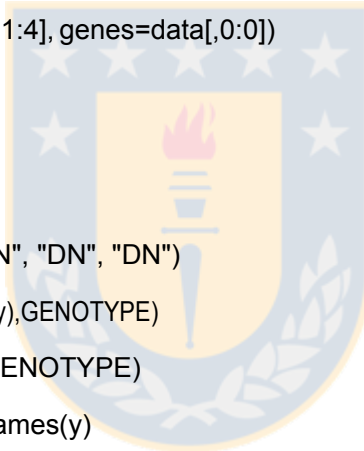
```

```

filter <- apply(data, 1, function(x) length(x[x>5])>=2)
filtered <- data[filter,]
x <- as.factor(rep(c("DW", "DKO"), WT=1, KO=3))
set <- newSeqExpressionSet(as.matrix(filtered), phenoData =
data.frame(conditions=c(rep("DW",1),rep("DN",3)), row.names=colnames(filtered)))
head(set)
library(RColorBrewer)
colors <- brewer.pal(3, "Set2")
plotRLE(set, outline=FALSE, ylim=c(-4, 4), col=colors[x])
plotPCA(set, col=colors[x], cex=1.2)

# Estimando los factores de variación no deseada utilizando residuales #
y <- DGEList(counts=data[,1:4], genes=data[,0:0])
y <- calcNormFactors(y)
y$samples
plotMDS(y)
GENOTYPE = c("DW", "DN", "DN", "DN")
data.frame(Sample=colnames(y), GENOTYPE)
design <- model.matrix(~GENOTYPE)
rownames(design) <- colnames(y)
design
y <- estimateGLMCommonDisp(y, design, verbose=TRUE)
y <- estimateGLMTrendedDisp(y, design)
y <- estimateGLMTagwiseDisp(y, design)
y
fit <- glmFit(y, design)
lrt <- glmLRT(fit, coef=2)
top <- topTags(lrt, n=nrow(set))$table
empirical <- rownames(set)[which(!(rownames(set) %in% rownames(top)[1:5000]))]
set2 <- RUVg(set, empirical, k=1)
pData(set2)

```





```
plotRLE(set2, outline=FALSE, ylim=c(-4, 4), col=colors[x])
plotPCA(set2, col=colors[x], cex=1.2)
write.table(normCounts(set2), file = "treatments-Norm.txt")
#####
```

El análisis de expresión diferencial de transcritos entre los genotipos silvestre (del inglés “wild-type”, WT) vs KO fue llevado a cabo mediante el uso del código edgeRun (Dimont et al. 2015), un paquete en R que implementa un test incondicional exacto más poderoso que la versión anterior del programa, edgeR (<http://cran.r-project.org/web/packages/edgeRun>). El código utilizado para estos efectos (WT vs KO en controles y tratamientos) fue el siguiente:



```

# Prueba de expresión diferencial en controles #

library(edgeRun)

setwd("C:/Users/Carlos/Desktop/RNA seq/HTSeq merge-bams/HTSeq analysis")

data<-read.table("controls-Norm.txt", header=T)

head(data)

dim(data)

y <- DGEList(counts=data[,1:4], genes=data[,0:0])

y <- calcNormFactors(y)

y$samples

plotMDS(y)

group = factor (c("1", "2", "2", "2"))

y <- DGEList(counts=data,
group=group) y <-
calcNormFactors(y)

y <- estimateCommonDisp(y)

y <- estimateTagwiseDisp(y)

de.edgeRun <- UCexactTest(y,upper=50000)

topTags(de.edgeRun)

out <- topTags(de.edgeRun, n=Inf, adjust.method="BH")

keep<-out$table$FDR<=0.05 & abs(out$table$logFC)>= 1

out[keep,]

write.table(out, file = "edgeRun-controls.csv", sep = ",", col.names = NA, qmethod =
"double")

#####

library(edgeRun)

setwd("C:/Users/Carlos/Desktop/RNA seq/HTSeq merge-bams/HTSeq analysis")

data<-read.table("treatments-Norm.txt", header=T)

head(data)

dim(data)

y <- DGEList(counts=data[,1:4], genes=data[,0:0])

```



```

y <- calcNormFactors(y)
y$samples
plotMDS(y)
group = factor (c("1","2", "2", "2"))
y <- DGEList(counts=data,group=group)
y <- calcNormFactors(y)
y <- estimateCommonDisp(y)
y <- estimateTagwiseDisp(y)
de.edgeRun <- UCexactTest(y,upper=50000)
topTags(de.edgeRun)
out <- topTags(de.edgeRun, n=Inf, adjust.method="BH")
keep<-out$table$FDR<=0.05&abs(out$table$logFC)>=1
out[keep,]
write.table(out, file = "edgeRun-treatments.csv", sep = ",", col.names = NA, qmethod =
"double")
#####

```

El archivo de salida arrojó la lista de genes diferencialmente expresados entre el genotipo silvestre versus el KO en condición control y bajo tratamiento con doxorubicina, con un valor p y una tasa de descubrimientos falsos correspondiente (del inglés “False Discovery Rate”, FDR). Se seleccionaron los genes con una tasa de descubrimiento falso menor a 0.001.

### **Llamado de variantes desde los datos de ARN-seq**

Los archivos de alineamiento de la secuenciación fueron subidos a la plataforma Galaxy, una plataforma web de código abierto para el análisis intensivo de datos de secuenciación masiva (<https://usegalaxy.org>). Los archivos BAM individuales provenientes del ARN-Seq fueron cargados a la plataforma Galaxy y organizados en base a las coordinadas genómicas (“sorting”). Después del ordenamiento, se realizó

llamada de variantes diploides simple a lo largo de todo el genoma utilizando la herramienta Freebayes (<https://github.com/ekg/freebayes>), una herramienta de llamado de variantes diseñada para el servidor. Las variantes obtenidas fueron filtradas utilizando el programa VCFlib (<https://github.com/vcflib/vcflib>) con los siguientes criterios: -f "DP> 10" (Profundidad sobre 10 lecturas) y -f "QUAL> 30" (probabilidad mínima de error de escala de Phred superior a 30). Después de los filtros utilizados, se confeccionaron histogramas de las variantes resultantes mediante el uso de un script diseñado por nosotros, implementado en el lenguaje de programación R (véase `genotype_variants.R` en resultados). Para la identificación de variantes comunes en el KO no presente en la contraparte de tipo silvestre (WT), utilizamos varias herramientas de VCFlib toolkit disponible en Galaxy. Las variantes de los archivos KO coleccionadas en el formato VCF (acrónimo del inglés “**V**ariant **C**alling **F**ormat”) fueron intersecadas utilizando el programa VCF-VCF intersect (en base al genoma de referencia mm10) y anotadas en base a los genotipos de la contraparte silvestre mediante el programa “VCFannotateGenotypes”. El archivo resultante de la anotación de variantes fue filtrado seleccionando las líneas no anotadas con la contraparte silvestre (Filter and Sort). Finalmente, se obtuvo un archivo de salida en formato VCF con las variantes ligadas al genotipo KO.

### **Visualización de variantes en R**

Como se mencionó anteriormente, desarrollamos un script escrito en el lenguaje de programación R (`genotype_variants.R`) para la visualización adecuada de las variantes a lo largo de los cromosomas en experimentos de secuenciación en ratón. El script para estos efectos toma como entrada la intersección de archivos VCF provenientes de genotipos silvestre y KO en formato `.txt` y produce un archivo de salida consistente en un gráfico de frecuencia de variantes por cromosoma para ambos genotipos. El script también incluye la detección estadística de cromosomas

con variantes ligadas al genotipo KO. En concordancia con lo anterior, testeamos la distribución de frecuencia de las variantes con la prueba estadística de Cochran-Armitage, disponible en el paquete DescTools, implementado en el programa estadístico R (<https://cran.r-project.org/web/packages/DescTools/index.html>). Las variantes detectadas se agruparon cada 10 millones de pares de bases de acuerdo con sus coordenadas cromosómicas, se ordenan en una tabla de contingencia y se graficaron. Después de esto, se implementó la prueba de tendencia de Cochran-Armitage para identificar cromosomas que contienen variantes ligadas al genotipo KO, basadas en la distribución de frecuencia de ambos genotipos. Los gráficos son compatibles con el paquete ggplot2, también implementado en R (<https://cran.r-project.org/web/packages/ggplot2/index.html>).



## IV. RESULTADOS

### OBJETIVO 1: **Determinar respuestas transcripcionales dependientes de SALL2**

1.1 Definir blancos transcripcionales específicos de SALL2 *in vitro* e *in vivo*.

#### **Experimento de ARN-seq en células MEF *Sall2* silvestres y KO**

Con el objeto de responder inicialmente esta interrogante, se realizó un experimento de ARN-seq en células MEF silvestres y KO para *Sall2* tratadas o no con la droga doxorrubicina a una concentración 1  $\mu$ M por 16 horas. El tratamiento busca ejercer estrés genotóxico en células MEF tal como se describe en (Kenzelmann Broz et al. 2013) e identificar genes blanco de *Sall2* en condiciones normales y bajo estrés genotóxico. El tiempo de tratamiento se eligió en base a estudios previos de laboratorio que mostraban la inducción de *Sall2* a 16 horas de tratamiento (Escobar et al. 2015). RNA total fue aislado desde cada condición experimental y procesado para secuenciamiento de acuerdo con lo descrito en **“Preparación de muestras para ARN-seq.”** sección METODOS. Este ARN fue utilizado para la construcción de las librerías de secuenciación según procedimientos de “Genomic Core Facility” de Cambridge University, y la secuenciación fue efectuada en la máquina NextSeq500 (illumina), con lecturas de 150 pares de bases de largo, de tipo “single end”.

El alineamiento de las lecturas obtenidas al ensamblaje mm10 correspondiente al genoma de ratón fue calculado desde los archivos en formato BAM utilizando el comando SAMtools flagstats, en BASH. La implementación del comando fue la siguiente (ejemplo para la muestra MEF WT-Control):

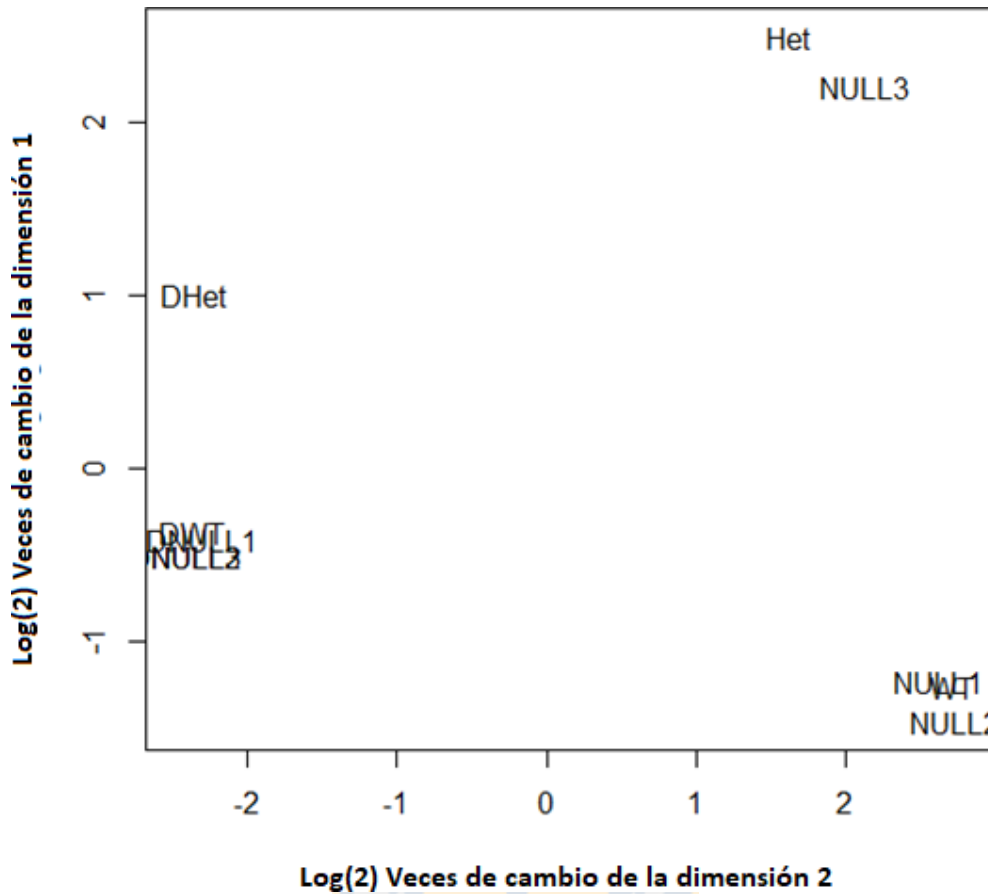
```
samtools flagstat WT-Control
```

Las lecturas alineadas por muestra fueron las siguientes:

Muestra	Número de lecturas	% de alineamiento (mm10)
WT-Control (RP5)	37,048,909	95.1
Het-Control (RP7)	38,304,413	93.22
KO-1-Control (RP1)	36,729,837	95.41
KO-2-Control (RP3)	35,120,348	94.1
KO-3-Control (RP9)	36,302,287	93.98
WT-DoxoR (RP6)	37,110,439	94.3
Het-DoxoR (RP8)	33,728,975	91.29
KO-1-DoxoR (RP2)	32,933,559	94.74
KO-2-DoxoR (RP4)	37,602,962	94.51
KO-3-DoxoR (RP10)	34,947,283	92.91

**Tabla 1: Grado de alineamiento del experimento de secuenciación en MEF silvestres y KO para *Sall2*.** Se muestra el número de lecturas alineadas al genoma de referencia y el correspondiente porcentaje de alineamiento con respecto al total de lecturas por muestra secuenciada. Fuente: Elaboración propia.

Con el objeto de reducir la dimensionalidad de los datos y visualizar la correlación de las muestras en términos de expresión génica, se recurrió al escalamiento multidimensional del experimento (derivado del inglés "Multidimensional scaling"), técnica que trata los valores de expresión génica de forma matricial y transpone los datos en base a dos coordenadas (dim1 y dim2), representándose la muestra analizada en un gráfico (Taguchi and Oono 2005). La técnica fue implementada con la herramienta edgeRun (Dimont et al. 2015), generándose el siguiente gráfico para las 10 muestras secuenciadas:



**Figura 8: Escalamiento multidimensional del experimento de ARN-seq.** La tabla de conteo de datos (“Count Table”) fue analizada con el paquete “edgeRun” en R y se generó por escalamiento multidimensional de las muestras secuenciadas. WT=Silvestre, Het=Heterocigoto, KO= “knockout”. El prefijo “D” se refiere a las muestras tratadas con doxorrubicina. Fuente: Elaboración propia.

Del análisis de escalamiento multidimensional se desprende que el tratamiento con doxorrubicina provoca una perturbación suficiente para separar muestras control de las muestras tratadas, pero no se evidencia clara separación entre genotipos.



## **Detección de variantes a nivel de genoma completo desde datos de ARN-seq**

El paso siguiente contemplaría comparar los transcriptomas de los genotipos silvestre (WT) versus KO para *Sal12* e identificar blancos transcripcionales en condiciones control y bajo estrés genotóxico. Como se comentó en la introducción, los aspectos genéticos son importantes al atribuir expresión génica y fenotipos a una mutación dada, por lo cual se decidió realizar una supervisión genética previa a estos análisis. Respecto a esto, se pueden identificar distintas fuentes de variación genética que ocurren en los modelos KO y deben ser consideradas en los experimentos llevados a cabo con ellos. Para abordar esta problemática, diseñamos un protocolo computacional para detectar introgresión de variantes en experimentos de ARN-seq. Este protocolo permite la identificación de variantes desde datos de ARN-seq, la cual puede ser implementada en la plataforma computacional Galaxy (Afgan et al. 2016). Si es que el genoma de la célula madre (ESC) en la cual se hizo la recombinación homóloga está disponible, las variantes que presentaría dicho genoma con respecto al genoma de referencia de ratón (C57BL/6J) pueden ser utilizadas para identificar y anotar dichas variantes en experimentos de secuenciación masiva. A la fecha, el proyecto de secuenciación de ratón (del inglés “Mouse Genome Project”, <http://www.sanger.ac.uk/science/data/mouse-genomes-project>) ha secuenciado y producido variantes curadas en al menos 36 genomas de cepas murinas (Keane et al. 2011; Doran et al. 2016). En base al repertorio de variantes que ofrece el proyecto, hemos implementado el siguiente protocolo de detección y caracterización de variantes en proyectos de secuenciación en ratón, incluyendo estudios con animales KO (véase figura 9).

Se escribió un script en R para la visualización de variantes por cromosoma desde animales silvestre (WT) y KO denominado “Genotype\_variants.R”. Este código al ser ejecutado ingresa como archivos de entrada las variantes desde ambos genotipos en formato VCF y entrega como archivos de salida un histograma de

variantes por coordenadas genómicas a lo largo del genoma completo de ratón, incluyendo la detección de cromosoma(s) con distribución diferencial de variantes entre genotipos. El cuerpo del código del programa que implementa al menos 6 librerías provenientes de bioconductor, se presenta a continuación:



```
##### Genotype_variants.R #####
```

```
library(DescTools)
```

```
library(ggplot2)
```

```
library(reshape2)
```

```
library(gridExtra)
```

```
library(grid)
```

```
library(dplyr)
```

```
WT_variants <- file.choose()
```

```
WT<-read.table(WT_variants, skip="###")
```

```
WT<-subset(WT, select = c(1, 2))
```

```
head(WT)
```

```
dim(WT)
```

```
KO_variants <- file.choose()
```

```
KO<-read.table(KO_variants, skip="###")
```

```
KO<-subset(KO, select = c(1, 2))
```

```
head(KO)
```

```
dim(KO)
```

```
##### "subsetting" cromosomas #####
```

```
WT_chr1<-subset(WT, V1 == "chr1")
```

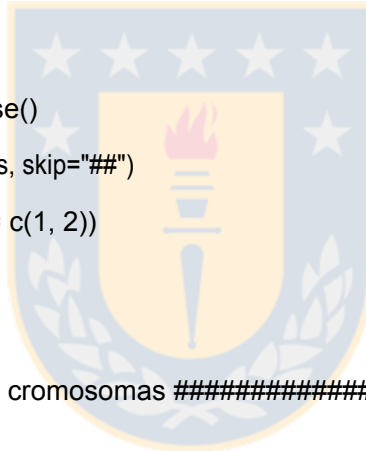
```
... (hasta chr19)...
```

```
WT_chrX<-subset(WT, V1 == "chrX")
```

```
KO_chr1<-subset(KO, V1 == "chr1")
```

```
... (hasta chr19)...
```

```
KO_chrX<-subset(KO, V1 == "chrX")
```



```
##### "subsetting" coordenadas genómicas #####
```

```
WT_chr1<-subset(WT_chr1, select = c(2))
```

```
...(hasta Chr19)...
```

```
WT_chrX<-subset(WT_chrX, select = c(2))
```

```
KO_chr1<-subset(KO_chr1, select = c(2))
```

```
...(hasta Chr19)...
```

```
KO_chrX<-subset(KO_chrX, select = c(2))
```

```
par(mfrow=c(2,10))
```

```
#generando histogramas de frecuencias de variantes de acuerdo con las coordenadas#
```

```
WT_chr1$Genotype<-"WT"
```

```
KO_chr1$Genotype<-"KO"
```

```
Chr1 <- rbind(WT_chr1, KO_chr1)
```

```
p1 <- ggplot(Chr1, aes(V2, fill = Genotype)) + geom_histogram(alpha = 0.5, aes(y =  
..count..), position = 'identity') + xlab("Position in Chr1") + ylab("Variants")  
+labs(title="Chr1") + theme(legend.position="none")
```

```
...(p2=Chr2, p3=Chr3 sucesivamente, hasta Chr19)...
```

```
WT_chrX$Genotype<-"WT"
```

```
KO_chrX$Genotype<-"KO"
```

```
ChrX <- rbind(WT_chrX, KO_chrX)
```

```
pX <- ggplot(ChrX, aes(V2, fill = Genotype)) + geom_histogram(alpha = 0.5, aes(y =  
..count..), position = 'identity') + xlab("Position in ChrX") + ylab("Variants")  
+labs(title="ChrX") + theme(legend.position="none")
```

```
#ajuste del eje X de acuerdo al histograma que presente la mayor cantidad de variantes#
```

```
max=max(max(layer_scales(p1)$y$range$range), max(layer_scales(p2)$y$range$range),  
max(layer_scales(p3)$y$range$range), max(layer_scales(p4)$y$range$range),  
max(layer_scales(p5)$y$range$range), max(layer_scales(p6)$y$range$range),  
max(layer_scales(p7)$y$range$range), max(layer_scales(p8)$y$range$range),  
max(layer_scales(p9)$y$range$range), max(layer_scales(p10)$y$range$range),  
max(layer_scales(p11)$y$range$range), max(layer_scales(p12)$y$range$range),
```

```

max(layer_scales(p13)$y$range$range),
max(layer_scales(p14)$y$range$range),max(layer_scales(p15)$y$range$range),
max(layer_scales(p16)$y$range$range), max(layer_scales(p17)$y$range$range),
max(layer_scales(p18)$y$range$range),
max(layer_scales(p19)$y$range$range), max(layer_scales(pX)$y$range$range))

```

```

p1 <- ggplot(Chr1, aes(V2, fill = Genotype)) + geom_histogram(alpha = 0.5, aes(y =
..count..), position = 'identity') + xlab("Position in Chr1") + ylab("Variants")
+labs(title="Chr1") + theme(legend.position="none") + ylim(0, max)

```

...(p2=Chr2, p3=Chr3 sucesivamente, hasta Chr19)...

```

pX <- ggplot(ChrX, aes(V2, fill = Genotype)) + geom_histogram(alpha = 0.5, aes(y =
..count..), position = 'identity') + xlab("Position in ChrX") + ylab("Variants")
+labs(title="ChrX") + theme(legend.position="none") + ylim(0, max)

```

#Ajuste de leyendas#

```

grid_arrange_shared_legend <- function(..., ncol = length(list(...)), nrow = 1, position =
c("bottom", "right")){

```

```

  plots <- list(...)

```

```

  position <- match.arg(position)

```

```

  g <- ggplotGrob(plots[[1]] + theme(legend.position = position))$grobs

```

```

  legend <- g[[which(sapply(g, function(x) x$name) == "guide-box")]]

```

```

  lheight <- sum(legend$height)

```

```

  lHeight <- sum(legend$Height)

```

```

  gl <- lapply(plots, function(x) x + theme(legend.position="none"))

```

```

  gl <- c(gl, ncol = ncol, nrow = nrow)

```

```

  combined <- switch(position,

```

```

    "bottom" = arrangeGrob(do.call(arrangeGrob, gl),

```

```

      legend,

```

```

      ncol = 1,

```

```

      heights = unit.c(unit(1, "npc") - lheight, lheight)),

```

```

heights = unit.c(unit(1, "npc") - lheight, lheight),
  "right"=arrangeGrob(do.call(arrangeGrob,gl),
    legend,
    ncol = 2,
    heights = unit.c(unit(1, "npc") - lHeight, lHeight)))

grid.newpage()
grid.draw(combined)

#return gtable invisibly
invisible(combined)
}
dsamp <- diamonds[sample(nrow(diamonds), 1000), ]

g<-grid_arrange_shared_legend(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12,p13,
p14, p15, p16, p17, p18, p19, pX, ncol = 3, nrow=7)
ggsave("graph.pdf", g, width=36, height=50, units="cm")

# Prueba estadística de Cochran-Armitage #
par(mar = rep(2, 4))
WTchr1<-hist(WT_chr1$V2, breaks=c(0,1e+7, 2e+7, 3e+7, 4e+7, 5e+7, 6e+7, 7e+7, 8e+7,
9e+7,10e+7,11e+7,12e+7,13e+7,14e+7,15e+7,16e+7,17e+7,18e+7,19e+7,20e+7))
KOchr1<-hist(KO_chr1$V2, breaks=c(0,1e+7, 2e+7, 3e+7, 4e+7, 5e+7, 6e+7, 7e+7, 8e+7,
9e+7, 10e+7, 11e+7, 12e+7, 13e+7, 14e+7, 15e+7, 16e+7, 17e+7, 18e+7, 19e+7, 20e+7))
Chr1=data.frame(WTchr1$counts, KOchr1$counts)
a=CochranArmitageTest(Chr1)
Chr1=a
...(Chr2=b, Chr3=c sucesivamente, hasta Chr19)...

```

```

WTchrX<-hist(WT_chrX$V2, breaks=c(0,1e+7, 2e+7, 3e+7, 4e+7, 5e+7, 6e+7, 7e+7, 8e+7,
9e+7, 10e+7, 11e+7, 12e+7, 13e+7, 14e+7, 15e+7, 16e+7, 17e+7, 18e+7))

KOchrX<-hist(KO_chrX$V2, breaks=c(0,1e+7, 2e+7, 3e+7, 4e+7, 5e+7, 6e+7, 7e+7, 8e+7,
9e+7, 10e+7, 11e+7, 12e+7, 13e+7, 14e+7, 15e+7, 16e+7, 17e+7, 18e+7))

ChrX=data.frame(WTchrX$counts, KOchrX$counts)

t=CochranArmitageTest(ChrX)

Chr19=t

p.value<-
c(a$p.value,b$p.value,c$p.value,d$p.value,e$p.value,f$p.value,g$p.value,h$p.value,i$p.
value,j$p.value,k$p.value,l$p.value,m$p.value,n$p.value,o$p.value,p$p.value,q$p.value
,r$p.value,s$p.value,t$p.value)

Chromosome<-
c("chr1","chr2","chr3","chr4","chr5","chr6","chr7","chr8","chr9","chr10","chr11","chr12
","chr13","chr14","chr15","chr16","chr17","chr18","chr19","chrX")

significance <- p.value < 0.05

matrix<-cbind(Chromosome, p.value, significance)

Summary<-data.frame(matrix)

Summary
pdf("Summary.pdf", height=11, width=8.5)

grid.table(Summary)

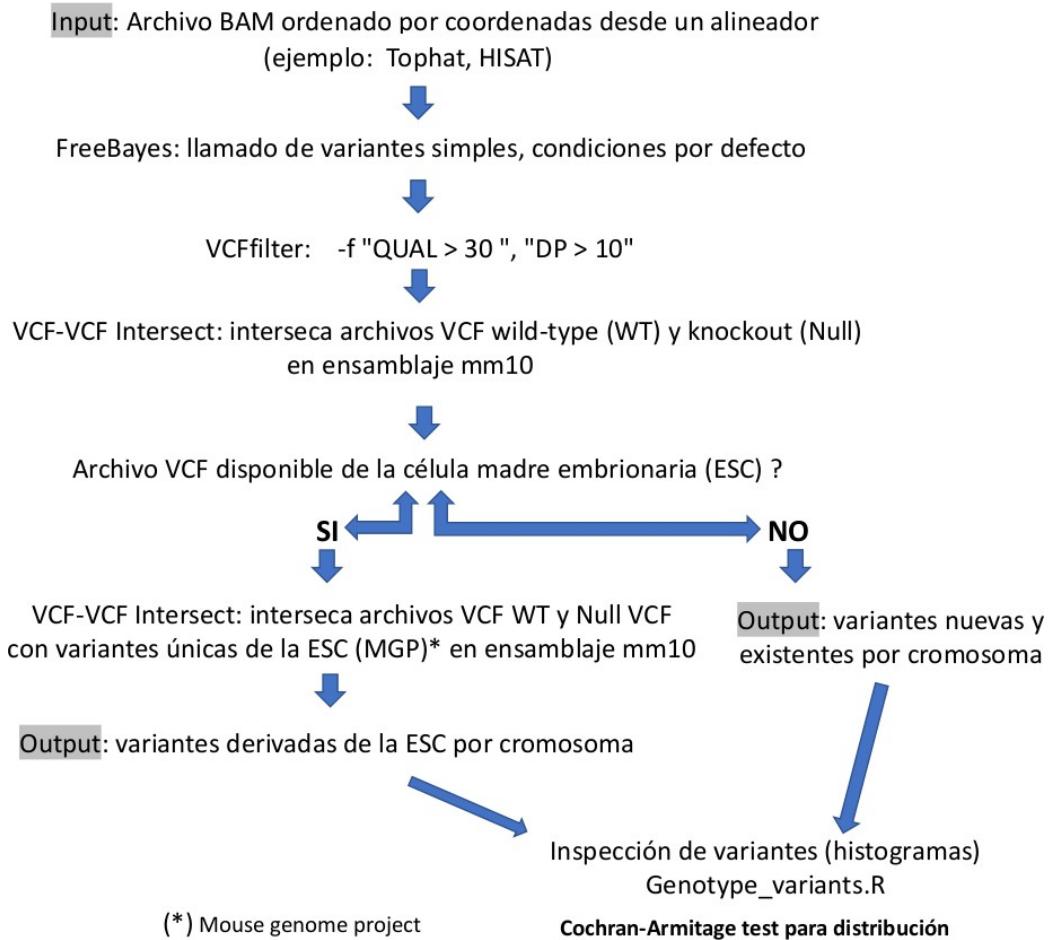
dev.off()

proc.time()

sessionInfo()

#####

```



**Sitio ftp de Mouse genome Project:**

`ftp://ftp-mouse.sanger.ac.uk/REL-1505-SNPs_Indels/strain_specific_vcfs/`

**Código UNIX para convertir archivos VCF:**

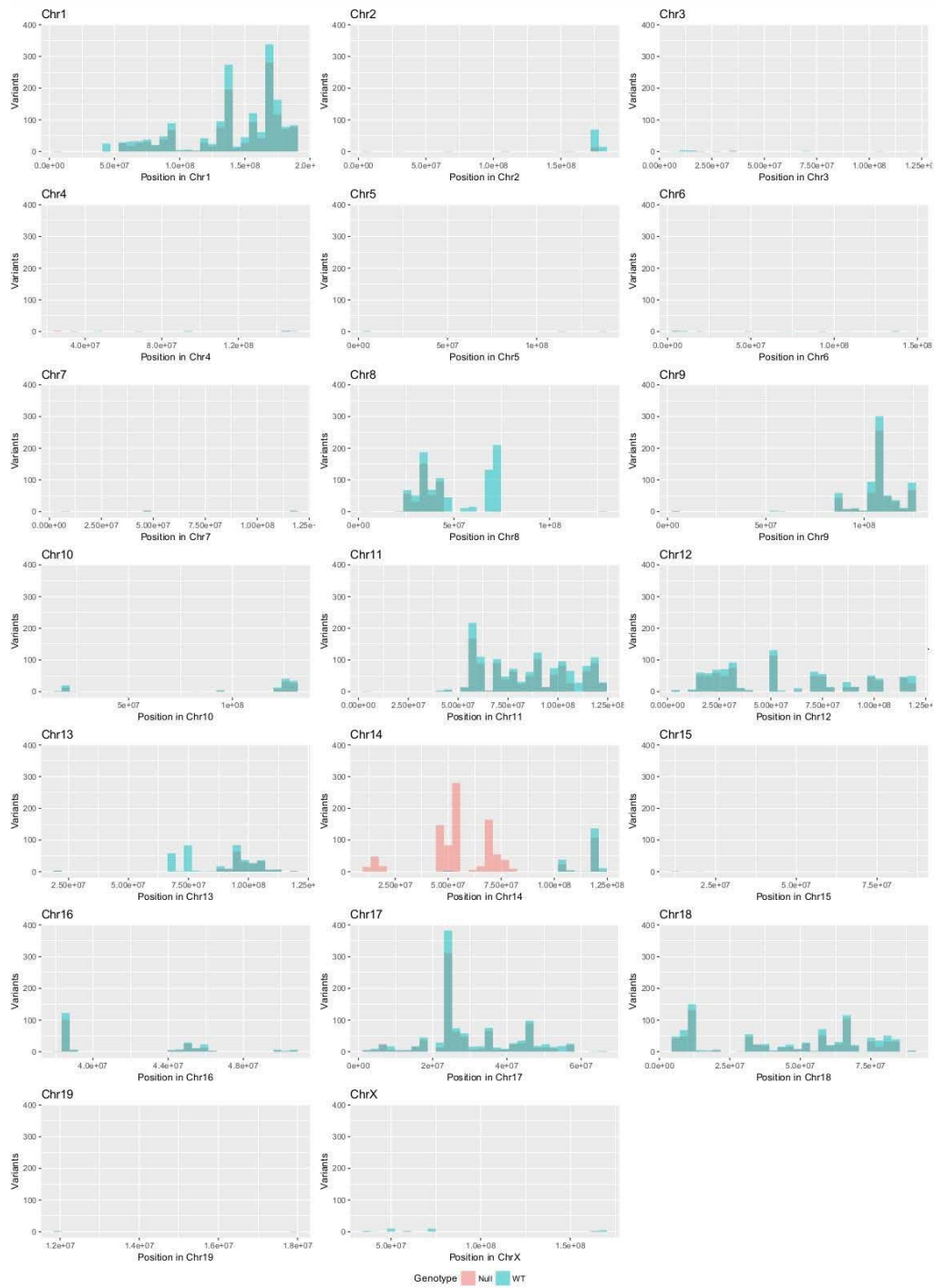
`awk '{if($0 !~ /^#/) print "chr"$0; else print $0}' MGP_VCF.vcf > MGP_VCF_with_chr.vcf`



**Figura 9: Protocolo computacional para la detección de variantes introgresadas derivadas de ESC en estudios de secuenciación masiva en ratón.** El archivo de entrada corresponde a un archivo BAM alineado por cada genotipo (usualmente WT y KO) en el genoma de ratón correspondiente (por ejemplo, un archivo de salida de la herramienta de alineamiento Tophat o HISAT2 en el ensamblaje de ratón mm10, formato BAM). El programa de llamado de variantes Freebayes (llamado de variantes simples, implementado en Galaxy) produce un archivo en formato VCF (del inglés “Variant Calling Format”) para cada archivo BAM respecto al genoma de referencia (usualmente el ensamblaje mm10 del genoma de la cepa C57BL/6J). Filtramos estos archivos VCF usando la librería VCFlib, con los siguientes parámetros: -f "QUAL > 30", "f" DP > 10 " (el parámetro “QUAL” representa calidad de la variante y “DP” la profundidad o el número de lecturas alineadas en la posición genómica de la variante). A continuación, el programa de intersección VCF-VCF interseca los archivos VCF de cada genotipo para obtener la variación promedio de cada genotipo (Se utiliza el ensamblaje mm10, con parámetros predeterminados del programa). Si el genoma de la ESC utilizado para la recombinación homóloga en el KO está disponible y las variantes se caracterizan correctamente, podemos utilizar estas variantes para identificar variantes introgresadas de la ESC en los archivos VCF de cada genotipo. Usamos archivos VCF disponibles en el proyecto del genoma del ratón (<http://www.sanger.ac.uk/science/data/mouse-genomes-project>) basados en la versión del genoma del ratón GRCm38, compatible con la construcción mm10 (versión REL-1505-SNPs\_Indels). En estos archivos VCF, se necesita un cambio de los números a los nombres de los cromosomas (véase el código UNIX). En el caso en que el genoma de la ESC no está disponible, se obtienen las variantes generadas por la deriva génica y derivadas de la ESC. Para identificar los cromosomas con una distribución diferencial de variantes entre genotipos, aplicamos la prueba de Cochran-Armitage para tendencia de las distribuciones de frecuencia. Fuente: Elaboración propia.

El modelo KO para *Sall2* utilizado para el experimento de secuenciación fue generado en células madre derivada de la línea 129P2OlaHsd (Sato et al. 2003) por lo cual se procedió a probar el script detectando estas variantes introgresadas en el genoma de ambos genotipos de este ratón. El gráfico de genoma completo del modelo

KO para *Sall2* anotado con variantes únicas de 129P2OlaHsd, se detalla a continuación (figura 10):



**Figura 10: Detección de variantes de la célula madre 129P2OlaHsd en el experimento de secuenciación del modelo KO para *Sall2*.** Histograma de las variantes únicas de 129P2OlaHsd ESC graficadas por cromosoma desde el ARN de células MEF *Sall2* silvestres y KO. Se agruparon las coordenadas genómicas de cada cromosoma cada 10 millones de bases y se graficaron las variantes en histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes de un embrión de tipo silvestre y las barras rojas representan las variantes promedio de tres embriones KO para *Sall2*. Fuente: Elaboración propia.

Como se puede visualizar en el gráfico, la introgresión del genoma de 129P2OlaHsd es patente en 10 cromosomas en ambos genotipos, destacando el cromosoma 14 en donde la recombinación homóloga ocurrió. La huella congénica asociada al genotipo KO es patente en el cromosoma 14, en donde se muestra que la acumulación de variantes en este genotipo es evidentemente superior a las del genotipo silvestre (wild type, WT). En resumen, las células provenientes de estos ratones tienen un fondo genético mixto entre C57BL/6J-129P2OlaHsd y el genotipo KO presenta huella congénica asociada.

Se procedió a probar el script usando como entrada la totalidad de las variantes del experimento realizado con el modelo KO para *Sall2*. Como prueba de concepto, se consideró adicionalmente las variantes colectadas desde los datos de ARN-Seq de cinco modelos KO, disponibles de forma pública en la base de datos GEO DataSets (<https://www.ncbi.nlm.nih.gov/gds>). Los experimentos de secuenciación seleccionados tienen los siguientes números de acceso: GSE71126, GSE81082, GSE47395, GSE65686 y GSE83555 para los modelos KO de los genes *Mepc2*, *Gtf2ird1*, *Stc1*, *Itch* y *Hnrnpd/AUF-1*, respectivamente. Previo a la obtención de las variantes, se substrajeron las variantes ocurrentes en el genoma C57BL/6J en cada estudio. El

gráfico de genoma completo del KO para *Sall2*, considerando la totalidad de las variantes por genotipo, se detalla a continuación (véase figura 11):

### Sall2 KO

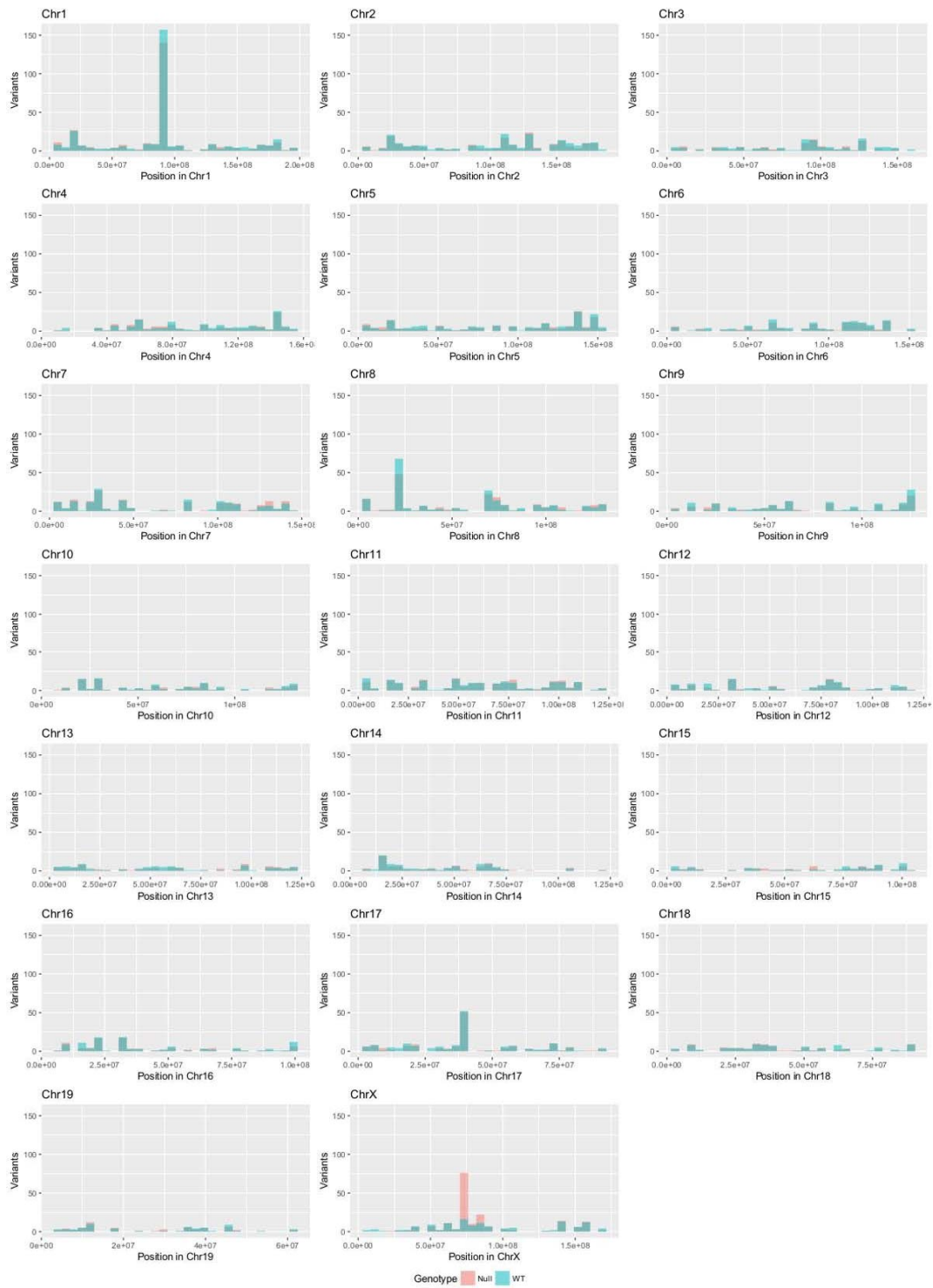


**Figura 11: Detección de la totalidad de las variantes en el experimento de secuenciación del modelo KO para *Sall2*.** Histograma de variantes nuevas y existentes por cromosoma desde el ARN de células MEF *Sall2* silvestres (WT) y KO. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes de un embrión de tipo *Sall2* silvestre y las barras rojas representan las variantes promedio de tres embriones KO para *Sall2*. Fuente: Elaboración propia.

Se puede apreciar que las variantes en el modelo KO para *Sall2* presentan un número y una distribución muy similar a las variantes anotadas de la figura 10, lo que sugiere que la mayoría de las variantes en este modelo provienen de las variantes de la célula madre derivada de 129P2OlaHsd.

Como se mencionó anteriormente, se exploró además la distribución de variantes en los modelos KO anteriormente mencionados. El gráfico de genoma completo del modelo KO para *Mecp2*, considerando la totalidad de las variantes por genotipo, se detalla a continuación (véase figura 12):

# Mecp2 KO

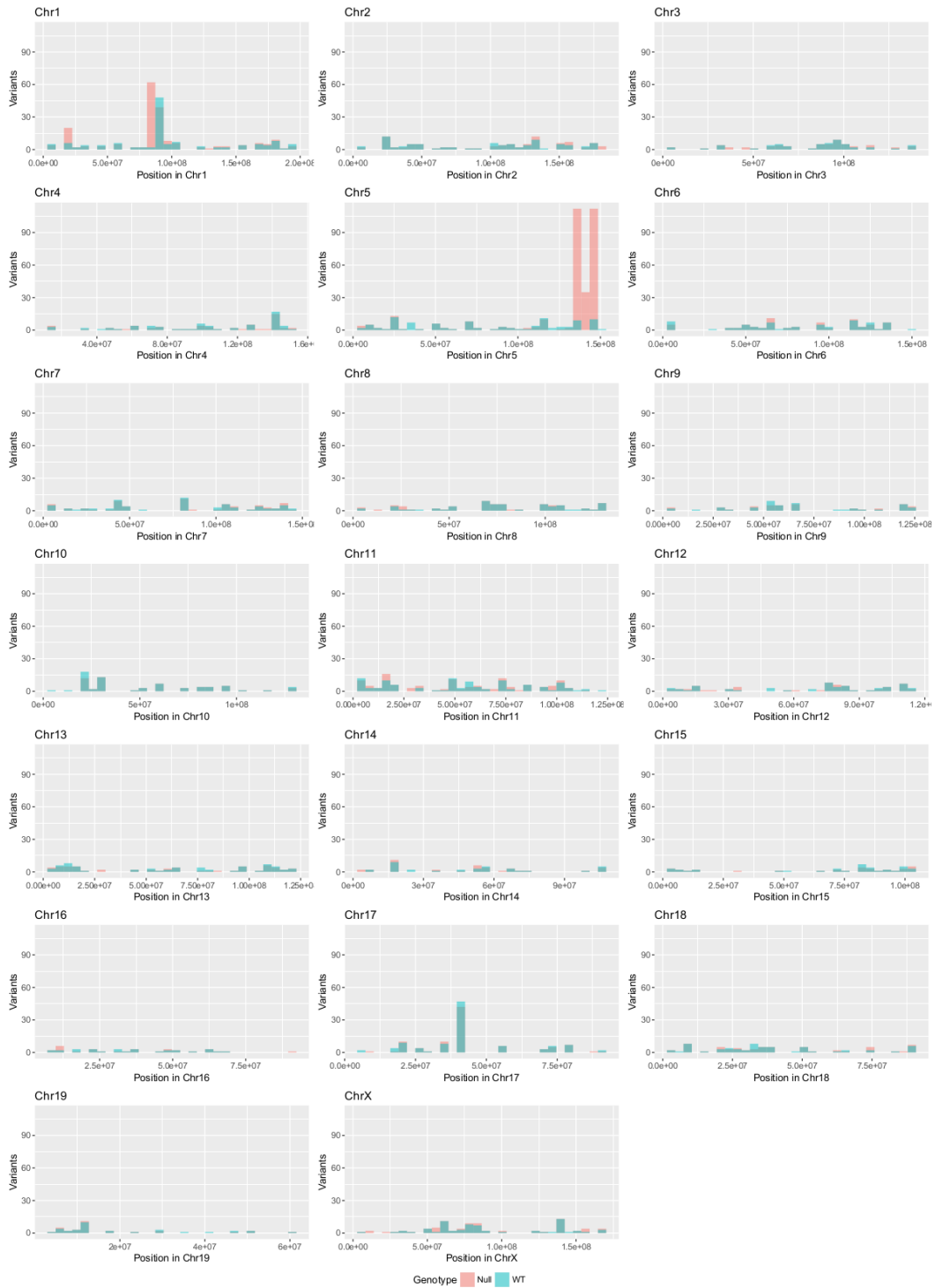


**Figura 12: Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Mecp2*** Histograma de variantes nuevas y existentes por cromosoma desde el ARN del neuroepitelio olfatorio de ratones *Mecp2* silvestre (WT) y KO. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes promedio de tres neuroepitelios *Mecp2* silvestres y las barras rojas representan las variantes promedio de tres neuroepitelios KO para *Mecp2*. Fuente: Elaboración propia.

En este estudio se puede observar que el modelo KO para *Mecp2* es congénico y fue retrocruzado correctamente con C57BL/6J, ya que las variantes de 129P2OlaHsd (ESC utilizada para la recombinación homóloga en este modelo) no son predominantes en el genoma recipiente C57BL/6J (véase figura 12). El gen *Mecp2* se ubica en el cromosoma X (ChrX:74026592-74085690, <http://www.informatics.jax.org/marker/MGI:99918>). Este cromosoma tiene una de las tasas más bajas de recombinación descritas (Jensen-Seaman et al. 2004), por lo que se espera que los eventos de recombinación homólogas sean bajos con C57BL/6J y que predominantemente el cromosoma X se componga en su mayoría del genoma de 129P29OlaHsd. Como se ve en la figura 12, las variantes asociadas al genotipo KO son patentes, al igual que en el caso del modelo KO para *Sall2*.

El gráfico de genoma completo del modelo KO para *Gtf2ird1*, considerando la totalidad de las variantes por genotipo, se detalla a continuación (véase figura 13):

# Gtf2ird1 KO



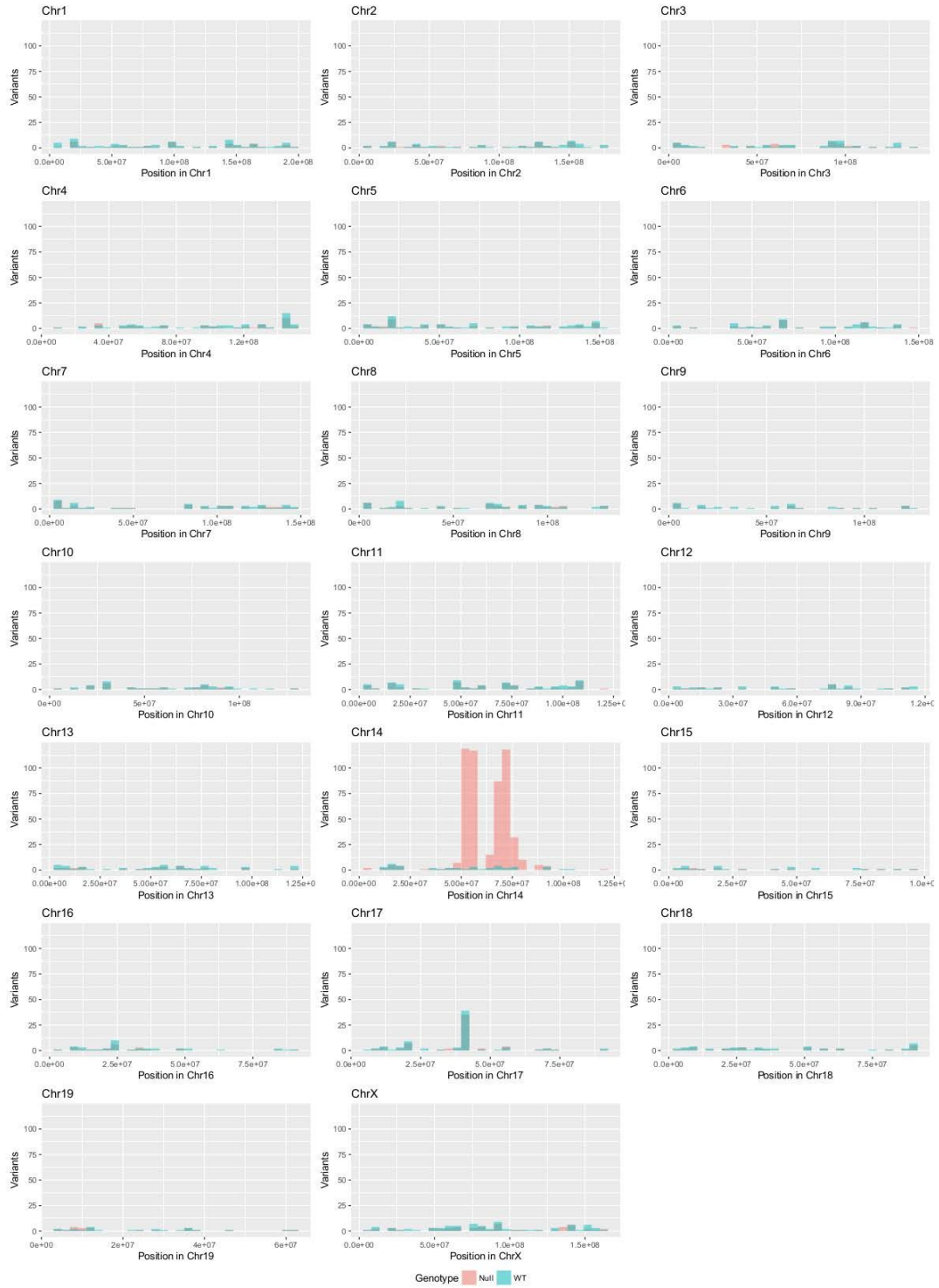


**Figura 13: Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Gtf2ird1*** Histograma de variantes nuevas y existentes por cromosoma desde el ARN de tejido epitelial de ratones *Gtf2ird1* silvestres (WT) y KO. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes promedio de tres epitelios de tipo *Gtf2ird1* silvestre y las barras rojas representan las variantes promedio de tres neuroepitelios KO para *Gtf2ird1*. Fuente: Elaboración propia.

Al igual que en el caso del modelo KO para *Mecp2*, el modelo KO para *Gtf2ird1* es congénico. Este modelo presenta una retrocruza correcta con el genoma recipiente C57BL/6J y marcada huella congénica en el cromosoma 5, en donde se ubica el gen que codifica para *Gtf2ird1* y donde ocurrió la recombinación homóloga (Chr5:134357656-134456716, <http://www.informatics.jax.org/marker/MGI:1861942>).

El gráfico de genoma completo del modelo KO para *Stc1*, considerando la totalidad de las variantes por genotipo, se detalla a continuación (véase figura 14):

# Stc1 KO

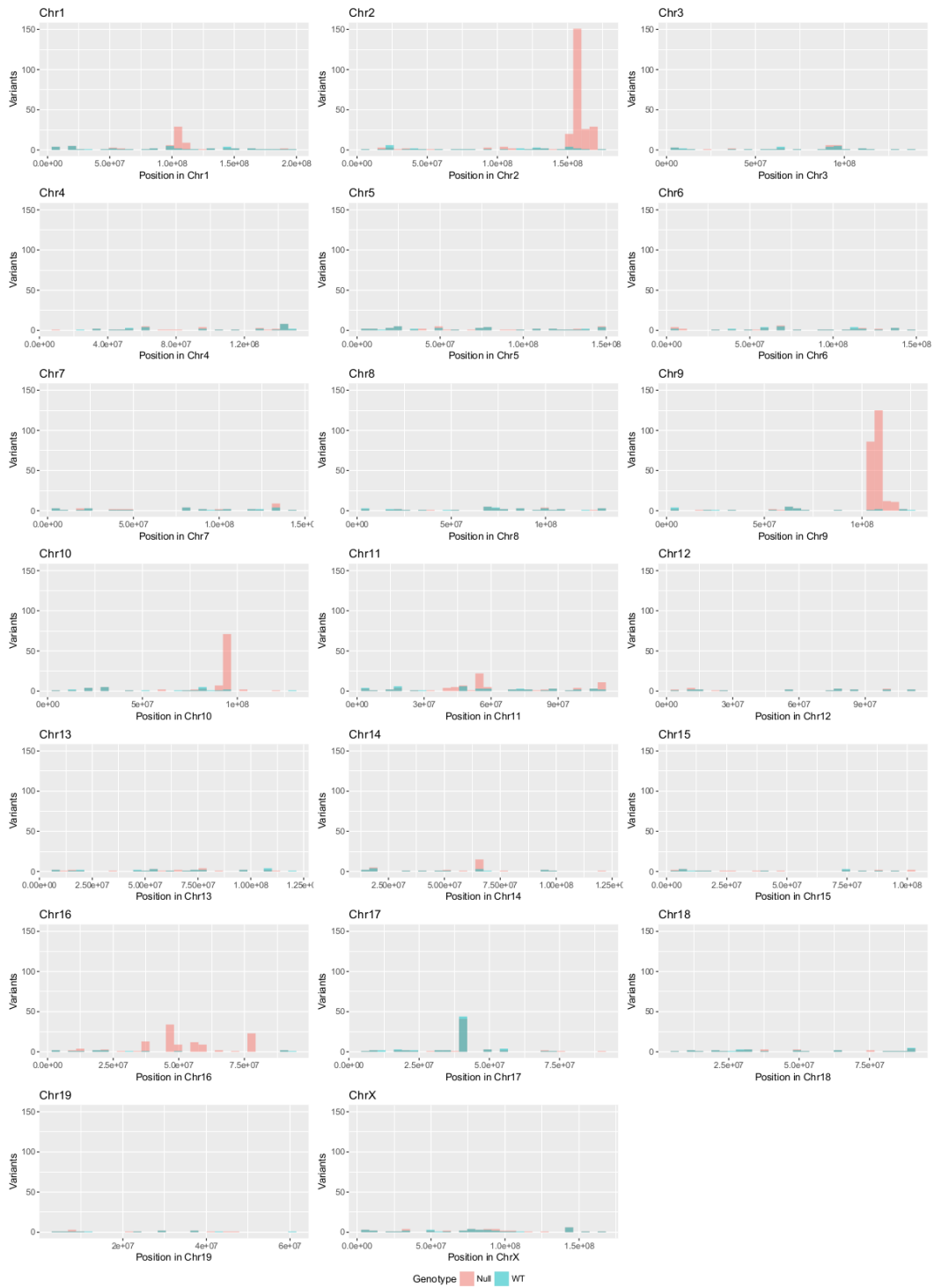


**Figura 14: Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Stc1*** Histograma de variantes nuevas y existentes por cromosoma desde el ARN de células MEF provenientes de ratones *Stc1* silvestres (WT) y KO. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes promedio de tres MEF de tipo *Stc1* silvestre y las barras rojas representan las variantes promedio de tres células MEF KO para *Stc1*. Fuente: Elaboración propia.

Al igual que en el caso de los modelos KO para *Mecp2* y *Gtf2ird1*, el modelo KO para *Stc1* es congénico. Este modelo presenta una retrocruza correcta con el genoma recipiente C57BL/6J y marcada huella congénica en el cromosoma 14, en donde se ubica el gen *Gtf2ird1* y donde ocurrió la recombinación homóloga (Chr14:69029289-69041401, <http://www.informatics.jax.org/marker/MGI:109131>).

El gráfico de genoma completo del modelo KO para *Itch*, considerando la totalidad de las variantes por genotipo, se detalla a continuación (véase figura 15):

# Itch KO



**Figura 15: Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Itch*** Histograma de variantes nuevas y existentes por cromosoma desde el ARN de la piel proveniente de ratones *Itch* silvestres (WT) y KO. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes promedio de 2 extracciones de piel de ratones *Itch* silvestre y las barras rojas representan las variantes promedio de 2 extracciones de piel de ratones KO para *Itch*. Fuente: Elaboración propia.

De este modelo se observa que, si bien existe retrocruza extensiva con C57BL/6J, se evidencian cuatro cromosomas con variantes ligadas al genotipo KO. El gen que codifica para *Itch* se encuentra en el cromosoma 2 (Chr2:155133509-155226855, <http://www.informatics.jax.org/marker/MGI:1202301>) y es uno de los cuatro cromosomas que presenta las variantes. La presencia de variantes ligadas al genotipo KO fuera del cromosoma donde se hizo la recombinación homóloga se puede deber a la mantención separada de ambos genotipos, eventos de contaminación cruzada entre cepas o la falta de retrocruza adecuada en la mantención de la mutación.

El gráfico de genoma completo del modelo KO para *Hnrnpd*, considerando la totalidad de las variantes por genotipo, se detalla a continuación (véase figura 16):

# Hnrnpd KO



**Figura 16: Detección de variantes nuevas y existentes en el experimento de secuenciación del modelo KO para *Hnrnpd* (*AUF-1*).** Histograma de variantes nuevas y existentes por cromosoma desde el ARN de células musculares satélite proveniente de ratones *Hnrnpd* silvestres (WT) y KO. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. Las barras azules representan variantes promedio de 3 extracciones de células de ratones *Hnrnpd* silvestre y las barras rojas representan las variantes promedio de 3 extracciones de células de ratones KO para *Hnrnpd*. Fuente: Elaboración propia.

De forma similar al modelo KO para *Sall2*, el modelo KO para *Hnrnpd* se mantiene en un fondo mixto, pero la distribución promedio de las variantes difiere en gran medida entre los genotipos silvestre y KO. Aunque la huella congénica no es evidente en el cromosoma 5, donde el gen *Hnrnpd* se ubica (Chr5:99955935-99978938, <http://www.informatics.jax.org/marker/MGI:101947>), la distribución de variantes es significativamente diferente a lo largo de los cromosomas, probablemente por falta de retrocruzamiento con el ratón recipiente C57BL/6J en las primeras generaciones. Si bien los gráficos ayudan a discernir de forma visual la distribución de variantes entre genotipos, se requiere de un método estadístico para detectar cromosomas con distribución diferencial de variantes entre genotipos. Para discernir de forma estadística si la distribución de variantes es significativamente diferente entre genotipos a lo largo de los cromosomas, el script anterior también implementa la prueba de Cochran-Armitage para la distribución de frecuencias. El resultado de la prueba en los modelos anteriormente estudiados se indica a continuación:

Prueba de Cochran-Armitage (valor p)						
	Gtf2ird1	Mecp2	Stc1	Itch	Sall2	Hnrnpd
Chr1	0.169	0.493	0.565	0.529	0.002	2.574e-18
Chr2	0.372	0.694	0.312	6.676e-15	0.001	3.553e-18
Chr3	0.868	0.421	0.464	0.822	0.338	0.001
Chr4	0.969	0.574	0.733	0.954	0.465	0.003
Chr5	1.141e-17	0.782	0.434	0.837	0.569	0.008
Chr6	0.805	0.975	0.893	0.460	0.325	0.338
Chr7	0.562	0.380	0.928	0.982	0.834	0.675
Chr8	0.382	0.293	0.595	0.604	4.007e-18	5.767e-14
Chr9	0.836	0.419	0.630	2.052e-16	0.109	3.416e-09
Chr10	0.561	0.604	0.621	1.270e-06	0.705	0.378
Chr11	0.880	0.381	0.950	0.168	0.876	9.013e-16
Chr12	0.594	0.701	0.753	0.588	0.555	0.739
Chr13	0.905	0.276	0.540	0.659	8.487e-05	7.6e-08
Chr14	0.608	0.520	2.789e-08	0.367	4.939e-60	1.084e-17
Chr15	0.799	0.881	0.952	0.702	0.942	0.653
Chr16	0.866	0.433	0.529	0.020	0.576	0.010
Chr17	0.740	0.730	0.688	0.341	0.312	0.197
Chr18	0.447	0.324	0.863	0.923	0.452	0.0002
Chr19	0.482	0.394	0.596	0.777	0.623	4.440e-20
ChrX	0.615	0.450	0.736	0.841	0.801	1

**Tabla 2. Prueba de Cochran-Armitage en los 6 experimentos de secuenciación analizados.** La prueba se implementó en el lenguaje de programación R (Genotype\_variants.R). Se considera estadísticamente significativo un valor p menor a 0.05, los cuales se denotaron en color rojo. Fuente: Elaboración propia.

Según la prueba y como se revisó en los gráficos, el modelo KO para *Gtf2ird1* muestra sólo un cromosoma con distribución diferencial de variantes, correspondiente al cromosoma con la huella congénica (cromosoma 5, valor P <0.0001). El modelo KO para *Mecp2* no presenta distribución diferencial de variantes en ningún cromosoma



según la prueba. Aun así, se esperan variantes ligadas al genotipo KO en el cromosoma X donde se localiza el gen *Mecp2*, debido a la naturaleza congénica de este modelo. Similar al modelo KO para *Gtf2ird1* y como se describió anteriormente en los gráficos, el modelo KO para el gen *Stc1* presenta una distribución diferencial de variantes sólo en el cromosoma congénico (cromosoma 14, valor  $P < 0.0001$ ). El modelo KO para el gen *Itch* presenta cuatro cromosomas con variantes ligadas al genotipo KO tal como se ilustra en el gráfico correspondiente (cromosoma 2, cromosoma 9, cromosoma 10 y cromosoma 16 con un valor de  $p < 0,0001$  para los primeros tres y  $P < 0,02$  para el último, respectivamente). El modelo KO para *Sall2* presenta cinco cromosomas con distribución diferencial de variantes, de los cuales el cromosoma 14 contendría la huella congénica y presenta el valor  $p$  más bajo, un fenómeno propio de cepas con fondo genético mixto (véase introducción). Finalmente, al igual que los ratones KO para *Sall2*, el modelo KO para *Hnrnpd* se mantiene en un fondo mixto pero la distribución de variantes es significativamente diferente en 12 cromosomas. En resumen, el script propuesto permite reconocer de forma gráfica la introgresión de variantes de la célula madre al genoma, incluyendo la detección estadística de cromosomas con distribución diferencial de variantes entre genotipos.

Revisamos la naturaleza de las variantes introgresadas en los modelos mediante la clasificación de su potencial efecto con la herramienta “variant effect predictor” (VEP, <https://www.ensembl.org/info/docs/tools/vep/index.html>). Centrándonos en el análisis de los archivos VCF del genotipo KO de cada experimento, encontramos que tanto el número como la proporción de variantes nuevas versus existentes varían entre los modelos KO, y las variantes nuevas (fijadas por la deriva génica) pueden representar más del 50% de las variantes totales, tal como en el caso de los modelos KO para *Mecp2* y *Gtf2ird1* (véase figura 17A). También notamos que el número de variantes de tipo “missense” (del inglés, con cambio de sentido) y “frameshift” (del inglés, interrupción del marco de lectura)

aumentan a medida que aumenta el número de variantes nuevas (véase figura 17B), existiendo una correlación positiva entre estas variables (p-Value = 0.0167, correlación de Spearman). Se ha descrito que es importante detectar las variantes nuevas fijadas por la deriva génica, ya que éstas pueden imponer nuevos fondos genéticos para la mutación en estudio. Este último problema, ha sido estudiado en colonias murinas por (Drake et al. 1998), estimando que se produce una tasa de  $1.1 \times 10^5$  mutaciones por generación, asumiendo  $10^3$  pares de bases por locus. En base a lo anterior, se puede estimar que habiendo aproximadamente 25000 genes en el genoma de ratón la tasa de mutaciones por generación se puede calcular de la siguiente forma:

$$\frac{\text{Número de mutaciones por generación}}{\text{Número de genes}} = \frac{1.1 \times 10^5}{25000} = 4.4$$

$$\left( \frac{\text{Número de mutaciones por generación}}{\text{Número de genes}} \right) = \frac{1.1 \times 10^5}{25000} = 4.4$$

$$\frac{4.4}{2} = 2.2$$

(1)

*Asumiendo genoma diploide*

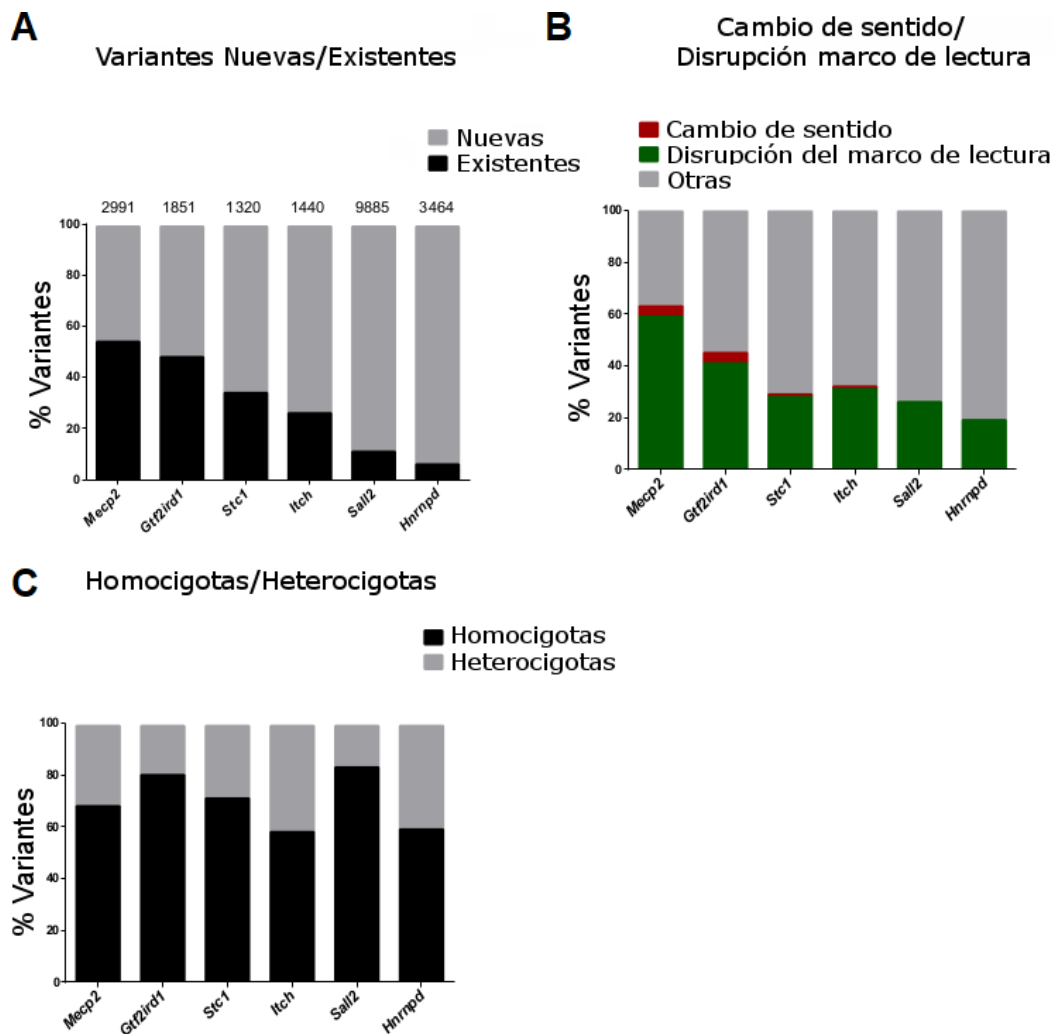
$$\frac{4.4}{2} = 2.2$$

(2)

Se desprende de (1) y (2) que una mutación puede fijarse en la colonia aproximadamente cada 2 generaciones, lo cual explicaría que la mantención de colonias por varias generaciones puede generar un fondo genético indeterminado para la mutación en estudio. Este último punto destaca la importancia en estudios con animales KO de no sólo detectar las variantes introgresadas de los genomas derivados de las cepas 129, sino que también las variantes nuevas.

Por otro lado, la relación de las variantes homocigotas versus heterocigotas entre los modelos KO es variable. Es el caso del modelo KO para *Hnrnpd/AUF-1* en

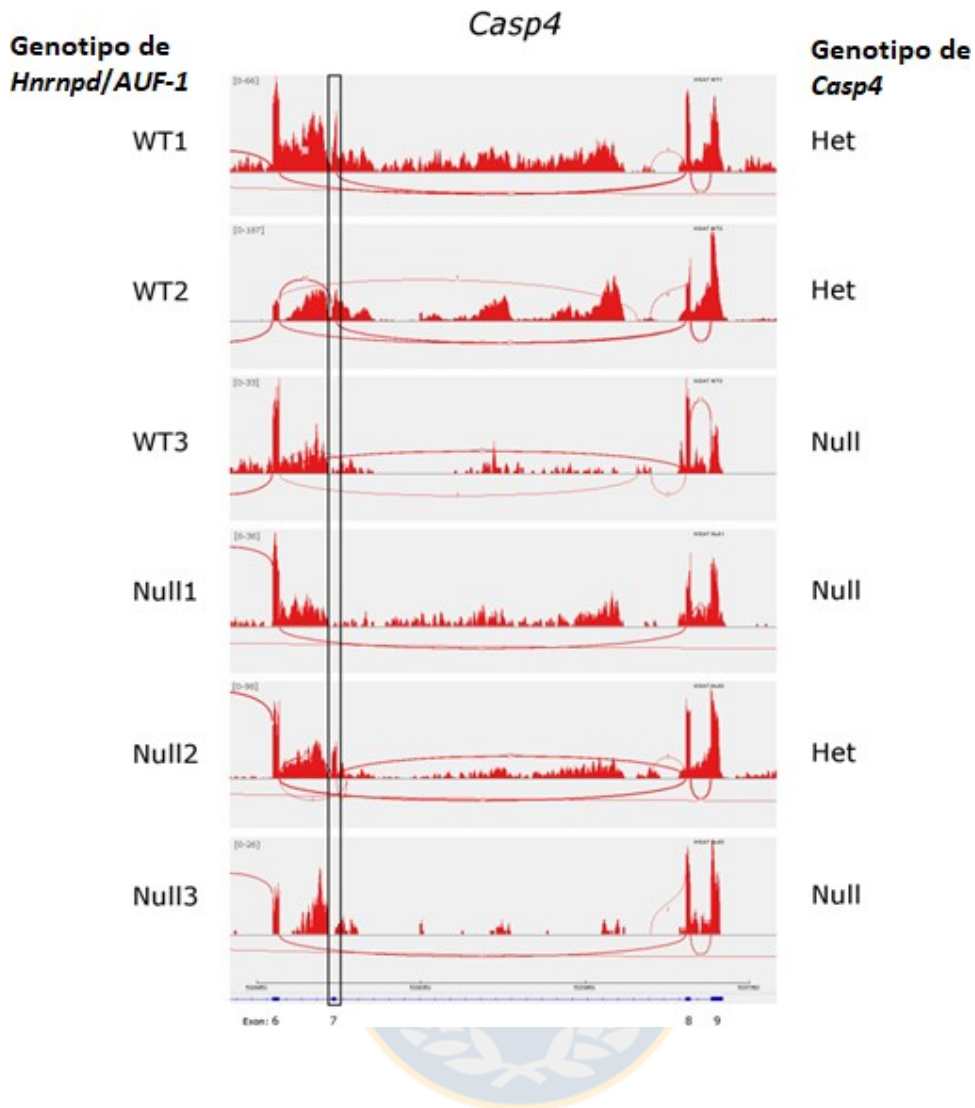
donde las variantes heterocigotas están casi ausentes en este modelo (véase figura 17C).



**Figura 17: Detección y distribución de variantes a nivel de genoma completo en los modelos KO analizados.** A) Gráfico de barras entrelazadas que muestra el porcentaje de variantes nuevas (barras negras) y de existentes (barras grises) caracterizadas por el algoritmo variant effect predictor (VEP) en cada modelo KO. El número total de variantes se denota arriba de cada barra. B) Porcentaje de variantes de cambio de marco de lectura (en rojo), variantes con cambio de sentido (en verde) y otras variantes (en gris) caracterizadas en cada modelo KO. C) La relación entre las variantes homocigotas (barras negras) y heterocigotas (barras grises) expresadas como porcentajes en cada modelo KO. Fuente: Elaboración propia.

Dado que casi la totalidad de las variantes en el modelo KO para *Hnrnpd* son existentes y derivadas de la célula madre (ESC), se espera encontrar mutaciones pasajeras inactivantes en las réplicas biológicas. Como ejemplo de lo anterior, revisamos las variantes del gen codificante para la caspasa 4 (*Casp4*) a través de las réplicas secuenciadas de este modelo. Este gen se encuentra ubicado en el cromosoma 9 y ha sido previamente reportado como inactivo en ratones derivados de las cepas 129 (Auerbach et al. 2000). La llamada de variantes en cada réplica biológica del experimento con el modelo KO para *Hnrnpd* reveló el genotipo del gen *Casp4*, evidenciando la mutación inactivante de *Casp4* en una muestra secuenciada de tipo *Hnrnpd* silvestre y en dos muestras KO (véase figura 18). También confirmamos esta observación por la falta de expresión del exón 7 del (ARNm) *Casp4*, un fenómeno anteriormente reportado en las cepas 129 (Kayagaki et al. 2011). Considerando todo lo anterior, además de las variantes ligadas al genotipo KO, el empleo de fondos mixtos para la mantención de mutaciones puede tener una profunda influencia en la expresión génica o en los fenotipos asociados con el modelo en estudio, como se revisó anteriormente en (Eisener-Dorman et al. 2009).

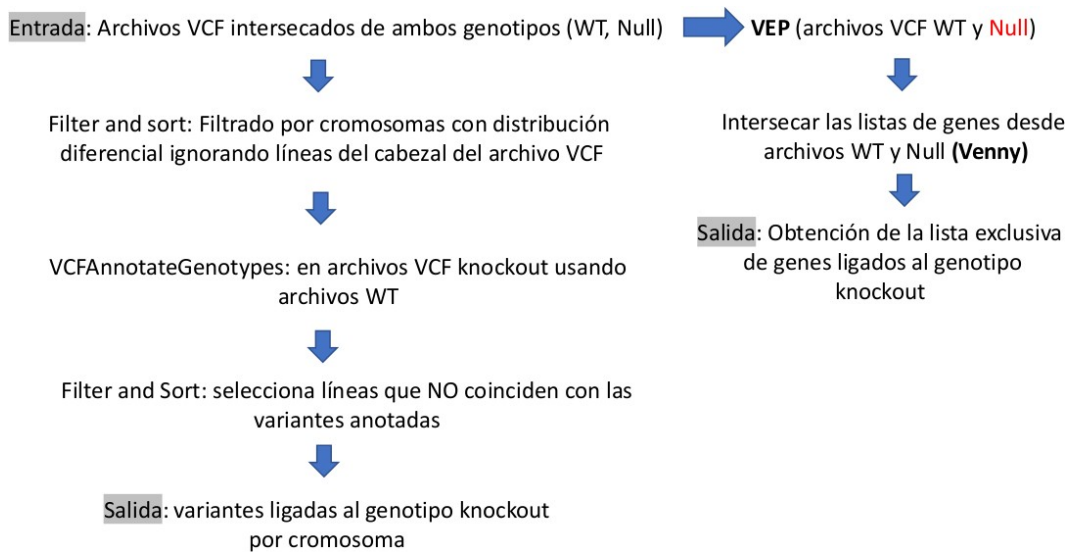
Por otro lado, se demostró que la deriva génica puede fijar variantes con cambio de sentido o disrupción del marco de lectura, independiente de la introgresión del genoma de las células madre derivadas de cepas 129. Dadas estas razones, interferencia genética es esperada en el modelo KO para *Sall2*; lo cual sesgaría el transcriptoma dependiente de *Sall2* debido a la introgresión del genoma de 129 en la cepa receptora C57BL/6J y la existencia de variantes ligadas al genotipo KO. Este último problema fue abordado en primera instancia detectando las variantes ligadas al genotipo KO con los genes correspondientes, lo que correspondería a la huella congénica. La detección de la huella congénica en cada KO se detalla en la siguiente sección.



**Figura 18: Detección de la mutación inactivante del gen *Casp4* a través de las réplicas biológicas del estudio de secuenciación del modelo KO para *Hnrnpd*.** Se muestran gráficos de Sashimi (<http://miso.readthedocs.io/en/fastmiso/sashimi.html>) originados a partir de tres réplicas biológicas de muestras de ARN de tipo silvestre y KO para el gen *Hnrnpd*. La expresión por base nitrogenada se muestra en el eje Y y los gráficos de Sashimi con las correspondientes coordenadas genómicas en el eje X. La estructura del gen se representa en la parte inferior en azul, obtenida del servidor UCSC (<https://genome.ucsc.edu/>). Obtuvimos los genotipos del gen *Casp4* de cada réplica con el llamador de variantes Freebayes y se adjudicó genotipo en base a la presencia de al menos un polimorfismo de un solo nucleótido (SNP). Destacamos en un rectángulo negro la expresión del exón 7, para denotar la presencia o ausencia de expresión en las muestras KO para el gen *Casp4*. Fuente: Elaboración propia.

### **Disección de variantes ligadas al genotipo KO: la huella congénica**

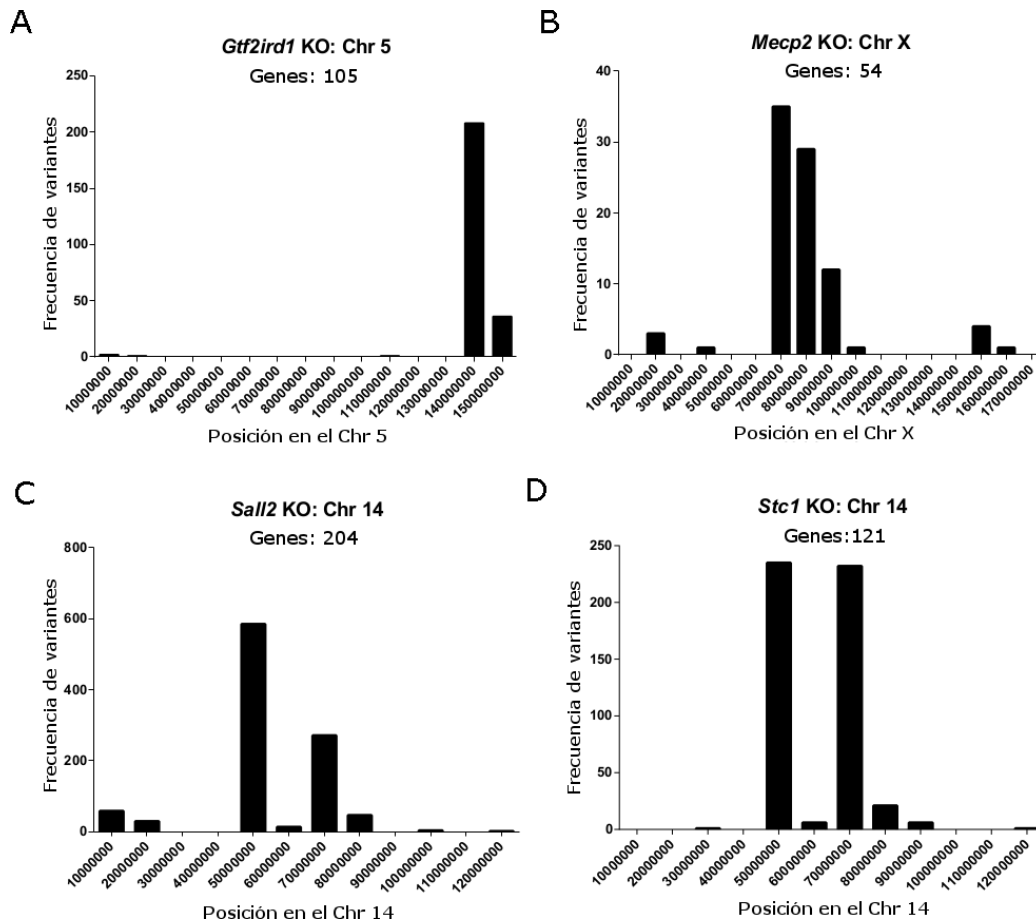
Dado que la existencia de variantes ligadas al genotipo KO conduce a una comparación errónea entre el genotipo silvestre y la contraparte KO, es importante detectar este sesgo. Implementamos un protocolo simple para estudiar la distribución y extensión de variantes ligadas al genotipo KO, anteriormente mencionado como la huella congénica (véase figura 19). Para recolectar las variantes ligadas al genotipo KO, ingresamos como entrada los archivos VCF intersecados de ambos genotipos en los modelos KO para *Gtf2ird1*, *Mecp2*, *Stc1* e *Itch*. También consideramos el modelo KO para *Sall2* debido a la presencia de variantes ligadas al genotipo KO en el cromosoma 14. Filtramos los cromosomas con una distribución de variantes significativa entre los genotipos mediante la prueba de Cochran-Armitage, y se realizó la anotación de genotipos en archivos VCF KO mediante el uso de archivos VCF de tipo silvestre. Después de seleccionar las líneas de los archivos VCF sin anotación, se obtuvo un archivo de salida con variante ligadas al genotipo KO por cromosoma. Luego se aplicó el algoritmo “variant effect predictor” (VEP) que determinó el efecto de estas variantes.



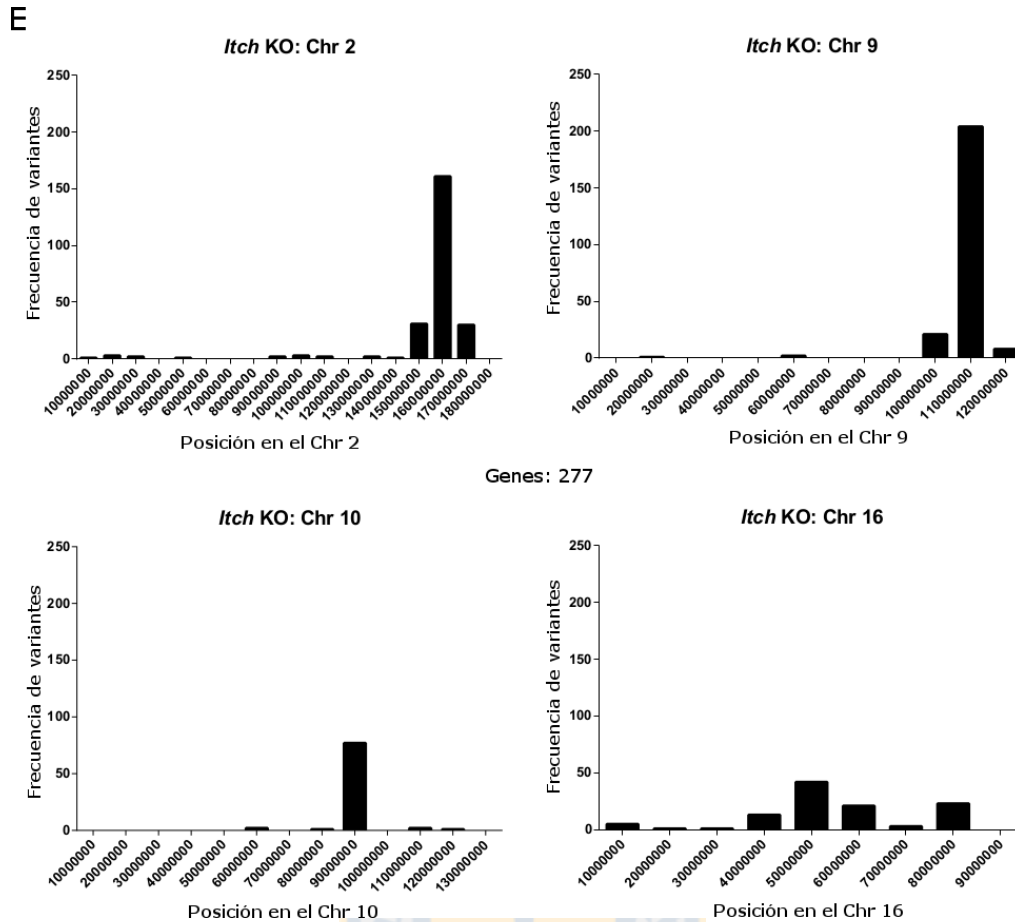
**Figura 19: Protocolo computacional para la detección de variantes ligadas al genotipo KO.** Implementación computacional para la detección de variantes ligadas al genotipo KO, basada en archivos VCF intersecados por genotipos como entrada. Ignoramos las líneas de encabezado del archivo VCF (programa “filter and sort”, omitiendo líneas de encabezado). Dado que el programa Freebayes genera 130 líneas de encabezado, en nuestros estudios se omitieron dichas líneas de cada archivo VCF intersecado. Realizamos la anotación de variantes con el programa VCFAnnotateGenotypes en archivos VCF KO utilizando la contraparte de tipo silvestre (WT). Después de seleccionar las líneas que no coinciden con las variantes anotadas, obtuvimos un archivo de salida que contenía las variantes ligadas al genotipo KO. Para obtener los genes ligados al genotipo KO, evaluamos la anotación adicional de variantes con el programa de predicción de efecto “variant effect predictor” (VEP) de Ensembl. Fuente: Elaboración propia.

Como era esperado de acuerdo con los gráficos de genoma completo, detectamos un solo cromosoma con variantes ligadas al genotipo KO en los modelos KO para *Gtf2ird1*, *Mecp2*, *Sall2* y *Stc1*, cada uno con un diverso número de variantes (véase las figuras 20A, B, C y D, respectivamente). Además, se puede observar que la huella congénica es distinta en extensión dependiendo del modelo KO. La distribución

de variantes en los modelos KO para *Stc1* y *Sall2* son similares porque ambos genes pertenecen al cromosoma 14 (véase las figuras 20C y 18D respectivamente). Como se discutió anteriormente en los gráficos de genoma completo, el modelo KO para *Itch* exhibe huella congénica en cuatro cromosomas, incluido el cromosoma 2, donde se realizó la recombinación homóloga (Chr 2, véase figura 20E). Adicionalmente notamos en este modelo que las variantes ligadas al genotipo KO fuera del cromosoma 2 son heterocigotas.







**Figura 20: Disección de variantes ligadas al genotipo KO: la huella congénica.** A) Histograma de las variantes ligadas al genotipo KO en el modelo KO para *Gtf2ird1*. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y luego graficamos las variantes en histogramas de frecuencia de acuerdo con estas posiciones. Indicamos en el histograma el número de genes congénicos que abarcan estas variantes. B) El mismo histograma que en (A) en el modelo KO para *Mecp2*, ubicado en el cromosoma X. C) El mismo histograma que en (A) en el modelo KO para *Sall2*, ubicado en el cromosoma 14. D) El mismo histograma que en (A) en el modelo KO para *Stc1*, ubicado en el cromosoma 14 E) Los mismos histogramas que en (A) en el modelo KO para *Itch*, ubicado en el cromosoma 2. Este último modelo KO presenta cuatro cromosomas con variantes ligadas al genotipo KO, incluido el cromosoma congénico (Chr 2). Se Indica en el histograma el número de genes congénicos que abarcan estas variantes. Fuente: Elaboración propia.

La clasificación de las variantes mediante la herramienta “variant effect predictor” (VEP) revela los genes congénicos en cada modelo KO. A modo de ejemplo de los genes identificados, se muestra la lista de los genes congénicos en el modelo KO para *Sall2* (ver Tabla 3).

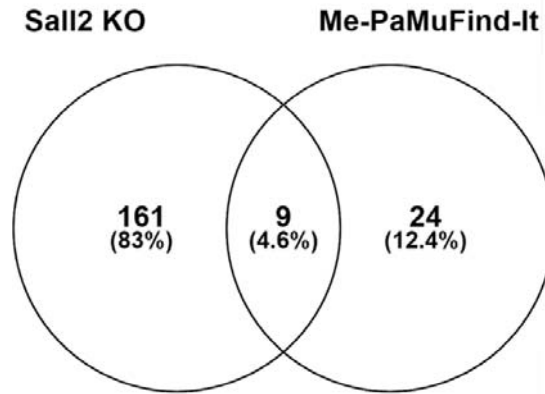


Ptprg	Cdkn3	Gm38316	Haus4	Slc25a37	Lpar6
AC102542.1	Gmfb	Pnp2	Ajuba	Loxl2	Itm2b
3830406C13Rik	Cgrrf1	Rnase4	Gm24705	Gm27177	Med4
AC154682.1	Samd4	Ang	Psemb5	R3hcc1	Nudt15
Thoc7	Gm10371	Mettl17	Mir686	Mir6950	Sucla2
Gm11100	Wdhd1	Slc39a2	Cdh24	Chmp7	Htr2a
Atxn7	Gm24378	Ndrp2	Psemb11	4930480K23Rik	Esd
Psmc6	Socs4	Arhgef40	Gm20726	Tnfrsf10b	Lrch1
Ii3ra	Mapk1ip1l	Gm16617	Acin1	AC162936.1	Lcp1
CT025689.6	Lgals3	Zfp219	Gm17606	Rhobtb2	Gm15629
Slc4a7	Dlgap5	Tmem253	Mir6948	Bin3	Zc3h13
Nek10	Fbxo34	Gm22354	4930579G18Rik	Ccar2	Cog3
Oxsm	Atg14	Hnrnp1	1700123O20Rik	Pdlim2	Gm24876
AC173482.6	Ktn1	Supt16	CT009512.1	9930012K11Rik	Gtf2f2
Ngly1	Gm37874	Rpgrip1	Bcl2l2	Sorbs3	Gpalpp1
AC173482.1	Tmem260	Chd8	Gm20521	Slc39a14	Tsc22d1
Top2b	Exoc5	Gm26590	Pabpn1	Gm44340	Lacc1
Rarb	Ap5m1	Snord8	Slc22a17	Polr3d	Ccdc122
Nr1d2	Naa30	Rab2b	Efs	Mir320	Akap11
Gm44469	Ttc5	Gm43766	Ngdn	Bmp1	Vwa8
Nkiras1	Gm10916	Tox4	Sacs	Reep4	
Rpl15	Parp2	Mettl3	Spryd7	Hr	
Ube2e1	Rpph1	Sall2	Wdfy2	Fam160b2	
Ube2e2	Tep1	Dad1	Msra	Nudt18	
Ddhd1	Gm26782	Abhd4	Extl3	Xpo7	
Mir5131	Osgep	Oxa1l	Elp3	Fndc3a	
Bmp4	Gm24689	Slc7a7	Ptk2b	Gm16549	
Gm15217	Apex1	Mmp14	Stmn4	Rcbtb2	
Gm15222	Tmem55b	Lrp10	Bnip3l	Rb1	
Cnih1	Pnp	Prmt5	Dock5	Gm17233	

**Tabla 3. Lista de los genes congénicos presentes en el modelo KO para *Sall2*.**

Los genes están ordenados por filas en base a las coordenadas del cromosoma 14 y abarcan más de 60 megabases del cromosoma 14. El gen *Sall2* se indica en el casillero amarillo. Fuente: Elaboración propia.

Se contrastaron estos resultados con la herramienta bioinformática Me-PaMuFind-It, disponible para estimar el número y posible efecto de las mutaciones pasajeras en genes congénicos de modelos KO (Vanden Berghe et al. 2015) (<http://bioit2.irc.ugent.be/prx/me-pamufind-it/>). La herramienta predijo menos del 5% de los genes congénicos en el modelo KO para *Sall2* (véase figura 21) confirmándose que el número de los genes congénicos en los modelos KO es variable y puede depender del número de retrocruzas con el ratón recipiente C57BL/6J; del esquema de cruzas empleado y el tipo de célula madre utilizada. En conclusión, este protocolo entrega un método de detección fiable de variantes ligadas al genotipo KO y confirman la variabilidad en el número de genes ligados al genotipo KO para cada modelo, poco probable de estimar *a priori*.



**Figura 21: Comparación de los genes congénicos detectados en el modelo KO para *Sall2* frente a la predicción de genes ligados al genotipo KO con el método Me-PaMuFind-it.** Diagrama de Venn que muestra el número de genes congénicos detectados por el protocolo computacional llevado a cabo como se menciona en la figura 17 frente a la predicción de la herramienta bioinformática Me-PaMuFind-it, implementada en Vanden Berghe T y col., Immunity 2015. Fuente: Elaboración propia.

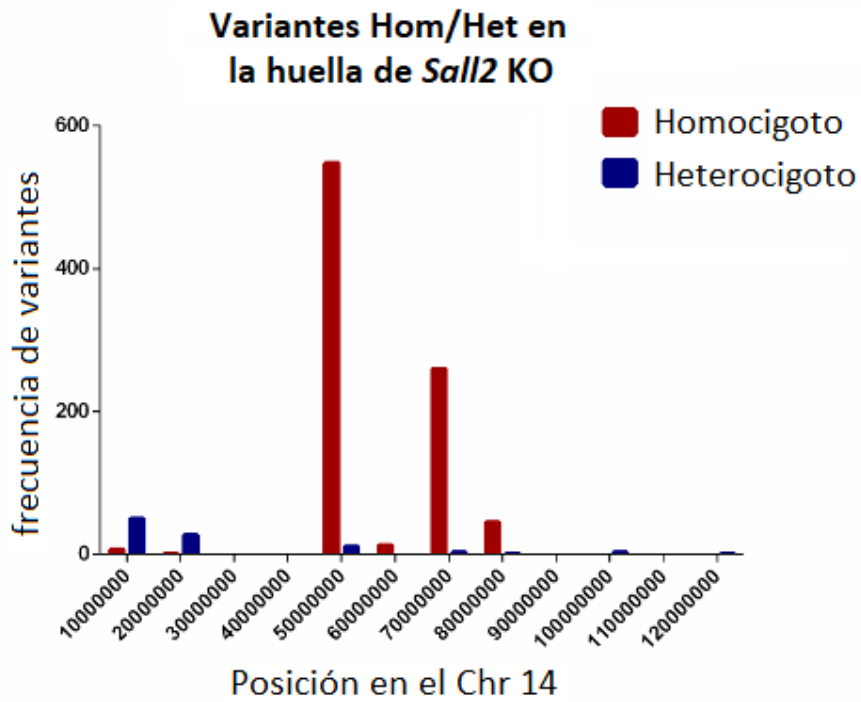
### **Ploidía de la huella congénica**

Razonamos que la herencia de las variantes ligadas al genotipo KO puede variar en ploidía entre las réplicas biológicas, sin asumir la herencia homocigótica de la huella congénica en todos los casos. Por esta razón, se abordó estudiar la ploidía de la huella congénica en todos los casos. Por esta razón, se abordó estudiar la ploidía de la huella congénica en los experimentos con los modelos KO analizados. Debido a razones comparativas, nos centramos en los modelos KO para los genes *Sall2* y *Stc1*, ambos genes ubicados en el cromosoma 14. En el modelo KO para *Sall2* se predijeron aproximadamente 968 variantes con un 90% de homocigosidad ligada al genotipo KO. Estas variantes se distribuyen de forma centrada en el cromosoma 14 (véase figura 22A). Para estudiar la herencia de estas variantes y caracterizar la ploidía de esta huella congénica, utilizamos las variantes del embrión *Sall2* heterocigoto que fue secuenciado en nuestro estudio. Razonamos que, si la herencia de la huella congénica es mendeliana, las variantes heterocigotas en este embrión contendrán las variantes

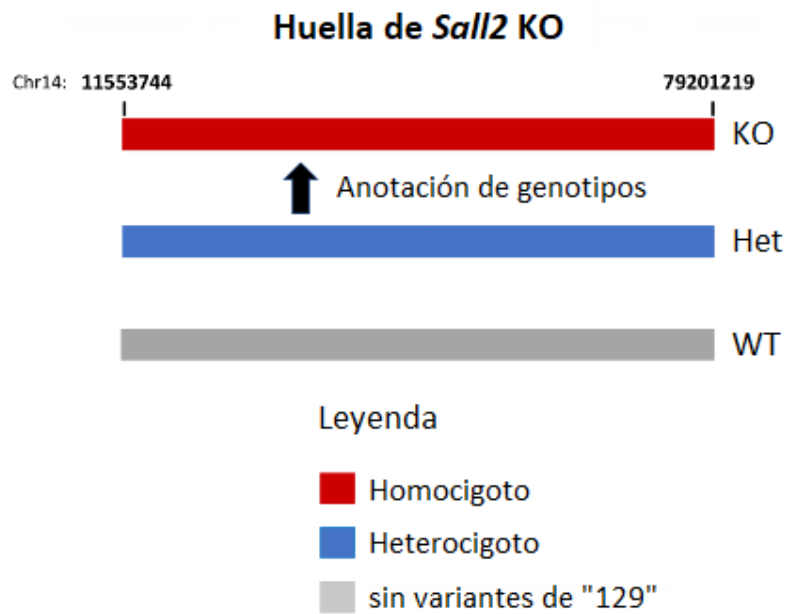
ligadas al genotipo KO (véase la figura 22B). Encontramos 169 variantes homocigotas y 1898 heterocigóticas ligadas al embrión heterocigoto para *Sall2*. Las variantes heterocigotas anteriormente encontradas también despliegan distribución orientada hacia el centro del cromosoma 14 (véase la figura. 22C). Usamos estas variantes heterocigóticas para anotar variantes en las 968 variantes detectadas de forma previa que están ligadas al genotipo KO. 873 variantes fueron anotadas utilizando este procedimiento (véase figura 22D), lo cual abarca casi el 95% de los genes ligados al genotipo KO anteriormente identificados (véase figura 22E). Por tanto, la herencia de las variantes ligadas al genotipo KO para *Sall2* son homocigóticas en su mayoría.



A

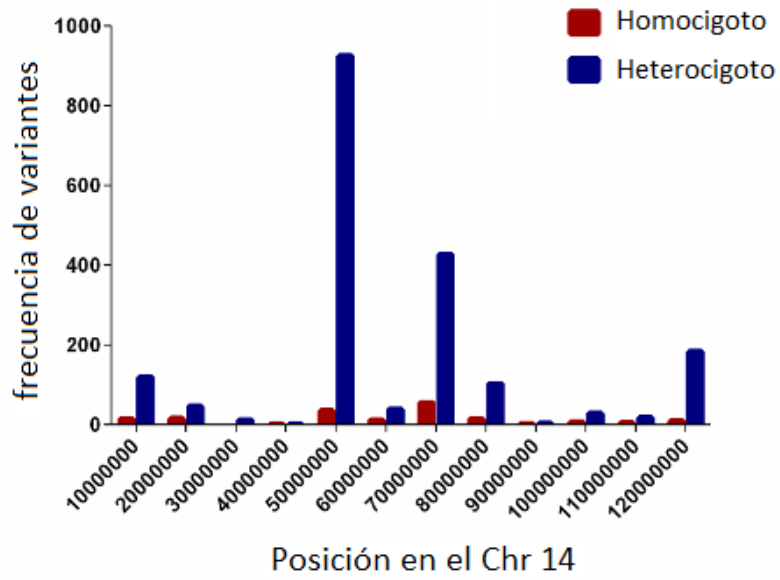


B



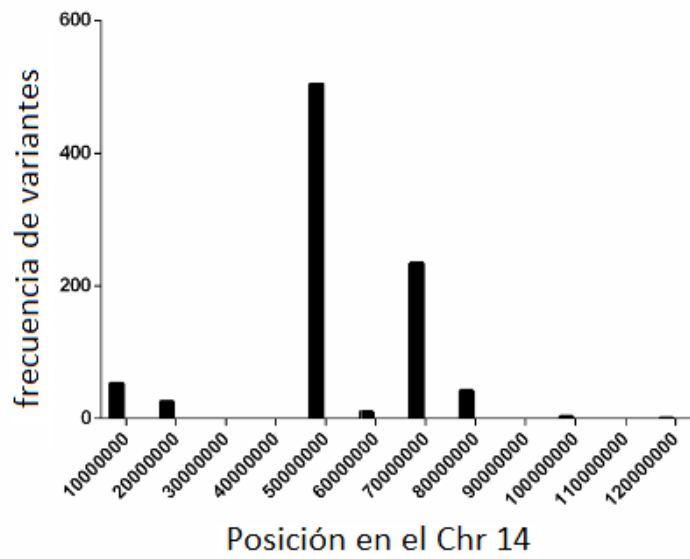
C

### Variantes Hom/Het en embrión *Sall2* Heterocigoto



D

### Anotación de genotipos en Null

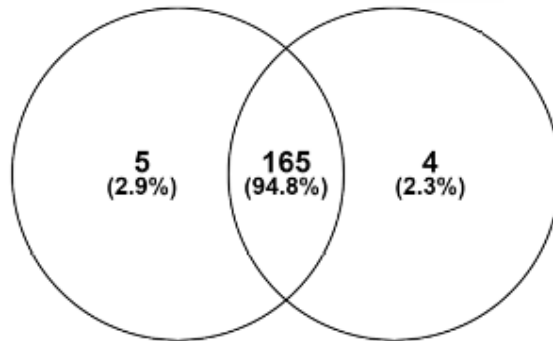




E

Substracción desde WT

Anotación desde  
embrión *Sall2* +/-

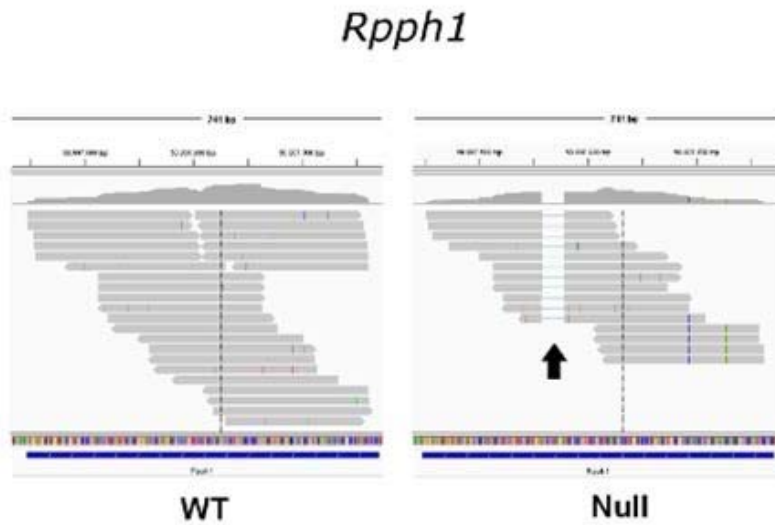


**Figura 22. Ploidía de la huella congénica en el modelo KO para *Sall2*.** A) Histograma de las variantes ligadas al genotipo KO que distingue variantes con ploidía homocigota (Hom, representada en barras rojas) y variantes con ploidía heterocigota (Het, representada con barras azules). Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes con cada ploidía como histogramas de frecuencia de acuerdo con estas posiciones. B) Esquema de la huella congénica de cada embrión KO para *Sall2*, ubicada en el cromosoma 14. Representamos cada ploidía con un color distintivo (segmento homocigótico en rojo y segmento heterocigótico en azul), marcando posiciones cromosómicas por encima de las barras. Se realizó una anotación genotípica de las variantes ligadas al genotipo KO utilizando las variantes heterocigóticas de un embrión heterocigoto para *Sall2*. C) Histograma de variantes homocigotas (representadas en barras rojas) y variantes heterocigóticas (representadas en barras azules) del embrión *Sall2* heterocigótico en el cromosoma 14. Las variantes que ocurren en el embrión de tipo silvestre se sustrajeron previamente de estas variantes. Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes con cada ploidía como histogramas de frecuencia de acuerdo con estas posiciones. D) Histograma de variantes anotadas de (A) con variantes heterocigóticas de (B). Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada genotipo como histogramas de frecuencia de acuerdo con estas posiciones. E) Diagrama de Venn que compara la identificación previa de los genes congénicos en el modelo KO para *Sall2* (véase la Tabla 3) con las variantes anotadas de (C), mostrando la superposición. El diagrama muestra los números de genes con los porcentajes correspondientes. Fuente: Elaboración propia.

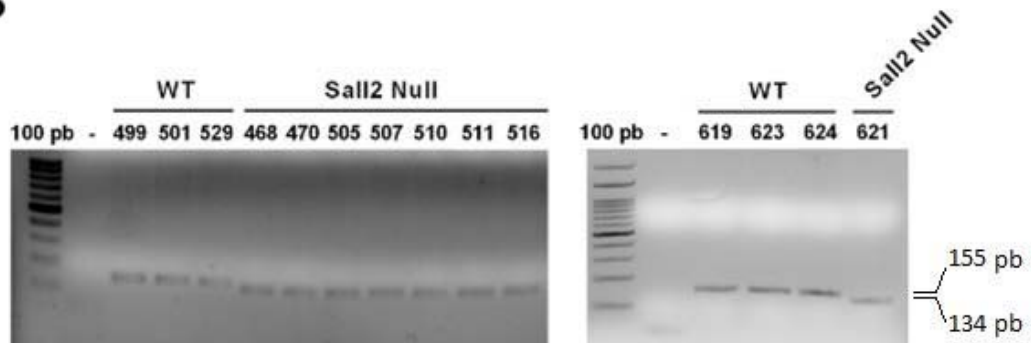
Para comprobar experimentalmente la introgresión del genoma de 129P2OlaHsd en la huella congénica del modelo KO para *Sall2*, notamos una delección de 21 pares de bases en el gen congénico *Rpph1*, útil como marcador de la huella congénica en el modelo KO para *Sall2* (Chr14: 50807449-50807767, véase la figura 23). Como se esperaba, la PCR desde el ADN genómico de ratones silvestres y KO para *Sall2* mostró la presencia de esta delección en varias generaciones de ratones KO (véase figura 23B). La secuenciación de Sanger de todo el gen congénico *Rpph1* en el

último retrocruzamiento realizado, confirmó que el origen del gen *Rpph1* proviene de 129P2OlaHsd, según el proyecto “Mouse Genome Project” (véase figura 23C).

**A**



**B**



C

```

Human      GTCCCTGCGGGGTACCTCACCTCAGCCATTGAACTCACTTCGCTGGCCGTGAGTCTGTT
Seq_Sall2_Null
Mouse     G-----CATTGAACTCGCACTGCTGGCCGTGAGTCTGTT
          G---CCCCGGGAGGTGCCTCACCTCAGCCATTGAACTCGCACTGCTGGCCGTGAGTCTGTT
          ***** * *****
          268                               236

          ↓

Human      CCAAGCTCCGGCAAAGGAGGCATCCGCCGGGCCCTCCCCGAA--GGGCGGGGTCCACGG
Seq_Sall2_Null
Mouse     CCAAG-----CTCCGGCGAGGGAGGCATCCGCCCGCACTCGGGG
          CCAAGCTCCGGCGAGGGAGGCATCCGCCCGCACTCGGGG
          ***** * * * * * * * * * *
          230                               208                               175

Human      CATCTCCTGCCAGTCTGACCTCGCGGGAGCCCCGTTCTCTGGGAACTCACCTCCCCGA
Seq_Sall2_Null
Mouse     CTTCTCCTGCCAGTCTGCCAAGCGTAGAGCCCTGCTCTCTGGGAACTCACCTCCCCG-
          CTTCTCCTGCCAGTCTGCCAAGCGTAGAGCCCTGCTCTCTGGGAACTCACCTCCCCG-
          * ***** ** * * * * * * * * * *
          174                               116

Human      AGCTCAGGGAGAGCCCTGTTAGGGCCGCTCTGGCCCTAGTCTCAGACCTCCCAAGGGA
Seq_Sall2_Null
Mouse     --CTCGGGGAGAGCCCGGTTAGGGCCGCGGGAGGCCAGTCTCAGACCTCCCAAGGGG
          --CTCGGGGAGAGCCCGGTTAGGGCCGCGGGAGGCCAAGTCTCAGACCTCCCAAGGGG
          *** ***** ***** * * * * * * * * * *
          115                               58

Human      CATGGGAGTGGAGTGACAGGACGCACTCAGCTCGTGGCCCACTGATGAGCTTCCCTCCG
Seq_Sall2_Null
Mouse     CGAAGAGT-----GACGAGCACTCAGCACGTGGCCCGCTGATGAGCTTCCCTCCG
          CGAAGAGT-----GACACGCACTCAGCACGTGGCCCGCTGATGAGCTTCCCTCCG
          * * * * * * * * * * * * * * * * * * * * * *
          57                               7

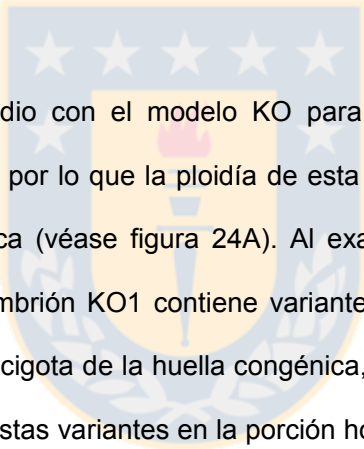
          ← RPPH1
          RNA

Human      CCCTATGGGAAAAGTGGTCTCATACAGAACTATAAGATTCCCAAATCCAAAGACATTT
Seq_Sall2_Null
Mouse     CCACCTTTAGGGTGTAG---ACCGGCCGCACTATAAGG-CTCGAAAGAGGAATAAATTT
          CCACCTTTAGGATGTAG---ACCGGCCGCACTATAAGG-CTCGAAAGAGGAATAAATTT
          ** + * * * * * * * * * * * * * * * *
          1

Human      CACGTTTATGGTGATTTCCAGAACACATAGCGACATGCAAAATATTGCAAGG-CGCCACT
Seq_Sall2_Null
Mouse     TACGTTTATGGGTGATTTCCACAAAGCACAGCGCGTAATTTGCATGCTCTATCCCAAG
          TACGTTTATGGGTGATTTCCACAAAGCACAGCGCGTAATTTGCATGCTCTATCCCAAG
          ***** ** * * * * * * * * * * * * * *
          +1
          →

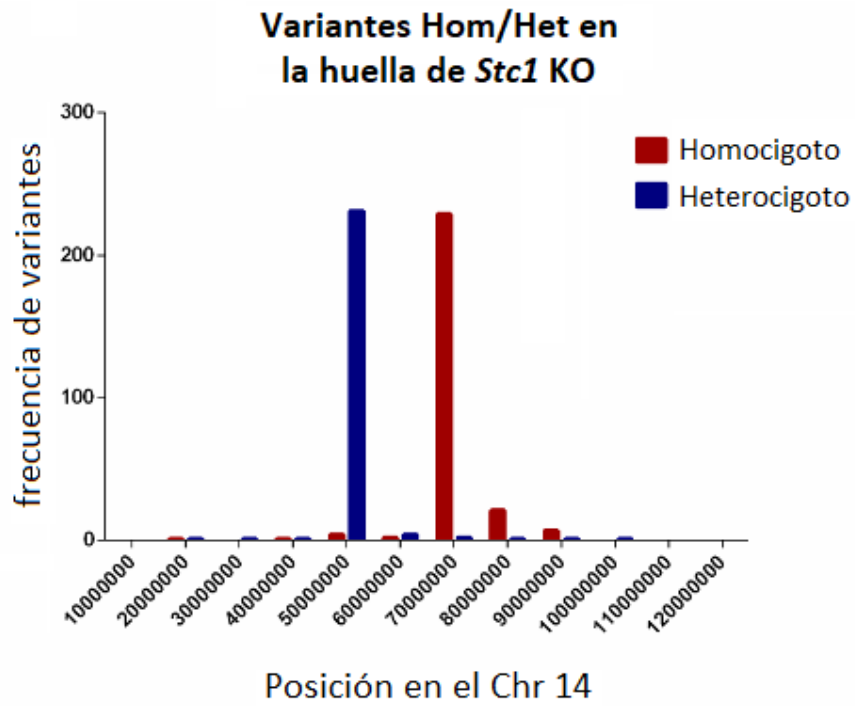
Human      CCCCTGTCCCTCACAGCCATCTTCTGCCAGGGCGCACGC GCGCTGGGTGTTCCCGCCTA
Seq_Sall2_Null
Mouse     CTCCTGTCCAGACAAGAAGCCC-----GCGCATCCGGGCA-----
          CTCCTGTCCAGACAAGAAGCCC-----GCGCATCCGGGCA-----
          * ***** * * * * * * * * * * * * * *
  
```

**Figura 23: Confirmación genómica de las variantes de ratones KO para *Sall2* mediante el uso del gen congénico *Rpph1* como marcador.** A) Imagen de la herramienta de visualización IGV (del inglés Integrative Genomics Viewer) del gen *Rpph1* de tipo silvestre y congénico. La flecha negra representa la delección genómica utilizada como marcador del origen 129P2OlaHsd en los experimentos de PCR. B) PCR desde ADN genómico usando partidores que amplifican la región del gen *Rpph1* que contiene la delección genómica de (A). Se genotiparon seis embriones de tipo silvestre y ocho embriones KO para *Sall2* en tres generaciones sucesivas. C) La secuenciación de Sanger de todo el gen *Rpph1* congénico del embrión 621 (*Sall2* KO). La secuencia anterior se alineó con el gen *RPPH1* humano (humano) y la secuencia de referencia del gen *Rpph1* proveniente de C57BL/6J (ratón). Destacamos en amarillo los polimorfismos entre la secuenciación y el genoma de referencia del ratón proveniente de C57BL/6J. Fuente: Elaboración propia.

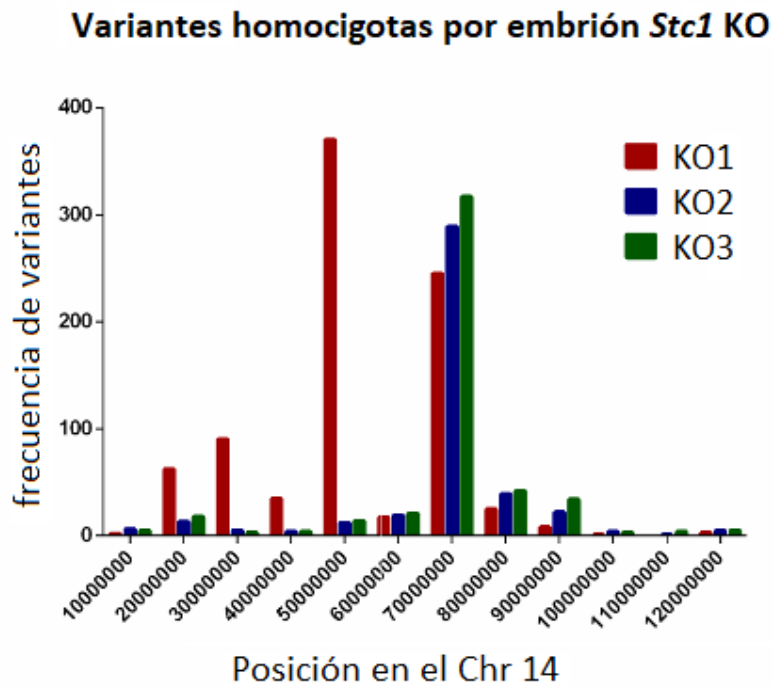


En el caso del estudio con el modelo KO para *Stc1*, casi la mitad de las variantes son heterocigotas, por lo que la ploidía de esta huella tiene una distribución heterocigótica y homocigótica (véase figura 24A). Al examinar la distribución de las variantes por embrión, el embrión KO1 contiene variantes homocigóticas tanto en la porción homocigota y heterocigota de la huella congénica, mientras que los embriones KO2 y KO3 solo contienen estas variantes en la porción homocigótica (figura 24B). Por el contrario, los embriones KO2 y KO3 contienen variantes heterocigóticas en la porción heterocigota de la huella congénica, mientras que KO1 apenas tiene este tipo de variantes (figura 24C). Por lo tanto, el embrión KO1 es homocigoto para ambas partes de la huella, mientras que KO2 y KO3 no lo son. La figura 22D muestra un resumen de la ploidía en cada embrión KO para *Stc1*, que evidencia la variabilidad de la ploidía en la huella congénica. Se incluye un embrión silvestre que contiene variantes heterocigotas a lo largo de toda la huella congénica. Todos estos análisis sugieren que la herencia de la huella congénica puede ser compleja y no puede asumirse como homocigótica en todos los casos.

A

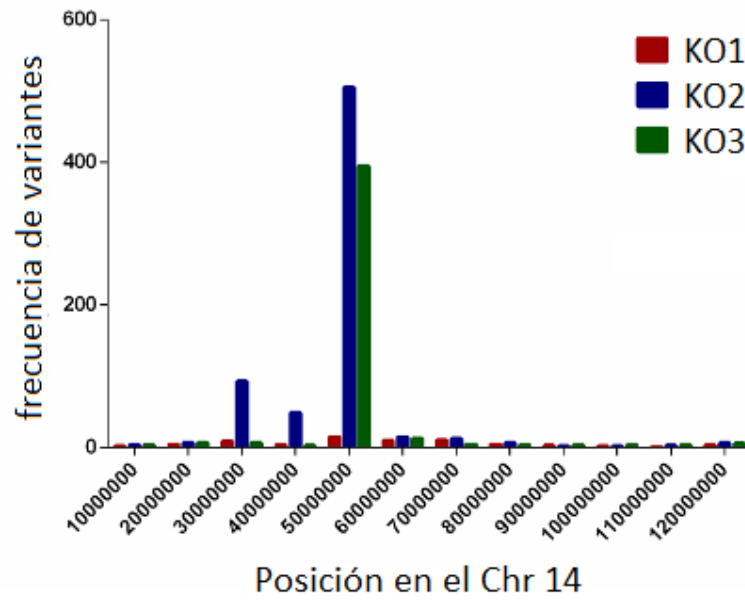


B



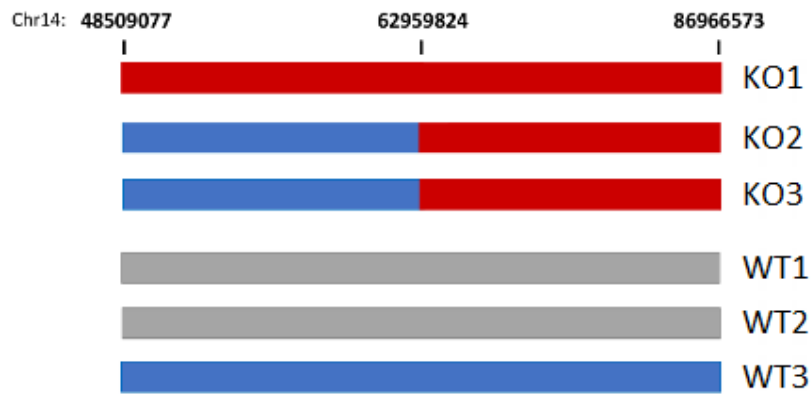
C

### Variantes heterocigotas por embrión *Stc1* KO



D

### Huella de *Stc1* KO



#### Leyenda

- Homocigoto
- Heterocigoto
- sin variantes de "129"

**Figura 24. Ploidía de la huella congénica en el modelo KO para *Stc1*.** A) Histograma de variantes ligadas al genotipo *Stc1* KO, que diferencian variantes homocigotas (barras rojas) de variantes heterocigotas (barras azules). Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes con cada ploidía como histogramas de frecuencia de acuerdo con estas posiciones. B) Histograma de variantes homocigóticas por embrión, ligadas al genotipo *Stc1* KO. Representamos variantes de cada embrión con un color distinto (KO1 = rojo, KO2 = azul, KO3 = verde). Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada embrión como histogramas de frecuencia de acuerdo con estas posiciones. C) Histograma de variantes heterocigóticas por embrión, ligadas al genotipo *Stc1* KO. Representamos variantes de cada embrión con un color distinto (KO1 = rojo, KO2 = azul, KO3 = verde). Agrupamos las coordenadas genómicas de cada cromosoma cada 10 millones de bases, y graficamos las variantes de cada embrión como histogramas de frecuencia de acuerdo con estas posiciones. D) Esquema que representa la huella congénica de cada embrión en el modelo KO para *Stc1*, ubicado en Chr 14. Representamos cada ploidía con un color distintivo (segmento homocigótico en rojo, segmento heterocigótico en azul) y se marcaron las posiciones cromosómicas por encima de las barras. Fuente: Elaboración propia.

### **La huella congénica influye en la expresión génica de las células MEF KO para *Sall2***

A continuación, investigamos la influencia de la huella congénica en la expresión génica de las células MEF KO para *Sall2*. Para estudiar la introgresión del genoma 129P2OlaHsd en el Chr 14 de las células MEF KO para *Sall2*, se comparó el transcriptoma de MEF silvestre versus MEF *Sall2* KO, identificando 520 genes diferencialmente expresados con una tasa nominal de descubrimientos falsos menor a 0.01 (del inglés “**False Discovery Rate**”, FDR <0,01) (Figura 25A). El análisis de ontología genética muestra varios grupos de genes implicados en las respuestas

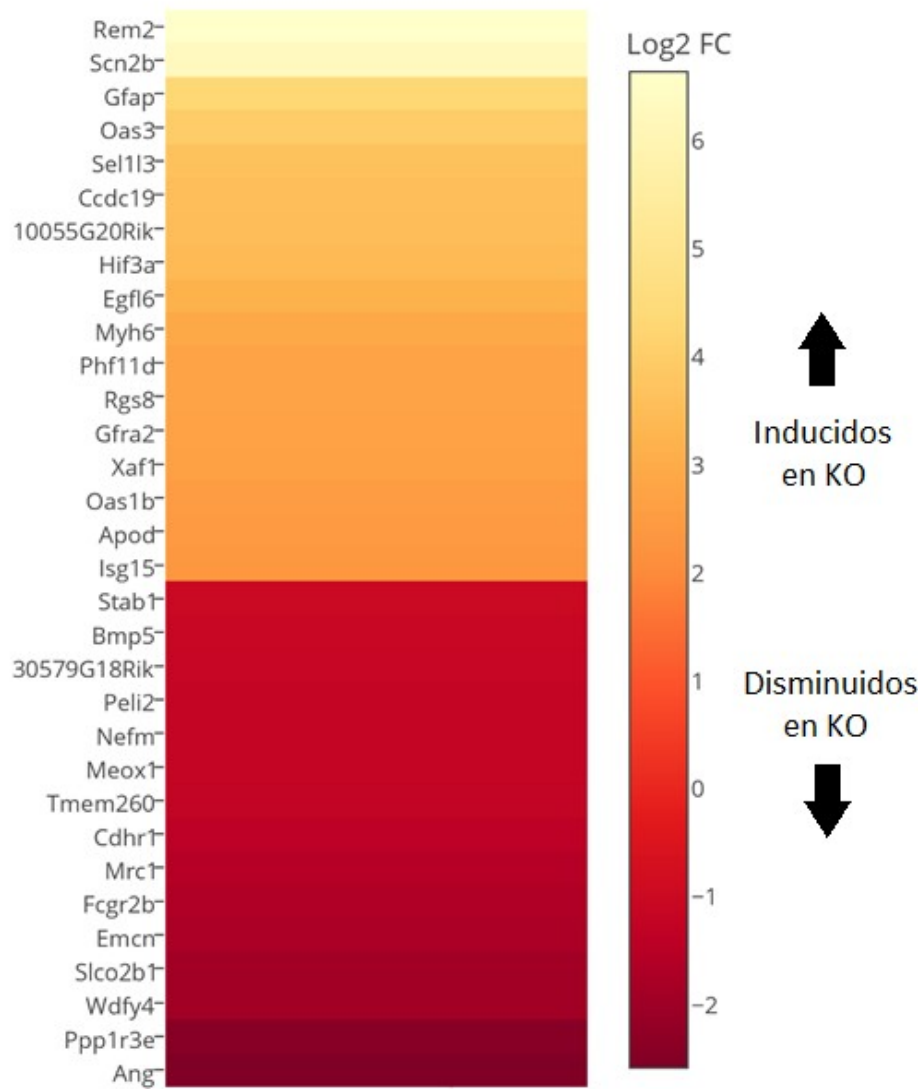


inmunitarias e inflamatorias, la respuesta al virus y la adhesión celular, entre otros, lo que sugiere la asociación de *Sall2* con estas funciones biológicas (Figura 25B).

A

## ARN-Seq

*Sall2* WT vs Null



## B Ontología genética (GO)

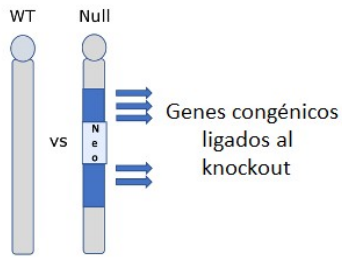


**Figura 25. Análisis de los genes diferencialmente expresados en MEFs silvestre versus KO para *Sall2*.** A) Lista de genes diferencialmente expresados con FDR<0.01 y con veces de cambio superior a 0.4. Se muestra la escala de veces de cambio en logaritmo en base 2, en donde los genes reprimidos en el KO se denotan en rojo. Fuente: Elaboración propia.

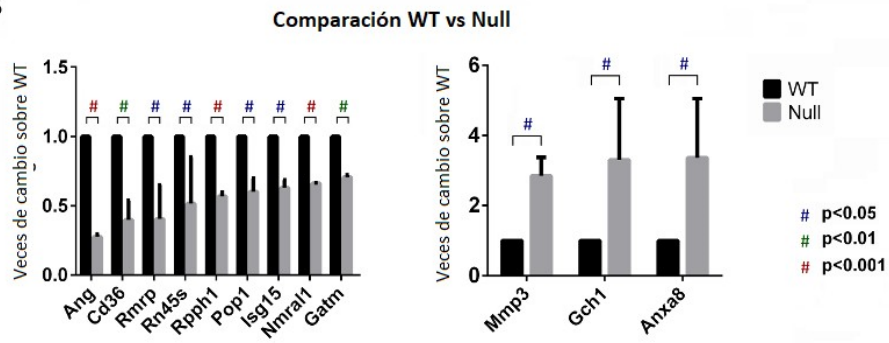
Confirmamos varios genes diferencialmente expresados por PCR cuantitativa en tiempo real (qPCR), que incluyen genes no codificantes tales como *Rmrp*, el precursor ribosomal *Rn45s* y genes localizados en la huella congénica tales como Angiogenina (*Ang*), el componente catalítico de la ribonucleasa P (*Rpph1*), *Gch1* y *Anxa8* (figura 26B). Para abordar si estos genes son sensibles a la pérdida de *Sall2*,

silenciamos *Sall2* en células MEF silvestres (WT) usando un ARN corto de horquilla (véase shRNA en materiales y métodos) y realizamos qPCR. Esta comparación es genéticamente válida porque prueba la perturbación de *Sall2* dentro del mismo fondo genético (figura 26A). De todos los genes probados, solo *Gch1*, *Rpph1*, *Ang* y *Cd36* respondieron al silenciamiento de *Sall2*, pero con cambios en la expresión diferentes a los obtenidos a partir del análisis de datos de ARN-seq (figura 26D). Estas diferencias en las veces de cambio probablemente se deben a la interferencia genética de los genes sensibles a *Sall2* con la huella congénica, presente en el modelo KO para *Sall2*. Con el objetivo de hacer una validación cruzada de la expresión de los genes sensibles a la pérdida y ganancia de *Sall2*, utilizamos datos de un estudio de microarreglos hecho en células madre embrionarias (MC1, derivadas de la línea congénica 129S6/SvEvTac) que inducían *Sall2* (Yamamizu et al. 2016). Este estudio permitió la validación cruzada de otros 37 genes diferencialmente expresados del experimento de ARN-Seq (Figura 26E, izquierda). A partir de esta comparación, 15 de estos genes presentaron en su expresión veces de cambio opuestas y consistentes entre ambos estudios (figura 26E, derecha). Evaluamos dos de estos genes diferencialmente expresados por qPCR, confirmando las tendencias observadas en el experimento de ARN-Seq y en el estudio de microarreglo (figura 26F).

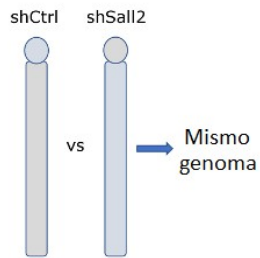
A



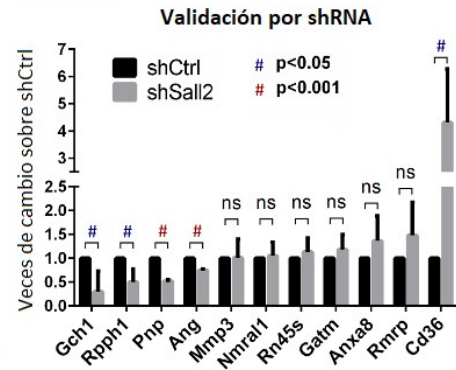
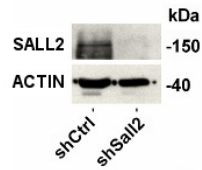
B



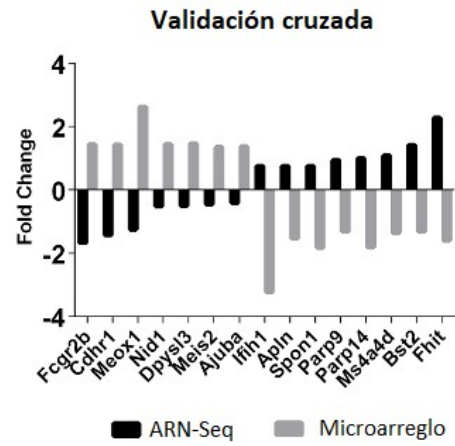
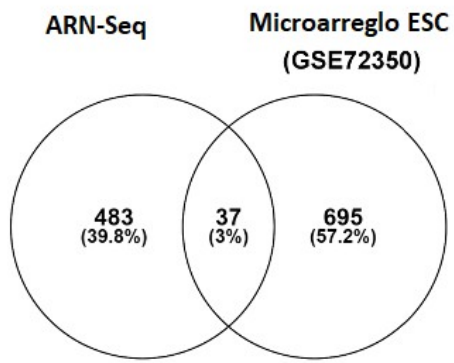
C



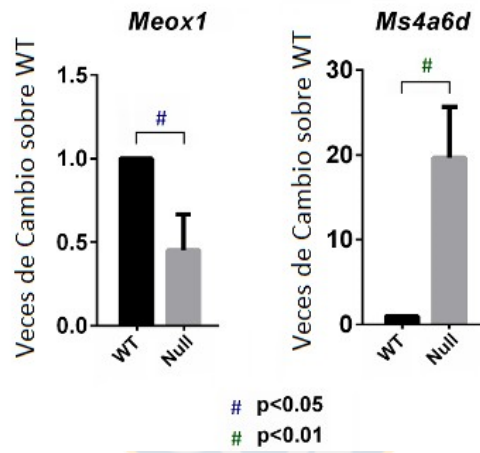
D



E



F



**Figura 26. La huella congénica influye en la expresión génica de las células MEF KO para *Sall2*.** A) esquema que representa la comparación entre las células MEF *Sall2* silvestre versus KO. El color gris denota la cepa receptora C57BL/6J y los segmentos azules representan el genoma introgresado de la célula madre embrionaria con los genes congénicos correspondientes que flanquean el locus de *Sall2* (véase el segmento cuadrado “Neo” que denota el casete de neomicina utilizado en la recombinación homóloga). B) Validación de varios genes diferencialmente expresados reprimidos (izquierda) o inducidos (derecha) entre las células MEF inmortalizadas *Sall2* silvestre y KO. Se muestran los niveles de expresión para cada gen normalizados con el ARNm de *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el genotipo silvestre (N = 3, los datos se representan como medias  $\pm$  s.e.m.). azul # P <0.05; verde # P <0.01; rojo # P <0.001 frente al genotipo silvestre; Prueba de T de Student. C) Esquema que representa la comparación entre las células MEF con el silenciador control (del inglés “scramble”, shCtrl) y en silenciadas para *Sall2* (shSall2). El color gris denota la cepa receptora C57BL/6J, mostrando que esta comparación es genéticamente correcta. D) Izquierda: “Western Blot” representativo para las proteínas *Sall2* y actina en las células MEF inmortalizadas con el silenciador control (shCtrl) y en células silenciadas para *Sall2* (shSall2). Derecha: PCR cuantitativa en tiempo real de los genes testeados en (B). Se muestran los niveles de expresión normalizados para cada gen con el ARNm de *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el silenciador control (shCtrl) (N = 3, los datos se representan como medias  $\pm$  s.e.m.). azul # P <0.05; rojo # P <0.001 versus shCtrl; Prueba T de Student. E) Izquierda: diagrama de Venn que compara los genes diferencialmente expresados encontrados en el KO para *Sall2* (FDR <0.01, veces de cambio > 0.35 en log2) con los genes diferencialmente expresados provenientes del microarreglo de inducción de *Sall2* en células madre embrionarias (ESC) (veces de cambio de expresión > 1.3, código de acceso GeoDatasets GSE72350). La superposición revela 37 genes diferencialmente expresados que son comunes. Derecha: Validación cruzada de los genes diferencialmente expresados de la superposición anterior mediante las veces de cambio de expresión registrados en ambos estudios. Consideramos sólo los genes diferencialmente expresados con veces de cambio de expresión opuestas en ambos estudios. F) Validación mediante qPCR de dos de los genes diferencialmente expresados anteriormente descritos (*Meox1* y *Ms4a6d*, “head to head” con *Ms4a4d*) en MEFs silvestres o KO para *Sall2* (N = 3; los datos se representan como medias  $\pm$  s.e.m.). azul # P <0.05; verde # P <0,01 frente a WT; Prueba T de Student. Fuente: Elaboración propia.

Las diferencias en la expresión génica entre el modelo knockout para *Sall2* y el modelo del silenciamiento, se resumen en la siguiente tabla:

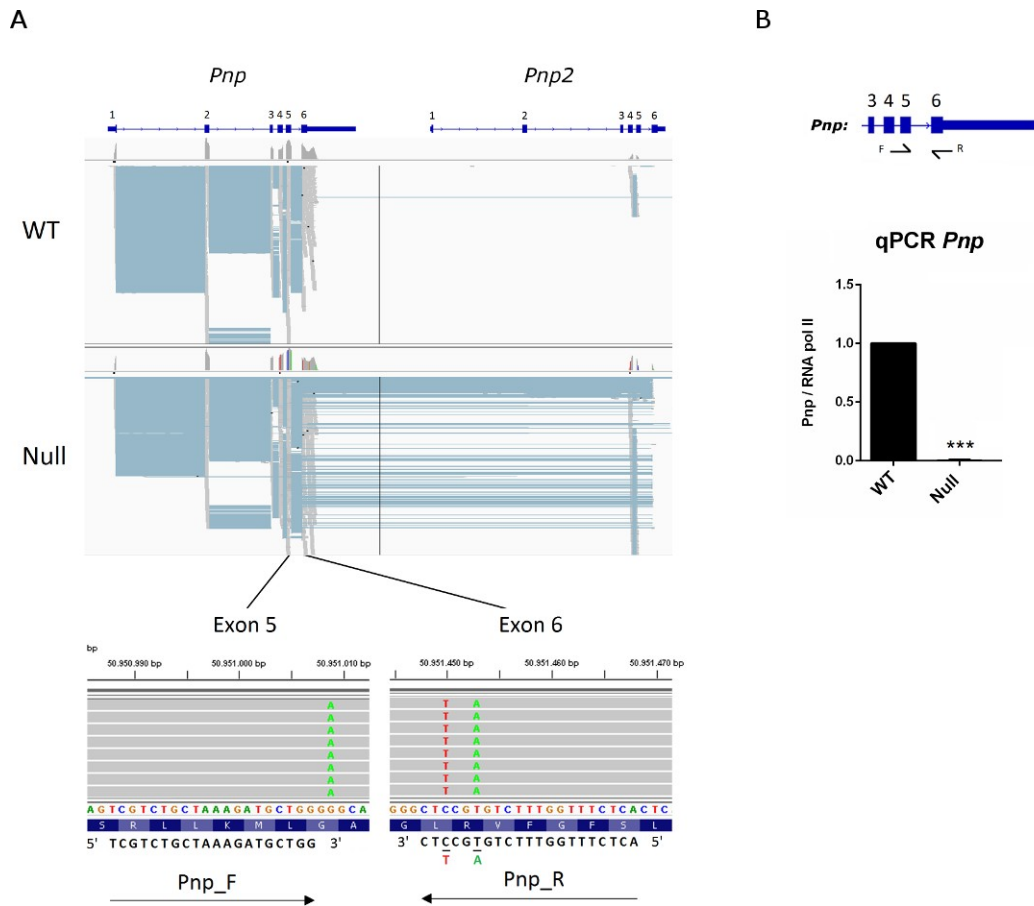
Genes	Knockout	Silenciamiento
Ang	↓	↓
Cd36	↓	↑
Rmrp	↓	-
Rn45s	↓	-
Rpph1	↓	↓
Nmral1	↓	-
Gatm	↓	-
Mmp3	↑	-
Gch1	↑	-
Anxa8	↑	-
Pnp	-	↓

**Tabla 4: Diferencias en la expresión génica entre el modelo *Sall2* “knockout” y el modelo de silenciamiento para *Sall2*.** Los genes reprimidos en cada modelo se indican con una flecha hacia abajo y los genes inducidos se indican con una flecha hacia arriba. Genes sin cambios significativos se denotan con una raya horizontal. Los genes con tendencias similares en cada modelo se destacan con amarillo. Fuente: Elaboración propia.

El cribado de genes dentro de la región congénica mediante silenciamiento con shRNA reveló el gen *Pnp* como un gen diferencialmente expresado, sin embargo, este gen no apareció en el análisis inicial de ARN-Seq (Figura 26D). Aunque el análisis de PCR cuantitativa en tiempo real mostró inicialmente que el (ARNm) *Pnp* está regulado negativamente en células MEF *Sall2* KO, este resultado fue un artefacto resultante de una pobre hibridación del partidor inverso debido a mutaciones en la región genómica, confundiendo el análisis de *Pnp* en las células MEF *Sall2* KO (véase figura 27).



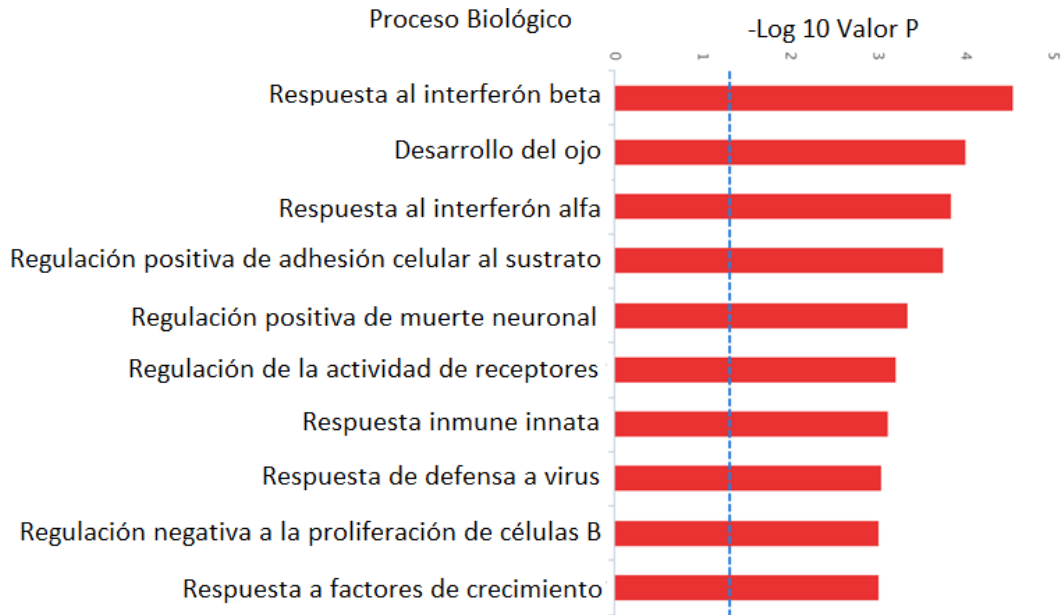




**Figura 27. Detección del (ARNm) *Pnp* en células MEF *Sall2* KO.** A) Imagen de la herramienta de visualización IGV (del inglés Integrative Genomics Viewer) del (ARNm) de *Pnp/Pnp2* presente en células MEF *Sall2* silvestres (WT) y KO. Los números representan los exones de ambos genes y se indica en cada exón los partidores para la detección de los transcritos (Exón 5 = Sentido, Exón 6 = Reverso). Magnificamos las mutaciones en la región de hibridación de ambos partidores (en rojo y verde), y lo subrayamos alrededor del partidore inverso en el exón 6 de *Pnp*. B) Superior: Esquema del gen *Pnp* y la posición de los partidores para la PCR en tiempo real. Inferior: PCR cuantitativa en tiempo real del ARNm de *Pnp* en células MEF *Sall2* silvestres (WT) y KO. Se muestran los niveles de expresión de *Pnp* normalizados con *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el genotipo silvestre (WT) (N = 4, \*\*\* P <0,001, frente a WT, Prueba T de Student). Fuente: Elaboración propia.

Los 15 genes diferencialmente expresados provenientes de la superposición de nuestro experimento con el estudio de microarreglos, recapitularon parcialmente la ontología génica inicialmente encontrada (respuesta a interferón beta, respuesta, inmune innata, respuesta de defensa frente a virus y adhesión celular, véase la figura 28).

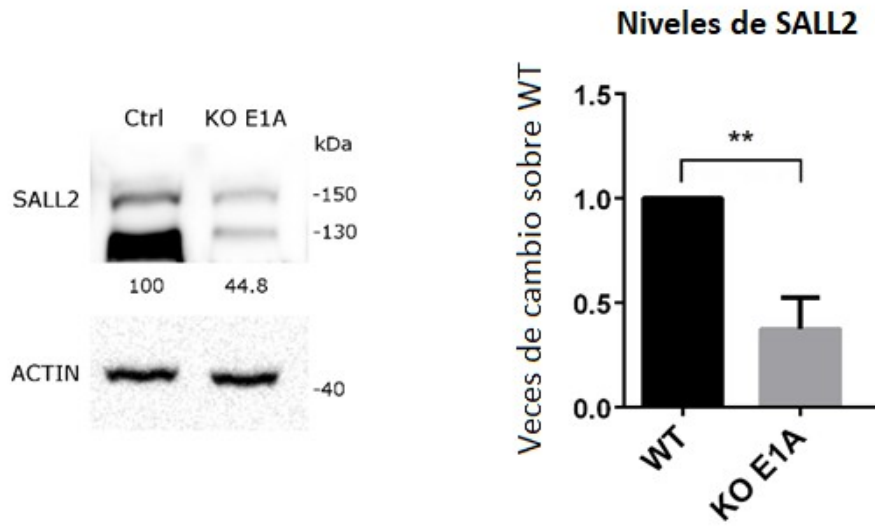




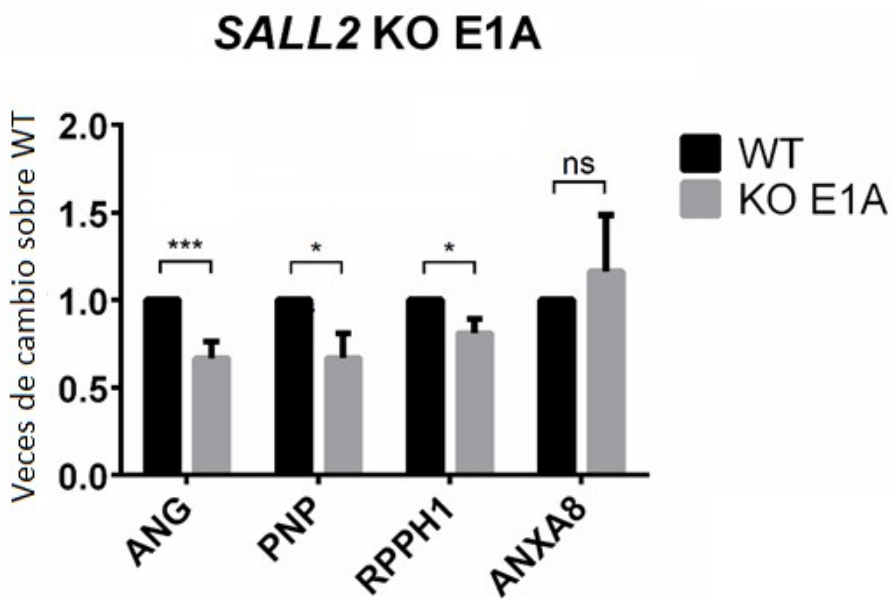
**Figura 28. Ontología genética para los 15 genes provenientes de la superposición del experimento de ARN-Seq con el estudio de microarreglos de la inducción de *Sall2*.** Se recapitularon los términos de respuesta a interferón beta, respuesta inmune innata, respuesta de defensa frente a virus y adhesión celular con respecto a la figura 25B. Fuente: Elaboración propia.

Adicional a la validación cruzada que realizamos en ratón, hemos validado la represión transcripcional de los genes *ANG*, *PNP* y *RPPH1* utilizando un modelo CRISPR/Cas9 para la isoforma E1A de *SALL2* en células HEK293. Cabe destacar que esta es la isoforma de *SALL2* más ubicua (véase introducción). (figura 29).

A



B



**Figura 29. Validación cruzada de los genes diferencialmente expresados sensibles a SALL2 en células humanas HEK293.** Generamos un modelo KO de la isoforma principal codificada por el gen SALL2 mediante CRISPR/Cas9 (isoforma E1A) para la validación cruzada los genes sensibles a SALL2. A) “Western Blot” representativo para las proteínas SALL2 y ACTINA en clones individuales HEK293 transfectados con el CRISPR Control (sin sgRNA) o el CRISPR para la isoforma SALL2 E1A. Cuantificación de las bandas de proteína SALL2 normalizadas contra ACTINA en el gráfico de barras a la derecha (N=4 para control (WT), N=4 para CRISPR E1A (KO E1A), \*\* P <0.01 versus WT; Prueba de T de Student) B) Validación de genes disminuidos en el KO de la isoforma E1A de SALL2 (KO E1A) en células HEK293. Se muestran los niveles de expresión de cada gen normalizados con el (ARNm) del gen que codifica la ARN polimerasa II (gen POLR2A). Expresamos los valores de expresión como veces de cambio sobre HEK293 silvestres (WT) (N=3, los datos se representan como medias  $\pm$  s.d.). \*\*\* P <0.001; \* P <0.05; ns, no significativo frente a WT; Prueba T de Student. Fuente: Elaboración propia.

Estos resultados confirman que la huella congénica y su interacción con el fondo genético influyen en el análisis del transcriptoma de los modelos KO, por lo tanto, la validación de los genes blanco, sensibles al gen en estudio, deben ser sometidos a validación cruzada mediante enfoques adicionales. Adicionalmente, la validación de los genes diferencialmente expresados por PCR cuantitativa en tiempo real puede estar sesgada, ya que los polimorfismos que se presentan en los segmentos introgrados de las células madre en las cepas receptoras C57BL/6J, pueden confundir el análisis (ver el caso de la validación por PCR cuantitativa en tiempo real del (ARNm) *Pnp* en el modelo KO para *Sall2*, figura 27).

En síntesis, se identificaron 15 genes diferencialmente expresados que son sensibles a la pérdida y ganancia de función de *Sall2*, mediante el estudio de ARN-Seq y microarreglo. También se validaron genes ubicados en la región congénica (*Ang*, *Rpph1*, *Pnp*) tanto en células iMEF silenciadas para *Sall2* como en células humanas KO para la isoforma E1A de SALL2. Sin embargo, una parte importante de

los genes diferencialmente expresados en el modelo KO para *Sall2* corresponderían a la interacción de la huella congénica con el fondo genético, sin contar los genes congénicos que están ligados al genotipo *Sall2* KO *per se*. Se ha reportado que los genes congénicos son los genes más diferencialmente expresados en los estudios con KO congénicos y su expresión estaría ligada al fondo genético. Este punto enfatiza la precaución con el uso de los modelos KO congénicos cuando se atribuye efectos fenotípicos a los genes (Schalkwyk et al. 2007). Dado que *Ang* es un gen congénico en el modelo KO para *Sall2* y uno de los genes más diferencialmente expresados en este modelo, potencialmente podría comportarse como un gen modificador de fenotipos en el modelo KO para *Sall2* (del inglés “Modifier Gene”). En base a los argumentos anteriores, se implementó un método computacional para detectar este tipo de genes en estudios con modelos KO.

### **Cribado de loci de carácter cuantitativo de expresión (eQTL) en la región congénica del modelo KO para *Sall2***

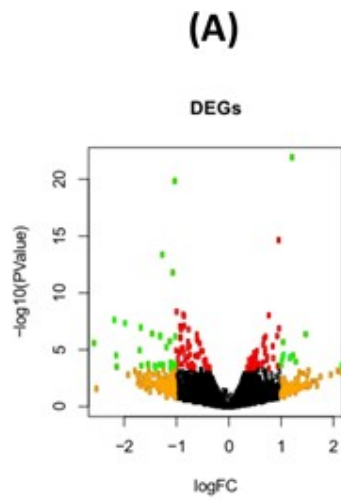
Como se mencionaba, las metodologías anteriores no tienen en cuenta la detección de potenciales genes modificadores (del inglés “Modifier genes”) que actuarían como loci de carácter cuantitativo de expresión (eQTL) en la región congénica de los ratones KO para *Sall2*. En genética, estos loci se refieren a genes cuya variación alélica está asociada con la variación de un carácter cuantitativo, tal como su expresión génica y los correspondientes fenotipos asociados (Masinde et al. 2001). En el caso de las células MEF KO para *Sall2*, alrededor de 170 genes en el Chr 14 son congénicos; por lo tanto, se espera interferencia genética en el modelo KO

para *Sall2* a través de estos (véase Tabla 3). De los 170 genes, 58 tienen mutaciones con cambio de sentido y 17 son genes diferencialmente expresados (véase Tabla 5).

Ang	Rnase4
Tmem260	Samd4
4930579G18Rik	Fbxo34
Haus4	Ajuba
Sall2	Pabpn1
Loxl2	Supt16
Mmp14	Rpl15
Slc7a7	Hr
Osgep	

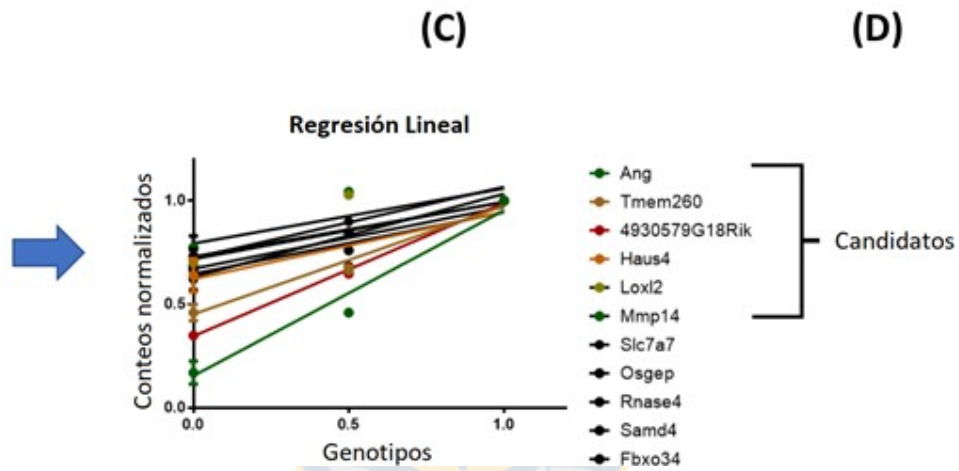
**Tabla 5. Genes diferencialmente expresados que se localizan en la huella congénica del modelo KO para *Sall2*.** Se muestran los 17 genes diferencialmente expresados que se localizan en la huella (FDR < 0.01, veces de cambio > 0.4). Los genes que contienen polimorfismos con cambio de sentido se denotaron en rojo. Fuente: Elaboración propia.

Para identificar posibles genes modificadores en la región congénica del modelo KO para *Sall2*, aplicamos la estrategia de búsqueda de loci de carácter cuantitativo de expresión (eQTL) (Majewski and Pastinen 2011). Propusimos un protocolo para la detección de genes candidatos como eQTLs, en función de la dependencia lineal de su expresión al genotipo, tal como se ha descrito anteriormente (Lagarrigue et al. 2013). Obtuvimos los recuentos de las lecturas de secuenciación de los 15 genes de la tabla 5 en MEF con genotipos *Sall2* silvestre (WT), heterocigoto (Het) y KO, respectivamente. Estos recuentos fueron previamente normalizados por la estrategia RUVSeq (Bioconductor) y probados en regresión lineal contra valores arbitrarios para cada genotipo (WT = 1, Het = 0.5 y KO = 0). Esta estrategia se detalla a continuación en la figura 30.



**(B)**

DEGs	WT	Het	Null
Ang	1	0.4601406	0.1101591
4930579G18Rik	1	0.6471568	0.3349866
Tmem260	1	0.6636857	0.4597086
Slc7a7	1	0.898505	0.6257595
Haus4	1	0.6751145	0.6205285
Rnase4	1	0.6820965	0.5683943
Osgep	1	0.7595198	0.6337372
Loxl2	1	1.028426	0.7696849
Samd4	1	0.8454538	0.709273
Fbxo34	1	0.8469745	0.6821615
Mmp14	1	1.043518	0.8292157






**Figura 30: Protocolo para la detección de loci de carácter cuantitativo de expresión (eQTL) a partir de datos de ARN-Seq.** Diseñamos un protocolo basado en la dependencia lineal de la expresión génica al genotipo, para la detección de loci de carácter cuantitativo de expresión (eQTL) en regiones congénicas. (A) Seleccionamos como entrada genes congénicos diferencialmente expresados (DEG congénicos) del experimento de ARN-Seq. (B) Se normalizaron los recuentos de las lecturas de secuenciación de los genes congénicos diferencialmente expresados en los genotipos silvestre (WT), heterocigoto (Het) y KO (por ejemplo, mediante el paquete RUVSeq, disponible en Bioconductor) y se usaron para el análisis de regresión lineal. (C) Se obtuvieron las regresiones lineales de recuentos de las lecturas de secuenciación de los genes anteriormente seleccionados a través de los genotipos, para los cuales previamente se definieron valores arbitrarios (WT = 1, Het = 0.5, KO = 0). (D) Los genes con pendientes significativamente distintas de cero se seleccionan de acuerdo con el valor de  $p < 0,05$  (representado como pendientes de color en el gráfico en (C)). Después de este procedimiento, obtuvimos como resultado una lista de genes con eQTL candidatos. Fuente: Elaboración propia.

Primero, realizamos el análisis global de la expresión génica en el experimento de ARN-Seq en MEFs silvestres o KO para *Sall2*, con o sin tratamiento con doxorubicina (1  $\mu\text{M}$  por 16 horas). Las muestras tratadas con esta droga se incluyeron en este análisis, ya que se ha reportado que esta droga aumenta el recambio de nucleosomas alrededor de los promotores de genes activos, generando una perturbación global de la transcripción (Yang et al. 2013). El tratamiento per sé aumentó la expresión génica en la región congénica, aumentando el número de genes expresados con las variantes correspondientes (ver figura 31 y tabla 6, respectivamente).



**Figura 31: Impacto del tratamiento con doxorubicina como una perturbación global de la expresión génica e histograma de las variantes ligadas al genotipo *Sal12* KO con y sin el tratamiento con doxorubicina.** A) Obtuvimos la razón de la expresión génica (logaritmo 2) desde los archivos bam entre las muestras *Sal12* silvestre con o sin tratamiento con doxorubicina (1  $\mu$ M durante 16 horas) mediante el programa bamCompare del paquete DeepTools, disponible en el servidor principal de la plataforma Galaxy. Mostramos hasta 10 megabases del Chr 14 a través del servidor USCS, donde se encuentra la huella congénica. (Escala = 5 megabases, ensamblaje de genoma de ratón mm10). Se denota el gen *Sal12* en rojo. B) Impacto en la detección de variantes ligadas al genotipo KO después del tratamiento con doxorubicina en MEFs *Sal12* KO. Se graficó el promedio de las variantes ligadas al genotipo KO (Chr 14) con o sin tratamiento con doxorubicina. Adicionalmente a las variantes, aumentó el número de genes congénitos expresados (véase Tabla 6). Fuente: Elaboración propia.



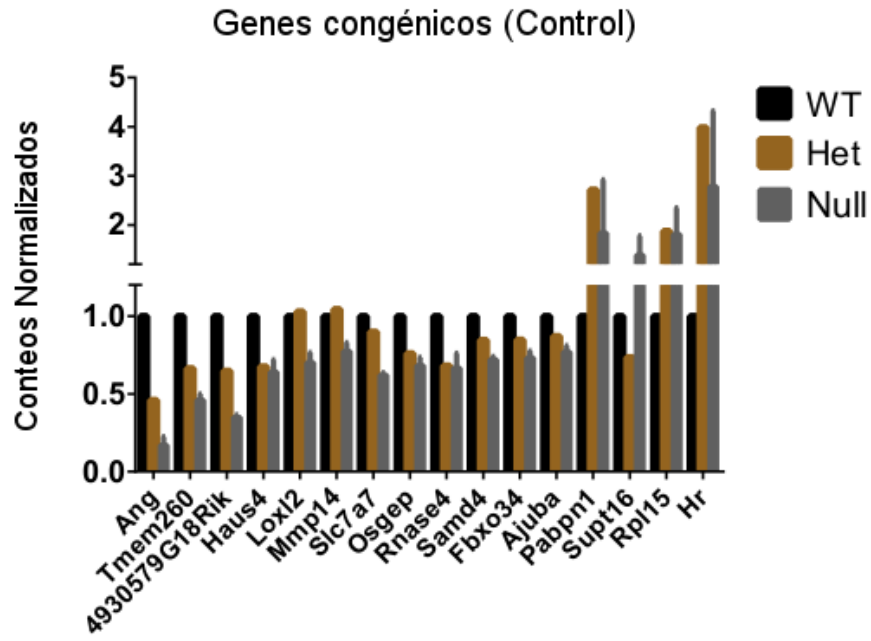
AC154515.2	Mrpl52	Zfx2	Stc1	Rgcc
CU024892.2	4931414P19Rik	Gm20687	Gm21451	Pcdh17
AC190403.1	Slc7a8	Thtpa	Gm16867	<b>Mycbp2</b>
Gm15601	<b>Homez</b>	Atp8a2	<b>Entpd4b</b>	Tgds
Gch1	Gm10332	Gm9013	Gm27174	Gm26773
Peli2	Ppp1r3e	Fam124a	Gm27176	Gm4675
<b>3632451O06Rik</b>	Gm37271	<b>Zfp395</b>	Egr3	Sox21
Klhl33	<b>Cmtm5</b>	Ppp2r2a	Phyhip	Dzip1
A930018M24Rik	Il25	Kctd9	<b>Dmtn</b>	Cldn10
Gm23758	Myh6	Gnrh1	Serp2	

**Tabla 6. Genes congénicos adicionales expresados en el tratamiento con doxorubicina en las MEF provenientes del modelo KO para *Sal12*.** Mediante la clasificación de las variantes ligadas al genotipo KO encontradas bajo el tratamiento con doxorubicina (Figura 31B), se dedujeron los nuevos genes expresados en la región congénica del modelo KO para *Sal12* bajo el tratamiento con doxorubicina, previamente no encontrados en la condición control. Se muestra en rojo, los genes que contienen variantes con cambio de sentido de acuerdo con la clasificación del programa VEP (del inglés, “**V**ariant **E**ffect **P**redictor”). Fuente: Elaboración propia.

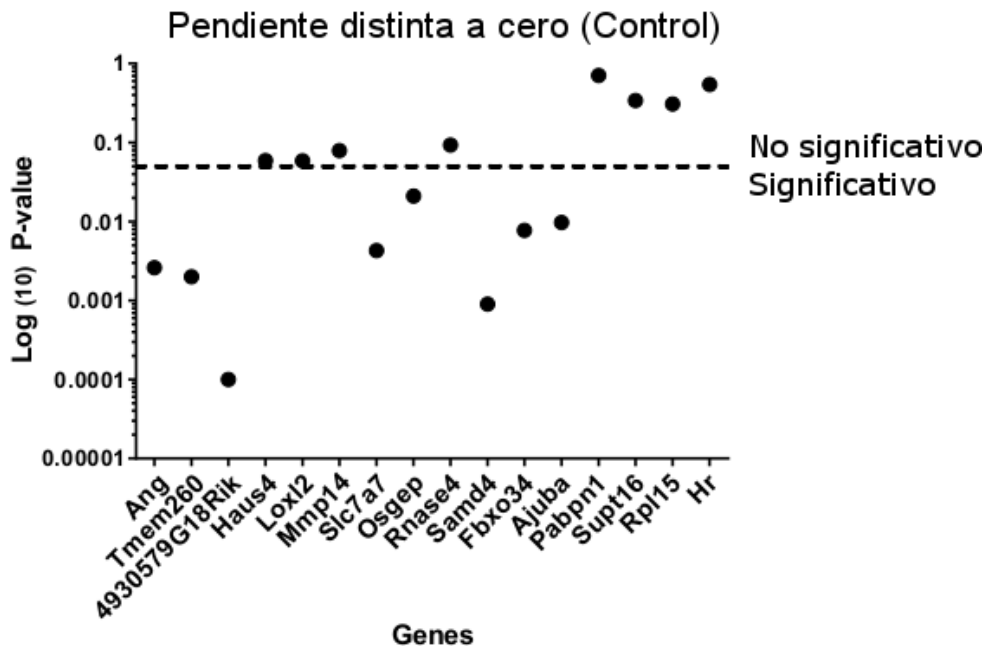
Once genes congénicos diferencialmente expresados (condición control) fueron ordenados acorde a las veces de cambio de expresión entre los genotipos silvestre versus KO (ver figura 32A), de los cuales siete mostraron dependencia lineal al genotipo mediante regresión lineal (ver figura 32B). La perturbación global con doxorubicina alteró la expresión de estos genes, y el número de genes diferencialmente expresados con dependencia genética (ver figura 32C y 32D, respectivamente). Cuatro de estos once genes mostraron dependencia genética tanto en el control como en el tratamiento con doxorubicina (*Ang*, *Tmem260*, *4930579G18Rik* y *Osgep*, respectivamente, ver figura 32E) y angiogenina (*Ang*) es el gen más diferencialmente expresado en ambos casos (ver figura 32F).



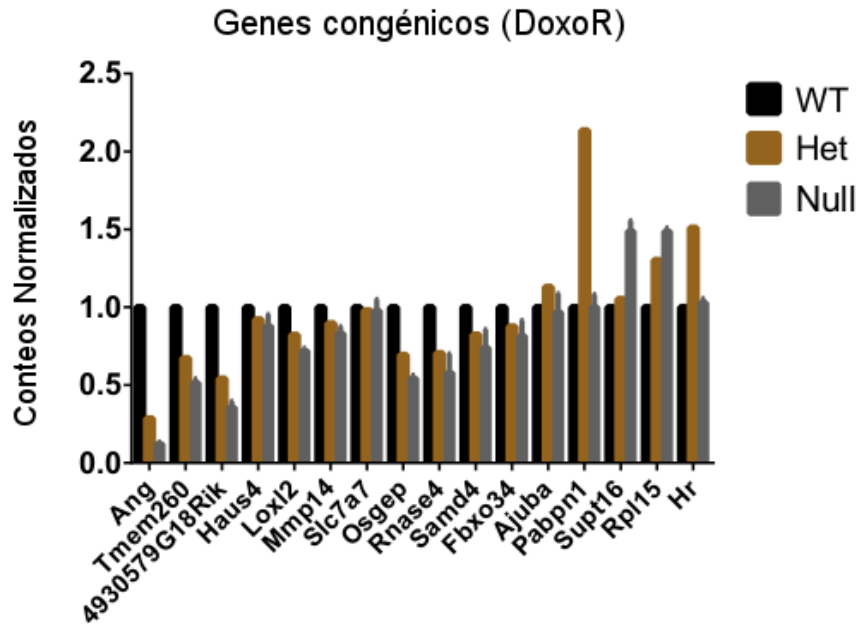
A



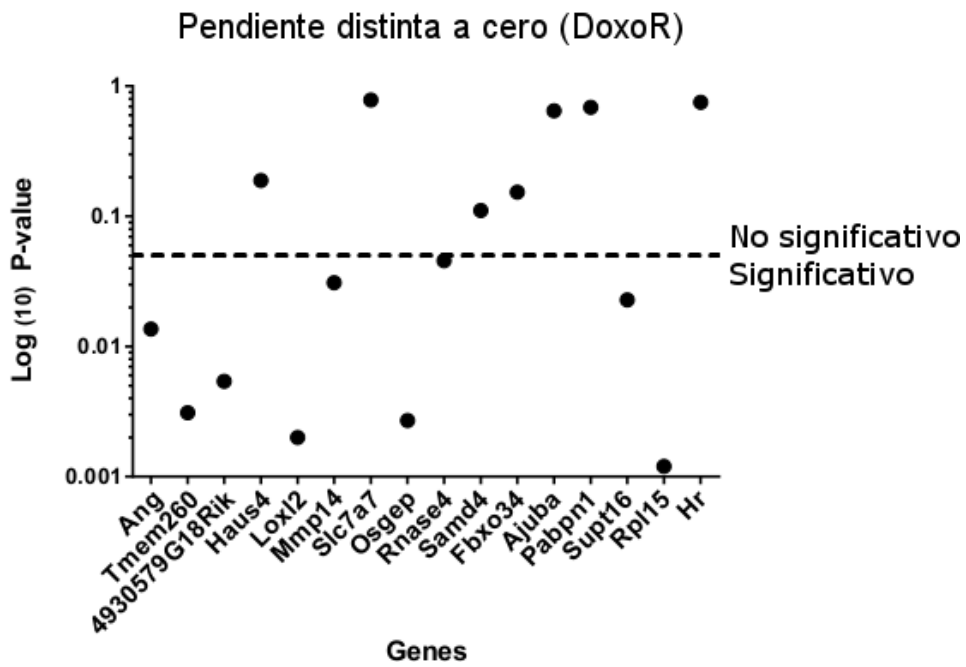
B



C



D

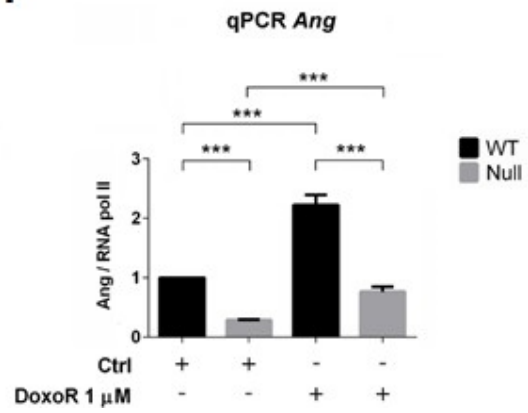


# E

Genes candidatos en ambos experimentos:

- Ang
- Tmem260
- 4930579G18Rik
- Osgep

# F



**Figura 32: Detección de loci con rasgos cuantitativos de expresión (eQTL) en el modelo KO para *Sall2* con o sin perturbación con doxorubicina.** A) Recuento de las lecturas de secuenciación de los genes congénicos diferencialmente expresados a través de los genotipos silvestre (WT), heterocigotos (Het) y knockout (KO) de las muestras de ARN-Seq en condición control (el genotipo silvestre fue normalizado arbitrariamente contra 1). Ordenamos los genes en el gráfico basado en el cambio de expresión obtenido a partir del análisis de ARN-Seq. B) Genes diferencialmente expresados con eQTL significativos en control después del análisis de regresión lineal como se describe en la figura 30. Los genes con pendiente significativamente diferente de cero son candidatos para eQTL y potenciales genes modificadores de fenotipos en el modelo KO para *Sall2*. Establecimos el umbral de valor P en 0.05 usando una línea punteada (significancia < 0.05). C) Mismo análisis que en (A) para genes congénicos diferencialmente expresados bajo la perturbación con doxorubicina (muestras tratadas con doxorubicina del experimento de ARN-Seq). Ordenamos los genes en el gráfico basado en el cambio de expresión obtenido a partir del análisis de ARN-Seq. D) Mismo análisis en (B) para los genes congénicos diferencialmente expresados de (A) bajo la perturbación con doxorubicina. E) Genes candidatos con eQTL basados en valores de P significativos tanto en control como con la perturbación con doxorubicina. Acorde con este criterio, cuatro genes son potenciales modificadores de fenotipo en el modelo KO para *Sall2* (*Ang*, *Tmem260*, *4930579G18Rik* y *Osgep*). F) Validación por PCR cuantitativa en tiempo real (qPCR) del ARNm de *Ang* en MEF *Sall2* silvestre y KO con o sin tratamiento con doxorubicina. Se muestran los niveles de expresión del (ARNm) *Ang* normalizados con *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el genotipo silvestre (WT) con y sin tratamiento con doxorubicina (N = 4; los datos se representan como medias  $\pm$  s.e.m.). \*\*\* P < 0,001 frente al tratamiento con WT tratadas o no con doxorubicina; Prueba T de Student. Fuente: Elaboración propia.

### **Angiogenina es un potencial gen modificador de fenotipos en el modelo KO para *Sall2***

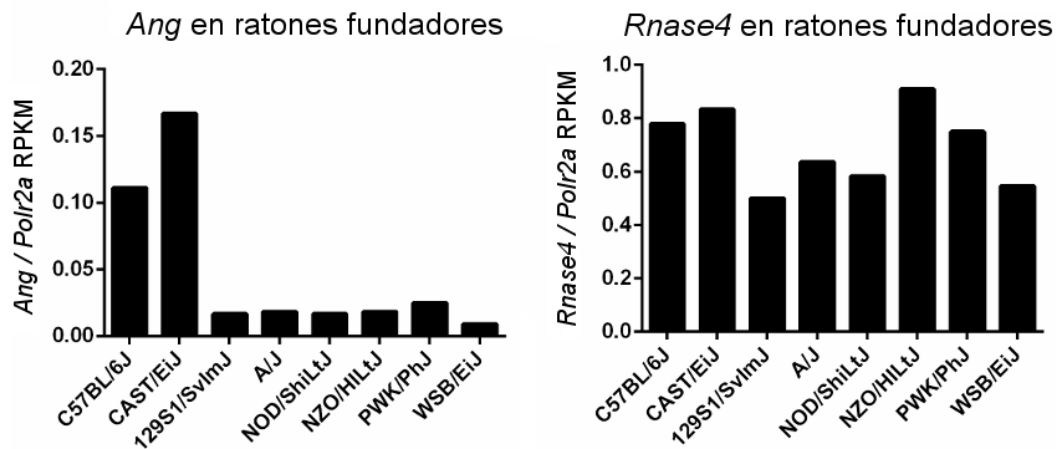
El cribado de loci de carácter cuantitativo de expresión (eQTL) reveló a angiogenina (*Ang*) como uno de los potenciales genes modificadores de fenotipo en el modelo KO para *Sall2*. El resultado de la figura 32F sugiere que el promotor de *Ang* congénico; el cual controla los genes *Ang* y *Rnase4* (Sheng et al. 2014a) es funcional,



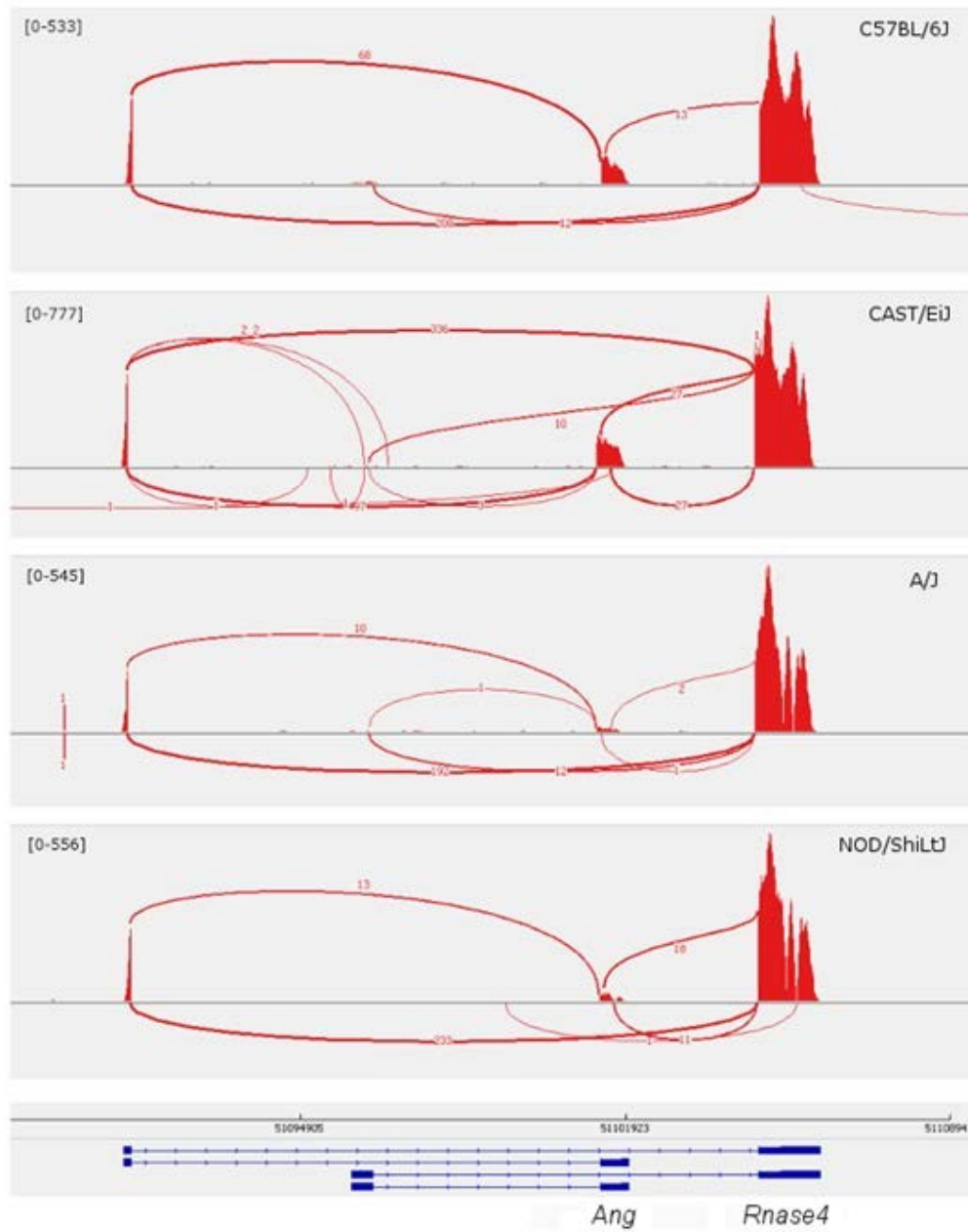
dada la inducción de angiogenina en ambos genotipos; pero la transcripción de *Ang* es baja en la cepa 129P2OlaHsd. De acuerdo con esta suposición, datos de un estudio de ARN-Seq del cuerpo estriado de las ocho cepas fundadoras en ratón (SRA project ID: PRJNA228935) (Zheng et al. 2015) reveló que la expresión de *Ang* es notablemente baja en seis de las ocho cepas, excepto en cepas C57BL/6J y CAST/EiJ, en donde *Ang* se expresa con valores superiores en comparación con el grupo (método ROUT, Q=0.1% (Motulsky and Brown 2006)). Al contrario, este efecto no es patente en la expresión de *Rnase4* (véase figura 33A). Los diagramas de Sashimi de los datos de ARN-Seq entre los fundadores de los ratones apoyan el “bypass” de la transcripción de *Ang* dentro del gen que codifica para *Rnase4* en seis cepas fundadoras, incluida la cepa que dio origen a la cepa 129 (véase figura 33B).

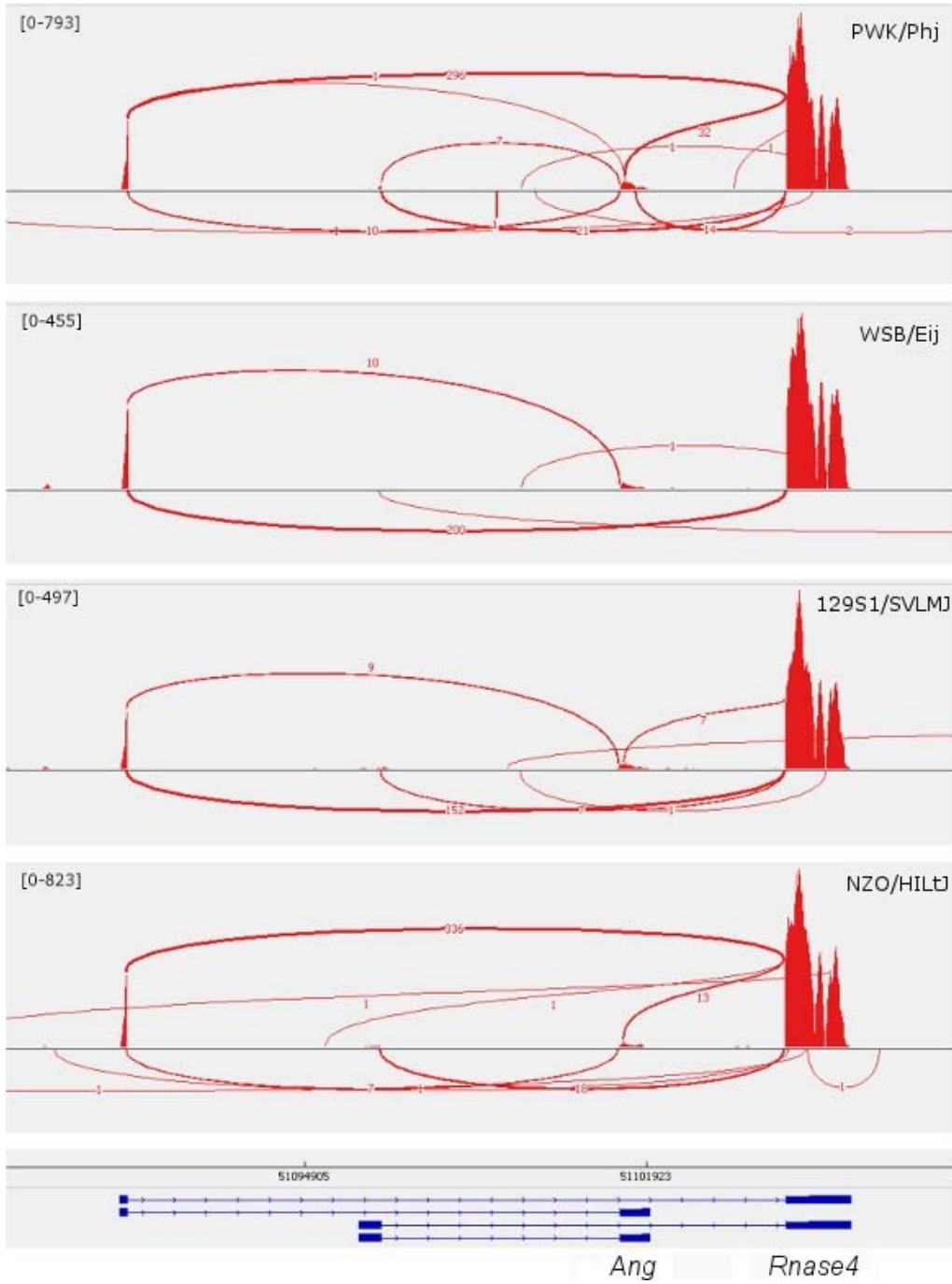


A



B

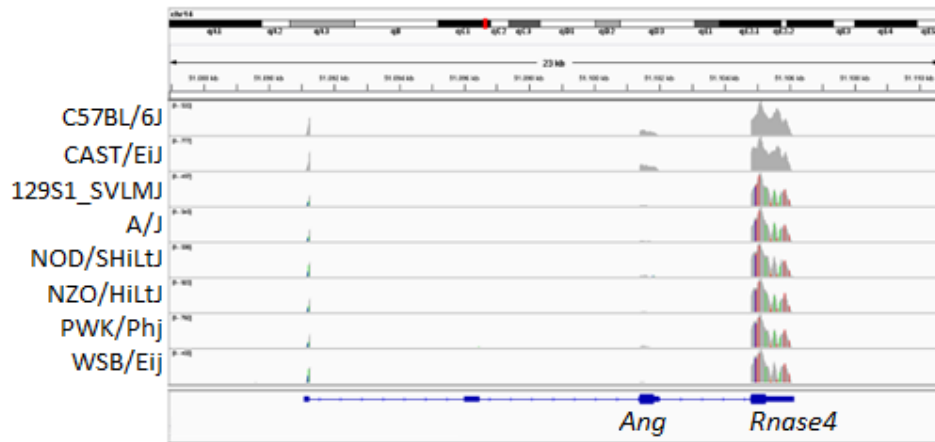




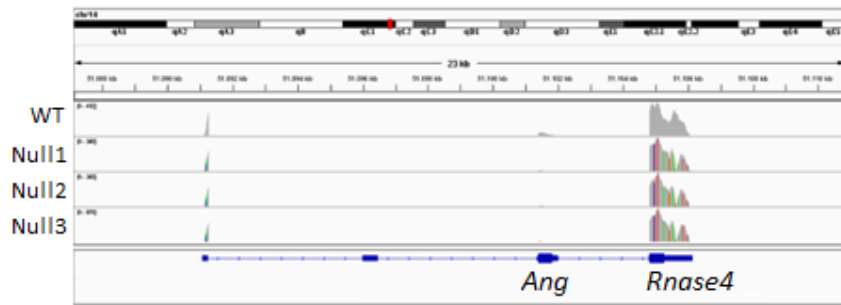
**Figura 33: *Ang* congénico es un eQTL en comparación con C57BL/6J.** A) Izquierda: recuento de lecturas de secuenciación normalizadas en kilobase por millón (RPKM) de Angiogenina (*Ang*) contra RNA Pol II (*Polr2A*) a través de las ocho cepas fundadoras de los ratones. Derecha: Lecturas normalizadas por kilobase por millón (RPKM) del gen *Rnase4* contra ARN Pol II (*Polr2A*) en las ocho cepas fundadoras. B) Gráficos de Sashimi de la expresión del gen *Ang/Rnase4* en las ocho cepas fundadoras, incluyendo C57BL/6J (superior) y la cepa 129S1/SvImJ (inferior) de la cual se derivaron las células madre embrionarias (ESC). Las coordenadas genómicas se ubicaron en el eje x, la expresión por base se graficó en el eje y, y la estructura del gen se representó en la parte inferior (en azul, obtenida del servidor USCS). Fuente: Elaboración propia.

Notamos que las cepas con bajos niveles de *Ang* en el estudio de los cuerpos estriados presentaron varias variantes a lo largo del gen *Ang/Rnase4*, ausentes en las cepas C57BL/6J y CAST/EiJ. Estas variantes también están presentes en el gen *Ang* congénico en MEF *Sall2* KO, pero ausentes en la contraparte silvestre; sugiriendo una asociación de estas variantes a la baja expresión de ARNm de *Ang* congénico. Además, otro estudio de ARN-Seq del hipocampo de ratones C57BL/6J versus 129S1/SvImJ (Vied et al. 2016) (acceso GEO DataSets GSE76567) muestra una fuerte disminución del ARNm de *Ang* en comparación con los ratones C57BL/6J, una tendencia que también confirmamos experimentalmente en la corteza de los ratones *Sall2* KO por qPCR (véase figura 34).

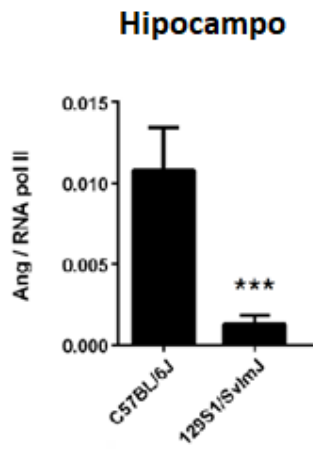
A



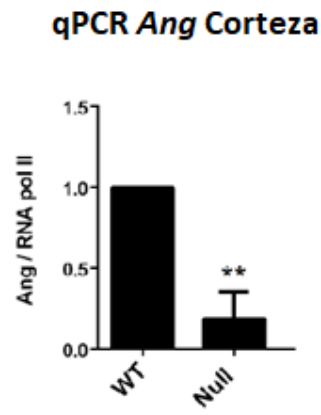
B



C



D



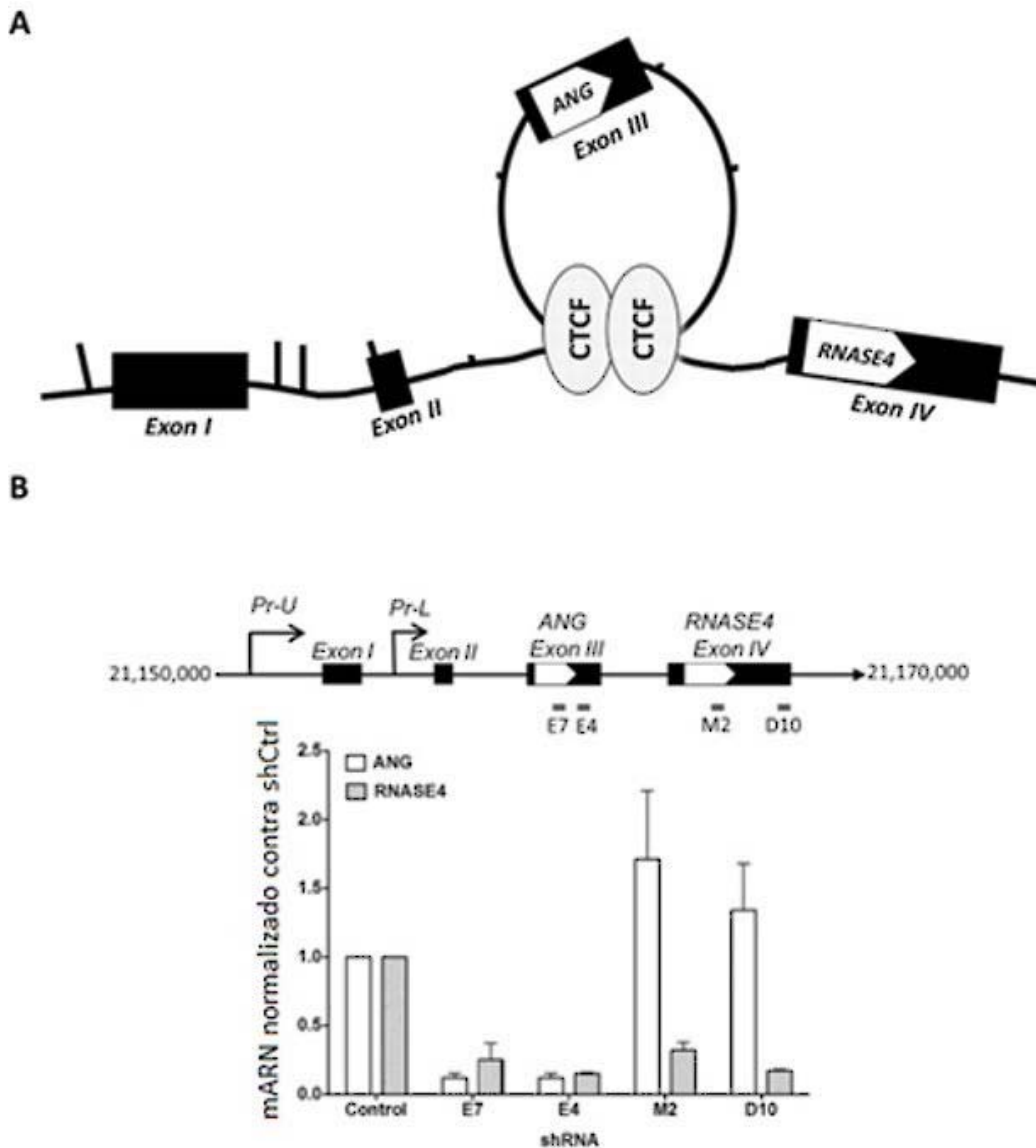
**Figura 34: Disminución del (ARNm) *Ang* desde cepas “129”.**

A) Izquierda: captura de pantalla desde la herramienta de visualización IGV (del inglés “Integrative Genomics Viewer”) de la expresión del gen *Ang/Rnase4* a en el estudio del cuerpo estriado de las cepas fundadoras. Las cepas C57BL/6J y 129S1/SvImJ se ubican en la parte superior. Los polimorfismos se denotan con colores y el modelo del gen se muestra en azul, obtenido desde el servidor UCSC. B) Mismas capturas de pantalla que en (A) a través del experimento de ARN-Seq en MEF *Sall2* silvestres y KO. C) Recuento de las lecturas de secuenciación del (ARNm) *Ang* a partir de hipocampo de ratones C57BL/6J y 129S1/SvImJ normalizados frente al recuentos de *Polr2A* (GSE76567, N = 6, \*\*\* P <0,001, frente a C57BL/6J, Prueba T de Student). E) PCR cuantitativa en tiempo real del ARNm de *Ang* en la corteza procedente de ratones *Sall2* silvestres y KO. Se muestran los niveles de expresión de (ARNm) *Ang* normalizados contra *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el genotipo silvestre (WT) (N = 3; los datos se representan como medias  $\pm$  s.e.m.). \*\*\* P <0.001 frente a WT; Prueba T de Student. Fuente: Elaboración propia.

En base al estudio de la actividad transcripcional del gen *ANG/RNASE4* en células humanas conducido por (Sheng et al. 2014a), se identificó que la proteína CTCF controla la expresión de *ANG* mediante un “loop” intragénico conservado. Se evidenció que al silenciar CTCF, los niveles de *ANG* disminuyeron de forma concomitante con los de *RNASE4*, por lo cual se hipotetizó que la expresión de *ANG* es requerida para la expresión de *RNASE4*. Esto se confirmó con el silenciamiento de *ANG*, el cual disminuyó la expresión de ambos genes. El silenciamiento de *RNASE4* por otra parte, no logró este efecto y sólo silenció el transcrito de *RNASE4*. Lo anterior se ilustra en la figura 35, en donde se concluye que estos genes están altamente correlacionados en expresión, dada la arquitectura transcripcional del gen *ANG/RNASE4*. La correlación de la expresión de *ANG/RNASE4* ya había sido reportado en varios estudios (Egesten et al. 1997; Futami et al. 1997; Dyer and Rosenberg 2005; Yang et al. 2008; Wheeler et al. 2012; Gupta et al. 2016), fenómeno

no observado en nuestro estudio, en donde los niveles transcripcionales entre *Ang* y *Rnase4* no se correlacionan. Esto apoya nuestra hipótesis de que la expresión de *Ang* está controlada por un eQTL presente en cepas murinas derivadas de “129”, pero no en C57BL/6J o CAST/EiJ; las últimas desprovistas de los polimorfismos encontrados en las seis cepas fundadoras con niveles bajos de ARNm de *Ang*.

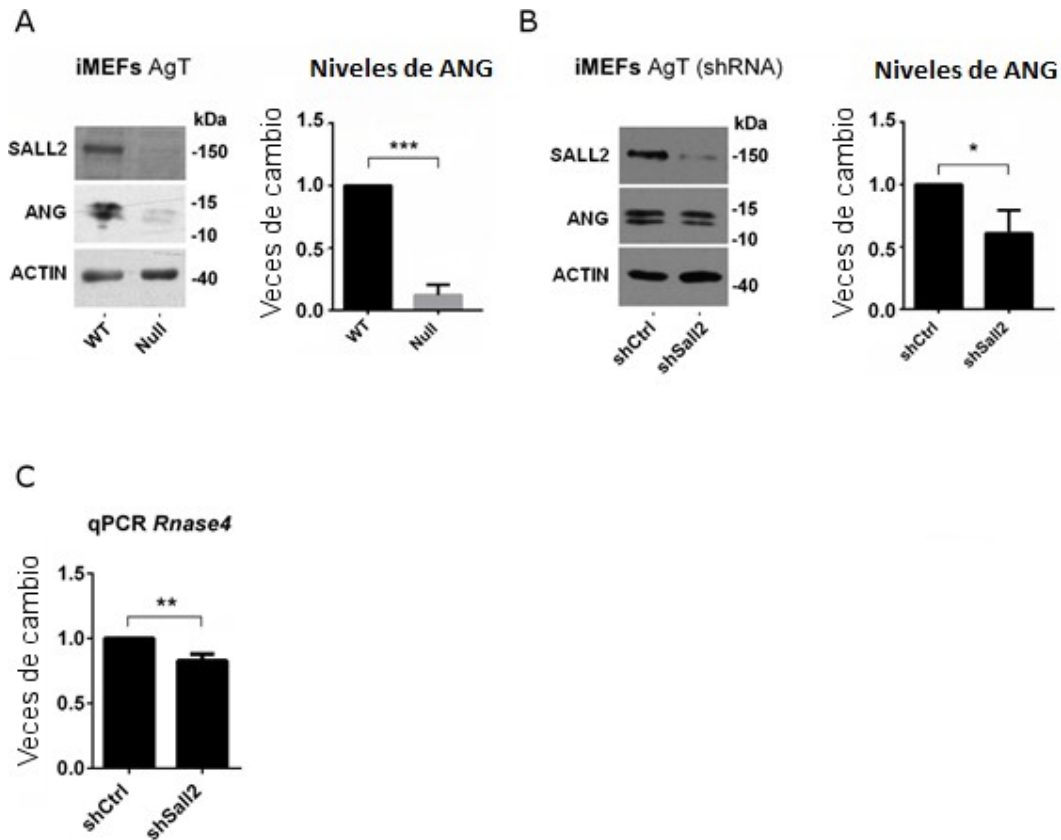




**Figura 35: La expresión de *RNASE4* depende de la expresión de *ANG*.** A) Formación de un “loop” intragénico que controla la expresión de *ANG*. (Sheng et al. 2014a) identificó mediante la técnica 3C que este “loop” es necesario para la transcripción de *ANG* probablemente mediante un mecanismo que involucra pausa transcripcional de la ARN-pol II. B) El silenciamiento de *ANG* en células DU145 de cáncer de próstata disminuye concomitantemente los niveles de *RNASE4*, pero no ocurre a la reversa (véase silenciadores “E7” Y “E4” para *ANG* versus silenciadores “M2” y “D10” para *RNASE4*). La cuantificación de los transcritos fue realizada mediante qPCR. Fuente: Elaboración propia.

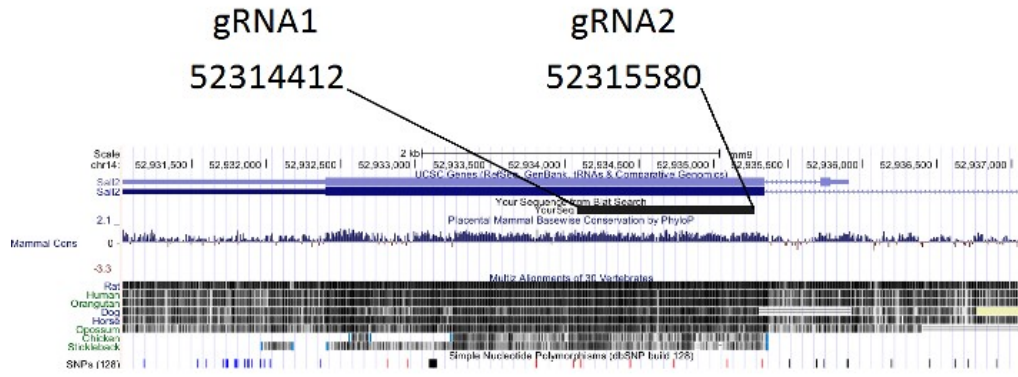


Mediante análisis de “Western Blot”, confirmamos una fuerte disminución de los niveles de proteína ANG en MEF *Sall2* KO, de acuerdo con los bajos niveles del ARNm de *Ang* detectados anteriormente por qPCR (véase figura 36A). Por el contrario, se observa una ligera disminución de los niveles de proteína ANG en las células silenciadas para *Sall2* (véase figura 36B), junto con una leve represión transcripcional del ARNm de *Rnase4* en este modelo (véase figura 36C). De forma similar, un modelo de silenciamiento de *Sall2* mediado por CRISPR/Cas9 en células MEF *Sall2* silvestres mostró una leve represión transcripcional de *Ang* (véase figura 37). Estos resultados sugieren que SALL2 regula transcripcionalmente al gen *Ang/Rnase4*, pero la expresión de *Ang* también se ve afectada por variantes congénicas presentes en el modelo KO para *Sall2*. En síntesis, el gen *Ang* congénico es transcrito en bajos niveles debido a determinantes genéticos heredados de 129P2OlaHsd, de alguna manera enmascarando una potencial regulación transcripcional dependiente de *Sall2*. Por lo tanto, *Ang* podría clasificarse como un potencial gen modificador del fenotipo en MEF *Sall2* KO.

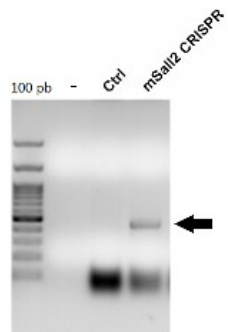


**Figura 36. Regulación del gen *Ang/Rnase4* en células KO y silenciadas para *Sall2*.** A) “Western Blot” representativo para SALL2, ANG y ACTIN en células MEF *Sall2* silvestres y KO. Cuantificación de las bandas de proteína ANG en los gráficos de barras a la derecha (N = 3, \*\*\* P <0,001 versus WT, Prueba T de Student). B) El mismo análisis de (A) en células con el silenciador control (shCtrl) y células silenciadas para *Sall2* (shSall2). Cuantificación de las bandas de proteína ANG en los gráficos de barras a la derecha (N = 3, \*\*\* P <0.001 versus shCtrl; Prueba T de Student). C) Validación de los niveles de *Rnase4* en el modelo de MEF silenciado para *Sall2*. Se muestran los niveles de expresión de *Rnase4* normalizados con *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el silenciador control (shCtrl) (N = 3, los datos se representan como medias  $\pm$  s.e.m.). N = 3, \*\* P <0.01 versus shCtrl; Prueba T de Student. Fuente: Elaboración propia.

# A



# B



C

E2: 52315642



```
Query   CTC- TTGTCTTCTCCCAGGTGATGCTAGCGAGGAAGACCA-----
Sall12  CTCCTTGTCTTCTCCCAGGTGATGCTAGCGAGGAAGACCA-----
          *** *****
Query   -----
Sall12  TGCTGCGCACAATTCTCTGACCCGACCGAATTCCTCGCTCACCAGAACTCATGTTGCACT
Query   -----
Sall12  GACCCACCGGTAATGGTGATAATTGGAGGCCAGGAGAATCCCAGCAACTCTTCAGCCTCC
Query   -----
Sall12  TCTGCGCCCCGACCAGAGGGCCACAGTAGGTCCCAGGTCATGGATACAGAGCACAGCAAT
Query   -----
Sall12  CCCCAGATTCTGGGTCCTCTGGGGCCCCGGATCCCCTTGGGGGCCAGAGCGGAGGGGA
Query   -----
Sall12  GAGGAATCTTCTGGGCAATTCCTGGTCTGCTGCCACAGGTACAGCGGCTGGGGGAGGTGGG
Query   -----
Sall12  GGCCTTATCTTGCCAGTCCCAAGCTGGGAGCAACCCATTACCTCCAGAATCCACTCCT
Query   -----
Sall12  GCACCCCTCCTCCCCACCTCCCCCTCCCCCTCCAGGTGTAGGCAGTGGCCACTTGAAC
Query   -----
Sall12  ATCCTCTGATCTTGAAGAGTTGCGGGTCTGCTGACGAGCGCCAGATTCACCAGATGCAG
Query   -----
Sall12  ATGACTGAACAAATCTGCCGCCAGGTGCTGCTACTTGGCTCCTTGGGGCAGACCGTGGGT
Query   -----
Sall12  GCCCCTGCCAGTCCCTCAGAGCTACCTGGGACAGGGGCTGCCTCTTCCACCAAGCCCCTA
Query   -----
Sall12  CTGCCTCTTTAGTCCCATCAAGCCAGCGCAAACGGCAAGACTGGCATCTTCTCTCT
```

Query -----  
Sal12 TCGTCATCCTCCTCCTCTGGAGCTGAACCGCCTAAGCAGGCTTTCTTCCACCTTTACCAT

Query -----  
Sal12 CCACTGGGATCACAGCATCCCTTCTCTGTAGGAGGGGTGGGCGGAGCCACAAACCCACC

Query -----  
Sal12 CCTGCCCTTCCCCTGCGCTGCCAGGCAGTACGGATCAGCTGATTGCTTACCTCATCTG

Query -----  
Sal12 GCATTCCCAGGCACCACCGGACTCCTGGCAGCTCAGTGTCTTGGGGCAGCAAGGGGCCTT

Query -----  
Sal12 GAGGCTGCTGCCTCCCCAGGGCTCCTGAAGCCAAAGAACGGAAGTGGTGAACGGGCTAT

Query -----  
Sal12 GGGGAAGTGATCAGTTCCTTGGAGAAACCCGGTGGGAAGGCACAAATGCCGCTTTGTGCA

Query -----  
Sal12 AAAGTATTCGGCAGTGACAGCGCCCTGCAGATCCACCTTCGTTCCCACTGGTGAGAGG

Query -----  
Sal12 CCCTATAAGTGCAACGTCTGTGGTAACCGTTTCACAACTCGGGGCAACCTCAAAGTACAT

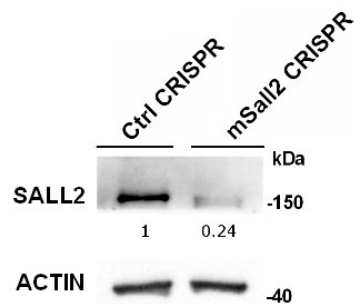
Query -----  
Sal12 TTCACCGGCATCGTGAGAAGTACCCACATGTGCAAATGAATCCACATCCAGTACCGGAG  
\*\*\*\*\*



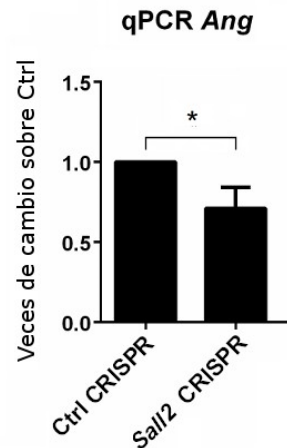
E2: 52314428

Query -----  
Sal12 CACCTAGACTACGTCATCACCAGCAGTGGGCTGCCTTACGGAATGTCTGTGCCACCAGAG  
CACCTAGACTACGTCATCACCAGCAGTGGGCTGCCTTACGGAATGTCTGTGCCACCAGAG  
\*\*\*\*\*

D



E



**Figura 37. Detección del ARNm de *Ang* desde un modelo con la delección del gen murino *Sall2* mediante CRISPR/Cas9.** A) Diseñamos un doble corte de CRISPR para eliminar un segmento del gen *Sall2*. Los dos CRISPR (denotados como gRNA uno y dos) se dirigieron al exón más extenso del gen murino *Sall2* (exón 2). Identificamos la delección deseada desde el ADN genómico de un grupo de células iMEF, electroporadas con los plásmidos para generar la delección en el gen *Sall2* (amplicón a 500 pares de bases en el carril mSall2, indicado con una flecha negra). B) Amplicón de 500 pb generado por PCR que detecta la delección mediante CRISPR como se explicó en (A). C) Alineamiento del resultado de la secuenciación de sanger del amplicón purificado desde el gel en (B) el cual contiene la delección genómica generada en el gen *Sall2* (posición cromosómica 52314428-52315642 en el ensamblaje mm10 del genoma de ratón). Destacamos las secuencias codificantes del exón dos del gen *Sall2* murino en amarillo. D) A) “Western Blot” representativo para SALL2 y ACTIN en células control y silenciadas para *Sall2* mediante los CRISPR mencionados, realizado en iMEFs *Sall2* silvestres. E) Detección del ARNm de *Ang* en el modelo anterior. Se muestran los niveles de expresión de *Ang* normalizados con *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre el CRISPR control (N = 3, \* P < 0.05 versus CRISPR control, Prueba T de Student). Fuente: Elaboración propia.

Con respecto al objetivo 1.1, se ha demostrado que existe un patrón de expresión único de genes en el modelo KO para *Sall2*, el cual refleja genes que son sensibles tanto a la pérdida de *Sall2* como los que son sensibles a la huella congénica que acompaña la mutación para *Sall2*. Además, se identificaron genes congénicos como loci de carácter cuantitativo de expresión, ligados al genotipo KO. La

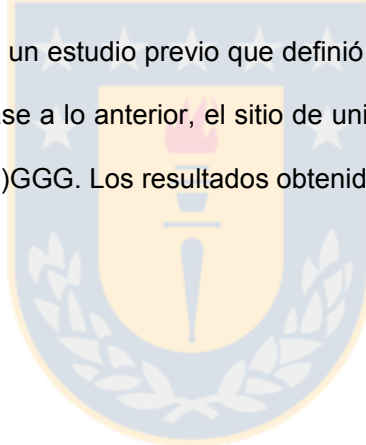
identificación de estos genes es importante, ya que pueden ser potenciales modificadores de fenotipo en animales KO como se ha reportado previamente (Nadeau 2001; Nadeau 2003). En base a lo anterior, se identificó a *Ang* como uno de los genes más diferencialmente expresados en el estudio de ARN-Seq y como un potencial gen modificador en el modelo KO para *Sall2*. Los experimentos de validación cruzada con shRNA, incluyendo CRISPR/Cas9 en células iMEF y HEK293, identificaron a angiogenina como un gen reprimido frente a la pérdida de SALL2. Esto último también es apoyado por la represión transcripcional de *Rnase4* en el modelo de silenciamiento de *Sall2*, gen que es controlado por el mismo promotor que *Ang*. Sin embargo, estos experimentos no explican la baja transcripción de angiogenina frente a *Rnase4* en el modelo KO para *Sall2*, por lo cual el gen *Ang* congénico estaría siendo transcrito en bajos niveles por determinantes genéticos. En base a este último punto, si bien los genes validados pueden ser específicos al fondo genético estudiado sobre la pérdida de *Sall2*, se realizó validación cruzada independiente del fondo genético (como por ejemplo mediante el estudio de microarreglos en fondo 129S6). A continuación, se abordará si los genes sometidos a validación cruzada contienen sitios de unión para SALL2 en sus regiones regulatorias y si angiogenina es un blanco transcripcional de SALL2, al reclutarse en la región promotora de este gen.

## **Objetivo 1: Determinar respuestas transcripcionales dependientes de *Sall2***

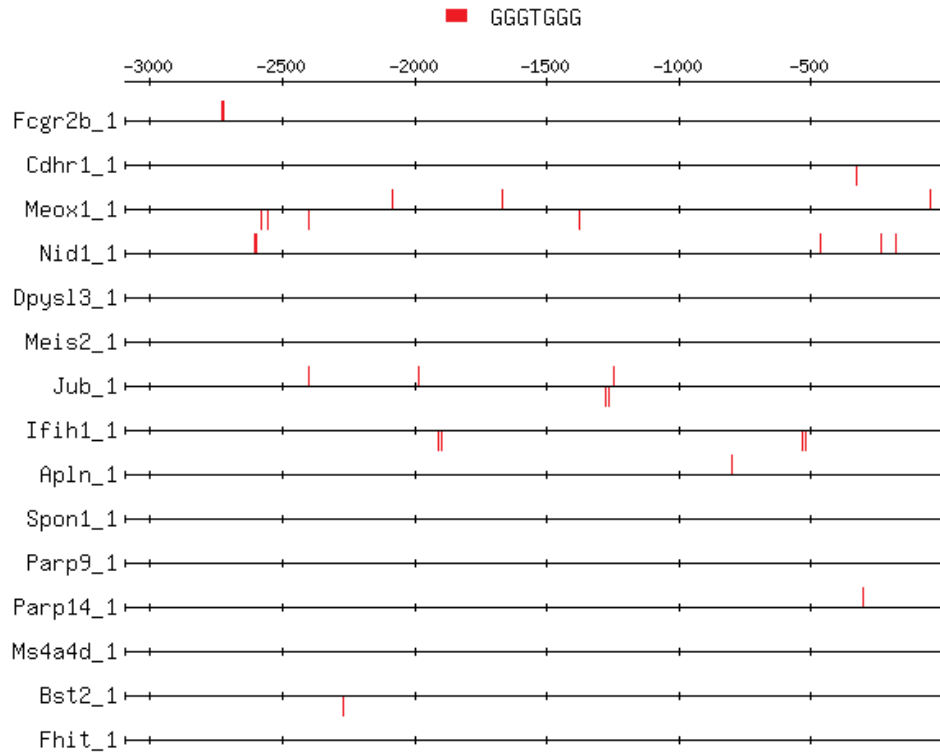
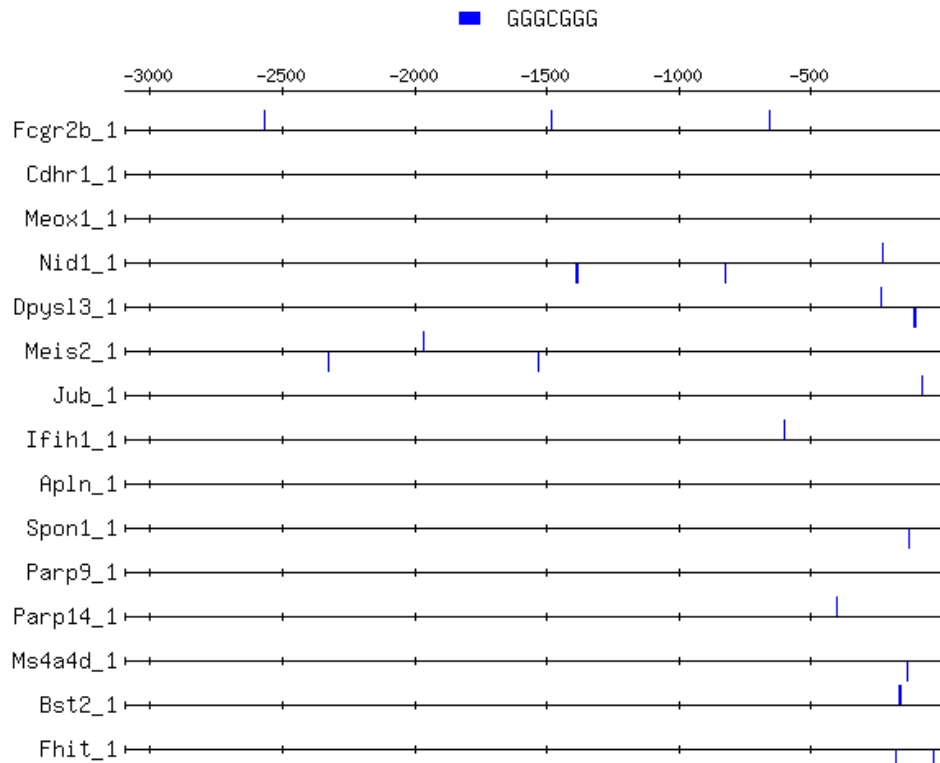
1.2 Investigar el efecto de *Sall2* en regiones promotoras de genes blanco.

### **Búsqueda de sitios de unión de SALL2 en genes diferencialmente expresados**

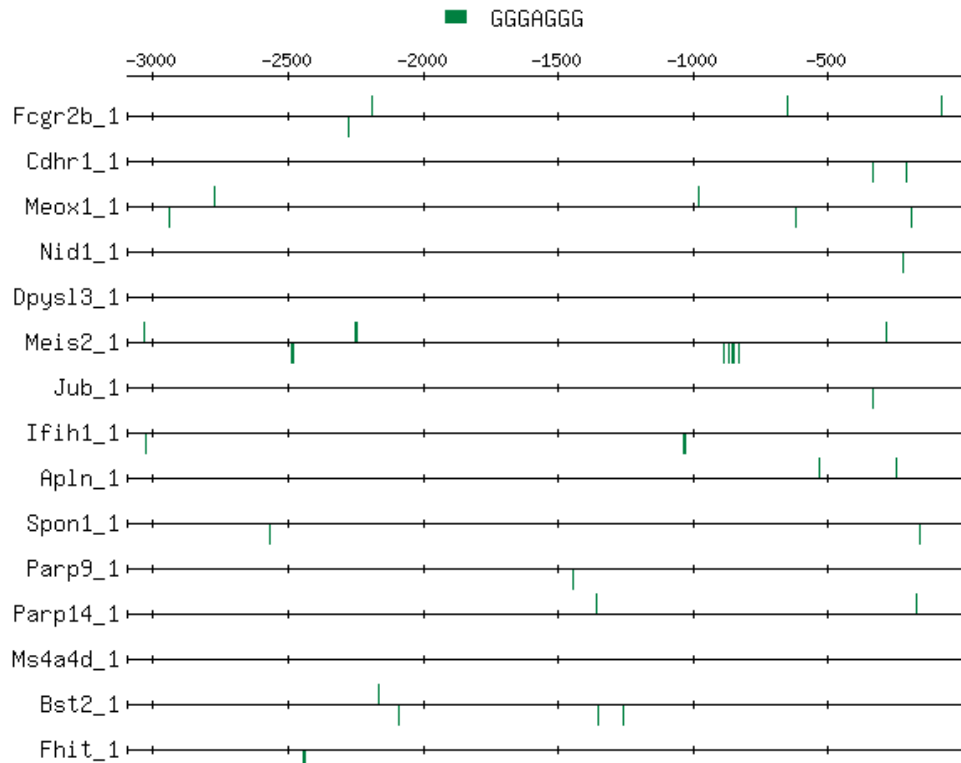
Con el objeto de identificar regiones regulatorias con sitios de unión de SALL2 en genes blanco-putativos, se realizó una búsqueda *in silico* de los sitios de unión para SALL2 en los 15 genes diferencialmente expresados provenientes de la validación cruzada entre el estudio de ARN-Seq y el microarreglo. Para ello, se empleó la base de datos RSAT ([www.rsat.eu/](http://www.rsat.eu/)) que proporciona una serie de programas informáticos diseñados específicamente para la detección de secuencias regulatorias en varios genomas procariontes y eucariontes. Se seleccionaron tres motivos de unión para la proteína SALL2 basados en un estudio previo que definió el sitio de unión de SALL2 *in vitro* (Gu et al. 2011). En base a lo anterior, el sitio de unión seleccionado para SALL2 fue el siguiente: GGG(T|A|C)GGG. Los resultados obtenidos se ilustran en la figura 38.





**A****B**

C



**Figura 38: Búsqueda de sitios de unión de SALL2 en los promotores de genes sensibles a SALL2.** Seleccionamos los promotores de los 15 genes diferencialmente expresados sometidos a validación cruzada entre nuestro estudio de ARN-seq (FDR <0.01, veces de cambio > 0.35) y el estudio de microarreglos en células madre embrionarias (ESC), para analizar sitios de unión de SALL2 en las regiones reguladoras respectivas. Se extrajo la secuencia de cada uno de los genes en el rango de -3000 a cero con respecto al sitio de inicio de la transcripción (TSS) y se buscó el siguiente motivo de unión a SALL2 (GGG(C|T|A)GGG) con el programa RSAT (<http://www.rsat.UE/>). Se muestra el promotor de cada gen con posiciones marcadas de A) motivo de unión GGGTGGG (sitios de unión representado en rojo) B) motivo de unión GGGCGGG (sitios de unión representado en azul) y C) motivo de unión GGGAGGG (sitios de unión representado en verde). Fuente: Elaboración propia.

En base a lo anterior, se identificó que cada gen contiene al menos un sitio de unión para SALL2 en al menos dos de las combinaciones buscadas. Adicionalmente se investigó el promotor del gen angiogenina (*Ang*) murino dado que este gen se validó como gen diferencialmente expresado por PCR en tiempo real tanto en el modelo KO para *Sall2* como en el modelo silenciado. Los sitios de unión de SALL2 encontrados en el promotor del gen angiogenina murino, no conservados en el promotor humano de *ANG* fueron los siguientes:

A

>Ang; rango -154 to 0.

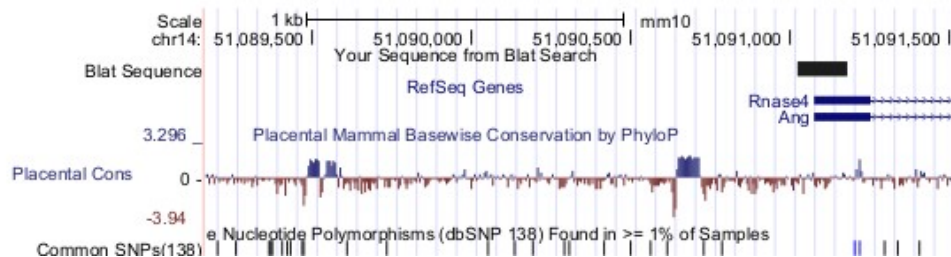
**GGGAGGG**GCATCCTTGAAGAATGGGGTTCGGTCCCAGCCACCGAGGCTGGAACG  
**CCCACCC**TCCACTCCTCTGCACCCCGCCCAATCTGTTTTCGTGGCCAGGCAACGCCGA  
 CCTCACCCATATCGGGGACGAGATTCCAAAGGAAATTTCA

-153: GGGAGGG

-99: CCCACCC

-76: CCCGCC

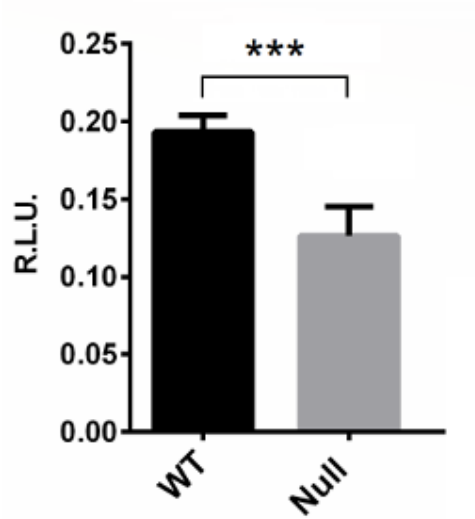
B



**Figura 39. Sitios de unión de SALL2 identificados en el promotor murino que controla el gen *Ang/Rnase4*.** A) La secuencia mostrada corresponde al promotor mínimo de *Ang/Rnase4*. Los sitios de unión para SALL2 están subrayados y se representa las secuencias canónicas de unión con letras en negrita. La posición de cada sitio con respecto al sitio de inicio de la transcripción está indicada debajo de la secuencia. B) Búsqueda “Blat” (<https://genome.ucsc.edu/cgi-bin/hgBlat>) para la secuencia en (A) en el genoma de ratón mm10, visualizada mediante el visor de genomas UCSC (<https://genome.ucsc.edu/>). Como se indica en el “track” de la conservación de organismos placentarios (véase “Placental Cons”), la región promotora buscada no presenta conservación. Fuente: Elaboración propia.

Consistente con una regulación transcripcional del gen *Ang/Rnase4* por SALL2, el promotor *Ang/Rnase4* contiene un grupo de tres sitios de unión para SALL2 alrededor del sitio de inicio de la transcripción (figura 39). Se procedió a clonar 1231 pares de bases río arriba del promotor que controla el gen *Ang/Rnase4* y que contiene estas tres secuencias, para estudiar su actividad transcripcional en células iMEF silvestres o *Sall2* KO. El promotor mostró menos activación en células *Sall2* KO frente a las células *Sall2* silvestres (véase figura 40); niveles consistentes con la leve represión de *Ang/Rnase4* y los niveles de proteína ANG en las células silenciadas para *Sall2*. Sin embargo, la disminución de la actividad transcripcional del promotor en células iMEF *Sall2* KO no recapituló la fuerte disminución de *Ang* observado en estas células, lo cual confirma que la baja transcripción del gen *Ang* congénico se debe a un efecto genético, sobre un efecto transcripcional ligado a SALL2.

### Promotor de *Ang/Rnase4*



**Figura 40: Ensayo de luciferasa con el promotor que controla el gen *Ang/Rnase4* murino electroporado en MEF *Sall2* silvestres y KO.** Se electroporó un promotor consistente en 1321 pares de bases río arriba del sitio de inicio de la transcripción del gen *Ang/Rnase4*, en conjunto con un plásmido que codifica para el gen  $\beta$ -gal. Después de 24 horas, se midieron las actividades de luciferasa y  $\beta$ -galactosidasa. R.L.U = unidades de luminiscencia relativas a  $\beta$ -gal (N = 3, \*\*\* P <0,001 frente a WT; Prueba T de Student). Fuente: Elaboración propia.

### Angiogenina como blanco transcripcional de SALL2

Dada la disponibilidad de anticuerpos contra SALL2 humano para ensayos de inmunoprecipitación de cromatina (ChIP), se decidió realizar ChIP en líneas celulares humanas de cáncer para probar si *ANG* es un blanco transcripcional de SALL2. Lo anterior se justifica en base a que angiogenina responde transcripcionalmente a la pérdida de *Sall2* en células de ratón y humano (MEF y HEK293, respectivamente). Se seleccionó la línea celular k562 (leucemia mieloide crónica) por presentar niveles

detectables de SALL2 acorde con la base de datos human protein atlas (<https://www.proteinatlas.org/ENSG00000165821-SALL2/cell>). Antes de proceder con la CHIP, se realizó un análisis del promotor de *ANG/RNASE4* en búsqueda de sitios conservados para SALL2. Mediante el uso de la base de datos eukaryotic promoter database (EPD, <https://epd.vital-it.ch/index.php>), se extrajo la secuencia promotora del gen *ANG/RNASE4* para realizar búsqueda de motivos de unión de SALL2, tal como se describió anteriormente en la base de datos RSAT. En este procedimiento se identificaron dos sitios de unión a SALL2 ubicados a -1437 y -1291 pares de bases río arriba del sitio de inicio de la transcripción. Estos resultados se presentan en la figura 41



>FP017729 ANG\_1 :+U EU:NC; range -1440 to 0.  
 GATT**GGGAGGG**AAAGCGCAGGAAGTGAGGACATCAGGGAGCCCGGGCCTAAGCCTGGAGGC  
 CCGCCCTCTCCGGCGTTCACAGTTGCGCAGTCAGAATTGTGTGCGTCCCACCGCCTCAGC  
 CCCCAGTTCTCCGACGCCAGGTGCCGGT**CCCACCC**TACACACAAACACACGCATCCAC  
 GCTCGCGAGGCCCTGGGTCCGCGCCAGTGAGGCCAACAGCCCTCGGGCGCCAGAGGA  
 TGAGCAGCCTCGGTATGCCCGAGTTTCAGGGGCGCCTGCTCTGCTTGGGAGCCCTCAGT  
 GAGGCAGGGGAATCCAGATTTAGAAAGGTTGTCTCAGAACCTCGGACGCAGGACTGTCT  
 TACAAGCTGGCTTTGTGACTCTGCAGTAATTACTTCCGGGTATAAAGAGATTTCTTTCT  
 TTCCCGTCTCTGGTTAAGGATGTCTCGAGTAGGGATTTTTAGGAGAGTTGTTTCCGTGCT  
 TCCTTCGATAGCTCAGCTGGTAGAGCGGAGGACTGTAGTACTTAATGTGTGGTCATCCTT  
 AGGTCGCTGGTTCGATTCCGGCTCGAAGGAGACGCTGCTGTTTTGGGGTTTACGGTTGCT  
 CTTCTGTTTGTGCAAAAATCTCTGTATCTTGACGTGATCAAATCTGTTTCTGTCTCTCAC  
 CCTTGTCCGAAGGAAACAGCTTAGCCGCGAGGTGGGCTGCGTTAATCGCGTCTTGAAAAG  
 TTGGTTACCTTTCTGAAAGATAGTAGACCCGAATCGTCCCACCGTTGGGGGAAGGCAC TG  
 TTGTCTTACCGAAGTGTGTTATATTTTGGCTTCTACGTTAAGTCCAAGGCCGGCAGATA  
 TATACGTATCGCGGTCTGTCTGGGGCTCAGTACCCTACAGCCTTGCAACCGAGCTCAG  
 CTCTGGCCCTCGATCTCTATTTAGCCCTCCAGGTTCCCTCCAGGCCCTAGCGGGCTT  
 AAATCGCTCCCGGGCTCAGTGAGACACAAGCATGTGACCTAAAGTCTAAAAAGCTGCCT  
 TCCTTCACCCCATCTGCATCTGGATCGAGGAGAATAAGCCAGAGAATAAGCCACACTTG  
 AAACCTGGTGAAGATGGAAGGAATAACTAAATAAGGCCTAAGGGAATGTCACAACAGAG  
 GAAAAGAAAACAAAACAAAAACCTTCAGAACAGCAGGCGGAAAAGAGAAAATTCAGAGTGA  
 TAGGAAGCACCCTCAGCTACAGTGGCTCGTTGGTCTAGGGGTATGATTCTCGCTTTGGG  
 TCGAGAGGTCCCGGTTCAAATCCCGACGAGCCCTCGTGGCTACTGTTTTTCCACCCC  
 CTTTTGTCAACTACTGAAAAAAGTCTCCGTTAAATTTGAAACACAACTGTCTCGCGAT  
 GGGCTGTCTGCCAAAGAAAGCAGCAGCGAATAAGTACGTGGCAGATGGTGTCTCGACC

-1437: GGGAGGG  
 -1291: CCCACCC (GGGTGGG)



**Figura 41. Sitios de unión de SALL2 identificados en el promotor humano que controla el gen *ANG/RNASE4*.** Se extrajeron 1440 pares de bases río arriba del inicio de la transcripción del promotor *ANG/RNASE4* mediante el uso de la base de datos EPD, como se indicó anteriormente. Los sitios de unión para SALL2 se identificaron con la base de datos RSAT (puestos en negrita) y se destacaron con color amarillo en la secuencia del promotor. La posición de cada sitio con respecto al sitio de inicio de la transcripción está indicada debajo de la secuencia. Fuente: Elaboración propia.

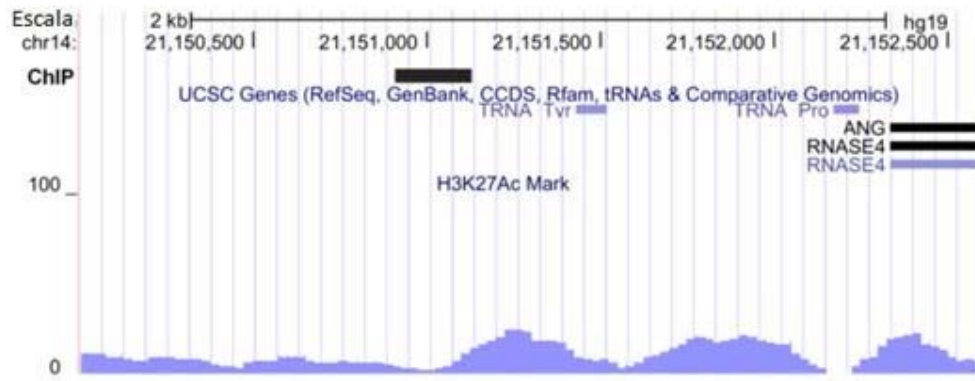
En base a esta identificación, se diseñaron partidores que amplificaran 25 pares de bases río arriba y río debajo del sitio CCCACCC (amplicón denominado “CCCACCC”) para los experimentos de ChIP. Adicionalmente, se incluyeron partidores que hibridan -10000 pares de bases río arriba del sitio de inicio de la transcripción del gen *ANG/RNASE4*; región sin sitios de unión detectables para SALL2, como control de

inespecificidad. Con el objeto de corroborar la regulación transcripcional del gen *ANG/RNASE4* por *SALL2*, se generó un KO total para *SALL2* en células K562 mediante CRISPR/Cas9 (ver plásmidos para CRISPR/Cas9 en materiales y métodos); evaluando si el (ARNm) *ANG* es sensible a la pérdida de *SALL2* en este tipo celular. Los resultados de lo anterior fueron los siguientes:



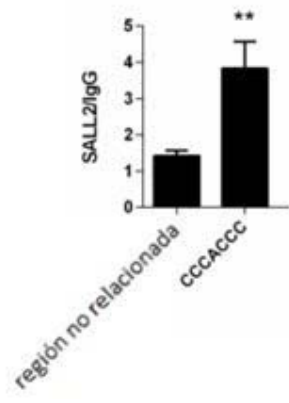


A



B

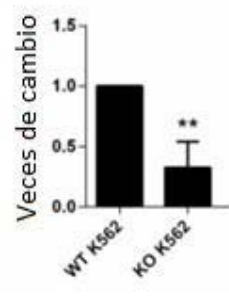
ChIP SALL2 K562



C



qPCR ANG K562



**Figura 42: Identificación del gen *ANG/RNASE4* como un blanco transcripcional de *SALL2* en células K562.** A) Captura de pantalla del servidor UCSC genome browser (<https://genome.ucsc.edu/>) visualizando la región promotora del gen *ANG/RNASE4* con una escala de 2000 pares de bases. Se denota la posición del amplicon “CCCACCC” mediante BLAT en el track denominado “ChIP”. Se incluye la marca de la acetilación de la lisina 27 de la histona H3 como marcador de la actividad transcripcional de este promotor. B) Ensayo de inmunoprecipitación de cromatina (ChIP) en el promotor de *ANG/RNASE4*. Se inmunoprecipitó la cromatina de células K562 con el anticuerpo anti *SALL2* (Bethyl laboratories) o IgG normal de conejo y se detectó con partidores que amplificaban en una región no relacionada o en el sitio de unión “CCCACCC” identificado por bioinformática (N = 3, \*\* P <0,01 versus la región no relacionada, Prueba T de Student). C) Izquierda: “Western Blot” representativo para *SALL2* y *ACTIN* de células K562 silvestres y KO para *SALL2*. El KO fue generado mediante una delección por CRISPR/Cas9 utilizando los vectores *SALL2* y *Jun\_SALL2*, tal como se detalló en la figura 37 en células MEF. Las secuencias utilizadas en los vectores de CRISPR/Cas9 para deletar el gen *SALL2* humano son homólogas para el gen de ratón (*Sall2*). Derecha: PCR cuantitativa en tiempo real de *ANG* en células K562 *SALL2* silvestres (WT K562) y KO (KO K562). Se muestran los niveles de expresión de *ANG* normalizados con *POLR2A*. Expresamos los valores como veces de cambio sobre el genotipo silvestre (WT K562) (N =3, \*\* P <0,01, frente a WT K562, Prueba T de Student). Fuente: Elaboración propia.

Si bien angiogenina se puede clasificar como un nuevo blanco transcripcional del gen *SALL2*, se discutió en el primer objetivo que este gen puede catalogarse a su vez como un loci de carácter cuantitativo de expresión (eQTL) y un potencial gen modificador del fenotipo en el modelo KO para *Sall2*. Como se revisó, este gen es uno de los genes más diferencialmente expresados en la comparación *Sall2* silvestre vs KO y la mayoría de los experimentos estuvieron orientados hacia el estudio y validación de este gen. Este problema ya se había reportado anteriormente en modelos KO congénicos (Schalkwyk et al. 2007; Vanden Berghe et al. 2015; Szabo et al. 2016) y se describe que este tipo de genes confunde el análisis del gen KO en

estudio. En adelante, se demostrará que angiogenina es un gen modificador de al menos dos fenotipos en el modelo KO para *Sall2*.



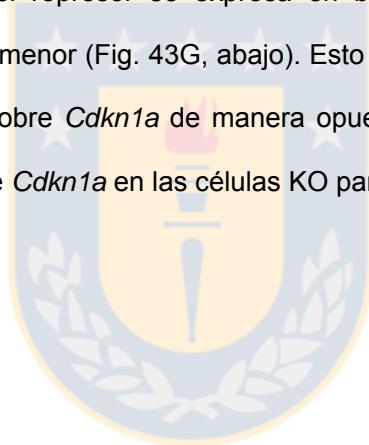
**Objetivo 2: Seleccionar blanco(s) transcripcional(es) de SALL2 asociados a fenotipo canceroso y realizar ensayos funcionales de estos genes y su dependencia de SALL2.**

2.1 Analizar el efecto de pérdida/ganancia de la función de genes asociados con *Sall2* provenientes del objetivo 1 y corroborar efectos celulares.

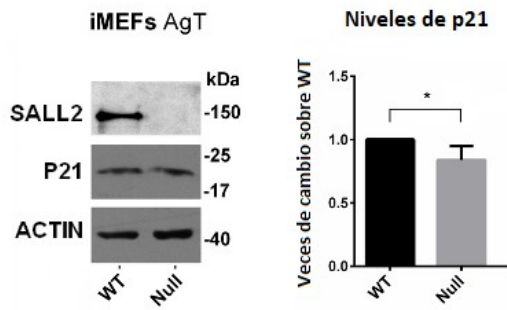
### **Interferencia genética de *Cdkn1a* por angiogenina, un blanco canónico de SALL2**

En ejemplo de cómo los genes congénicos pueden actuar como genes modificadores de fenotipos o afectar la expresión génica en los modelos KO; nos centramos en el gen *Cdkn1a* (p21<sup>CIP/WAF</sup>) que es regulado tanto por *Sall2* como por *Ang*. En primer lugar, se sabe que SALL2 induce la expresión de *Cdkn1a* en neuronas, células epiteliales ováricas y en MEF bajo el tratamiento con estrés genotóxico (Li et al. 2004; Pincheira et al. 2009; Escobar et al. 2015). Por otro lado, ANG regula negativamente a p21<sup>CIP/WAF</sup> a través de la degradación de p53 en células humanas (Sadagopan et al. 2012; Yeo et al. 2017). En nuestro estudio *Cdkn1a* no se detectó como un gen diferencialmente expresado en el análisis de ARN-Seq al comparar MEF *Sall2* silvestre versus KO. Por lo tanto, es probable que se produzca interferencia genética con la transcripción de *Cdkn1a* en ratones *Sall2* KO, debido a la existencia de la huella congénica. Notamos mínimos cambios en los niveles de proteína p21 y en el ARNm de *Cdkn1a* probado por “Western Blot” y qPCR respectivamente, entre MEF *Sall2* silvestre y KO (figura 43A y 43B, respectivamente). Por el contrario, el silenciamiento de *Sall2* mediado por shRNA en MEF *Sall2* silvestre mostró una fuerte disminución de los niveles de proteína p21 y ARNm de *Cdkn1a*, en consistencia con reportes anteriores (Pincheira et al. 2009; Miao et al. 2017) (figura 43C y 43D, respectivamente). Hipotetizamos que la fuerte disminución de ANG explica en parte los niveles inalterados de *Cdkn1a* y proteína p21 en MEF *Sall2* KO. Respecto a lo

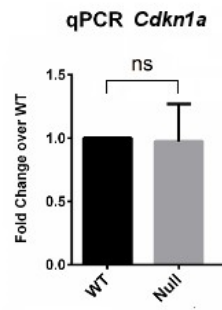
anterior, la expresión de la proteína ANG madura tanto en MEF *Sall2* silvestre y KO regula negativamente los niveles de proteína p21 y ARNm de *Cdkn1a*, suponiendo una regulación transcripcional negativa de *Cdkn1a* por ANG (Fig. 43E y 43F, respectivamente). Sorprendentemente, no vimos este efecto en P53 murino (TRP53), como se describió anteriormente (Sadagopan et al. 2012). En base a esto se propuso un modelo de regulación de *Cdkn1a* basado en SALL2 y ANG como reguladores opuestos. En el modelo *Sall2* silenciado, se observa una leve disminución de los niveles de proteína ANG, por lo que el activador de *Cdkn1a* (SALL2) está en niveles muy bajos mientras que el represor (ANG) está presente, regulando negativamente el (ARNm) *Cdkn1a* (Fig. 43G, arriba). Por el contrario, en el modelo KO para *Sall2*, el activador está ausente y el represor se expresa en bajos niveles, por lo cual la inhibición sobre *Cdkn1a* es menor (Fig. 43G, abajo). Esto sugiere que ANG ejerce una regulación transcripcional sobre *Cdkn1a* de manera opuesta a como lo hace SALL2, interfiriendo la regulación de *Cdkn1a* en las células KO para *Sall2*.



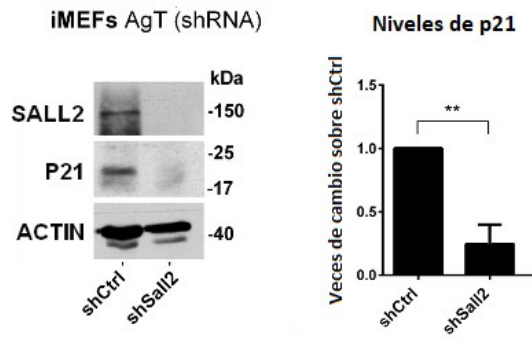
A



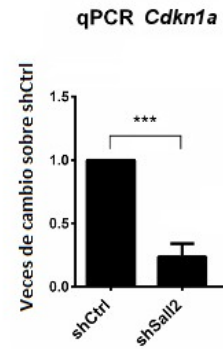
B



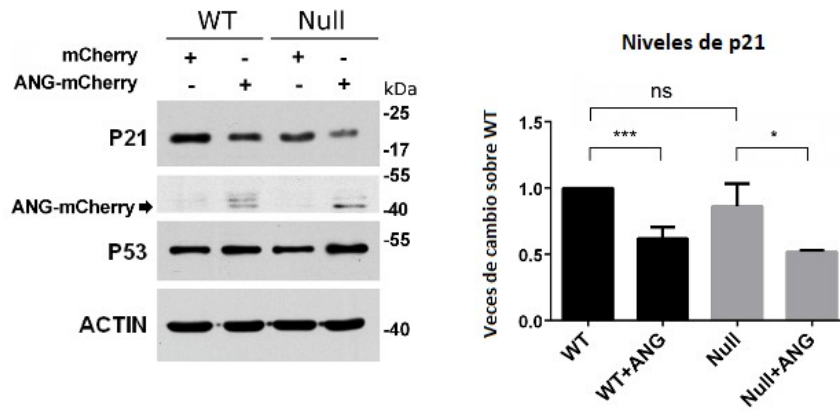
C



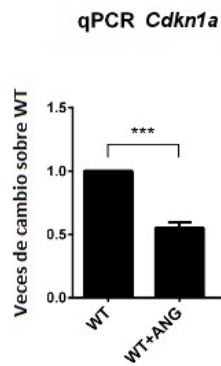
D



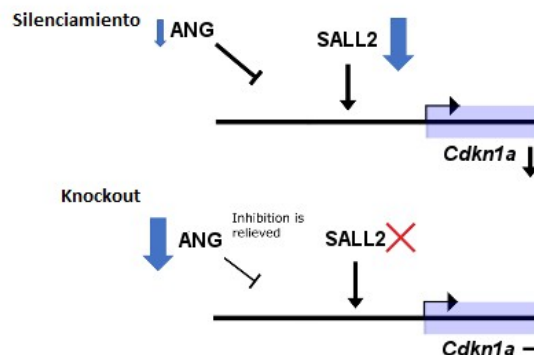
E



F



G



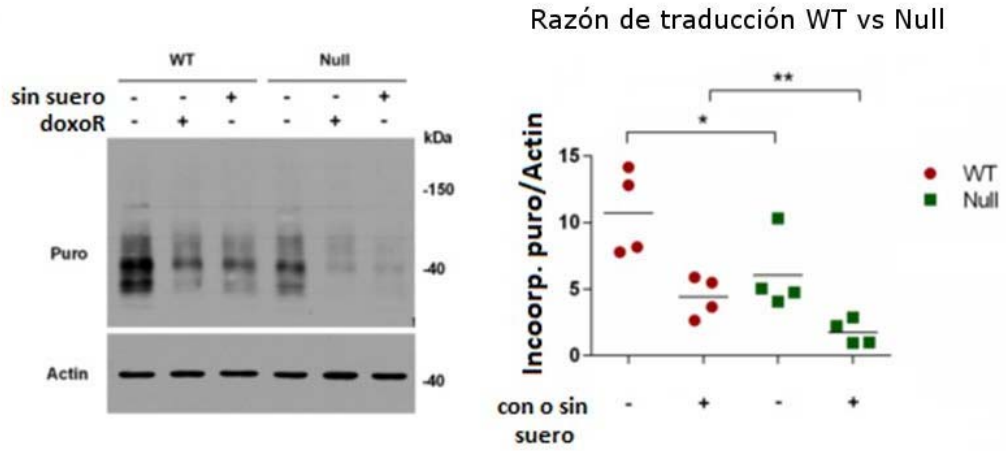
**Figura 43. Interferencia genética de *Cdkn1a* por angiogenina, un blanco canónico de SALL2.** A) “Western Blot” representativo para SALL2, P21 y ACTINA en células MEF *Sall2* silvestres y KO. Se muestra la cuantificación de las bandas de proteína P21 normalizadas contra ACTINA en el gráfico de barras a la derecha (N = 3, \*\*\* P <0,001 versus WT, Prueba T de Student). B) Validación de los niveles de *Cdkn1a* en células *Sall2* silvestres versus KO. Se muestran los niveles de expresión de *Cdkn1a* normalizados contra *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre MEF *Sall2* silvestres (N = 3, \* P <0.05 versus WT, Prueba T de Student). C) El mismo análisis de (A) en células MEF con el silenciador control (shCtrl) y con el silenciador para *Sall2* (shSall2). Se muestra la cuantificación de las bandas de proteína P21 normalizadas contra ACTINA en el gráfico de barras a la derecha (N = 3, \*\*\* P <0,001 frente a shCtrl, Prueba T de Student). D) El mismo análisis que en (B) para *Cdkn1a* en células MEF con el silenciador control (shCtrl) versus el silenciador para *Sall2* (shSall2). Expresamos los valores de expresión como veces de cambio sobre el shCtrl (N = 3, \*\*\* P <0.001 frente a shCtrl; Prueba T de Student). E) “Western Blot” representativo para P21, ANG-mCherry, P53 y ACTINA en iMEFs *Sall2* WT y KO. Donde se indica, se electroporaron plásmidos que codifican para mCherry o ANG-mCherry. Los lisados celulares se obtuvieron después de 16 horas. Cuantificación de las bandas de proteína p21 normalizadas contra ACTINA en el gráfico de barras a la derecha (N = 4 para WT, N = 3 para nulo, \*\*\* P <0.001, \* P <0.05, ns = no significativo versus WT o nulo; Prueba T de Student). F) Niveles de expresión de *Cdkn1a* normalizados contra *Polr2A*. Expresamos los valores de expresión como veces de cambio sobre mCherry (N = 3; los datos se representan como medias  $\pm$  s.d. \*\*\*, P <0.001 versus mCherry; Prueba T de Student). G) Esquema que propone un modelo de regulación de *Cdkn1a* basado en SALL2 y ANG como reguladores opuestos. Arriba: en el modelo con el silenciador para *Sall2*, una leve disminución del represor de *Cdkn1a* (ANG, pequeña flecha azul) y una fuerte disminución del activador (SALL2, flecha azul agrandada) ejercen una represión transcripcional sobre *Cdkn1a*. Abajo: en el modelo KO para *Sall2*, el activador de *Cdkn1a* está ausente (SALL2, cruz roja) y se expresan niveles bajos del represor (ANG, flecha azul agrandada), disminuyendo la represión transcripcional de los niveles de *Cdkn1a*. Fuente: Elaboración propia.

## Otros efectos fenotípicos en MEF *Sall2* KO inherentes a su expresión génica

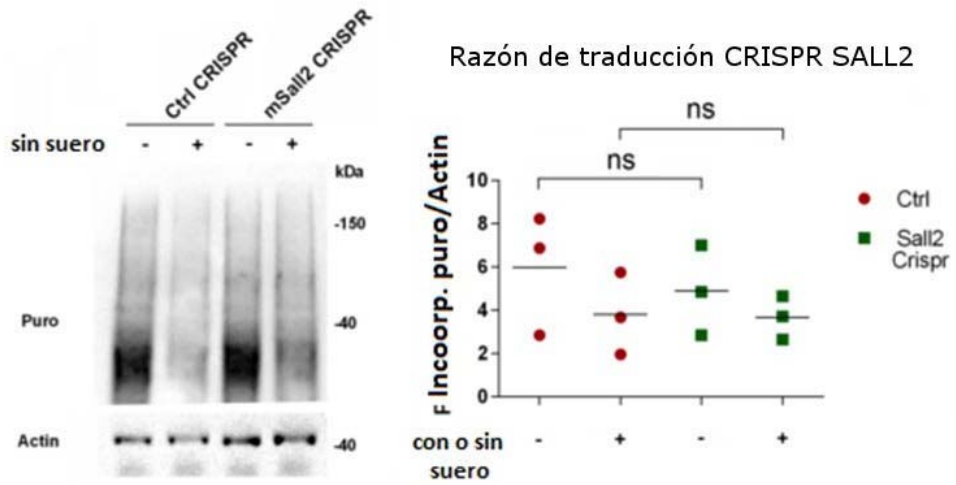
Como se indicó en la figura 26, existe un patrón de expresión génica único en MEFs KO para *Sall2* que incluye la represión transcripcional del precursor ribosomal *Rn45s* (Grozdanov et al. 2003). Dado que el procesamiento de este transcrito policistrónico da origen a tres de las cuatro subunidades ribosomales 40 y 60S en células de ratón (Henras et al. 2015), se investigaron los niveles de traducción en células MEF *Sall2* silvestre y KO, respectivamente. Los niveles de traducción se midieron mediante la técnica SUnSET (Schmidt et al. 2009; Goodman and Hornberger 2013) en células MEF *Sall2* silvestres y KO, bajo tratamiento con estrés genotóxico y privación de suero como condiciones de estrés celular (figura 44). Los niveles de traducción disminuyeron con el estrés en comparación con el control en ambos genotipos, pero las células *Sall2* KO tienen menos proporción de traducción per sé en comparación con la contraparte silvestre (figura 44A, carriles 4,5 y 6 en comparación con 1,2 y 3). Estos resultados van en consonancia con la represión transcripcional del precursor *Rn45s* en células MEF *Sall2* KO. Por el contrario, el modelo de CRISPR para *Sall2*, generado en MEF *Sall2* silvestres, no recapituló los resultados observados en MEF *Sall2* KO, en donde la razón de traducción no varió significativamente entre células CRISPR control y células CRISPR *Sall2*. Esto es esperado, ya que se demostró que el silenciamiento de *Sall2* no ejerce efecto en la transcripción del precursor *Rn45s* (véase figura 26).



A



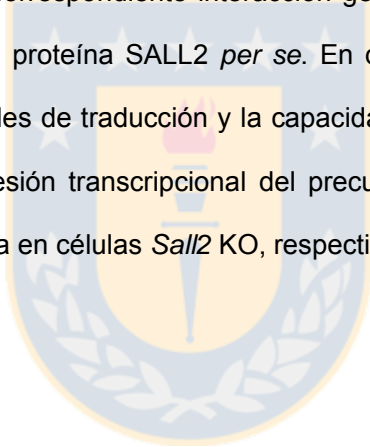
B



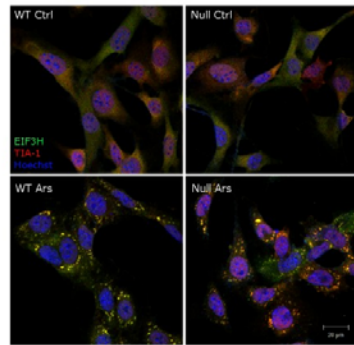
**Figura 44. Razón de traducción en el modelo MEF *Sall2* KO versus el modelo de CRISPR/Cas9 para *Sall2*.** A) Ensayo SUnSET en iMEFs *Sall2* silvestre y KO. *Izquierda:* “Western Blot” representativo para calcular la razón de traducción relativo (Puro=incorporación de puromicina) frente a actina. Donde se indica, las células fueron privadas de suero durante 9 horas o se trataron con doxorubicina (1  $\mu$ M) durante 9 horas. Después de esto las células tratadas con cada condición se pulsaron con puromicina (10  $\mu$ M) durante 10 minutos y se prepararon lisados para inmunodetección. *Derecha:* cuantificación de la señal de puromicina normalizada con actina en el gráfico a la derecha (N = 4, \* P <0,05, \*\* P <0,01, frente a WT, prueba T de Student). B) Ensayo SUnSET en iMEFs CRISPR control y CRISPR para *Sall2*. *Izquierda:* “Western Blot” representativo para calcular la razón de traducción relativo (Puro=incorporación de puromicina) frente a actina. Donde se indica, las células fueron privadas de suero durante 9 horas. Después de esto las células tratadas con cada condición se pulsaron con puromicina (10  $\mu$ M) durante 10 minutos y se prepararon lisados para inmunodetección. *Derecha:* cuantificación de la señal de puromicina normalizada con actina en el gráfico a la derecha (N = 3, ns = no significativo frente a Control CRISPR, prueba T de Student). Fuente: Elaboración propia.

Se ha descrito que angiogenina es una ribonucleasa secretada que cliva tRNA para iniciar una respuesta a estrés en células de mamífero (Emara et al. 2010). Estos tRNA a su vez inician el ensamblaje de los gránulos de estrés, ribonucleoproteínas de mRNA mensajero no traducidos que se forman a partir de mRNA estancados en la iniciación de la traducción bajo estrés (Protter and Parker 2016). Se ha demostrado que el ensamblaje de los gránulos de estrés puede ser inducido *in vitro*, mediante el tratamiento con arsenito de sodio (McEwen et al. 2005). En base a esto, se realizaron experimentos con arsenito de sodio para evaluar la capacidad de la formación de gránulos de estrés en células MEF *Sall2* silvestres y KO, dada la expresión diferencial de angiogenina en el modelo KO para *Sall2*. La detección de los gránulos de estrés se llevó a cabo mediante inmunocitoquímica, detectando la colocación de las proteínas TIA-1 e EIF3H como marcadores de gránulos (Kedersha et al. 1999; Kedersha and Anderson 2007). La formación de gránulos de estrés está disminuida en

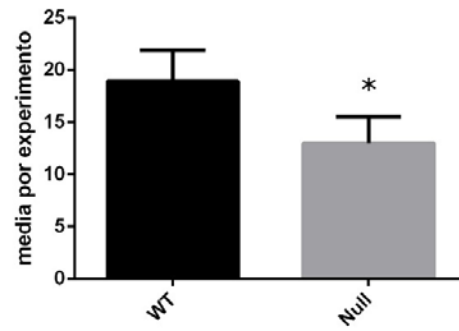
iMEF *Sall2* KO (figura 45A), en acuerdo con los bajos niveles de angiogenina en estas células. Al contrario, la formación de gránulos de estrés no está significativamente disminuida en el modelo de silenciamiento de *Sall2* (figura 45B). Con el objeto de comprobar si el restablecimiento de angiogenina puede recuperar la capacidad de formación de gránulos de estrés en células iMEF *Sall2* KO, se electroporó el plásmido mCherry como control o mCherry-ANG que expresa la forma madura de la proteína ANG, en células iMEF *Sall2* KO. La formación de gránulos de estrés se recuperó con la electroporación de angiogenina, como se muestra en la figura 45C. En síntesis, los experimentos de traducción y formación de gránulos de estrés revelan que los efectos fenotípicos revisados en células iMEF *Sall2* KO son ejercidos por la huella congénica en el cromosoma 14 (y la correspondiente interacción genética de esta huella con el fondo genético) y no por la proteína SALL2 *per se*. En demostración de este último punto, se revisaron los niveles de traducción y la capacidad de formación de gránulos de estrés debido a la represión transcripcional del precursor ribosomal *Rn45s* y los bajos niveles de angiogenina en células *Sall2* KO, respectivamente.



A

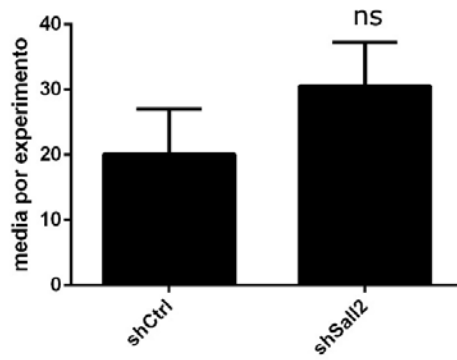


Gránulos de estrés *Sall2* WT vs Null

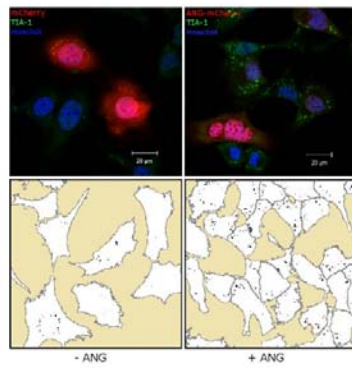


B

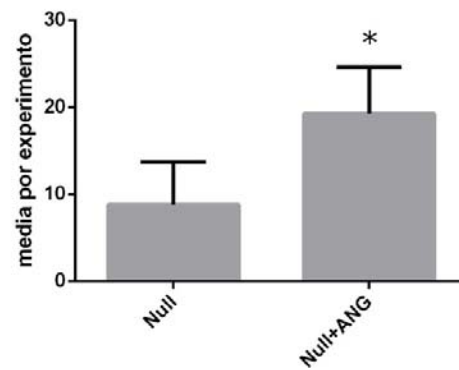
Gránulos de estrés shCtrl vs shSall2



C



Recuperación de gránulos con ANG



**Figura 45: Capacidad de formación de gránulos de estrés (SG) en células *Sall2* KO versus el modelo silenciado para *Sall2*.**

A) Ensayo de ensamblaje de gránulos de estrés (SG) en células iMEF *Sall2* silvestre y KO. Las células de ambos genotipos se trataron o no con arsenito de sodio 500  $\mu$ M durante 45 minutos (Ars) y se realizó Inmunocitoquímica usando anti TIA-1 (rojo), anti EIF3H (verde) y Hoechst para la tinción de ácido nucleico. Barra blanca = 20  $\mu$ M. Cuantificación de SG por célula en los gráficos a la izquierda. Se consideró para el análisis la media de cada experimento (N = 4, \* P <0,05, frente a WT, prueba T de Student). D) Mismo análisis que (A) pero usando células iMEF con el silenciador control (shCtrl) o el silenciador para *Sall2* (sh*Sall2*) (N = 3, ns= no significativo versus shCtrl, prueba T de Student). E) Mismo análisis que (A) pero usando células iMEF *Sall2* KO electroporadas con mCherry o ANG-mCherry (como se indica con letras rojas) y se sometidas a tratamiento con arsenito de sodio. Abajo de las imágenes del microscopio, se representan los límites de las células con SG (puntos negros) para visualizar el efecto ANG en la formación de los gránulos. (N = 4, \* P <0,05, frente a mCherry; prueba T de Student). Fuente: Elaboración propia.

En síntesis, como prueba de la interferencia genética en el modelo KO para *Sall2* se revisaron fenotipos que son exclusivos para este modelo, los cuales no pudieron ser recapitulados en pruebas de pérdida de función de *Sall2*. Sin embargo, estos fenotipos pudieron ser recuperados mediante el rescate con un gen congénico. Tal fue el caso de angiogenina (*Ang*), un gen congénico que rescató el fenotipo observado con  $p21^{CIP1/WAF1}$  y los gránulos de estrés en células MEF *Sall2* KO. Las diferencias fenotípicas se pueden resumir en la siguiente tabla:

Característica	Knockout	Silenciamiento
Variantes ligadas al KO	+	-
Angiogenina ( <i>Ang</i> )	↓↓↓	↓
P21 ( <i>Cdkn1a</i> )	-	↓↓↓
lncRNA DEGs	↓↓↓ (Rn45s, Rmrp, Rpph1)	↓ (Rpph1)
Niveles de traducción	↓↓↓	-
Ensamblaje de gránulos de estrés	↓	-

**Tabla 7: Diferencias en los fenotipos caracterizados entre el modelo *Sall2* “knockout” y el modelo de silenciamiento para *Sall2*.** Se indica en la tabla un símbolo (+) o flechas si se detecta un fenotipo en el modelo indicado. En el caso que este fenotipo no se detecte en los modelos caracterizados se denota un símbolo (-). El modelo “knockout” para *Sall2* contiene variantes ligadas al genotipo knockout, bajos niveles de angiogenina, bajos niveles de lncRNAs, traducción y ensamblaje de gránulos de estrés, mientras que el modelo de silenciamiento de *Sall2* discrepa del modelo anterior al no poseer variantes ligadas y sólo una baja modesta en angiogenina y Rpph1, entre los fenotipos mencionados. Al contrario, el modelo de silenciamiento de *Sall2* expresa niveles bajos de P21, fenotipo no observado en el modelo *Sall2* knockout. Fuente: Elaboración propia.

## V. DISCUSIÓN

Como se ha descrito anteriormente, la genética de los ratones KO es un tema crítico que hay que considerar al hacer conclusiones determinadas de la expresión génica y los fenotipos asociados con un gen en estudio. Como se revisó aquí y por otros, el origen de las células madre embrionarias (ESC) en experimentos de recombinación homóloga, las posteriores rondas de retrocruzamiento con el genoma receptor C57BL/6 (subcepa N o J) y el mantenimiento de la mutación en estudio pueden variar de modelo KO a KO, creando un fondo y determinantes genéticos únicos en el modelo donde se estudia un gen de interés. Estos sesgos, incluida la fijación de variantes nuevas debidas a la deriva genética, deberían controlarse en experimentos que hacen uso de ratones KO con el fin de obtener conclusiones correctas y reproducibilidad a través de los estudios. Como ejemplo de lo anterior, un caso revisado en la literatura son los modelos KO para el factor de transcripción IX (*Nfix*) (Pekarik and Izpisua Belmonte 2008). En un primer modelo KO para este gen desarrollado por un grupo en la Universidad de Freiburg (Driller et al. 2007), se reportó como fenotipo de los animales KO la hidrocefalia, agenesia parcial del cuerpo caloso y deformidades espinales que se debieron a un retraso en la osificación de los cuerpos vertebrales. También se registraron defectos femorales y los animales KO generalmente morían alrededor del día postnatal (P) 21-28 debido a estos defectos. Al contrario, un segundo grupo en la Universidad de Nueva York en Buffalo generó un modelo KO para *Nfix* en donde no se reportaron alteraciones tan severas (Campbell et al. 2008). No se observó agenesia callosal en estos animales ni muerte al estadio P21-P28, pero presentaron deformidades en la corteza cingulada y en el hipocampo. Se discute que tales diferencias en los dos modelos KO se debería al uso de distintas células madre, ya que el modelo KO de Freiburg se realizó en células madre embrionarias derivadas de la cepa 129S1/Sv y el modelo KO de Buffalo se realizó en células madre derivadas de la cepa 129S4/SvJae. Ambas cepas son derivadas de la

línea 129S1/SvImJ, pero divergieron en subcepas distintas por la deriva génica. También se discute que el modelo KO de Freiburg no es congénico y contiene una parte importante del genoma de la célula madre de origen, ya que la agenesia del cuerpo caloso y la hidrocefalia son fenotipos reportados de las cepas derivadas de 129 (Wahlsten 1982). Como se mencionó en la introducción, este fenómeno también fue reportado en el estudio realizado por (Bohm et al. 2008) con el modelo KO para *Sall2*, en donde se reportaron defectos en el cierre del tubo neural en los animales *Sall2* KO con los fondos genéticos 129SV/J, 129SV/J-CD1 y 129SV/J-NZW. Por el contrario, estos defectos no fueron registrados en los animales *Sall2* KO con los fondos genéticos mixtos 129SV/J-C57BL/6, 129SV/J-DBA ni en el modelo KO generado por (Sato et al. 2003) con fondo genético mixto 129P2OlaHsd-C57BL/6 (utilizado en esta tesis). Por ende, los fenotipos asociados con los modelos KO congénicos y los que se mantienen en fondo genético mixto pueden estar fuertemente ligados al fondo genético en donde se mantiene la mutación.

Por otro lado, se ha reportado que el uso de animales KO congénicos puede adicionar fenotipos nuevos a la mutación en estudio. En el estudio de (Vanden Berghe et al. 2015) se ha reportado que la mutación inactivante del gen *Casp11* (proveniente de cepas 129) reside a menos de cinco centimorgan en 60 modelos KO distintos y atribuye resistencia a la muerte por inyección de lipopolisacárido bacteriano (LPS), entre otros fenotipos. Dado que las retrocruzas de los ratones KO con el genoma receptor C57BL/6 puede reemplazar el genoma de 129 por C57BL/6 a lo largo de los cromosomas, el genoma de la célula madre donde se realizó la recombinación homóloga persiste en las regiones circundantes a la mutación en estudio. Respecto a lo anterior, en la práctica se ha registrado que al menos 10 centimorgan de la célula madre embrionaria pueden flanquear el gen en estudio (Lusis et al. 2007), problema que se ha denominado en la literatura como “huella congénica” (Schalkwyk et al. 2007). Este fenómeno se reconoció en la mitad de los años noventa por distintos



grupos (Gerlai 1996; Crawley et al. 1997; Simpson et al. 1997) pero el problema fue ampliamente ignorado en el uso y estudio de los ratones KO. Distintos reportes han demostrado que las variantes derivadas de cepas 129 en regiones codificantes causan fenotipos o respuestas celulares aberrantes. Como ejemplos de lo anterior, se ha estudiado que la señalización del receptor DAP12 (*Tyrobp*) involucrada en la respuesta inmune de las células asesinas naturales (del inglés “natural killer”) es deficiente en células 129/J, 129Sv y ratones que contengan la región genómica donde localiza *Tyrobp* derivado de cepas 129. La vía de señalización de *Tyrobp* carece de variantes inactivantes, pero se demostró que el origen congénico de estas proteínas es suficiente para causar una señalización aberrante (McVicar et al. 2002). Ejemplos similares a este caso se han demostrado con el estudio del filamento intermedio plakinina (*Bfsp2*) y los efectos en el citoesqueleto del ojo (Alizadeh et al. 2004; Sandilands et al. 2004); la señalización dependiente de *Disc1* y desórdenes neurológicos asociados (Gomez-Sintes et al. 2014); la mutación del factor de licenciamento *Ctd1* y los efectos en la replicación del ADN (Coulombe et al. 2013), entre otros. En base a estos argumentos, es importante detectar la introgresión del genoma de 129 y otro tipo de cepas en la cepa receptora C57BL/6 a modo de supervisión genética de los modelos KO y otro tipo de animales transgénicos. Como prueba de concepto de lo anterior, identificamos desde el estudio de ARN-Seq en el modelo KO para *Hnrnpd/AUF-1* la presencia de la mutación espontánea para el gen *Casp4*, como se describe en el estudio de (Vanden Berghe et al. 2015). Adicionalmente, detectamos que esta mutación tiene una introgresión diferencial a través de las réplicas biológicas secuenciadas, perjudicando potencialmente las conclusiones de este estudio.

Respecto a lo anterior, se ha considerado realizar control genético de los modelos murinos en uso, dado que la contaminación genética en ratón ha sido reportada como un problema frecuente (Foster and Balk 1982; Benavides 1999; Nitzki

et al. 2007). En el marco de esta tesis, se propuso realizar la supervisión genética de los modelos KO/genéticamente modificados a nivel de genoma completo mediante el uso de una estrategia basada en el llamado de variantes desde datos de ARN-Seq. El procedimiento presentado se basó en detectar la introgresión de variantes en los genes que están siendo expresados en la célula/tejido en estudio, e identificar las variantes derivadas de la célula madre embrionaria (ESC) en genomas recipientes tales como C57BL/6. Se demostró que la estrategia fue eficiente en identificar modelos KO congénicos o de fondo genético mixto en cinco estudios que utilizaban modelos KO, incluyendo nuestro experimento de secuenciación con el modelo KO para *Sall2* (*Mepc2*, *Gtf2ird1*, *Stc1*, *Itch*, *Hnrnpd/AUF-1* y *Sall2*, respectivamente). Para fundamentar esto, se diseñó un protocolo computacional que permite obtener tanto las variantes congénicas como las fijadas por la deriva génica y visualizar estos polimorfismos a nivel de genoma completo. Lo anterior se logró mediante un programa que implementó la visualización de la frecuencia de variantes basado en las coordenadas cromosómicas de ratón (véase `genotype_variants.R`). Mediante el uso de este programa se pudo estimar visualmente tanto la distribución como la extensión de las variantes introgresadas en el genoma de los modelos KO, incluyendo la extensión de la huella congénica.

Posterior a esto, se caracterizó la naturaleza de las variantes introgresadas mediante el uso del programa Variant Effect Predictor (VEP) y las variantes encontradas se clasificaron como nuevas o existentes. Se evidenció que las variantes nuevas son muy variables entre los modelos KO, incluso sobrepasando las variantes introgresadas desde la célula madre embrionaria. Esto representa otro sesgo, ya que demostró que este tipo de variantes en conjunto con las que presentan disrupción del marco de lectura (del inglés “frameshift”) se correlacionan en número, imponiéndose de esta forma nuevos fondos genéticos para la mutación en estudio. Por tanto, el

método propuesto para la detección de estas variantes puede emplearse como control de la deriva génica en modelos genéticamente modificados.

Adicionalmente a los procedimientos anteriores, se implementó un protocolo computacional para identificar el conjunto de variantes ligadas al genotipo KO, lo que se reconoce en la literatura como la huella congénica. Teniendo en cuenta todas las fuentes mencionadas de variabilidad genética, determinamos que el número y la naturaleza de las variantes que conforman la huella congénica en cada modelo KO depende del modelo en sí mismo y es poco probable que se pueda predecir *a priori*. Como se muestra en la figura 20, el número de genes congénicos varía de KO a KO y en un caso en particular se encontraron genes congénicos fuera del cromosoma en donde se realizó la recombinación homóloga (véase el modelo KO para *Itch*). La razón de esta observación en el modelo KO para *Itch* podría implicar el mantenimiento de ambos genotipos por separado, la falta de retrocruzas con el genoma receptor C57BL/6 o la presencia de contaminación genética con otras cepas murinas, probando que nuestro protocolo es útil para estos efectos. Como se discutió al comienzo, la presencia de genes congénicos en modelos KO puede confundir los fenotipos y la expresión génica atribuida a la mutación en estudio, por lo tanto, es importante detectar este tipo de genes de forma adecuada a través de los estudios con animales genéticamente modificados. Los paneles que genotifican polimorfismos de un solo nucleótido (del inglés single nucleotide polymorphism genotyping panels) han demostrado ser útiles para realizar esta tarea, pero no detectan ploidía ni el número total de variantes introgresadas, incluyendo las variantes fijadas por la deriva génica.

Después de obtener las variantes ligadas al genotipo KO por sustracción con la contraparte silvestre, se propuso emplear embriones heterocigotos para confirmar la extensión de la huella congénica. Lo anterior se basa en que la mayoría de las variantes congénicas cercanas a la mutación en estudio son homocigóticas, por tanto, las variantes heterocigóticas de un embrión heterocigoto pueden ayudar a discriminar

este tipo de variantes. El empleo de este método fue exitoso para detectar variantes en modelo KO para *Sall2* como se muestra en la figura 20, confirmando más de 35 megabases de huella congénica en este modelo. Sin embargo, se pueden encontrar escenarios más complejos de ploidía, tal como en el caso del modelo KO para *Stc1*, también ubicado en el cromosoma 14. En este modelo, casi la mitad de la huella congénica tiene ploidía heterocigota y está introgresada con distinto genotipo entre los embriones KO. En base a esto, se evidencia que es importante discriminar la ploidía de la huella congénica en los modelos KO y determinar las réplicas biológicas que contienen su introgresión. Este último punto es particularmente importante en términos de reproducibilidad de los estudios con animales KO.

Posterior a la identificación computacional de las variantes introgresadas en los modelos KO, procedimos a demostrar experimentalmente la influencia de la huella congénica en el modelo KO para *Sall2*, en términos de expresión génica. Al silenciar *Sall2* dentro del mismo fondo genético (MEF *Sall2* silvestre en fondo mixto C57BL/6-129P2OlaHsd) varios genes diferencialmente expresados en la comparación *Sall2* silvestre versus KO no pudieron ser confirmados en los experimentos de silenciamiento en MEFs. Esto último se explicaría debido a la influencia del genoma 129P2OlaHsd introgresado en el cromosoma 14 de las células *Sall2* KO, el cual generaría un patrón único de expresión génica debido a la presencia de los genes congénicos. Este fenómeno también ocurrió a la inversa, en donde se pudieron detectar genes diferencialmente expresados en el modelo de silenciamiento de *Sall2*, pero no en el modelo KO. Lo último es el caso del gen *Pnp*, gen ubicado en la huella congénica del modelo KO para *Sall2* que está diferencialmente expresado en el modelo de silenciamiento de *Sall2*, pero no en el modelo KO para *Sall2*. Como un sesgo adicional causado por la variación genética, no se pudo validar *Pnp* como un gen diferencialmente expresado en el modelo KO para *Sall2* debido a polimorfismos en la región de hibridación de los partidores (véase figura 27). En cuanto a los niveles

de expresión de este gen en el modelo KO para *Sall2*, la naturaleza congénica de *Pnp* probablemente explicaría por qué no es un gen diferencialmente expresado en las células MEF *Sall2* KO. De hecho, los polimorfismos en las regiones reguladoras de genes pueden modificar su regulación transcripcional, creando o eliminando sitio de unión para factores de transcripción u otros elementos reguladores transcripcionales como se ha revisado en la literatura (De Gobbi et al. 2006; Menendez et al. 2006; Wang et al. 2006). Sin embargo, pudimos realizar validación cruzada de los genes sensibles a la pérdida o ganancia de función de *Sall2* identificados a partir del experimento de ARN-Seq, mediante el uso de un análisis de microarreglos en células madre embrionarias que sobreexpresaban SALL2. De este análisis se identificaron 37 genes de los cuales 15 presentaron veces de cambio de expresión opuestos en ambos estudios, como se explicó en la figura 24, evidenciando esta cohorte de genes como sensibles a la pérdida/ganancia de SALL2. De la cohorte anterior, dos genes diferencialmente expresados fueron validados por PCR cuantitativa en tiempo real con el objeto de validar experimentalmente estas tendencias. De forma similar, se validaron genes diferencialmente expresados en un modelo de células HEK293 que carecían de la isoforma E1A de SALL2, confirmando las tendencias vistas en el modelo de silenciamiento de *Sall2*. En síntesis, utilizando el modelo KO para *Sall2* como ejemplo, se confirma que la huella congénica y su interacción con el fondo genético influye en el análisis del transcriptoma de los modelos KO, lo que requiere siempre la validación cruzada de los genes diferencialmente expresados mediante enfoques adicionales.

Dado que los genes congénicos se comportan como genes diferencialmente expresados, definimos una estrategia para identificar genes que se expresan dependiendo del genotipo de las variantes en la región congénica. Como se discutió anteriormente, esto permitiría descubrir potenciales genes modificadores en el genoma introgresado de la célula madre con respecto al genoma receptor C57BL/6J. Con el

objeto de abordar computacionalmente este problema, aplicamos la estrategia eQTL (cribado de loci de carácter cuantitativo de expresión) en los datos de secuenciación de ARN-Seq de las MEF *Sall2* silvestre, heterocigotos y KO. En esta búsqueda, también se incluyó el tratamiento con doxorubicina de estas células como una perturbación ambiental de los nucleosomas (Yang et al. 2013). Con este procedimiento descubrimos cuatro eQTLs candidatos como posibles genes modificadores tanto en el control como en el tratamiento con doxorubicina. Uno de ellos corresponde al gen angiogenina, que se transcribe en bajos niveles en MEF *Sall2* KO en comparación con la contraparte silvestre. Nuestros estudios revelaron que la baja transcripción del (ARNm) *Ang* en el modelo KO para *Sall2* es probablemente causada por componentes genéticos heredados de la célula madre, pero también por la ausencia del factor de transcripción SALL2. Lo anterior se resolvió en primera instancia investigando los niveles del gen *Rnase4*, un gen que comparte el mismo promotor que *Ang*, tanto en nuestro experimento de secuenciación como en datos públicamente disponibles. Se mostró que el ARNm de *Ang* estaba fuertemente disminuido con respecto a los niveles del ARNm de *Rnase4* en dichos estudios, descartando así que la disminución de *Ang* se deba a una regulación transcripcional a nivel de promotor. Lo anterior se probó mediante la electroporación del promotor de *Ang/Rnase4* en células MEF *Sall2* silvestres y KO, en donde no se recapituló la represión observada para *Ang* en células *Sall2* KO, pero sí la leve represión detectada para *Rnase4* en el modelo silenciado de *Sall2*. En base a este último punto, se desarrolló el objetivo 1.2 de esta tesis, al probar que este gen es un blanco transcripcional de SALL2 en células de leucemia mieloide crónica K562, confirmando que la proteína SALL2 es reclutada *in vivo* en el promotor de *ANG/RNASE4*. Lo anterior tendría relevancia en el contexto de cáncer, ya que *ANG* es un blanco terapéutico de la angiogénesis tumoral (Piccoli et al. 1998; Miyake et al. 2015). Estas evidencias apoyan que *Ang* es un blanco transcripcional de SALL2, pero la regulación de este gen en el modelo KO para *Sall2* se controla por determinantes

genéticos ajenos a la regulación mediada por SALL2. Esto último clasificaría a *Ang* como un potencial gen modificador del fenotipo en este modelo KO.

En base a estos hallazgos, se procedió a demostrar que el gen *Ang* congénico se comporta como un gen modificador del fenotipo en el modelo KO para *Sall2*. Esto se probó en primera instancia mediante el análisis de la interferencia genética de un blanco conocido de SALL2, *Cdkn1a*. Los experimentos de silenciamiento de *Sall2* mostraron que los niveles de proteína p21 y el ARNm de *Cdkn1a* estaban fuertemente disminuidos en las células silenciadas para *Sall2*, un efecto no observado en el modelo KO para *Sall2*, probablemente debido a la interferencia genética de gen(es) congénico(s). Dado que reportes previos en células humanas demuestran que ANG regula negativamente a p21, realizamos experimentos restableciendo angiogenina en células MEF *Sall2* KO, con el objeto de comprobar si la introducción de angiogenina disminuía los niveles de *Cdkn1a* y la proteína p21 en estas células. Efectivamente se comprobó que la introducción de ANG disminuye los niveles de proteína p21 y ARNm de *Cdkn1a* en ambos genotipos, lo que sugiere otro nivel de regulación en el gen *Cdkn1a*, basado en SALL2 como activador y ANG como represor. Sin embargo, no se descarta que los niveles de *Cdkn1a* en el modelo KO para *Sall2* puedan ser consecuencia de un efecto poligénico y no solo a los bajos niveles de ANG detectados en las células *Sall2* KO.

Se abordó el objetivo 2 evaluando fenotipos asociados al patrón de expresión génica en el modelo KO para *Sall2*. Como se mostró, el precursor transcripcional *Rn45s* y *Ang* están reprimidos, esperando observar diferencias en la traducción de proteínas ente células *Sall2* silvestre versus KO. Se ha descrito que *Ang* regula positivamente la transcripción de este precursor, por lo cual se explica que este precursor esté disminuido en células MEF *Sall2* KO (Xu et al. 2003; Tsuji et al. 2005; Sheng et al. 2014b). Los niveles de traducción están disminuidos en MEF *Sall2* KO con respecto a la contraparte silvestre, consistente con la represión transcripcional de

estos genes. Con el objeto de comprobar si un modelo silenciado para *Sall2* puede exhibir este fenotipo, se llevó a cabo el silenciamiento mediado por CRISPR/Cas9 en células MEF inmortalizadas *Sall2* KO, debido a que la resistencia del casete de puromicina en el modelo silenciado por shRNA pudo confundir el análisis de incorporación de puromicina. Se observó que la traducción no está alterada significativamente en las células con el silenciamiento mediado por CRISPR/Cas9 de *Sall2*, lo que corrobora que este fenotipo es exclusivo del modelo KO para *Sall2*. Además, el ensamblaje de gránulos de estrés está disminuido en células KO para *Sall2*. Nuevamente este fenotipo no se pudo corroborar en células silenciadas para *Sall2*, pero sí mediante el restablecimiento de ANG en células KO para *Sall2*. Por lo tanto, es probable que los fenotipos revisados en las células MEF KO para *Sall2* sean ejercidos por genes congénicos, que pueden interferir con las funciones normales de este gen. Hay que indicar que esto no implica que la carencia de *Sall2* no genere fenotipo alguno, sino que es probable que los fenotipos revisados sean exclusivos del modelo KO estudiado y por ende no puedan ser replicados mediante ensayos de pérdida de *Sall2* en la contraparte silvestre. Debido a las limitaciones mencionadas en ratones congénicos KO, las conclusiones a nivel de expresión génica y a nivel fenotípico pueden ser engañosas, pero también estos modelos pueden abrir la oportunidad de descubrir nuevas interacciones genéticas, tales como el eje *Ang/p21* demostrado en esta tesis. La selección de una cepa apropiada y la prevención de las fuentes genéticas de variabilidad mediante una adecuada supervisión genética son aspectos críticos de cualquier proyecto de investigación *in vivo*. Incluso por razones técnicas, los polimorfismos en los genes expresados deberían detectarse si se ejecuta validación por PCR cuantitativa en tiempo real, dado el caso de la expresión del gen *Pnp* en células *Sall2* KO. De acuerdo con (Vanden Berghe et al. 2015), también hacemos hincapié en generar modelos KO mediante tecnologías de ingeniería del genoma tales como CRISPR/Cas9, con el objeto de obtener conclusiones fiables sobre los efectos del gen en estudio o al menos corroborar los fenotipos observados



con varios modelos. Nuestro trabajo resume un esquema experimental para la supervisión genética de modelos genéticamente modificados, la cual incluye la caracterización del fondo genético a nivel de genoma completo, la validación cruzada de genes sensibles al gen de estudio, la identificación de potenciales genes modificadores y otras estrategias para comprender los resultados fenotípicos en estudios con animales KO (véase figura anexa I).

## VI. CONCLUSIONES GENERALES

- I) El fondo genético de los ratones KO y genéticamente modificados es un factor crítico por considerar cuando se asocia fenotipos y/o expresión génica a la mutación en estudio. Por tanto, una adecuada supervisión genética de estos modelos es requerida.
- II) Diseñamos un protocolo computacional basado en datos de ARN-Seq para detectar la introgresión del genoma de la célula madre embrionaria utilizada en la generación de los modelos genéticamente modificados. Para estos efectos, se confeccionaron gráficos de genoma completo, que permiten visualizar el grado de introgresión genética de las variantes derivadas de las células madre y la deriva génica. Adicionalmente se formularon protocolos para detectar la presencia de la huella congénica con su correspondiente ploidía.
- III) Mediante el uso del modelo KO para *Sall2* y la validación cruzada con otros modelos murinos incluyendo células humanas, pudimos identificar una cohorte de genes que son sensibles a la pérdida y ganancia de función de SALL2. Adicionalmente identificamos a angiogenina (*ANG*) como un blanco transcripcional de SALL2 en células humanas.

- IV) Identificamos a angiogenina murino (*Ang*) como un gen sensible a la pérdida de *SALL2*, pero a la vez como un gen modificador de fenotipos (del inglés “Modifier gene”) en el modelo KO para *Sall2*. Esto se demostró en el análisis de fenotipos en la formación de los gránulos de estrés y traducción de proteínas revisados en el objetivo 2.



## VII. AGRADECIMIENTOS

Quisiera agradecer por el desarrollo de esta tesis a los siguientes financiamientos:

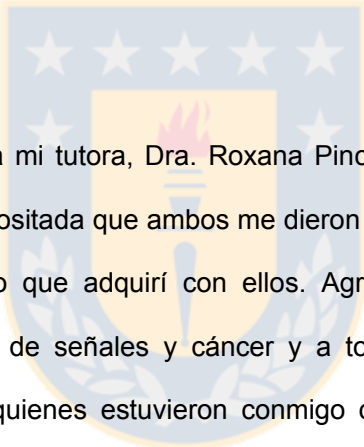
Financiamiento FONDECYT 1151031

Beca CONICYT para estudios de doctorado en Chile número 21130230 año 2013

Gastos operacionales asociados a la beca CONICYT N° 21130230

Financiamiento de la pasantía doctoral asociada a la beca CONICYT N° 21130230

Dirección de Postgrado de la Universidad de Concepción por el financiamiento de congresos y el último semestre del doctorado, ambas acciones gestionadas por el Dr. Juan Pablo Henríquez.



Luego, quisiera agradecer a mi tutora, Dra. Roxana Pincheira y su esposo, Dr. Ariel Castro, por la confianza depositada que ambos me dieron para realizar este proyecto y por el aprendizaje científico que adquirí con ellos. Agradezco a mi laboratorio, el laboratorio de transducción de señales y cáncer y a todos sus integrantes, tantos pasados como presentes, quienes estuvieron conmigo durante todo mi proceso de formación, desde pregrado hasta ahora. Agradezco además al Dr. Sylvain Marcellini y a la Dra. Teresa Caprile por las extensas discusiones científicas brindadas.

Agradezco a mis amigos, en especial a Francisco y Antonia, los cuales siempre estuvieron a mi lado y fueron un gran soporte. Gracias a su apoyo, sin duda cumplí este proceso.

Finalmente quisiera agradecer a mis padres y mi hermana, que siempre me apoyaron de forma incondicional y lo siguen haciendo.

## VIII. BIBLIOGRAFÍA

- Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Eberhard C et al. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**: W3-W10.
- Alagaratnam S, Lind GE, Kraggerud SM, Lothe RA, Skotheim RI. 2011. The testicular germ cell tumour transcriptome. *Int J Androl* **34**: e133-150; discussion e150-131.
- Alizadeh A, Clark J, Seeberger T, Hess J, Blankenship T, FitzGerald PG. 2004. Characterization of a mutation in the lens-specific CP49 in the 129 strain of mouse. *Invest Ophthalmol Vis Sci* **45**: 884-891.
- Atchley WR, Fitch WM. 1991. Gene trees and the origins of inbred strains of mice. *Science* **254**: 554-558.
- Auerbach W, Dunmore JH, Fairchild-Huntress V, Fang Q, Auerbach AB, Huszar D, Joyner AL. 2000. Establishment and chimera analysis of 129/SvEv- and C57BL/6-derived mouse embryonic stem cell lines. *Biotechniques* **29**: 1024-1028, 1030, 1032.
- Bandera CA, Takahashi H, Behbakht K, Liu PC, LiVolsi VA, Benjamin I, Morgan MA, King SA, Rubin SC, Boyd J. 1997. Deletion mapping of two potential chromosome 14 tumor suppressor gene loci in ovarian carcinoma. *Cancer Res* **57**: 513-515.
- Baribault H, Penner J, Iozzo RV, Wilson-Heiner M. 1994. Colorectal hyperplasia and inflammation in keratin 8-deficient FVB/N mice. *Genes Dev* **8**: 2964-2973.
- Benavides FJ. 1999. Genetic Contamination of an SJL/J Mouse Colony: Rapid Detection by PCR-based Microsatellite Analysis. *Contemp Top Lab Anim Sci* **38**: 54-55.
- Bittner JJ, Huesby RA, Visscher MB, Ball ZB, Smith F. 1944. Mammary Cancer and Mammary Structure in Inbred Stocks of Mice and Their Hybrids. *Science* **99**: 83-85.
- Bohm J, Buck A, Borozdin W, Mannan AU, Matysiak-Scholze U, Adham I, Schulz-Schaeffer W, Floss T, Wurst W, Kohlhase J et al. 2008. Sall1, sall2, and sall4 are required for neural tube closure in mice. *Am J Pathol* **173**: 1455-1463.
- Brilliant MH, Ching A, Nakatsu Y, Eicher EM. 1994. The original pink-eyed dilution mutation (p) arose in Asiatic mice: implications for the H4 minor histocompatibility antigen, Myod1 regulation and the origin of inbred strains. *Genetics* **138**: 203-211.
- Campbell CE, Piper M, Plachez C, Yeh YT, Baizer JS, Osinski JM, Litwack ED, Richards LJ, Gronostajski RM. 2008. The transcription factor Nfix is essential for normal brain development. *BMC Dev Biol* **8**: 52.
- Casellas J, Medrano JF. 2008. Within-generation mutation variance for litter size in inbred mice. *Genetics* **179**: 2147-2155.
- Clarke LL, Grubb BR, Gabriel SE, Smithies O, Koller BH, Boucher RC. 1992. Defective epithelial chloride transport in a gene-targeted mouse model of cystic fibrosis. *Science* **257**: 1125-1128.
- Coulombe P, Gregoire D, Tsanov N, Mechali M. 2013. A spontaneous Cdt1 mutation in 129 mouse strains reveals a regulatory domain restraining replication licensing. *Nat Commun* **4**: 2065.
- Crawley JN, Belknap JK, Collins A, Crabbe JC, Frankel W, Henderson N, Hitzemann RJ, Maxson SC, Miner LL, Silva AJ et al. 1997. Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl)* **132**: 107-124.
- Crusio WE. 2004. Flanking gene and genetic background problems in genetically manipulated mice. *Biol Psychiatry* **56**: 381-385.
- de Celis JF, Barrio R. 2009. Regulation and function of Spalt proteins during animal development. *Int J Dev Biol* **53**: 1385-1398.

- De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, Vernimmen D, Yoshinaga Y, de Jong P et al. 2006. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. *Science* **312**: 1215-1217.
- Dimont E, Shi J, Kirchner R, Hide W. 2015. edgeRun: an R package for sensitive, functionally relevant differential expression discovery using an unconditional exact test. *Bioinformatics* **31**: 2589-2590.
- Doetschman T. 2009. Influence of genetic background on genetically engineered mouse phenotypes. *Methods Mol Biol* **530**: 423-433.
- Doran AG, Wong K, Flint J, Adams DJ, Hunter KW, Keane TM. 2016. Deep genome sequencing and variation analysis of 13 inbred mouse strains defines candidate phenotypic alleles, private variation and homozygous truncating mutations. *Genome Biol* **17**: 167.
- Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics* **148**: 1667-1686.
- Driller K, Pagenstecher A, Uhl M, Omran H, Berlis A, Grunder A, Sippel AE. 2007. Nuclear factor I X deficiency causes brain malformation and severe skeletal defects. *Mol Cell Biol* **27**: 3855-3867.
- Dyer KD, Rosenberg HF. 2005. The mouse RNase 4 and RNase 5/ang 1 locus utilizes dual promoters for tissue-specific expression. *Nucleic Acids Res* **33**: 1077-1086.
- Egesten A, Dyer KD, Batten D, Domachowske JB, Rosenberg HF. 1997. Ribonucleases and host defense: identification, localization and gene expression in adherent monocytes in vitro. *Biochim Biophys Acta* **1358**: 255-260.
- Eisener-Dorman AF, Lawrence DA, Bolivar VJ. 2009. Cautionary insights on knockout mouse studies: the gene or not the gene? *Brain Behav Immun* **23**: 318-324.
- Emara MM, Ivanov P, Hickman T, Dawra N, Tisdale S, Kedersha N, Hu GF, Anderson P. 2010. Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J Biol Chem* **285**: 10959-10968.
- Escobar D, Hepp MI, Farkas C, Campos T, Sodir NM, Morales M, Alvarez CI, Swigart L, Evan GI, Gutierrez JL et al. 2015. Sall2 is required for proapoptotic Noxa expression and genotoxic stress-induced apoptosis by doxorubicin. *Cell Death Dis* **6**: e1816.
- Esteller M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**: 5427-5440.
- Estilo CL, P Oc, Talbot S, Socci ND, Carlson DL, Ghossein R, Williams T, Yonekawa Y, Ramanathan Y, Boyle JO et al. 2009. Oral tongue cancer gene expression profiling: Identification of novel potential prognosticators by oligonucleotide microarray analysis. *BMC Cancer* **9**: 11.
- Foster HL, Balk MW. 1982. Histocompatibility and isoenzyme differences in commercially supplied BALB/c mice: a reply. *Science* **217**: 381.
- Futami J, Tsushima Y, Murato Y, Tada H, Sasaki J, Seno M, Yamada H. 1997. Tissue-specific expression of pancreatic-type RNases and RNase inhibitor in humans. *DNA Cell Biol* **16**: 413-419.
- Gerlai R. 1996. Gene-targeting studies of mammalian behavior: is it the mutation or the background genotype? *Trends Neurosci* **19**: 177-181.
- Gerlai R. 2001. Gene targeting: technical confounds and potential solutions in behavioral brain research. *Behav Brain Res* **125**: 13-21.
- Gerlai R. 2016. Gene Targeting Using Homologous Recombination in Embryonic Stem Cells: The Future for Behavior Genetics? *Front Genet* **7**: 43.
- Gomez-Sintes R, Kvajjo M, Gogos JA, Lucas JJ. 2014. Mice with a naturally occurring DISC1 mutation display a broad spectrum of behaviors associated to psychiatric disorders. *Front Behav Neurosci* **8**: 253.
- Goodman CA, Hornberger TA. 2013. Measuring protein synthesis with SUNSET: a valid alternative to traditional techniques? *Exerc Sport Sci Rev* **41**: 107-115.

- Grozdanov P, Georgiev O, Karagyozev L. 2003. Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer. *Genomics* **82**: 637-643.
- Gu H, Li D, Sung CK, Yim H, Troke P, Benjamin T. 2011. DNA-binding and regulatory properties of the transcription factor and putative tumor suppressor p150(Sal2). *Biochim Biophys Acta* **1809**: 276-283.
- Guan C, Ye C, Yang X, Gao J. 2010. A review of current large-scale mouse knockout efforts. *Genesis* **48**: 73-85.
- Gupta SK, Haigh BJ, Wheeler TT. 2016. Abundance of RNase4 and RNase5 mRNA and protein in host defence related tissues and secretions in cattle. *Biochem Biophys Res Commun* **478**: 261-267.
- Henras AK, Plisson-Chastang C, O'Donohue MF, Chakraborty A, Gleizes PE. 2015. An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley Interdiscip Rev RNA* **6**: 225-242.
- Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, Bild A, Iversen ES, Liao M, Chen CM et al. 2003. Gene expression predictors of breast cancer outcomes. *Lancet* **361**: 1590-1596.
- Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res* **14**: 528-538.
- Kallapur S, Ormsby I, Doetschman T. 1999. Strain dependency of TGFbeta1 function during embryogenesis. *Mol Reprod Dev* **52**: 341-349.
- Kayagaki N, Warming S, Lamkanfi M, Vande Walle L, Louie S, Dong J, Newton K, Qu Y, Liu J, Heldens S et al. 2011. Non-canonical inflammasome activation targets caspase-11. *Nature* **479**: 117-121.
- Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M et al. 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**: 289-294.
- Kedersha N, Anderson P. 2007. Mammalian stress granules and processing bodies. *Methods Enzymol* **431**: 61-81.
- Kedersha NL, Gupta M, Li W, Miller I, Anderson P. 1999. RNA-binding proteins TIA-1 and TIAR link the phosphorylation of eIF-2 alpha to the assembly of mammalian stress granules. *J Cell Biol* **147**: 1431-1442.
- Kelberman D, Islam L, Lakowski J, Bacchelli C, Chanudet E, Lescai F, Patel A, Stupka E, Buck A, Wolf S et al. 2014. Mutation of SALL2 causes recessive ocular coloboma in humans and mice. *Hum Mol Genet* **23**: 2511-2526.
- Kenzelmann Broz D, Spano Mello S, Biegging KT, Jiang D, Dusek RL, Brady CA, Sidow A, Attardi LD. 2013. Global genomic profiling reveals an extensive p53-regulated autophagy program contributing to key p53 responses. *Genes Dev* **27**: 1016-1031.
- Kohlhase J, Schuh R, Dowe G, Kuhnlein RP, Jackle H, Schroeder B, Schulz-Schaeffer W, Kretzschmar HA, Kohler A, Muller U et al. 1996. Isolation, characterization, and organ-specific expression of two novel human zinc finger genes related to the Drosophila gene spalt. *Genomics* **38**: 291-298.
- Koks S, Soomets U, Paya-Cano JL, Fernandes C, Luuk H, Plaas M, Terasmaa A, Tillmann V, Noormets K, Vasar E et al. 2009. Wfs1 gene deletion causes growth retardation in mice and interferes with the growth hormone pathway. *Physiol Genomics* **37**: 249-259.
- Kuhnlein RP, Frommer G, Friedrich M, Gonzalez-Gaitan M, Weber A, Wagner-Bernholz JF, Gehring WJ, Jackle H, Schuh R. 1994. spalt encodes an evolutionarily conserved zinc finger protein of novel structure which provides homeotic gene function in the head and tail region of the Drosophila embryo. *EMBO J* **13**: 168-179.
- Kuhnlein RP, Schuh R. 1996. Dual function of the region-specific homeotic gene spalt during Drosophila tracheal system development. *Development* **122**: 2215-2223.
- Lagarrigue S, Martin L, Hormozdiari F, Roux PF, Pan C, van Nas A, Demeure O, Cantor R, Ghazalpour A, Eskin E et al. 2013. Analysis of allele-specific expression in mouse liver

- by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. *Genetics* **195**: 1157-1166.
- Li CM, Guo M, Borczuk A, Powell CA, Wei M, Thaker HM, Friedman R, Klein U, Tycko B. 2002. Gene expression in Wilms' tumor mimics the earliest committed stage in the metanephric mesenchymal-epithelial transition. *Am J Pathol* **160**: 2181-2190.
- Li D, Dower K, Ma Y, Tian Y, Benjamin TL. 2001. A tumor host range selection procedure identifies p150(sal2) as a target of polyoma virus large T antigen. *Proc Natl Acad Sci U S A* **98**: 14619-14624.
- Li D, Tian Y, Ma Y, Benjamin T. 2004. p150(Sal2) is a p53-independent regulator of p21(WAF1/CIP). *Mol Cell Biol* **24**: 3885-3893.
- Lilue J, Muller UB, Steinfeldt T, Howard JC. 2013. Reciprocal virulence and resistance polymorphism in the relationship between *Toxoplasma gondii* and the house mouse. *Elife* **2**: e01298.
- Limaye A, Hall B, Kulkarni AB. 2009. Manipulation of mouse embryonic stem cells for knockout mouse production. *Curr Protoc Cell Biol* **Chapter 19**: Unit 19 13 19 13 11-24.
- Luo J, Wang W, Tang Y, Zhou D, Gao Y, Zhang Q, Zhou X, Zhu H, Xing L, Yu J. 2017. mRNA and methylation profiling of radioresistant esophageal cancer cells: the involvement of Sall2 in acquired aggressive phenotypes. *J Cancer* **8**: 646-656.
- Lusis AJ, Yu J, Wang SS. 2007. The problem of passenger genes in transgenic mice. *Arterioscler Thromb Vasc Biol* **27**: 2100-2103.
- Ma Y, Li D, Chai L, Luciani AM, Ford D, Morgan J, Maizel AL. 2001. Cloning and characterization of two promoters for the human HSAL2 gene and their transcriptional repression by the Wilms tumor suppressor gene product. *J Biol Chem* **276**: 48223-48230.
- Majewski J, Pastinen T. 2011. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet* **27**: 72-79.
- Masinde GL, Li X, Gu W, Davidson H, Mohan S, Baylink DJ. 2001. Identification of wound healing/regeneration quantitative trait loci (QTL) at multiple time points that explain seventy percent of variance in (MRL/MpJ and SJL/J) mice F2 population. *Genome Res* **11**: 2027-2033.
- McEwen E, Kedersha N, Song B, Scheuner D, Gilks N, Han A, Chen JJ, Anderson P, Kaufman RJ. 2005. Heme-regulated inhibitor kinase-mediated phosphorylation of eukaryotic translation initiation factor 2 inhibits translation, induces stress granule formation, and mediates survival upon arsenite exposure. *J Biol Chem* **280**: 16925-16933.
- McVicar DW, Winkler-Pickett R, Taylor LS, Makrigiannis A, Bennett M, Anderson SK, Ortaldo JR. 2002. Aberrant DAP12 signaling in the 129 strain of mice: implications for the analysis of gene-targeted mice. *J Immunol* **169**: 1721-1728.
- Menendez D, Krysiak O, Inga A, Krysiak B, Resnick MA, Schonfelder G. 2006. A SNP in the flt-1 promoter integrates the VEGF system into the p53 transcriptional network. *Proc Natl Acad Sci U S A* **103**: 1406-1411.
- Miao F, Zhang X, Cao Y, Wang Y, Zhang X. 2017. Effect of siRNA-silencing of SALL2 gene on growth, migration and invasion of human ovarian carcinoma A2780 cells. *BMC Cancer* **17**: 838.
- Miyake M, Goodison S, Lawton A, Gomes-Giacoia E, Rosser CJ. 2015. Angiogenin promotes tumoral growth and angiogenesis by regulating matrix metalloproteinase-2 expression via the ERK1/2 pathway. *Oncogene* **34**: 890-901.
- Motulsky HJ, Brown RE. 2006. Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**: 123.
- Nadeau JH. 2001. Modifier genes in mice and humans. *Nat Rev Genet* **2**: 165-174.
- Nadeau JH. 2003. Modifier genes and protective alleles in humans and mice. *Curr Opin Genet Dev* **13**: 290-295.

- Nielsen TO, Hsu FD, O'Connell JX, Gilks CB, Sorensen PH, Linn S, West RB, Liu CL, Botstein D, Brown PO et al. 2003. Tissue microarray validation of epidermal growth factor receptor and SALL2 in synovial sarcoma with comparison to tumors of similar histology. *Am J Pathol* **163**: 1449-1456.
- Nishioka Y, Dolan BM, Zahed L, Prado V, Tyson H. 1994. Molecular evolution of a Y-chromosomal repetitive sequence family in the genus *Mus*. *Mol Biol Evol* **11**: 146-153.
- Nitzki F, Kruger A, Reifenberg K, Wojnowski L, Hahn H. 2007. Identification of a genetic contamination in a commercial mouse strain using two panels of polymorphic markers. *Lab Anim* **41**: 218-228.
- Noyes HA, Agaba M, Anderson S, Archibald AL, Brass A, Gibson J, Hall L, Hulme H, Oh SJ, Kemp S. 2010. Genotype and expression analysis of two inbred mouse strains and two derived congenic strains suggest that most gene expression is trans regulated and sensitive to genetic background. *BMC Genomics* **11**: 361.
- Paigen K. 2003. One hundred years of mouse genetics: an intellectual history. II. The molecular revolution (1981-2002). *Genetics* **163**: 1227-1235.
- Pekarik V, Izpisua Belmonte JC. 2008. NFIX--one gene, two knockouts, multiple effects. *J Biol* **7**: 29.
- Perry WL, Hustad CM, Swing DA, O'Sullivan TN, Jenkins NA, Copeland NG. 1998. The itchy locus encodes a novel ubiquitin protein ligase that is disrupted in a 18H mice. *Nat Genet* **18**: 143-146.
- Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner G, Sargent EE, Asquith S, Crew V, Johnson KA, Robinson P et al. 2004. An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* **14**: 1806-1811.
- Piccoli R, Olson KA, Vallee BL, Fett JW. 1998. Chimeric anti-angiogenin antibody cAb 26-2F inhibits the formation of human breast cancer xenografts in athymic mice. *Proc Natl Acad Sci U S A* **95**: 4579-4583.
- Pincheira R, Baerwald M, Dunbar JD, Donner DB. 2009. Sall2 is a novel p75NTR-interacting protein that links NGF signalling to cell cycle progression and neurite outgrowth. *EMBO J* **28**: 261-273.
- Protter DS, Parker R. 2016. Principles and Properties of Stress Granules. *Trends Cell Biol* **26**: 668-679.
- Provost E, Rhee J, Leach SD. 2007. Viral 2A peptides allow expression of multiple proteins from a single ORF in transgenic zebrafish embryos. *Genesis* **45**: 625-629.
- Risso D, Ngai J, Speed TP, Dudoit S. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* **32**: 896-902.
- Rivera J, Tessarollo L. 2008. Genetic background and the dilemma of translating mouse studies to humans. *Immunity* **28**: 1-4.
- Sadagopan S, Veettil MV, Chakraborty S, Sharma-Walia N, Paudel N, Bottero V, Chandran B. 2012. Angiogenin functionally interacts with p53 and regulates p53-mediated apoptosis and cell survival. *Oncogene* **31**: 4835-4847.
- Sandilands A, Wang X, Hutcheson AM, James J, Prescott AR, Wegener A, Pekny M, Gong X, Quinlan RA. 2004. Bfsp2 mutation found in mouse 129 strains causes the loss of CP49' and induces vimentin-dependent changes in the lens fibre cell cytoskeleton. *Exp Eye Res* **78**: 875-889.
- Sarbassov DD, Guertin DA, Ali SM, Sabatini DM. 2005. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**: 1098-1101.
- Sato A, Matsumoto Y, Koide U, Kataoka Y, Yoshida N, Yokota T, Asashima M, Nishinakamura R. 2003. Zinc finger protein sall2 is not essential for embryonic and kidney development. *Mol Cell Biol* **23**: 62-69.
- Schalkwyk LC, Fernandes C, Nash MW, Kurrikoff K, Vasar E, Koks S. 2007. Interpretation of knockout experiments: the congenic footprint. *Genes Brain Behav* **6**: 299-303.



- Schmidt EK, Clavarino G, Ceppi M, Pierre P. 2009. SUnSET, a nonradioactive method to monitor protein synthesis. *Nat Methods* **6**: 275-277.
- Seong E, Saunders TL, Stewart CL, Burmeister M. 2004. To knockout in 129 or in C57BL/6: that is the question. *Trends Genet* **20**: 59-62.
- Sheng J, Luo C, Jiang Y, Hinds PW, Xu Z, Hu GF. 2014a. Transcription of angiogenin and ribonuclease 4 is regulated by RNA polymerase III elements and a CCCTC binding factor (CTCF)-dependent intragenic chromatin loop. *J Biol Chem* **289**: 12520-12534.
- Sheng J, Yu W, Gao X, Xu Z, Hu GF. 2014b. Angiogenin stimulates ribosomal RNA transcription by epigenetic activation of the ribosomal DNA promoter. *J Cell Physiol* **229**: 521-529.
- Simpson EM, Linder CC, Sargent EE, Davisson MT, Mobraaten LE, Sharp JJ. 1997. Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nat Genet* **16**: 19-27.
- Sterken R, Kiryluk K. 2010. The genetics of albuminuria: from haplotype association mapping in mice to genetic causation in humans. *Kidney Int* **77**: 173-175.
- Stewart SA, Dykxhoorn DM, Palliser D, Mizuno H, Yu EY, An DS, Sabatini DM, Chen IS, Hahn WC, Sharp PA et al. 2003. Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**: 493-501.
- Sung CK, Li D, Andrews E, Drapkin R, Benjamin T. 2013. Promoter methylation of the SALL2 tumor suppressor gene in ovarian cancers. *Mol Oncol* **7**: 419-427.
- Sung CK, Yim H, Gu H, Li D, Andrews E, Duraisamy S, Li C, Drapkin R, Benjamin T. 2012. The polyoma virus large T binding protein p150 is a transcriptional repressor of c-MYC. *PLoS One* **7**: e46486.
- Suva ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, Riggi N, Chi AS, Cahill DP, Nahed BV et al. 2014. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**: 580-594.
- Szabo R, Samson AL, Lawrence DA, Medcalf RL, Bugge TH. 2016. Passenger mutations and aberrant gene expression in congenic tissue plasminogen activator-deficient mouse strains. *J Thromb Haemost* **14**: 1618-1628.
- Taguchi YH, Oono Y. 2005. Relational patterns of gene expression via non-metric multidimensional scaling analysis. *Bioinformatics* **21**: 730-740.
- Taketo M, Schroeder AC, Mobraaten LE, Gunning KB, Hanten G, Fox RR, Roderick TH, Stewart CL, Lilly F, Hansen CT et al. 1991. FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc Natl Acad Sci U S A* **88**: 2065-2069.
- Threadgill DW, Dlugosz AA, Hansen LA, Tennenbaum T, Lichti U, Yee D, LaMantia C, Mourton T, Herrup K, Harris RC et al. 1995. Targeted disruption of mouse EGF receptor: effect of genetic background on mutant phenotype. *Science* **269**: 230-234.
- Threadgill DW, Yee D, Matin A, Nadeau JH, Magnuson T. 1997. Genealogy of the 129 inbred strains: 129/SvJ is a contaminated inbred strain. *Mamm Genome* **8**: 390-393.
- Tsuji T, Sun Y, Kishimoto K, Olson KA, Liu S, Hirukawa S, Hu GF. 2005. Angiogenin is translocated to the nucleus of HeLa cells and is involved in ribosomal RNA transcription and cell proliferation. *Cancer Res* **65**: 1352-1360.
- Vanden Berghe T, Hulpiau P, Martens L, Vandenbroucke RE, Van Wonterghem E, Perry SW, Bruggeman I, Divert T, Choi SM, Vuylsteke M et al. 2015. Passenger Mutations Confound Interpretation of All Genetically Modified Congenic Mice. *Immunity* **43**: 200-209.
- Vied C, Ray S, Badger CD, Bundy JL, Arbeitman MN, Nowakowski RS. 2016. Transcriptomic analysis of the hippocampus from six inbred strains of mice suggests a basis for sex-specific susceptibility and severity of neurological disorders. *J Comp Neurol* **524**: 2696-2710.
- Wahlsten D. 1982. Deficiency of corpus callosum varies with strain and supplier of the mice. *Brain Res* **239**: 329-347.

- Wang H, Parry S, Macones G, Sammel MD, Kuivaniemi H, Tromp G, Argyropoulos G, Halder I, Shriver MD, Romero R et al. 2006. A functional SNP in the promoter of the SERPINH1 gene increases risk of preterm premature rupture of membranes in African Americans. *Proc Natl Acad Sci U S A* **103**: 13463-13467.
- Weissmann C, Flechsig E. 2003. PrP knock-out and PrP transgenic mice in prion research. *Br Med Bull* **66**: 43-60.
- Wheeler TT, Smolenski GA, Harris DP, Gupta SK, Haigh BJ, Broadhurst MK, Molenaar AJ, Stelwagen K. 2012. Host-defence-related proteins in cows' milk. *Animal* **6**: 415-422.
- Xu ZP, Tsuji T, Riordan JF, Hu GF. 2003. Identification and characterization of an angiogenin-binding DNA sequence that stimulates luciferase reporter gene expression. *Biochemistry* **42**: 121-128.
- Yamamizu K, Sharov AA, Piao Y, Amano M, Yu H, Nishiyama A, Dudekula DB, Schlessinger D, Ko MS. 2016. Generation and gene expression profiling of 48 transcription-factor-inducible mouse embryonic stem cell lines. *Sci Rep* **6**: 25667.
- Yang F, Kemp CJ, Henikoff S. 2013. Doxorubicin enhances nucleosome turnover around promoters. *Curr Biol* **23**: 782-787.
- Yang W, Ng P, Zhao M, Wong TK, Yiu SM, Lau YL. 2008. Promoter-sharing by different genes in human genome--CPNE1 and RBM12 gene pair as an example. *BMC Genomics* **9**: 456.
- Yeo KJ, Jee JG, Hwang E, Kim EH, Jeon YH, Cheong HK. 2017. Interaction between human angiogenin and the p53 TAD2 domain and its implication for inhibitor discovery. *FEBS Lett* **591**: 3916-3925.
- Zalvide J, DeCaprio JA. 1995. Role of pRb-related proteins in simian virus 40 large-T-antigen-mediated transformation. *Mol Cell Biol* **15**: 5800-5810.
- Zheng CL, Wilmot B, Walter NA, Oberbeck D, Kawane S, Searles RP, McWeeney SK, Hitzemann R. 2015. Splicing landscape of the eight collaborative cross founder strains. *BMC Genomics* **16**: 52.
- Zhu JY, Abate M, Rice PW, Cole CN. 1991. The ability of simian virus 40 large T antigen to immortalize primary mouse embryo fibroblasts cosegregates with its ability to bind to p53. *J Virol* **65**: 6872-6880.

## VIII. ANEXOS

### i. PUBLICACIONES

**Farkas C**, Martins CP, Escobar D, Hepp MI, Donner DB, Castro AF and R Pincheira. (2013) Wild Type p53 Transcriptionally Represses the SALL2 Transcription Factor under Genotoxic Stress. *PLoS ONE* 8(9): e73817. doi:10.1371/journal.pone.0073817

D Escobar, M I Hepp, **C Farkas**, T Campos, N M Sodikin, M Morales, C I Álvarez, L Swigart, G I Evan, J L Gutiérrez, R Nishinakamura, A F Castro and R Pincheira. Sall2 is required for proapoptotic Noxa expression and genotoxic stress-induced apoptosis by doxorubicin. *Cell Death and Disease* (2015) **6**, e1816; doi:10.1038/cddis.2015.165

Hermosilla VE, Hepp MI, Escobar D, **Farkas C**, Riffo EN, Castro AF, Pincheira R. "Developmental SALL2 transcription factor: a new player in cancer." *Carcinogenesis*. 2017 Jul 1;38(7):680-690. doi: 10.1093/carcin/bgx036.

SALL2 represses cyclins D1 and E1 expression, and restrains G1/S cell cycle transition and cancer-related phenotypes. Viviana Hermosilla, Ginessa Salgado, Elizabeth Riffo, David Escobar, Matías Hepp, **Carlos Farkas**, Mario Galindo, Violeta Morín, María A. García-Robles, Ariel F. Castro, and Roxana Pincheira. *Mol Oncol*. 2018 Apr 24. doi: 10.1002/1878-0261.12308. [Epub ahead of print]

A Trichostatin A (TSA)/Sp1-mediated mechanism for the regulation of SALL2 tumor suppressor in Jurkat T cells. Matias Hepp, David Escobar\*, **Carlos Farkas\***, Viviana Hermosilla, Claudia Alvarez, Jose Gutierrez, Ariel Castro, Roxana Pincheira. *Biochim Biophys Acta*. 2018 May 18. pii: S1874-9399(18)30028-2. doi: 10.1016/j.bbagr.2018.05.002. [Epub ahead of print]

A genomic strategy for the background characterization of GEM models with the aid of expression data. **Farkas C.**, Fuentes-Villalobos F., Rebolledo-Jaramillo B., Makova KD., Nekrutenko A., Benavides F., Castro AF and Pincheira R. BMC genomics (Manuscrito en revisión).

ii. **PRESENTACIONES A CONGRESOS DURANTE LA PERMANENCIA EN EL PROGRAMA DE DOCTORADO**

Septiembre 2016 *"Angiogenina controla la traducción de proteínas de una manera dependiente de Sall2"*, presentada en la 16ª conferencia de Informática Genómica, Wellcome Genome Campus, Hinxton, Cambridge, Reino Unido..

Octubre 2015 *"Papel del factor de transcripción Sall2 bajo estrés genotóxico: un enfoque transcriptómico"*, presentado en la Cuarta Conferencia Internacional AACR sobre Fronteras en la Investigación Básica del Cáncer, Filadelfia, Pensilvania, Estados Unidos.

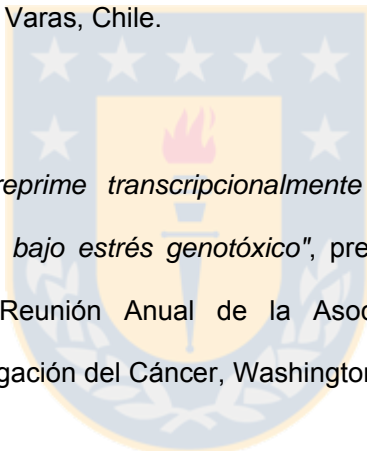
Septiembre 2015 *"Identificación de los objetivos transcripcionales Sall2 en respuesta al estrés genotóxico"*, presentado en la XXXVIII Reunión Anual de la Sociedad de Bioquímica y Biología

Molecular de Chile (Puerto Varas, Chile).

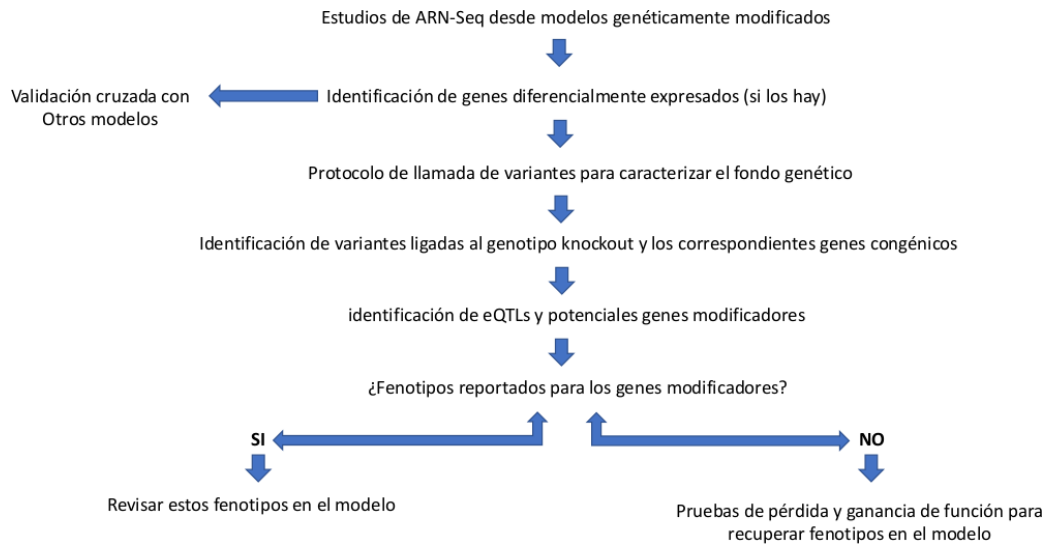
Agosto-Sept. 2014 *"El factor de transcripción Sall2 está involucrado en las respuestas de estrés celular"* (co-autor), presentado en el Congreso Internacional FEBS-EMBO, París, Francia.

Noviembre 2013 *"P53 y SP1 regulan transcripcionalmente el factor de transcripción SALL2"*, presentado en la XII Asociación Panamericana de Bioquímica y Biología Molecular (PABMB), Puerto Varas, Chile.

Abril 2013 *"p53 reprime transcripcionalmente el factor de transcripción SALL2 bajo estrés genotóxico"*, presentado en las Actas de la 104a Reunión Anual de la Asociación Americana para la Investigación del Cáncer, Washington, DC, Estados Unidos.



**Figura Anexo I** Protocolo para la supervisión genética de modelos genéticamente modificados (GEMs)



Después de la caracterización de los genes diferencialmente expresados en los modelos genéticamente modificados (si están presentes), el silenciamiento y/o la validación cruzada con otras estrategias pueden ayudar a discriminar genes dependientes del gen en estudio versus genes dependientes del fondo genético. Luego, la estrategia de llamado de variantes desde los datos de ARN-Seq pueden supervisar genéticamente el modelo empleado, incluyendo la identificación de variantes ligadas al genotipo KO y los genes congénitos correspondientes. Adicionalmente, la detección de loci con carácter cuantitativos de expresión pueden ayudar a identificar potenciales genes modificadores de fenotipos. Si los genes modificadores identificados registran fenotipos revisados en la literatura, estos fenotipos se deben revisar en el modelo en uso. Por el contrario, se deben iniciar estudios de pérdida o ganancia de función tanto con el gen en estudio como con los genes modificadores y así comprender los resultados fenotípicos en los modelos genéticamente modificados. Fuente: Elaboración propia.