



**TESIS DE DOCTORADO EN  
CIENCIAS BIOLÓGICAS**  
**Biología Celular y Molecular**



# **Rol de los ARNs no codificantes largos en la espermatogénesis**

**María Fernanda Trovero**

**Noviembre 2020**

**Orientadora:**

**Dra. Adriana Geisinger**

**Co-orientadores:**

**Dra. Rosana Rodríguez**

**Dr. José Sotelo Silveira**

**Tribunal:**

**Dra. Rossana Sapiro**

**Dr. Bernardo Bertoni**

**Dr. Pablo Smircich**



## **Agradecimientos**

Esta tesis no hubiese sido posible sin la colaboración o el apoyo de todas estas personas, y por eso a ellos va dedicada:

A mi tutora Adriana y co-tutora Rosana, por el apoyo incondicional aún en momentos de “crisis del no me da nada”, por darme para adelante con las ideas, los viajes, las becas y con presentarme a todo, siempre.

A mi co-tutor Coya por el apoyo y los valiosos consejos, y a su equipo de estudiantes e investigadores por la ayuda en los análisis informáticos.

A mi casi-tutor germano-uruguayo Ricardo Benavente por recibirme y bancarme tantas veces en su laboratorio, y en su casa, junto a Mercedes. A todo su equipo, especialmente Irene, Marie y Roberta, por ayudarme con los ensayos durante las pasantías. Y a toda la gente que tuve el placer de conocer en Alemania por hacer mis estadías tan amenas.

A Fede Santiñaque, el técnico de citometría más dedicado y detallista que conozco, gran compañero y colega, un gustazo trabajar con él y charlar durante las tardes de sorting.

A Mateo François, porque me aceptó como orientadora, y parte de los resultados de esta tesis también son suyos.

A Lucía Canclini por los varios tips que me dio, y muy especialmente a Gustavo Folle por darme la oportunidad no solamente de continuar en el Instituto, sino también en este proyecto de investigación, y por introducirme en el mundo del cultivo celular.

A todo el grupo de Biología Molecular de la Reproducción, por tirar todos para el mismo lado siempre.

A todos y cada uno de mis compañeros de laboratorio por el aguante del día a día.

A mis padres, porque siempre apostaron a mi educación y formación ante todo, y sin ellos seguro que nada de esto sería posible.

A la familia que confió en que podía llegar.

A mis amigos del liceo, de la Facu, del baile y de la vida que me aguantan y acompañan en todas, en las fáciles, pero sobre todo en las difíciles.

Al tribunal de la Tesis, Rossana, Bernardo y Pablo, por haber aceptado corregir este trabajo. También a Nibia Berois, que siguió todo el proceso de la tesis formando parte de la CAS.

Y a todo el que de alguna manera aportó su granito de arena para que este trabajo saliera adelante.

Nada más y nada menos que un ¡GRACIASSS ENORME!

Un agradecimiento más que especial a todos los financiadores y colaboradores:

A la ANII, por la beca de posgrado nacional, y la beca de movilidad para realizar una de las pasantías en Alemania. También por el proyecto de investigación FCE en el cual se enmarcó mi doctorado.

A Boehringer Ingelheim, por colaborar con una de mis estadías en Alemania.

A la DAAD, por la beca para una de las pasantías en Alemania.

Al programa PEDECIBA de la UdelaR, por las alícuotas y por colaborar con una de las estadías en Alemania.

A la CSIC, por el proyecto Grupos I+D del cual formo parte.

A la Universidad de Würzburg, porque a través del Prof. Benavente, siempre me abrió las puertas y me dio mi espacio para trabajar.

Al IIBCE donde llevé a cabo la mayor parte del trabajo experimental, y a la Facultad de Ciencias donde me formé.



## **Índice de contenidos**

|                                                                                        |    |
|----------------------------------------------------------------------------------------|----|
| <b><u>1. Resumen</u></b> .....                                                         | 11 |
| <b><u>2. Introducción, antecedentes y objetivos</u></b> .....                          | 14 |
| 2.1: Espermatogénesis .....                                                            | 15 |
| 2.1.1: Generalidades de la espermatogénesis .....                                      | 15 |
| 2.1.2: Meiosis .....                                                                   | 17 |
| 2.1.3: Espermiogénesis .....                                                           | 19 |
| 2.2: ARNs no codificantes .....                                                        | 21 |
| 2.2.1: Un poco de historia sobre los ARNs no codificantes .....                        | 21 |
| 2.2.2: ncRNAs cortos y largos .....                                                    | 22 |
| 2.2.3: Origen y conservación de los lncRNAs .....                                      | 25 |
| 2.2.4: Localización sub-celular de los lncRNAs .....                                   | 27 |
| 2.2.5: Clasificación de los lncRNAs en distintos biotipos .....                        | 28 |
| 2.2.6: Características generales de los lncRNAs .....                                  | 31 |
| 2.2.7: Funciones generales de los lncRNAs .....                                        | 32 |
| 2.2.8: lncRNAs y epigenética .....                                                     | 34 |
| 2.2.9: lncRNAs asociados a enfermedades y patologías .....                             | 37 |
| 2.3: lncRNAs y espermatogénesis .....                                                  | 38 |
| 2.3.1: Los lncRNAs testiculares .....                                                  | 38 |
| 2.3.2: lncRNAs y patologías testiculares .....                                         | 44 |
| 2.4: Antecedentes de nuestro laboratorio .....                                         | 45 |
| 2.4.1: Metodologías de purificación de células espermatogénicas .....                  | 45 |
| 2.4.2: Análisis del transcriptoma codificante de la<br>espermatogénesis de ratón ..... | 47 |
| 2.5: Hipótesis y objetivos .....                                                       | 48 |
| 2.5.1: ¿Por qué estudiar los lncRNAs en la espermatogénesis? ...                       | 48 |
| 2.5.2: Hipótesis .....                                                                 | 49 |
| 2.5.3: Objetivo general .....                                                          | 50 |
| 2.5.4: Objetivos específicos .....                                                     | 50 |
| <b><u>3. Materiales y métodos</u></b> .....                                            | 51 |
| 3.1: Animales .....                                                                    | 53 |
| 3.2: Clasificación celular por citometría de flujo .....                               | 53 |

|                                                                                                                                                               |    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.2.1: Preparación de suspensiones celulares .....                                                                                                            | 53 |
| 3.2.2: Citometría de flujo .....                                                                                                                              | 54 |
| 3.2.3: Análisis de la pureza de las fracciones celulares obtenidas                                                                                            | 55 |
| 3.3: Extracciones de ARN, construcción de genotecas y<br>secuenciación masiva .....                                                                           | 56 |
| 3.3.1: Extracción de ARN y cuantificación .....                                                                                                               | 56 |
| 3.3.2: Construcción de librerías de secuenciación hebra-<br>específicas, controles cuanti- y cualitativos, y secuenciación del ARN<br>( <i>RNAseq</i> ) ..... | 56 |
| 3.4: Análisis <i>in silico</i> de los datos de <i>RNAseq</i> .....                                                                                            | 56 |
| 3.4.1: Procesamiento de las secuencias obtenidas y generación<br>de datos transcriptómicos .....                                                              | 57 |
| 3.4.2: Análisis de la reproducibilidad de las réplicas .....                                                                                                  | 58 |
| 3.4.3: Construcción de diagramas de Venn y mapas de expresión                                                                                                 | 58 |
| 3.4.4: Análisis de largo y número de exones, biotipos, y<br>distribución cromosómica de los lncRNAs .....                                                     | 58 |
| 3.4.5: Análisis de co-expresión con genes codificantes vecinos o<br>solapantes .....                                                                          | 59 |
| 3.4.6: Análisis de ontología de los genes codificantes vecinos .....                                                                                          | 59 |
| 3.4.7: Análisis de conservación en humanos .....                                                                                                              | 60 |
| 3.4.8: Re-análisis de datos de otros trabajos para comparación<br>con nuestras listas .....                                                                   | 60 |
| 3.4.9: Análisis de datos de lncRNAs utilizando la base de datos de<br>CLS ( <i>Capture Long-Read Sequence</i> ) .....                                         | 60 |
| 3.5: Validaciones de patrones de expresión y de co-expresión con<br>genes codificantes mediante RT-qPCR .....                                                 | 61 |
| 3.6: Hibridación <i>in situ</i> e inmunolocalización .....                                                                                                    | 62 |
| 3.6.1: Preparación de criosecciones de tejido testicular .....                                                                                                | 62 |
| 3.6.2: Diseño y construcción de sondas de ARN con tecnología<br><i>Stellaris</i> ® .....                                                                      | 63 |
| 3.6.3: Hibridación <i>in situ</i> de ARN (RNA-FISH) sobre criosecciones<br>de testículo .....                                                                 | 64 |

|                                                                                                                                     |     |
|-------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.6.4: Detección de proteínas por inmunofluorescencia (IF) sobre<br>criosecciones de testículo – protocolo <i>Stellaris</i> ® ..... | 64  |
| 3.7: Microscopía y análisis de imágenes .....                                                                                       | 65  |
| 3.7.1: Microscopía confocal .....                                                                                                   | 65  |
| 3.7.2: Microscopía de alta resolución .....                                                                                         | 65  |
| 3.7.3: Análisis de imágenes con Fiji .....                                                                                          | 66  |
| 3.8: Cultivos de sobrevida de células germinales .....                                                                              | 68  |
| 3.8.1: Preparación de suspensión celular mediante disgregación<br>enzimática del tejido .....                                       | 68  |
| 3.8.2: Clasificación de las fracciones 4C y RS por citometría de<br>flujo .....                                                     | 69  |
| 3.8.3: Análisis de la viabilidad de los cultivos de sobrevida a las 48<br>horas .....                                               | 69  |
| <b>4. Resultados</b> .....                                                                                                          | 71  |
| 4.1: Capítulo 1: LncRNAs y su vinculación a patologías testiculares .<br>Artículo Trovero & Geisinger, 2019 .....                   | 73  |
| 4.2: Capítulo 2: Fracciones celulares puras y librerías hebra-<br>específicas para <i>RNAseq</i> .....                              | 93  |
| 4.2.1: Obtención de fracciones celulares puras de las poblaciones<br>espermatogénicas de interés por citometría de flujo .....      | 94  |
| 4.2.2: Construcción de librerías hebra-específicas para<br>secuenciación masiva .....                                               | 97  |
| 4.3: Capítulo 3: Transcriptoma global de lncRNAs a lo largo de la<br>espermatogénesis de ratón .....                                | 100 |
| 4.3.1: Análisis general de los datos de secuenciación obtenidos ....                                                                | 101 |
| 4.3.2: LncRNAs expresados y diferenciales en las distintas<br>poblaciones de células testiculares .....                             | 102 |
| 4.3.3: Comparación de los resultados con los de otros trabajos .....                                                                | 106 |
| 4.3.4: Caracterización global del transcriptoma de la<br>espermatogénesis del ratón .....                                           | 108 |
| 4.3.5: Co-expresión de lncRNAs con genes codificantes .....                                                                         | 110 |
| 4.3.6: Conservación de pares de genes lncRNA/gen codificante<br>entre ratón y humano .....                                          | 112 |

|                                                                                                         |     |
|---------------------------------------------------------------------------------------------------------|-----|
| 4.3.7: Validaciones por RT-qPCR .....                                                                   | 113 |
| Artículo Trovero <i>et al.</i> , 2020 .....                                                             | 114 |
| 4.4: Capítulo 4: Localización celular y sub-celular de lncRNAs en<br>criosecciones testiculares .....   | 147 |
| 4.4.1: Sondas de ARN <i>Stellaris</i> ® y tipos de microscopía .....                                    | 148 |
| 4.4.2: <i>Malat1</i> como control, y su señal nuclear .....                                             | 150 |
| 4.4.3: Localización de lncRNAs en el cuerpo cromatoide .....                                            | 153 |
| 4.4.4: Co-localización de un mRNA y su lncRNA antisentido<br>solapante, en el cuerpo cromatoide .....   | 155 |
| 4.4.5: Localización del lncRNA antisentido <i>Kcnmb4os1</i> en las<br>células en profase meiótica ..... | 160 |
| 4.5: Capítulo 5: Cultivos celulares de sobrevida: resultados<br>preliminares .....                      | 163 |
| <u>5. Discusión</u> .....                                                                               | 167 |
| 5.1: Acerca de nuestros abordajes metodológicos .....                                                   | 168 |
| 5.2: Algunas consideraciones en relación a la comparación con otros<br>trabajos .....                   | 172 |
| 5.3: La meiosis y (principalmente) la espermiogénesis como fuentes<br>de lncRNAs .....                  | 174 |
| 5.4: De la caracterización global del transcriptoma de lncRNAs .....                                    | 177 |
| 5.5: ¿Existe alguna relación entre la expresión de lncRNAs y la de<br>genes codificantes? .....         | 178 |
| 5.6: ¿Todos los caminos conducen al cuerpo cromatoide? .....                                            | 183 |
| 5.7: ¿Podrán participar lncRNAs en la regulación de procesos<br>meióticos? .....                        | 186 |
| 5.8: Avances en los cultivos celulares de sobrevida de línea germinal .....                             | 186 |
| <u>6. Conclusiones y perspectivas</u> .....                                                             | 190 |
| 6.1: Conclusiones .....                                                                                 | 191 |
| 6.2: Perspectivas .....                                                                                 | 192 |
| <u>7. Bibliografía</u> .....                                                                            | 195 |
| <u>8. Anexos</u> .....                                                                                  | 225 |
| 8.1: Anexo 1: Soluciones, tampones y medios .....                                                       | 226 |

|                                                                                                             |     |
|-------------------------------------------------------------------------------------------------------------|-----|
| 8.2: Anexo 2: Tabla de lncRNAs diferenciales con FDR p-valor $\leq 0,05$ y $ 2  > \log_2 FC \geq  1 $ ..... | 228 |
|-------------------------------------------------------------------------------------------------------------|-----|

## **Índice de tablas**

|                                                                                                                                                                                    |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabla 1. Cebadores diseñados y utilizados en la tesis .....                                                                                                                        | 62  |
| Tabla 2. Secuencias de las sondas de ARN <i>Stellaris</i> ® diseñadas .....                                                                                                        | 66  |
| Tabla 3. Rendimiento de las purificaciones celulares obtenidas por citometría de flujo .....                                                                                       | 95  |
| Tabla 4. Top 10 lncRNAs diferencialmente sobre-expresados o reprimidos (mayor FC) en las distintas transiciones de etapa (FDR p-valor $\leq 0,05$ y $\log_2 (FC) \geq  2 $ ) ..... | 104 |
| Tabla 5. Características de los lncRNAs o genes seleccionados y de las sondas de ARN diseñadas .....                                                                               | 149 |
| Tabla 6. Características de las fracciones celulares clasificadas por citometría de flujo, partiendo de suspensiones celulares obtenidas por dos métodos diferentes .....          | 165 |
| Tabla 7. Viabilidad de los cultivos celulares de sobrevida a las 24 y 48 horas .....                                                                                               | 165 |

## **Índice de figuras**

|                                                                                                 |    |
|-------------------------------------------------------------------------------------------------|----|
| Figura 1. Cortes de testículo de ratón adulto .....                                             | 16 |
| Figura 2. Esquema de la espermatogénesis del ratón .....                                        | 16 |
| Figura 3. Esquema resumido de la profase I de la meiosis masculina .....                        | 18 |
| Figura 4. Clasificación de los lncRNAs en diferentes biotipos según posición en el genoma ..... | 29 |
| Figura 5. lncRNAs y epigenética .....                                                           | 36 |
| Figura 6. Genes de lncRNAs predichos en distintos tejidos para 5 especies .....                 | 39 |
| Figura 7. Apareamiento homólogo en el <i>locus sme2</i> de levadura .....                       | 43 |

|                                                                                                                                                                    |     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 8. Esquema del flujo de trabajo de la tesis .....                                                                                                           | 52  |
| Figura 9. Esquema de la espermatogénesis del ratón, donde se señalan las poblaciones celulares purificadas en rojo .....                                           | 95  |
| Figura 10. Histogramas y gráficos de puntos ( <i>dot plots</i> ) de la clasificación en flujo de las distintas poblaciones celulares de interés .....              | 96  |
| Figura 11. Compilado de imágenes de células RS purificadas por citometría de flujo vistas al microscopio confocal .....                                            | 97  |
| Figura 12. Electroferogramas de las corridas en el <i>Bioanalyzer</i> .....                                                                                        | 99  |
| Figura 13. LncRNAs expresados utilizando los <i>pipelines</i> de CLC y HSB .....                                                                                   | 102 |
| Figura 14. Gráfico de análisis de componentes principales (PCA) .....                                                                                              | 103 |
| Figura 15. Diagramas de Venn de los genes de <i>lncRNAs</i> DE (izquierda: sobre-expresados; derecha: reprimidos) ( $FDR \leq 0,05$ y $\log_2 FC \geq  1 $ ) ..... | 104 |
| Figura 16. Localización de <i>Malat1</i> .....                                                                                                                     | 152 |
| Figura 17. Localización sub-celular de lncRNAs .....                                                                                                               | 154 |
| Figura 18. Cuerpo cromatoide .....                                                                                                                                 | 155 |
| Figura 19. LncRNAs y cuerpo cromatoide .....                                                                                                                       | 156 |
| Figura 20. <i>Kcnmb4</i> en el cuerpo cromatoide .....                                                                                                             | 157 |
| Figura 21. Co-localización de <i>Kcnmb4os1</i> y <i>kcnmb4</i> en el cuerpo cromatoide .....                                                                       | 158 |
| Figura 22. Figura 6a y b, tomada y modificada del artículo anexado en sección 4.3 [Trovero <i>et al.</i> , 2020] .....                                             | 159 |
| Figura 23. <i>Kcnmb4os1</i> en el núcleo de espermatoцитos paquiténicos .....                                                                                      | 160 |
| Figura 24. <i>Kcnmb4os1</i> y el complejo sinaptonémico .....                                                                                                      | 161 |
| Figura 25. Esquema de la espermatogénesis del ratón, donde se representan las señales obtenidas para las distintas sondas .....                                    | 162 |
| Figura 26. Viabilidad de los cultivos de sobrevida .....                                                                                                           | 166 |



# **1. RESUMEN**

Los ARNs no codificantes largos (lncRNAs) han adquirido un gran interés en los últimos tiempos, conforme se fueron describiendo sus roles en distintos procesos biológicos y mecanismos celulares, así como también han sido asociados a diferentes patologías. Se ha visto que el testículo es el órgano o tejido que presenta mayor expresión de lncRNAs; por lo tanto, constituye un sistema ideal para el estudio de estas moléculas. Particularmente, se han descrito funciones para algunos lncRNAs en distintos eventos que ocurren durante la espermatogénesis. Es así que el objetivo general de esta tesis fue la caracterización global del transcriptoma de lncRNAs a lo largo de la espermatogénesis, utilizando ratón como modelo, y tratar de dilucidar sus posibles roles en el proceso.

Con este objetivo, se obtuvieron poblaciones celulares purificadas de distintas etapas de la espermatogénesis del ratón mediante una metodología desarrollada por nuestro grupo de investigación, que se basa en la citometría de flujo y alcanza niveles altísimos de pureza. A partir de ARN total de estas poblaciones se estudiaron sus transcriptomas mediante secuenciación masiva, conservando la información direccional de los transcritos. Los datos de la secuenciación fueron analizados mediante *pipelines* bioinformáticos alternativos. A través de diferentes análisis *in silico*, se caracterizó el transcriptoma de los lncRNAs (en cuanto a biotipos, número de exones y largo de transcritos, distribución cromosómica, conservación), y se estudió la co-expresión con genes codificantes vecinos y solapantes. Además, se validaron la expresión y co-expresión mediante RT-qPCR. A continuación, se seleccionaron algunos lncRNAs candidatos, y se localizaron sub-celularmente empleando hibridaciones *in situ* de ARN (*RNA-FISH*), en simultáneo con la co-localización de distintas proteínas.

Brevemente, la mayoría de los lncRNAs se expresa en etapas post-meióticas, y sus biotipos mayoritarios son intergénicos (lincRNA) o antisentido. En general, poseen un par de exones y largo menor a 3.000 nucleótidos. Se encontraron distribuidos en todos los cromosomas excepto en el Y, y en promedio un 50% están conservados en humano. El 95% de los lncRNAs antisentido muestra una correlación en la expresión con sus genes codificantes solapantes, mientras que ese porcentaje es bastante menor (no más de un 65%) para los lincRNAs y sus genes vecinos cercanos. Sorprendentemente, todos los lncRNAs de expresión

diferencial durante la espermatogénesis escogidos para las *RNA-FISH* muestran una localización en el cuerpo cromatoide, un organelo perinuclear no membranoso típico de las células post-meióticas, que se cree vinculado a regulación postranscripcional y al almacenamiento (y/o secuestro) de distintas moléculas. Nuestros hallazgos apoyan la idea de que muchos lncRNAs de la espermatogénesis estarían relacionados con los procesos de regulación postranscripcional durante el desarrollo de los espermatozoides, mediados por el cuerpo cromatoide. Sin embargo, para poder conocer las funciones de los lncRNAs y sus mecanismos de acción, es preciso realizar ensayos funcionales. En ese sentido, iniciamos la puesta a punto de un cultivo de sobrevida de células germinales, sobre los cuales realizar los mencionados ensayos.

En suma, los aportes de este trabajo resultan novedosos, a la vez que nos dan el puntapié inicial para que continuemos investigando acerca de los mecanismos de acción y sus roles en relación con la reproducción y la espermatogénesis. Si bien se han descrito importantes funciones en la reproducción y la espermatogénesis para algunos lncRNAs, poco se sabe aún sobre la participación de estos transcritos en eventos clave durante esos procesos, ya que es un área de investigación aún en sus inicios.

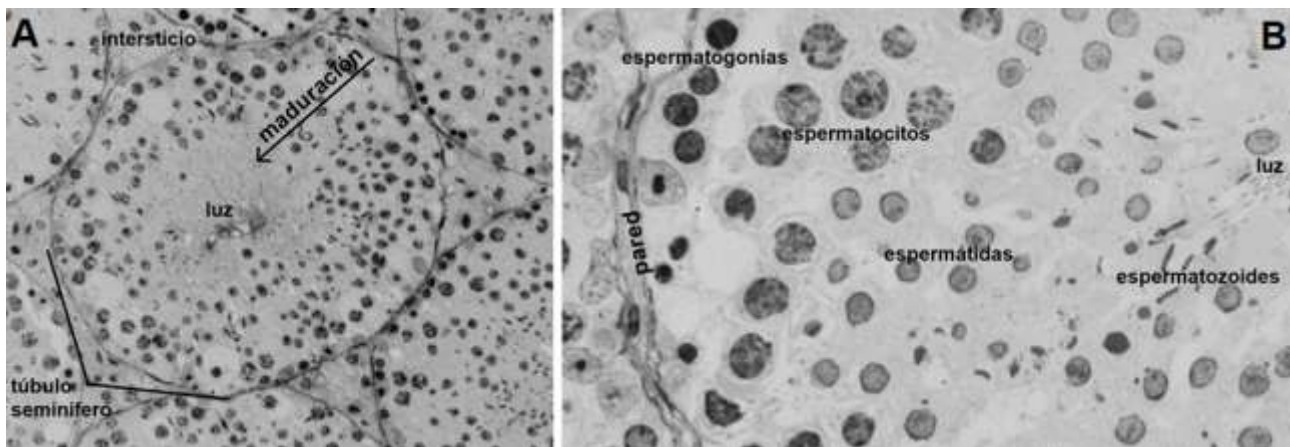
## **2. INTRODUCCIÓN,** **ANTECEDENTES Y** **OBJETIVOS**

## 2.1: Espermatogénesis

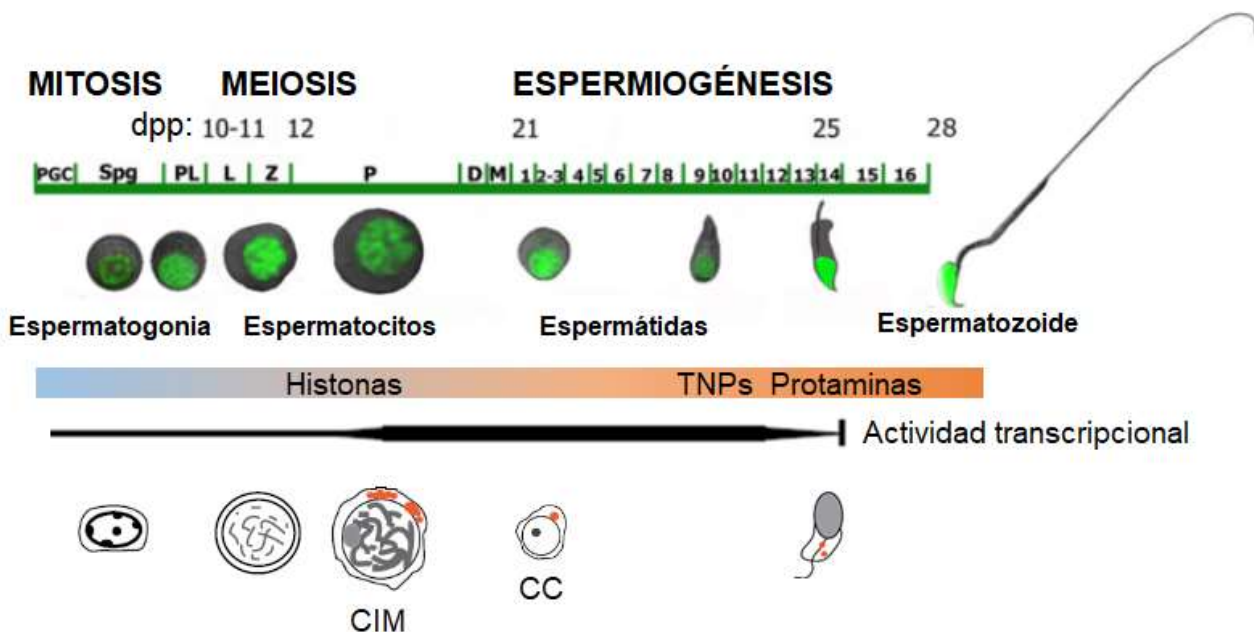
### **2.1.1: Generalidades de la espermatogénesis**

Los testículos, órganos reproductores masculinos, están formados por múltiples túbulos seminíferos separados por espacio intersticial. Dentro de éstos co-existen más de 30 tipos celulares distintos, que incluyen la línea germinal y las células somáticas (Leydig, Sertoli, mioides, fibroblastos, linfocitos, entre otras). Hacia la membrana basal de los túbulos se encuentran las células germinales más inmaduras (espermatogonias) y las células de Sertoli. Estas últimas proveen a la línea germinal del nicho apropiado, el microambiente y los nutrientes para el proceso espermatogénico [Petersen & Soder, 2006]. A medida que nos acercamos a la luz del túbulo encontramos células más maduras en las distintas etapas del proceso espermatogénico (espermatoцитos, espermátidas) y en la luz, los espermatozoides maduros (Figura 1). El testículo adulto se caracteriza por la presencia de todos los tipos celulares de la línea germinal co-existiendo, con ondas de diferenciación propulsadas por la señalización con ácido retinoico (AR) [Hogarth *et al.*, 2015].

La espermatogénesis (del griego, “*spermato*”: semilla y “*genesis*”: generación) es un proceso de proliferación y diferenciación celular que ocurre en la vida adulta, consistente en la formación de gametos masculinos. El mismo se compone de tres programas de expresión génica sucesivos pero co-existentes en los túbulos seminíferos del testículo: a) la proliferación mitótica de las espermatogonias, células precursoras de línea germinal (que en el ratón se clasifican en de tipo A, intermedias y de tipo B), a la vez que se mantiene una pequeña población de células madre espermatogénicas; b) división meiótica, durante la cual se provee a los espermatoцитos, células germinales meióticas, de la maquinaria de apareamiento, recombinación homóloga y segregación; y c) espermiogénesis, etapa en la que las espermátidas redondas, células ya haploides, sufren su diferenciación en espermatozoides maduros (Figura 2) [Erickson, 1990; Kleene, 2001; Bolcun-Filas & Handel, 2018]. Cada uno de estos eventos depende de la expresión correcta y precisa de genes específicos, participando diferentes factores de crecimiento y epigenéticos [Bettegowda & Wilkinson, 2010; Zhou *et al.*, 2019].



**Figura 1.** Cortes de testículo de ratón adulto. (A) Vista panorámica de un túbulo seminífero. La flecha indica el sentido de maduración de las células. (B) Ampliación del corte de testículo; se diferencian los distintos tipos celulares de línea germinal.



**Figura 2.** Esquema de la espermatogénesis del ratón, donde se representa el momento de reemplazo de histonas, la actividad transcripcional, y la aparición y desarrollo del cuerpo cromatoide. dpp: días post-parto. PGC: células germinales primordiales. Spg: espermatogonias. PL: pre-leptoteno. L: leptoteno. Z: citogeno. P: paquiteno. D: diploteno. M: divisiones meióticas. TNPs: proteínas de transición. CIM: cemento intermitocondrial. CC: cuerpo cromatoide. Tomado y modificado de da Cruz *et al.* [2016] y Meikar *et al.* [2011].

### 2.1.2: Meiosis

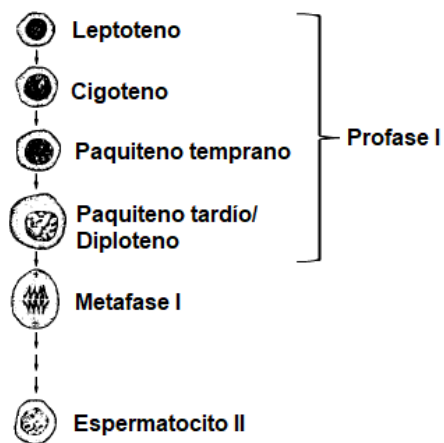
La meiosis en el contexto de la gametogénesis, por su parte, es un proceso de división celular especial de la línea germinal, consistente en una única duplicación de ADN seguida de dos divisiones celulares consecutivas (meiosis I y II), que dan lugar finalmente a células haploides y con contenido de ADN C. La señalización del AR es esencial para la entrada en meiosis, y la Meiosina junto con Stra8, expresados en respuesta al AR, actúan como factores de transcripción activando genes de meiosis. Ambos en conjunto son importantes en la supresión del programa mitótico y coordinación del ciclo celular para la entrada en meiosis, controlando el destino celular de las células germinales [Ishiguro *et al.*, 2020].

Luego de la replicación del ADN durante la mitosis de las espermatogonias, ocurren dos divisiones meióticas consecutivas: la primera es reductiva, reduciendo el contenido de ADN de los espermatocitos de 4C (número cromosómico 2N) a 2C (1N), al separar los pares de cromosomas homólogos. La segunda división meiótica es equitativa, separando las cromátidas hermanas, y reduciendo la cantidad de ADN a 1C en las espermátidas (manteniendo el número cromosómico haploide) [Bolcun-Filas & Handel, 2018].

Dentro del proceso de meiosis pueden diferenciarse varias etapas, caracterizadas por eventos únicos que en ellas suceden: profase I, metafase I, anafase I, telofase I, profase II, metafase II, anafase II y telofase II.

La profase I, que es la más larga y extensa de todas, se sub-divide a su vez en diferentes estadios: leptoteno, cigoteno, paquiteno, diploteno y diacinesis (Figura 3). La estructura que caracteriza a la profase I es el complejo sinaptonémico (CS), estructura macroproteica mediadora del apareamiento entre cromosomas homólogos, en la cual pueden distinguirse dos elementos laterales (ELs), un elemento central y filamentos transversos que los conectan dándole un aspecto escaleriforme [Page & Hawley, 2004; Fraune *et al.*, 2012; Schücker *et al.*, 2015]. Durante el leptoteno se ensamblan los elementos axiales (EAs, que posteriormente constituirán los ELs de los CSs), los telómeros se unen a la envoltura nuclear [Esponda & Gimenez-Martín, 1972], y se desplazan sobre la misma constituyendo una agrupación característica conocida como *bouquet* [Scherthan, 2007], que posiblemente favorece el acercamiento de los cromosomas homólogos. El apareamiento de éstos comienza hacia la entrada

en cigoteno, observándose coexistencia de EAs con tramos ya ensamblados de CS. En el estadio de paquiteno los CSs se encuentran totalmente ensamblados de un extremo cromosómico al otro, y ocurre la recombinación entre cromosomas homólogos (*crossing-over*), mientras que en diploteno se desensambla el CS [Bolcun-Filas & Handel, 2018]. En lepto/citogeno pueden verse los nódulos de recombinación tempranos, que marcarían sitios vinculados al intercambio de hebra; hacia paquiteno aparecen los nódulos de recombinación tardíos, que indicarían los lugares donde efectivamente ocurre la recombinación; y en diploteno, al desensamblarse el CS, los cromosomas homólogos permanecen unidos únicamente por los quiasmas, que serían resabios de esos sitios de recombinación [Moens *et al.*, 2007; Handel & Schimenti, 2010].



**Figura 3.** Esquema resumido de la profase I de la meiosis masculina.

En la etapa de espermatocitos paquiténicos ocurre lo que llamamos inactivación de cromosomas sexuales en meiosis (*meiotic sex chromosome inactivation*: MSCI), durante la cual se silencian transcripcionalmente los genes de los cromosomas X e Y mediante una condensación cromatínica (excepto una pequeña zona de homología entre ambos conocida como región pseudoautosómica), y se empaquetan en una estructura compacta denominada cuerpo sexual o X-Y [Cloutier & Turner, 2010]. El apareamiento y la recombinación de cromosomas homólogos son esenciales para asegurar la correcta segregación de los cromosomas a las células hijas en la meiosis [Gerton & Hawley, 2005]. Existen puntos de control en la meiosis (*checkpoints*) para

monitorear todo el proceso, siendo uno de los principales el de la salida de paquiteno. Alteraciones en el apareamiento o recombinación homólogos activan este punto de control, generando un arresto meiótico que puede derivar en patologías e infertilidad [Roeder & Bailis, 2000; Cohen *et al.*, 2006].

El complejo sinaptonémico es una estructura proteica altamente conservada en la evolución, presente en todos los eucariotas desde las levaduras hasta el hombre. En los mamíferos hasta el momento se han descrito 8 proteínas que conforman este complejo: SYCP2 y SYCP3 en los ELs; SYCE1, SYCE2, SYCE3, TEX12 y SIX6OS en el elemento central; y SYCP1 en los filamentos transversales [Costa *et al.*, 2005; De Vries *et al.*, 2005; Hamer *et al.*, 2006; Winkel *et al.*, 2009; Alsheimer *et al.*, 2010; Schramm *et al.*, 2011; Gómez-H *et al.*, 2016]. Varios de estos elementos están conservados, llegando a encontrarse incluso en cnidarios [Fraune *et al.*, 2012; Fraune *et al.*, 2014]. Sabemos que las proteínas SYCP1 y SYCP3 son capaces de unirse al ADN [Syrjänen *et al.*, 2014; Duncie *et al.*, 2018; Bollschweiler *et al.*, 2019] y que SYCP2, el otro componente de los ELs, presenta también dominios potenciales de unión a ADN [Winkel *et al.*, 2009]. Un trabajo ha reportado que a los ELs de esta estructura proteica se unen bucles de cromatina que contendrían repetidos de ADN cortos y largos (*SINES*, *LINES*, etc.), que podrían servir de anclaje del cromosoma al CS [Hernández-Hernández *et al.*, 2008]. Si bien se han descubierto los componentes proteicos del complejo, no está claro cuál es el mecanismo mediante el cual el mismo participa en los procesos de apareamiento y recombinación de los cromosomas homólogos [Kleckner, 2006]. Se ha sugerido también la presencia de ARNs no codificantes en el CS, aunque se ignora su identidad o función por ser un campo aún inexplorado.

### **2.1.3: Espermiogénesis**

La espermiogénesis es un proceso de diferenciación terminal muy complejo que sufren las espermátidas redondas hasta llegar a espermatozoides maduros. Estos dramáticos cambios incluyen un altísimo nivel de compactación de la cromatina, desarrollo de flagelo y acrosoma, elongación nuclear junto con un reordenamiento mitocondrial, y exclusión de la mayor parte del citoplasma [Tanaka & Baba, 2005; White-Cooper & Davidson, 2011]. Además, ocurre una remodelación de la cromatina única en la espermiogénesis, donde

aproximadamente el 90% de las histonas canónicas son sustituidas por variantes específicas de testículo; posteriormente, sobre el final de la espermiogénesis, la mayor parte de las histonas son reemplazadas por proteínas de transición en primer lugar y luego por protaminas [Hao *et al.*, 2019] (ver Figura 2). Una pequeña fracción de las histonas (5-15%, dependiendo de la especie) es retenida en la cromatina bajo la forma de nucleosomas, y se especula que la herencia de estas histonas provee al espermatozoide de marcas epigénéticas importantes para regular la expresión génica en la fertilización [White-Cooper & Davidson, 2011]. La sustitución de las histonas hace que la cromatina del espermatozoide alcance niveles únicos de compactación. Esos niveles tan elevados de compactación a su vez permiten una gran reducción en el tamaño nuclear (una cabeza de espermatozoide es unas 50 veces menor que un núcleo promedio) favoreciendo la hidrodinámica, característica fundamental teniendo en cuenta el largo trayecto que dichas células deben recorrer en el tracto femenino. Además, la compactación protegería a la cromatina del daño en el ADN durante el trayecto [Braun, 2001]. No obstante, como consecuencia de la extrema compactación y del reemplazo de histonas (con la subsiguiente pérdida de marcas epigenéticas) estas células son incapaces de transcribir, produciéndose un silenciamiento transcripcional. La línea germinal masculina ha desarrollado mecanismos de regulación postranscripcional y traduccional complejos y diversos, y se cree que ello obedece a la necesidad de utilizarlos en etapas tardías de la espermiogénesis, de modo de permitir la síntesis de proteínas en ausencia de transcripción [Geisinger, 2008; Gan *et al.*, 2013]. Algunas de estas estrategias de regulación pueden incluir el secuestro de ARNs mensajeros (mRNAs) en ribonucleoproteínas, unión de represores en regiones UTRs, regulación del largo de poliadenilación de mRNAs, entre otros [Kleene, 2001].

A lo largo de la diferenciación de las células germinales masculinas aparecen varios gránulos u organelos sin membrana en el citoplasma de las mismas, que empaquetan ARN y proteínas de unión a ARN. Estos gránulos, conocidos como gránulos germinales o “*nuage*”, participan en la localización, estabilidad y traducción de ARNs, así como en la producción de pequeños ARNs no codificantes (ncRNAs) y en el control de la organización cromatínica [Meikar *et al.*, 2011; Meikar *et al.*, 2014; Lehtiniemi & Kotaja, 2018]. Es así que otra posible estrategia de regulación génica particular de las espermátidas podría ser la

retención de ARNs en el cuerpo cromatoide (CC), el gránulo citoplásmico no membranoso perinuclear mejor caracterizado en mamíferos. El CC es una estructura específica de las células post-meióticas, y no existe en ningún otro tejido ni tipo celular. Aparece en los espermátocitos en paquíteno como gránulos dispersos, y su precursor sería el cemento intermitocondrial, aunque este hecho es bastante discutido. En las espermátidas redondas el CC se condensa en un único gránulo perinuclear, manteniéndose hasta la elongación del núcleo celular [Kotaja *et al.*, 2006; Kotaja & Sassone-Corsi, 2007; Chuma *et al.*, 2009; Meikar *et al.*, 2011; Meikar *et al.*, 2014] (ver Figura 2). Sin embargo, existen autores que plantean que la formación de estos organelos es discontinua entre ambos tipos celulares (espermátocitos y espermátidas redondas) [Onohara *et al.*, 2010]. Los rápidos movimientos no aleatorios en el citoplasma y la vinculación con los poros nucleares, caracterizan al CC [Parvinen, 2005].

El CC se conoce desde hace mucho tiempo atrás dado que es visible al microscopio óptico, y ha resultado enigmático para los científicos durante décadas ya que si bien se han propuesto diferentes roles para este organelo, su función aún es desconocida. Recientemente se ha propuesto que es un depósito que acumula proteínas de unión a ARN, proteínas relacionadas a la regulación del ARN y su procesamiento, ARNs de diversos tipos (piwi-RNAs, ARNs no codificantes cortos y largos, algunos mRNAs, etc.) [Meikar *et al.*, 2014]. El CC alberga también los componentes de la maquinaria del decaimiento de mRNAs con codones de terminación prematuros (*non-sense mediated decay*, NMD), que no solamente eliminaría los transcriptos aberrantes, sino también algunos completos y funcionales colaborando en la sincronización fina de los distintos procesos celulares [Lehtiniemi & Kotaja, 2018]. Es posible entonces que el CC esté involucrado en la regulación postranscripcional en las espermátidas, y su rol incluya el control, el procesamiento y/o la degradación de ARNs.

## 2.2: ARNs no codificantes

### **2.2.1: Un poco de historia sobre los ARNs no codificantes**

En el año 1955 George Palade identificó el primer ARN no codificante (ncRNA) como parte de una ribonucleoproteína muy abundante en el citoplasma: el ARN

ribosomal (rRNA). También hacia finales de la década del 1950 Hoagland y Zamecnik descubrieron la molécula que participa en la traducción de ARN a aminoácidos: el ARN de transferencia (tRNA) [Jarroux *et al.*, 2017]. Treinta años después se describió el primer ncRNA regulatorio, y en la década de los 90s se descubrieron y caracterizaron algunos famosos lncRNAs como *Xist*, *BC200* o *H19* en la era pre-genómica. Pero no fue sino hasta luego del Proyecto Genoma Humano, que se vio que sólo un 2% del mismo codifica proteínas, y el resto se consideró ncRNA “basura” o ruido transcripcional [Jarroux *et al.*, 2017; Tsagakis *et al.*, 2020]. A partir de ese momento comenzaron a conocerse y caracterizarse múltiples ncRNAs, relacionados a tantas funciones celulares y procesos biológicos como las proteínas. Es así como se caía el dogma central de la biología molecular sostenido por décadas: el ADN se transcribe a ARN, y éste se traduce a proteínas.

### **2.2.2: ncRNAs cortos y largos**

Alrededor de dos tercios del genoma humano sufren una transcripción generalizada, lo que contrasta con que solamente un 2% del mismo sea traducido a proteínas [Djebali *et al.*, 2012]. Aquellos transcriptos a partir de los cuales no se generan proteínas, se denominan ARNs no codificantes, que podemos dividir en ARNs no codificantes cortos (sncRNAs) y largos (lncRNAs).

Los ARNs pequeños o cortos (sncRNAs) pueden ser de varios tipos. Los ARNs pequeños que interactúan con la proteína Piwi (piRNAs) tienen entre 29 y 30 nucleótidos, son expresados casi exclusivamente en la línea germinal masculina de mamíferos [Girard *et al.*, 2006], y se los ha visto, por ejemplo, involucrados en el silenciamiento de retrotransposones [Homolka *et al.*, 2011]. Se observó que mutaciones en cualquiera de los componentes de la vía de los piRNAs lleva a infertilidad masculina [Lehtiniemi & Kotaja, 2018]. Por ejemplo, TOPAZ1 es una proteína que participa en la vía de los piRNAs y está involucrada en la espermatogénesis, contribuyendo al silenciamiento de transposones y a la estabilidad genómica. Su delección no tiene efecto en la fertilidad de ratones hembras, pero sí en los machos donde hay arresto meiótico de espermatoцитos paquiténicos [Luangpraseuth-Prosper *et al.*, 2015].

Los microARNs (miRNAs) son otra clase de pequeños ARNs endógenos y simple hebra de aproximadamente 21-25 nucleótidos de extensión, cuyas

funciones principales se han relacionado a la degradación o inhibición de la traducción de mRNAs [Liu *et al.*, 2018]. Se ha visto que los miRNAs pueden escapar a la inactivación de los cromosomas sexuales en paquitenio, y por ende, se han sugerido como reguladores de dicho proceso [Homolka *et al.*, 2011]. Estos sncRNAs han sido extensivamente revisados en distintos trabajos y han demostrado ser indispensables en el desarrollo de las células germinales masculinas [Luk *et al.*, 2014].

Los lncRNAs han sido menos estudiados, y solamente en los últimos años han comenzado a estudiarse más exhaustivamente, al descubrirse que pueden desempeñar funciones relevantes. Por oposición a los sncRNAs, se definen como aquellos ncRNAs de tamaño mayor a 200 pb. El corte arbitrario en un largo de 200 pb surgió por razones experimentales, ya que los protocolos de purificación de ARN y de secuenciación masiva de ARNs (*RNAseq*) en sus inicios excluían los ARNs no codificantes pequeños de esa manera, y por lo tanto a los mayores de 200 pb se les conoció como ARNs no codificantes largos [Kapranov *et al.*, 2007; Mercer *et al.*, 2009; Atkinson *et al.*, 2012]. También se los ha definido como moléculas de más de 2 Kb con un potencial codificante menor a 100 aminoácidos (aa) [Cao, 2014].

Es preciso mencionar que la premisa de “no generar proteínas funcionales” para considerarse ncRNAs se ha debilitado en los últimos tiempos. En base fundamentalmente a estudios de *ribosome-profiling*, se ha puesto de manifiesto que algunos lncRNAs poseen potencial traduccional, y se han observado ribosomas unidos a lncRNAs [Quinn & Chang, 2016; Li *et al.*, 2017]. El 40% de los lncRNAs de las células humanas codificarían oligopéptidos mayores a 10 aa, mientras que un tercio de los lncRNAs de las células madre espermatogénicas de ratón albergarían al menos un marco abierto de lectura (ORF) [Li *et al.*, 2017]. Para algunos de estos pequeños péptidos se han descrito funciones: DWORF de ratón contrarresta el efecto de inhibidores de la ATPasa  $\text{Ca}^{2+}$  del retículo sarcoplásmico (SERCA); Spar regula negativamente la actividad de la proteína diana de la rapamicina mTORC1 [Li *et al.*, 2017]. Como toda excepción que confirma la regla, se ha visto que por azar algunos lncRNAs pueden contener ORFs de más de 100 aa. Por lo tanto, un criterio posible sería admitir que no existe una separación clara entre transcritos codificantes y no-codificantes, y nombrar a los ncRNAs como aquellos transcritos que es “poco probable” que

codifiquen para proteínas funcionales [Ulitsky & Bartel, 2013]. Asimismo, un mismo ARN puede contener o generar un transcripto codificante de proteínas y uno no codificante, por lo que la definición se vuelve aún más confusa [Wu & Du, 2017].

Otros ncRNAs como elementos repetidos en el genoma de mamíferos, por ejemplo los retrotransposones, componen entre el 30 y el 50% del mismo. Estos presentan actividad transcripcional, son tejido-específicos y están cerca de transcriptos codificantes de proteínas. Esto podría indicar su participación en la expresión de proteínas vía promotores alternativos o en la regulación postranscripcional, como reguladores en *cis*, en la formación de sncRNAs doble hebra [Faulkner & Carninci, 2009; Atkinson *et al.*, 2012]. Se ha visto por ejemplo que los elementos transponibles LINE1 son necesarios para inducir la condensación de la cromatina, pero además, se ha demostrado el papel del ARN LINE1 actuando como lncRNA (*lncRNA-like*) en la renovación de las células madre embrionarias y el desarrollo pre-implantatorio del embrión [Honson & Macfarlan, 2018]. Los pseudogenes también serían una clase de elementos repetidos que se transcriben como lncRNAs [Atkinson *et al.*, 2012].

Otra clase de ARNs no codificantes endógenos que se ha descubierto en los últimos años son los ARNs circulares (circRNAs). A diferencia de los lncRNAs que son lineales, los circRNAs son producidos por un procesamiento alternativo no canónico, y forman un bucle circular con una unión covalente que los hace más estables y resistentes a la degradación por nucleasas; además, son altamente tejido-específicos. Por estos motivos se cree que podrían estar relacionados a diversas funciones, y de hecho se ha visto que cumplen roles importantes en muchos procesos biológicos, e incluso se los ha asociado a patologías como la azoospermia, es decir, ausencia de espermatozoides en el semen [Hentze & Preiss, 2013; Memczak *et al.*, 2013; Li *et al.*, 2017; Quan & Li, 2018; Ge *et al.*, 2019; Zhou *et al.*, 2019]. Curiosamente, la expresión del circRNA derivado del gen *Sry* (*circSry*) en el testículo adulto está acoplada a la primera onda espermatogénica y a las espermátidas redondas, mientras que en la cresta genital durante el desarrollo embrionario *Sry* se expresa como un mRNA lineal [Quan & Li, 2018].

En procariotas se estima que menos del 25% de su ADN es no codificante, en eucariotas simples entre el 25 y el 50%, mientras que en los más complejos como

hongos, plantas y animales supera el 50%, alcanzando incluso el 98% en humanos [Amaral *et al.*, 2008]. En el año 2016, Zheng y colaboradores desarrollaron un *pipeline* que les permitió identificar más de 176.600 lncRNAs en 14 especies; de ellos cerca de 24.300 eran específicos de primates, unos 5.200 específicos de roedores, y 55 de ellos mostraron ortología entre humano y pez cebra. Además anotaron 14.800 ARNs circulares (circRNAs) en humanos, de los cuales 1.260 eran ortólogos con circRNAs de ratón [Zheng *et al.*, 2016].

### **2.2.3: Origen y conservación de los lncRNAs**

Los lncRNAs surgen a través de diferentes mecanismos que incluyen duplicaciones en tándem de ADN o ARN de secuencias genómicas, pérdida de potencial codificante de genes codificantes para proteínas, evolución adaptativa de ADN no codificante o de elementos transponibles, yuxtaposición y reestructuración durante la recombinación cromosómica [Ponting *et al.*, 2009; Marques & Ponting, 2014; Ulitsky, 2016; Jarroux *et al.*, 2017; Liu *et al.*, 2018].

Los lncRNAs funcionales no tienen alta conservación de secuencia primaria, sugiriendo que principalmente tienen funciones reguladoras y que evolucionan más rápida y fluidamente que las proteínas, cuyos genes son altamente conservados a lo largo de los vertebrados [Firth *et al.*, 2006; Ulitsky & Bartel, 2013]. Solamente un 72% de los lncRNAs intergénicos de humano se expresan también en mono macaco, comparado con el 98% de proteínas codificantes expresadas en todos los primates [Necsulea *et al.*, 2014]. En contraste con los mRNAs e incluso con algunos otros tipos de ncRNAs, los lncRNAs de mamíferos no tienen ortólogos en otras especies fuera de los vertebrados. Solamente para un 6% de los lncRNAs de pez cebra su secuencia puede detectarse en lncRNAs humanos o murinos, 12% de lncRNAs entre humano y ratón, y un 60% dentro de los roedores [Ulitsky & Bartel, 2013]. Sin embargo, Hon y colaboradores [2017] encontraron un enriquecimiento en lncRNAs conservados implicados en rasgos de GWAS (estudios de asociación de genoma completo) y eQTLs (expresión cuantitativa de rasgos).

Resulta interesante que un grupo de lncRNAs conservados expresados en células madre embrionarias en humano se procesan diferente que sus homólogos en el ratón, resultando en diferente localización sub-celular y función. Dentro de ellos se identificaron 122 lncRNAs con secuencia conservada y 229

con posición conservada entre ambas especies de mamíferos, que en el ratón se retenían principalmente en el núcleo, mientras en humano eran procesados y exportados preferentemente al citoplasma [Sharma & Carninci, 2020].

El criterio de conservación primaria de secuencia puede resultar muy estricto cuando consideramos los lncRNAs, pero otras características como estructura, función o sintenia pueden ser factores más aplicables al estudio de la evolución de estas moléculas [Diederichs, 2014]. Se ha sugerido que los lncRNAs más conservados son aquellos que participan en el desarrollo, contando con una fuerte conservación espacio-temporal y expresión sinténica [Ulitsky *et al.*, 2011; Necsulea *et al.*, 2014; Washietl *et al.*, 2014]. Por otra parte, si bien todo el gen de un lncRNA tiende a no ser conservado, algunos fragmentos cortos de secuencia conservados se pueden encontrar sobre todo en el extremo 5' [Mercer *et al.*, 2009; Hezroni *et al.*, 2015]. Algunos lncRNAs también pueden contener algunos elementos conservados, como ser elementos regulatorios, sitios de unión a factores de transcripción, señales de localización nuclear o motivos de *splicing* [Ulitsky, 2016; Melé *et al.*, 2017].

Los lncRNAs tienen la capacidad de actuar directamente como ligandos de proteínas o de mediar la unión de las proteínas a sus blancos en el ADN o ARN. Si bien esta propiedad es compartida con sncRNAs, los lncRNAs se pueden plegar y formar estructuras secundarias que faciliten estas uniones ARN- o ADN-proteínas [Ulitsky & Bartel, 2013; Fatica & Bozzoni, 2014]. Si los lncRNAs tuvieran regiones estructuradas críticas para su función, éstas deberían mantenerse conservadas en la evolución. Sin embargo, esas estructuras predichas están depletadas o pobremente enriquecidas en exones de lncRNAs, por lo que las estructuras secundarias conservadas parecen ocupar un lugar muy pequeño en el transcriptoma de lncRNAs de los vertebrados [Ulitsky & Bartel, 2013]. Algunas excepciones son la triple estructura que estabiliza el extremo 3' en los lncRNAs *Malat1* y *Neat1*, la estructura en bucle de *roX*, o el repetido *RepA* en *Xist* [Ulitsky, 2016]. Por otra parte, Rivas y colaboradores [2017] desarrollaron un método computacional, *R-scape*, que analiza una alineación de secuencia de ARN múltiple y cuantifica el soporte estadístico para la conservación evolutiva de una estructura secundaria de ARN, y no hallaron una significancia estadística que apoye la idea de una estructura conservada, por ejemplo, para los tan estudiados lncRNAs *Xist*, *SRA* o *HOTAIR*.

Dado que los lncRNAs son poco conservados en su secuencia y se expresan a bajos niveles, se cree que actuarían agregados más que individualmente, o que el solo evento de su transcripción cumpliría una función [Ulitsky & Bartel, 2013; Quinn & Chang, 2016]. Un ejemplo claro es el del lncRNA *Airn*, que es antisentido al promotor del receptor de insulina Igf2r, y cuyo evento de transcripción es importante para el silenciamiento de dicho receptor, pero no así el transcripto de 118 kb en sí mismo [Latos *et al.*, 2012; Ulitsky, 2016]. Incluso se ha visto que promotores de los lncRNAs actúan como potenciadores de genes próximos [Engreitz *et al.*, 2016].

Existe evidencia creciente de que los lncRNAs no están sujetos a presiones de selección a nivel de secuencia primaria, pero sí a restricciones selectivas en la estructura génica, sobre promotores y patrones de *splicing* [Nitsche & Stadler, 2017]. Otro aspecto a destacar, es que se han identificado lncRNAs de muy baja abundancia pero que cumplen roles fundamentales. Tal es el caso del lncRNA *TERC*, que se expresa en muy bajos niveles y sin embargo tiene un rol crítico manteniendo el largo de los cromosomas, extendiendo los telómeros en las células de mamíferos [Rinn & Chang, 2020].

#### **2.2.4: Localización sub-celular de los lncRNAs**

Según su localización a nivel celular, los lncRNAs pueden ser nucleares o citoplásmicos [Ponting *et al.*, 2009]. Debido a los roles encontrados en un inicio para los lncRNAs, se pensaba que estaban principalmente localizados en el núcleo [Khalil *et al.*, 2009]. Sin embargo, en otros trabajos se vio que algunos lncRNAs tienen localización fundamentalmente citoplásmica [Bao *et al.*, 2013].

Los lncRNAs nucleares pueden cumplir funciones como por ejemplo guiar a los modificadores de cromatina a sus blancos en el genoma, o reclutar metilasas de histonas para la represión transcripcional [Fatica & Bozzoni, 2014].

Los lncRNAs citoplásmicos usualmente tienen complementariedad con transcriptos que se originan de sus mismos *loci* o de *loci* independientes. Uno de los modos en que actúan estos lncRNAs es compitiendo con ARN endógeno, uniendo miRNAs como si fueran una esponja, y evitando la acción de los mismos [Salmena *et al.*, 2011; Fatica & Bozzoni, 2014]. Ejemplo de esto es el caso de uno de los lncRNAs más estudiados, *Malat1*. En el testículo, *Malat1* promueve la apoptosis e inhibe la proliferación celular en lesiones de reperfusión luego de

una isquemia (IRI) a través de la unión del miRNA miR-214, lo que impide la inhibición de TRPV4, que es una proteína que actúa en la IRI [Li *et al.*, 2018; revisado en Trovero & Geisinger, 2019]. Otro ejemplo es el del lncRNA AK015322, altamente expresado en células madre espermato goniales, que sirve como señuelo para el microRNA miR-19b-3p actuando como antagonista a su función, que es la de reprimir el factor de transcripción ETV5. ETV5 es esencial para la auto-renovación de las células madre espermato goniales [Hu *et al.*, 2017].

Aunque los lncRNAs tienen un sesgo en su localización por estar más enriquecidos en el núcleo en relación a los mRNAs, hay más lncRNAs en cuanto a número absoluto de transcritos en el citoplasma [Ulitsky & Bartel, 2013; Quinn & Chang, 2016]. Kapranov *et al.* [2007] sugirieron que alrededor del 50% de los lncRNAs de líneas celulares humanas se localizarían tanto en el núcleo como en el citoplasma, un 30% únicamente en el núcleo, y un 15% únicamente en el citoplasma. Esto puede indicar que los lncRNAs cumplen funciones tanto a nivel transcripcional (relacionados a potenciadores y promotores, o reclutando y coordinando otros factores) como a nivel postranscripcional (regulando procesamiento alternativo, transporte, traducción o degradación de mRNAs) [Atkinson *et al.*, 2012].

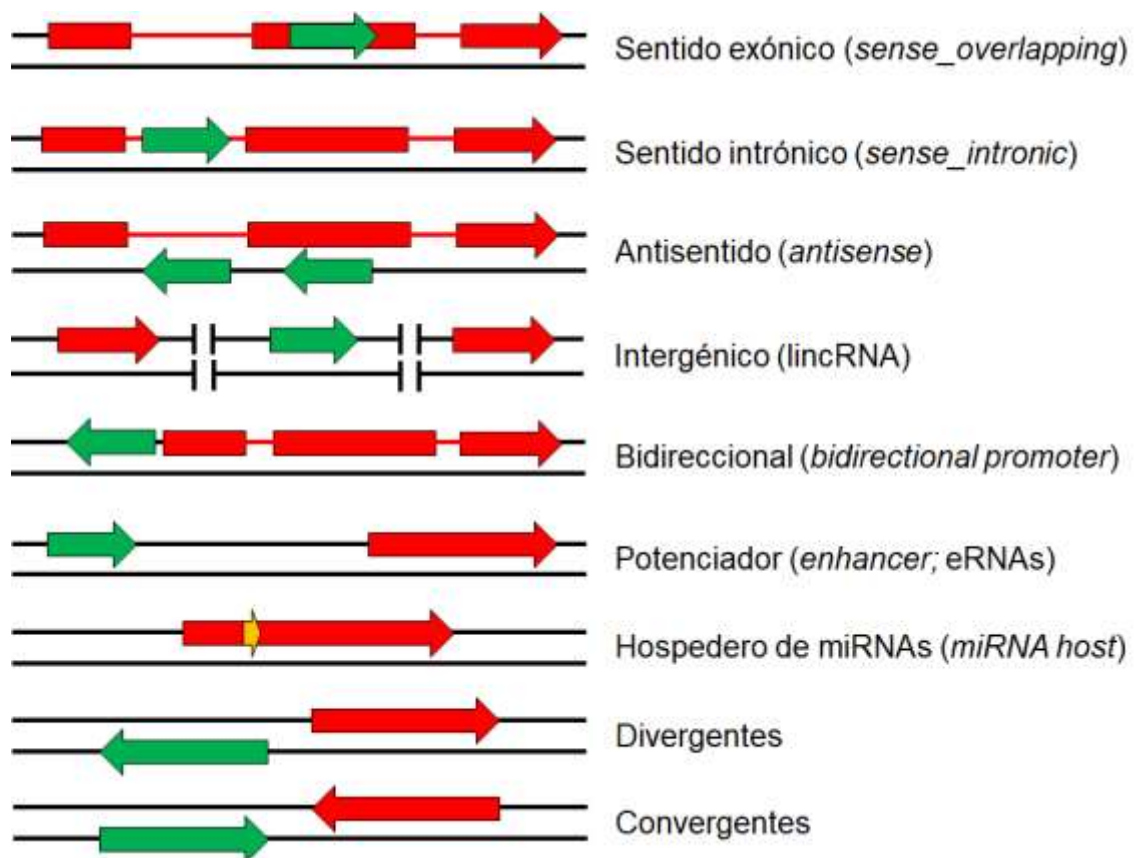
#### **2.2.5: Clasificación de lncRNAs en distintos biotipos**

Existen muchas formas o criterios para clasificar a los lncRNAs, como largo, función, localización, orientación, etc. Un criterio muy utilizado es de acuerdo a su posicionamiento en el genoma con respecto a otros transcritos (Figura 4).

De acuerdo a este criterio, los lncRNAs pueden ser clasificados como sentido exónicos o intrónicos, antisentido, intergénicos o bidireccionales [Ponting *et al.*, 2009]. Una clasificación un poco más abarcativa los categoriza en: divergentes o convergentes, intrónicos, intergénicos, sentido o antisentido solapantes, potenciadores, y hospederos de miRNAs [Wu & Du, 2017].

Las tecnologías de NGS (*Next Generation Sequencing*) han revelado que la ARN polimerasa II puede iniciar la transcripción en forma bidireccional, generándose lncRNAs a partir de promotores bidireccionales. Se ha sugerido que este evento puede favorecer la expresión de genes codificantes de proteínas ya que se genera un entorno de cromatina no compactada. No está claro

igualmente, si estos promotores bidireccionales son realmente funcionales o si tienen roles regulatorios [Atkinson *et al.*, 2012]. Esta tecnología también ha puesto de manifiesto la transcripción de lncRNAs desde potenciadores (eRNAs), lo que en general precede a una inducción del gen codificante adyacente [Atkinson *et al.*, 2012]. Este proceso es mucho más frecuente de lo que se pensaba anteriormente. En un trabajo en el que se estudiaron lncRNAs en células humanas, se identificó una subclase de lncRNAs que cumplen funciones de tipo potenciador en la expresión de genes codificantes, ya que su supresión reveló la influencia positiva que éstos tenían sobre genes codificantes vecinos [Orom *et al.*, 2010].



**Figura 4.** Clasificación de lncRNAs en diferentes biotipos según posición en el genoma. Verde: lncRNA. Rojo: gen codificante. Amarillo: miRNA. LincRNA: *long intergenic noncoding RNA*.

La transcripción intergénica, es decir, aquella que no solapa con ningún gen codificante de proteínas, produce lncRNAs intergénicos (lincRNAs), que son

aquellos que se localizan a más de 1 Kb de distancia de otros genes vecinos [Atkinson *et al.*, 2012; Tsagakis *et al.*, 2020]. Muchos estudios generales se han centrado en los lincRNAs [Ponjavic *et al.*, 2007; Guttman *et al.*, 2009; Khalil *et al.*, 2009; Orom *et al.*, 2010], ya que al no solapar con otros transcritos facilitan la manipulación experimental y el análisis computacional. Además, es más sencillo atribuirles función ya que las consecuencias de mutaciones genómicas (generadas en estudios funcionales, por ej. mediante CRISPR/Cas) pueden asignárseles inequívocamente [Tsagakis *et al.*, 2020].

Se ha sugerido que alrededor del 72% del genoma humano genera transcritos antisentido no codificantes, esto es, transcritos que surgen de la hebra opuesta a la hebra sentido, ya sea solapando con transcritos codificantes o no [Pelechano & Steinmetz, 2013; Barman *et al.*, 2019]. Del mismo modo, se ha propuesto que más del 70% de los transcritos en el genoma del ratón generaría transcritos antisentido [Barman *et al.*, 2019]. Los lncRNAs antisentido pueden ser solapantes de exones o de intrones de otros transcritos [Ma *et al.*, 2013]. Si bien se cree que los lncRNAs antisentido se acumulan en el núcleo, algunos de ellos se han encontrado en el citoplasma y en las mitocondrias [Djebali *et al.*, 2012]. Estos transcritos son generados desde promotores independientes, bidireccionales o crípticos. Los transcritos antisentido pueden regular la expresión local del gen que les da origen en *cis*, o bien en *trans* en genes de otros cromosomas; pero también el acto de la transcripción en sí mismo puede ser el regulador de la expresión génica, regulando la misma en todos los niveles: transcripción, traducción y degradación de mRNAs [Pelechano & Steinmetz, 2013; Engreitz *et al.*, 2016; Kopp & Mendell, 2018; Barman *et al.*, 2019]. Se propone que los transcritos antisentido también pueden tener un efecto activador, protegiendo a los promotores de la metilación del ADN, a través de la formación de un bucle híbrido de ADN-ARN durante la transcripción. Los lncRNAs antisentido pueden también regular las isoformas que se producen desde el transcrito codificante sentido, por ejemplo enmascarando sitios de *splicing* alternativo específicos y previniendo su procesamiento [Pelechano & Steinmetz, 2013].

### 2.2.6: Características generales de los lncRNAs

Muchas características de los lncRNAs son comunes entre todos los vertebrados en que se han estudiado: son relativamente cortos, de bajo número de exones (aunque frecuentemente los exones son más largos que los de los mRNAs), baja conservación de secuencia, poco abundantes, y con expresión altamente restringida en espacio y tiempo [Cabili *et al.*, 2011; Pauli *et al.*, 2012; Derrien *et al.*, 2012; Sun *et al.*, 2013; Ulitsky & Bartel, 2013; Nakagawa & Kageyama, 2014; Quinn & Chang, 2016; Ran *et al.*, 2016]. Se cree que su bajo contenido en GC podría explicar en parte sus niveles generalmente inferiores de expresión con respecto a los mRNAs [Niazi & Valadkhan, 2012]. Se ha reportado también que los lncRNAs tienden a concentrarse en las cercanías de genes codificantes asociados a distintos procesos biológicos [Cabili *et al.*, 2011].

En general, por lo que se ha visto, se unen menos cantidad de factores de transcripción a las regiones promotoras de lncRNAs que de mRNAs, a la vez que tienen menos cantidad de modificaciones de histonas. Sin embargo, los eventos de unión de factores de transcripción parecen ser más conservados en los promotores de lncRNAs que en los de los mRNAs [Rinn & Chang, 2020].

Como hemos mencionado, una característica sobresaliente de los lncRNAs es que en general son poco abundantes en comparación con los mRNAs y pueden ser poliadenilados o carecer de cola poli-A [Rinn *et al.*, 2007; Ponting *et al.*, 2009; Gupta *et al.*, 2010]. En algunos reportes se ha subestimado la abundancia de lncRNAs no poliadenilados [Mercer *et al.*, 2009], al generar ADN copia con oligodTs que únicamente hibridan con las colas poliadeniladas. Sin embargo, otros autores opinan que el utilizar ARN total en lugar de únicamente poliadenilado no agrega números sustanciales de lncRNAs al análisis, y por el contrario adiciona ruido al mismo [Ulitsky, 2016].

Los lncRNAs sufren procesamiento alternativo por corte y empalme (*splicing*), aunque éste sería más ineficiente que en los mRNAs [Ulitsky & Bartel, 2013; Melé *et al.*, 2017]. Finalmente, la inmensa mayoría de los lncRNAs tiene patrones de expresión tejido-específicos, y de ellos un gran número son específicos del testículo [Cabili *et al.*, 2011], como veremos más adelante.

### 2.2.7: Funciones generales de los lncRNAs

Como hemos mencionado, los lncRNAs pueden actuar en *cis* o en *trans* [Ponjavic *et al.*, 2009; Pauli *et al.*, 2012]. No obstante, se desconoce cómo son retenidos en el sitio de transcripción o cómo encuentran sus blancos a distancia. Se ha propuesto que podría ser a través de proteínas, por formación de tripletes ARN-ADN, o mediante estructuras en el ARN [Fatica & Bozzoni, 2014]. Muchas veces incluso se han asignado diferentes funciones mediante distintos mecanismos al mismo lncRNA, dependiendo de su localización en cada tipo celular y de las moléculas con las que interactúa [Morlando & Fatica, 2018]. Por ejemplo, se reportó para el lncRNA *roX* una función en la compensación de dosis del cromosoma X del macho en *Drosophila* como parte del complejo MSL y, a su vez, debido a su locación genómica, el rol de andamio para la unión y ensamblaje del complejo MSL [Ilik & Akhtar, 2009].

En términos globales, se considera que los lncRNAs pueden actuar de varias formas diferentes: competidores/inhibidores, activadores/reclutadores, precursores, o potenciadores [Luk *et al.*, 2014; Wu & Du, 2017]. Como competidores, los lncRNAs pueden unirse a proteínas de unión al ADN y evitar su acción, o pueden unirse por complementariedad al ADN y evitar que se unan dichas proteínas, o bien evitar la unión de miRNAs a mRNAs evadiendo su posterior degradación (ejemplos: *Xist*, *PANDA*, *RMST*). Como activadores, pueden reclutar modificadores de cromatina y de histonas en sitios específicos actuando como andamio, y activarlos (tal sería el caso de *HOTAIR* o *Evf2*). Como precursores, los lncRNAs pueden ser procesados por RNasas como Drosha o Dicer y generar sncRNAs (ej.: *H19*, *HongrES2*). Como potenciadores *on/off*, cuando se transcriben desde una región potenciadora, interactúan en el contacto promotor-potenciador, inhibiendo la transcripción del gen codificante [Luk *et al.*, 2014; Wu & Du, 2017].

Existen múltiples trabajos donde se ha estudiado a los lncRNAs en relación a tan diversos procesos celulares y biológicos como remodelación de la cromatina, establecimiento de impronta de genes, expresión génica tanto de mRNAs como de miRNAs, transcripción, procesamiento postranscripcional, procesamiento alternativo por corte y empalme, traducción, tráfico intracelular, neurogénesis, embriogénesis, proliferación, diferenciación y migración celulares, ciclo celular, pluripotencialidad celular, respuesta inmune, apoptosis, respuesta a shock

térmico, entre muchos otros [Dinger *et al.*, 2008; Mercer *et al.*, 2009; Hung & Chang, 2010; Koziol & Rinn, 2010; Valadkhan & Nilsen, 2010; Pauli *et al.*, 2011; Pauli *et al.*, 2012; Ma *et al.*, 2013; Sun *et al.*, 2013; Marchese & Huarte, 2014; Lü *et al.*, 2015].

Un número creciente de estudios caracterizó a diferentes lncRNAs con roles importantes en el desarrollo, como *Xist*, *TUG1*, *PINC*, *Evf2*, y *HOTAIR* [Dinger *et al.*, 2008]. En el trabajo publicado por Sarropoulos y colegas [2019], se estudió la dinámica de la expresión de lncRNAs en siete órganos diferentes, en distintos estadios del desarrollo desde la organogénesis hasta la adultez, y para siete especies distintas de vertebrados. Tras examinar la contribución a los programas de expresión del desarrollo por parte de los lncRNAs, concluyeron que existe una transición en la expresión de lncRNAs durante el proceso, desde una amplia expresión de lncRNAs más conservados hacia un incremento de lncRNAs linaje- y órgano-específicos [Sarropoulos *et al.*, 2019].

Por otra parte, un trabajo de hace un tiempo concluyó que los lncRNAs expresados en el cerebro estarían más conservados que el promedio, y enriquecidos en estructura secundaria. Además, estarían localizados preferiblemente adyacentes a genes de proteínas expresadas en el cerebro, o que participan en la regulación transcripcional o desarrollo del sistema nervioso [Ponjavic *et al.*, 2009].

Los lncRNAs también pueden participar de la organización de estructuras nucleares para el procesamiento de ARN denominadas “*nuclear speckles*”. Estos son dominios sub-nucleares no membranosos que contienen maquinaria de procesamiento de pre-mRNAs y ncRNAs, y que participan en las distintas etapas de transcripción, *splicing* y transporte de mRNAs [Tripathi *et al.*, 2012; Fritah *et al.*, 2014].

En plantas como *Arabidopsis*, donde se han descrito unos 40.000 lncRNAs siendo un 75% antisentido y un 15% lincRNAs, o en otras como arroz, maíz, trigo o manzana, se han identificado miles de lncRNAs que participan en distintos procesos biológicos como vernalización, floración, expresión génica, regulación del fotoperíodo, regulación hormonal, homeostasis de fosfatos, organogénesis, respuesta a estrés abiótico, entre otros [Shafiq *et al.*, 2016; Wang & Chekanova, 2017].

### 2.2.8: LncRNAs y epigenética

Como ya hemos mencionado, muchos lncRNAs actúan como reguladores epigenéticos. Pueden ejercer su efecto controlando la expresión de mRNAs individuales, o incluso de cromosomas enteros. Varios de ellos usan de sustrato a la cromatina para cumplir con sus funciones [Rinn *et al.*, 2007; Pontier & Gribnau, 2011], en tanto otros reclutan complejos modificadores de cromatina a sitios específicos, tales como los del complejo represivo Polycomb 1 y 2 (PRC1 y PRC2) G9a, REST/Co-REST, TLS/CBP/p300, etc. [Sun *et al.*, 2013]. Esta interacción puede suceder de tres maneras: a) los lncRNAs pueden ser cebos para secuestrar modificadores de cromatina, b) los lncRNAs pueden ser guías que dirigen los modificadores a sitios específicos de la cromatina, y c) los lncRNAs pueden actuar como andamio para reclutar más de una proteína modificadora de cromatina [Marchese & Huarte, 2014].

Los lncRNAs también pueden modular la metilación del ADN en regiones específicas del genoma. Un ejemplo es el del *lincRNA-p21*, un lncRNA inducido por el factor p53 que impide la reprogramación celular. Actúa manteniendo marcas represivas en las histonas o en el ADN metilado en los promotores de genes de pluripotencialidad, interaccionando con las metiltransferasas de histonas SETDB1 y de ADN DNMT1, respectivamente [Zhao *et al.*, 2016].

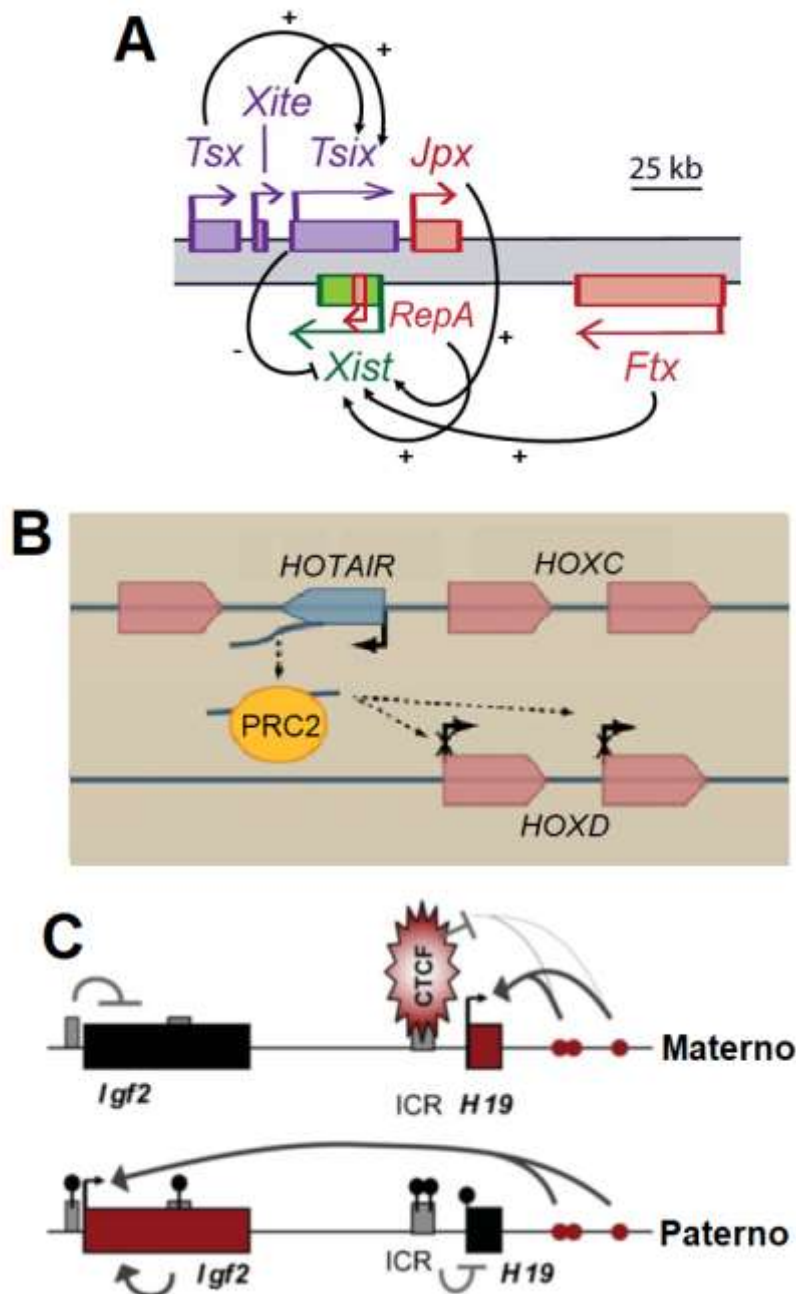
El lncRNA *Xist* es responsable del mecanismo conocido como “compensación de dosis”, que consiste en la inactivación de uno de los cromosomas X en las hembras de los mamíferos, evitando así la síntesis del doble de la cantidad de ARN y proteína a partir de los genes localizados sobre este cromosoma, en hembras que en machos. *Xist* se expresa desde el cromosoma X y a su vez él mismo es su blanco, uniendo una proteína del grupo Polycomb (PRC2) mediante su secuencia repetida en 5’ “Repeat A” (*RepA*), lo que induce trimetilación de la lisina 27 de la histona 3 (H3K27me3), e inicia el silenciamiento de todo el cromosoma X [Ponting *et al.*, 2009; Pontier & Gribnau, 2011; Froberg *et al.*, 2013]. Así como la expresión de *Xist* determina el cromosoma X inactivo (Xi) que va a formar una masa heterocromatínica única denominada cuerpo de Barr, la expresión del lncRNA *Tsix* lo hace con el cromosoma X activo (Xa) [Loda & Heard, 2019]. *Tsix* es un lncRNA antisentido y antagonista de *Xist*, ya que regula la expresión de *Xist* modulando el estado de la cromatina y la metilación del ADN en su promotor [Pontier & Gribnau, 2011; Froberg *et al.*, 2013]. También

participan en el proceso los lncRNAs *Jpx*, que remueve la proteína represora transcripcional CTCF del promotor de *Xist*, *RepA*, que también se une a PRC2, y *Ftx*; todos éstos se consideran activadores de *Xist*. *Xite*, por su parte, es un potenciador de *Tsix* [Froberg *et al.*, 2013; Marchese & Huarte, 2014]. Finalmente, en este proceso de inactivación del cromosoma X, participa también el lncRNA *Tsx*, cuya expresión aumenta hacia paquiteno y disminuye en la entrada a espermiogénesis, y cuya delección lleva a la apoptosis de espermátocitos paquiténicos reduciendo la fertilidad [Anguera *et al.*, 2011] (Figura 5A). Curiosamente, se ha visto que los ratones macho mutantes para *Tsx* tienen afectado su comportamiento y aprendizaje [Anguera *et al.*, 2011]. Una hipótesis que se mantiene en discusión es la idea de que los repetidos largos de ADN (*LINES*), presentes en alta densidad en el cromosoma X humano, tengan un papel en la dispersión de *Xist* en el cromosoma X, interaccionando directamente con los *LINES* del Xi [Pontier & Gribnau, 2011]. Vemos entonces, como el mecanismo de compensación de dosis nos brinda un ejemplo de inactivación epigenética a nivel cromosómico, regulado de manera protagónica por lncRNAs.

Otro ejemplo muy estudiado es el del lncRNA *HOTAIR*. Se ha demostrado que el patrón de expresión del clúster HOX-C depende de *HOTAIR*, que recluta complejos que modifican la cromatina (PRC2) metilando la H3K27 y desmetilando la H3K4 (histona 3 metilada en la lisina 4), mediante su unión al complejo demetilasa LSD1-CoREST. A su vez, los genes del clúster de HOX-D son silenciados en presencia de *HOTAIR* [Rinn *et al.*, 2007; Gupta *et al.*, 2010; Marchese & Huarte, 2014]. El caso de *HOTAIR* es un claro ejemplo de lncRNAs actuando en *trans*; se transcribe desde el clúster de HOX-C, y actúa como represor del clúster HOX-D que está en un cromosoma diferente [Rinn *et al.*, 2007] (Figura 5B).

En el ratón se han identificado más de 150 genes con impronta, que usualmente se localizan en clústeres de tamaño variable. Estos clústeres contienen al menos uno o dos lncRNAs cuya expresión es inversa a la de los genes codificantes del mismo [Kanduri, 2016]. Un ejemplo clásico es la metilación del alelo paterno del gen *H19*. *H19* se expresa en el cromosoma 11 humano solamente a partir del alelo materno, y codifica un lncRNA de igual nombre. El factor CTCF se une a la región de control de imprintado (ICR) no metilada de *H19* y permite que se exprese *H19* pero no *Igf2* (factor de

crecimiento insulínico tipo 2) desde el alelo materno. En el alelo paterno sucede lo opuesto: la región ICR está hipermetilada, por lo que no se une CTCF, y entonces no se expresa *H19* pero sí lo hace *Igf2* [Gabory *et al.*, 2010] (Figura 5C).



**Figura 5.** LncRNAs y epigenética. (A) LncRNAs que participan en la compensación de dosis (tomado y modificado de Anguera *et al.* [2011]). (B) Mecanismo de acción de *HOTAIR* (tomado y modificado de Ponting *et al.* [2009]). (C) Expresión de *H19/Igf2* desde los alelos materno y paterno (tomado y modificado de Gabory *et al.* [2010]).

Por otra parte, existen diferencias a nivel de las modificaciones de histonas que presentan los promotores de lincRNAs en comparación con los de mRNAs, como por ejemplo, que los promotores de lincRNAs están totalmente desprovistos de marcas epigenéticas excepto por la H3K9me3 (marca que inactiva la cromatina) [Melé *et al.*, 2017].

### **2.2.9: LncRNAs asociados a enfermedades y patologías**

La desregulación en la expresión de los lncRNAs puede tener alto impacto en diferentes patologías, como por ejemplo la tumorigénesis, desde la proliferación, resistencia a apoptosis, hasta la metástasis [Morlando & Fatica, 2018], ya que pueden controlar la expresión de otros genes relacionados a esta enfermedad. Mediante ensamblaje *de novo* de lecturas provenientes de secuenciación masiva de muestras humanas de pacientes sanos y con diferentes tipos de cáncer, se identificaron unos 2.550 transcritos nuevos, principalmente lncRNAs. Muchos de ellos exhibían expresión específica, o estaban silenciados en algunos tipos de cáncer [Kazemian *et al.*, 2015].

En los tumores, los lncRNAs pueden actuar tanto como oncogenes o como supresores de tumores [Lü *et al.*, 2015]. Un claro ejemplo de esto es *Xist*, enormemente reprimido en tumor de ovario, lo que conlleva a la sobreexpresión de inhibidores apoptóticos que hacen a la célula tumoral resistente a ciertas drogas [Liu *et al.*, 2018]. Se ha visto también la co-expresión de *H19* con oncogenes en cáncer de ovario y endometrio; la pérdida de impronta en hembras hace que *H19* se sobreexpresa, favoreciendo la migración e invasión de células tumorales [Liu *et al.*, 2018]. *Malat1* se ha relacionado a todo tipo de cánceres, que incluyen cáncer de cuello uterino, carcinoma hepatocelular, de mama, de colon y recto, de ovario, osteosarcoma, etc. Su delección suprime la proliferación, invasión y metástasis de células tumorales [Liu *et al.*, 2018]. El lncRNA *TTN-AS1* promueve la proliferación celular, la formación de colonias y la invasión y migración de las células de cáncer de ovario. Funciona como competidor uniéndose como “esponja” al miR-139-5p, para elevar los niveles de la proteína ROCK2 [Liu *et al.*, 2020 (a)].

Alrededor del 40% de los lncRNAs del genoma humano se expresa en cerebro, donde tienen papeles importantes en la neurogénesis y la plasticidad sináptica, por lo que no resulta sorprendente que su desregulación esté

involucrada en enfermedades neurodegenerativas [Tsagakis *et al.*, 2020]. El incremento en la expresión del transcrito antisentido a la  $\beta$ -secretasa-1 (*BACE1*) ha sido implicado en la progresión del Alzheimer [Ponting *et al.*, 2009]. Asimismo, el lncRNA humano *BC200* o su ortólogo en ratón *BC1* son altamente expresados en el cerebro, regulan la traducción de mRNAs y el mantenimiento de la plasticidad sináptica a largo plazo, y también se han visto involucrados en la enfermedad de Alzheimer [Tsagakis *et al.*, 2020]. La participación de lncRNAs se ha puesto también en evidencia en otras enfermedades neurológicas y neurodegenerativas como el Parkinson [Tsagakis *et al.*, 2020], la esclerosis múltiple [Han *et al.*, 2020], o la ansiedad y desarrollo de otros desórdenes neuropsiquiátricos [Spadaro *et al.*, 2015].

Además, se ha asociado la desregulación o la expresión de lncRNAs a distintas enfermedades como el lupus [Taheri *et al.*, 2020 (a)], psoriasis [Li *et al.*, 2020], aterosclerosis [Liu *et al.*, 2017 (b)], osteoporosis [Guo *et al.*, 2020], enfermedades cardíacas [Kreutzer *et al.*, 2020], artritis reumatoide [Taheri *et al.*, 2020 (b)], entre muchas otras patologías.

Finalmente, dada la condición de ser tejido-específicos, los lncRNAs podrían considerarse como potenciales biomarcadores y/o blancos de terapias génicas dirigidas ante diferentes tipos de patologías [Wu & Du, 2017]. En el cáncer de mama por ejemplo, se encontraron cinco firmas moleculares compuestas de lncRNAs y genes codificantes, que podrían servir como herramientas de pronóstico para tratamientos personalizados, ya que también permiten distinguir entre pacientes de riesgo alto y leve [Liu *et al.*, 2020 (b)].

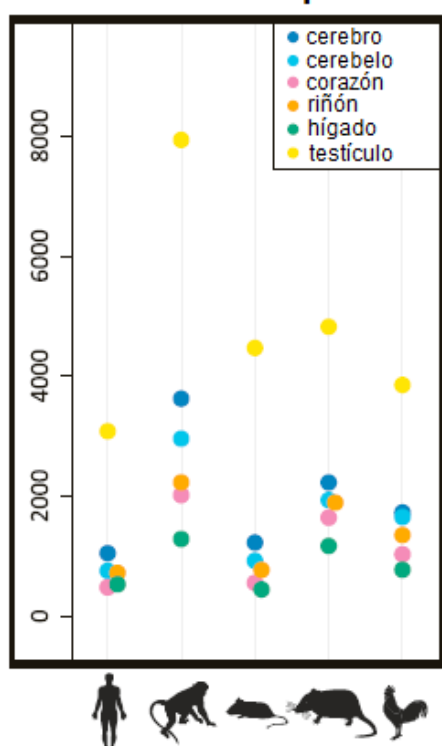
## 2.3: LncRNAs y espermatogénesis

### **2.3.1: Los lncRNAs testiculares**

Soumillon *et al.* [2013] llevaron a cabo un estudio del transcriptoma de diferentes órganos para distintas especies de vertebrados, y concluyeron que el testículo de mamíferos (y también de aves) es el órgano que tiene más amplia transcripción del genoma, tanto de lncRNAs como de genes codificantes. Más aún: la cantidad de lncRNAs expresados es abrumadoramente mayor en el testículo que en otros órganos y tejidos, para todas las especies estudiadas

(Figura 6). Observaron además que las espermátidas y los espermatoцитos son los tipos celulares que expresan el mayor número de lncRNAs, en primer y segundo lugar, respectivamente. En general, los patrones de expresión de los lncRNAs diferencialmente expresados a lo largo del desarrollo del testículo son conservados entre el ratón y la rata, y muestran más conservación evolutiva que el promedio de los lncRNAs, sobre todo en sus promotores [Darbellay & Necsulea, 2020].

### Genes de lncRNAs predichos



**Figura 6.** Genes de lncRNAs predichos en distintos tejidos para 5 especies (humano, mono, ratón, zarigüeya y gallo) (tomado y modificado de Soumillon *et al.* [2013]).

Muchos lncRNAs han sido caracterizados y se les han asignado funciones en distintas etapas de la espermatogénesis, o se han vinculado a la reproducción. Se ha reportado que el silenciamiento del lncRNA *H19* produce una tendencia al descenso del número de células en los túbulos seminíferos. Este lncRNA afecta la expresión del receptor de insulina IGF-1R en células de Sertoli y espermatoγogénicas, donde mantiene la supervivencia celular y tiene funciones esenciales en la reproducción masculina [Sahlu *et al.*, 2020]. *HongrES2* es un lncRNA que se expresa en el núcleo de células testiculares de rata, donde es

procesado en un miRNA, y se lo relaciona a la capacitación y maduración de los espermatozoides [Luk *et al.*, 2014; Liu *et al.*, 2018]. El lncRNA *Neat1* (*nuclear paraspeckle assembly transcript 1*) forma los dominios nucleares (*speckles*), y junto con otros complejos de ARNs modifica la transcripción de mRNAs. Se encontró a *Neat1* expresado en el testículo de la rata, y se lo identificó como un regulador de la espermatogénesis [Liu *et al.*, 2018].

El gen *Prss42/Tessp-2* forma parte de un clúster de genes, los cuales codifican para los componentes de una serín-proteasa crucial para la progresión meiótica. Las regiones flanqueantes al lncRNA testículo-específico *lncRNA-HSVIII* adyacente al gen *Prss42/Tessp-2* en ratón, interaccionan directamente con el promotor de dicho gen actuando como potenciadores para la activación de su transcripción [Yoneda *et al.*, 2016].

El lncRNA *Mrhl* (*meiotic recombination hot-spot locus*) es expresado en el testículo y regula la espermatogénesis a través de una conexión con la vía de señalización Wnt: *Mrhl* se reprime cuando se activa la vía Wnt, y eso permite la progresión meiótica, con la participación del factor de transcripción SOX8 como el enlace regulador en este proceso [Luk *et al.*, 2014; Kataruka *et al.*, 2017]. Además del mecanismo antes mencionado, *Mrhl* también regula la espermatogénesis por un segundo mecanismo que consiste en su fragmentación por la proteína Drosha en ARNs de 80 nucleótidos, que se localizan en el núcleo de células espermátogoniales e interaccionarían con la cromatina [Luk *et al.*, 2014; Liu *et al.*, 2018].

El lncRNA *Gm2044* funciona como una “esponja” para la unión del miRNA miR-355-3p, que a su vez afecta la expresión de su blanco SYCP1, codificante para una proteína del CS. Mediante análisis bioinformáticos se logró predecir dos sitios de unión para el factor de transcripción A-MYB en la región promotora del gen codificante para el lncRNA *Gm2044*, y se ha demostrado que dicha unión promueve su expresión. Es decir que, indirectamente, A-MYB aumenta los niveles de SYCP1 [Liang *et al.*, 2020].

Por otra parte, *lncRNA-Tcam1* (*lncRNA-testicular cell adhesion molecule 1*) y *R53* son otros lncRNAs a los cuales se han atribuido funciones regulatorias en la meiosis [Kurihara *et al.*, 2017; Nakajima *et al.*, 2017].

Se han visto muchos lncRNAs reguladores que actúan en las células madre embrionarias (ESC) de humano y roedor [Luk *et al.*, 2014]. Dos transcriptos

específicos de espermatogonias, *Spga-lncRNAs 1 y 2*, han sido sugeridos como esenciales en el mantenimiento del estado de diferenciación de las mismas [Luk *et al.*, 2014].

El *lncRNA033862* es un transcripto antisentido al gen del receptor alfa-1 del factor neurotrófico GDNF (*Gfra1*), y regula los niveles de expresión de este receptor interaccionando con la cromatina de *Gfra1*. GDNF es un factor de crecimiento requerido para la renovación y supervivencia de las células madre espermatogénicas (SSCs), y dado que la reducción del *lncRNA033862* perjudica la sobrevivencia de las SSCs, este lncRNA se considera esencial para la repoblación testicular luego de un trasplante [Li *et al.*, 2016].

*Dmr (Dmrt1-related gene)* es otro lncRNA específico del testículo, que es capaz de formar un transcripto quimérico interrumpiendo el mRNA de *Dmrt1*. Este último es un factor de transcripción esencial para promover el desarrollo de las espermatogonias y prevenir la meiosis prematura reprimiendo a la proteína STRA8. Se cree entonces que la supresión de *Dmrt1* por *Dmr* podría estar involucrada en la transición entre mitosis y meiosis en el desarrollo de las células germinales [Luk *et al.*, 2014].

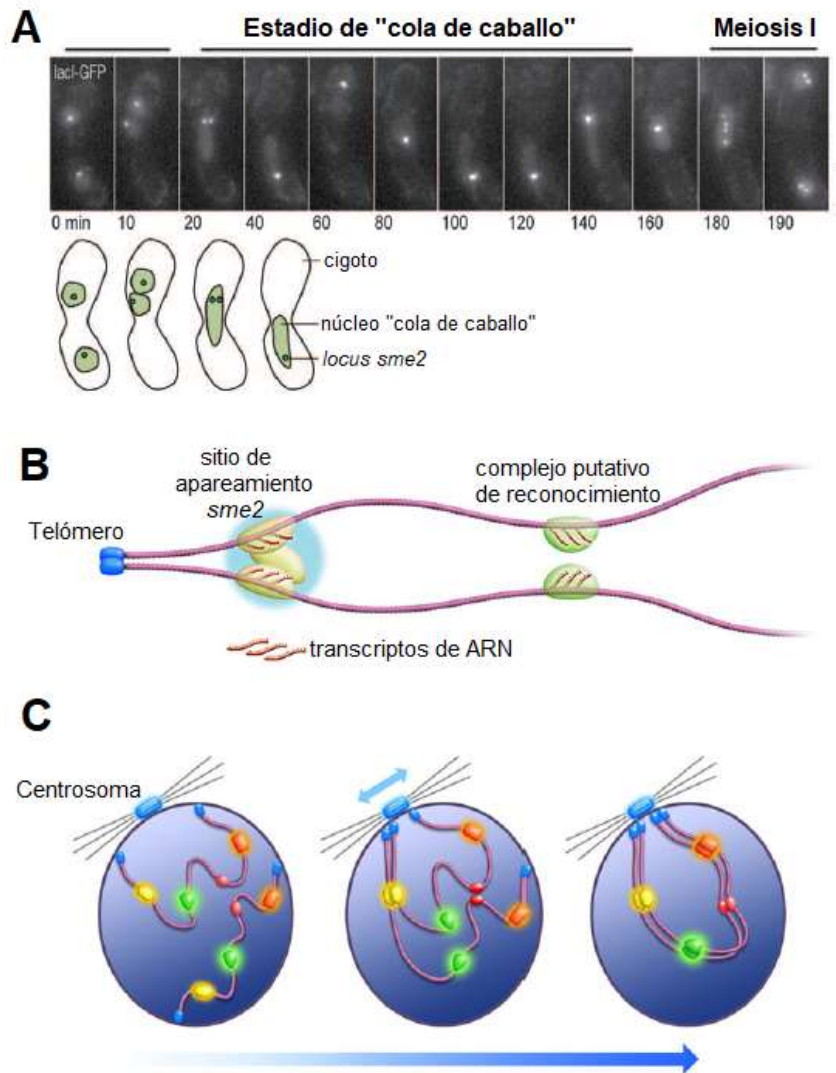
El lncRNA *hsa-lncRNA12238*, conservado en la evolución, es altamente expresado en el testículo humano, y se ha sugerido su intervención en la espermatogénesis, reconocimiento del ovocito por el espermatozoide, y otros procesos reproductivos [Sahlu *et al.*, 2020]. La distribución citoplásmica de los lncRNAs *1700108J01Rik* y *1700101O22Rik* en el testículo del ratón, también sugiere una función en la regulación postranscripcional [Song *et al.*, 2018].

A pesar de la represión transcripcional, incluso los espermatozoides contienen una gran variedad de ARNs, muchos de los cuales son lncRNAs. Algunos de ellos se han identificado como esenciales para la movilidad y la fertilidad espermática [McSwiggin & O'Doherty, 2018]. Existen lncRNAs que pueden ser incluso especie-específicos y reguladores de la espermatogénesis, como se vio en el caso de dos cepas de ratones donde encontraron diferencias en la expresión de ncRNAs en el transcriptoma testicular [Homolka *et al.*, 2011].

Recientemente se ha creado una base de datos de lncRNAs relacionados a la línea germinal, *Germ-lncRNA*, que incluye un catálogo de lncRNAs de tres tipos celulares: espermatogonias tipo A (células madre de espermatogonias), espermátocitos paquiténicos y espermátidas redondas [Luk *et al.*, 2015]. En esta

base recopilan información de repositorios como *Ensembl*, *RefSeq*, UCSC, NONCODE y *fRNAdb*; también incluyen datos de regulación, y de proteínas y miRNAs blancos de lncRNAs.

Finalmente, deseamos destacar la posibilidad de que lncRNAs participen en el reconocimiento de homologías y apareamiento homólogo durante la profase meiótica. Como se explicó anteriormente en la sección 2.1.2, existe un agrupamiento de los telómeros seguido de un apareamiento de cromosomas homólogos previo a la recombinación que ocurre durante la meiosis. Este proceso no puede asegurar la especificidad en dicho apareamiento, por lo que debe haber algún mecanismo que facilite el reconocimiento de los pares homólogos. El único mecanismo de reconocimiento que se ha descrito es en la meiosis de las levaduras, y aparece mediado por un lncRNA. En la levadura *Schizosaccharomyces pombe* se encontró que un ARN retenido en el cromosoma juega un rol en el alineamiento de los cromosomas homólogos [Ding *et al.*, 2012 (a)]; el *locus sme2* codifica para un ARN no codificante poliadenilado que es específico de meiosis, denominado *MeiRNA*. Se encontró que desde este locus se generan dos ARNs no codificantes: uno pequeño de 0,5 Kb, *MeiRNA-S*, y otro de 1,5 Kb desde el extremo 3', *MeiRNA-L* [Ding *et al.*, 2012 (a); Ding *et al.*, 2013]. Además, Mei2 es una proteína esencial para la entrada en meiosis, que es reclutada al *locus sme2* y se une al *MeiRNA-S*, secuestrando a su vez a la proteína Mmi1. *MeiRNA-L* es fundamental para el apareamiento robusto de cromosomas homólogos, formando todos juntos un “cuerpo de ARN” (similar a una “nube” de transcripto) alrededor de este *locus (Mei2-dot)* [Ding *et al.*, 2013]. Curiosamente, para que el *MeiRNA* se localice en su propio *locus* para formar el *Mei2-dot* depende de la asociación con Mmi1; pero éste es un supresor de la meiosis, eliminando los transcriptos meióticos en células vegetativas. Es así que este punto de ARN promueve la meiosis bloqueando a Mmi1 al inicio del proceso, permitiendo a los transcriptos meióticos escapar de la degradación mediada por Mmi1 [Shichino *et al.*, 2014]. Cuando se deleta el *locus sme2* no hay apareamiento en este punto, y cuando el *locus* se transpone a otro sitio, ocurre apareamiento en ese sitio ectópico, por lo que se evidenció su rol en el alineamiento y apareamiento de cromosomas homólogos [Ding *et al.*, 2012 (b)] (Figura 7).



**Figura 7.** Apareamiento homólogo en el *locus sme2* de levadura. (A) Proyección 3D de un *stack* y esquema del núcleo "cola de caballo" de una observación temporal del *locus sme2*. Se aprecia como los dos puntos de *sme2* confluyen en un único punto durante el apareamiento homólogo (tomado y modificado de Ding *et al.* [2012 (b)]). Modelo del reconocimiento y apareamiento del *locus sme2* (B) y de cromosomas homólogos en general (C) mediados por RNAs (tomado y modificado de Ding *et al.* [2013]).

Más recientemente, se describieron otros dos *loci* de apareamiento en la levadura (A55 y A37, además de *sme2*), en los cuales se forman *foci* donde se acumulan complejos lncRNAs-proteínas, los que median el reconocimiento y posterior apareamiento robusto de cromosomas homólogos. Para la formación

de estos *foci* participan tanto proteínas Smp (*sme2 RNA-associated protein*), como lncRNAs (*MeiRNA*, *omt3 RNA* y lncRNA584) [Ding *et al.* 2019]. Creemos que mecanismos mediados por lncRNAs podrían participar también en el reconocimiento de homologías y apareamiento en los eucariotas superiores (y de hecho algunas de las proteínas que actuarían en estos procesos en *S. pombe* están conservadas en los eucariotas superiores; [Ding *et al.*, 2019]), pero los mismos son desconocidos hasta ahora.

### **2.3.2: LncRNAs y patologías testiculares**

Últimamente, han comenzado a surgir evidencias de la participación de lncRNAs en el desarrollo de patologías testiculares, como infertilidad y otras.

Doce mil setecientos lncRNAs (casi un 8% anotados y 92% transcritos nuevos) fueron detectados como diferencialmente expresados en un análisis de expresión entre pacientes con astenozoospermia (es decir, con reducción de la motilidad espermática) y pacientes normales, algunos de ellos formando parte de redes de co-expresión junto con mRNAs y miRNAs [Lu *et al.*, 2020]. Otro trabajo detectó un importante número de lncRNAs diferencialmente expresados entre pacientes con azoospermia no obstructiva (es decir, ausencia de espermatozoides por una causa diferente de una obstrucción mecánica de los conductos de transporte de los mismos) e hipoespermatogénesis (niveles de espermatogénesis por debajo de lo normal), en comparación con pacientes normales [Lü *et al.*, 2015].

El lncRNA *Gm2044* ha sido encontrado altamente expresado en azoospermia no obstructiva con arresto de espermatogonias, al igual que el microRNA miR-202, ya que *Gm2044* es el gen hospedero del miR-202. miR-202 es un supresor de *Rbfox2*, cuya función está involucrada en la auto-renovación y diferenciación de las células madre espermatogoniales [Liang *et al.*, 2019]. En otro estudio sobre el perfil de expresión de lncRNAs en pacientes con azoospermia no obstructiva y los roles de los mismos como competidores endógenos, se identificó al lncRNA *LINC00467* como regulador de la expresión de genes de espermatogénesis compitiendo por la unión a miRNAs [Bo *et al.*, 2020].

*NLC1-C (narcolepsy candidate-region 1 gene)* es un lncRNA citoplásmico expresado en espermatogonias y espermatocitos tempranos, promoviendo el crecimiento celular. Su expresión es más baja en pacientes infértiles con arresto

meiótico [Liu *et al.*, 2018]. *NLC1-C* se une a la Nucleolina y también constituye un ejemplo de lncRNAs que actúan como “esponjas” uniendo miRNAs (en este caso, los precursores de miR-320a y miR-383). Cuando en condiciones anormales *NLC1-C* se acumula en el núcleo de los espermatoцитos, se une a los precursores de los miRNAs, que no pueden ser exportados al citoplasma para su procesamiento, lo que lleva a la hiperproliferación de células germinales y en última instancia a la infertilidad masculina [Lü *et al.*, 2015].

El lncRNA *Tslrn1* (1700019B21Rik) es específico del testículo y está codificado en el cromosoma X. Ratones mutantes con dicho gen deletado no ven afectada su fertilidad, pero sin embargo muestran una reducción significativa en el conteo espermático [Wichman *et al.*, 2017].

Dos estudios han relacionado la expresión aberrante de lncRNAs con el síndrome de Klinefelter (SK) (consiste en la presencia de por lo menos un cromosoma X extra en hombres), y particularmente con el lncRNA *GAS5*, sobreexpresado en testículos de pacientes con el síndrome. Este lncRNA tiene relación con procesos celulares como crecimiento, proliferación, desarrollo, señalización e interacción, por lo que podría relacionarse con la pérdida de células germinales en testículos con SK, aunque esto necesita de más estudios [Winge *et al.*, 2018; Salemi *et al.*, 2019].

Del mismo modo, se ha reportado desregulación de lncRNAs en los tumores de testículo. En particular, se ha detectado la expresión de *Xist* y *H19* en los cánceres de células germinales testiculares [Okamoto, 2012]. Otro lncRNA *SPRY4-IT1*, también aparece vinculado al cáncer de células germinales testiculares, y se ha propuesto que actuaría como oncogén [Das *et al.*, 2018].

## 2.4: Antecedentes de nuestro laboratorio

### **2.4.1: Metodologías de purificación de células espermatoгénicas**

A pesar de los muchos intentos llevados a cabo en las últimas décadas, poco progreso se ha logrado en el establecimiento de sistemas eficientes para cultivo celular *in vitro* de línea germinal [Reuter *et al.*, 2012]. Tal como describimos en la sección 2.1.1, co-existen en el testículo más de 30 tipos celulares diferentes, incluyendo células somáticas y germinales. Pero la baja frecuencia de algunos

tipos celulares como espermatogonias o células en profase meiótica temprana, en relación a otras como células paquiténicas y espermatidas, hace difícil la purificación de esos estadios [Bellvé *et al.*, 1977]. Más aún: a pesar de que muchos de los genes expresados por las células de Sertoli (indispensables para la nutrición, el soporte y el mantenimiento de la línea germinal) han sido identificados, sus funciones específicas en la interrelación con las células germinales permanecen aún desconocidas [Petersen & Soder, 2006]. Por lo tanto, no se han logrado reproducir *in vitro* las condiciones del tejido *in vivo*. Por estos motivos, se vuelve necesario trabajar con el organismo. Dada la gran heterogeneidad celular, lo que se traduce en un obstáculo a la hora de realizar estudios moleculares de poblaciones específicas, muchos esfuerzos se han puesto en el desarrollo de metodologías de purificación celular.

A lo largo del tiempo se han desarrollado varios protocolos para la purificación de poblaciones celulares testiculares. El proceso de sedimentación “*STA-PUT*” involucra acción enzimática para disociación celular seguida de una sedimentación por tamaño celular en una unidad de gravedad y en gradiente de seroalbúmina bovina (BSA) [Bellvé, 1993]. A modo de ejemplo, Bao y colaboradores [2013] obtuvieron sus poblaciones celulares por *STA-PUT* a partir de testículos de ratones de edades crecientes, y posteriormente realizaron un microarreglo con secuencias de lncRNAs y mRNAs.

Por su parte, el método de purificación celular mediante elutriación centrífuga es muy similar al *STA-PUT*, ya que también se basa en la separación de partículas por su velocidad de sedimentación. Cuando la densidad del fluido es menor que la de la célula, ésta sedimenta en dirección gravitacional; cuando el fluido está en reposo, la célula sedimenta al fondo de la cámara; cuando el flujo del fluido es igual y opuesto al de sedimentación de la célula, ésta se mantiene estacionaria. Esto se realiza en un rotor de elutriación [Meistrich, 1977]. En el trabajo de Soumillon y colegas [2013] utilizaron la elutriación para obtener fracciones de espermatocitos paquiténicos y espermatidas redondas, cuya pureza estimada fue de 70% y 90%, respectivamente.

La elutriación alcanza separaciones equivalentes al *STA-PUT* en tiempos más cortos y con mayores números celulares [Meistrich, 1977]. Sin embargo, con ambos procedimientos se obtienen principalmente células paquiténicas y

espermátidas redondas, y ambas poblaciones con un enriquecimiento moderado [Getun *et al.*, 2011].

Más recientemente surge una metodología para purificar células mediante citometría de flujo, y de hecho se ha vuelto la tecnología dominante para la obtención de fracciones enriquecidas o puras de células de testículo, con purezas superiores al 95% [Geisinger & Rodríguez-Casuriaga, 2010]. Esta técnica permite además la obtención de poblaciones celulares de distintas etapas de la meiosis, que incluyen las más tempranas (como leptoteno o cigoteno) [Getun *et al.*, 2011; Gaysinskaya & Bortvin, 2015]. Para la separación e identificación de las diferentes poblaciones celulares se basa en distintos parámetros como ser el tamaño y morfología celulares, complejidad interna, y contenido y compactación del ADN (2C, 4C o C). Para el marcado del ADN se emplea tinción con colorantes fluorescentes, usualmente el *Hoechst*, que tiene por desventaja su excitación en longitud de onda de UV [Getun *et al.*, 2011; Gaysinskaya & Bortvin, 2015], lo que podría ocasionar daño a los ácidos nucleicos [Rodríguez-Casuriaga *et al.*, 2014].

En nuestro laboratorio se desarrolló un método para la separación de células de distintos estadios de la espermatogénesis que se basa precisamente en la citometría de flujo, con altísimo nivel de pureza. Comienza con una disgregación mecánica del tejido testicular, lo cual evita el uso de enzimas ayudando a preservar las macromoléculas. Posteriormente la suspensión celular se tiñe con el colorante vital *Vybrant DyeCycle Green* (VDG) que es excitado en el espectro visible, evitando la luz UV. Finalmente se clasifican las poblaciones celulares por citometría de flujo, con purezas que alcanzan el 98-99% [Rodríguez-Casuriaga *et al.*, 2014; Geisinger & Rodríguez-Casuriaga, 2017].

#### **2.4.2: Análisis del transcriptoma codificante de la espermatogénesis del ratón**

Empleando el protocolo de purificación celular de células espermatogénicas de roedor desarrollado por nuestro grupo de trabajo, se logró analizar el transcriptoma codificante de distintas etapas de la espermatogénesis del ratón mediante *RNAseq* [da Cruz *et al.*, 2016]. Este trabajo logró incluir por primera vez el transcriptoma de las células en profase meiótica temprana (leptoteno/cigoteno), de altísimo interés dado que es en esas etapas en que

ocurren el alineamiento y apareamiento de los cromosomas homólogos. Entre los resultados más destacados del mismo se observó: que una proporción de los genes del proceso meiótico ya están encendidos al inicio de la meiosis; que un alto número de genes tienen pico de expresión en profase meiótica temprana, refutando trabajos antiguos que consideraban que estas células eran transcripcionalmente inertes [e.g. Monesi, 1964]; que los espermátocitos y las espermátidas redondas se caracterizan por la transcripción temprana de genes y el acopio de mRNAs cuyos productos serán utilizados más tardíamente en la espermiogénesis; y que un clúster de genes ligados al cromosoma X escapan a la MSCI [da Cruz *et al.*, 2016]. En particular, este trabajo muestra el enorme alcance de los mecanismos de regulación postranscripcional que operan en las células de la línea germinal masculina, en los que muchos genes son transcritos y sus mRNAs acopiados incluso por semanas, para su posterior utilización para la síntesis de proteínas en las etapas finales de la espermiogénesis, cuando la transcripción ya ha cesado [da Cruz *et al.*, 2016].

Si bien en este trabajo se logró caracterizar la expresión génica durante diferentes etapas de la espermatogénesis del ratón con mucha precisión, las genotecas utilizadas para secuenciación masiva eran no direccionales. Estas genotecas estándar de *RNAseq* no preservan información de la dirección transcripcional. Sin embargo, existen métodos de *RNAseq* hebra-específicos, que permiten revelar la transcripción solapante de ambas hebras. Esto, que carece de relevancia para la caracterización de mRNAs, es esencial para identificar lncRNAs, ya que muchos de ellos se transcriben de la hebra opuesta a otros transcritos, es decir, como lncRNAs antisentido (*antisense*) [Atkinson *et al.*, 2012].

## 2.5: Hipótesis y objetivos

### **2.5.1: ¿Por qué estudiar los lncRNAs en la espermatogénesis?**

El problema de la infertilidad afecta a aproximadamente el 15% de las parejas y, lejos de la creencia por mucho tiempo sostenida de que la misma era un problema femenino, se reconoce actualmente que al menos la mitad de los casos se deben a causas masculinas [Sharlip *et al.*, 2002; McSwiggin & O'Doherty,

2018; Liang *et al.*, 2019]. Algunas causas descritas que pueden contribuir a la infertilidad masculina incluyen anomalías epigenéticas, metilación del ADN, o los ARNs presentes en los espermatozoides [McSwiggin & O'Doherty, 2018]. Esta infertilidad se manifiesta como una reducción en el conteo espermático o ausencia de espermatozoides (oligozoospermia o azoospermia), decremento de la movilidad espermática (astenozoospermia), espermatozoides con morfología defectuosa (teratozoospermia), disfunción sexual, inflamación del tracto reproductivo, o por la combinación de varios defectos [Dong *et al.*, 2018; Liang *et al.*, 2019]. Se asume que entre el 30 y el 50% de los casos de infertilidad masculina están relacionados a causas genéticas [Zorrilla & Yatsenko, 2013], si bien el 75% son diagnosticados como idiopáticos (de causa desconocida) por desconocerse los mecanismos moleculares que los afectan [Lü *et al.*, 2015; revisado en Trovero & Geisinger, 2019]. Por otra parte, los tumores de células germinales testiculares constituyen el tipo de tumor sólido más frecuente en hombres jóvenes en el mundo desarrollado, y su incidencia ha crecido durante las últimas décadas, lo cual resulta altamente preocupante [Znaor *et al.*, 2014; revisado en Trovero & Geisinger, 2019].

En consecuencia, es importante la caracterización de todos aquellos genes, ya sea que generen transcritos codificantes o no codificantes, necesarios para el desarrollo de los gametos masculinos y que están presentes en condiciones normales en las diferentes células del testículo. La participación de lncRNAs en la regulación de procesos esenciales, así como su desregulación en diversas patologías, en conjunto con el hecho de que el testículo es el tejido donde se han detectado mayores niveles de expresión de los mismos, ponen de manifiesto la necesidad de profundizar el conocimiento acerca de estas moléculas, y de los roles que desempeñan en la espermatogénesis. Más aún, precisamente los elevados niveles de expresión de lncRNAs en la espermatogénesis, hacen de este proceso de diferenciación terminal un sistema ideal para la identificación y estudio de lncRNAs.

### **2.5.2: Hipótesis**

Nuestra hipótesis de trabajo es que existen lncRNAs que se expresan diferencialmente en alguna/s de las etapas estudiadas de la espermatogénesis del ratón, estadios donde cumplen distintas funciones mediante diferentes

mecanismos. Algunos de ellos participan en los procesos que suceden durante la meiosis, como por ejemplo a nivel del apareamiento y recombinación de cromosomas homólogos, mientras otros cumplen roles en la diferenciación de las espermátidas durante la espermiogénesis, participando por ejemplo en los mecanismos de regulación postranscripcional, entre otros.

### **2.5.3: Objetivo general**

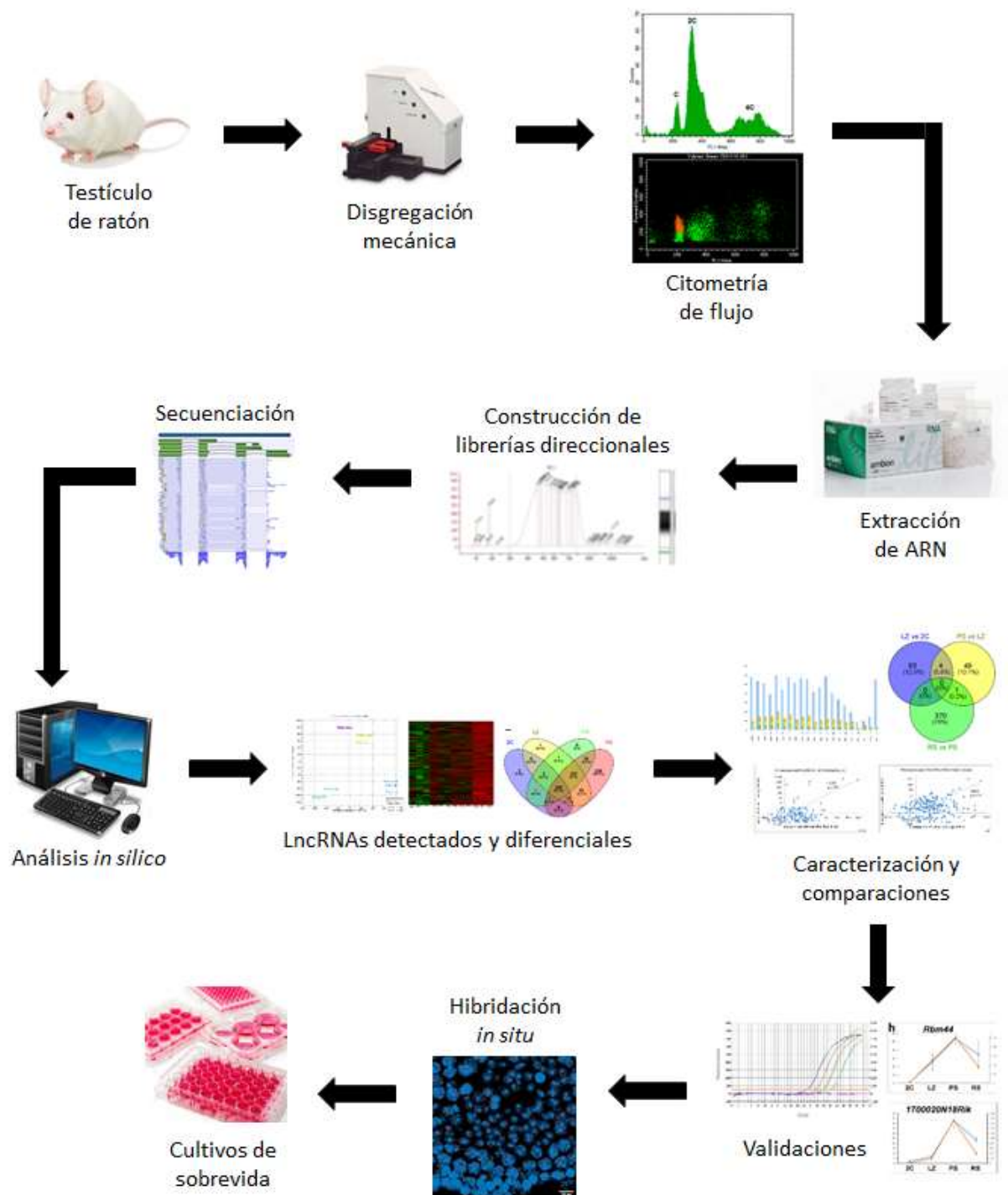
El objetivo general de esta tesis fue caracterizar el transcriptoma de los ARNs no codificantes largos en distintas fases de la espermatogénesis (pre-meiosis, meiosis y post-meiosis). Con ello, contribuir al conocimiento de las bases moleculares de la espermatogénesis del ratón.

### **2.5.4: Objetivos específicos**

Como objetivos específicos, nos planteamos:

- a) Obtener fracciones celulares puras, representantes de pre-meiosis, profase meiótica temprana, profase media/tardía y post-meiosis del testículo del ratón.
- b) Detectar los lncRNAs expresados y diferenciales de las etapas mencionadas en el objetivo anterior.
- c) Describir el transcriptoma global de lncRNAs en cuanto a sus características de transcritos, distribución cromosómica, co-expresión con genes codificantes, conservación, entre otras. Comparar con otros trabajos.
- d) Estudiar la localización de lncRNAs candidatos que posean ciertas características deseadas o de interés, contribuyendo a la caracterización de los mismos.

# **3. MATERIALES** **Y MÉTODOS**



**Figura 8.** Esquema del flujo de trabajo de la tesis.

En la figura 8 se muestra de modo esquemático, el flujo general del trabajo realizado.

### 3.1: Animales

Los ratones (*Mus musculus*) machos de la cepa *CD1-Swiss* se obtuvieron del bioterio del Instituto de Higiene (Facultad de Medicina, UdelaR, Uruguay). Los animales fueron sacrificados mediante dislocación cervical, de acuerdo a la Ley Nacional de Experimentación Animal uruguaya (ley 18.611). El protocolo de experimentación fue aprobado por la Comisión Nacional de Experimentación Animal del Instituto de Investigaciones Biológicas Clemente Estable (protocolo 001/02/2012; código: 008/11; <http://www.cnea.org.uy/index.php/instituciones/registro/10>).

Inmediatamente luego de la eutanasia, se realizó la disección de los testículos, se colocaron en tampón PBS 1X (ver Anexo 1) tratado con dimetil pirocarbonato (DMPC) 0,1%, y rápidamente se siguió con los distintos procedimientos.

### 3.2: Clasificación celular por citometría de flujo

#### **3.2.1: Preparación de suspensiones celulares**

Para la obtención de poblaciones celulares de distintas etapas de la espermatogénesis se partió de ratones de edades crecientes. Para la clasificación de la población de espermatogonias mezclada con células somáticas (que denominamos 2C, por su contenido 2C de ADN) se utilizaron machos de 12-14 días post-parto (dpp). Para purificar la población de células en leptoteno-cigoteno (LZ, es decir profase meiótica temprana) empleamos ratones de 15-18 dpp; usamos machos de 16-19 dpp para clasificar la población de células paquiténicas (PS, por *pachytene spermatocytes*, es decir, profase meiótica media/tardía), y finalmente ratones de 22-24 dpp para obtener la fracción de espermátidas redondas (RS, por *round spermatids*, es decir, células post-meióticas, en etapa de espermiogénesis). Vale aclarar que si bien la fracción 2C es una fracción mezclada ya que contiene varios tipos celulares, constituye una población control de células somáticas y pre-meióticas.

Luego de la eutanasia y disección testicular, la túnica albugínea fue removida de los testículos, y se prosiguió con un protocolo desarrollado previamente en nuestro laboratorio para preparar las suspensiones celulares [Rodríguez-Casuriaga *et al.*, 2011; Rodríguez-Casuriaga *et al.*, 2013; Rodríguez-Casuriaga *et al.*, 2014; da Cruz *et al.*, 2016; Geisinger & Rodríguez-Casuriaga, 2017]. Brevemente, el protocolo

comenzó con la disgregación mecánica del tejido en 1 ml de medio *Dulbecco's Modified Eagle's Medium* (DMEM; Capricorn Scientific GmbH) suplementado con 10% de suero fetal bovino (SFB), utilizando el sistema *BD™ Medimachine System* (Becton Dickinson, USA). En este punto se introdujo una modificación al protocolo únicamente para la obtención de la fracción 2C: una incubación con colagenasa 0,6 U/ml por 15 minutos a temperatura ambiente antes de la disgregación, ya que comprobamos que favorece la liberación de espermatogonias. A continuación se realizaron dos filtrados en mallas de nylon de 50 µm de poro, entre los cuales se le adicionó 0,2% final de ácido 2-naftol-6,8 disulfónico (NDA) a la suspensión celular para disminuir la agregación de células. Se realizó un recuento celular en cámara de Neubauer para ajustar la concentración a  $1 \times 10^6$  células/ml, en medio DMEM suplementado. Las suspensiones celulares testiculares fueron posteriormente teñidas con el colorante vital de ADN *Vybrant DyeCycle Green* (VDG) (Invitrogen, Life Technologies, Carlsbad, CA), durante 1 hora, incubadas a 35°C con leve agitación a 80 revoluciones por minuto (rpm). Finalmente, se mantuvieron a temperatura ambiente por 30 minutos, previo a la clasificación.

### **3.2.2: Citometría de flujo**

Las fracciones celulares fueron clasificadas en un citómetro de flujo y clasificador celular *FACSVantage* (Becton Dickinson, USA), utilizando un láser iónico de argón (Innova 304, Coherent, USA) emitiendo a 488 nm de longitud de onda (100 mW de potencia). Se seleccionó una boquilla de 70 µm y el protocolo citométrico empleado fue basado en lo reportado previamente por nuestro grupo [Rodríguez-Casuriaga *et al.*, 2014]. La modalidad de clasificación celular fue configurada en modo *Counter*, que alcanza una pureza superior al 95%. Se realizaron *pools* de hasta un máximo de 5 animales para la obtención de cada fracción. Se obtuvieron 3 triplicados biológicos de las 4 fracciones celulares, lo que hace un total de 12 muestras, que se recolectaron en PBS 1X tratado con DMPC 0,1%. Las mismas se centrifugaron a 500 g por 10 minutos y a 4°C, se congelaron rápidamente en nitrógeno líquido y se guardaron a -80°C hasta su posterior uso.

La viabilidad de las fracciones celulares se estimó mediante tinción con Ioduro de Propidio (IP). Para ello, se incubaron 499 µl de suspensión celular con 1 µl IP a 1 mg/ml durante 15 minutos a temperatura ambiente, previo al análisis citométrico. En otros casos, la viabilidad celular se estimó mediante tinción con azul de tripán al

0,08%, incubado por 5 minutos a temperatura ambiente, con posterior conteo celular en el microscopio óptico.

### **3.2.3: Análisis de la pureza de las fracciones celulares obtenidas**

La pureza de las fracciones LZ y PS ya había sido analizada en trabajos anteriores de nuestro grupo de investigación [Rodríguez-Casuriaga *et al.*, 2011; Rodríguez-Casuriaga *et al.*, 2014; Geisinger & Rodríguez-Casuriaga, 2017]. Por consiguiente, en esta instancia se corroboró la pureza de la fracción RS mediante la inmunodetección de la proteína MVH (*Mouse Vasa Homolog*, también conocida como DDX4) que además de presentar una señal difusa en el citoplasma, marca intensamente el cuerpo cromatoide, organelo típico de las células post-meióticas [Toyooka *et al.*, 2000]. Para ello se realizó una clasificación en placa con fondo de vidrio (D35-20-1.5-N, Cellvis, Mountain View, CA) por citometría de flujo (tal como fue explicado en los puntos 3.2.1 y 3.2.2) de células de la fracción RS. Las placas fueron previamente tratadas con una solución de poli-L-lisina 0,01% (Sigma-Aldrich) por 15 minutos a temperatura ambiente, y se dejaron secar toda la noche. Luego de la deposición de las células en las placas, se dejaron decantar las mismas por 90 minutos a 4°C en cámara húmeda. Se fijaron con paraformaldehído (PFA) 2% en PBS 1X durante 10 minutos a temperatura ambiente. Se realizaron 2 lavados de 5 minutos con PBS 1X a temperatura ambiente y en cámara húmeda. Las células se permeabilizaron con Tween20 al 0,1% en PBS 1X por 5 minutos a temperatura ambiente. A continuación se repitieron dos lavados con PBS 1X. Se incubaron con anticuerpo primario anti-MVH de conejo (ab13840, Abcam), diluido 1:2000 en PBS 1X en presencia de un cóctel de inhibidores de proteasas (P2714, Sigma), toda la noche a 4°C en cámara húmeda. Al día siguiente se realizaron 3 lavados con PBS 1X como fue explicado anteriormente. Se incubaron las placas con anticuerpo secundario anti-conejo conjugado a Alexa Flúor 488 diluido 1:1000 (A-11034, Invitrogen), en cámara húmeda y protegidos de la luz, por 30 minutos a 37°C. Se realizaron nuevamente dos lavados con PBS 1X. Las células se tiñeron con DAPI 1 µg/ml en PBS 1X, por 10 minutos a temperatura ambiente. Finalmente, las placas se montaron con una gota de medio de montaje *ProLong Diamond Antifade Mountant* (P36961, Invitrogen), se colocó cubreobjetos y se observaron al microscopio confocal (ver 3.13.1 más adelante).

### 3.3: Extracciones de ARN, construcción de genotecas y secuenciación masiva

#### **3.3.1: Extracción de ARN y cuantificación**

El ARN total de cada una de las 12 muestras fue extraído utilizando el kit *PureLink RNA Mini Kit* (Ambion, Life Technologies), siguiendo las indicaciones del fabricante. Se cuantificó mediante fluorimetría en el fluorómetro *Qubit 2.0* empleando el kit *RNA HS Assay kit* (Life Technologies), según las instrucciones del fabricante.

#### **3.3.2: Construcción de librerías de secuenciación hebra-específicas, controles cuanti- y cualitativos y secuenciación del ARN (RNAseq)**

Construimos librerías hebra-específicas para secuenciación masiva, utilizando el kit *Ovation RNA-Seq 1-16 for Mouse* (NuGEN, CA), siguiendo el protocolo del fabricante, con mínimas modificaciones. Se partió de 60 ng de ARN total de cada muestra sin fragmentar. Se amplificó el material con 16 ciclos de PCR, de acuerdo a las instrucciones. La concentración de las librerías se estimó por fluorimetría utilizando nuevamente el *Qubit 2.0* pero esta vez con el kit *dsDNA Assay kit* (Life Technologies). La calidad de las librerías construidas se testeó mediante corrida en un chip para ADN del analizador *2100 Bioanalyzer System* (Agilent, Santa Clara, CA). Para la secuenciación de las 12 librerías se utilizó el servicio de la empresa *Fasteris* (Suiza). Las mismas fueron secuenciadas en una plataforma *Illumina HiSeq4000*, desde ambos extremos, y generando lecturas de 150 pares de bases (pb). Todos los datos crudos fueron depositados en el repositorio de SRA (*Sequence Read Archive*) con el código PRJNA548952.

### 3.4: Análisis *in silico* de los datos de RNAseq

En líneas generales, lo primero que se realizó fue un filtrado de las lecturas, y luego se procedió al mapeo de las lecturas remanentes contra el genoma de referencia de ratón. En nuestro caso, se utilizó el genoma de la base de datos *Ensembl* versión 93 (Gm38.p6). Se calculó una estimación de la abundancia de los transcritos, y se analizó estadísticamente la expresión de los mismos en las diferentes muestras.

### 3.4.1: Procesamiento de las secuencias obtenidas y generación de datos transcriptómicos

Para el análisis de los datos de secuenciación se utilizaron dos abordajes en paralelo, uno basado en paquetes de *software* libre y otro en un paquete comercial, con el objetivo de realizar los análisis con aquellos datos confirmados por más de una metodología. El primero de ellos se basó en el uso del *software* comercial *CLC Genomics Workbench* 10.1.1 (CLC Bio, Aarhus). El filtrado de las lecturas se realizó con el mismo *software* considerando los siguientes parámetros: calidad  $Q > 20$ , largo mayor a 50 pb y se eliminaron los adaptadores *Illumina*. Con las lecturas que pasaron el filtrado se realizó un análisis de *RNAseq*, obteniendo las listas de genes expresados en cada muestra. Los parámetros del mapeo en el genoma de referencia fueron los siguientes: costo de apareamiento incorrecto de bases 2; costo de inserción de base 3; costo de delección de base 3; longitud de fracción 0,8; número máximo de mapeos por lectura 10. Finalmente se calcularon los genes expresados diferencialmente entre las distintas etapas de la espermatogénesis mediante un análisis estadístico basado en una normalización por TMM (*trimmed mean of M-values*), asumiendo una distribución binomial negativa y modelando los genes por un modelo lineal generalizado (GLM). Este análisis estadístico, basado en la prueba de Wald, compara todos los grupos de muestras de a pares, seleccionando nosotros aquellas comparaciones en orden cronológico siguiendo la primera onda espermatogénica (2C-LZ; LZ-PS; PS-RS).

El segundo abordaje empleado fue una combinación de paquetes de *software* libres *HISAT-StringTie-Ballgown* (HSB), adaptado del trabajo de Perteau *et al.* [2016]. El filtrado de las lecturas se llevó a cabo con el *TrimGalore* 0.4.3 (Babraham Bioinformatics, Cambridge), utilizando los mismos parámetros de calidad y largo que con CLC, y removiendo también los adaptadores *Illumina*. El mapeo contra el genoma de referencia fue realizado con el *software* *HISAT* 2.0 [Kim *et al.*, 2015], creando un genoma indexado, usando una referencia enmascarada. *StringTie* 1.3 [Perteau *et al.*, 2015] se utilizó para estimar la abundancia de los transcritos, y para el análisis estadístico se usó el paquete *Ballgown* [Frazee *et al.*, 2015] del repositorio *R*, también para comparaciones pareadas siguiendo la cronología de la onda espermatogénica.

Para considerar genes como expresados en una muestra, se tomaron como criterios una varianza (VAR) entre todas las muestras  $> 1$  y FPKM (fragmentos por millón de kilobases secuenciadas)  $\geq 0,1$ . Los criterios para seleccionar los genes

diferencialmente expresados (DE) fueron los siguientes: un valor estadístico FDR (*False Discovery Rate*)  $\leq 0,05$  [Benjamini & Hochberg, 1995] y un  $\log_2$  de FC (*Fold-Change*, el valor relativo del cambio en la expresión)  $\geq |2|$ .

Las listas de genes expresados y DE entre las distintas etapas espermatogénicas obtenidas mediante ambos procesamientos, fueron luego cruzadas con una lista de genes de lncRNAs anotados en el genoma de ratón en la base de datos *Ensembl* versión 93, obteniendo de esta manera las sub-listas conteniendo únicamente los genes de lncRNAs expresados y DE. Finalmente, se cruzaron las listas de ambos abordajes entre sí, y se prosiguió con los análisis únicamente con los genes de lncRNAs coincidentes.

#### **3.4.2: Análisis de la reproducibilidad de las réplicas**

Con el objetivo de verificar la reproducibilidad de las 12 muestras (las 4 poblaciones celulares por triplicado) se llevaron a cabo dos tipos de estudios. Por un lado, se realizó un análisis de componentes principales (PCA) en el programa CLC, que usa los valores normalizados de conteos por millón (CPM). Paralelamente, también se construyó una matriz de correlación para todas las muestras mediante la función *cor()* en la plataforma *R Bioconductor* 3.4.4 (<http://www.R-project.org>), calculando el coeficiente de Pearson con los valores normalizados de FPKMs.

#### **3.4.3: Construcción de diagramas de Venn y mapas de expresión**

Los diagramas de Venn fueron construidos utilizando la herramienta web *Venny* 2.1 (<http://bioinfogp.cnb.csic.es/tools/venny/>). Los mapas de expresión (*heatmaps*) se construyeron utilizando la función *heatmap()* del *R Bioconductor* versión 3.4.4 (<http://www.R-project.org>).

#### **3.4.4: Análisis de largo y número de exones, biotipos, y distribución cromosómica de los lncRNAs**

La clasificación de los lncRNAs según su largo y número de exones, se basó en la anotación presente en la base de datos del *Ensembl* para el genoma de ratón versión 93. Asimismo, la categorización de los lncRNAs según su biotipo se basó en la anotación de la misma base de datos, que los divide en 7 categorías: intergénicos (entre genes; *lincRNAs*), antisentido (que solapan con exones o intrones de genes codificantes en la hebra opuesta; *antisense*), sentido exónicos (solapan con exones

de genes codificantes en la misma hebra; *sense\_overlapping*), intrónicos sentido (solapan con intrones de genes codificantes en la misma hebra; *sense\_intronic*), lncRNAs macro (ARNs no codificantes sin procesar de varias kb de largo; *macro\_lncRNAs*), solapantes en 3' (solapan con regiones 3'-UTR de genes codificantes en la misma hebra, *3'-overlapping\_ncRNAs*) y promotores bidireccionales (originados de promotores de genes codificantes en sentido contrario de la hebra opuesta; *bidirectional\_promoter\_lncRNAs*).

Para determinar si los lncRNAs mostraban una tendencia en la distribución cromosómica, se empleó una prueba hipergeométrica (mediante la función *1-phyper()*) en el *R bioconductor* versión 3.4.4 (<http://www.R-project.org>), el cual calcula un p-valor estadístico para el enriquecimiento o depleción de genes en los distintos cromosomas, tomando en cuenta también los lncRNAs totales anotados y expresados en cada uno.

#### **3.4.5: Análisis de co-expresión con genes codificantes vecinos o solapantes**

Para los dos biotipos de lncRNAs más expresados, intergénicos y antisentido, se estudió la expresión de sus genes codificantes vecinos a < 300 kb (con cortes intermedios a < 100 kb, < 30 kb y < 10 kb) o solapantes, respectivamente. Para ello, buscamos las coordenadas génicas de nuestros lncRNAs y genes codificantes expresados en nuestras listas en la base de datos *Ensembl* v93 con la ayuda de la herramienta BioMart, y generamos listas en formato BED. Para los lincRNAs, agregamos en ambos extremos 300, 100, 30 o 10 Kb a las coordenadas de la lista, y sobre ellas aplicamos *BEDTools Intersect*, de forma de obtener los genes codificantes vecinos de lincRNAs a 300, 100, 30 y 10 Kb. Se trabajó con los pares de lncRNA/gen codificante en que ambos estuvieran expresados en nuestras listas, y además al menos uno de ellos fuera DE en alguna de las transiciones de estadios. Se calculó la correlación en la expresión de estos pares de genes mediante el coeficiente de Pearson con su p-valor estadístico correspondiente.

#### **3.4.6: Análisis de ontología de los genes codificantes vecinos**

También para el caso de los genes codificantes vecinos o solapantes de lincRNAs y lncRNAs antisentido se llevó a cabo una anotación funcional con la herramienta presente en la web *David Bioinformatics Resources* 6.8 (<https://david.ncifcrf.gov/>). Se incluyeron en este análisis las categorías de “componentes celulares”, “funciones moleculares”, “procesos biológicos” y “tejidos”.

### **3.4.7: Análisis de conservación en humanos**

Para evaluar si los lncRNAs expresados en el ratón durante la espermatogénesis también se encontraban en el genoma humano, ya sea por homología de secuencia (para lo cual se empleó la herramienta BLAST del NCBI, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) o por sintenia conservando la posición relativa en el genoma, nos basamos en la anotación del genoma de *Homo sapiens* del *Ensembl* versión 93 (GRCh38.p12). Paralelamente también se utilizó la anotación de la base de datos de humano *Chess* (<http://ccb.jhu.edu/chess/>) [Perte et al., 2018], pero solamente para el caso de los antisentido, ya que *Chess* no discrimina entre otros biotipos de lncRNAs.

### **3.4.8: Re-análisis de datos de otros trabajos de RNAseq para comparación con nuestras listas**

Los datos crudos de los trabajos de Lin et al. [2016] (PRJNA305485) y Wichman et al. [2017] (PRJNA393538) fueron descargados del repositorio ENA (*European Nucleotide Archive*). Para el primer caso, se utilizaron los datos de las muestras de duplicados de PS y RS (el trabajo de Lin et al. solamente comparó muestras por duplicado) ya que dicho trabajo no contaba con una fracción espermatogonias completa, comparable con nuestra fracción 2C. Para el segundo, se emplearon las muestras triplicadas de espermatogonias, PS y RS. Señalamos que nuestra fracción LZ no pudo ser comparada, ya que ninguno de ambos trabajos contaba con una fracción de espermatoцитos en profase temprana. Estos datos fueron re-analizados utilizando el paquete comercial *CLC Genomics Workbench* (ver 3.5.1), y luego las listas obtenidas fueron cruzadas con nuestras listas generadas en CLC de lncRNAs expresados. Se calculó el coeficiente de Pearson y su p-valor estadístico para analizar la correlación entre los datos. No se utilizó *Ballgown* para esta comparación dado que, como hemos mencionado, el trabajo de Lin et al. solamente contaba con 2 réplicas, y el uso de *Ballgown* está desaconsejado para menos de 3 réplicas.

### **3.4.9: Analisis de datos de lncRNAs utilizando la base de datos de CLS (Capture Long-Read Sequence)**

Por otra parte, se re-analizaron nuestros datos de RNAseq con el conjunto de programas *HISAT-StringTie-Ballgown* (ver 3.5.1) pero utilizando la anotación de

lncRNAs que surge del trabajo de Lagarde *et al.* [2017], en el cual se capturan y secuencian lecturas completas y más largas (CLS, *Capture Long-Read Sequence*). Esta anotación se encuentra disponible en un portal web (<https://public-docs.crg.es/riguigo/CLS/>), y se decidió trabajar con los transcriptos construidos con el método “*anchored*” surgidos del análisis de todos los tejidos de ratón. Del mismo portal se descargó un archivo de identificadores (IDs) de lncRNAs (“*Transcript-to-biotype*”) con los equivalentes de los IDs de *Ensembl* y los de esta anotación, para poder llevar a cabo las comparaciones con nuestras listas. Decidimos trabajar con los transcriptos definidos como “lncRNAs” en esta anotación, y consideramos aquellos cuya expresión fuera  $FPKM \geq 0,1$  y varianza  $\geq 1$ . En el caso de los DE, se trabajó con los transcriptos cuyo  $\log_2(FC) \geq |2|$ , y el  $p\text{-valor} < 0,01$  para las transiciones 2C-LZ y LZ-PS, o  $FDR\ p\text{-val} < 0,05$  para PS-RS. Es pertinente aclarar que el trabajo de Lagarde y colaboradores se centra en el estudio de lncRNAs únicamente, por lo que en la comparación con nuestro análisis con la anotación de *Ensembl* se consideraron solamente los lncRNAs. Se calculó el coeficiente de Pearson y su  $p\text{-valor}$  estadístico para analizar la correlación entre los análisis con ambas anotaciones.

### 3.5: Validaciones de patrones de expresión y de co-expresión con genes codificantes mediante RT-qPCR

Para las validaciones se procedió a la realización de nuevas purificaciones de cada una de las fracciones celulares 2C, LZ, PS y RS, por triplicado. Las poblaciones celulares fueron clasificadas utilizando el protocolo de clasificación celular por citometría de flujo descrito en el punto 3.2, pero mediante el uso de un citómetro *MoFlo Astrios EQ* (Beckman Coulter, USA), en modo “*Purify*” que permite la máxima pureza. Partiendo de 3.000 células purificadas por cada una de las fracciones celulares, se realizaron reacciones de RT-qPCR empleando el kit *Power SYBR Green Cell-to-Ct* (Ambion, Life Technologies), siguiendo las instrucciones del fabricante. La lisis de cada fracción celular se llevó a cabo en un volumen de 50  $\mu\text{l}$ , y se tomaron 22,5  $\mu\text{l}$  de cada reacción de lisis para las reacciones de transcripción reversa, en 50  $\mu\text{l}$  de volumen final. El paso de qPCR se realizó en el equipo *CFX96 Touch Real-Time PCR Detection System 1* (BioRad, Hercules, CA), partiendo de 2  $\mu\text{l}$  de ADN copia (ADNc) en 20  $\mu\text{l}$  de volumen final de reacción, y usando cebadores diseñados específicamente

con la herramienta *PrimerBLAST* del portal *NCBI* (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Los juegos de cebadores diseñados y empleados en las reacciones de RT-qPCR se presentan en la Tabla 1. Se realizó una cuantificación relativa por el método  $2^{-\Delta\Delta C_t}$  [Livak & Schmittgen, 2001], usando el mRNA del gen *Surf4* (2C: 55,99 FPKM, LZ: 59,91 FPKM, PS: 59,36 FPKM, RS: 45,08 FPKM) como gen normalizador y la condición 2C como calibradora. La eficiencia de los cebadores fue calculada, y en todos los casos fue superior al 93%.

**Tabla 1.** Cebadores para RT-qPCR diseñados y utilizados en la tesis.

|                      | Sentido (5'-3')          | Reverso (5'-3')          |
|----------------------|--------------------------|--------------------------|
| <i>Tnp1</i>          | CATCACAAGTGGGATCGGTA     | TCAAGAGAGGTGGAAGCAAGA    |
| <i>1700027A15Rik</i> | CAACCAGAACCTCACTGCTG     | GCTTCTTCAGCCAAAGCATGTG   |
| <i>Dazl</i>          | CTAGGCAGCCACCTCACG       | TCCATCCTAACATCAATTCCTCC  |
| <i>AY702103</i>      | TCCGACCTGCTGGGATTCTA     | ACTTTGGGTGCCTCAGATGG     |
| <i>Actb</i>          | GTGTGACGTTGACATCCGTAAAG  | CCGGACTCATCGTACTCCTG     |
| <i>Rbakdn</i>        | CAACAGCCCATCCTTTTGCAG    | CTGGACCACATAGAAGCCACC    |
| <i>Gapdhs</i>        | GCTGGCATCCTTGCTTACAC     | CATTGAGGGCAATTCCAGCC     |
| <i>Tmem147os</i>     | CCACCCCTTCTCCACTACTC     | AAGCACCCAACCCTCTGTAG     |
| <i>Pdzk1</i>         | AAGCTGGTGTCTGGCTGATG     | TCAAAGTGGCTGTCTCCCTC     |
| <i>4930442L01Rik</i> | AGCCTTTAAGACGTGGGAGC     | TGTGGAAACGGACCTCTCTG     |
| <i>Spata2l</i>       | CCTGTTCACTGATGGCCTGTG    | GCACATAGCCCCCTGAGAAG     |
| <i>4933417D19Rik</i> | GCACCACTGATGCCATACC      | GAGCCTATGCCTCCACCTTG     |
| <i>Lrp8</i>          | CATCAAACGGTGCAACCAGG     | TCATCACACACTTTCCCGCC     |
| <i>Lrp8os1</i>       | TCAAACCAAGGATCCCACCC     | CATGAGAGGATCTCGGCACG     |
| <i>1700047F07Rik</i> | TCGCCTACTCCTGCGTCTAC     | ATCGGTGCGAAGGACATGG      |
| <i>Fam181a</i>       | AAGGTGAAGTGTCTGTCTGTGGTC | CTGGACGCCAGGTTGACAAAATTC |
| <i>Gm29508</i>       | AGACAGACAATCGGATGGG      | AGATGCAGGTCTTCGTTGG      |
| <i>Rbm44</i>         | CAGGTGCTGTCAGAAGACCAC    | GCTGCGCCTAACTCCTTCTTAG   |
| <i>1700020N18Rik</i> | AGGATCGTGCCTACCTTCAAC    | ACAGGAGAAAGAGCAAGGGTC    |
| <i>Akap1</i>         | CGCTACGTGGACTATGGTGG     | CCAGTCGCACTGTAGCTTGTC    |
| <i>C030037D09Rik</i> | GAAGGCTACTACGTCACTGCAC   | GAGTGAAGGTAGCGCCATGTC    |
| <i>Nkx2-6</i>        | GGAACCCAGTAAAGCCTCGG     | CTCCAAGTCGGGAGATCGTG     |
| <i>1700092C10Rik</i> | CTGTCCTCCATTGGTTCCAC     | GTCCCTTCTGGGTTTGCTTGG    |
| <i>Lyzl4</i>         | CGGCCTCTTTCAGATTCGTG     | TTTCTTGGCACACTGGATCG     |
| <i>Lyzl4os</i>       | TGGGAAGAGCTGAGAGTAGCC    | ATGAAGCAGAGGAAGGAGAATCA  |
| <i>Surf4</i>         | TGCTTTGGGCTGTTTGAATC     | GGTTGGGACACCAGCAAACA     |

### 3.6: Hibridación *in situ* e inmunolocalización

#### **3.6.1: Preparación de criosecciones de tejido testicular**

A partir de testículos de ratones de 25 dpp, se generaron bloques de tejido para criosecciones. Posteriormente a la eutanasia y dislocación cervical, se realizó la

dissección de los testículos. Sin quitar la túnica albugínea, se cortaron al medio y colocaron en PFA 4% en solución PHEM 1X (ver Anexo 1) pH 7,4 durante 1 hora a temperatura ambiente. Se realizó un lavado en PHEM 1X por 10 minutos a temperatura ambiente. El tejido se deshidrató primeramente en una solución de sacarosa 15% en PHEM 1X a temperatura ambiente, hasta que el tejido se hundió en el fondo del tubo. Se traspasó luego a una solución de sacarosa 30% en PHEM 1X, y se dejó toda la noche a 4°C, de modo que el tejido nuevamente se hundiera al fondo del tubo. La sacarosa actúa como agente crioprotector. Se colocaron los fragmentos de tejido en soportes conteniendo medio de congelación Jung (Leica Microsystems Nussloch GmbH, Alemania), y se cubrieron por completo con el medio. Finalmente, se congelaron los bloques rápidamente en nitrógeno líquido, y luego se almacenaron a -80°C.

Las criosecciones de testículo se realizaron en un crióstato SLEE modelo MEV (SLEE medical GmbH, Alemania), con un espesor de 10 µm, y se depositaron sobre portaobjetos tratados previamente con poli-L-lisina 0,01% (Sigma-Aldrich) según los métodos convencionales. Los portaobjetos conteniendo los cortes se guardaron a -20°C hasta su utilización.

### **3.6.2: Diseño y construcción de sondas de ARN con tecnología Stellaris®**

Las sondas de ARN tecnología *Stellaris® RNA-FISH* (Biosearch Technologies) consisten en hasta 48 oligonucleótidos de unos 20 bases, conjugados a fluoróforos. Las mismas han sido usadas exitosamente para la detección de lncRNAs, como en el trabajo de Orjalo y Johansson [2016]. Las sondas fueron diseñadas por nosotros utilizando el diseñador *Stellaris® Probe Designer* versión 4.2 disponible *online* (<https://www.biosearchtech.com/stellaris-designer>), y luego se encargó la síntesis de las mismas a la propia empresa. Las sondas contra los transcritos de *1700001L05Rik*, *1700008K24Rik*, *1110020A21Rik*, *Rbakdn* y *Kcnmb4os1* se conjugaron al fluoróforo Quasar670, mientras la sonda contra *Kcnmb4* y el control positivo lncRNA *Malat1* (único que se adquirió pre-diseñado por la empresa) se conjugaron al fluoróforo Quasar570 (Tabla 2). Una vez recibidas, las mismas fueron diluidas en tampón TE pH 8,0 (ver Anexo 1) a una concentración de 12,5 µM siguiendo las indicaciones del fabricante. Finalmente, fueron alicuotadas y almacenadas a -20°C protegidas de la luz.

### **3.6.3: Hibridación *in situ* de ARN (RNA-FISH) sobre criosecciones de testículo**

Para la hibridación de las sondas *Stellaris*® seguimos el protocolo disponible en la página web de la empresa, recomendado por el fabricante para tejido criopreservado. Brevemente, los portaobjetos conteniendo las criosecciones testiculares guardados a -20°C se dejaron tomar temperatura ambiente. Se fijaron en solución de formaldehído (FA) 3,7% por 10 minutos a temperatura ambiente, y luego se lavaron 2 veces con PBS 1X por 5 minutos a temperatura ambiente. Los cortes se permeabilizaron en etanol 70% por al menos 1 hora a temperatura ambiente. A continuación se realizó un lavado por 5 minutos a temperatura ambiente con solución comercial *Wash Buffer A* (SMF-WA1-60, Biosearch Technologies) conteniendo 10% formamida desionizada. Se colocaron las sondas diluidas 1:100 en solución de hibridación comercial (SMF-HB1-10, Biosearch Technologies) con 10% formamida desionizada sobre los portaobjetos, incubándose 16 horas a 37°C en cámara húmeda y oscura. Al día siguiente, los portaobjetos se lavaron con solución comercial *Wash Buffer A* por 30 minutos a 37°C en cámara húmeda y oscura, y luego con solución comercial *Wash Buffer A* con DAPI 1 µg/ml por 30 minutos a 37°C en cámara húmeda y oscura. Para finalizar, se realizó un lavado con solución comercial *Wash Buffer B* (SMF-WB1-20, Biosearch Technologies) por 5 minutos a temperatura ambiente y en oscuridad, y se montaron en medio de montaje *ProLong Diamond Antifade Mountant* (P36961, Invitrogen).

### **3.6.4: Detección de proteínas por inmunofluorescencia (IF) secuencial sobre criosecciones de testículo – protocolo *Stellaris*®**

El protocolo disponible en la página web de Biosearch Technologies para la IF secuencial está descrito para cultivos de células adherentes, pero nosotros lo adaptamos a las criosecciones testiculares. La preparación previa de los portaobjetos con las criosecciones guardados a -20°C es igual al protocolo de *RNA-FISH* (sección 3.6.3), es decir que se dejaron tomar temperatura ambiente. Luego, todo el procedimiento se realizó a temperatura ambiente. Se fijaron en solución de FA 3,7% por 10 minutos, se lavaron 2 veces con PBS 1X por 5 minutos, y se permeabilizaron en etanol 70% por al menos 1 hora. A continuación, las criosecciones se permeabilizaron con Tritón X-100 al 0,1% en PBS 1X por 5 minutos. Luego de un lavado con PBS 1X por 5 min, se incubaron con anticuerpo primario en su dilución

correspondiente en PBS 1X, por 1 hora en cámara húmeda. Seguidamente se realizaron 3 lavados con PBS 1X, 10 minutos cada uno, y se incubó con anticuerpo secundario en su dilución correspondiente en PBS 1X, por 1 hora en cámara húmeda y oscura. Después de repetir los 3 lavados con PBS 1X, los portaobjetos se fijaron nuevamente con FA 3,7% por 10 minutos y finalmente se realizaron 2 lavados con PBS 1X de 5 minutos cada uno. Siguiendo el protocolo de la empresa, a continuación y en forma secuencial a la IF se llevaron a cabo las hibridaciones con las sondas de ARN, según el protocolo descrito más arriba (ver 3.6.3).

Los anticuerpos primarios utilizados fueron anti-MVH de conejo diluido 1:2000 (ab13840, Abcam) y anti-SYCP1 de conejo dilución 1:200 (cedido por el Prof. Dr. Ricardo Benavente). El anticuerpo secundario empleado fue anti-conejo conjugado a Alexa Flúor 488 (A-11034, Invitrogen).

### 3.7: Microscopía y análisis de imágenes

#### **3.7.1: Microscopía confocal**

Todas las imágenes de microscopía láser confocal fueron realizadas en el microscopio Zeiss LSM 800 (Carl Zeiss Microscopy, Alemania), con objetivo de inmersión en aceite de 63X y apertura numérica de 1,4, equipado con láseres 405 nm, 488 nm, 561 nm y 640 nm para excitación. El equipo cuenta con un sistema de super-resolución *AiryScan*. Las fotografías fueron tomadas con una cámara color digital *Axiocam 506*, tamaño de 1024 x 1024 píxeles, de secciones individuales o consecutivas en el eje z de distancia variable (*stacks*). Las mismas fueron procesadas inicialmente con el software *Zen 2.3* (Zeiss).

#### **3.7.2: Microscopía de alta resolución**

Las imágenes de alta resolución fueron visualizadas mediante microscopía estructurada de iluminación (*Structured Illumination Microscopy*, SIM) en un microscopio Elyra S.1 (Carl Zeiss Microscopy, Alemania), equipado con láseres de excitación de 405 nm, 488 nm, 561 nm y 642 nm. Las imágenes se tomaron con un objetivo de inmersión en aceite de 63X/1.4 DIC M27 con cámaras sCMOS y *Edge 5.5* (PCO, Alemania), tamaño 1024 x 1024 píxeles, tres rotaciones con 5 cambios de fase.

Las mismas fueron procesadas inicialmente con el software *Zen 2012 SP1* (Black Edition, Zeiss).

### 3.7.3: Análisis de imágenes con *Fiji*

El análisis de las imágenes de microscopía se llevó a cabo con el programa *Fiji ImageJ* [Schindelin *et al.*, 2012]. Las reconstrucciones de los *stacks* se realizaron con la herramienta de proyección en Z con máxima intensidad. La co-localización entre las sondas sentido *Kcnmb4* y antisentido *Kcnmb4os1* fue calculada usando el complemento *Colocalization Finder* del mismo programa.

**Tabla 2.** Secuencias de las sondas de ARN *Stellaris®* diseñadas.

| Gen (n° oligonucleótidos) | Oligonucleótidos (5'-3')                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
|---------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1700001L05Rik (48)        | ACCTCCTGAGATTAAGTGT; GTAGAAGATGGCCAGTATGT;<br>TCTATGTGGTTGCTAAGCAC; CGAATACGACCGTTGACCAA;<br>CATGTCTTCTTGTGAGCTA; GGCACCTTTAGGATGCATTA;<br>GATTAATGGGTCATGGTGGG; CCCATTCCATGTGTTAATTA;<br>AAAGATGAGGTGTGCAAGCC; GACAGTCACATGTACCCAAA;<br>ACTGGTCTGGAATTGGAGTG; TCCTTTGGGCTGAAAGGAAG;<br>AGGATCAGATAGGAGGTGTC; TAAGTGTCTCATCCTCTTTC;<br>TGAAGAGCTTGTGCTTTGTC; CTGGAAGGAGGGTGCAGAA;<br>AAGTGTGGCTCACTTATGTC; CCAGAACAGTGAGACAGTGT;<br>AGGCTCTGAAAACCTTGCT; CAGAGCAGACACAGGGAGAA;<br>ACTTGGGACCCAAAACAGTG; ACACAGGCATGTGGTGAATG;<br>TAGGCTAGAAGGTTGCACAG; CCATTTAGACCTTTTCTCTA;<br>AGAGGCTGAGTAGACAGACT; ACCACATCCTAGGGATTTAA;<br>AAGTACTCAAGTCTGGGTGG; AGGAAAGCAATGGTAGGGGG;<br>CTTACTCCCAAGAACTCGT; AATTGGTGTTCCCACTAGAA;<br>TCACATGCATCCTAAAGACC; TGGCTACATGCTAATGGACA;<br>GGAAGAGCGGGTCTAATGTA; GAGCTTTCAGAAAGCCCAA;<br>CACATCAACACAGGAGGCAG; TCTGCTTCCAAATCTTGACA;<br>TCCTTAGCTTTTTGCTAAGT; ATGGACAGTTACTCGGCAAG;<br>GACAAGCCACTCAAGGAAGG; TAGCCCGGTAAAGCAACAAG;<br>TTGACTTGTCTCTGTTGGTA; GTAGTCTCTGTCAGTACTTG;<br>TGTGCCTATAAAGTCCACAG; CGAACTGCTGCTTCAGATTA;<br>TAGCTTTCTTCTTGTTAGCA; ACCAATCACAAATGGCAGCG;<br>GCCATCAAAGATGTCAGCAA; AGCAGCTAAGATTCAGGCAA |

|                    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1700008K24Rik (29) | <p> TGAGAAGCCTCCTGCAAAAT; CAACCAGCTGATCAGCTAGG;<br/> T TACTTCGTGAAGAGTCCTC; TTCTTCCATCTTCACAATGG;<br/> TTCAAGCTGGATTTTGTGGG; TAGGGTAAGACCCCACTGAC;<br/> AGTGGAGATCATATGCTTGG; TTCTGGTTAGGACAGAAGAA;<br/> CTCCGCAAATAGAAGGGTAA; TCCAGGCCGATAGGCAATAG;<br/> ATGAGTGCTGAGTTATCTGC; CGTTATAGCTTGACAGCTGA;<br/> TTTCAGGGACTGCCTCAACA; TAGTTCCTACAAAATGGCGC;<br/> TGATTCCCTTGAGGTTGGCTT; AACAAAGTCTTCTTGGGGCG;<br/> ACCAGGCTAGAAACACATCA; ATCTTAACAGCTCTTCCTTT;<br/> TAAGGCAGATGGCTGGAGAG; TGAGGCTAAGTTTCCTTTTC;<br/> GTCTGGACAAGGAGCCAAAG; AGGTAACCTCCGTGGTAAAG;<br/> GAGGCAAGTTTCGATGTTGA; AAGAGGGAGAATACACCCAG;<br/> GGAGAGCAGTTTGGAGGAAT; TGAATTCCATCGGCTTTCTT;<br/> CTTTATAAAGGCCTGGGTGG; CATACATCCTGTGGTACGAG;<br/> GCAAAAGTCTTTATTGGGTC </p>                                                                                                                                                                       |
| Kcnmb4os1 (26)     | <p> CGCAGGAGATGGAAGGGG; CAGAGCCCACATTTGGTG;<br/> AAGCCGAGACCAGAGGTT; GTCAAGGCGGAAGCCATG;<br/> ATCTGTGCCAAGAGCCTG; TTCTCATCGTGGTCCTGA;<br/> GTGGTGGCGTTTGTGGTG; GATTGCTCTCCTGCACTG;<br/> TACAGCGCACACACGACG; AGACCAGAGGACGTTCTC;<br/> CTGTCCTCCACTCTCTTG; ACTCACTGGTCTTCCTTT;<br/> GAATGTGCTGGGCTGAGA; CTTGTCTGCTGCAGGTTG;<br/> ATTATTCCTACGCCGGA; GTCATCACTTTCAAGGGT;<br/> GGACGGTGTCTGCAATCA; ATGGGTGTGGAGCAGTCT;<br/> AGAGCTCACGGCTTTTGG; ATGGCATTCAGTGCCATC;<br/> GCTGGGTTGTTCTTATT; CCTCTTAGGAGCTCTTTC;<br/> GTGTGGTGGTTCCAGACT; TCGCTTCCTTCTCTCCTG;<br/> TGGATCCAACAGTTTCCG; AAGTGTTGCCCAAGATCC </p>                                                                                                                                                                                                                                                                                                           |
| Kcnmb4 (39)        | <p> CGTAAGACACCCTGAGCT; TCTTCGGCTTCCGTGTAC;<br/> AACAAAGCCGAGCCGGATG; GGATGCCGGAGACGATGA;<br/> GCCGAAGATGAAGAGCGA; AAGGCGGGACTGAGCCAG;<br/> CGTGGCTTGACAGATCCTG; GACAGCACGGTGCAGTTG;<br/> GAACACCTCGCCGATCTG; CCACAGGTGAAGGTGCAC;<br/> AGGGATACTGCGAGGTGC; GTTCACGTACACCTGGAC;<br/> TGGAGTTGGACTCGGAGT; TGGTCGCTGTGTAGCAGC;<br/> GTTGGTCAGGAGCTGGTG; GGATATAGGAGCACTTGG;<br/> TTTTCTCTCTTACAGGGC; GCTCTCCGAGTTCTTCTG;<br/> ACTGCTGCCAGTTCATGA; CCGATCTCATCTTTCCAA;<br/> GCAAGTGAATGGCTGGGA; CTGGTCTCTGATGCTGAT;<br/> CGCTGTAGGAGAACGTCC; AGCAATCTCGTCGTGTGT;<br/> CCAGAGGAAGCAGTGCAG; CACCACAAACGCCACCAC;<br/> GTCAGGACCACGATGAGA; CAGGCTCTTGGCACAGAT;<br/> GCGTTTCTTCATGGCTTC; AACCTCTGGTCTCGGCTT;<br/> CAGGCGAGTGCCGACGAG; ACCAAATGTGGGCTCTGC;<br/> GTTTCTGCTGCCGACGAG; CAGTCTTCCAACGTGCC;<br/> TACAGGCAGTTAAGTCCC; GTTGCTACAGCACATGGA;<br/> AGAGCTTGACCCTTTAGC; TAGTCACTCCATAGAGCT;<br/> ACCGACTTCTTTGAGGGT </p> |

|                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|---------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>1110020A21Rik</i> (25) | ACTCTACCTATGGCTACTAG; GGCCTAATTTGACTGACGGA;<br>AAGGATATCGGTCTTCGAGT; CGCCATTTGAACCAATCAGA;<br>AAGCCAATAGTCGGAGGATC; TGGTTTCTTGACGGACAAGT;<br>CCGAAGTGACGGACTTGGA; CCTTTTAGGATACTTTTGCG;<br>CTTGCCATCTACGAATACT; AGAGGTCAGAACTGCAGGGT;<br>TGTTCTGTTCTCCGCGAAGTG; TGCCCGTTGGTTTCTTTATG;<br>GGACCAATTTCCACCTATAA; CTTCTCTGGCCAATCAGAAG;<br>TTCTTCCAATGAGGGTCTAG; CACAGCTCTATGCCAAAGAG;<br>TAAGGTGTCCCAAGCAAAGG; CCCGTAATTCAATCTTCTAA;<br>CCTTATTTTGAAGGTGAGG; CCCTTATTTGTTGATCTTTA;<br>TCGAGGCAAGGGAATCCAAG; ATCAAGCAGTGCGAGACAGG;<br>CATATTGCATTTGAGTGTCT; GCCTAGTTACATCACTGAAT;<br>GTGAGTTGTATACAATCCGC |
| <i>Rbakdn</i> (25)        | CAGGTCAGTTGTGGCTCT; TCCCCAAGAACTGTCAGG;<br>CTGGTCAACGGTCTCCAG; CTCTGTGTCAGCCCTCAG;<br>TGCCTTCTCACTTCCGAG; TCTCTGGTACTGGGAAGC;<br>TGCACGCCACGTCCATTC; CAGGTGGCTTTTATTCCC;<br>GTCAGAGGACTTCCCTAC; GCAAAAGGATGGGCTGTT;<br>CATGACCCTGGAACAGGG; AGTTTTCTGGAGGTCCG;<br>TCTTCGAGTACGCTTGGA; GTCTCCTTCAGTTTGGAT;<br>CTTCTCTCCACATGCCTC; GAGCCAGCGAACGGGTTG;<br>ATGAGCCAGGAGGACAGC; GACCACATAGAAGCCACC;<br>TCATGAGGCTACAGGTCC; TGGGAGGAGTGTGGAGCT;<br>CGGAGGAGACCAGGGATT; CTTGAGCAGCTGAAGGGC;<br>GGATAATGGTGTCCACCC; GAGTTCCACATATGTCCC;<br>AATGGTGCTGCAGGTGGG                                                  |

### 3.8: Cultivos de sobrevida de células germinales

#### **3.8.1: Preparación de suspensión celular mediante disgregación enzimática del tejido**

El protocolo de preparación de las suspensiones celulares se basó en el descrito por Gaysinskaya y Bortvin [2015] con modificaciones. Luego de la eutanasia de ratones de 21-25 dpp y su disección testicular, los testículos fueron colectados en una placa de Petri estéril conteniendo PBS 1X tratado con DMPC 0,1%. Se removió la túnica albugínea e inmediatamente se los transfirió a un tubo estéril conteniendo 6 ml de solución DMEM/colagenasa/DNAsa fría (ver Anexo 1), y se los incubó en posición horizontal durante 5 minutos a 35°C con agitación a 60 rpm. Posteriormente se los pipeteó gentilmente 5 veces con pipeta Pasteur estéril, y se continuó incubando en iguales condiciones por 5 minutos más. Luego se colocó el tubo verticalmente por 2 minutos a temperatura ambiente, para permitir la sedimentación de túbulos conteniendo las células germinales. El sobrenadante enriquecido en células intersticiales, se retiró con pipeta Pasteur. Se adicionaron 6 ml de solución

DMEM/colagenasa/DNAsa/tripsina (ver Anexo 1) precalentada a 35°C, se pipeteó gentilmente 10 veces con pipeta Pasteur, y se incubó por 12 minutos a 35°C con agitación a 60 rpm. Posteriormente, se agregaron 60 µl de tripsina 2,5% al tubo, se pipeteó nuevamente 10 veces con pipeta Pasteur, y se incubó por otros 12 minutos a 35°C con agitación a 60 rpm. Se pipeteó suavemente la suspensión final 10 veces más con pipeta Pasteur. Luego se agregaron 600 µl de suero fetal bovino para inactivar la tripsina, y se mezcló por pipeteo unas 5 veces más. La suspensión se filtró por una malla de 80 µm de poro, y se realizó un conteo celular en cámara de Neubauer para ajustar la concentración a  $1 \times 10^6$  células/ml. Fue posteriormente teñida con el colorante vital de ADN VDG, e incubada durante 1 hora a 35°C con leve agitación a 60 rpm, protegida de la luz. Luego de ese tiempo, se volvió a filtrar por una malla de 50 µm, e incubó 15 minutos con IP a 2 µg/ml a temperatura ambiente.

### **3.8.2: Clasificación de las fracciones 4C y RS por citometría de flujo**

Las fracciones celulares fueron clasificadas en un citómetro de flujo y clasificador celular *MoFlo Astrios EQ* (Beckman Coulter, USA), en modo “*Purify*” que permite la máxima pureza. La deposición de las células clasificadas se realizó tanto en tubos estériles de 1,5 ml como en placa con fondo de vidrio (D35-20-1.5-N, Cellvis, Mountain View, CA) pre-tratadas con poli-L-lisina 0,01% por 15 minutos a temperatura ambiente, y secadas toda la noche, ambos soportes conteniendo medio de cultivo de sobrevida [La Salle *et al.*, 2009] (ver Anexo 1) precalentado a 35°C para la colecta y conteniendo IP a 2 µg/ml. Se colectaron 50.000 células de la población 4C (mezcla de todos los espermatoцитos I de la meiosis) o RS (espermátidas redondas) por cada tubo o placa. Se analizó la viabilidad a tiempo 0 (a la salida del citómetro) mediante visualización de las células colectadas en placa en el microscopio confocal Zeiss LSM 800 (ver 3.7.1).

### **3.8.3: Análisis de la viabilidad de los cultivos de sobrevida a las 48 horas**

Las fracciones de 4C o RS clasificadas, se incubaron a 32°C con atmósfera de CO<sub>2</sub> al 5% [La Salle *et al.*, 2009] en medio para cultivo de sobrevida (ver Anexo 1), conteniendo estreptomicina 200 U/µl y penicilina 50 mg/ml. En el caso de las células colectadas en placa, se incubaron en la misma placa con 4 ml finales de medio. En el caso de las células colectadas en tubos de 1,5 ml, las mismas fueron transferidas a tubos estériles de 15 ml conteniendo 4 ml finales de medio. Se realizaron 3 ensayos

independientes, y en cada uno se colectaron réplicas de placas o tubos para medir su viabilidad a las 24 y 48 horas de incubación. La viabilidad se estimó mediante conteo de células verdes (viables, teñidas con VDG) o rojas (muertas, teñidas con IP) en el microscopio confocal. Para los conteos de las células cultivadas en tubos, el contenido de los mismos fue transferido a placas previo a la observación al microscopio. Se contaron 10 campos por cada tiempo y fracción celular, en cada uno de los ensayos.

## **4. RESULTADOS**

Los resultados de este trabajo se presentarán a continuación divididos en cinco capítulos:

- un primer capítulo que abarca una revisión sobre los lncRNAs, sus funciones generales y su relación con diversas patologías, especialmente testiculares;
- un segundo capítulo sobre la optimización del método de purificación de las fracciones celulares y la construcción de las librerías de secuenciación hebra-específicas;
- un tercer capítulo que incluye todos los análisis bioinformáticos de los datos de la secuenciación, la caracterización del transcriptoma global de la espermatogénesis del ratón y confirmaciones por *RT-qPCR*;
- un cuarto capítulo donde se muestran las localizaciones celulares y sub-celulares de un grupo de lncRNAs de interés seleccionados; y
- un quinto capítulo que describe algunos resultados preliminares de la puesta a punto de un cultivo celular de sobrevida de células germinales a partir de células purificadas por citometría de flujo, para la realización de estudios funcionales de lncRNAs de espermatogénesis.

En cada capítulo se realizará una breve discusión específica de los resultados obtenidos, y luego se discutirá globalmente el trabajo en la siguiente sección (ver 5. Discusión).

Hacemos notar que la revisión del capítulo 1 fue publicada en la revista nacional arbitrada *AnFaMed* ("Anales de Facultad de Medicina") de UdelaR, y está anexada al final del capítulo correspondiente [Trovero & Geisinger, 2019]. Todos los resultados del tercer capítulo, y algunos del cuarto, fueron publicados en un artículo en la revista internacional arbitrada *RNA Biology*, igualmente anexado al finalizar el capítulo 3 [Trovero *et al.*, 2020].

# **LncRNAs y su vinculación a patologías testiculares**

## **4.1: CAPÍTULO 1**

Tal como fue mencionado en la introducción de esta tesis, en los últimos años ha emergido gran interés por los lncRNAs ya que se han encontrado evidencias de su participación en múltiples procesos biológicos, así como también con diversas patologías, incluyendo algunas testiculares. Existen varios trabajos donde se estudian individualmente diferentes lncRNAs relacionados a anomalías testiculares como cáncer o infertilidad, pero no encontramos una revisión donde se haga un estudio exhaustivo de la temática. Es así que realizamos una búsqueda de trabajos que trataran de lncRNAs, sus funciones y relación a patologías relacionadas a la reproducción masculina, y lo compilamos en una revisión publicada en la revista “Anales de Facultad de Medicina” en junio 2019 [Trovero & Geisinger, 2019], la cual se adjunta a continuación.

# Los ARNs no codificantes largos y su vinculación con las patologías testiculares

## Long Non-Coding RNAs and Their Involvement in Testicular Pathologies

## RNAs longos não-codificantes e sua ligação com patologias testiculares

María F. Trovero<sup>1</sup> y Adriana Geisinger<sup>2</sup>

### Resumen:

Si bien la porción del genoma destinada a la síntesis de proteínas es muy pequeña, actualmente se sabe que casi todo el genoma se expresa bajo forma de ARNs no codificantes. Entre dichos ARNs se encuentran los ARNs no codificantes largos (lncRNAs). Aunque los lncRNAs han sido muy poco estudiados, recientemente han comenzado a centrar la atención de los investigadores, al descubrirse que los mismos pueden desempeñar diversas funciones en la regulación de la expresión génica. Además, su vinculación con patologías ha comenzado a ser puesta de manifiesto. Curiosamente, la cantidad de lncRNAs presentes en el testículo es abrumadoramente mayor que en cualquier otro órgano o tejido estudiado. Los perfiles de expresión de estos lncRNAs varían significativamente a lo largo de la espermatogénesis, y algunas evidencias sugieren que al menos algunos de ellos podrían participar en el proceso de formación de células germinales masculinas. No obstante, el conocimiento sobre el tema es aún muy escaso. En este trabajo revisamos la información disponible sobre la expresión de lncRNAs en el testículo y sus posibles funciones. Asimismo, analizamos algunos ejemplos que ilustran la participación de lncRNAs en el desarrollo de patologías como la infertilidad y el cáncer testicular.

**Palabras clave:** ARNs no codificantes largos, lncRNAs, espermatogénesis, testículo, patología testicular.

### Abstract:

Although the portion of the genome devoted to protein synthesis is very small, it is now known that almost the entire genome is expressed as non-coding RNAs. Among them, there are long noncoding RNAs (lncRNAs). Despite that lncRNAs have been very poorly studied, they have recently started to focus the attention of researchers, as it has been found out that lncRNAs can perform diverse functions in the regulation of gene expression. Moreover, their involvement in pathologies is being revealed. Intriguingly, the amount of lncRNAs in the testis is overwhelmingly higher than in any other analyzed organ or tissue. LncRNA expression profiles

<sup>1</sup>Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Montevideo, Uruguay. ORCID: 0000-0001-9257-6979

<sup>2</sup>Sección Bioquímica/Biología Molecular, Facultad de Ciencias, Universidad de la República. Montevideo, Uruguay. Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE). Montevideo, Uruguay. ORCID: 0000-0002-3995-9185

Contacto: adriana.geisinger@gmail.com

significantly vary along spermatogenesis, and some evidence suggests that at least some of them could participate in the formation of male germ cells. However, knowledge on the subject is still very scarce. In this work we review the available information on the expression of lncRNAs in testis and their possible roles. We also analyze some examples that illustrate the participation of lncRNAs in the development of pathologies such as infertility and testicular cancer.

**Keywords:** Long Non-Coding Rnas, Lncrnas, Spermatogenesis, Testis, Testicular Pathology.

### Resumo:

Embora a porção do genoma usada para a síntese proteica seja muito pequena, sabe-se agora que quase todo o genoma é expresso na forma de RNAs não-codificantes. Entre esses RNAs estão os longos RNAs não codificantes (lncRNAs). Embora os lncRNAs tenham sido pouco estudados, eles recentemente começaram a focar a atenção dos pesquisadores, ao descobrirem que podem desempenhar diversas funções na regulação da expressão gênica. Além disso, sua ligação com as patologias começou a ser revelada. Curiosamente, a quantidade de lncRNAs presentes nos testículos é esmagadoramente maior do que em qualquer outro órgão ou tecido estudado. Os perfis de expressão destes lncRNAs variam significativamente ao longo da espermatogênese, e algumas evidências sugerem que pelo menos alguns deles poderiam participar no processo de formação de células germinativas masculinas. No entanto, o conhecimento sobre o assunto ainda é muito escasso. Neste trabalho, revisamos as informações disponíveis sobre a expressão de lncRNAs no testículo e suas possíveis funções. Também analisamos alguns exemplos que ilustram a participação dos lncRNAs no desenvolvimento de patologias como infertilidade e câncer testicular.

**Palavras-chave:** RNAs longos não codificantes, lncRNAs, espermatogênese, testículos, patologia testicular.

### *El descubrimiento de los ARNs no codificantes largos y sus características*

Durante varias décadas el dogma central de la biología molecular afirmaba que el ADN era leído a ARN (proceso conocido como transcripción) y a continuación dicha información era “traducida” para la síntesis de proteínas como producto final, para cumplir con los diversos procesos y funciones celulares. Estos ARNs que llevan el mensaje para la síntesis de proteínas son los ARNs mensajeros (ARNm). Se conocían además algunos ARNs que si bien se producían, no eran traducidos luego a proteína. Tal es el caso de los ARNs ribosomales y ARNs de transferencia, que constituyen partes esenciales en la maquinaria de la síntesis proteica.

Con el advenimiento de las nuevas tecnologías de secuenciación a gran escala (secuenciación masiva, o NGS por su sigla en inglés) se ha revelado que, sorprendentemente, la proporción de genoma transcripto a ARN pero que no produce proteínas (ARN no codificante, ya que no contiene el código que especifica su traducción a proteína) es mucho mayor de lo que previamente se estimaba. En humanos, por ejemplo, el ADN no codificante para proteínas alcanza el 98% del genoma, si bien la mayoría es activamente transcripto, es decir, se producen ARNs a partir del mismo<sup>(1)</sup>. Esto hace suponer que, dado el gasto energético que ello conlleva para las células, no todo este ARN no codificante sería simplemente “basura” o “ruido transcripcional”, sino que al menos parte de él

debería desempeñar alguna función en el organismo.

Estos transcriptos no codificantes son usualmente clasificados de acuerdo a su longitud, en ARNs no codificantes cortos y largos. Ejemplos del primer grupo incluyen ARNs ribosomales (rRNAs), micro ARNs (miRNAs), ARNs de interferencia o silenciamiento (siRNAs), y ARNs de interacción con proteínas PIWI (piRNAs)<sup>(2)</sup>. Estos pequeños ARNs, muchos de los cuales cumplen funciones regulatorias y están involucrados en procesos de silenciamiento o apagado de genes<sup>(3)</sup>, han sido ampliamente estudiados.

Por su parte, los ARNs no codificantes largos (lncRNAs por su nombre en inglés, *long noncoding RNAs*) son operacionalmente definidos por oposición a los cortos, como aquellos transcriptos de ARN que presentan al menos 200 nucleótidos de longitud<sup>(4)</sup>. Como regla general se asume que estos ARNs no producen proteínas, aunque se ha visto que algunos de ellos podrían dar lugar a pequeños péptidos funcionales<sup>(5)</sup>. A medida que se observaron cada vez más ARNs no codificantes en los organismos superiores, y que aumentó el número de lncRNAs identificados (alcanzando más del 65% en todo el genoma humano), los lncRNAs han comenzado a recibir mayor atención<sup>(6)</sup>.

Los lncRNAs comparten una serie de características comunes dentro de los vertebrados: presentan pocos exones (es decir, los segmentos que constituyen los ARNs maduros), tienen bajo contenido en las bases nucleotídicas guanina y citosina, están mucho menos conservados en las distintas especies de organismos que los ARNs codificantes, y son mucho menos abundantes que éstos<sup>(7)(8)(9)</sup>. Asimismo, se ha evidenciado que la mayor parte de ellos son generados por la ARN polimerasa II, la misma enzima que produce los ARNs mensajeros que luego darán lugar a las proteínas, y por lo tanto la mayoría posee una cola poliadenilada en su extremo 3' y un casquete o caperuza metilada en su extremo 5', que suelen ser características típicas de los ARNs mensajeros<sup>(10)</sup>. Por otra parte, también pueden generarse lncRNAs

por el procesamiento de otros transcriptos<sup>(2)</sup>. Una característica sumamente interesante de los lncRNAs es su patrón de expresión, altamente restringido en tiempo y espacio; es decir, que distintos lncRNAs aparecen en tipos celulares determinados o momentos determinados del desarrollo, con una especificidad mucho mayor de tejido/tipo celular/etapa del desarrollo, que los ARNs codificantes<sup>(2)(11)</sup>. A pesar de que en promedio son detectados con no más de una copia por célula, suelen ser abundantes en algunos tipos celulares específicos o incluso en compartimentos sub-celulares particulares, lo que les confiere aún mayor complejidad<sup>(12)</sup>.

### *Funciones y mecanismos de acción de los ARNs no codificantes largos*

Si bien el estudio de los lncRNAs ha sido históricamente relegado, en los últimos tiempos se ha comenzado a prestar particular atención a este subgrupo de ARNs, vinculándolos con variados procesos biológicos como diferenciación, proliferación y migración celular, apoptosis, respuesta inmune, desarrollo tisular, etc.<sup>(4)</sup>. Investigaciones recientes han mostrado que al menos algunos lncRNAs están asociados a procesos moleculares como regulación de la expresión de genes (a nivel de la transcripción, post-transcripcional, control de actividad de proteínas), modificaciones en la cromatina, organización de dominios nucleares, corte y empalme de ARNs mensajeros (*splicing*, en inglés), entre otros. Más aún, la alteración en la expresión de algunos lncRNAs se ha relacionado con la generación de tumores; incluso podrían actuar como oncogenes<sup>(13)(14)(15)</sup>.

Aunque los roles biológicos y mecanismos de acción de estos ARNs en su mayoría aún se desconocen, existen algunos modelos que intentan explicar su funcionamiento. De modo resumido, como se muestra en la Figura 1, se propone que podrían actuar como: *competidores*,

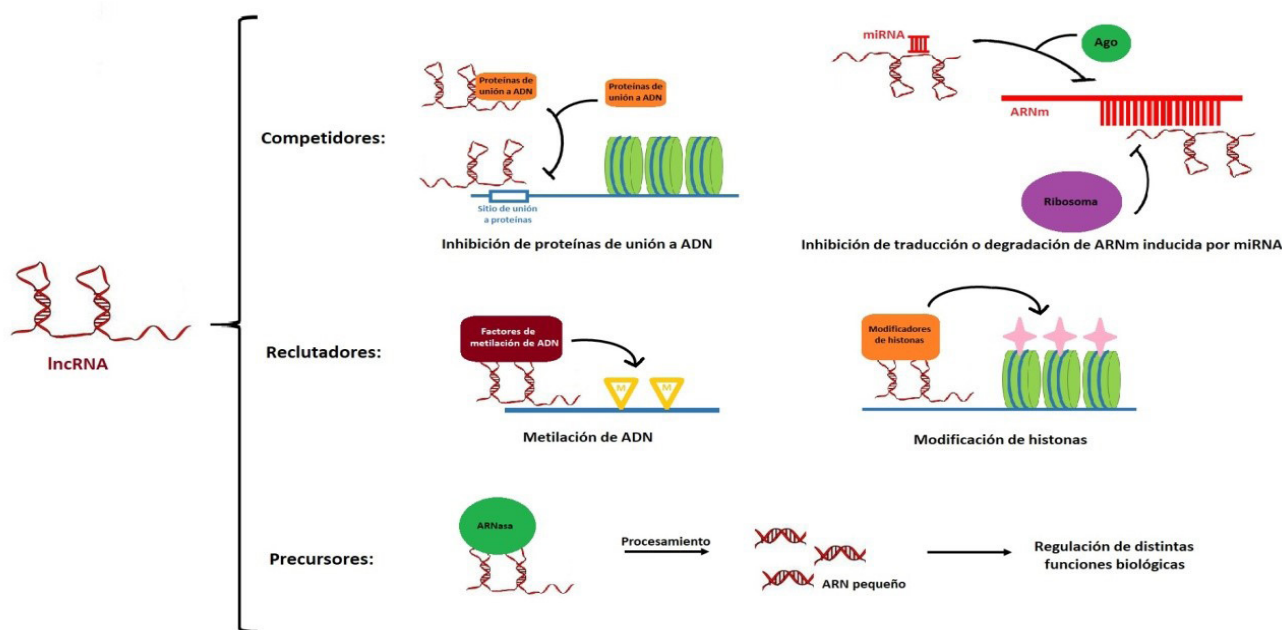


Figura 1. Posibles modos de acción de lncRNAs: competidores, reclutadores y precursores

compitiendo con otras moléculas por el sitio de unión a ADN o proteínas; *reclutadores* o *activadores*, activando modificadores del estado de la cromatina (modificaciones epigenéticas); y *precursores*, siendo procesados y dando lugar a pequeños ARNs, entre otros mecanismos<sup>(2)</sup>.

Uno de los lncRNAs cuyo rol ha sido más estudiado es el *ARN Xist* (en inglés, *X-inactive specific transcript*). La existencia de dos cromosomas X en las hembras de los mamíferos y sólo uno en los machos, acarrearía el problema de la síntesis del doble de cantidad de ARN y proteína a partir de los genes localizados sobre este cromosoma, en hembras que en machos. La evolución ha solucionado este inconveniente manteniendo activo sólo uno de los cromosomas X de las hembras, en tanto el otro está silenciado. Este mecanismo, que permite balancear la diferencia de expresión derivada del cromosoma X entre ambos sexos, es conocido como “compensación de dosis”<sup>(16)</sup>. El lncRNA *Xist* se transcribe en la vida embrionaria únicamente desde el futuro cromosoma X inactivo, y participa en el silenciamiento de dicho cromosoma en las hembras. El mecanismo de acción de *Xist* involucra su unión al cromosoma X formando una “nube” a su alrededor que cubre físicamente al

cromosoma con su propio transcripto, y reclutando otras proteínas y remodeladores de cromatina para lograr el silenciamiento<sup>(17)</sup>. El gen del lncRNA *Xist* presenta a su vez otro lncRNA antisentido, en la hebra de ADN opuesta, el *Tsix*, cuya expresión señala al cromosoma X activo. *Tsix* sirve de regulador de *Xist*, y es un claro ejemplo de represor antisentido de un gen en la hebra opuesta, que inhibe la expresión de *Xist* por varios mecanismos<sup>(18)</sup>.

Otro modo de acción bastante estudiado es el del lncRNA *HOTAIR*. Se trata de un ARN no codificante antisentido de los genes *Hox*, un grupo de genes involucrados en el desarrollo embrionario. El ARN *HOTAIR* se origina a partir del cluster *HoxC* (un cluster es un grupo de genes integrantes de una misma familia génica, que a menudo se localizan en una misma región del ADN), y mediante el reclutamiento de ciertos complejos proteicos que estimulan la compactación de la cromatina, silencia la expresión de la región alrededor de los genes *HoxD*<sup>(19)</sup>.

Como ejemplo de precursores, podemos citar al lncRNA *H19*, que desempeña un rol en la regulación negativa del peso corporal y la proliferación celular. Este ARN constituyó el primer transcripto con impronta identificado. El *imprinting* o impronta

genética es un fenómeno por el cual ciertos genes son expresados o no, según el sexo del progenitor. Ello conduce a la transcripción de una sola copia de un gen autosómico dependiendo del origen parental, en tanto el alelo residente en el cromosoma homólogo está silenciado. *H19*, que es de expresión materna y está “imprimado” (es decir, silenciado) en el cromosoma paterno, comparte un cluster con el gen *Igf2* (factor de crecimiento insulínico tipo 2), que por el contrario, es de expresión paterna y está silenciado en el cromosoma materno. Recientemente se determinó que *H19* es precursor de un micro ARN, el miR-675, tanto en humanos como en roedores. Esto parecería indicar que el lncRNA *H19* regula la expresión génica mediante un mecanismo basado en la acción de micro ARNs<sup>(19)(20)</sup>.

### *ARNs no codificantes largos y patologías*

Se ha sugerido que los lncRNAs estarían asociados a numerosas enfermedades, incluyendo cáncer de diferentes tipos<sup>(2)(21)</sup>. En ese sentido, si bien históricamente las causas de la ocurrencia y avance del cáncer se han relacionado principalmente con proteínas, frecuentemente se detecta desregulación de ciertos ARNs no codificantes en pacientes con cáncer<sup>(22)</sup>.

Con respecto a los lncRNAs, la expresión aberrante de *H19* se ha vinculado desde hace un tiempo a algunos síndromes como los de Beckwith-Wiedemann y Silver-Russell<sup>(23)</sup>. Evidencia creciente indica, además, que *H19* sería uno de los principales genes involucrados en el desarrollo de numerosos tipos de cáncer, al contribuir a la inestabilidad genómica, aceleración de la proliferación celular y aumento de la resistencia celular al estrés<sup>(24)</sup>.

*HOTAIR* también ha sido vinculado con diversos tipos de cáncer, incluyendo renal, hepático, de mama y esofágico<sup>(25)</sup>. Por otra parte, alteraciones en la expresión de *H19* y *HOTAIR* se han asociado con enfermedades

cardiovasculares y artritis reumatoidea<sup>(26)</sup>.

Algunos trabajos también han relacionado a *Xist* con tumorigénesis ya que se ha observado la desregulación de su expresión en cáncer de mama y ovario, aunque no puede descartarse que ello sea una consecuencia derivada de la inestabilidad epigenética global en estas patologías<sup>(18)(27)</sup>.

Entre los lncRNAs que pueden actuar como competidores, se ha visto que algunos pueden regular los niveles de ARNs mensajeros, compitiendo con micro ARNs. Un ejemplo lo constituye el lncRNA *Mir13HG*, que está regulado positivamente en un tipo de cáncer de pulmón (NSCLC, por sus siglas en inglés *Non Small Cell Lung Cancer*)<sup>(28)</sup>. *Mir13HG* es capaz de unirse a miR-214, un micro ARN que ha sido detectado en varios tipos de cáncer, y cuya sobreexpresión inhibe al NSCLC. El lncRNA, al unirse a miR-214 lo secuestra, promoviendo la migración e invasión de células de NSCLC. Dandan y colaboradores<sup>(28)</sup> observaron que silenciando este lncRNA en NSCLC se inhibía la migración, invasión y metástasis de células en dicho cáncer. Por este motivo, lo proponen como un marcador de NSCLC y posible blanco de terapia génica.

Se ha detectado también desregulación de lncRNAs en relación con la fertilidad femenina. En ese sentido, es sabido que una dieta alta en grasas afecta al sistema reproductor. Se ha reportado en ratones que la obesidad afecta la ovulación, el desarrollo ovárico y del ovocito. Además, puede impactar en el embarazo durante la implantación y el desarrollo del embrión, llevando a abortos y preeclampsia, entre otros<sup>(29)</sup>. Un trabajo reciente detectó desregulación de lncRNAs en ovarios de hembras de ratón obesas por dieta alta en grasas con respecto al control normal, que incluyó sobreexpresión de numerosos lncRNAs y subexpresión de muchos otros<sup>(30)</sup>.

También se ha observado desregulación de numerosos lncRNAs en el síndrome de ovario poliquístico, una de las causas más frecuentes de infertilidad no ovulatoria en hembras, con

una prevalencia del 5-10% en el mundo<sup>(31)</sup>. Particularmente, el lncRNA *HCG26* está sobreexpresado en células de la granulosa de ovarios poliquísticos<sup>(31)</sup>. El silenciamiento de este lncRNA conlleva una disminución en la viabilidad celular e inhibición de la progresión del ciclo celular, sugiriendo que la desregulación de este lncRNA contribuiría a la patogénesis del síndrome de ovario poliquístico, aunque los mecanismos moleculares permanecen aún sin aclarar<sup>(31)</sup>.

Finalmente mencionaremos la preeclampsia, una de las complicaciones más comunes del embarazo, que causa mortalidad fetal o prenatal. Se ha sugerido el rol de factores genéticos y epigenéticos en la patogénesis de la preeclampsia, por lo que diversos estudios se han centrado en el análisis de los lncRNAs. Se han observado patrones aberrantes en su expresión (incluyendo *H19*, *HOTAIR*, *Meg3*, *PVT1*, *Malat1*, entre varios otros), aunque los mecanismos subyacentes aún necesitan ser dilucidados<sup>(32)</sup>.

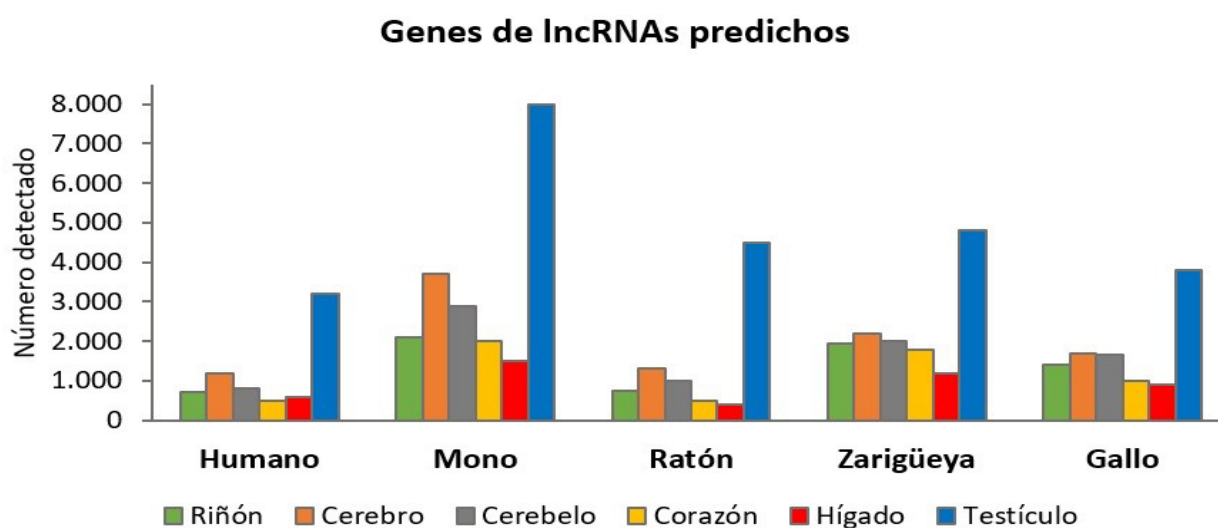
### ARNs no codificantes largos en el testículo

El testículo se caracteriza por un elevado nivel de diversidad y complejidad transcriptómicas. Exhibe el mayor número de ARNs mensajeros específicos de tejido entre todos los tejidos y órganos estudiados,

y una gran complejidad en sus mecanismos de regulación<sup>(33)(34)</sup>. Sin embargo, se ha visto que esta diferencia de amplitud en la generación de ARN a partir de ADN en el testículo es mucho más manifiesta aún para regiones intergénicas del ADN (como pseudogenes o elementos transponibles), ARNs no codificantes cortos y lncRNAs, que para los genes que dan lugar a proteínas<sup>(35)</sup>. Es decir, que el testículo es el órgano/tejido que presenta los mayores niveles generales de transcripción.

En particular, nos centraremos en los lncRNAs y el hecho de que su presencia es abrumadoramente mayor en el testículo que en otros órganos o tejidos. Es importante además, señalar que este fenómeno se ha observado para los testículos de todas las especies de vertebrados en las que se ha estudiado (Figura 2).

Dado que los lncRNAs en general presentan mayor especificidad de tejido y tipo celular que los ARNs mensajeros, se ha propuesto que su gran abundancia y diversidad en el testículo podría estar sugiriendo su implicancia en el desarrollo de células germinales masculinas<sup>(14)</sup>. En el mismo sentido, un estudio realizado en el año 2014 donde se analizó la evolución y conservación de lncRNAs de 11 especies diferentes y también su co-expresión con genes de proteínas, propone que los lncRNAs tendrían funciones asociadas a procesos esenciales, como por



**Figura 2.** El testículo es el órgano que presenta mayor número de lncRNAs predichos. Se muestran los lncRNAs detectados en seis órganos de cinco especies diferentes: humano, macaco Rhesus (mono), ratón, zarigüeya y gallo. Información extraída de Soumillon *et al.*, 2013<sup>(35)</sup>.

ejemplo la espermatogénesis<sup>(36)</sup>. Un dato derivado de otro estudio de conservación de lncRNAs testículo-específicos en el ratón, sugiere que la mayoría de los lncRNAs tendrían una baja conservación de secuencia primaria entre especies, y una rápida evolución<sup>(14)</sup>.

Diversos estudios han identificado y caracterizado parcialmente un elevado número de lncRNAs en el testículo del ratón<sup>(14)(35)(37)(38)(39)(40)</sup>, humano<sup>(41)(42)</sup>, u otras especies<sup>(43)(44)</sup>. En general estos estudios sugieren que al menos parte de dichos lncRNAs podrían tener un rol regulador o de mantenimiento, afectando al proceso de la espermatogénesis.

De entre los distintos tipos celulares que componen el testículo, la expresión de lncRNAs es más pronunciada en las espermatidas redondas (células posmeióticas), y en los espermátocitos I (células cursando la primera profase meiótica) que en otros tipos de células testiculares<sup>(35)</sup>. Nuestro grupo ha observado que la gran mayoría de los lncRNAs testiculares se encuentran en las espermatidas redondas, y los mayores cambios en los patrones de expresión de lncRNAs se detectan en estas células (Trovero *et al.*, manuscrito en revisión). Durante la espermiogénesis (proceso por el cual las espermatidas redondas pasan por una serie de etapas hasta transformarse en espermatozoides), las espermatidas están sujetas a dramáticos cambios en el remodelado de su cromatina, los cuales a su vez requieren de la utilización de diversos mecanismos peculiares de regulación de la expresión génica<sup>(33)</sup><sup>(34)</sup>. Si bien es probable que muchos de los lncRNAs presentes en las espermatidas representen simplemente “ruido” transcripcional, resulta muy atractiva la hipótesis de que su elevada presencia pudiera estar vinculada, al menos en parte, a la regulación de esos dramáticos cambios cromatínicos y mecanismos de expresión génica, particulares de estas células (Trovero *et al.*, manuscrito en revisión).

Como hemos mencionado más arriba, los lncRNAs revisten un gran interés porque se ha comprobado que contribuyen al silenciamiento génico, inactivación

del cromosoma X, establecimiento de la impronta genética, y desarrollo. Sin embargo, muy poco se sabe hasta el momento sobre su biogénesis y función en general, y mucho menos aún de su rol en las células germinales masculinas<sup>(37)</sup>. A pesar de ello, existen algunos casos de lncRNAs ya estudiados en detalle, y para los cuales se ha descubierto una función en el testículo y/o la espermatogénesis.

*Mrhl* (en inglés, *Meiotic recombination hot spot locus*), es un lncRNA nuclear expresado en el testículo, que regula la vía de señalización Wnt, vía que desempeña un rol esencial en la espermatogénesis de los mamíferos<sup>(45)</sup>. *HongrES2* regula la maduración espermática, y tendría un rol en la capacitación del espermatozoide<sup>(46)</sup>. *Tsx* (en inglés, *Testis specific X-linked*) es un lncRNA expresado robustamente en células germinales (específicamente en la primera profase meiótica, en estadio de paquiteno), células madre embrionarias y cerebro, que escapa a la inactivación del cromosoma X, y cuya delección en ratones aumenta la apoptosis celular durante el paquiteno, derivando en machos con menor tamaño testicular<sup>(47)</sup>.

Por otra parte, se han detectado más de 240 lncRNAs específicos de células madre espermatogénicas (SSC), lo que podría indicar su participación en el mantenimiento y auto-renovación de estas células a través de la regulación de proteínas o micro ARNs<sup>(37)</sup>. El balance entre la generación y la diferenciación de SSC es indispensable para la espermatogénesis y la fertilidad masculina<sup>(48)</sup>. Entre los factores participantes en este proceso, GDNF (en inglés, *Glial-Derived Neutrophic Factor*) es un factor de crecimiento de las SSC necesario para su proliferación y mantenimiento. En las espermatogonias tempranas de ratón, incluyendo las SSC, se ha identificado un lncRNA específico (*LncRNA033862*) que resultó esencial para regular el destino de las SSC. Este lncRNA regula la expresión del gen *Gfra1*, un receptor del factor GDNF, localizándose dentro de

*Gfra1* y actuando como antisentido a dicho gen.

### *Vinculación con patologías testiculares*

La infertilidad es un problema de salud mundial que afecta aproximadamente a una de cada 6 ó 7 parejas<sup>(49)</sup>. La mitad de estos casos se debe a factores masculinos, aunque el 75% de los pacientes son diagnosticados con causas idiopáticas por desconocerse los mecanismos moleculares que los afectan<sup>(13)</sup>. Trabajos recientes han comenzado a revelar un vínculo entre la alteración en la expresión de los patrones de lncRNAs en el testículo, y la infertilidad masculina.

Una de las principales causas de infertilidad en machos es la falta de motilidad de los espermatozoides (astenozoospermia). Un análisis del semen de toros con alta y baja motilidad espermática, dio como resultado una expresión diferencial de lncRNAs: de alrededor de 11.500 lncRNAs identificados en los espermatozoides, unos 2.570 estaban expresados diferencialmente entre ambos grupos, incluyendo 57 lncRNAs sobreexpresados y 2.460 reprimidos en el grupo con baja motilidad<sup>(50)</sup>. En general, se considera que muchos de los ARNs presentes en los espermatozoides son una “huella” de eventos ocurridos durante la espermatogénesis, y podrían permitir evaluar el estado de fertilidad del individuo<sup>(51)</sup>.

En un estudio reciente sobre el patrón de expresión de lncRNAs en humanos, se identificaron más de 29.000 lncRNAs presentes en espermatozoides y testículo. De entre los ARNs hallados, se identificaron unos 6.300 genes de lncRNAs sobreexpresados y alrededor de 3.400 lncRNAs reprimidos específicamente en muestras de espermatozoides con astenozoospermia, en comparación con muestras normales. Cuando se analizaron los potenciales genes blanco (codificantes para proteínas) de estos lncRNAs, se comprobó un enriquecimiento en genes vinculados a biosíntesis de macromoléculas, espermatogénesis, generación de gametos, metabolismo de ARN, y a vías como la de digestión

de proteínas, degradación de ARN, adhesión focal, metabolismo de ácidos grasos y señalización<sup>(51)</sup>. Esto sugiere entonces que el metabolismo de proteínas, ácidos grasos y ARN, así como de otras macromoléculas, podrían cumplir roles esenciales en los procesos de espermatogénesis, y su alteración contribuir a la ocurrencia de astenozoospermia.

Por otra parte, un factor de riesgo para la espermatogénesis y la fertilidad masculina es, al igual que para la fertilidad femenina, la obesidad. La obesidad afecta el volumen, la concentración y la motilidad espermática, además de estar también asociada al incremento de daño en el ADN espermático<sup>(52)</sup>. Diversos estudios se han focalizado en este problema. Un ejemplo es el trabajo de An y colaboradores<sup>(52)</sup>, en el que se identificaron 973 lncRNAs diferencialmente expresados entre muestras de espermatozoides de ratas con obesidad inducida comparadas con ratas normales, revelando una asociación entre los lncRNAs, la reproducción masculina y la obesidad.

Los cánceres de células germinales testiculares (TGCT, del inglés, *Testicular Germ Cell Tumours*), son el tipo de tumor sólido más frecuente en hombres jóvenes en el mundo desarrollado, y su incidencia ha crecido durante las últimas décadas<sup>(53)</sup>. No obstante, la ausencia de modelos de laboratorio apropiados para su estudio, así como las dificultades para el establecimiento de cultivos primarios de células germinales masculinas humanas, representan un obstáculo para el avance del conocimiento acerca de la patogénesis de este tipo de tumores<sup>(54)</sup>. Aún así, se ha reportado alteración en el TGCT en algunos genes codificantes, y también en ARNs no codificantes. En particular, se ha detectado expresión del lncRNA *Xist* en todos los tipos histológicos de TGCT<sup>(55)</sup>. Del mismo modo, el lncRNA H19 estaría implicado en este tipo de tumores, dado que los patrones de impronta en la región 5' de este gen se encuentran borrados por completo en el TGCT; esto marca una diferencia con los tumores de tejidos somáticos, en que la

impronta disminuye pero no se borra por completo<sup>(55)</sup>.

Los lncRNAs de testículo también pueden estar relacionados con la diabetes mellitus tipo II (TDM2), ya que esta enfermedad tiene efectos estructurales y funcionales sobre el sistema reproductor masculino. La TDM2 paterna influencia las modificaciones epigenéticas durante la espermatogénesis, y esta disregulación epigenética puede ser heredada a las siguientes generaciones por la línea germinal, aumentando el riesgo de TDM2 en la descendencia<sup>(56)</sup>. Jiang y colaboradores<sup>(57)</sup> señalaron perfiles de expresión aberrantes de lncRNAs en espermatozoides de ratones diabéticos en comparación con ratones no diabéticos, encontrando unos 7.720 lncRNAs diferencialmente expresados entre ambos grupos (alrededor de 4.310 sobreexpresados y 3.410 regulados negativamente).

Sin embargo, son muy pocos aún los casos en que se ha estudiado la relación de lncRNAs específicos con patologías testiculares. A continuación ilustramos algunos ejemplos de trabajos en los que se vincula a lncRNAs en particular, con el desarrollo de patologías testiculares:

#### *NLC1-C y su vinculación con la azoospermia no obstructiva*

Una gran proporción de los casos de infertilidad masculina presenta arresto en la maduración testicular (MA) y azoospermia no obstructiva (NOA), u oligozoospermia severa. En pacientes con NOA se observó un patrón de expresión diferencial de lncRNAs en comparación con pacientes control normales. Se detectaron un total de 757 lncRNAs reprimidos y 475 lncRNAs sobreexpresados en pacientes con MA, y 2.370 lncRNAs reprimidos y 163 lncRNAs sobreexpresados en pacientes con hipoespermatogénesis (es decir, con niveles de espermatogénesis por debajo de lo normal)<sup>(13)</sup>.

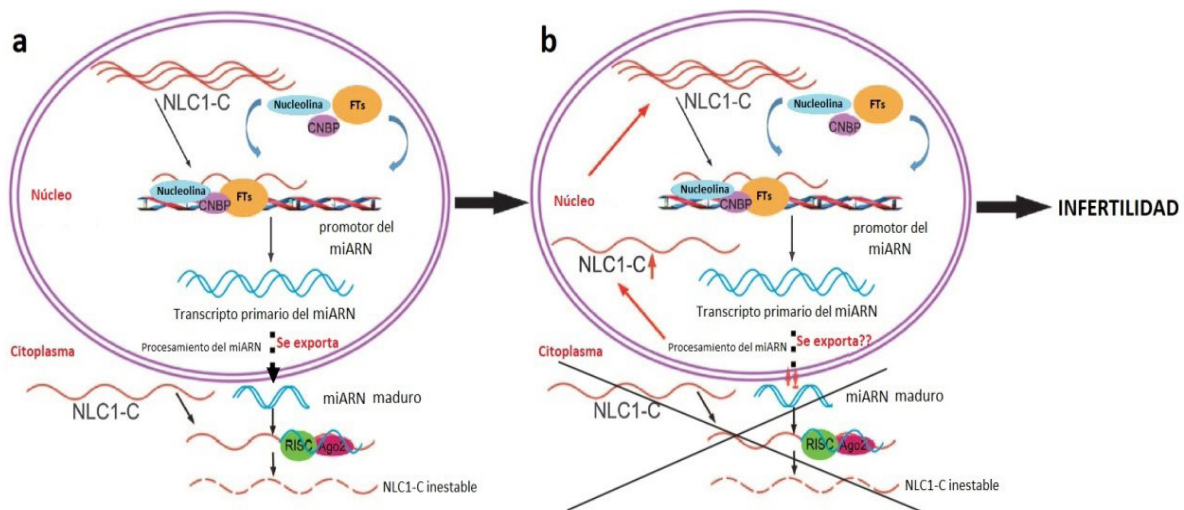
Entre los lncRNAs regulados negativamente en el grupo de pacientes con NOA, se encuentra el lncRNA

*NLC1-C* (en inglés, *NarcoLepsy Candidate-region 1*). Se presume que este ARN tendría una función en las etapas tempranas de la espermatogénesis regulando la proliferación o muerte de células germinales, ya que está predominantemente localizado en el citoplasma de las espermatogonias y espermatozoides tempranos en los testículos con espermatogénesis normal. Sin embargo, en los casos de testículos infértiles, el ARN *NLC1-C* se acumula en el núcleo de estas células<sup>(13)</sup>.

*NLC1-C* es capaz de unirse a proteínas, todas ellas asociadas al procesamiento de ARNs mensajeros inmaduros. Una de estas proteínas es la nucleolina, una proteína nucleolar cuya expresión coincide con la de *NLC1-C*, encontrándose ambos subexpresados en pacientes con MA y co-localizando en el núcleo celular en los testículos de estos pacientes<sup>(13)</sup>.

Por otra parte, el ARN *NLC1-C* regula micro ARNs, siendo un blanco directo para la unión del miR-320a y el miR-383, y actuando como secuestrador de dichos micro ARNs. Tanto el silenciamiento de *NLC1-C* como el de la nucleolina incrementan la expresión de ambos micro ARNs, y ambos unidos (*NLC1-C* y nucleolina) reprimen la expresión de los mismos. La acumulación de *NLC1-C* en el núcleo de las espermatogonias y espermatozoides primarios, con la consecuente represión de la expresión de miR-320a y miR-383, resulta en una proliferación hiperactiva de células germinales que lleva a la infertilidad<sup>(13)</sup>. Además, se ha visto que la acumulación de *NLC1-C* en el núcleo y su unión a la proteína nucleolina, promoverían también la proliferación celular en el carcinoma testicular embrionario<sup>(13)</sup>.

En suma, se propone un modelo para la acción de este lncRNA, que es ilustrado en la Figura 3. En condiciones normales, el lncRNA *NLC1-C* inhibe a los micro ARNs miR-320a y miR-383, a través de su unión a la nucleolina en el núcleo de las espermatogonias y de los espermatozoides primarios. Al ser exportado al citoplasma junto con los precursores de estos micro ARNs, los mismos son allí procesados en micro ARNs

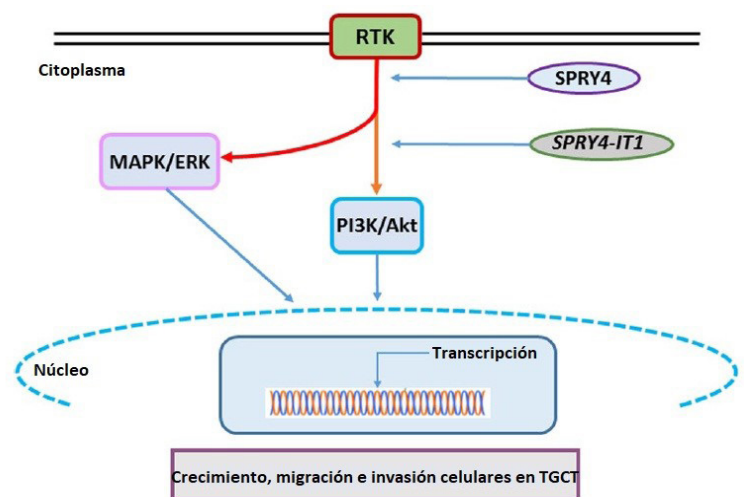


**Figura 3.** Modelo de la desregulación del lncRNA *NLC1-C* durante la espermatogénesis, que deriva en azoospermia no obstructiva e infertilidad. (a) En el testículo normal, *NLC1-C* es exportado al citoplasma junto a los precursores de los micro ARNs, que son procesados en el citoplasma. (b) Bajo desregulación de *NLC1-C*, éste se acumula en el núcleo, y junto a la nucleolina, inhibiría la producción de los precursores de los micro ARNs. Tomado y modificado de Lü et al., 2015<sup>(13)</sup>.

maduros que se unen a *NLC1-C* para regular la espermatogénesis. Sin embargo, cuando el *NLC1-C* está regulado negativamente en el citoplasma (no está claro aún si lo que falla es la exportación desde el núcleo al citoplasma) y es acumulado en el núcleo reprimiendo a los transcritos precursores de miR-320a y miR-383, lleva a la hiperactiva proliferación de las células germinales mencionada mediante su unión a la nucleolina, lo que en última instancia deriva en infertilidad masculina<sup>(13)</sup>.

#### *SPRY4-IT1* y los tumores de células germinales testiculares

Un estudio realizado empleando líneas celulares de carcinoma embrionario derivado de tejido metastásico de TGCT, detectó sobreexpresión del gen codificante para la proteína *SPRY4* y de su lncRNA intrónico *SPRY4-IT1* en muestras de TGCT en comparación con muestras de testículo normal. Al ser un lncRNA intrónico, el ARN *SPRY4-IT1* se produce a partir de una región del gen *SPRY4* que no produce proteína. Los patrones de expresión del gen codificante y del lncRNA difirieron notablemente entre los distintos tipos de TGCT: mientras la expresión del gen codificante era mayor en los tumores de saco vitelino, carcinoma embrionario y coriocarcinoma, los niveles de expresión del



**Figura 4.** Modelo de activación de las vías MAPK/ERK y PI3K/Akt por *SPRY4* y el lncRNA *SPRY4-IT1*, en células con TGCT. RTK: receptores de tirosín-kinasas. Tomado y modificado de Das et al., 2018<sup>(54)</sup>.

lncRNA eran mayores en teratomas. Las diferencias en el patrón de expresión entre el gen codificante y el lncRNA, junto con el hecho de que la disrupción de uno de ellos no alteraba la expresión del otro, que el lncRNA presentaba mayor sensibilidad a factores de crecimiento que el gen codificante, y que la degradación de ambos transcritos estaba regulada independientemente, sugieren que ambos funcionarían como transcritos independientes<sup>(54)</sup>.

Un descenso en los niveles de expresión de *SPRY4* y de *SPRY4-IT1* resultó en disminución de la viabilidad, proliferación, migración y capacidad de invasión celular. Por otra parte, la supresión de la

expresión tanto del gen codificante como del lncRNA mostró un efecto inhibitorio sobre las vías mediadas por receptores de tirosín-kinasas (RTK), al inhibirse principalmente la fosforilación de la kinasa Akt y, en menor medida, la fosforilación de las kinasas ERK1/2 (esta última fue dependiente del producto del gen codificante pero no del lncRNA) <sup>(54)</sup>. Estos factores son parte de las cascadas de señalización MAPK / ERK y PI3K/ Akt, involucradas en la homeostasis del crecimiento y diferenciación celular.

En conclusión, los resultados sugieren que tanto el gen codificante *SPRY4* como su lncRNA *SPRY4-IT1* actuarían como oncogenes en el TGCT, y que el bloque de la expresión de ambos inhibe el crecimiento de TGCT, al inhibir la activación de la vía PI3 / Akt (Figura 4). De todos modos, aún restan muchos estudios para comprender mejor los mecanismos de acción de estos genes y, en particular, del lncRNA *SPRY4-IT1*, en la patogénesis del cáncer testicular.

### *Malat1 y su participación en la isquemia testicular*

La torsión testicular afecta principalmente a hombres adolescentes y adultos, e involucra una torsión aberrante del cordón espermático que conlleva isquemia. Si esta falta de irrigación sanguínea es prolongada puede ocasionar daños irreparables, por lo que debe ser tratada rápidamente. Pero a su vez la detorsión puede provocar daño por reperfusión que puede ser aún más grave, lo que se conoce como daño por reperfusión de isquemia (en inglés, *Ischemia-Reperfusion Injury* o IRI) <sup>(58)</sup>.

Se estudió el rol del lncRNA *Malat1* en IRI testicular en ratones, observándose que su expresión se incrementa abruptamente luego de una hora de isquemia y 8 horas de reperfusión. Además, se comprobó que la sobreexpresión de este lncRNA inhibía notablemente la viabilidad y proliferación celular, promoviendo la apoptosis en las células testiculares espermatogénicas <sup>(58)</sup>.

Como hemos visto, los lncRNAs pueden unirse a

micro ARNs, modulando a su vez la represión de los blancos de los micro ARNs. El micro ARN miR-214 está inversamente expresado en IRI testicular con respecto a *Malat1*, y la sobreexpresión de *Malat1* reduce significativamente la expresión de miR-214, aunque este efecto no es recíproco. Ello indica que miR-214 es un blanco inhibitorio del lncRNA *Malat1* durante la progresión del IRI testicular <sup>(58)</sup>. A su vez, la sobreexpresión de *Malat1* promueve la expresión del gen *TRPV4* (en inglés, *Transient Receptor Potential Vanilloid 4*), un canal catiónico no selectivo cuya excesiva activación se ha visto en IRIs en miocardio, cerebro y pulmón; sin embargo, la presencia de miR-214 interrumpe este efecto. Se ha demostrado que la inducción de miR-214 inhibe la expresión de *TRPV4*, en tanto la represión del micro ARN aumenta la expresión del gen. En conjunto, la regulación de *TRPV4* por *Malat1* requiere de la actividad de miR-214 <sup>(58)</sup>.

En resumen, la sobreexpresión de *Malat1* es una característica molecular de la IRI testicular, que aparece asociada positivamente a la muerte celular inducida en IRI testicular, y negativamente a la proliferación celular a diferentes tiempos de reperfusión. Se identificó al micro ARN miR-214 como blanco de *Malat1*, y éste último inhibiría la expresión de miR-214. Por el contrario, el gen *TRPV4* sería un blanco inhibitorio del miR-214 que actúa como sensor para la IRI testicular <sup>(58)</sup>.

### *LncRNAs desregulados y síndrome de Klinefelter*

El síndrome de Klinefelter (SK) es un fenómeno común en hombres, con una frecuencia aproximada de 1:600 <sup>(59)</sup>. Este síndrome se caracteriza por la presencia de al menos un cromosoma X extra, a consecuencia de la no disyunción del par de cromosomas X durante alguna de las divisiones meióticas incluidas en el proceso de formación de los gametos masculinos. El 90 % de los casos del SK no corresponde a mosaicos sino que se trata, por ejemplo,

de individuos con cariotipo XXY, siendo éste el cariotipo mayoritario. Si bien los hombres con SK pueden presentar focos de espermatogénesis en sus testículos, el 90% de los casos no-mosaico presenta oligozoospermia severa o azoospermia completa<sup>(60)</sup>.

Recientemente se publicaron dos trabajos que relacionan al SK con lncRNAs. Winge y colaboradores<sup>(61)</sup> analizaron testículos fetales con SK y normales, obtenidos de abortos. Observaron una pérdida significativa de células germinales en los testículos fetales con SK en comparación con los normales, que es una característica propia de los testículos adultos con SK. Los autores sugieren por lo tanto que esta pérdida comenzaría en el feto, por una falla en la diferenciación de los gonocitos hacia pre-espermatogonias. Al analizar el transcriptoma (es decir, todo el ARN presente) en los testículos fetales con SK, y compararlos con testículos normales, encontraron que entre los transcritos diferenciales entre ambas condiciones había un enriquecimiento de lncRNAs ubicados en regiones intergénicas (o sea, regiones que no contienen genes codificantes para proteínas) en el genoma de los testículos con SK. En menor medida, también observaron un enriquecimiento en lncRNAs antisentido (transcritos a partir de la hebra de ADN opuesta a un gen que da lugar a una proteína) y lncRNAs sentido-intrónicos (en la misma hebra que un gen de una proteína, pero ubicados en la región intrónica del mismo). Sólo 2 de los 37 lncRNAs intergénicos expresados diferencialmente en el testículo con SK correspondieron a la transcripción de secuencias del cromosoma X. Uno de ellos es el lncRNA *Xist*, un resultado esperable ya que como fue explicado anteriormente, *Xist* colabora en la inactivación del X, y se genera a partir del mismo; en consecuencia, si hay un X supernumerario, no es extraño encontrar sobreexpresión de *Xist*. Este artículo propone que la pérdida de las células germinales en el feto de los pacientes con SK podría vincularse con la expresión aberrante de lncRNAs<sup>(61)</sup>.

El segundo artículo publicado recientemente en relación a SK y lncRNAs se centra en el lncRNA *GAS5* (en inglés, *Growth Arrest-Specific 5*). En este trabajo se compararon testículos con SK y testículos normales de igual edad, encontrándose 4.448 genes expresados diferencialmente en SK, entre los cuales el locus *GAS5* estaba sobreexpresado<sup>(62)</sup>. Este lncRNA está asociado a diversos procesos biológicos como la represión de la acción de glucocorticoides, enfermedades inflamatorias y autoinmunes, apoptosis vascular, aterosclerosis, y proliferación de células madre germinales femeninas<sup>(62)</sup>. Debido a su relación con los eventos de crecimiento, proliferación y desarrollo celular, señalización, e interacción célula-célula, se sugiere un posible rol para *GAS5* en la pérdida de células germinales en los testículos de pacientes con SK. De todos modos, si realmente su sobreexpresión es causa de esta característica fenotípica en la enfermedad, es algo que aún queda por investigar<sup>(62)</sup>.

### Los antígenos "cancer-testis"

Un caso diferente aunque muy interesante, es el de los "cancer-testis antigens". Si bien en sentido estricto no deberíamos incluirlos en esta revisión dado que no se encuentran involucrados en patologías testiculares, elegimos mencionarlos por lo curioso de los mismos. Se trata de un grupo de genes de expresión normal exclusivamente en el testículo, y cuya expresión en otros tejidos está asociada al cáncer, hecho del cual deriva su nombre.

Dentro de este grupo de los *cancer-testis antigens* se incluyen algunos lncRNAs. Un ejemplo lo constituye el lncRNA *DBF4B*, una isoforma no codificante del gen *DBF4B*, de expresión diferencial de espermatogénesis pero inhibida en espermatocitos, y que es regulada positivamente en el cáncer gástrico. Otro caso es el del lncRNA antisentido *DLG1*, que es específico del testículo y se encuentra sobreexpresado normalmente en

los espermatoцитos. Sin embargo, su expresión se detecta también en carcinoma escamoso, melanoma maligno, y enfermedad de Hodgkin. El lncRNA antisentido *KIRREL3-2* es también diferencial de espermatogénesis y sobreexpresado normalmente en los espermatoцитos, con un posible rol en la meiosis. Su presencia se ha detectado también en linfoma, enfermedad de Hodgkin, adenocarcinoma papilar, y melanoma maligno<sup>(63)</sup>.

En 2018 Cheng y colaboradores<sup>(64)</sup> relacionaron al lncRNA *THOR* (en inglés, *Testis-associated Highly conserved Oncogenic long non-coding RNA*) con el carcinoma hepatocelular (HCC), ya que lo detectaron altamente expresado en muestras de pacientes con dicha patología. Observaron inhibición del crecimiento y de la metástasis de HCC mediante la depleción de *THOR*.

El lncRNA *TDGRI* (en inglés, *Testis Development-Related Gene 1*) mostró elevada expresión en tejidos de cáncer epitelial de ovario (EOC), observándose que la regulación negativa de este lncRNA suprimía la proliferación, migración e invasión de células de EOC. Se ha propuesto que *TDGRI* actuaría induciendo el EOC mediante la reducción de la expresión del micro ARN miR-93 y sus genes blanco, por lo que no sólo podría considerarse un marcador de EOC, sino también un posible blanco terapéutico<sup>(65)</sup>.

### *Perspectivas y posibles aplicaciones de los ARNs no codificantes largos*

Con el surgimiento de las nuevas tecnologías de secuenciación, ha quedado en claro que la proporción de ADN que codifica para proteínas representa el porcentaje más pequeño del genoma, y que la inmensa mayoría corresponde a ARNs no codificantes. Dentro de éstos, un grupo mayoritario lo conforman los lncRNAs, cuyo descubrimiento ha cambiado dramáticamente la visión sobre la genética. Últimamente se les ha comenzado a prestar particular atención ya que, como hemos visto, se

ha acumulado evidencia creciente que demuestra su vínculo con diversos procesos biológicos, principalmente con funciones regulatorias, incluida la espermatogénesis (recordemos que el testículo es el tejido que expresa el mayor número de lncRNAs). En este trabajo, hemos revisado algunos de los diferentes estudios que los han relacionado también a diversas patologías. En este sentido, se ha propuesto a los lncRNAs como posibles biomarcadores para las diferentes afecciones, y también como posibles blancos de terapias génicas.

En particular para el caso del cáncer, el método más común de obtención de muestras tumorales ha sido la biopsia de tejido, que conlleva riesgo y es muy dolorosa para el paciente, además de económicamente costosa. Existe una nueva metodología de diagnóstico, la “biopsia líquida”, que es menos invasiva, reduce los riesgos, y, dependiendo del tipo de cáncer en cuestión, podría ser tomada de sangre, orina, saliva, entre otros, permitiendo detectar biomarcadores circulantes, tales como proteínas y ácidos nucleicos<sup>(66)</sup>. A pesar de este avance en los métodos diagnósticos, las pruebas basadas en detección de proteínas no son suficientemente selectivas y sensibles. Por otra parte, el análisis en el ADN de mutaciones en genes de cáncer no es específico de tejido, por lo que se necesitan análisis adicionales. En consecuencia, se está trabajando en la búsqueda de biomarcadores alternativos para el diagnóstico, que sean confiables y eficientes sobre todo en etapas tempranas de la tumorigénesis<sup>(66)</sup>.

El transcriptoma circulante representa una rica fuente de biomarcadores para el cáncer, incluyendo tanto ARNs codificantes como no codificantes. Entre estos últimos, los micro ARNs han sido estudiados exhaustivamente como potenciales biomarcadores. Sin embargo, en este último tiempo también ha comenzado a ponerse interés en otros tipos de ARNs no codificantes, como los lncRNAs<sup>(67)</sup>.

Los lncRNAs pueden funcionar como oncogenes (causan el crecimiento descontrolado de células

normales y las convierten en tumorales) o como supresores tumorales (protegen a las células de transformarse en células cancerosas). La transcripción aberrante de los mismos se ha confirmado como un predictor confiable de cáncer. Además, en múltiples tipos de cáncer se han encontrado lncRNAs circulantes en plasma, suero, orina, así como también en exosomas<sup>(67)</sup>.

El *PCA3* (en inglés, *Prostate Cancer Antigen 3*) es un lncRNA específico de próstata que se sobreexpresa en el cáncer de próstata, y es utilizado en una prueba diagnóstica de esta enfermedad donde se mide su concentración y la relación con otro biomarcador, en muestras de orina<sup>(68)</sup>. *MALAT1* se ha estudiado exhaustivamente como lncRNA circulante, habiéndose encontrado en plasma y suero de pacientes con cáncer de próstata, y en exosomas de pacientes con otros tipos de cáncer. Asimismo se han identificado otros lncRNAs circulantes en pacientes con cáncer, incluyendo *HOTAIR* y *GAS5*, entre otros<sup>(67)</sup>.

Además de su potencial uso como biomarcadores, los lncRNAs podrían ser empleados como blanco de terapias, debido a su gran especificidad de tejido, célula o tumor<sup>(69)</sup>. Esto podría llevarse a cabo mediante degradación postranscripcional del ARN, modulación de los lncRNAs por bloqueo de su promotor o utilizando técnicas de edición genómica efectivas y sin blancos inespecíficos como lo es la tecnología CRISPR, o mediante pérdida de función del lncRNA ya sea con inhibidores de interacciones ARN-proteínas o impidiendo la formación de estructuras secundarias en el ARN<sup>(69)(70)</sup>. A modo de ejemplo, un reporte reciente indica que la inhibición del lncRNA *H19* podría constituir una terapia efectiva para el tratamiento del cáncer de páncreas<sup>(71)</sup>.

A pesar de todos los esfuerzos puestos en el estudio de los lncRNAs y en la elucidación de sus funciones, el mundo de los lncRNAs está aún en sus inicios y mucho queda aún por hacer para lograr identificar y describir sus mecanismos moleculares de acción.

Sin dudas, los próximos años marcarán importantes avances en el conocimiento de este novedoso grupo de ARNs, y en su eventual utilidad para el diagnóstico y la terapia. Dada la gran abundancia de los lncRNAs a nivel testicular, estos avances serán especialmente relevantes para el mejor conocimiento de los eventos moleculares subyacentes al complejo proceso de la espermatogénesis, así como para el diagnóstico y tratamiento de patologías como la infertilidad y el cáncer testicular.

### Agradecimientos:

*Este trabajo contó con el apoyo de CSIC (UDELAR, Programa Grupos I+D 2018 “Bases moleculares de la gametogénesis masculina: abordaje multidisciplinario con metodologías y equipamiento de avanzada”) y ANII (Agencia Nacional de Investigación e Innovación, proyecto FCE\_1\_2014\_1\_104251).*

### Referencias:

1. Mattick JS. The central role of RNA in human development and cognition. *FEBS Lett.* 2011;585(11):1600-16.
2. Luk AC, Chan WY, Rennert OM, Lee TL. Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptome studies. *Reproduction.* 2014;147(5):131-41.
3. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet.* 2009;10(2):94-108.
4. Atkinson SR, Marguerat S, Bähler J. Exploring long non-coding RNAs through sequencing. *Semin Cell Dev Biol.* 2012;23(2):200-205.
5. Li LJ, Leng RX, Fan YG, Pan HF, Ye DQ. Translation of noncoding RNAs: Focus on lncRNAs, pri-miRNAs, and circRNAs. *Exp Cell Res.* 2017;361(1):1-8.
6. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long

- noncoding RNAs in the human transcriptome. *Nat Genet.* 2015;47(3):199-208.
7. Dinger ME, Amaral PP, Mercer TR, Pang KC, Bruce SJ, Gardiner BB, et al. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 2008;18(9):1433-45.
  8. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci USA.* 2008;105(2):716-21.
  9. Pauli A, Valen E, Lin MF, Garber M, Vastenhouw NL, Levin JZ, et al. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res.* 2012;22(3):577-91.
  10. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009;458(7235):223-7.
  11. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011;25(18):1915-27.
  12. de Hoon M, Shin JW, Carninci P. Paradigm shifts in genomics through the Fantom projects. *Mamm Genome.* 2015;26(9-10):391-402.
  13. Lü M, Tian H, Cao YX, He X, Chen L, Song X, et al. Downregulation of miR-320a/383-sponge-like long noncoding RNA NLC1-C (narcolepsy candidate-region 1 genes) is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation. *Cell Death Dis.* 2015;6:e1960.
  14. Hong SH, Kwon JT, Kim J, Jeong J, Kim J, Lee S, et al. Profiling of testis-specific long noncoding RNAs in mice. *BMC Genomics.* 2018;19:539.
  15. Sanchez Calle A, Kawamura Y, Yamamoto Y, Takeshita F, Ochiya T. Emerging roles of long non-coding RNA in cancer. *Cancer Sci.* 2018;109(7):2093-2100.
  16. Kamikawa Y, Donohoe ME. The dynamics of X-chromosome inactivation in mouse development. *Mol Reprod Dev.* 2014;81(2):141-7.
  17. Pontier DB, Gribnau J. Xist regulation and function eXplored. *Hum Genet.* 2011;130(2):223-36.
  18. Froberg JE, Yang L, Lee JT. Guided by RNAs: X-inactivation as a model for lncRNA function. *J Mol Biol.* 2013;425(19): 3698-706.
  19. Wei JW, Huang K, Yang C, Kang CS. Non-coding RNAs as regulators in epigenetics. *Oncol Rep.* 2017;37(1):3-9.
  20. Cai X, Cullen BR. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA.* 2007;13(3):313-6.
  21. Bhartiya D, Kapoor S, Jalali S, Sati S, Kaushik K, Sachidanandan C, et al. Conceptual approaches for lncRNA drug discovery and future strategies. *Expert Opin Drug Discov.* 2012;7(6):503-13.
  22. Calin GA, Liu CG, Ferracin M, Hyslop T, Spizzo R, Sevignani C, et al. Ultranconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell.* 2007;12(3):215-29.
  23. Azzi S, Abi Habib W, Netchine I. Beckwith-Wiedemann and Russell-Silver syndromes: from new molecular insights to the comprehension of imprinting regulation. *Curr Opin Endocrinol Diabetes Obes.* 2017;21(1):30-8.
  24. Raveh E, Matouk IJ, Gilon M, Hochberg A. The H19 long non-coding RNA in cancer initiation, progression and metastasis - A

- proposed unifying theory. *Mol Cancer*. 2015;14:184.
25. Dong X, He X, Guan A, Huang W, Jia H, Huang Y, et al. Long non-coding RNA Hotair promotes gastric cancer progression via miR-217-GPC5 axis. *Life Sci*. 2019;217:271-82.
26. Barman P, Reddy D, Bhaumik SR. Mechanisms of antisense transcription initiation with implications in gene expression, genomic integrity and disease pathogenesis. *Noncoding RNA*. 2019;5(1):e11.
27. Weakley SM, Wang H, Yao Q, Chen C. Expression and function of a large non-coding RNA gene XIST in human cancer. *World J Surg*. 2011;35(8):1751-6.
28. Dandan W, Jianliang C, Haiyan H, Hang M, Xuedong L. Long noncoding RNA MIR31HG is activated by SP1 and promotes cell migration and invasion by sponging miR-214 in NSCLC. *Gene*. 2019;692:223-30.
29. Brewer CJ, Balen AH. The adverse effect of obesity on conception and implantation. *Reproduction*. 2010;140(3):347-64.
30. Huang BB, Liu XC, Qin XY, Chen J, Ren PG, Deng WF, et al. Effect of high-fat diet on immature female mice and messenger and noncoding RNA expression profiling in ovary and white adipose tissue. *Reprod Sci*. 2018.
31. Liu YD, Li Y, Feng SX, Ye DS, Chen X, Zhou XY, et al. Long noncoding RNAs: potential regulators involved in the pathogenesis of polycystic ovary syndrome. *Endocrinology*. 2017;158(11):3890-9.
32. Moradi MT, Rahimi Z, Vaisi-Raygani A. New insight into the role of long non-coding RNAs in the pathogenesis of preeclampsia. *Hypertens Pregnancy*. 2019;38(1):41-51.
33. Geisinger A. Spermatogenesis in Mammals: a very peculiar cell differentiation process. *Cell Differentiation Research Developments*. In: Ivanova LB, editor. New York: Nova Publishers; 2008. p. 97-123.
34. Da Cruz I, Rodríguez-Casuriaga R, Santiñaque FF, Farías J, Curti G, Capiano CA, et al. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to-postmeiotic-related processes at pachytene stage. *BMC Genomics*. 2016;17:294.
35. Soumillon M, Necseulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep*. 2013;3(6):2179-90.
36. Necseulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*. 2014;505(7485):635-40.
37. Liang M, Li W, Tian H, Hu T, Wang L, Lin Y, et al. Sequential expression of long noncoding RNAs as mRNA gene expression in specific stages of mouse spermatogenesis. *Sci Rep*. 2014;4:5966.
38. Bao J, Wu J, Schuster AS, Hennig GW, Yan W. Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline. *Biol Reprod*. 2013;89(5):107.
39. Lin X, Han M, Cheng L, Chen J, Zhang Z, Shen T, et al. Expression dynamics, relationships, and transcriptional regulations of diverse transcripts in mouse spermatogenic cells. *RNA Biol*. 2016;13(10):1011-24.
40. Wichman L, Somasundaram S, Breindel C, Valerio DM, McCarrey JR, Hodges CA, et al. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol Reprod*. 2017;97(2):313-23.
41. Zhu Z, Li C, Yang S, Tian R, Wang J, Yuan Q, et al. Dynamics of the transcriptome during human spermatogenesis: predicting the

- potential key genes regulating male gametes generation. *Sci Rep.* 2016;6:19069.
42. Jan SZ, Vormer TL, Jongejan A, Röling MD, Silber SJ, de Rooij DG, et al. Unraveling transcriptome dynamics in human spermatogenesis. *Development.* 2017;144(20):3659-73.
  43. Ran M, Chen B, Li Z, Wu M, Liu X, He C, et al. Systematic identification of long noncoding RNAs in immature and mature porcine testes. *Biol Reprod.* 2016;94(4):77.
  44. Liu Y, Sun Y, Li Y, Bai H, Xue F, Xu S, et al. Analyses of Long Non-Coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. *Sci Rep.* 2017;7(1):9055.
  45. Akhade VS, Arun G, Donakonda S, Rao MR. Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells. *RNA Biol.* 2014;11(10):1262-79.
  46. Ni MJ, Hu ZH, Liu Q, Liu MF, Lu M, Zhang JS, et al. Identification and characterization of a novel non-coding RNA involved in sperm maturation. *Plos One.* 2011;6(10):e26053.
  47. Anguera MC, Ma W, Clift D, Namekawa S, Kelleher III RJ, Lee JT. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *Plos Genet.* 2011;7(9):e1002248.
  48. Li L, Wang M, Wu X, Geng L, Xue Y, Wei X, et al. A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells. *Cell Death Dis.* 2016;7:e2140.
  49. National Collaborating Centre for Women's and Children's Health. *Fertility: Assessment and Treatment for People with Fertility Problems.* London: Royal College of Obstetricians & Gynaecologists (UK), 2013.
  50. Wang X, Yang C, Guo F, Zhang Y, Ju Z, Jiang Q, et al. Integrated analysis of mRNA and long noncoding RNAs in the semen from Holstein bulls with high and low sperm motility. *Sci Rep.* 2019;9(1):2092.
  51. Zhang X, Zhang P, Song D, Xiong S, Zhang H, Fu J, et al. Expression profiles and characteristics of human lncRNAs in normal and asthenozoospermia sperm. *Biol Reprod.* 2018;0(0):1-12.
  52. An T, Fan H, Liu YF, Pan YY, Liu YK, Mo FF, et al. The difference in expression of long noncoding RNAs in rat semen induced by high-fat diet was associated with metabolic pathways. *Peer J.* 2017;5:e3518.
  53. Znaor A, Lortet-Tieulent J, Jemal A, Bray F. International variations and trends in testicular cancer incidence and mortality. *Eur Urol.* 2014;65(6):1095-106.
  54. Das MK, Furu K, Evensen HF, Haugen ØP, Haugen TB. Knockdown of SPRY4 and SPRY4-IT1 inhibits cell growth and phosphorylation of Akt in human testicular germ cell tumours. *Sci Rep.* 2018;8(1):2462.
  55. Okamoto K. Epigenetics: a way to understand the origin and biology of testicular germ cell tumors. *Int J Urol.* 2012;19(6):504-11.
  56. Ding GL, Liu Y, Liu ME, Pan JX, Guo MX, Sheng JZ, et al. The effects of diabetes on male fertility and epigenetic regulation during spermatogenesis. *Asian J Androl.* 2015;17(6):948-53.
  57. Jiang GJ, Zhang T, An T, Zhao DD, Yang XY, Zhang DW, et al. Differential expression of long noncoding RNAs between sperm samples from diabetic and non-diabetic mice. *Plos One.* 2016;11(4):e0154028.
  58. Li W, Ning JZ, Cheng F, Yu WM, Rao T, Ruan Y, et al. MALAT1 Promotes cell apoptosis and suppresses cell proliferation in testicular ischemia-reperfusion injury by sponging miR-214 to modulate TRPV4 expression.

- Cell Physiol Biochem. 2018;46(2):802-14.
59. Bojesen A, Juul S, Gravholt CH. Prenatal and postnatal prevalence of Klinefelter syndrome: a national registry study. *J Clin Endocrinol Metab.* 2013;88(2):622-6.
60. De Sanctis V, Ciccone S. Fertility preservation in adolescents with Klinefelter's syndrome. *Pediatr Endocrinol Rev.* 2010;(Suppl 1):178-81.
61. Winge SB, Dalgaard MD, Jensen JM, Graem N, Schierup MH, Juul A, et al. Transcriptome profiling of fetal Klinefelter testis tissue reveals a possible involvement of long non-coding RNAs in gonocyte maturation. *Human Mol Genetics.* 2018;27(3):430-9.
62. Salemi M, Cannarella R, Condorelli RA, Cimino L, Ridolfo F, Giurato G, et al. Evidence for long noncoding RNA GAS5 up-regulation in patients with Klinefelter síndrome. *BMC Med Genet.* 2019;20:4.
63. Dianatpour A, Ghafouri-Fard S. Long non coding RNA expression intersecting cancer and spermatogenesis: a systematic review. *Asian Pac J Cancer Prev.* 2017;18(10):2601-10.
64. Cheng Z, Lei Z, Yang P, Si A, Xiang D, Zhou J, et al. Long non-coding RNA THOR promotes cell proliferation and metastasis in hepatocellular carcinoma. *Gene.* 2018;678:129-36.
65. Chen S, Wang L, Sun KX, Xiu Y, Zong ZH, Chen X, et al. The role of long non-coding RNA TDRG1 in epithelial ovarian carcinoma tumorigenesis and progression through miR-93/RhoC pathway. *Mol Carcinogenesis.* 2017;57(2): 225-34.
66. Miranda-Castro R, de-los-Santos-Álvarez N, Lobo-Castañón MJ. Long noncoding RNAs: from genomic junk to rising stars in the early detection of cancer. *Anal Bioanal Chem.* 2019;0(0):1-11.
67. Sole C, Arnaiz E, Manterola L, Otaegui D, Lawrie CH. The circulating transcriptome as a source of cancer liquid biopsy biomarkers. *Semin Cancer Biol.* 2019;pii: S1044-579X(18)30100-7.
68. Sartori DA, Chan DW. Biomarkers in prostate cancer: what's new? *Curr Opin Oncol.* 2014;26(3):259-64.
69. Arun G, Diermeir SD, Spector DL. Therapeutic targeting of long non-coding RNAs in cancer. *Trends Mol Med.* 2018;24(3):257-77.
70. Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, et al. RNA targeting with CRISPR-Cas13a. *Nature.* 2017;550(7675):280-4.
71. Yoshimura H, Matsuda Y, Yamamoto M, Kamiya S, Ishiwata T. Expression and role of long non-coding RNA H19 in carcinogenesis. *Front Biosci.* 2018;23:614-25.

# **Fracciones celulares puras y librerías hebra- específicas para *RNAseq***

## **4.2: CAPÍTULO 2**

#### **4.2.1: Obtención de fracciones celulares puras de las poblaciones espermatogénicas de interés por citometría de flujo**

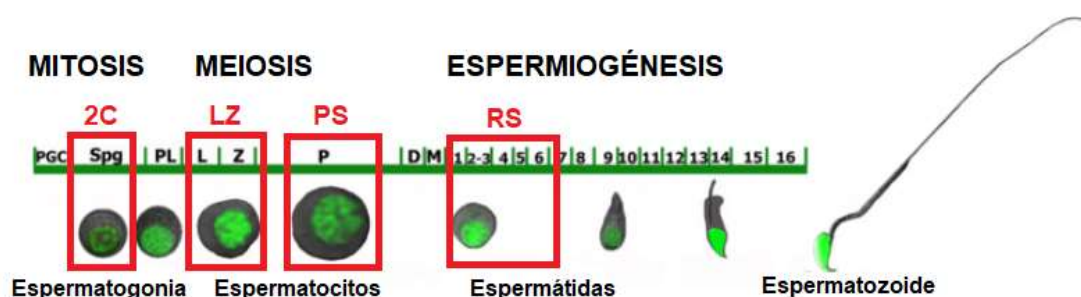
Mediante la metodología de purificación de fracciones celulares por citometría de flujo desarrollada por nuestro grupo de investigación, se lograron obtener 12 fracciones correspondientes a 4 poblaciones celulares por triplicado: 2C, LZ, PS y RS (Figura 9), similares a las que habían sido obtenidas en el trabajo previo del grupo, de da Cruz y cols. [da Cruz *et al.*, 2016], pero con pequeñas diferencias en las edades de los ratones utilizados. Esto fue particularmente relevante en relación a la purificación de la fracción de PS. El proceso de disgregación mecánica del tejido puede generar multinucleados, que se producen dada la naturaleza sincitial del epitelio seminífero, donde las células que se originaron a partir de una misma espermatogonia quedan unidas por puentes citoplásmicos [Meistrich, 1977]. Los multinucleados de células RS, formados por el “englobamiento” de cuatro espermátidas, pueden llegar a clasificarse como población 4C (sobre todo PS, ya que pasan a tener un contenido de ADN y tamaño similares), representando un contaminante para dicha población. Por este motivo, la elección de edades más tempranas (en que aún no han aparecido RS en el desarrollo de la onda espermatogénica) para la clasificación de PS con respecto al trabajo de da Cruz, nos permitió asegurarnos de la ausencia total de RS y por tanto de sus multinucleados, eliminando toda posible contaminación de la fracción PS.

Una segunda modificación introducida durante el desarrollo de esta tesis, fue la incorporación de un tratamiento con collagenasa durante la disgregación celular. El objetivo era lograr un mejor rendimiento de la población 2C, particularmente de la capa más externa de células que a menudo resulta más difícil de liberar por encontrarse adherida a la pared de los túbulos seminíferos. La introducción del tratamiento con collagenasa mejoró el rendimiento de la misma sin perjudicar la viabilidad celular:

-2C purificada sin tratamiento con collagenasa:  $35,6 \pm 6\%$  de los eventos totales (representación de ese tipo celular en el total de la suspensión); de ellas,  $29,8 \pm 6\%$  eran células viables.

-2C purificada con tratamiento con collagenasa:  $56,6 \pm 1\%$  de los eventos totales; de ellas,  $43,4 \pm 0,8\%$  eran células viables.

De esta forma pudimos liberar más cantidad de células 2C, que de otra forma habrían quedado unidas a la pared de los túbulos, perdiéndose en la disgregación.



**Figura 9.** Esquema de la espermatogénesis del ratón, donde se señalan las poblaciones celulares purificadas en rojo. Tomado y modificado de da Cruz *et al.*, [2016].

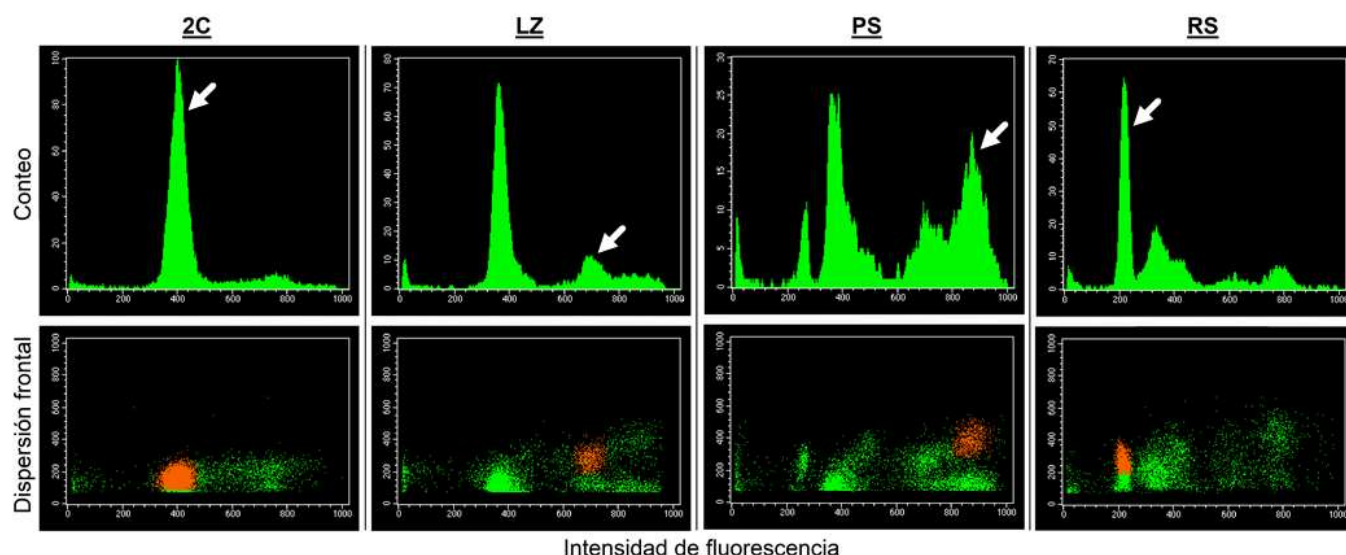
En cuanto al rendimiento celular por ratón, existió una alta variabilidad debido tanto a la variación y maduración individuales de los ratones, como al rango de edades o la población celular a clasificar. Los valores alcanzados se pueden encontrar en la Tabla 3.

**Tabla 3.** Rendimiento de las purificaciones celulares obtenidas por citometría de flujo.

| Fracción celular | % eventos totales* | Células obtenidas / ratón |
|------------------|--------------------|---------------------------|
| 2C (12-14 dpp)   | 48,6 - 64,8 %      | 97.000 - 258.000          |
| LZ (15-18 dpp)   | 3,6 - 6,9 %        | 2.800 - 17.000            |
| PS (16-19 dpp)   | 2,0 - 11,0 %       | 5.800 - 33.000            |
| RS (22-24 dpp)   | 11,0 - 29,5 %      | 170.000 - 265.000         |

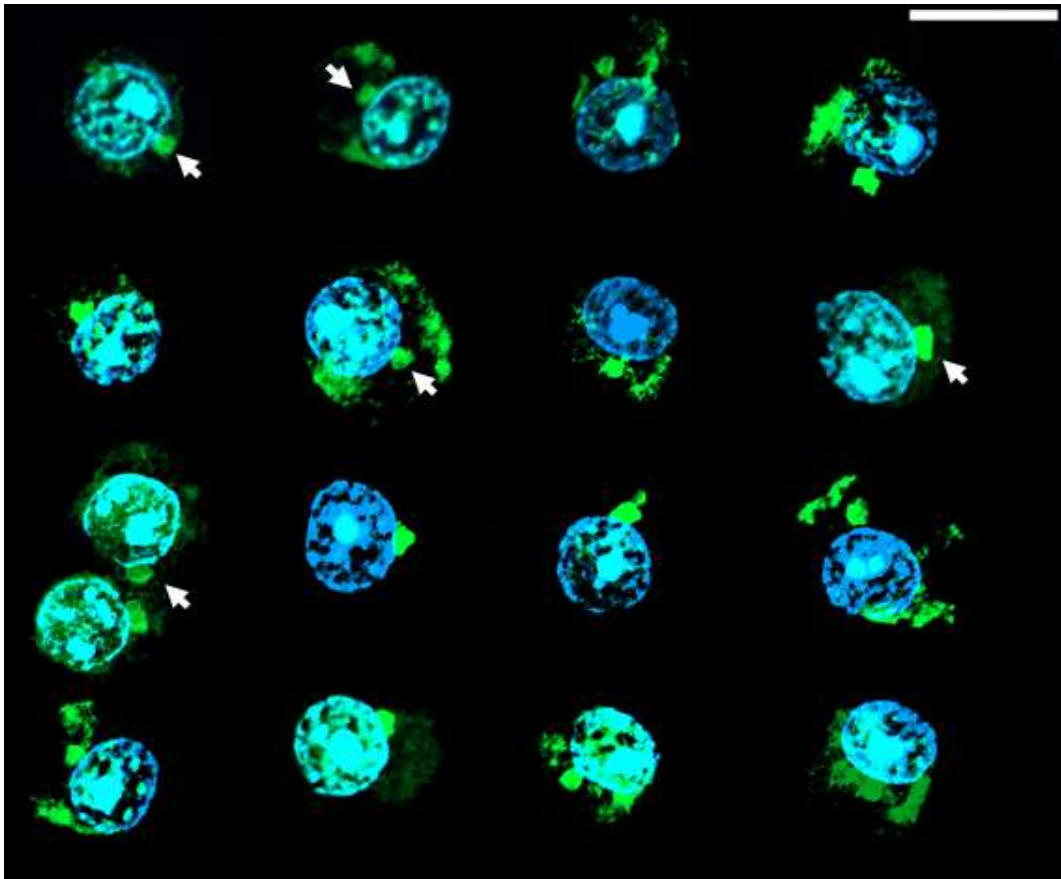
\*Refiere a la representación de ese tipo celular en el total de la suspensión testicular.

La purificación de las células se realizó con exclusión de dobletes y de restos celulares no viables. Los gráficos correspondientes a la clasificación de las distintas poblaciones celulares pueden verse en la Figura 10.



**Figura 10.** Histogramas y gráficos de puntos (*dot plots*) de la clasificación en flujo de las distintas poblaciones celulares de interés (2C, LZ, PS, RS). Los picos señalados con las flechas blancas y las nubes de puntos rojos se corresponden con la fracción purificada en cada caso.

Sabemos de antemano que 2C no es una fracción pura, ya que es una mezcla de células somáticas y de línea germinal; por otra parte, dado que en trabajos anteriores se demostró la pureza de las fracciones LZ y PS [Rodríguez-Casuriaga *et al.*, 2014; Geisinger & Rodríguez-Casuriaga, 2017], en esta ocasión nos enfocamos en verificar la pureza de la fracción RS, que en trabajos anteriores se había realizado exclusivamente por morfología, pero no con marcadores específicos. Para ello realizamos una purificación por citometría de flujo de la población RS con deposición en placa, y sobre ella una inmunodetección de la proteína MVH, que además de presentar una marca difusa en el citoplasma, marca el cuerpo cromatoide, organelo sin membrana típico de células post-meióticas [Meikar *et al.*, 2014]. De esta manera, pudimos observar que la fracción RS tiene una pureza de ~99% (Figura 11).



**Figura 11.** Compilado de imágenes de células RS purificadas por citometría de flujo vistas al microscopio confocal. Anticuerpo anti-MVH (ab13840) en verde, con señal citoplásmica y específica en el cuerpo cromatoide (flechas blancas). Núcleos contrateñidos con DAPI en azul. Barra: 10  $\mu$ m.

#### 4.2.2: Construcción de librerías hebra-específicas para secuenciación masiva

El primer paso para la construcción de las librerías para *RNAseq* consistió en extraer ARN total de buena calidad de las 12 fracciones celulares obtenidas anteriormente. Esto no es trivial, ya que se trata de células que han sido sometidas al estrés de disgregación mecánica, y pasaje por el citómetro de flujo. El rendimiento de ARN es variable y depende de cada tipo celular, ya que cada uno contiene diferentes cantidades de ARN. En promedio, las cantidades de células requeridas para obtener 100 ng de ARN total, fueron las siguientes:

- 2C: 99.200-179.500 células
- LZ: 46.600-100.900 células

-PS: 38.800-47.000 células

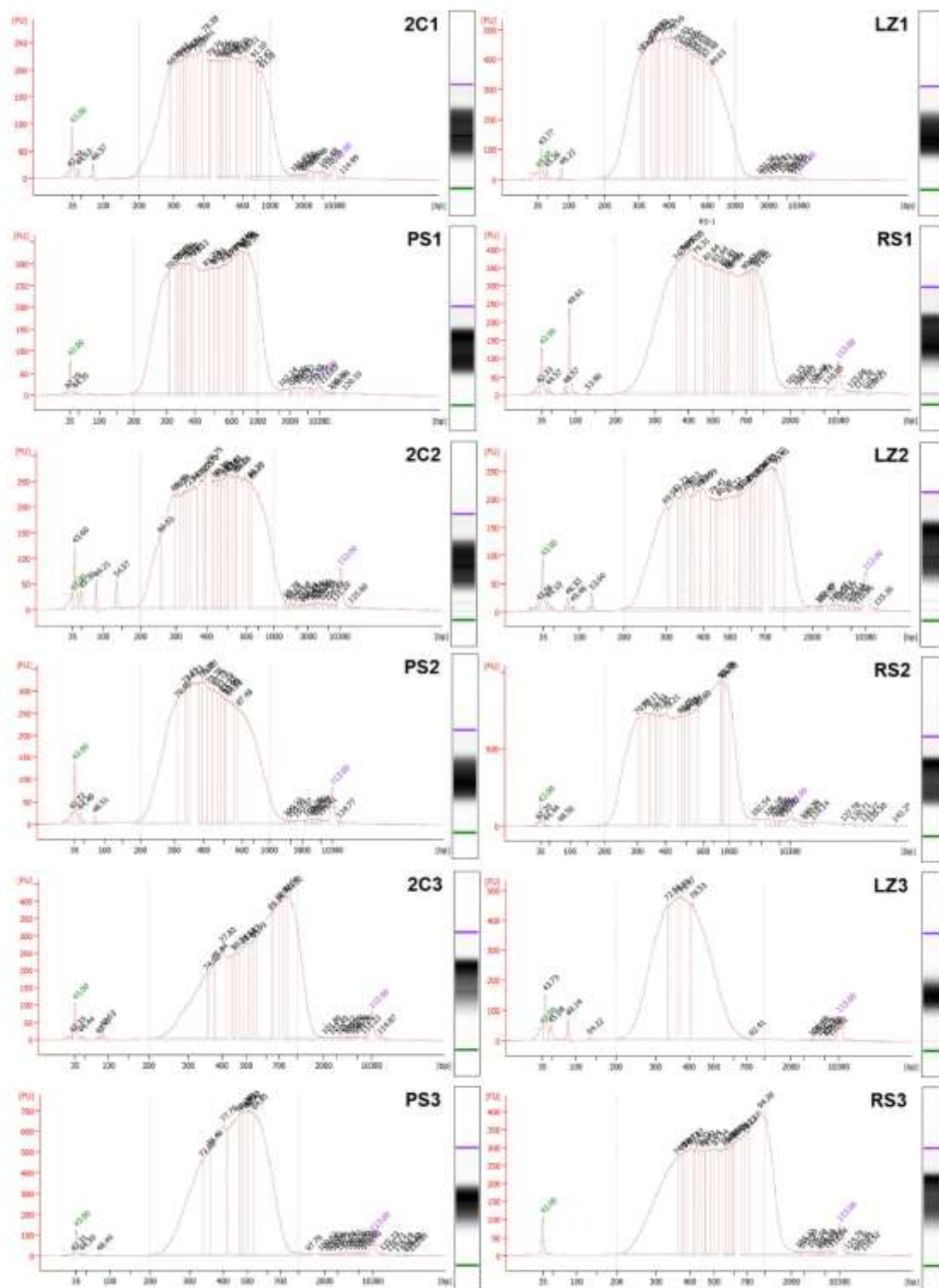
-RS: 140.700-237.900 células

Esto pone en evidencia el elevado contenido de ARN de los PS, que son células de gran tamaño y con importante actividad transcripcional.

Luego de obtenido el ARN, se procedió a la construcción de librerías hebra-específicas. Cabe mencionar la importancia de que las librerías sean direccionales, debido a que existe un alto porcentaje de ARNs antisentido a otros genes en la hebra opuesta, por lo que este tipo de librerías permite diferenciar la expresión de cada uno, ya que distingue las lecturas provenientes de cada hebra. El kit comercial utilizado para su construcción presenta además otras dos ventajas: a) permite una cantidad de ARN de partida muy baja (desde 10 ng) ya que amplifica la muestra mediante PCR a bajo número de ciclos (16), lo que es importante en nuestro caso dado que las células purificadas por citometría de flujo son escasas y difíciles de obtener; y b) depleta de ARNs ribosomales y mitocondriales específicos de ratón mediante unos adaptadores y cuentas magnéticas, lo que excluye del análisis ARNs que no son de interés en el estudio de los lncRNAs expresados y diferenciales.

Partimos de 60 ng de ARN total de entrada y el rendimiento obtenido fue variable, resultando entre 140 y 1.510 ng de ADNc por librería. La calidad de las mismas fue evaluada mediante corrida en un chip para ADN en el *Bioanalyzer 2100 Agilent*, cuyos electroferogramas se muestran en la figura 12.

Estas librerías fueron enviadas a la empresa *Fasteris* (Suiza) para su secuenciación.



**Figura 12.** Electroferogramas de las corridas en el *Bioanalyzer* para control de calidad de las 12 librerías de *RNAseq* construidas.

# **Transcriptoma global de lncRNAs a lo largo de la espermatogénesis del ratón**

## **4.3: CAPÍTULO 3**

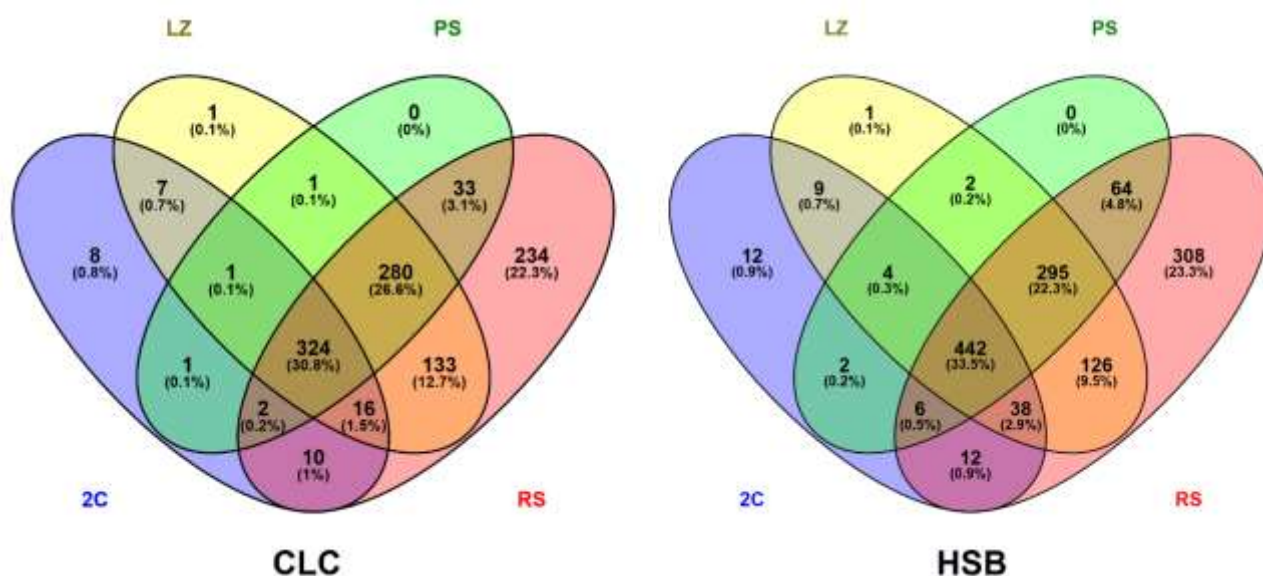
El capítulo 3 de esta tesis comprende el trabajo más extenso, y trata sobre la caracterización global del transcriptoma de lncRNAs en las distintas etapas estudiadas de la espermatogénesis. Seguidamente se mencionarán brevemente los resultados más importantes, ya que la gran mayoría son descritos y discutidos en profundidad en el artículo publicado en la revista *RNA Biology* que titulamos “*Revealing stage-specific expression patterns of long noncoding RNAs along mouse spermatogenesis*” [Trovero *et al.*, 2020], y que se adjunta a continuación.

#### **4.3.1: Análisis general de los datos de secuenciación obtenidos**

Se obtuvieron en promedio 150 millones de lecturas crudas secuenciadas por cada muestra. Mapearon contra el genoma de referencia entre 84 y 94% de las lecturas filtradas remanentes, con ambos abordajes empleados (*TrimGalore-HISAT-StringTie-Ballgown*, de ahora en adelante “HSB”; y *CLC Genomics Workbench*, de aquí en más “CLC”) (Tabla Suplementaria S1 del artículo).

Los datos de secuenciación fueron analizados mediante los dos abordajes alternativos en paralelo, CLC y HSB. Alrededor de un 60% de los genes de lncRNAs expresados coinciden entre ambos análisis (878 lncRNAs), con una varianza entre muestras  $> 1$  y  $FPKM \geq 0,1$  (83,5% de los genes expresados identificados con CLC coincidieron con los de HSB; Figura 1A del artículo y Tabla Suplementaria S2 del artículo). Los resultados de cada abordaje por separado se muestran en la figura 13. Cuando variamos el corte de expresión a  $FPKM \geq 0,3$ ,  $0,5$  ó  $1$  haciéndolo más restrictivo, se mantuvieron los mismos genes de lncRNAs en la lista de expresados. Esto indicó la confiabilidad y solidez del grupo de lncRNAs con el que se continuó realizando los análisis.

Un 0,22% y 3,28% de las lecturas mapeó contra el ARN ribosomal y mitocondrial, respectivamente. Estos resultados nos permitieron concluir que la depleción de estos ARNs con el kit de construcción de librerías fue muy eficiente.



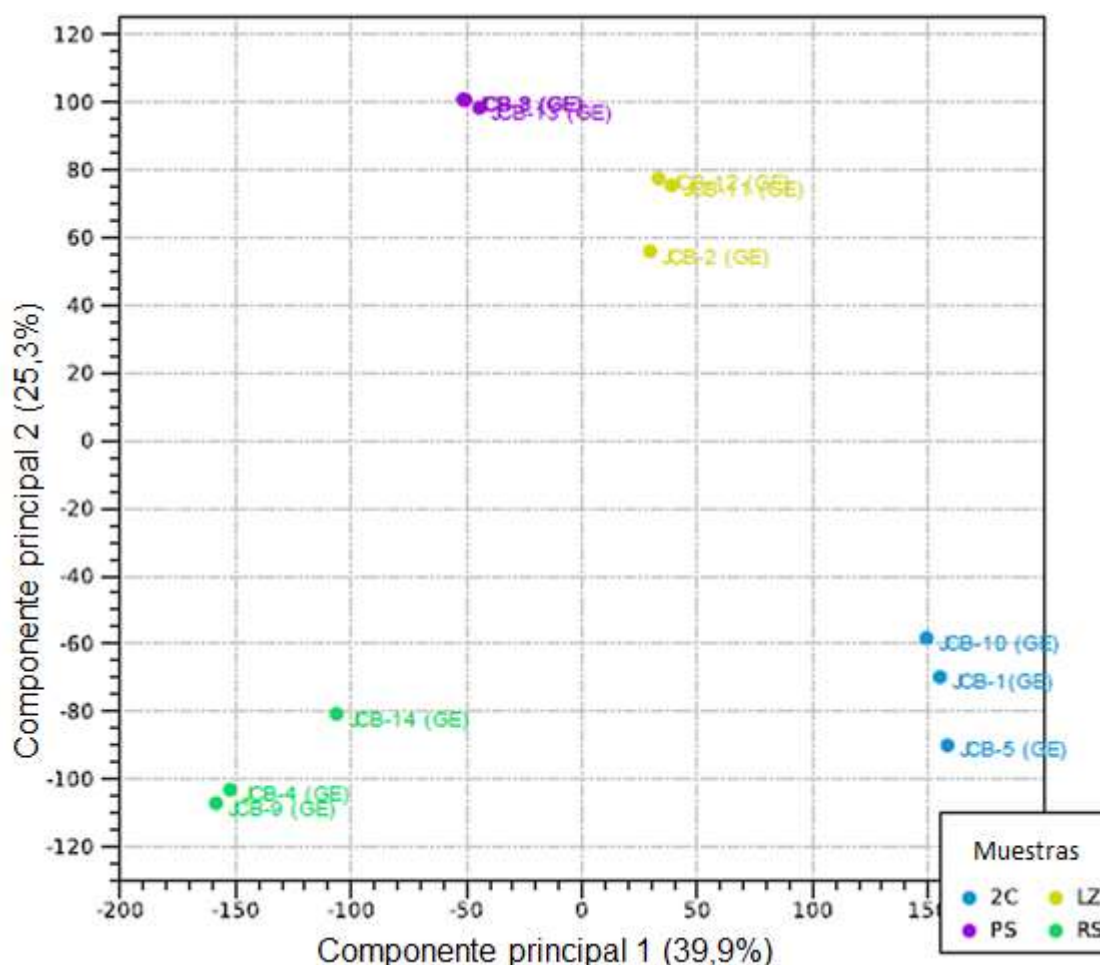
**Figura 13.** LncRNAs expresados con los parámetros de varianza entre muestras  $> 1$  y FPKM  $\geq 0,1$ , detectados mediante los análisis con CLC y HSB.

Adicionalmente, las réplicas de cada muestra fueron altamente reproducibles, lo que puede verse en el gráfico del análisis de componentes principales (PCA) (Figura 14) y en la matriz de correlación (Figura Suplementaria S2 del artículo). Esto se reflejó en el hecho de que las réplicas de una misma población se agruparon entre sí, separadas de las demás muestras. Con esto concluimos que tanto la metodología de purificación celular, como las librerías construidas y su secuenciación fueron exitosas.

#### 4.3.2: LncRNAs expresados y diferenciales en las distintas poblaciones de células testiculares

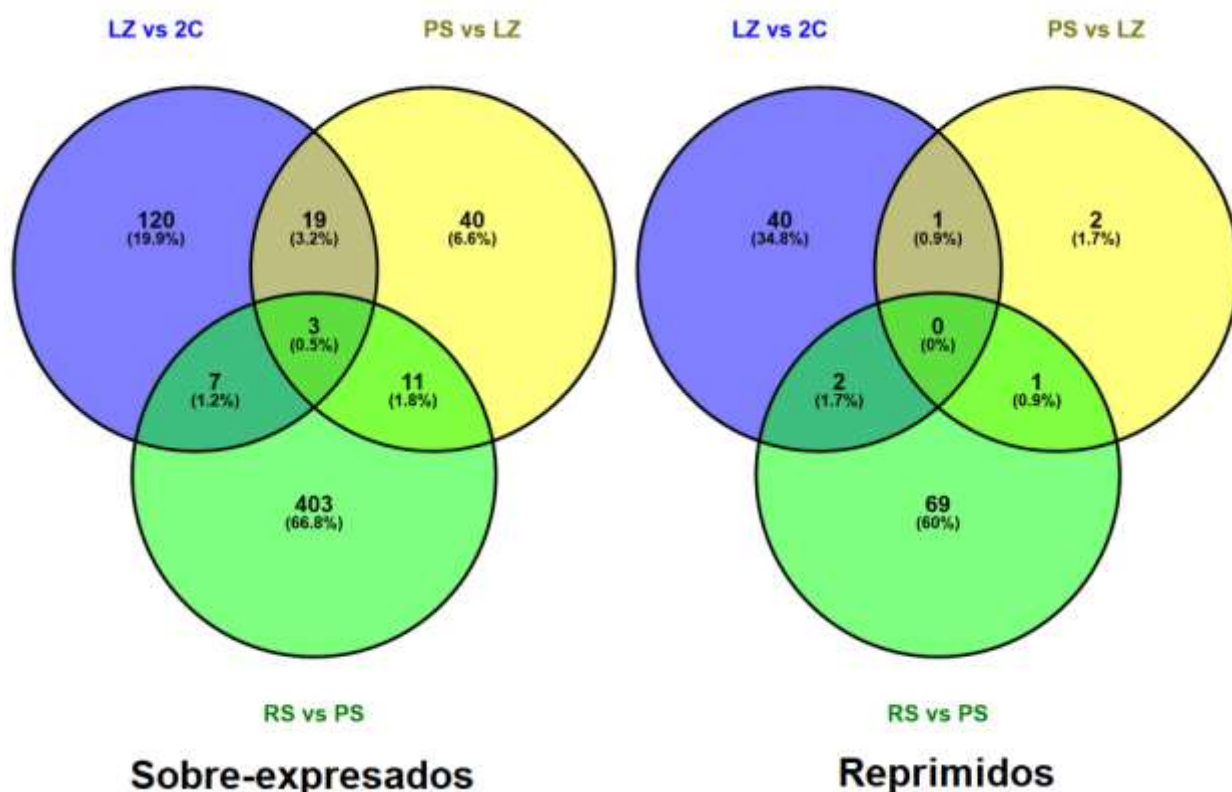
La etapa que presentó mayor cantidad de genes de lncRNAs expresados fue la correspondiente a la población de células post-meióticas: se identificaron 284 expresados en 2C, 595 en LZ, 528 en PS y 863 en RS (Figura 1B del artículo). Sucedió lo mismo con los lncRNAs diferencialmente expresados (DE): se encontraron 96 genes de lncRNAs DE entre 2C y LZ, 58 entre LZ y PS, y 411 entre PS y RS, tomando como criterio un FDR  $p\text{-valor} \leq 0,05$  y  $\log_2(\text{FC}) \geq |2|$  (Figura 2 del artículo). Cuando el umbral de corte es llevado a  $\log_2 \text{FC} \geq |1|$

detectamos 192 genes de lncRNAs DE entre 2C y LZ, 77 entre LZ y PS, y 496 entre PS y RS (Figura 15 y Tabla en Anexo 2 de la Tesis).



**Figura 14.** Gráfico de análisis de componentes principales (PCA) para las 12 muestras.

El número de genes de lncRNAs sobreexpresados fue mayor que el de subexpresados, para todas las transiciones (Figuras 2A y 2B del artículo). Los diez lncRNAs más diferencialmente sobreexpresados y reprimidos en cada transición de etapas se presentan en la Tabla 4 (excepto para los lncRNAs reprimidos en la transición de LZ a PS, en que el número observado fue menor de 10). No está de más aclarar que el corte para la expresión diferencial está arbitrariamente marcado en  $\log_2(FC) \geq |2|$ , lo que equivale a una diferencia en el FC de  $|4|$ . La mayoría son transcriptos para los cuales no se han atribuido funciones hasta el momento.



**Figura 15.** Diagramas de Venn de los genes de lncRNAs DE (izquierda: sobre-expresados; derecha: reprimidos) surgidos de la intersección de ambos abordajes, cuando usamos como criterios de corte estadístico FDR  $p\text{-valor} \leq 0,05$  y de expresión  $\log_2 FC \geq |1|$ .

**Tabla 4.** Top 10 lncRNAs diferencialmente sobre-expresados o reprimidos (mayor FC) en las distintas transiciones de etapa (FDR  $p\text{-valor} \leq 0,05$  y  $\log_2 (FC) \geq |2|$ ).

| LncRNA               | ID                  | Cromosoma | Biotipo         | FDRp-val | $\log_2 FC$ |
|----------------------|---------------------|-----------|-----------------|----------|-------------|
| <b>LZ vs 2C</b>      |                     |           |                 |          |             |
| <i>Scml1</i>         | ENSMUSG00000086540  | X         | lincRNA         | 0,003    | 4,951       |
| <i>1700001L05Rik</i> | ENSMUSG00000075511  | 15        | lincRNA         | 0,003    | 4,764       |
| <i>AC125141.3</i>    | ENSMUSG000000116879 | 17        | lincRNA         | 0,005    | 4,739       |
| <i>Gm11837</i>       | ENSMUSG00000086587  | 4         | Antisentido     | 0,003    | 4,638       |
| <i>1700008K24Rik</i> | ENSMUSG000000101012 | 17        | lincRNA         | 0,012    | 4,185       |
| <i>Gm17232</i>       | ENSMUSG00000090374  | 14        | Antisentido     | 0,002    | 4,045       |
| <i>Gm20621</i>       | ENSMUSG00000093394  | 15        | sentido exónico | 0,007    | 3,814       |
| <i>Gm14244</i>       | ENSMUSG00000087162  | 2         | Antisentido     | 0,041    | 3,811       |
| <i>9330102E08Rik</i> | ENSMUSG000000107879 | 6         | lincRNA         | 0,004    | 3,795       |

|                      |                    |    |                   |       |        |
|----------------------|--------------------|----|-------------------|-------|--------|
| <i>Gm2762</i>        | ENSMUSG00000107008 | 13 | lincRNA           | 0,009 | 3,620  |
| <i>Gm28119</i>       | ENSMUSG00000100158 | 9  | sentido intrónico | 0,022 | -3,286 |
| <i>Malat1</i>        | ENSMUSG00000092341 | 19 | lincRNA           | 0,002 | -3,295 |
| <i>Gm10399</i>       | ENSMUSG00000072688 | 5  | Antisentido       | 0,006 | -3,312 |
| <i>Gm16008</i>       | ENSMUSG00000087381 | 4  | Antisentido       | 0,011 | -3,323 |
| <i>Gm2044</i>        | ENSMUSG00000085354 | 7  | lincRNA           | 0,010 | -3,433 |
| <i>H19</i>           | ENSMUSG00000000031 | 7  | lincRNA           | 0,002 | -3,818 |
| <i>Gm12963</i>       | ENSMUSG00000085517 | 4  | Antisentido       | 0,002 | -4,000 |
| <i>A530020G20Rik</i> | ENSMUSG00000097124 | 3  | Antisentido       | 0,004 | -4,432 |
| <i>Snhg18</i>        | ENSMUSG00000096956 | 15 | lincRNA           | 0,002 | -4,730 |
| <i>Meg3</i>          | ENSMUSG00000021268 | 12 | lincRNA           | 0,002 | -4,964 |
| <b>PS vs LZ</b>      |                    |    |                   |       |        |
| <i>1700026J12Rik</i> | ENSMUSG00000073976 | 16 | lincRNA           | 0,036 | 5,027  |
| <i>Gm43579</i>       | ENSMUSG00000107191 | 5  | Antisentido       | 0,024 | 4,776  |
| <i>Proscos</i>       | ENSMUSG00000091154 | 8  | Antisentido       | 0,041 | 4,440  |
| <i>1700010K23Rik</i> | ENSMUSG00000092599 | 16 | Antisentido       | 0,048 | 4,286  |
| <i>4930488L21Rik</i> | ENSMUSG00000097520 | 8  | Antisentido       | 0,026 | 4,108  |
| <i>1700021F13Rik</i> | ENSMUSG00000107053 | 5  | lincRNA           | 0,044 | 4,070  |
| <i>1700018L02Rik</i> | ENSMUSG00000100075 | 19 | Antisentido       | 0,026 | 4,051  |
| <i>AC160993.1</i>    | ENSMUSG00000116576 | 16 | lincRNA           | 0,017 | 3,991  |
| <i>Gm10339</i>       | ENSMUSG00000071591 | 14 | lincRNA           | 0,033 | 3,990  |
| <i>Gm43352</i>       | ENSMUSG00000106086 | 3  | Antisentido       | 0,028 | 3,979  |
| <i>Gm17552</i>       | ENSMUSG00000097233 | 19 | lincRNA           | 0,027 | -2,240 |
| <i>Gm15265</i>       | ENSMUSG00000086288 | 3  | Antisentido       | 0,031 | -2,565 |
| <i>Gm13375</i>       | ENSMUSG00000075514 | 2  | Antisentido       | 0,031 | -2,851 |
| <i>Scml1</i>         | ENSMUSG00000086540 | X  | lincRNA           | 0,045 | -4,105 |
| <b>RS vs PS</b>      |                    |    |                   |       |        |
| <i>1700041C23Rik</i> | ENSMUSG00000100282 | 9  | lincRNA           | 0,003 | 7,857  |
| <i>1700061N14Rik</i> | ENSMUSG00000110464 | 8  | lincRNA           | 0,007 | 7,591  |
| <i>BC061195</i>      | ENSMUSG00000096153 | X  | lincRNA           | 0,001 | 7,443  |
| <i>1700003E24Rik</i> | ENSMUSG00000095110 | X  | lincRNA           | 0,003 | 7,434  |
| <i>Gm28364</i>       | ENSMUSG00000101503 | 1  | Antisentido       | 0,007 | 7,204  |
| <i>Smkr-ps</i>       | ENSMUSG00000045709 | 6  | lincRNA           | 0,002 | 7,041  |
| <i>1700052I22Rik</i> | ENSMUSG00000101912 | 12 | lincRNA           | 0,003 | 6,989  |
| <i>1700031L13Rik</i> | ENSMUSG00000099863 | 5  | lincRNA           | 0,001 | 6,854  |

|                      |                    |    |             |       |        |
|----------------------|--------------------|----|-------------|-------|--------|
| <i>Gm44145</i>       | ENSMUSG00000107453 | 6  | Antisentido | 0,005 | 6,821  |
| <i>1700095J12Rik</i> | ENSMUSG00000084804 | 4  | Antisentido | 0,002 | 6,749  |
| <i>Gm10619</i>       | ENSMUSG00000074067 | 7  | lincRNA     | 0,044 | -4,394 |
| <i>Gm14244</i>       | ENSMUSG00000087162 | 2  | Antisentido | 0,004 | -5,518 |
| <i>Gm21269</i>       | ENSMUSG00000108448 | 7  | lincRNA     | 0,005 | -5,641 |
| <i>Gm21221</i>       | ENSMUSG00000106740 | 5  | lincRNA     | 0,001 | -5,807 |
| <i>4933435G04Rik</i> | ENSMUSG00000108345 | 7  | lincRNA     | 0,001 | -5,948 |
| <i>AC125141.2</i>    | ENSMUSG00000116751 | 17 | lincRNA     | 0,002 | -5,996 |
| <i>Gm26917</i>       | ENSMUSG00000097971 | 17 | lincRNA     | 0,024 | -6,504 |
| <i>AC125141.3</i>    | ENSMUSG00000116879 | 17 | lincRNA     | 0,004 | -6,570 |
| <i>Gm42418</i>       | ENSMUSG00000098178 | 17 | lincRNA     | 0,003 | -6,667 |
| <i>1700023F02Rik</i> | ENSMUSG00000100000 | 10 | lincRNA     | 0,018 | -6,988 |

---

Como hemos mencionado previamente, las células de la fase de espermiogénesis (RS) se caracterizan por sufrir cambios drásticos como la adquisición del flagelo, formación del acrosoma, reorganización mitocondrial, elongación nuclear y, a nivel de la cromatina, altísima compactación debido al reemplazo de histonas por protaminas, llevando al silenciamiento transcripcional [Kleene, 2001]. En todos estos procesos podrían participar los lncRNAs, que son especialmente abundantes en el testículo [Soumillon *et al.*, 2013]. En otros trabajos también se ha sugerido que las células post-meióticas son las que expresan mayor número de lncRNAs en el ratón y otras especies [Laiho *et al.*, 2013; Soumillon *et al.*, 2013; Rolland *et al.*, 2019].

#### 4.3.3: Comparación de los resultados con los de otros trabajos

Comparamos nuestros resultados con los trabajos de otros autores, ya que un modo habitual de validación de resultados es la comparación con resultados obtenidos en trabajos previos. No obstante, hay pocos estudios reportados de identificación de lncRNAs en el testículo del ratón, y los mismos utilizaron testículo completo y no poblaciones celulares aisladas [Laiho *et al.*, 2013; Hong *et al.*, 2018] o emplearon microarreglos [Bao *et al.*, 2013; Sun *et al.*, 2013; Liang *et al.*, 2014]. En un trabajo se purificaron células de estadios específicos de la

espermatogénesis para secuenciación de célula única, pero se basó en la sincronización artificial de la espermatogénesis mediante una combinación del compuesto WIN 18,446 que inhibe a la enzima retinaldehído deshidrogenasa, y ácido retinoico [Chen *et al.*, 2018]. Por lo tanto, no se estudió la expresión génica durante la progresión natural de la espermatogénesis. Finalmente, escogimos para comparar con nuestros datos, los trabajos de Lin *et al.* [2016] y Wichman *et al.* [2017], únicos dos trabajos en los que se identificaron lncRNAs de distintas etapas de la espermatogénesis del ratón, empleando poblaciones celulares enriquecidas. Como dato importante, en ambos trabajos se partió de muestras celulares enriquecidas por *STA-PUT* (es decir, menor pureza que las nuestras), y ninguno de los dos contó con una población LZ. Además, en el trabajo de Lin *et al.* [2016] se trabajó con 2 réplicas de cada población y no 3. Esto trajo algunas limitaciones a la hora de comparar los datos (ver Materiales y Métodos).

Cuando comparamos nuestros resultados con los datos del trabajo de Lin y colaboradores, re-analizados con el paquete comercial CLC, encontramos una correlación de Pearson de  $r = 0,38$  a  $0,47$  con un valor estadístico  $p\text{-valor} < 10^{-5}$ . Asimismo, cuando los comparamos con el re-análisis de los datos de Wichman *et al.*, la correlación fue de  $r = 0,49$  a  $0,67$  con  $p\text{-valor} < 10^{-5}$  (Figura Suplementaria S3 del artículo). Esto indicaría que, a pesar de las diferencias metodológicas en la obtención de las muestras y el número de réplicas, los resultados globales correlacionaron significativamente con nuestros resultados.

Por otra parte, Lagarde *et al.* [2017] desarrollaron una metodología para la anotación de lncRNAs de tamaño completo mediante la secuenciación de fragmentos largos (CLS). Si bien utilizaron oligodTs para capturar el ARN blanco, por lo que su análisis se centra principalmente en lncRNAs poliadenilados como los lincRNAs, nos resultó de interés re-analizar nuestros datos con la anotación generada en dicho trabajo. El análisis de nuestros datos de secuenciación con el abordaje HSB cuando variamos el genoma de referencia de *Ensembl* por la anotación de Lagarde *et al.*, específicamente para lincRNAs, detectó como expresados 1.020 lncRNAs en 2C, 1.497 en LZ, 1.392 en PS y 2.036 en RS, coincidiendo entre 34 y 36% con nuestros datos utilizando el genoma *Ensembl* (correlación de  $r = 0,68$  a  $0,83$ ,  $p < 10^{-5}$ ) (Figura Suplementaria S6 y Tabla Suplementaria S3 del artículo). Si bien es cierto que con esta anotación de CLS detectamos mayor número de lncRNAs expresados, las conclusiones a las que

arribamos coinciden, ya que sigue siendo RS la fracción con mayor expresión de lncRNAs, seguido de LZ y PS, y por último la pre-meiótica. Además, si consideramos que la coincidencia entre la anotación de CLS y *Gencode* (misma anotación que *Ensembl*) es muy baja (20%), nuestros niveles de coincidencia pueden considerarse como muy buenos. Este resultado también refuerza la solidez de nuestros datos, ya que no dependen en gran medida de la anotación usada. Por otro lado, como era de esperar, los transcritos identificados con la anotación CLS, en general, fueron más largos que con la anotación de *Ensembl* (1.032 y 808 pb promedio, respectivamente).

#### **4.3.4: Caracterización global del transcriptoma de la espermatogénesis del ratón**

Respecto a la caracterización de los lncRNAs expresados y DE en las distintas poblaciones celulares, más del 50% de los transcritos de lncRNAs presentó menos de 1.000 nucleótidos, y alrededor de un 34% entre 1.000 y 3.000. A su vez, la amplia mayoría de los lncRNAs expresados contenía menos de 5 exones y solamente una pequeña proporción que no supera en ningún caso el 1,5%, tenía más de 10 exones (Figuras 3B y 3C del artículo). Esto fue cierto para todos los tipos celulares, y se cumplió también para el subconjunto de los lncRNAs DE (Figuras Suplementarias S5a y S5b del artículo). *A priori*, no habría por qué esperar sesgos en estas características entre los diferentes tipos celulares testiculares. Sin embargo, Chalmel *et al.* [2014] reportaron que en la rata los lncRNAs con pico de expresión en meiosis eran particularmente más largos. Nuestros resultados no apoyan dicha conclusión, al menos para el ratón. Por otra parte, nuestros resultados muestran que, en general, los lncRNAs del testículo son más cortos y poseen menor número de exones que los mRNAs, lo que estaría de acuerdo con lo que se ha propuesto para lncRNAs en general [Cabili *et al.*, 2011; Derrien *et al.*, 2012] y coincide para lo reportado para testículo total en cerdo [Ran *et al.*, 2016] y gallo [Liu *et al.*, 2017 (a)].

Los lincRNAs fueron el biotipo expresado predominantemente en todas las poblaciones celulares estudiadas, seguidos de los lncRNAs antisentido, los promotores bidireccionales, y en mucho menor medida lncRNAs sentido exónicos e intrónicos. Cabe mencionar que no detectamos ningún lncRNA macro ni solapante en 3' (Figuras 3D del artículo). Esto se cumplió también para el

subconjunto de los lncRNAs DE (Figura Suplementaria S5c del artículo), y no resulta sorprendente ya que los dos biotipos mayoritarios se corresponden con la mayor parte de los lncRNAs anotados en el genoma del ratón. Estos resultados coinciden con lo reportado por Bao *et al.* [2013] también en ratón, y con lo observado para gallo [Liu *et al.* 2017 (a)]. Sin embargo, la proporción de lncRNAs antisentido detectada en nuestras listas es llamativamente alta (este punto se discutirá en la sección 5.5 de la Discusión).

En relación a la distribución cromosómica, observamos una importante depleción en la expresión de lncRNAs del cromosoma Y (test hipergeométrico  $p < 10^{-9}$ ) en nuestras listas, a pesar de que este cromosoma, más allá de su tamaño comparativamente pequeño, posee gran cantidad de lncRNAs anotados. Ello fue válido para los lncRNAs de las distintas etapas de la espermatogénesis estudiadas (Figura 3A del artículo). Esto es algo que particularmente llamó nuestra atención, pero hasta el momento no hemos podido encontrarle una explicación, aunque coincide con el resultado observado para otra especie (ver 5.Discusión). En cuanto al cromosoma X, si bien vemos un cambio en la expresión de los lncRNAs, que en líneas generales disminuye en PS y vuelve a aumentar en RS, no podemos sacar conclusiones definitivas dados los bajos números de lncRNAs detectados expresándose desde este cromosoma (Figura Suplementaria S4 del artículo). De igual manera, este puñado de lncRNAs parecería no escapar a la MSCI. No todos los genes del cromosoma X son inactivados durante la MSCI; por ejemplo, los genes de la región pseudoautosómica (PAR) donde los cromosomas X e Y son homólogos y que es responsable del apareamiento X-Y durante la meiosis, así como otros genes individuales del X, escapan a dicha inactivación [Pontier & Gribnau, 2011]. En algunos trabajos se ha visto que existen clústeres de algunos genes de mRNAs y lncRNAs del X que escaparían a la inactivación durante paquiteno [da Cruz *et al.*, 2016; Wichman *et al.*, 2017]. Se ha descrito también un número importante de miRNAs del cromosoma X localizados en la periferia del cuerpo XY con niveles constantes y transcripción continua durante la meiosis, sugiriendo funciones relevantes en el proceso [Sosa *et al.*, 2015]. Es decir que, de los lncRNAs del cromosoma X que aparecen en nuestras listas, los mismos parecieran comportarse más parecido a los mRNAs en su expresión, que en su mayoría no escapan a la MSCI, y no tanto a los muchos miRNAs que sí lo hacen.

#### 4.3.5: Co-expresión de lncRNAs con genes codificantes

Para los dos biotipos más expresados en nuestras listas, lincRNAs y lncRNAs antisentido, decidimos evaluar si podíamos establecer una posible correlación en su comportamiento a lo largo de la espermatogénesis, con la de genes codificantes vecinos o solapantes, respectivamente. Examinando el contexto genómico de ARNs no codificantes relativo a genes codificantes de función conocida, y en conjunto con datos de expresión, es posible predecir roles para el par de transcriptos codificante y no codificante [Dinger *et al.*, 2008]. Sorprendentemente, para aproximadamente el 95% de los pares de lncRNAs antisentido y sus genes codificantes solapantes encontramos una correlación en la expresión (Figura 4A del artículo). Más aún, 81,5% de los pares exhibían un comportamiento similar a lo largo de las transiciones entre etapas de la espermatogénesis estudiadas. Dentro de este grupo, en el 88,75% (72% del total) ambos estaban sobreexpresados (“up-up”;  $r = 0,81$ ,  $p < 10^{-5}$ ), mientras que en 18,5% (alrededor del 7% del total) ambos estaban reprimidos (“down-down”;  $r = 0,54$ ,  $p < 0,05$ ). Por otra parte, para el 13% restante de los pares correlacionados se estableció una correlación inversa ( $r = -0,70$ ,  $p < 10^{-4}$ ), es decir, que su comportamiento era opuesto a lo largo de las distintas transiciones de etapas (cuando uno aumenta su expresión el otro disminuye, y viceversa; ver Figura 4A del artículo). Sólo para un 5% de los pares gen codificante/lncRNA antisentido solapante, no se pudo establecer una correlación. Adicionalmente, para al menos el 61% de los genes codificantes solapantes y co-expresados con lncRNAs antisentido DE en nuestras listas, se encontró que se les ha asignado una función testículo-específica, o exhiben un patrón de expresión restringido al testículo (Tabla Suplementaria S4 del artículo). Una posible explicación es que estos lncRNAs antisentido tengan funciones regulatorias sobre sus genes solapantes, o bien alternativamente que se expresen simplemente debido al estado descompactado de la cromatina en esa región, que lo favorezca. Esto se discutirá detalladamente en la sección 5.5 de la Discusión.

En lo que respecta al análisis de co-expresión con los lincRNAs, seleccionamos los genes codificantes vecinos a diferentes distancias basándonos en otros artículos:  $< 300$  Kb [Orom *et al.*, 2010],  $< 100$  Kb [Wang *et al.*, 2016; Liu *et al.*, 2017 (a)],  $< 30$  Kb [Bao *et al.*, 2013] y  $< 10$  Kb [Hong *et al.*,

2018], y decidimos estudiar cómo la correlación variaba a medida que cambiaba la distancia, dado que no existían precedentes de un estudio semejante. Los pares lincRNA/gen codificante también debían cumplir con la condición de que ambos genes del par estuvieran expresados en nuestras listas y uno de ellos fuera DE en alguna etapa. Considerando los lincRNAs y sus genes vecinos codificantes situados a menos de 300 Kb de distancia, pudimos comprobar que para el 57% de los pares hay una correlación evidente en su expresión (Figura 4B del artículo). Este porcentaje aumenta conforme la distancia entre el par de genes disminuye (65% para lincRNA/gen codificante localizados a  $< 100$  Kb). Dentro de este porcentaje de pares que pueden ser correlacionados, encontramos que el 32% de los mismos situados a  $< 300$  Kb son “up-up” ( $r = 0,78$ ,  $p < 10^{-5}$ ), 42% a  $< 100$  Kb ( $r = 0,75$ ,  $p < 10^{-5}$ ), y 53% a  $< 30$  Kb ( $r = 0,85$ ,  $p < 10^{-5}$ ). Contrariamente, el número de pares de genes (lincRNA/gen codificante) cuyo patrón de expresión no pudo ser correlacionado (es decir, que no exhiben un comportamiento coherente entre ambos a lo largo de las diferentes transiciones de etapas), fue mayor cuanto más grande era la distancia entre ambos (43% para pares situados a  $< 300$  Kb, 35% para aquellos a  $< 100$  Kb; ver Figura 4B del artículo). Estos resultados señalarían que la correlación de los pares co-expresados de lincRNAs y sus genes codificantes vecinos es mayor a medida que la distancia entre ellos se reduce. De todas formas, el porcentaje de los pares de genes con patrón de expresión similar fue mucho menor en este caso, que para los lncRNAs antisentido y sus genes codificantes solapantes.

En relación con el punto anterior, si la distribución de lincRNAs a lo largo del genoma fuera al azar y homogénea, a medida que reducimos la distancia al tercio (300, 100, 30 y 10 Kb), la cantidad de lincRNAs alrededor de un gen codificante cualquiera debería reducirse en la misma proporción. Sin embargo, la cantidad de lincRNAs expresados observados es superior a la esperada en todos los casos (Figura 4C del artículo). Esto sugiere que los lincRNAs se encontrarían más concentrados en las cercanías de genes codificantes. En el mismo sentido, en un par de trabajos anteriores [Cabili *et al.*, 2011; Ulitsky & Bartel, 2013] se encontró que lincRNAs testiculares se acumulan a distancias menores a 10 Kb de genes codificantes.

Los resultados del análisis de ontología para los genes codificantes vecinos de lincRNAs o solapantes de lncRNAs antisentido muestran un enriquecimiento

moderado ( $p < 0,05$ ) en los términos “espermatogénesis”, “motilidad espermática”, “movimiento basado en microtúbulos”, “actividad motora de microtúbulos”; y un enriquecimiento ( $p < 0,01$ ) en los términos “ensamblaje nucleosómico”, “regulación positiva de la expresión génica”, “silenciamiento génico”, “unión a ADN nucleosomal” y “vesícula acrosómica”. Concordantemente, “testículo” fue el término altamente enriquecido ( $p < 10^{-27}$ ) en la categoría de tejidos.

#### **4.3.6: Conservación de pares de genes lncRNA/gen codificante entre ratón y humano**

Como se mencionó en la Introducción, existe un consenso en el hecho de que los lncRNAs son menos conservados entre especies que los genes codificantes de proteínas. En ese sentido, nos preguntamos si los lncRNAs que están co-expresados con genes codificantes en la espermatogénesis están conservados en humanos, lo que reforzaría la idea de una posible relación funcional. Con ese objetivo, analizamos en particular si los pares de genes de lncRNA antisentido/gen codificante solapante expresados en ratón estaban también presentes en humano, utilizando dos anotaciones diferentes (*Ensembl* y *Chess*). Encontramos un lncRNA antisentido para el 46% de los genes codificantes anotado en humano en el *Ensembl*, y para el 49% anotado en *Chess*. Este porcentaje es mayor que el porcentaje de menos del 20% de coincidencia que se ha sugerido en otros trabajos para pares de genes sentido/antisentido entre humano y ratón en general [Breschi *et al.*, 2017].

Analizamos luego si los pares de genes de lincRNA/gen codificante vecino situados a  $< 30$  Kb y expresados en la espermatogénesis del ratón conservaban una posición sinténica en el genoma humano, utilizando la anotación de *Ensembl* (como se indicó en el punto 3.4.7 de Materiales y Métodos, *Chess* no se pudo usar en este caso ya que no identifica específicamente lincRNAs). En el 59% de los casos encontramos un lincRNA a menos de 30 Kb de las secuencias codificantes homólogas al gen de ratón, en humano. En términos generales, este porcentaje parece superior a los reportados en otros trabajos para lincRNAs en general (no específicamente testiculares), o para otras especies [Cabili *et al.*, 2011; Ran *et al.*, 2016].

#### 4.3.7: Validaciones por RT-qPCR

Al momento de validar los resultados mediante RT-qPCR, elegimos confirmar no solamente la correlación con los resultados del *RNAseq*, sino también la correlación entre los patrones de expresión de los pares de genes (lncRNA/gen codificante) mencionados. Con ese objetivo, se escogieron 13 pares de lncRNA/gen codificante, tanto de antisentido/solapante como de lincRNA/gen vecino, para su verificación mediante RT-qPCR. Estos pares de genes fueron seleccionados de modo que nos permitieran validar genes DE en todas las etapas (2C, LZ, PS y RS), representantes de todos los patrones de expresión (“up-up”, “down-down”, “up-down”, etc.) y buscando en lo posible que el gen codificante asociado al par tuviera alguna función conocida o sospechada en el testículo, o expresión específica en el testículo (la descripción detallada de cada gen seleccionado y su función en la espermatogénesis, puede encontrarse en el artículo anexo a continuación). *Actb*, por su parte, es el gen de la  $\beta$ -actina formadora de citoesqueleto, y este gen se seleccionó como marcador de las células no germinales. En este sentido, nos interesa señalar que el gen de  $\beta$ -actina es uno de los más utilizados como gen normalizador en experimentos de RT-qPCR. Sin embargo, nuestros resultados muestran cómo su expresión disminuye de manera abrupta en la espermatogénesis, dejando en claro que este gen no debe ser utilizado como marcador a lo largo del proceso espermatogénico. En consecuencia, realizamos un análisis exhaustivo de nuestras listas triplicadas de genes codificantes expresados en todas las condiciones, en busca de genes apropiados para ser utilizados como normalizador. Finalmente, escogimos el gen *Surf4*, que resultó ser un buen gen normalizador por su escasa variación entre las diferentes poblaciones celulares estudiadas (ver 3.5 en Materiales y Métodos). Por obvias razones, dado que la mayoría de los pares lncRNA/gen co-expresados muestran un patrón “up-up” en la etapa RS, gran parte de los pares seleccionados son de este tipo. Estos análisis dieron resultados altamente consistentes, y nos permitieron corroborar todas las condiciones estudiadas, así como la co-expresión entre los genes codificantes y sus lncRNAs solapantes o cercanos (Figura 5 del artículo).

RESEARCH PAPER



# Revealing stage-specific expression patterns of long noncoding RNAs along mouse spermatogenesis

María F. Trovero <sup>a</sup>, Rosana Rodríguez-Casuriaga <sup>a,b</sup>, Carlos Romeo <sup>c</sup>, Federico F. Santiñaque<sup>d</sup>, Mateo François<sup>a</sup>, Gustavo A. Folle <sup>d,e</sup>, Ricardo Benavente <sup>f</sup>, José R. Sotelo-Silveira <sup>c,g</sup>, and Adriana Geisinger <sup>a,b</sup>

<sup>a</sup>Department of Molecular Biology, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Montevideo, Uruguay; <sup>b</sup>Biochemistry-Molecular Biology, Facultad de Ciencias, Universidad de la República (UdelAR), Montevideo, Uruguay; <sup>c</sup>Department of Genomics, IIBCE, Montevideo, Uruguay; <sup>d</sup>Flow Cytometry and Cell Sorting Core, IIBCE, Montevideo, Uruguay; <sup>e</sup>Department of Genetics, IIBCE, Montevideo, Uruguay; <sup>f</sup>Department of Cell and Developmental Biology, Biocenter, University of Würzburg, Würzburg, Germany; <sup>g</sup>Department of Cell and Molecular Biology, Facultad de Ciencias, Universidad de la República (UdelAR), Montevideo, Uruguay

## ABSTRACT

The discovery of a large number of long noncoding RNAs (lncRNAs), and the finding that they may play key roles in different biological processes, have started to provide a new perspective in the understanding of gene regulation. It has been shown that the testes express the highest amount of lncRNAs among different vertebrate tissues. However, although some studies have addressed the characterization of lncRNAs along spermatogenesis, an exhaustive analysis of the differential expression of lncRNAs at its different stages is still lacking. Here, we present the results for lncRNA transcriptome profiling along mouse spermatogenesis, employing highly pure flow sorted spermatogenic stage-specific cell populations, strand-specific RNAseq, and a combination of up-to-date bioinformatic pipelines for analysis. We found that the vast majority of testicular lncRNA genes are expressed at post-meiotic stages (i.e. spermiogenesis), which are characterized by extensive post-transcriptional regulation. lncRNAs at different spermatogenic stages shared common traits in terms of transcript length, exon number, and biotypes. Most lncRNAs were lincRNAs, followed by a high representation of antisense (AS) lncRNAs. Co-expression analyses showed a high correlation along the different spermatogenic stage transitions between the expression patterns of AS lncRNAs and their overlapping protein-coding genes, raising possible clues about lncRNA-related regulatory mechanisms. Interestingly, we observed the co-localization of an AS lncRNA and its host sense mRNA in the chromatoid body, a round spermatid-specific organelle that has been proposed as a reservoir of RNA-related regulatory machinery. An additional, intriguing observation is the almost complete lack of detectable expression for Y-linked testicular lncRNAs, despite that a high number of lncRNA genes are annotated for this chromosome.

## ARTICLE HISTORY

Received 14 June 2019  
Revised 28 November 2019  
Accepted 29 November 2019

## KEYWORDS

lncRNAs; spermatogenesis; testis; RNAseq; transcriptome

## Introduction

Mammalian testes are very complex organs, composed of over 30 different cell types. As a consequence of this complexity, gene expression studies have required the development of methods for spermatogenic-specific cell types isolation such as unit gravity sedimentation (Staput) and centrifugal elutriation, that allow an enrichment in certain cell types [1,2]. Later on, FACS (Fluorescence-Activated Cell Sorting) appeared as the dominant technology as it enables the obtaining of highly pure male germline-specific cell populations [3,4].

Spermatogenesis is an intricate differentiation process that can be divided into three main consecutive yet overlapping phases (Supplementary Fig. S1A): the mitotic proliferation of spermatogonia (the precursors of meiotic cells), meiotic divisions, and spermiogenesis. Meiosis is characterized by significant unique events such as the alignment and pairing of homologous chromosomes that take place during early meiotic prophase I (i.e. leptotene and zygotene stages), and recombination (crossing-over) at the following prophase stage (i.e.

pachytene) [5]. On the other hand, during spermiogenesis round spermatids (the cells that result from the meiotic divisions) go through several differentiation stages until becoming sperm, and this process is accompanied by dramatic changes: acquisition of a flagellum, loss of most cytoplasm, nuclear elongation, acrosome formation, reorganization of mitochondria, and chromatin compaction. In particular, chromatin compaction is achieved through the replacement of most histones first by transition proteins and then by protamines, and this replacement results in transcriptional silencing at the later spermiogenesis stages [6].

Due to transcriptional silencing, spermatogenic cells at earlier stages have developed a panoply of mechanisms for post-transcriptional regulation and translational delay, as a strategy to regulate the time of synthesis for proteins that are required later on by elongated spermatids and sperm [6–8]. The involved regulatory mechanisms may be as diverse as mRNA sequestration as free ribonucleoprotein particles, binding of repressors to UTRs of testis-specific transcripts, regulation of

**CONTACT** Adriana Geisinger  [adriana.geisinger@gmail.com](mailto:adriana.geisinger@gmail.com)  Department of Molecular Biology, Instituto de Investigaciones Biológicas Clemente Estable (IIBCE), Montevideo 11,600, Uruguay; José R. Sotelo-Silveira  [sotelosj@gmail.com](mailto:sotelosj@gmail.com)  Department of Genomics, IIBCE, Montevideo 11,600, Uruguay  
 Supplemental data for this article can be accessed [here](#).

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

poly(A) tails length, and others [6]. Another such strategies may be the retention of RNAs in the chromatoid body, a membrane-less spermatid-specific perinuclear granule that contains lncRNAs, piRNAs, mRNAs, and proteins for mRNA processing, and whose role is probably linked to the control of RNA-related processes [9]. In addition, translational control mechanisms may involve noncoding RNAs. In fact, a distinctive feature of testis is the extensive transcription of noncoding RNAs, among which long noncoding RNAs (lncRNAs) stand out [10].

lncRNAs have been defined as RNAs longer than 200 nucleotides (as opposed to small noncoding RNAs) that are not translated into proteins [11,12]. However, as the definition is arbitrary, some lncRNAs may be smaller than 200 nt [13], and, moreover, a minority of them might actually encode short functional peptides [14]. In general, lncRNAs share some traits among vertebrates, such as the fact that they are mostly transcribed by RNA polymerase II, are often alternatively spliced, and many of them are polyadenylated and capped [15–18]. Besides, some of their characteristics comprise low abundance, low sequence conservation, and highly tissue- and developmental-specific expression patterns [17–21]. According to their location, they have been classified as overlapping protein-coding genes in sense or antisense direction, corresponding to bidirectional promoter transcripts, intronic, intergenic (lincRNAs), associated with enhancers or repetitive regions, or 3'-overlapping [11,17,22]. Some lncRNA genes may host miRNAs [22]. Besides, some long, unspliced macro lncRNAs have been described [23].

The high representation of lncRNAs in mammalian genomes, together with their restricted expression patterns, suggests that at least part of them should be functional [24,25]. In fact, roles have been elucidated for different lncRNAs in relation to transcriptional activation or repression; chromatin remodelling; regulation of mRNAs splicing, editing and degradation; modulation of the activity or abundance of proteins or RNAs; regulation of miRNA activity; and other functions [11,22,26,27]. So far, lncRNAs have been implicated in processes as diverse as the maintenance of pluripotency, embryonic development and organ differentiation, imprinting, organization of cell architecture, and apoptosis, among others. Moreover, several lncRNAs have been related to the development of different diseases, including cancer [20,26–28].

Different reports coincide with the fact that the testes express the largest numbers of lncRNAs among different tissues [10,17,20,21,29]. A number of studies have addressed the characterization of testicular lncRNAs in different mammalian species, disclosing that they tend to share the general traits described for lncRNAs [30–32], and revealing that at least some testicular lncRNAs are functional during spermatogenesis [33,34]. A handful of papers addressed differential lncRNAs expression along mouse spermatogenesis, using microarrays [35,36] or RNAseq [10,37–39]. These studies employed whole testes of animals of increasing ages [35,37], or spermatogenic stage-specific cell populations isolated by Staptut or elutriation [10,36,38,39] that only allow the obtainment of enriched (but not pure) cell populations [4]. In some of these studies, lncRNAs were not the main research objective but they arose as a by-product, and therefore their

characterization was not exhaustive [37]. Additionally, concerning RNAseq studies, data analysis has recently incorporated more accurate tools for the study of lncRNAs [40] than the ones used in most previous papers. An interesting recent paper reported single-cell RNA-sequencing of 20 different spermatogenic cell subtypes [41]. While this report represents an important advance as it included some previously unpurified cell types, this was achieved by manipulating the spermatogenic process through combining transgenic labelling with synchronization of the cycle of the seminiferous epithelium by means of WIN 18,446/retinoic acid. Besides, lncRNAs were not extensively characterized.

In the past, we have reported the collection of highly pure stage-specific spermatogenic cell populations from mouse by flow cytometry sorting [42,43]. We used these cell populations for transcriptome analyses, establishing stage-specific gene expression signatures along spermatogenesis with an unprecedented reliability in the profiling [4]. However, the employed RNAseq libraries were not strand-specific, thus limiting their usefulness to characterize the whole repertoire of lncRNAs as a high proportion of them are antisense to coding transcripts and therefore require strand-specific libraries for their reliable detection [44].

In the present work, we aimed to characterize the lncRNAs expressed in highly pure stage-specific testicular cell populations obtained with our tailored flow cytometry purification protocols, using deep sequencing of strand-specific RNAseq libraries and current bioinformatics tools. We hereby present a comprehensive analysis of the differential expression of lncRNAs at distinct stages of mouse spermatogenesis. The obtained data provide a highly reliable information set available to the community for future studies in the field.

## Results

To address the characterization of differential gene expression of lncRNAs throughout spermatogenesis, we performed transcriptome analysis of strand-specific RNAseq libraries generated from highly purified testicular cell populations by flow sorting (Supplementary Fig. S1B). The cell populations, representative of landmark stages along mouse spermatogenesis, were the same as in a previous report from our group, in which the coding transcriptome was disclosed [4]: 2C (a heterogeneous population with 2C DNA content, consisting of spermatogonia and testicular somatic cells); LZ (leptotyzotene spermatocytes); PS (pachytene spermatocytes); and RS (round spermatids).

Sequence analysis of the different libraries yielded on average 150 million reads per sample. Between 84.21% and 94.15% of the reads conserved after trimming aligned with the reference genome (Supplementary Table S1). Up to 0.22% reads mapped on ribosomal RNA and 3.28% on mitochondrial RNA, indicating that the depletion step was successful. Principal component analysis and correlation matrix showed high reproducibility between biological replicas (Supplementary Fig. S2), as well as with our previous RNAseq study (data not shown). Overall, the obtained data is a very deep set of reads with robust reproducibility, useful to characterize even lowly expressed lncRNAs.

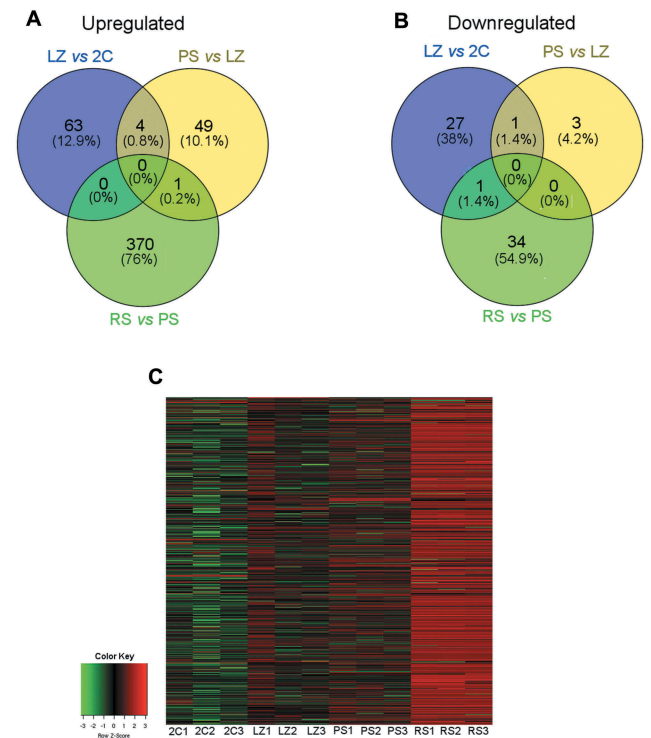
**Round spermatids contain the highest number of lncRNAs among different testicular cell populations**

The obtained data was analysed in parallel with two pipelines, CLC Genomics Workbench and Hisat/StringTie/Balgon. We only considered mappings onto previously annotated lncRNAs that arose from the intersection of both pipelines, had a variance across samples of more than 1, and presented more than 0.1 FPKM. Although this may turn out to be a strict criterion excluding a high number of reads, it yields very reliable results. Notably, 83.5% of the lncRNA genes detected as expressed with CLC, coincided with those detected with Hisat/StringTie/Balgon (Fig. 1A). Our stringent approach detected the presence of transcripts corresponding to 878 annotated lncRNA genes in the full dataset that includes the four studied cell populations (See Fig. 1A, and Supplementary Table S2). This represents 10% of all annotated lncRNA genes in Ensembl database (Gm38.p6 release 93). Among them, 28.4% (249 lncRNAs) were common to all the analysed cell populations. Two hundred and eighty-four lncRNAs were present in the 2C cell population, 595 in LZ, 528 in PS, and 863 in RS (Fig. 1B).

Concerning differentially expressed (DE) lncRNA genes, pairwise comparisons between cell populations in chronological order along the progression of the spermatogenic wave rendered 96 annotated DE lncRNA genes in LZ vs 2C population, 58 in PS vs LZ, and 411 in RS vs PS (Fig. 2, and Supplementary Table S2), considering a  $\log_2$  fold-change  $\geq |2|$  cut-off and an FDR p-value  $\leq 0.05$  as significant. Importantly, these numbers did not change when the expression cut-off was raised from 0.1 FPKM to 1 FPKM. Thus, the highest number of testicular lncRNA genes is expressed in spermiogenesis, and much more so with regards to the DE ones.

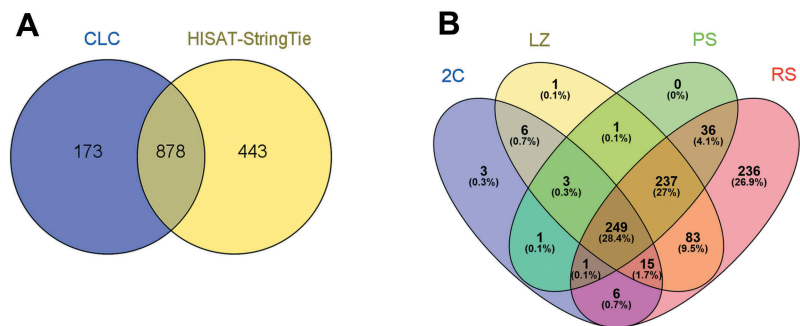
Among the DE lncRNAs, the number of upregulated genes (Fig. 2A) was notoriously higher than that of downregulated genes (Fig. 2B), for all the pairwise comparisons. Despite this, RS exhibited the highest number of upregulated and downregulated DE lncRNA genes as clearly depicted in the heat-maps, where the highest number of DE lncRNAs in RS is evident (Fig. 2C).

Then, we set out to contrast our data with those from other studies. In order to use the most comparable available datasets, we chose the results from Lin *et al.* [38] and Wichman *et al.* [39],



**Figure 2.** Representation of the DE lncRNA genes between pairwise sample comparisons of the four testicular cell populations in chronological order along spermatogenesis. The following comparisons were performed: LZ vs 2C; PS vs LZ; RS vs PS ( $\log_2$  fold-change  $\geq |2|$ , FDR p-value correction  $\leq 0.05$ ). Only those genes whose variance across samples was more than 1 and FPKM  $\geq 0.1$ , were considered. (A) Venn diagram of upregulated lncRNA genes between stage transitions. (B) Venn diagram of downregulated lncRNA genes between stage transitions. (C) Heat map of expression levels of lncRNA genes in the four cell populations, with three biological replicas each. All lncRNA genes detected as differential in at least one stage transition were included. Z-score values are coded on the green-to-red scale (high expression: red; low expression: green).

as both studies employed strand-specific RNAseq of cell populations enriched in different stages of mouse spermatogenesis. However, on the other hand, notorious differences exist between these works and ours in the method for cell purification (both these studies used Staput, which involves longer cell collection times and renders lower purity), the age of the animals selected for the collection of the different fractions, and the pipelines used for data analysis. In spite of this, after re-analysing their raw data we found a significant correlation between expression data



**Figure 1.** lncRNA genes expressed in the four testicular cell populations. (A) Venn diagram showing the annotated lncRNA genes detected as expressed in the four cell populations with two different pipelines, and that passed all the filters. Only lncRNAs whose variance across samples was more than 1 and FPKM  $\geq 0.1$  in the three biological replicas, were considered. (B). Venn diagram showing the lncRNA genes expressed in each of the cell populations (2C, LZ, PS, and RS). Separate and overlapping expression between samples is shown. The considered parameters were the same as in (A).

from their experiments and ours for the different cell populations ( $r = 0.49\text{--}0.67$ ,  $p < 10^{-5}$  for the dataset from ref [39];  $r = 0.38\text{--}0.47$ ,  $p < 10^{-5}$  for ref [38]; Supplementary Fig. S3).

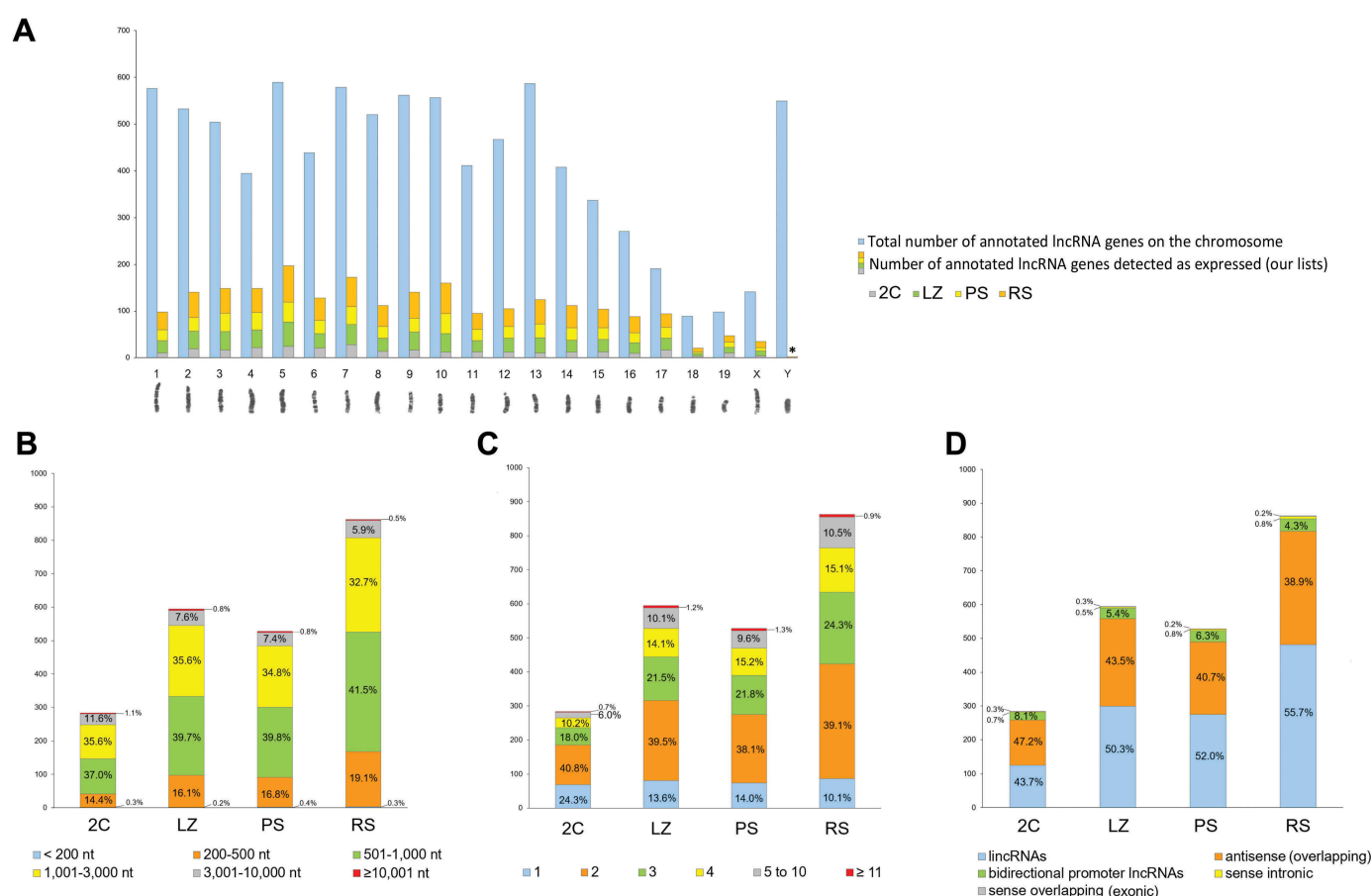
### Characterization of lncRNAs in the different testicular cell populations

In order to characterize the spermatogenic lncRNAs, we first analysed the chromosomal distribution of the lncRNA genes appearing as expressed in our lists (see Supplementary Table S2). In particular, it is noticeable that in spite of the small size of the Y chromosome, a very high number of lncRNAs are annotated on this chromosome in Ensembl database. Strikingly, we observed a very strong depletion in the number of Y-linked lncRNA genes expressed in testis (hypergeometric test  $p < 10^{-9}$ ; see Methods), and this was true for the lncRNAs from the four cell populations (Fig. 3A). On the other hand, a relatively low number of lncRNA genes are annotated on the X chromosome. In relation to testicular X-linked lncRNAs, we observed a switch off of lncRNAs in PS, and their switch on in RS (Supplementary Fig. S4). Nevertheless, the numbers are too small to attribute this finding to meiotic sex chromosome inactivation (MSCI). Moreover, the time course of testicular X-linked lncRNAs expression could

simply reflect the general behaviour of most testis-expressed lncRNAs, i.e. upregulation in RS.

We then analysed some features of the lncRNAs in the four cell populations, such as the average transcript length, number of exons, and biotypes (see Supplementary Table S2). Less than 20% of the expressed lncRNAs were smaller than 500 nt, around 75% had between 501 and 3,000 nt, and only a small proportion (between 6% and 13%) were larger than 3,000 nt, for the four populations (Fig. 3B). Concerning exon number, most lncRNAs had less than five exons, with a high proportion of them (around 40%) containing only two exons (Fig. 3C). When only the DE lncRNA genes were considered, these characteristics were observed for the different stage transitions as well (Supplementary Fig. S5 A, B).

Regarding biotypes, we used the categorization from Ensembl that classifies lncRNAs into lincRNAs (long intergenic non-coding RNAs), antisense (AS) lncRNAs (overlapping a protein-coding gene on the opposite strand), sense overlapping, sense intronic, macro lncRNAs (unspliced ncRNAs of several Kb in size), 3'-overlapping ncRNAs (on the 3'-UTR of a protein-coding locus on the same strand), and bidirectional promoter lncRNAs (originating from promoter regions of protein-coding genes, in the opposite direction on the other strand).



**Figure 3.** Characterization of lncRNAs in the different testicular cell populations. (A) Chromosomal distribution of lncRNAs. The total number of lncRNA genes annotated for each chromosome, and the number of lncRNA genes identified as expressed per chromosome in our lists for the 2C, LZ, PS and RS cell populations are shown. A mouse karyotype is presented below the graphic, as a reference of chromosome size. \* $p < 10^{-9}$ . (B) Length distribution of the expressed lncRNA genes in 2C, LZ, PS and RS cell populations. (C) Exon number distribution of the expressed lncRNA genes in the four testicular cell populations. (D) Biotype distribution of the expressed lncRNA genes in the four testicular cell populations.

LincRNA genes were the most abundantly expressed in our lists, followed by AS lncRNAs (Fig. 3D). This is not unexpected, as these have been reported to be the two most highly represented categories among lncRNAs in general [17], and particularly in testis as well [29,35,38]. However, the proportion of AS lncRNAs overlapping protein-coding genes in our lists was surprisingly high in relation to the small proportion of the genome occupied by protein-coding genes (see Fig. 3D). As stated above for other features, this was true for the four cell populations, and both for the total expressed and DE lncRNA genes (Supplementary Fig. S5C).

In particular, concerning lincRNAs, Lagarde *et al.* (2017) developed an experimental reannotation of the GENCODE intergenic polyA+ lncRNAs by means of RNA Capture Long Seq (CLS) [18]. We then re-analysed our sequencing data for lincRNAs, employing CLS annotation as a reference. Using Hisat/Stringtie/Balloon, we detected 1,020 lncRNAs expressed in the 2C cell population, 1,497 in LZ, 1,392 in PS, and 2,036 in RS (Supplementary Table S3). However, it is important to point out that due to existent redundancies, the number of unique lincRNA species at each stage must be lower (see Discussion). The analysis of our data with both annotations showed coincidence levels that ranged from 34.1% to 36.9% (Supplementary Fig. S6A) with high correlations (Supplementary Fig. S6B) for the four testicular cell populations. This is a good coincidence level, considering that the overall coincidence between CLS and GENCODE transcript catalogues from mouse is 20% [18]. In relation to DE lincRNAs, using CLS annotation we observed 52 DE lincRNAs at the LZ vs 2C transition, 54 in PS vs LZ, and 572 in RS vs PS. Thus, in general terms, the re-analysis of our data with CLS annotation confirms for the case of lincRNAs, the highest number in RS, both for the expressed and for the upregulated genes. On the other hand, as expected, transcripts identified with CLS reference were, in general, longer than those identified using Ensembl (median = 1,032 and 808 for CLS and Ensembl, respectively).

### Co-expression of lncRNAs with overlapping and neighbouring coding genes

Several studies have shown that AS transcripts can interfere with sense transcription of protein-coding genes by regulating gene expression and/or genome integrity, and exerting their effect in *cis* or *trans*, either locally or distally. Nevertheless, due to their genomic arrangement, it is believed that more frequently they act locally, in *cis* (e.g. by a sense-AS self-regulatory mechanism) [28,45].

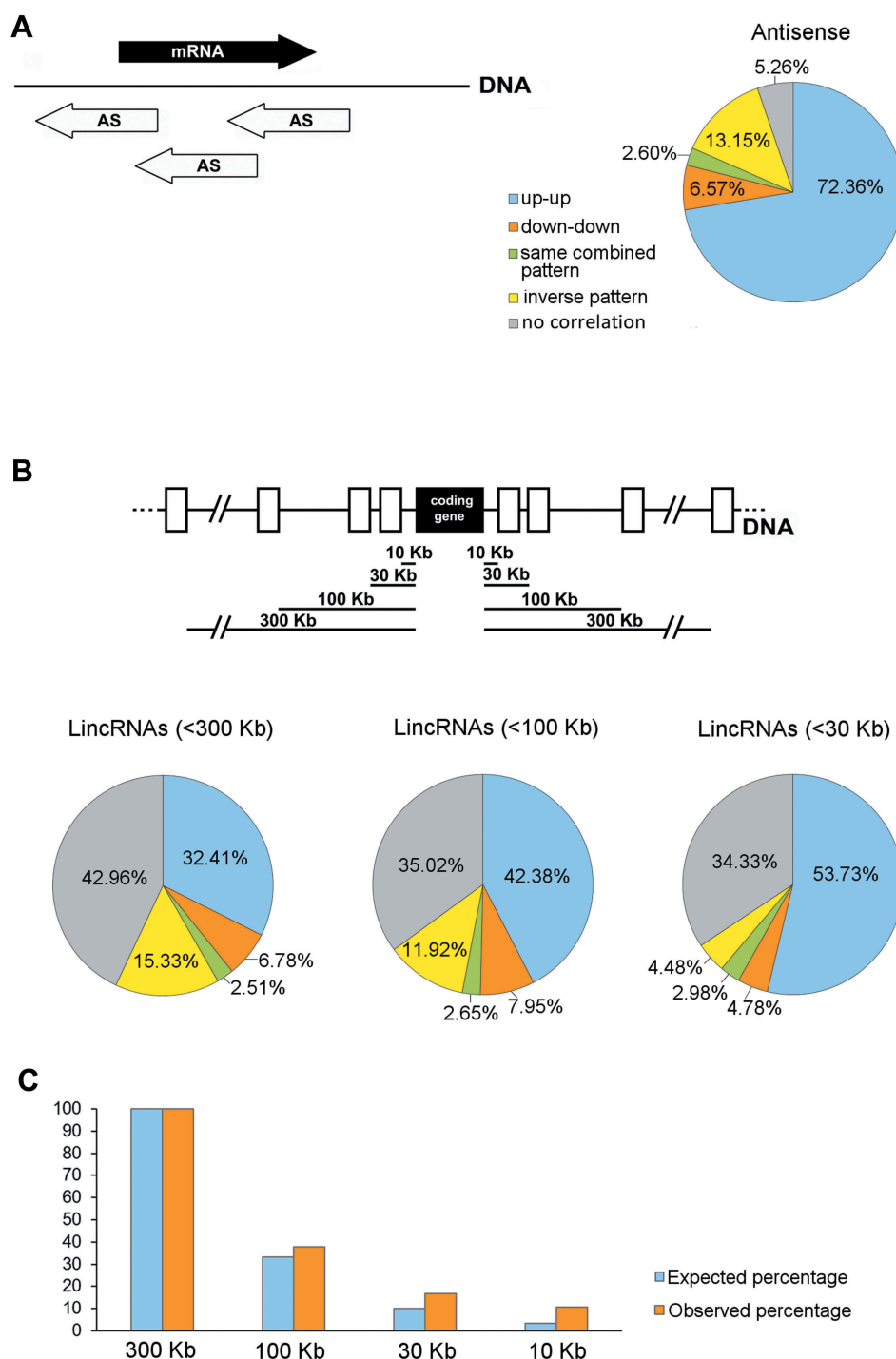
The high representation of AS lncRNAs in our lists, prompted us to analyse their co-expression with overlapping protein-coding genes. We found that for 85.5% of the expressed AS lncRNAs, the host protein-coding gene also appeared as expressed in our lists (Supplementary Table S4). Moreover, 81.5% of the DE AS-overlapping lncRNAs followed the same expression pattern as their cognate protein-coding genes (i.e. both either upregulated or downregulated at the same spermatogenic stage transition/s). Interestingly, for over 72% of the co-expressed gene pairs (i.e. 88.75% among those following the same expression pattern), their expression

pattern was 'up-up' (both the coding gene and the overlapping AS were upregulated;  $r = 0.81$ ,  $p < 10^{-5}$ ) (Fig. 4A, and Supplementary Table S4). In 13% of the gene pairs, an inverse correlation between the AS and its host mRNA gene was observed (i.e. one was upregulated and the other was downregulated at the same stage transition/s; see Fig. 4A;  $r = -0.70$ ,  $p < 10^{-4}$ ), while in only 5% of the cases the expression pattern between the AS lncRNA and its host coding gene could not be correlated ( $r = 0.20$ ,  $p = 0.52$ ).

Additionally, in at least 61% of the differentially co-expressed gene pairs, a testis-specific role or testis-restricted expression pattern for the protein-coding gene has been described (see Supplementary Table S4). Gene ontology (GO) analysis of the coding genes in the DE gene pairs showed a moderate enrichment ( $p < 0.05$ ) in the terms 'spermatogenesis', 'sperm motility', and 'microtubule-based movement' for the biological process category, and 'microtubule motor activity' for the molecular function category (data not shown).

We then evaluated the co-expression of lincRNAs with their neighbour protein-coding genes, as lincRNAs were the most highly represented lncRNA biotype in our lists. An increasing number of studies have shown that lincRNAs can affect the expression of their neighbouring protein-coding genes or other target genes by acting through different mechanisms, either in *cis* or in *trans* [27,46,47]. Thus, in order to identify possible functional relations, different studies have correlated lincRNAs with neighbour mRNA loci located at varying distances such as <300 Kb [47], <100 Kb [48,49], <30 Kb [35], or <10 Kb [29], depending on the study.

In this work we have analysed the co-expression of lincRNAs with coding genes (Supplementary Table S5), starting at a distance of <300 Kb. For 43% of the testis-expressed lincRNAs, there was at least one neighbour protein-coding gene whose transcript was also present in our lists of expressed genes. 41.7% of the DE gene pairs showed the same expression pattern, among which 77.7% were 'up-up' (32.41% of all the co-expressed pairs;  $r = 0.78$ ,  $p < 10^{-5}$ ). On the other hand, 15.33% of the gene pairs showed an inverse correlation ( $r = -0.71$ ,  $p < 10^{-5}$ ). Conversely, 43% of the gene pairs showed a very low correlation coefficient all along the stage transitions ( $r = -0.25$ ,  $p < 10^{-5}$ ; Fig. 4B). When only pairs located at <100 Kb were considered, the proportion of pairs with the same expression pattern along stage transitions raised to 53% (80% of which were 'up-up', i.e. 42.38% of the total;  $r = 0.75$ ,  $p < 10^{-5}$ ), while the percentage of gene pairs whose expression pattern could not be correlated ( $r = -0.07$ ,  $p = 0.51$ ) turned out to be 35% (see Fig. 4B). When we further narrowed the distances to <30 Kb, the percentage of pairs showing the same behaviour increased to 61.49% (of which almost 88% were 'up-up', i.e. 53.73% of the total;  $r = 0.85$ ,  $p < 10^{-5}$ ), although the percentage of those with no correlation ( $r = -0.06$ ,  $p = 0.64$ ) remained unchanged (see Fig. 4B). No significant differences were found when the distances were further reduced to <10 Kb (not shown). GO analysis of the coding genes co-expressed with neighbouring lincRNAs showed an enrichment ( $p < 0.01$ ) in 'nucleosome assembly', 'positive regulation of gene expression', and 'gene silencing' (biological process), and 'nucleosomal DNA binding' (molecular function), among other categories (not shown).



**Figure 4.** Co-expression of testicular lincRNA genes with protein-coding genes along the different spermatogenic stage transitions. (A) Representation of the percentages of the different co-expression patterns of antisense (AS) lincRNAs and their host protein-coding genes. The schematic diagram shows different positions of overlapping AS lincRNAs (white arrows) in relation to their cognate co-expressed mRNAs (black arrow). Expression patterns were classified as: up-up (both the AS and the coding gene are upregulated); down-down (both the AS and the coding gene are downregulated); same combined pattern (both the AS and the coding gene follow the same combined behaviour along spermatogenesis, e.g. both upregulated at a certain stage transition and then both downregulated at another stage transition, or vice versa); inverse pattern (the AS is upregulated when the coding gene is downregulated, or vice versa); and no correlation (the expression patterns of the AS and the coding gene were not correlatable). (B) Representation of the percentages of the different co-expression patterns of lincRNAs and neighbouring protein-coding genes. The results for gene pairs where the lincRNA is located at <300 Kb, <100 Kb, and <30 Kb distance from the co-expressed coding gene are shown. The top diagram illustrates the different distances at which co-expression patterns between lincRNAs (white rectangles) and protein-coding genes (black rectangle) were analysed. The categories of the expression patterns are the same as in (A). (C) Distribution of the distance of lincRNAs from their co-expressed neighbouring coding genes. The number of gene pairs located at <300 Kb from the coding gene is taken as 100%. The percentage of lincRNAs located at varying distances from the coding gene is shown. The expected percentages are those that would be observed at <100 Kb, <30 Kb and <10 Kb, if the distribution was even. For all the co-expression analyses, only those pairs in which both genes were expressed in our transcriptomes and at least one of them was DE, were considered.

In summary, the smaller the distance between lincRNAs and their neighbouring coding genes, the greater the trend to follow the same expression patterns along the different spermatogenic stage transitions. Anyway, the percentage of gene pairs that exhibited a similar behaviour was much higher for sense/AS pairs than for those involving lincRNAs. Likewise, while the proportion of the AS lincRNAs whose expression pattern could not be correlated with that of their cognate coding genes was very small, this percentage was significantly higher for lincRNAs. Despite this, most co-expressed lincRNA/coding gene pairs – both for AS and lincRNAs – were upregulated in RS. When we performed GO analysis for the coding genes in the RS-upregulated co-expressed pairs (including both AS and lincRNAs), we observed an enrichment in basically the same categories as above, while some other spermiogenic-specific categories such as ‘acrosomal vesicle’, also appeared significantly enriched ( $p < 0.01$ ). Above all, in all GO analyses there was a highly significant enrichment in the term ‘testis’ for the tissue category ( $p < 10^{-27}$ ), indicating that most DE coding genes that are co-expressed with lincRNAs along spermatogenesis are testis-specific.

We also observed that the distribution of lincRNAs in relation to their co-expressed neighbour coding genes was not uniform: 44.37% of the lincRNAs positioned at <100 Kb from a co-expressed coding gene were within <30 Kb distance, and 64.17% of those at <30 Kb were actually located at <10 Kb. Thus, testis-expressed lincRNAs are more concentrated within 10 Kb of co-expressed protein-coding genes, in accordance with what was observed in a couple of previous reports for lincRNAs in general [20,50] (Fig. 4C).

Next, we asked if the lincRNAs that are co-expressed with coding genes along spermatogenesis are conserved in humans, which would reinforce the idea of a possible functional relationship. For this purpose, we first used our co-expression lists of testicular AS lincRNAs and coding genes, and searched for annotated AS lincRNAs for the homologous coding genes in the human genome. By using two different human databases in parallel (Ensembl and Chess), we found annotated AS lincRNAs in human for 46% and 49% of them, respectively (see Supplementary Table S4). These percentages would be higher than the overall proportion of orthologous sense-AS pairs between mouse and human, which has been suggested to be less than 20% [51].

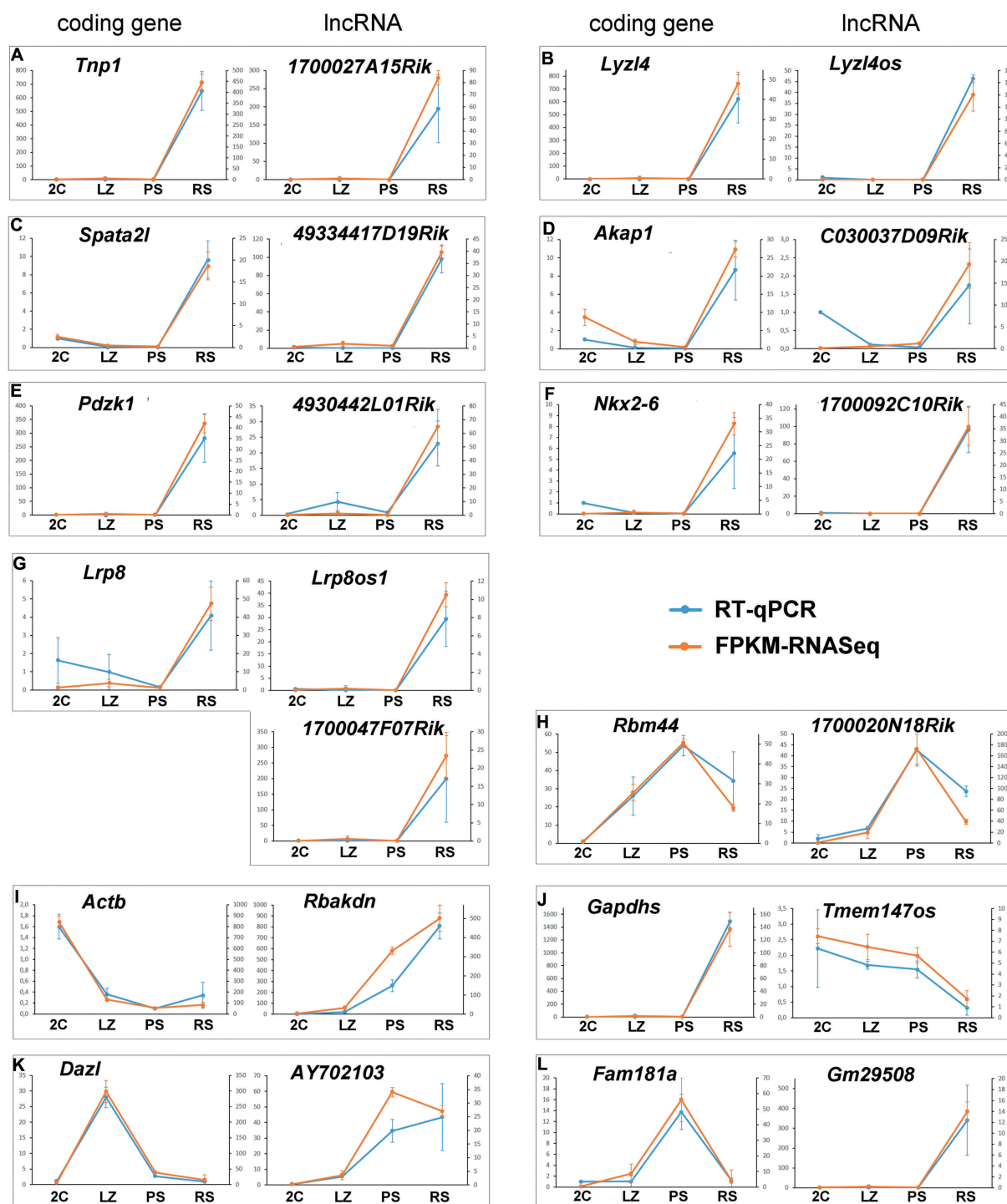
We then selected our co-expression lists of coding genes within <30 Kb from lincRNAs in mouse, and searched for the existence of homologous lincRNAs at syntenic positions in the human genome, as it has been suggested that the maintenance of the genomic position of lincRNAs relative to protein-coding genes might be important in determining their function [52]. Despite the fact that in most cases we located highly homologous DNA sequences in the human genome and near the same coding genes as in mouse, in general, we found no evidence of transcription for those specific sequences in human. On the other hand, in 59% of the cases, we found an annotated lincRNA within <30 Kb of the homologous coding gene in the human genome (see Supplementary Table S5), although blast searches showed no significant sequence homology with the mouse lincRNAs located at syntenic positions. In this regard, it has been stated that some lincRNAs can be orthologs located at conserved genomic locations, yet perhaps their sequences may be too

divergent to be detected with the existing tools, or their function may not depend on the strict nucleotide sequence [50,53]. The observed percentage of synteny is also higher than the ones suggested for lincRNAs between mouse and human in other contexts. As an example, it has been reported that only 10% of the DE lincRNAs upon activation of the innate immune response in human showed syntenic versions in mouse [52]. Although our findings could seem to go against previous reports suggesting low evolutionary conservation of testicular lincRNAs [21,38], those studies did not specifically refer to the conservation of testicular co-expressed pairs. Anyway, in principle, we cannot draw definitive conclusions about the existence of a significant conservation of co-expression patterns along spermatogenesis, as we do not know whether human syntenic lincRNAs exhibit the same co-expression patterns as in mouse.

### Validation of the co-expression patterns between lincRNAs and protein-coding genes

In order to validate the co-expression patterns of lincRNAs with coding genes from our RNAseq data, we selected 13 pairwise lincRNA-coding gene combinations for the analysis of their expression levels *via* RT-qPCR. On one side we tried to choose genes that would be upregulated at the different stage transitions and, on the other side, we aimed at selecting genes representative of different expression profiles (e.g. ‘up-up’, ‘down-up’, etc.). However, as for the vast majority of co-expressed pairs – both for those including AS and lincRNAs – their expression pattern was ‘up-up’ in RS, most of the selected pairs for confirmation were of this type. To exemplify this correlation, and focusing on pairs whose coding genes would encode proteins with known spermatid- or sperm-specific roles, we chose *Tnp1* (transition protein one), *Lyzl4* (Lysozyme-like 4), *Spata2l* (Spermatogenesis associated 2 like), *Akap1* (A-kinase anchoring protein 1), *Pdzk1* (PDZ domain containing 1), *Nkx2-6* (NK2 homeobox 6), and *Lrp8* (LDL receptor related protein 8, also known as apoER2) (Fig. 5 A-G). TNP1 participates in the replacement of histones by protamines in elongating spermatids [6]. LYZL4 is a sperm-related protein involved in fertilization [54], while SPATA2L is a paralog of SPATA2 (a necroptosis-involved protein), which is required for full fertility in mouse [55]. AKAP1 anchors protein-kinase A to mitochondria in sperm [56], while PDZK1 (localized at the middle piece of the sperm tail) and LRP8 are epididymal proteins required for normal sperm morphology and motility [57,58]. Although no specific bibliographic information is available for *Nkx2-6*, its expression pattern is highly restricted to adult testis [59] (<https://www.ncbi.nlm.nih.gov/gene/18092#gene-expression>), as has been also shown for a closely related member of its same gene family [60]. *Tnp1*, *Lyzl4*, *Spata2l*, *Pdzk1*, and *Nkx2-6* have a co-expressed AS, while *Akap1* has a co-expressed neighbour lincRNA, and *Lrp8* has both a co-expressed AS and a neighbour lincRNA (see Supplementary Tables S4 and S5).

To exemplify a positive correlation for gene pairs with an expression peak in PS we chose *Rbm44* (RNA-binding protein 44), an intercellular bridge component of pachytene and secondary spermatocytes [61], and its neighbour lincRNA (Fig. 5H).



**Figure 5.** Validation of the dynamic co-expression patterns of 13 selected gene pairs of lncRNAs and protein-coding genes. Gene pairs representative of the different co-expression profiles are shown. The graphics show the consistency between RNAseq data and RT-qPCR. The left axis indicates the relative quantification (RQ) of RT-qPCR, while the right axis indicates the FPKM values for RNAseq. The error bars correspond to  $\pm$ SD.

In order to show some inverse correlations, we selected *Actb* (beta-actin), *Gapdhs* (Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic), *Dazl* (Deleted in azoospermia like), and *Fam181a* (Family with sequence similarity 181 member A).

*Actb* was chosen as an example of an inverse correlation where the coding gene is downregulated from the 2C population while its neighbour lincRNA, *Rbakdn*, shows an opposite expression pattern (Fig. 5I). Regarding *Rbakdn*, an interesting

feature we found is that this lincRNA is conserved between mouse and human, and in both species its expression is testis-restricted [59,62] (see <https://www.ncbi.nlm.nih.gov/gene/100042605> and <https://www.ncbi.nlm.nih.gov/gene/389458>, respectively). Moreover, both in mouse and human *Rbakdn* is located near the same genes, including *Act b*.

*Gapdhs*, a spermiogenesis-specific counterpart of *Gapdh* that is required for sperm motility and male fertility [63],

was selected to exemplify an inverse correlation between a coding gene that is upregulated from PS to RS and its AS, with the opposite expression pattern (Fig. 5J).

We have previously shown that *Dazl*, which encodes a germ cell-specific RNA-binding protein required for the differentiation of germ cells, shows a marked expression peak in LZ and abruptly decreases at the LZ-to-PS transition [4]. Conversely, its neighbour lincRNA is upregulated at the LZ-to-PS transition, coinciding with the decline of *Dazl* mRNA (Fig. 5K). Finally, we selected *Fam181a* because it is overexpressed both in mouse and human testes [59,62] (<https://www.ncbi.nlm.nih.gov/gene/100504156> and <https://www.ncbi.nlm.nih.gov/gene/90050>), and a QTL related to fertility in cattle that overlaps this gene has been detected [64]. *Fam181a* exemplifies an inverse correlation for a gene with an expression peak in PS and downregulated at the PS-to-RS transition, and whose AS expression starts coincidentally with the mRNA decline (Fig. 5L).

The dynamic co-expression patterns of all genes were highly consistent with RNAseq analyses (see Fig. 5), showing the high reliability of the data in our lists.

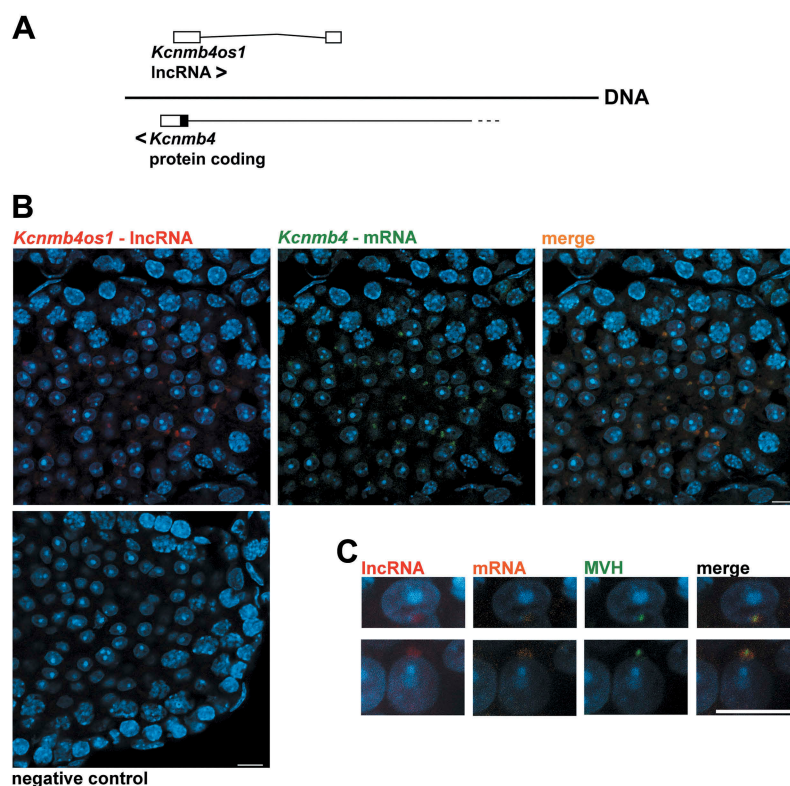
#### **An antisense lncRNA and its overlapping mRNA co-localize in the chromatoid body of round spermatids**

Next, in order to characterize a possible relation between AS lncRNAs and host mRNAs, we chose one such pairs for *in situ* hybridization using Stellaris® RNA-FISH probes [65]. The

selected RNAs were *Kcnmb4* mRNA, which encodes a regulatory subunit of a calcium-activated potassium channel [66] and its overlapping AS lncRNA, *Kcnmb4os1* (Fig. 6A). The expression of both the sense and AS transcripts is differential of testis compared to other mouse tissues [59] (<https://www.ncbi.nlm.nih.gov/gene/58802> and <https://www.ncbi.nlm.nih.gov/gene/?term=kcnmb4os1>, respectively). Both probes gave positive signals that co-localized in the chromatoid body of RS (Fig. 6B; colocalization Pearson's correlation coefficient 0.989), as shown by co-staining for MVH (DDX4), a well-characterized chromatoid body marker [9] (Fig. 6C). On the other hand, a probe against widely studied lncRNA *Malat1* [67] that was used as a positive control, gave the expected localization pattern in the nuclei of somatic testicular cells (Supplementary Fig. S7).

#### **Discussion**

Recent studies have identified thousands of testis-expressed lncRNAs in mouse [10,29,35,38]. However, the estimation of the exact number of lncRNAs is far more complicated than for coding transcripts. Besides the fact that coding potential cannot be taken into account for evaluation, lncRNAs are in general less abundant compared to mRNAs [17,20], and therefore it is difficult to set a baseline above which a lncRNA gene is considered as transcribed. Another drawback is that the annotation of lncRNAs is much less refined than that of coding transcripts. In particular, a disadvantage



**Figure 6.** Fluorescent *in situ* hybridization (FISH) analysis of *Kcnmb4* mRNA and its overlapping *Kcnmb4os1* AS lncRNA in mouse testis. (A) Diagram showing the structure and orientations of *Kcnmb4* and *Kcnmb4os1* transcripts. White rectangles represent non-coding exons. For the sense transcript, only the last coding exon is shown (coding region in black, 3' UTR in white). (B) Co-localization of the sense and AS transcripts in the chromatoid body of RS. Red: AS lncRNA; green: mRNA. The right frame shows co-staining with both probes. (C) Co-immunostaining with anti-MVH antibody (green) in RS as a marker for the chromatoid body. The AS is shown in red, while the mRNA is shown in orange in this case. The image to the right shows the triple staining. All the sections were co-stained with DAPI. Bars: 10  $\mu$ m.

of microarray-based studies is that due to the lack of complete annotations microarray probes are highly redundant [38], which may lead to overestimations of lncRNA numbers. Concerning RNAseq, although most analyses have employed dedicated software tools, more recently new software tools that provide more accurate overall results have been developed [40]. Another fact that complicates the scene is that in most studies it is not clearly specified whether lncRNA species (including transcript variants) or lncRNA genes were accounted for. In a different study (to be published elsewhere), we found around 60,000 unannotated transcripts that corresponded to non-coding RNAs (including splice variants) expressed in all the cell populations.

In this paper, we decided to work with annotated lncRNA genes. Besides, we deliberately decided to privilege reliability at the expense of amount of detected genes. The reliability of our data is based in the first place, in the method for cell classification, which yields highly pure stage-specific testicular cell populations [4]. These were combined with strand-specific RNAseq, which is essential for the accurate identification of AS lncRNAs [44,45]. This differentiates our stage-specific data from those from some other reports that used methods for cell-type enrichment [10,36,38,39] or whole testes [35,37], and some of them in combination with microarrays for transcriptome analysis [35,36]. Moreover, we have used the newest available bioinformatics tools for RNAseq analysis [40], and we have only selected those genes that arose from the intersection of two different pipelines. Besides, we have focused on lncRNA genes and not on lncRNA transcripts (i.e. we have not considered splice variants); this highly reduces the number of identified species. In addition, all DE genes were selected using FDR p-value correction, which corrects type I error thus reducing the number of false positives in the reported lists [68,69]. Hence, although we are aware that we are dealing with only a subset of all testis-expressed lncRNAs, we are convinced of the reliability of the results presented here, that correspond to those lncRNAs for which the evidence of their expression patterns is robust. We want to denote that despite the marked differences in the sampled cell populations between our study and others (see Results), we found significant correlations of detected genes with those studies we chose for comparison [38,39].

Additional support for the robustness and reproducibility of our data also comes from RT-qPCR analyses, which allowed confirming all the lncRNA expression patterns we chose for validation, as well as their co-expression with coding genes. Needless to say, as the raw data is deposited at the SRA, it is available for re-analysis by means of new or less conservative approaches.

A very remarkable result of our study is that the great majority of lncRNA genes in mouse testis are expressed in RS, which is in accordance with a couple of previous reports [10,38], but not with others [36,39]. Interestingly, we found that the difference in favour of RS was even much greater for the case of the DE lncRNA genes, indicating that for most lncRNAs that are present at different spermatogenic stages, their expression levels significantly raise after meiotic prophase.

We have analysed the lncRNA populations in the different spermatogenic cell types. Our analyses show that although the molecules that compose the lncRNA populations significantly vary at each spermatogenic stage, they all share the same general characteristics, with most lncRNAs being between 500 and 3,000 nt in length, and having less than 5 exons. These features are shared with testicular lncRNAs from other species, such as pig [31] and chicken [49]. On the other hand, these results differ from those of Chalmel *et al.* [30] that reported that in rat lncRNAs with an expression peak in meiosis were exceptionally long.

Most testicular lncRNAs in our lists are lincRNAs followed by AS, which has been also observed in some other studies, both for mouse [35,38] and chicken [49]. Interestingly, however, we have found that in all the cell populations the percentage of AS lncRNAs was surprisingly high, most probably due to the fact that strand-specific RNAseq contributes to their reliable identification. In this regard, it has been estimated that about 32% of lncRNAs in human would be AS to protein-coding genes [17], suggesting that regulation through AS lncRNAs is a commonly used mechanism [22,45].

Our results indicate that for the vast majority of AS lncRNAs that were co-expressed with their host coding genes in testis, there was a high correlation between the expression pattern of the sense and AS, all along the analysed spermatogenic stage transitions. Moreover, in most cases there was a positive correlation, and both the coding gene and the AS were upregulated. This, in the first place, suggests that the existing permissive, transcription-compatible chromatin state, would facilitate transcription from the other strand [28]. On the other hand, some studies in other tissues have revealed that AS lncRNAs transcription/transcripts can interfere with sense coding transcripts at different levels and in different ways, i.e. by acting at the initiation of transcription, co-transcriptionally, or post-transcriptionally, and exerting either activating or repressing effects [28,45]. In case at least some of the overlapping AS lncRNAs in testis modulate the expression of their host coding genes, the fact that most pairs are positively co-expressed could suggest a mechanism for regulation. A possibility is that AS lncRNAs transcription/transcripts mostly carry out a positive regulation on the expression of coding genes. Although we cannot exclude this possibility and, indeed, regulatory mechanisms of this type have been described in other tissues [70–72], the fact that the great majority of the co-expression of testicular AS lncRNAs with their cognate coding genes takes place in RS, raises another attractive hypothesis. As stated above, RS are characterized by extensive post-transcriptional regulation, among which translational delay stands out as a strategy through which a high amount of mRNAs are sequestered by diverse mechanisms [4,6]. This allows to regulate the translation time for sperm-related proteins [73], whose premature production in many cases would be detrimental [74,75]. Furthermore, an associated trait to the high post-transcriptional regulation levels in spermatogenic cells, and mainly of RS, is the existence of widespread transcriptional activity [10], which may be accompanied by inefficient translation as a mechanism to prevent protein overexpression [6]. Thus, it is tempting to speculate that at least in some cases, testicular AS lncRNAs

could somehow act on their complementary coding transcripts by sequestering them for translational repression, and/or eventually stabilizing them for delayed translation. An example of an AS lncRNA that enhances the stability of its sense mRNA in the brain is *BACE1-AS*, which has been associated with Alzheimer's disease [76].

The idea that at least some AS lncRNAs could modulate gene expression in RS by sequestering and/or stabilizing mRNAs for delayed translation, may be supported by the fact that many of the coding genes that appeared in our lists as upregulated in RS and co-expressed with their AS, encode proteins that are used in elongated spermatids or mature sperm. One such protein is TNPI; in this regard, it is well known that the mRNAs for protamines and transition proteins are translationally repressed in RS, and their premature translation causes spermatogenic arrest and infertility [74,75].

It is interesting that we identified an AS lncRNA (*Kcnmb4os1*) that co-localizes with its overlapping sense mRNA (*Kcnmb4*) in the chromatoid body of RS. While the true functions of this spermatid-specific structure remain intriguing, the available evidence points to a role in RNA-related processes such as the storing of repressed mRNAs and, more recently, also in the degradation of transcripts *via* non-sense mediated decay (NMD) [9,77]. However, how RNAs are targeted to the chromatoid body is presently unclear [77]. In particular, KCNMB4 is a regulatory subunit of a calcium-activated potassium channel, which modulates calcium sensitivity [66]. It is well established that ion channels play essential roles in sperm motility, sperm activation, acrosome reaction, and fertilization [78]. Moreover, although thus far the specific function of KCNMB4 in testis is not clear, its modulatory role in spermatogenesis has been suggested [79]. A nice hypothesis that deserves to be explored would be that *Kcnmb4os1* somehow interacts with the sense transcript, with the consequence of its targeting to the chromatoid body. If this were true, maybe it could be part of a tuning mechanism to modulate ion channel function in germ cells. It will be interesting to analyse the localization patterns of more AS/sense RNA pairs in RS, in order to determine if co-localization in the chromatoid body is an extended phenomenon.

Regarding lincRNAs co-expressed with neighbour coding genes, the proportion of them whose expression pattern was consistent with that of a coding gene in the four testicular cell populations was significantly lower than that for AS lncRNAs. This most probably indicates that many lincRNAs that are expressed in the same tissue and even at the same time as nearby coding genes, are not really co-regulated with them. In many cases, the co-expression pattern could simply represent transcriptional noise. This could be particularly so in the male germline, as a consequence of the widespread promiscuous transcription that operates in these cells [10]. On the other hand, the role of an important number of lincRNAs in modulating gene expression in different tissues has been undoubtedly shown [27,50,70,80]; however, we still have no clues about what proportion of them may have a biological role.

Coincidentally with our observations concerning AS, most of the lincRNAs/neighbour gene pairs whose expression pattern followed a similar behaviour in all the stage transitions, showed

upregulation in RS. The proportion of pairwise co-expressed lincRNAs/neighbour coding genes whose pattern was 'up-up', increased as the distance between the neighbouring genes decreased. Interestingly, our results both regarding the impact of the distance on co-expression patterns, and the positive correlation of AS with coding genes, are strikingly coincident with those obtained by Derrien et al. [17], through a bioinformatic analysis of the human lncRNA GENCODE annotation.

Specifically, in relation to lincRNAs, we also used our RNAseq data for re-analysis with a new, long-transcript annotation of intergenic lncRNAs (CLS) [18]. Although we found some redundancies that make direct comparisons difficult (e.g. many IDs from CLS correspond to the same GENCODE ID; many testicular transcripts also appear in other tissue/s with a different ID), this analysis produced an interesting wealth of new information. In particular, it is allowed to assign over 2,000 gene IDs corresponding to long-read intergenic transcripts, to specific spermatogenic stages.

Another intriguing fact is the almost complete depletion of testicular lncRNAs from the Y chromosome, despite the high number of annotated lncRNAs from this chromosome. It is interesting that a transcriptomic analysis of the chromatoid body of RS detected non-coding transcripts derived from all the chromosomes but the Y chromosome [9]. We here extend this finding to the whole transcriptome from RS and, moreover, to testicular lncRNAs in general. We currently lack an explanation for this, but we do not relate it to MSCI as we observed this depletion in the four testicular cell populations, and not specifically in relation to the pachytene stage. Definitely, this curious fact will deserve further investigation.

Finally, we have noted that in a number of cases, lncRNAs annotated as AS overlapping, are in fact very close adjacent, but non-overlapping neighbours to coding genes (and therefore, *sensu stricto* they should be classified as lincRNAs). Furthermore, while attempting to conduct conservation studies, we have observed that although for many mice lncRNAs the homologous DNA sequences exist in other mammalian species, there are no lncRNAs annotated for those specific sequences in the other species. This raises the question of whether none of those homologous sequences is transcribed in the other species, or if at least for some of them, the homologous lncRNAs may have not been annotated yet. No doubt, the years to come will represent a breakthrough in the research of lncRNAs in testis as annotations are optimized and, most importantly, as their roles in relation to the modulation of gene expression in spermatogenesis and fertility start to be elucidated.

## Methods

### Ethics statement

All animal procedures were performed following the recommendations of the Uruguayan National Commission of Animal Experimentation (CNEA), approved experimental protocol 001/02/2012 (code: 008/11; <http://www.cnea.org.uy/index.php/instituciones/registro/10>). Animals were euthanized by cervical dislocation, in accordance with the National Law of Animal Experimentation 18,611 (Uruguay).

## Animals

Male CD-1 Swiss mice (*Mus musculus*) at different ages were obtained from the animal facility at Instituto de Higiene of Facultad de Medicina (UdelaR, Montevideo, Uruguay). Immediately after euthanasia, testes were dissected and tunica albuginea was removed before proceeding to the preparation of testicular cell suspensions. Testes for *in situ* hybridization were processed with the tunica albuginea.

## Preparation of cellular suspensions and sorting by flow cytometry

Testicular cell suspensions were prepared by a procedure described earlier in our laboratory [42,43,81]. We introduced a brief modification to the preparation of the 2C cell population, i.e. a treatment with 0.6 U/mL collagenase for 15 min at room temperature before mechanical disaggregation step. Cells were counted in a Neubauer chamber and resuspended at a concentration of  $1 \times 10^6$  cells/mL in Dubelcco's Modified Eagle's medium (DMEM) supplemented with 10% foetal calf serum. Testicular cell suspensions were stained with Vybrant DyeCycle Green (Invitrogen, Life Technologies, Carlsbad, CA), as previously described [43]. Samples were analysed and sorted with a FACS Vantage flow cytometer (Beckton Dickinson, CA) furnished with an argon ion laser (Coherent, Innova 304) tuned at 488 nm of excitation wavelength (100 mW), and using a 70  $\mu$ m nozzle. The protocol for cell analysis and sorting was the same as reported earlier [4], with a slight difference in the ages of the animals used for the obtainment of each cell population. The 2C cell population was classified from a testicular cell suspension of a pool of up to five individuals ageing 12–14 days post-partum (dpp), LZ cell population was obtained from 15 to 18 dpp animals, PS from 16 to 19 dpp, and RS from 22 to 24 dpp animals. Of note, due to the age of the animals employed for the classification of the 2C population, this fraction does not include spermatocytes II. Sorting was set in Counter mode that yields the highest purity achieved by the equipment (>95%), with three sorted drops as envelope. Twelve samples were obtained (four different cell populations, with three biological replicates each), collected in PBS treated with 0.1% DMPC, spun down (500 g, 10 min, 4°C), deep frozen in liquid nitrogen, and stored at –80°C until use.

The purity of each sorted fraction was assessed, first using laser confocal and differential interference contrast (DIC) microscopy for the estimation of chromatin characteristics (based on VDG green fluorescence) as well as cell and nuclear size and morphology, as in previous reports from our group [42,43]. Besides, the purity of the LZ and PS sorted fractions was confirmed by immunodetection with an antibody against SYCP3 (a marker of the synaptonemal complex; Acris Antibodies, 1:100) on spread cells as instructed [82], to monitor the advance of meiotic prophase. The purity of the RS fraction was confirmed with an antibody against MVH, a marker of the chromatoid body (Abcam ab13840, 1:2,000) (see Supplementary Fig. S1).

## RNA extraction and construction of sequencing libraries

Total RNA from each of the 12 samples was extracted with PureLink RNA Mini Kit (Ambion, Life Technologies), following manufacturer's recommendations. RNA quantitation was performed by fluorometry using Qubit 2.0 and RNA HS Assay kit (Life Technologies). We used Ovation RNA-Seq System 1–16 for Mouse kit (NuGEN) to generate strand-specific sequencing libraries. In brief, 60 ng of total RNA was used as input, and the libraries were constructed without fragmentation of the RNA samples. The material was amplified for 16 PCR cycles, according to the instructions from the kit. Library concentration was measured by fluorometry with Qubit 2.0 and dsDNA Assay Kit (Life Technologies), and library quality was assessed on a 2100 Bioanalyzer system (Agilent, Santa Clara, CA). Libraries were sequenced at Fasteris (Switzerland) on an Illumina HiSeq4000 platform, and 150 bp paired-end reads were generated.

## LncRNA data analysis using Ensembl database

Raw data was processed using two different pipelines in parallel. The first used software package was CLC Genomics Workbench 10.1.1 (CLC bio). Raw reads were trimmed by quality ( $Q > 20$ ), length (more than 50 bp), and Illumina adaptors were removed. RNAseq analysis was performed with CLC to obtain lists of differentially expressed (DE) genes among spermatogenic stages, using the *M. musculus* Ensembl database (Gm38.p6 release 93) as reference genome. Mapping to the reference genome was performed using the following parameters: mismatch cost 2; insertion cost 3; deletion cost 3; length fraction 0.8; similarity fraction 0.8; and maximum number of hits per read 10. Our analyses were based on expressed lncRNA genes and not on lncRNA species (i.e. lncRNA splice variants were not considered). Differential gene expression between the four testicular cell populations was obtained by means of 'DE for RNAseq analysis' tool included in CLC Genomics Workbench (based on TMM normalization, assuming a negative binomial distribution, and modelling each gene by a Generalized Linear Model [GLM]). The statistical analysis across all group pairs (which relies on Wald test) was done by pairwise comparisons in chronological order of appearance along the first spermatogenic wave (LZ vs 2C; PS vs LZ; RS vs PS), retaining genes with  $\log_2$  fold-change  $\geq |2|$ , and False Discovery Rate (FDR) p-value correction  $\leq 0.05$ . We also filtered by variance, making a selection of those genes whose variance across samples was more than 1, and RPKM  $\geq 0.1$ .

The second pipeline used in parallel was based on free access software. We trimmed the sequences using Trim Galore, and employing the same parameters as in CLC. Clean reads were aligned to the *M. musculus* reference genome of Ensembl database (Gm38.p6 release 93; index was created using a masked reference), with Hisat 2.0 [83]. Aligned reads were assembled into transcripts and counted with StringTie 1.3 [84]. Differential gene expression (again, by pairwise comparisons in chronological order of appearance along the first spermatogenic wave) was analysed with Ballgown [85]. DE genes were identified with the following

parameters:  $\log_2$  fold-change  $\geq |2|$ , FDR p-value correction  $\leq 0.05$ , and again considering only those genes whose variance across samples was more than 1 and FPKM  $\geq 0.1$ . All RNAseq raw data were deposited in SRA repository, PRJNA548952.

A list of annotated lncRNAs available in Ensembl database (release 93) was downloaded. This list was used for filtering the DE gene results, in order to obtain a sub-list of annotated DE lncRNAs. lncRNAs that passed all the filters, and were detected as expressed or DE by both pipelines, were kept for further analysis.

Principal Component Analysis (PCA) was generated using 'PCA for RNAseq' CLC tool (that uses normalized log CPM [Counts Per Million] values as input). Matrix correlation was constructed in R bioconductor, calculating Pearson's correlation coefficient between FPKM expression of every transcript in each of the 12 samples.

Venn diagrams were constructed using Venny 2.1 (<http://bioinfogp.cnb.csic.es/tools/venny/>). For heatmaps construction, we used R bioconductor version 3.4.4 (<http://www.R-project.org>).

### Chromosomal distribution

To determine whether lncRNAs had preferential chromosome location, hypergeometric tests were performed in R bioconductor for the lncRNA genes expressed in each cell population, and enrichment/depletion p-values were calculated.

### Co-expression analysis

We analysed the co-expression of overlapping antisense (AS) lncRNAs with their host coding genes, and lncRNAs with their neighbour coding genes. For this purpose, we searched the coordinates of our lists of expressed coding- and lncRNA genes by means of BioMart tool in Ensembl database. We generated two lists (coding and non-coding) in BED file format. For lincRNAs, we added 300 Kb, 100 Kb, 30 Kb, or 10 Kb distances to both ends. BEDTools Intersect was applied over both lists (<https://github.com/arq5x/bedtools/blob/master/docs/content/tools/intersect.rst>) to obtain neighbour coding genes. The results were merged with the tables of DE genes. We only kept those pairs in which both genes were expressed in our transcriptomes, and at least one of them was DE. We used the Pearson's correlation test to calculate the correlation coefficients with their corresponding p-values for the lncRNA/coding gene pairs.

Gene ontology analyses for the coding genes overlapping AS lncRNAs or neighbouring lincRNAs were conducted using the functional annotation present at David Bioinformatics Resources 6.8 website (<https://david.ncifcrf.gov/>).

For syntenic analysis of lncRNAs in human, we used Ensembl Human Genome (GRCh38.p12, release 93). Chess database (<http://ccb.jhu.edu/chess/>) [86] was used in parallel for the analysis of gene pairs involving AS lncRNAs. As Chess does not discriminate between lncRNA subtypes except for AS, it was not employed for the syntenic analysis of gene pairs involving lincRNAs.

### Data comparison with other RNAseq studies

For comparison of our lncRNA lists with those from other reports, we downloaded datasets from refs [38] and [39]. We re-processed the raw data from both studies using CLC, as the study from ref [38] had been only performed with two biological replicates, and the use of Ballgown is not recommended for less than three replicates. In order to make data comparable, the same parameters as indicated above for our sequencing data were employed. The generated lists of lncRNAs expressed at each stage were crossed with our CLC-generated lists. We used the Pearson's correlation test to calculate the correlation coefficients with their corresponding statistical significance for the coincidences.

### lncRNA data analysis using CLS database

For the study with CLS annotation reference [18], our sequencing data were analysed with Hisat/StringTie/Ballgown, using those transcripts built with 'anchored' method as reference, and all tissues-derived annotation provided at the portal for CLS data (<https://public-docs.crg.es/rguigo/CLS/>). We associated the IDs and biotypes from ref [18] to those equivalent from GENCODE, by means of 'Transcript-to-biotype' file that is available at the web portal of CLS annotation. We particularly worked with those transcripts defined as 'lncRNAs', and considering those transcripts with variance  $\geq 1$  and FPKM  $\geq 0.1$ . For the DE transcripts, we considered those with  $\log_2$  (FC)  $\geq |2|$ . We took p-values  $\leq 0.01$  for the LZ vs 2C and PS vs LZ transitions, and FDR  $\leq 0.05$  for the RS vs PS transition as statistical values. We then compared the obtained data with ours (analysed with the same pipeline but using Ensembl v93 as reference, and only considering lincRNAs). Pearson's correlation coefficients with their statistical significance were calculated for the coincidences.

### RT-qPCR validation

For confirmative RT-qPCR, 3,000-cell fractions from 2C, LZ, PS and RS populations were sorted as explained above, but using a MoFlo Astrios EQ (Beckman Coulter) in Purify mode. Generation of cell lysates, reverse transcription and qPCR were performed using Power SYBR Green Cells-to-Ct kit (Ambion, Life Technologies) following the instructions of the manufacturer. For qPCR step, 2  $\mu$ L cDNA in 20  $\mu$ L final volume reaction mix was used. All the reactions were made in a CFX96 Touch Real-Time PCR Detection System 1 (BioRad, Hercules, CA), with three biological replicates each.

As the expression levels of commonly used control genes such as *Gapdh* and *Actb* significantly vary across the different testicular cell populations (e.g. see Fig. 5I), we chose *Surf4* (Surfeit gene 4) as normalizing gene because it exhibited similar expression levels in the four cell populations in our RNAseq data (55.99  $\pm$  2.91 FPKM in 2C; 59.91  $\pm$  10.31 FPKM in LZ; 59.36  $\pm$  5.76 FPKM in PS; 45.09  $\pm$  4.38 FPKM in RS). The coding genes and their AS lncRNAs or neighbour lincRNAs selected for confirmation by RT-qPCR are shown

in Fig. 5, and all especially designed primers are listed in Supplementary Table S6.

Amplification efficiency of the primers was >93%. We made relative expression quantification using the  $2^{-\Delta\Delta Ct}$  method, and 2C as calibrator condition.

### Stellaris® RNA fluorescence *in situ* hybridization

We designed probe sets using the Stellaris® FISH probe designer (<https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer>), and had them synthesized (Biosearch Technologies) as follows: probes against *Kcnmb4* were conjugated to Quasar570, while probes against *Kcnmb4os1* were conjugated to Quasar670. Ready-to-use *Malat1* probes conjugated to Quasar570 (SMF-3008-19) were purchased from the same company.

Testes of male mice at 25 dpp were cut in halves, fixed in 4% paraformaldehyde in PHEM buffer (60 mM PIPES pH 7.4, 25 mM HEPES, 10 mM EGTA, 2 mM MgCl<sub>2</sub>) for 1 h at room temperature, and cryoprotected as previously described [87]. Sections of 10 µm in thickness were obtained, transferred to poly-L-lysine-coated slides, and kept at -20°C.

*In situ* hybridization was performed following the protocols as indicated by the manufacturer for 'frozen tissues' [65], and using the commercial buffers from the company.

For those cases where FISH and co-immunostaining were performed on the same sections, immunodetection was done before the *in situ* hybridization. Anti-MVH antibody (Abcam, ab13840) was used at 1:2,000 dilution, followed by incubation with secondary anti-rabbit Alexa 488 antibody (A-11034, Invitrogen) 1:1,000. We followed the protocol as suggested by Biosearch Technologies for 'sequential IF'. All the protocols are in Biosearch Technologies web page (<https://www.biosearchtech.com/support/resources/stellaris-protocols>).

The sections were examined under a Zeiss LSM 800 confocal microscope, and photographed with Axiocam 506 colour digital camera (Carl Zeiss Microscopy, Germany). Overlapping between sense and AS probes was analysed by using Colocalization Finder plugin in Fiji ImageJ.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work was supported by Agencia Nacional de Investigación e Innovación (ANII, Uruguay) under grant FCE-1-2014-1-104251 to AG (including a PhD scholarship to MFT), and Comisión Sectorial de Investigación Científica (CSIC), UDELAR (Uruguay), under an I+D Groups grant to AG and RB. MFT was awarded with a Boehringer Ingelheim Fonds travel grant, and a short-term research grant from DAAD. Flow Cytometry purifications with Astrios EQ were carried out under ANII grant PEC\_1\_2016\_1\_133123.

### ORCID

María F. Trovero  <http://orcid.org/0000-0001-9257-6979>  
Rosana Rodríguez-Casuriaga  <http://orcid.org/0000-0002-7656-0288>  
Carlos Romeo  <http://orcid.org/0000-0002-4727-4098>

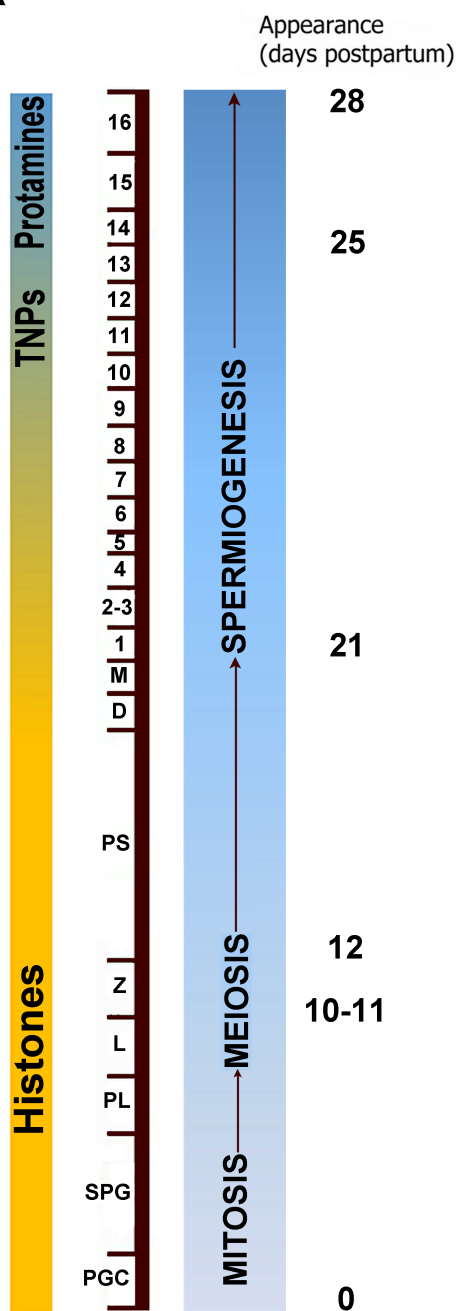
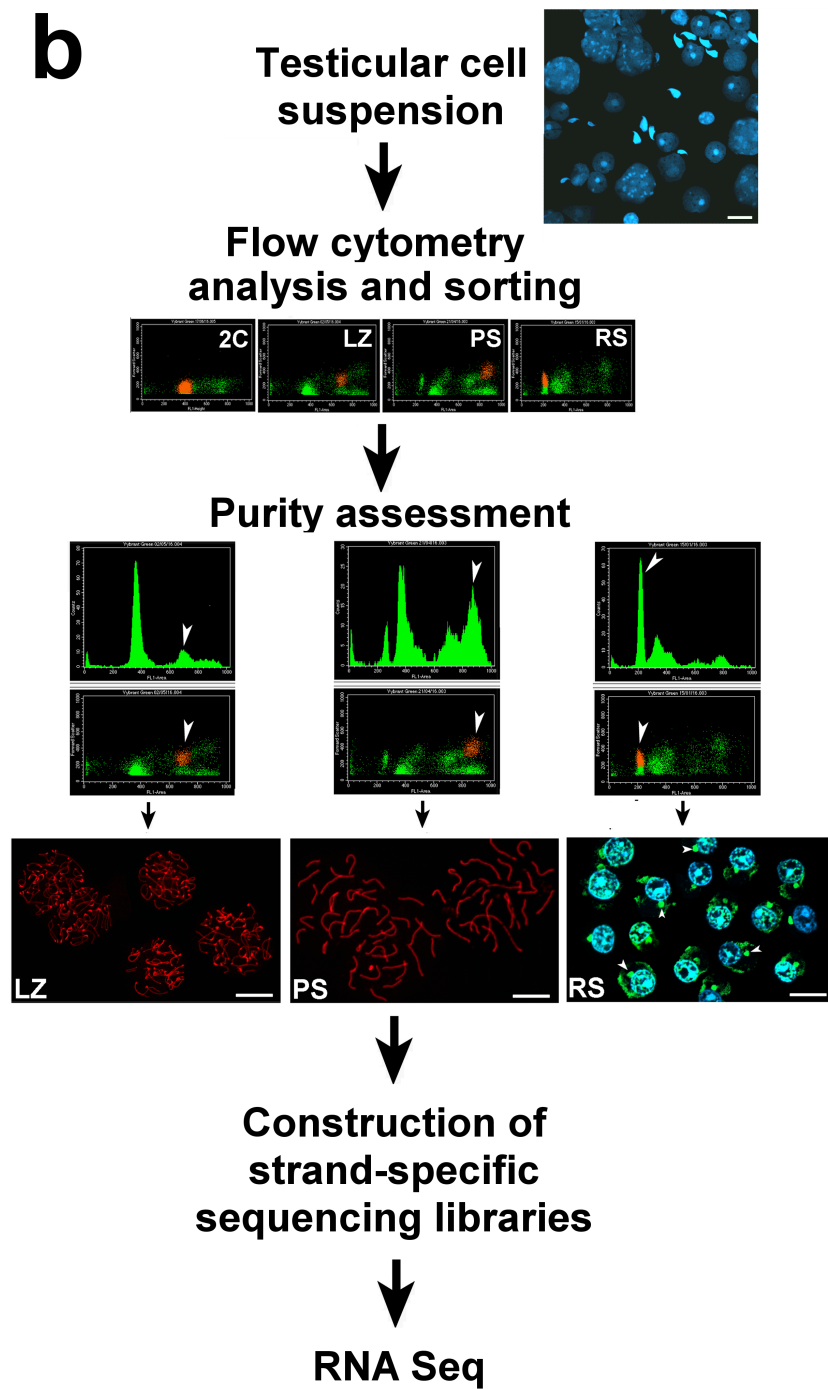
Gustavo A. Folle  <http://orcid.org/0000-0002-8205-1615>  
Ricardo Benavente  <http://orcid.org/0000-0001-6361-0672>  
José R. Sotelo-Silveira  <http://orcid.org/0000-0002-4758-8556>  
Adriana Geisinger  <http://orcid.org/0000-0002-3995-9185>

### References

- [1] Romrell LJ, Bellvé AR, Fawcett DW. Separation of mouse spermatogenic cells by sedimentation velocity. *Dev Biol.* 1976;19:119–131.
- [2] Meistrich ML. Separation of spermatogenic cells and nuclei from rodent testes. *Methods Cell Biol.* 1977;15:15–54.
- [3] Gaysinskaya V, Bortvin A. Flow cytometry of murine spermatocytes. *Curr Protoc Cytom.* 2015;72:7.44.1–7.44.24.
- [4] Da Cruz I, Rodríguez-Casuriaga R, Santiñaque FF, et al. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to-postmeiotic-related processes at pachytene stage. *BMC Genomics.* 2016;17(294). DOI:10.1186/s12864-016-2618-1
- [5] Bolcun-Filas E, Handel MA. Meiosis: the chromosomal foundation of reproduction. *Biol Reprod.* 2018;99:112–126.
- [6] Kleene KC. A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mech Dev.* 2001;106:3–23.
- [7] Geisinger A. Spermatogenesis in mammals: a very peculiar cell differentiation process. In *Cell Differentiation Research Developments*. Ivanova LB, ed. New York, NY: Nova Publishers; 2008. p. 97–123.
- [8] Gan H, Cai T, Lin X, et al. Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol Cell Proteomics.* 2013;12:1144–1157.
- [9] Meikar O, Vagin VV, Chalmel F, et al. An atlas of chromatoid body components. *RNA.* 2014;20:483–495.
- [10] Soumilion M, Necseulea A, Weier M, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* 2013;3:2179–2190.
- [11] Atkinson SR, Marguerat S, Bähler J. Exploring long non-coding RNAs through sequencing. *Semin Cell Dev Biol.* 2012;23:200–205.
- [12] Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem.* 2012;81:145–166.
- [13] Amaral PP, Clark MB, Gascoigne DK, et al. LncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res.* 2011;39:D146–151.
- [14] Li LJ, Leng RX, Fan YG, et al. Translation of noncoding RNAs: focus on lncRNAs, pri-miRNAs, and circRNAs. *Exp Cell Res.* 2017;361:1–8.
- [15] Ginger MR, Shore AN, Contreras A, et al. A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci.* 2006;103:5781–5786.
- [16] Mehler MF, Mattick JS. Noncoding RNAs and RNA editing in brain development, functional diversification, and neurological disease. *Physiol Rev.* 2007;87:799–823.
- [17] Derrien T, Johnson R, Bussotti G, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 2012;22:1775–1789.
- [18] Lagarde J, Uszczynska-Ratajczak B, Carbonell S, et al. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet.* 2017;49:1731–1740.
- [19] Wilhelm BT, Marguerat S, Watt S, et al. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature.* 2008;453:1239–1243.
- [20] Cabili MN, Trapnell C, Goff L, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011;25:1915–1927.
- [21] Necseulea A, Soumilion M, Warnefors M, et al. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature.* 2014;505:635–640.
- [22] Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol.* 2013;10:925–933.

- [23] Guenzl PM, Barlow DP. Macro lncRNAs: a new layer of cis-regulatory information in the mammalian genome. *RNA Biol.* **2012**;9:731–741.
- [24] Carninci P. The transcriptional landscape of the mammalian genome. *Science.* **2005**;309:1559–1563.
- [25] Hon CC, Ramilowski JA, Harshbarger J, et al. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature.* **2017**;543:199–204.
- [26] Wu T, Du Y. LncRNAs: from basic research to medical application. *Int J Biol Sci.* **2017**;13:295–307.
- [27] Kopp F, Mendell JT. Functional classification and experimental dissection of long noncoding RNAs. *Cell.* **2018**;172:393–407.
- [28] Barman P, Reddy D, Bhaumik SR. Mechanisms of antisense transcription initiation with implications in gene expression, genomic integrity and disease pathogenesis. *Noncoding RNA.* **2019**;5:E11.
- [29] Hong SH, Kwon JT, Kim J, et al. Profiling of testis-specific long noncoding RNAs in mice. *BMC Genomics.* **2018**;19(1). DOI:10.1186/s12864-018-4931-3
- [30] Chalmel F, Lardenois A, Evrard B, et al. High-resolution profiling of novel transcribed regions during rat spermatogenesis. *Biol Reprod.* **2014**;91:5.
- [31] Ran M, Chen B, Li Z, et al. Systematic identification of long noncoding RNAs in immature and mature porcine testes. *Biol Reprod.* **2016**;94:77.
- [32] Jan SZ, Vormer TL, Jongejan A, et al. Unraveling transcriptome dynamics in human spermatogenesis. *Development.* **2017**;144:3659–3673.
- [33] Luk AC, Chan WY, Rennert OM, et al. Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptome studies. *Reproduction.* **2014**;147:R131–41.
- [34] Nakajima R, Sato T, Ogawa T, et al. A noncoding RNA containing a SINE-B1 motif associates with meiotic metaphase chromatin and has an indispensable function during spermatogenesis. *PLoS One.* **2017**;12:e0179585.
- [35] Bao J, Wu J, Schuster AS, et al. Expression profiling reveals developmentally regulated lncRNA repertoire in the mouse male germline. *Biol Reprod.* **2013**;89:107.
- [36] Liang M, Li W, Tian H, et al. Sequential expression of long noncoding RNA as mRNA gene expression in specific stages of mouse spermatogenesis. *Sci Rep.* **2014**;4:5966.
- [37] Laiho A, Kotaja N, Gyenesei A, et al. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One.* **2013**;8:e61558.
- [38] Lin X, Han M, Cheng L, et al. Expression dynamics, relationships, and transcriptional regulations of diverse transcripts in mouse spermatogenic cells. *RNA Biol.* **2016**;13:1011–1024.
- [39] Wichman L, Somasundaram S, Breindel C, et al. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol Reprod.* **2017**;97:313–323.
- [40] Pertea M, Kim D, Pertea GM, et al. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* **2016**;11:1650–1667.
- [41] Chen Y, Zheng Y, Gao Y, et al. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res.* **2018**;28:879–896.
- [42] Rodríguez-Casuriaga R, Santiñaque FF, Folle GA, et al. Rapid preparation of rodent testicular cell suspensions and spermatogenic stages purification by flow cytometry using a novel blue-laser-excitable vital dye. *MethodsX.* **2014**;1:239–243.
- [43] Geisinger A, Rodríguez-Casuriaga R. Flow cytometry for the isolation and characterization of rodent meiocytes. *Methods Molec Biol.* **2017**;1471:217–230.
- [44] Ilott NE, Ponting CP. Predicting long non-coding RNAs using RNA sequencing. *Methods.* **2013**;63:50–59.
- [45] Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. *Nat Rev Genet.* **2013**;14:880–893.
- [46] Khalil AM, Guttman M, Huarte M, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA.* **2009**;106:11667–11672.
- [47] Ørom UA, Derrien T, Beringer M, et al. Long noncoding RNAs with enhancer-like function in human cells. *Cell.* **2010**;143:46–58.
- [48] Wang Y, Xue S, Liu X, et al. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Sci Rep.* **2016**;6:20238.
- [49] Liu Y, Sun Y, Li Y, et al. Analyses of long non-coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. *Sci Rep.* **2017**;22:9055.
- [50] Ulitsky I, Bartel DP. LincRNAs: genomics, evolution, and mechanisms. *Cell.* **2013**;154:26–46.
- [51] Breschi A, Gingeras TR, Guigó R. Comparative transcriptomics in human and mouse. *Nat Rev Genet.* **2017**;18:425–440.
- [52] Roux BT, Heward JA, Donnelly LE, et al. Catalog of differentially expressed long non-coding RNA following activation of human and mouse innate immune response. *Front Immunol.* **2017**;8:1038.
- [53] Hezroni H, Koppstein D, Schwartz M, et al. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.* **2015**;11:1110–1122.
- [54] Sun R, Shen R, Li J, et al. Lyzl4, a novel mouse sperm-related protein, is involved in fertilization. *Acta Biochim Biophys Sin (Shanghai).* **2011**;43:346–353.
- [55] Zhao J, Zhao J, Xu G, et al. Deletion of Spata2 by CRISPR/Cas9n causes increased inhibin alpha expression and attenuated fertility in male mice. *Biol Reprod.* **2017**;97:497–513.
- [56] Lin RY, Moss SB, Rubin CS. Characterization of S-AKAP84, a novel developmentally regulated A kinase anchor protein of male germ cells. *J Biol Chem.* **1995**;270:27804–27811.
- [57] Andersen OM, Yeung CH, Vorum H, et al. Essential role of the apolipoprotein E receptor-2 in sperm development. *J Biol Chem.* **2003**;278:23989–23995.
- [58] Liang AJ, Wang GS, Ping P, et al. The expression of the new epididymal luminal protein of PDZ domain containing 1 is decreased in asthenozoospermia. *Asian J Androl.* **2018**;20:154–159.
- [59] Yue F, Cheng Y, Breschi A, et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature.* **2014**;515:355–364.
- [60] Wang CC, Brodnicki T, Copeland NG, et al. Conserved linkage of NK-2 homeobox gene pairs Nkx2-2/2-4 and Nkx2-1/2-9 in mammals. *Mamm Genome.* **2000**;11:466–468.
- [61] Iwamori T, Lin YN, Ma L, et al. Identification and characterization of RBM44 as a novel intercellular bridge protein. *PLoS One.* **2011**;6:e17066.
- [62] Fagerberg L, Hallström BM, Oksvold P, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics.* **2014**;13:397–406.
- [63] Miki K, Qu W, Goulding EH, et al. Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility. *Proc Natl Acad Sci USA.* **2004**;101:16501–16506.
- [64] Nayeri S, Sargolzaei M, Abo-Ismael MK, et al. Genome-wide association for milk production and female fertility traits in Canadian dairy Holstein cattle. *BMC Genet.* **2016**;17:75.
- [65] Orjalo AV Jr, Johansson HE. Stellaris® RNA fluorescence in situ hybridization for the simultaneous detection of immature and mature long noncoding RNAs in adherent cells. *Methods Mol Biol.* **2016**;1402:119–134.
- [66] Weiger TM, Holmqvist MH, Levitan IB, et al. A novel nervous system beta subunit that downregulates human large conductance calcium-dependent potassium channels. *J Neurosci.* **2000**;20:3563–3570.
- [67] West JA, Davis C, Sunwoo H, et al. The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. *Mol Cell.* **2014**;55:791–802.
- [68] Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA.* **2003**;100:9440–9445.

- [69] Conesa A. A survey of best practices for RNA-seq data analysis. *Genome Biol.* **2016**;17:13.
- [70] Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol.* **2012**;19:1068–1075.
- [71] Liu Z, Zhao P, Han Y, et al. lncRNA FEZF1-AS1 is associated with prognosis in lung adenocarcinoma and promotes cell proliferation, migration, and invasion. *Oncol Res.* **2018**;27:39–45.
- [72] Yung Y, Ophir L, Yerushalmi GM, et al. HAS2-AS1 is a novel LH/hCG target gene regulating HAS2 expression and enhancing cumulus cells migration. *J Ovarian Res.* **2019**;12:21.
- [73] Chowdhury TA, Kleene KC. Identification of potential regulatory elements in the 5' and 3' UTRs of 12 translationally regulated mRNAs in mammalian spermatids by comparative genomics. *J Androl.* **2012**;33:244–256.
- [74] Lee K, Haugen HS, Clegg CH, et al. Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proc Natl Acad Sci USA.* **1995**;92(26):12451–12455.
- [75] Tseden K, Topaloglu Ö, Meinhardt A, et al. Premature translation of transition protein 2 mRNA causes sperm abnormalities and male infertility. *Mol Reprod Dev.* **2007**;74:273–279.
- [76] Faghihi MA. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol.* **2010**;11:R56.
- [77] Lehtiniemi T, Kotaja N. Germ granule-mediated RNA regulation in male germ cells. *Reproduction.* **2018**;155:R77–R91.
- [78] Shukla KK, Mahdi AA, Rajender S. Ion channels in sperm physiology and male fertility and infertility. *J Androl.* **2012**;33:777–788.
- [79] Yang CT, Zeng XH, Xia XM, et al. Interactions between beta subunits of the KCNMB family and Slo3: beta4 selectively modulates Slo3 expression and function. *PLoS One.* **2009**;4:e6135.
- [80] Fatica A, Bozzoni I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet.* **2014**;15:7–21.
- [81] Rodríguez-Casuriaga R, Folle GA, Santiñaque F, et al. Simple and efficient technique for the preparation of testicular cell suspensions. *J Vis Exp.* **2013**;78. DOI:10.3791/50102.
- [82] Rodríguez-Casuriaga R, Geisinger A, Santiñaque FF, et al. High-purity flow sorting of early meocytes based on DNA analysis of guinea pig spermatogenic cells. *Cytometry A.* **2011**;79:625–634.
- [83] Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* **2015**;12:357–360.
- [84] Pertea M, Pertea GM, Antonescu CM, et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* **2015**;33:290–295.
- [85] Frazee AC, Pertea G, Jaffe AE, et al. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol.* **2015**;33:243–266.
- [86] Pertea M, Shumate A, Pertea G, et al. CHES: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol.* **2018**;19:208.
- [87] Capovano CA, Wettstein R, Kun A, et al. *Spats 1 (Srsp1)* is differentially expressed during testis development of the rat. *Gene Expr Patterns.* **2010**;10:1–8.

**a****b**

**Supplementary Figure S1. Timing of spermatogenesis in mouse, and graphical outline of the workflow.** (a) Diagram representing the main three spermatogenic phases. The different cell types (left) and the onset of some of them in days postpartum (right) are shown. The substitution of histones - first by transition proteins (TNPs) and then by protamines - is also represented. PGC: primordial germ cells; SPG: spermatogonia; PL: preleptotene; L: leptotene; Z: zygotene; PS: pachytene; D: diplotene; M: meiotic divisions; 1-16 represent the different spermatid stages. (b) Schematic diagram of the followed experimental procedure until RNAseq. Note the heterogeneity of the cell suspensions (top DAPI-stained photograph). In the cytometric profiles, the arrowheads point at the peaks for LZ, PS and RS cell populations in the histograms, and the respective regions selected for sorting in the dot plots in each case (orange). Immunofluorescent analysis of the LZ, PS and RS sorted fractions is shown below. Anti-SYCP3 antibody (red) was used for purity assessment on spread spermatocyte cells of the LZ and PS sorted fractions, while anti-MVH (green) was employed for the analysis of the purity of the RS fraction. In the latter, nuclei appear counterstained with DAPI, and the arrowheads point at chromatoid bodies. In all the micrographs the bars correspond to 10  $\mu$ m.

Principal component scatter plot

Principal component 1 (39.9%)

Principal component 2 (25.3%)

Condition (Node color)

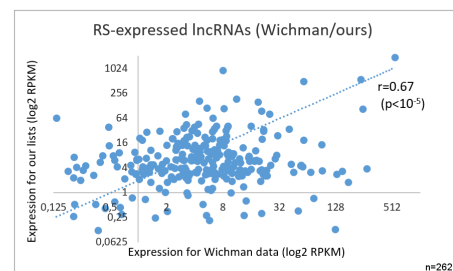
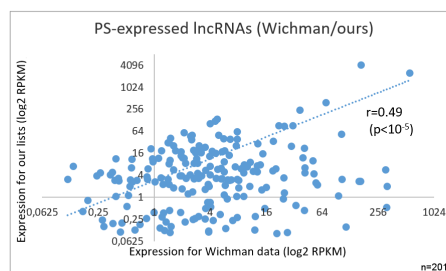
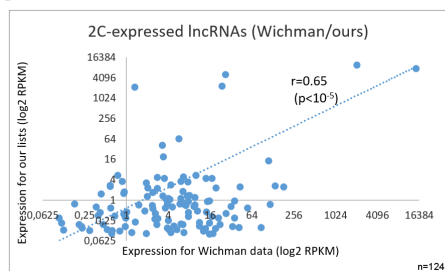
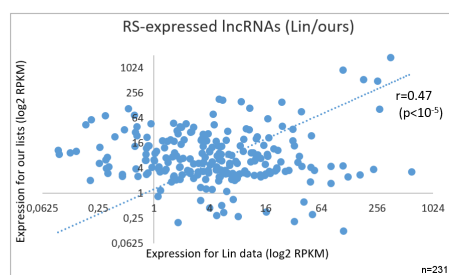
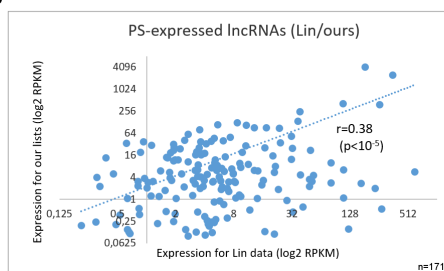
- 2C (blue)
- LZ (yellow)
- PS (purple)
- RS (green)

Nodes labeled in the plot:

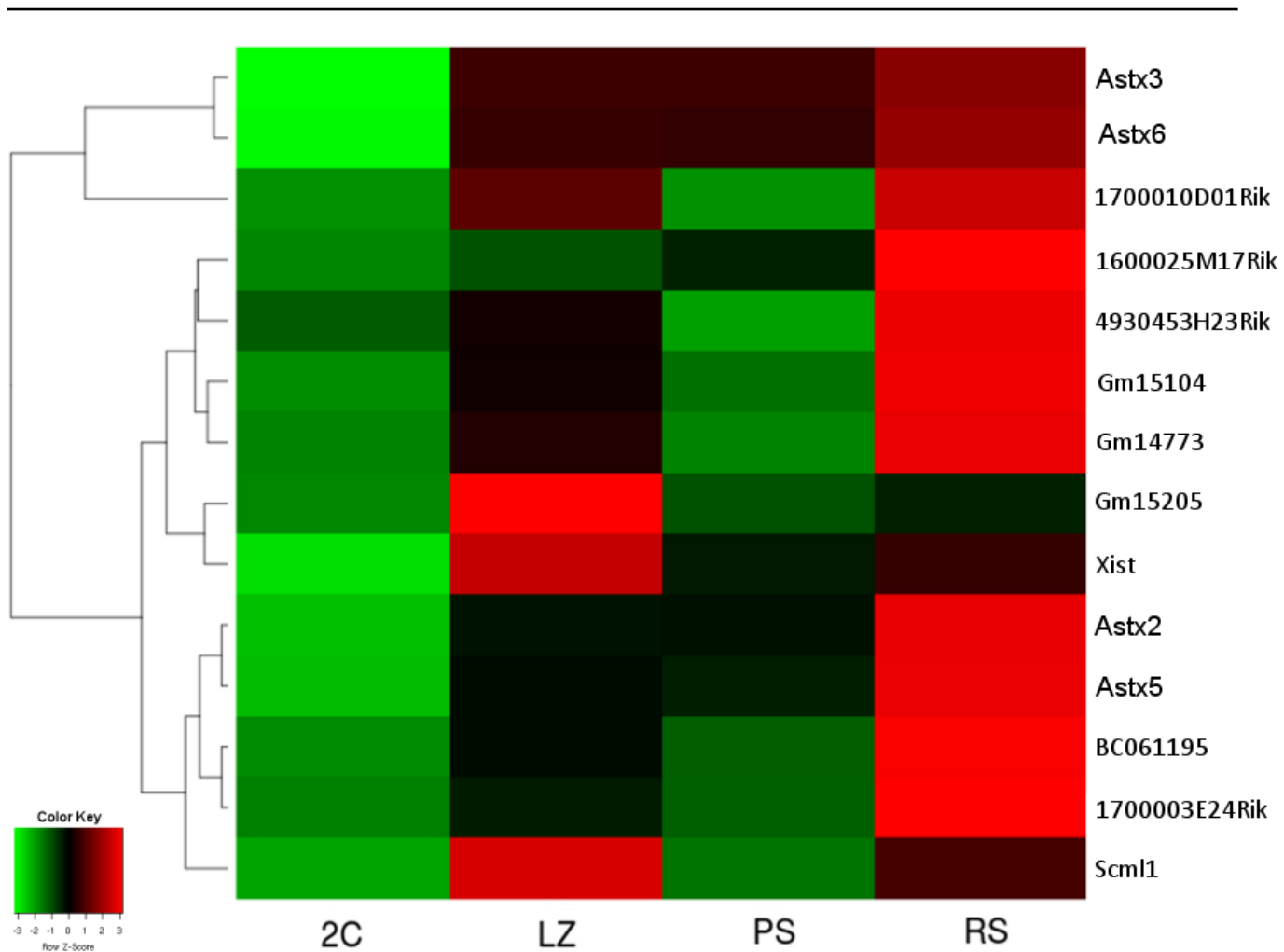
- 2C-3 (9E)
- 2C-2 (9E)
- 2C-1 (9E)
- 2C-4 (9E)
- 2C-5 (9E)
- 2C-10 (9E)
- 2C-11 (9E)
- 2C-12 (9E)
- 2C-13 (9E)
- 2C-14 (9E)
- 2C-15 (9E)
- 2C-16 (9E)
- 2C-17 (9E)
- 2C-18 (9E)
- 2C-19 (9E)
- 2C-20 (9E)
- 2C-21 (9E)
- 2C-22 (9E)
- 2C-23 (9E)
- 2C-24 (9E)
- 2C-25 (9E)
- 2C-26 (9E)
- 2C-27 (9E)
- 2C-28 (9E)
- 2C-29 (9E)
- 2C-30 (9E)
- 2C-31 (9E)
- 2C-32 (9E)
- 2C-33 (9E)
- 2C-34 (9E)
- 2C-35 (9E)
- 2C-36 (9E)
- 2C-37 (9E)
- 2C-38 (9E)
- 2C-39 (9E)
- 2C-40 (9E)
- 2C-41 (9E)
- 2C-42 (9E)
- 2C-43 (9E)
- 2C-44 (9E)
- 2C-45 (9E)
- 2C-46 (9E)
- 2C-47 (9E)
- 2C-48 (9E)
- 2C-49 (9E)
- 2C-50 (9E)
- 2C-51 (9E)
- 2C-52 (9E)
- 2C-53 (9E)
- 2C-54 (9E)
- 2C-55 (9E)
- 2C-56 (9E)
- 2C-57 (9E)
- 2C-58 (9E)
- 2C-59 (9E)
- 2C-60 (9E)
- 2C-61 (9E)
- 2C-62 (9E)
- 2C-63 (9E)
- 2C-64 (9E)
- 2C-65 (9E)
- 2C-66 (9E)
- 2C-67 (9E)
- 2C-68 (9E)
- 2C-69 (9E)
- 2C-70 (9E)
- 2C-71 (9E)
- 2C-72 (9E)
- 2C-73 (9E)
- 2C-74 (9E)
- 2C-75 (9E)
- 2C-76 (9E)
- 2C-77 (9E)
- 2C-78 (9E)
- 2C-79 (9E)
- 2C-80 (9E)
- 2C-81 (9E)
- 2C-82 (9E)
- 2C-83 (9E)
- 2C-84 (9E)
- 2C-85 (9E)
- 2C-86 (9E)
- 2C-87 (9E)
- 2C-88 (9E)
- 2C-89 (9E)
- 2C-90 (9E)
- 2C-91 (9E)
- 2C-92 (9E)
- 2C-93 (9E)
- 2C-94 (9E)
- 2C-95 (9E)
- 2C-96 (9E)
- 2C-97 (9E)
- 2C-98 (9E)
- 2C-99 (9E)
- 2C-100 (9E)
- 2C-101 (9E)
- 2C-102 (9E)
- 2C-103 (9E)
- 2C-104 (9E)
- 2C-105 (9E)
- 2C-106 (9E)
- 2C-107 (9E)
- 2C-108 (9E)
- 2C-109 (9E)
- 2C-110 (9E)
- 2C-111 (9E)
- 2C-112 (9E)
- 2C-113 (9E)
- 2C-114 (9E)
- 2C-115 (9E)
- 2C-116 (9E)
- 2C-117 (9E)
- 2C-118 (9E)
- 2C-119 (9E)
- 2C-120 (9E)
- 2C-121 (9E)
- 2C-122 (9E)
- 2C-123 (9E)
- 2C-124 (9E)
- 2C-125 (9E)
- 2C-126 (9E)
- 2C-127 (9E)
- 2C-128 (9E)
- 2C-129 (9E)
- 2C-130 (9E)
- 2C-131 (9E)
- 2C-132 (9E)
- 2C-133 (9E)
- 2C-134 (9E)
- 2C-135 (9E)
- 2C-136 (9E)
- 2C-137 (9E)
- 2C-138 (9E)
- 2C-139 (9E)
- 2C-140 (9E)
- 2C-141 (9E)
- 2C-142 (9E)
- 2C-143 (9E)
- 2C-144 (9E)
- 2C-145 (9E)
- 2C-146 (9E)
- 2C-147 (9E)
- 2C-148 (9E)
- 2C-149 (9E)
- 2C-150 (9E)
- 2C-151 (9E)
- 2C-152 (9E)
- 2C-153 (9E)
- 2C-154 (9E)
- 2C-155 (9E)
- 2C-156 (9E)
- 2C-157 (9E)
- 2C-158 (9E)
- 2C-159 (9E)
- 2C-160 (9E)
- 2C-161 (9E)
- 2C-162 (9E)
- 2C-163 (9E)
- 2C-164 (9E)
- 2C-165 (9E)
- 2C-166 (9E)
- 2C-167 (9E)
- 2C-168 (9E)
- 2C-169 (9E)
- 2C-170 (9E)
- 2C-171 (9E)
- 2C-172 (9E)
- 2C-173 (9E)
- 2C-174 (9E)
- 2C-175 (9E)
- 2C-176 (9E)
- 2C-177 (9E)
- 2C-178 (9E)
- 2C-179 (9E)
- 2C-180 (9E)
- 2C-181 (9E)
- 2C-182 (9E)
- 2C-183 (9E)
- 2C-184 (9E)
- 2C-185 (9E)
- 2C-186 (9E)
- 2C-187 (9E)
- 2C-188 (9E)
- 2C-189 (9E)
- 2C-190 (9E)
- 2C-191 (9E)
- 2C-192 (9E)
- 2C-193 (9E)
- 2C-194 (9E)
- 2C-195 (9E)
- 2C-196 (9E)
- 2C-197 (9E)
- 2C-198 (9E)
- 2C-199 (9E)
- 2C-200 (9E)
- 2C-201 (9E)
- 2C-202 (9E)
- 2C-203 (9E)
- 2C-204 (9E)
- 2C-205 (9E)
- 2C-206 (9E)
- 2C-207 (9E)
- 2C-208 (9E)
- 2C-209 (9E)
- 2C-210 (9E)
- 2C-211 (9E)
- 2C-212 (9E)
- 2C-213 (9E)
- 2C-214 (9E)
- 2C-215 (9E)
- 2C-216 (9E)
- 2C-217 (9E)
- 2C-218 (9E)
- 2C-219 (9E)
- 2C-220 (9E)
- 2C-221 (9E)
- 2C-222 (9E)
- 2C-223 (9E)
- 2C-224 (9E)
- 2C-225 (9E)
- 2C-226 (9E)
- 2C-227 (9E)
- 2C-228 (9E)
- 2C-229 (9E)
- 2C-230 (9E)
- 2C-231 (9E)
- 2C-232 (9E)
- 2C-233 (9E)
- 2C-234 (9E)
- 2C-235 (9E)
- 2C-236 (9E)
- 2C-237 (9E)
- 2C-238 (9E)
- 2C-239 (9E)
- 2C-240 (9E)
- 2C-241 (9E)
- 2C-242 (9E)
- 2C-243 (9E)
- 2C-244 (9E)
- 2C-245 (9E)
- 2C-246 (9E)
- 2C-247 (9E)
- 2C-248 (9E)
- 2C-249 (9E)
- 2C-250 (9E)
- 2C-251 (9E)
- 2C-252 (9E)
- 2C-253 (9E)
- 2C-254 (9E)
- 2C-255 (9E)
- 2C-256 (9E)
- 2C-257 (9E)
- 2C-258 (9E)
- 2C-259 (9E)
- 2C-260 (9E)
- 2C-261 (9E)
- 2C-262 (9E)
- 2C-263 (9E)
- 2C-264 (9E)
- 2C-265 (9E)
- 2C-266 (9E)
- 2C-267 (9E)
- 2C-268 (9E)
- 2C-269 (9E)
- 2C-270 (9E)
- 2C-271 (9E)
- 2C-272 (9E

|        | 2C         |            |            | LZ         |            |            | PS         |            |            | RS         |            |            |
|--------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
|        | JCB_1      | JCB_5      | JCB_5      | JCB_11     | JCB_12     | JCB_2      | JCB_13     | JCB_3      | JCB_8      | JCB_14     | JCB_9      | JCB_9      |
| JCB_1  | 1          | 0.91175584 | 0.85544446 | 0.6280226  | 0.55246422 | 0.56914239 | 0.61010299 | 0.59589408 | 0.58162048 | 0.55551333 | 0.48131195 | 0.41209398 |
| JCB_10 | 0.91175584 | 1          | 0.83088912 | 0.8309036  | 0.78544351 | 0.8038084  | 0.64200953 | 0.65994048 | 0.61793502 | 0.5784805  | 0.47856971 | 0.38407039 |
| JCB_5  | 0.85544446 | 0.83088912 | 1          | 0.7546629  | 0.72602674 | 0.73131213 | 0.57539081 | 0.62044688 | 0.59223761 | 0.53625011 | 0.4186418  | 0.35273136 |
| JCB_11 | 0.6280226  | 0.8309036  | 0.7546629  | 1          | 0.97065359 | 0.98453115 | 0.73221796 | 0.79903002 | 0.74927956 | 0.62181937 | 0.42216074 | 0.39638788 |
| JCB_12 | 0.55246422 | 0.78544351 | 0.72602674 | 0.97065359 | 1          | 0.99511092 | 0.60678696 | 0.69920804 | 0.64428355 | 0.54087776 | 0.27863332 | 0.27039761 |
| JCB_2  | 0.56914239 | 0.83088912 | 0.73131213 | 0.98453115 | 0.99511092 | 1          | 0.62799755 | 0.71427274 | 0.65656903 | 0.53004447 | 0.35446524 | 0.29542925 |
| JCB_13 | 0.61010299 | 0.64200953 | 0.57539081 | 0.73221796 | 0.60678696 | 0.62799755 | 1          | 0.97438568 | 0.97359562 | 0.86921644 | 0.7833178  | 0.77116158 |
| JCB_3  | 0.59558982 | 0.65940048 | 0.62044688 | 0.79903002 | 0.69920804 | 0.71427274 | 0.97438568 | 1          | 0.99287891 | 0.86434574 | 0.72702206 | 0.6962087  |
| JCB_8  | 0.58162048 | 0.61793502 | 0.59223761 | 0.74927956 | 0.64238355 | 0.65656903 | 0.97359562 | 0.99287891 | 1          | 0.86742674 | 0.73755064 | 0.76150965 |
| JCB_14 | 0.55551333 | 0.7848005  | 0.53625011 | 0.62181937 | 0.50487776 | 0.53004447 | 0.86921644 | 0.86434574 | 0.86742674 | 1          | 0.94204414 | 0.96243143 |
| JCB_9  | 0.48131195 | 0.47856971 | 0.4186418  | 0.42216074 | 0.39638788 | 0.34156454 | 0.7833178  | 0.71702206 | 0.73755064 | 0.94204414 | 1          | 0.96632491 |
| JCB_9  | 0.41209398 | 0.38407039 | 0.35273136 | 0.39638788 | 0.27039761 | 0.29542925 | 0.77116158 | 0.74929622 | 0.67150965 | 0.96624343 | 0.96632491 | 1          |

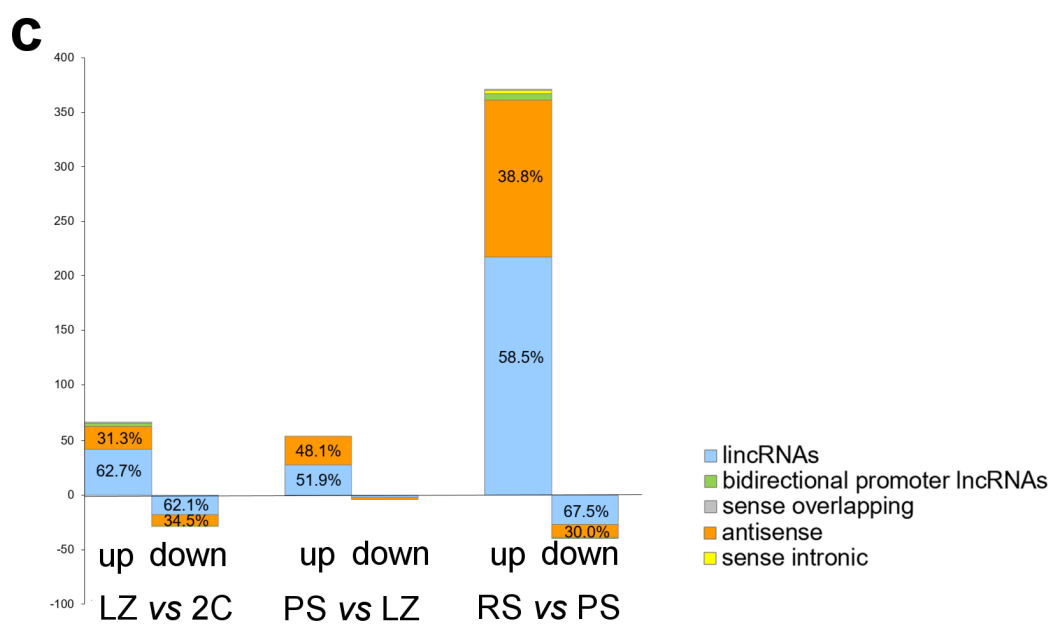
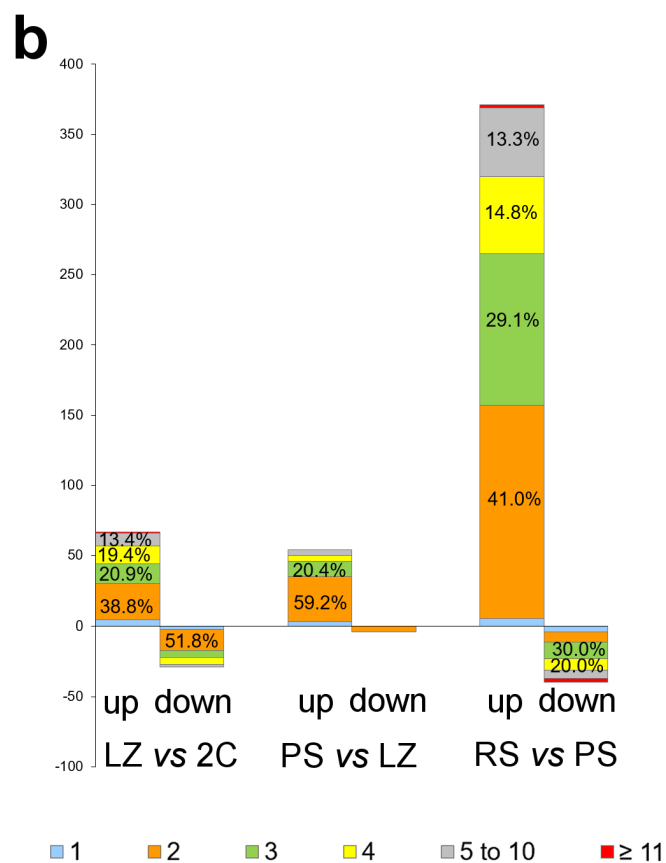
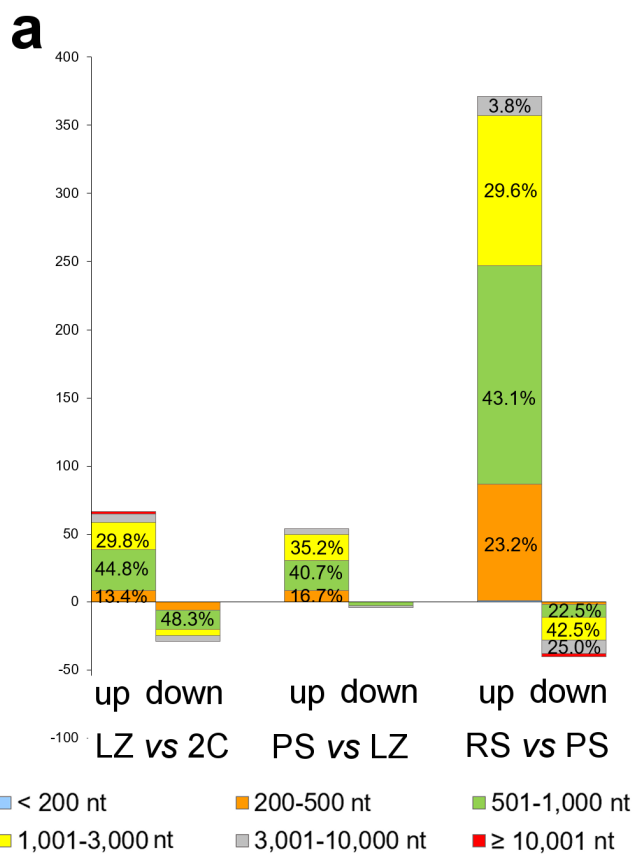
**Supplementary Figure S2. Reproducibility of the RNAseq results between biological replicas.** (a) Principal component analysis (PCA) of the RNAseq data. The data for the 2C, LZ, PS and RS samples (3 replicas each) are presented, showing that the RNAseq results are highly reproducible. (b) Correlation matrix for the 4 cell populations with 3 biological replicas each.

**a****b**

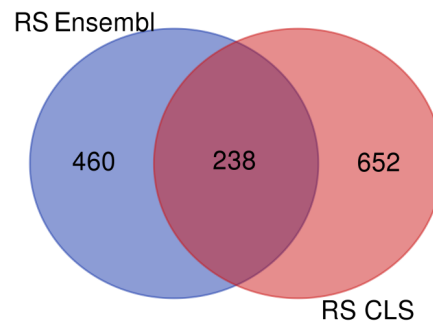
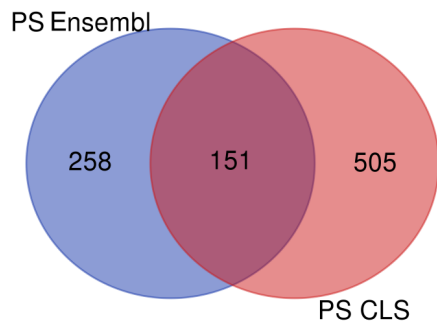
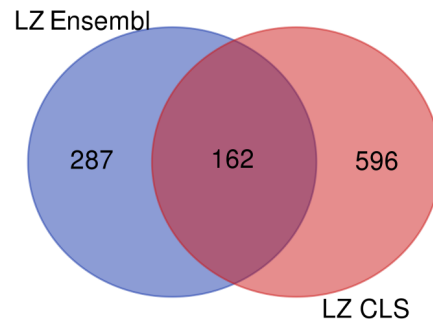
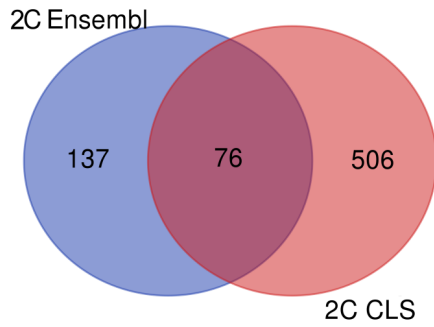
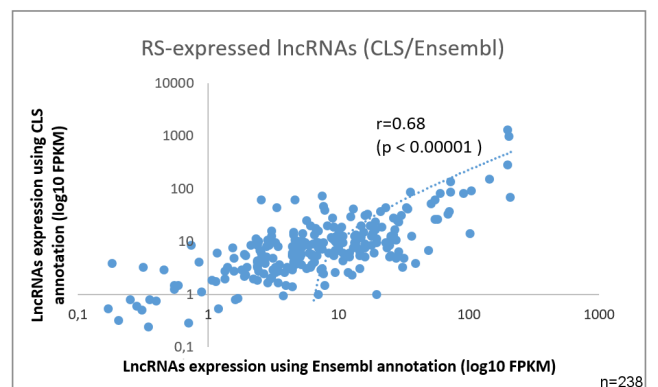
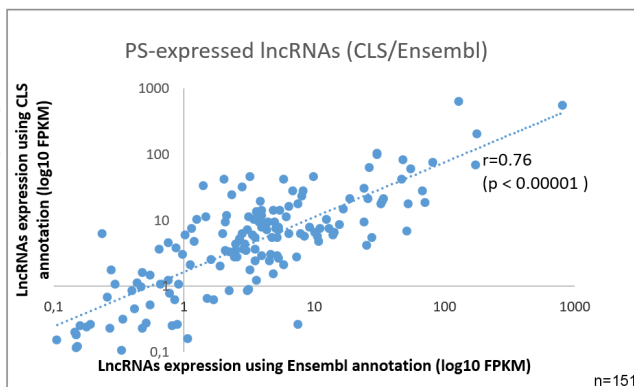
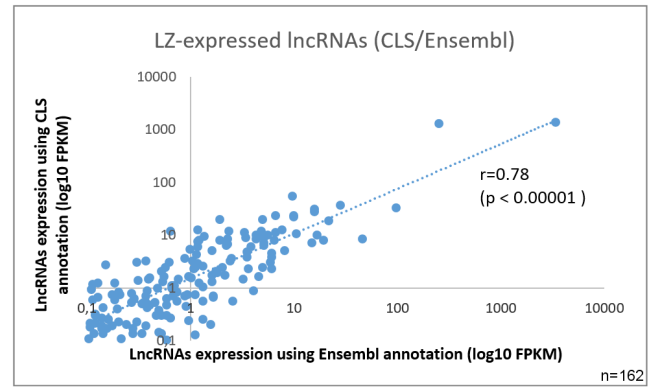
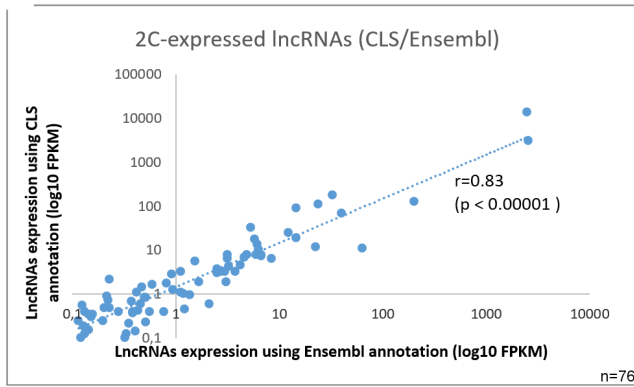
**Supplementary Figure S3. Comparisons of our lncRNAs expression data with those from other reports in which stage-specific-enriched cell populations were obtained by Staput.** The source were our lists of lncRNAs detected as expressed in each cell population, and generated with CLC Genomics Workbench. Scatter plots of  $\log_2$  RPKM are presented. Pearson's correlation coefficients with their p values are shown in each case. (a) Datasets were downloaded from Ref. [39]. For the 2C comparison, we crossed our 2C results (which contain spermatogonia and somatic cells) with their spermatogonia list. (b) Datasets were downloaded from Ref. [38]. Due to high differences in the criteria for collecting the 2C cell populations, only PS and RS were used for comparison in this case. Of note, as none of both studies had a LZ cell population for RNAseq, data from our LZ cell population could not be used for comparison.



**Supplementary Figure S4. Heat map of expression levels of X-linked lncRNA genes in the four cell populations.** All lncRNA genes detected as differential of at least one cell population were included. Average values of the 3 biological replicas for each testicular cell population are presented. Z-score values are coded on the green-to-red scale (high expression: red; low expression: green).



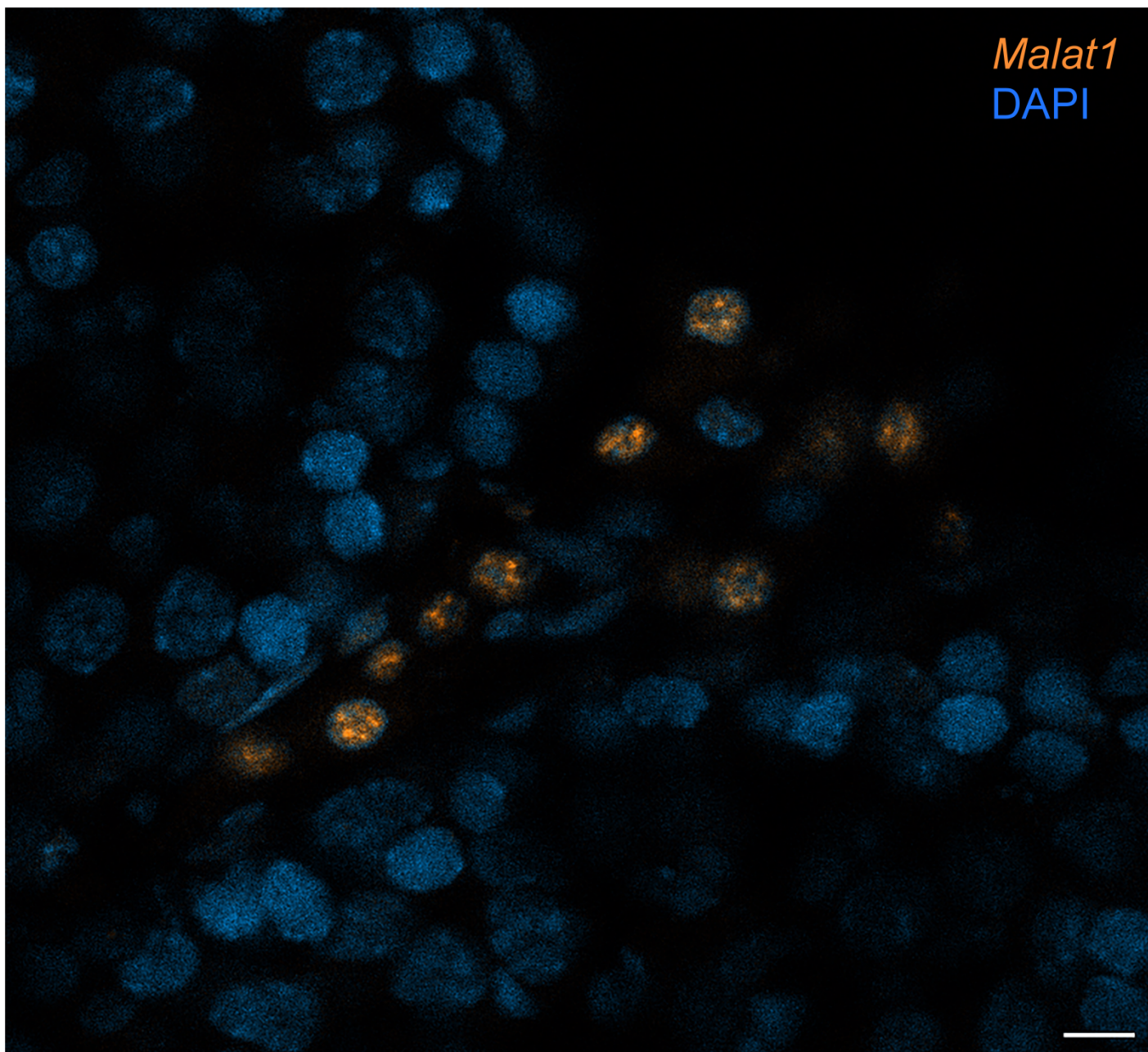
**Supplementary Figure S5. Characterization of DE lncRNA genes in the different spermatogenic stage transitions.** (a) Length distribution of the DE lncRNA genes in LZ vs 2C, PS vs LZ, and RS vs PS. (b) Exon number distribution of the DE lncRNA genes in the different spermatogenic stage transitions. (c) Biotype distribution of the DE lncRNA genes in the different stage transitions.

**a****b**

**Supplementary Figure S6. Coincidences in the expression of lincRNAs, using CLS and Ensembl annotations.** (a) Venn diagrams showing the number of coincidences in each cell population, for unique elements (i.e. redundancies were eliminated). (b) Scatter plots of  $\log_{10}$  FPKM for the coincidences. Pearson's correlation coefficients with their p values are shown in each case.

---

*Malat1*  
DAPI



**Supplementary Figure S7. Fluorescent *in situ* hybridization (FISH) showing the localization of lncRNA *Malat1* in mouse testis, as a positive control.** As expected, *Malat1* (orange) is localized in the nuclei of testicular interstitial cells. The section was co-stained with DAPI. Bar: 10  $\mu$ m.

## Supplementary Tables:

Download from <https://www.tandfonline.com/doi/suppl/10.1080/15476286.2019.1700332>.

### Supplementary Table S1. Sequencing yield of individual samples with two different pipelines.

#### TrimGalore-HISAT-StringTie-Ballgown

|                  | 2C1       | 2C2         | 2C3         | LZ1         | LZ2         | LZ3         | PS1         | PS2         | PS3         | RS1         | RS2         | RS3        |
|------------------|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------------|
| # raw reads      | 150246252 | 158005892   | 133016306   | 173271980   | 113303014   | 357455088   | 142606922   | 163342124   | 130996728   | 112925946   | 159968036   | 30748548   |
| # trimmed reads  | 1520832   | 3354539     | 737762      | 1570936     | 3871660     | 1564341     | 1215343     | 572954      | 475538      | 1587524     | 1587987     | 139723     |
| # remanent reads | 148725420 | 154651353   | 132278544   | 171701044   | 109431354   | 355890747   | 141391579   | 162769170   | 130521190   | 111338422   | 158380049   | 30608825   |
| # mapping reads  | 130833752 | 130231904,4 | 117092967,1 | 148263851,5 | 94296997,74 | 321867591,6 | 128440110,4 | 150089451,7 | 120275276,6 | 100917145,7 | 143983302,5 | 27599977,5 |
| % mapping reads  | 88        | 84          | 89          | 86          | 86          | 90          | 91          | 92          | 92          | 91          | 91          | 90         |

#### CLC Genomics Workbench

|                  | 2C1         | 2C2         | 2C3         | LZ1       | LZ2        | LZ3         | PS1         | PS2         | PS3         | RS1         | RS2         | RS3         |
|------------------|-------------|-------------|-------------|-----------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| # raw reads      | 150246252   | 158005892   | 133016306   | 173271980 | 113303014  | 357455088   | 142606922   | 163342124   | 130996728   | 112925946   | 159968036   | 30748548    |
| # trimmed reads  | 492452      | 26078262    | 6869309     | 514806    | 15495172   | 21787592    | 11303582    | 16390737    | 7097657     | 5546772     | 17530606    | 3378219     |
| # remanent reads | 149753800   | 131927630   | 126146997   | 172757174 | 97807842   | 335667496   | 131303340   | 146951387   | 123899071   | 107379174   | 142437430   | 27370329    |
| # mapping reads  | 138746895,7 | 120872094,6 | 117316707,2 | 159368493 | 88956232,3 | 313312040,8 | 122466625,2 | 138354730,9 | 116626195,5 | 97027821,63 | 129204992,8 | 24778358,84 |
| % mapping reads  | 93          | 92          | 93          | 92        | 91         | 93          | 93          | 94          | 94          | 90          | 91          | 91          |

**Supplementary Table S2. Identity and general characteristics of the 878 expressed lncRNA genes.**

**Supplementary Table S3. List of expressed lncRNAs (intergenic), obtained from the analysis using CLS annotation by Lagarde *et al.* [18] as a reference, Hisat/Stringtie/Balloon pipeline, and filtered by variance >1 and FPKM >0.1.**

**Supplementary Table S4. List of AS lncRNAs and their co-expressed overlapping coding genes, with their expression levels (FPKM) in the four testicular cell populations.**

**Supplementary Table S5. List of lincRNAs and their co-expressed neighbour coding genes at < 300 Kb, < 100 Kb, < 30 Kb, and < 10 Kb distance, with their expression levels (FPKM) in the four testicular cell populations.**

**Supplementary Table S6. List of primers used for RT-qPCR.**

| Gene/lncRNA   | Forward (5'-3')          | Reverse (5'-3')          |
|---------------|--------------------------|--------------------------|
| Tnp1          | CATCACAAGTGGGATCGGTA     | TCAAGAGAGGTGGAAGCAAGA    |
| 1700027A15Rik | CAACCAGAACCTCACTGCTG     | GCTTCTCAGCCAAAGCATGTG    |
| Dazl          | CTAGGCAGCCACCTCACG       | TCCATCCTAACATCAATTCCTCC  |
| AY702103      | TCCGACCTGCTGGGATTCTA     | ACTTTGGGTGCCTCAGATGG     |
| Actb          | GTGTGACGTTGACATCCGTAAAG  | CCGGACTCATCGTACTCCTG     |
| Rbakdn        | CAACAGCCCATCCTTTTGACG    | CTGGACCACATAGAAGCCACC    |
| Gapdhs        | GCTGGCATCCTTGCTTACAC     | CATTGAGGGCAATTCCAGCC     |
| Tmem147os     | CCACCCCTTCTCCACTACTC     | AAGCACCCAACCCTCTGTAG     |
| Pdzk1         | AAGCTGGTGTTCTGGCTGATG    | TCAAAGTGGCTGTCTCCCTC     |
| 4930442L01Rik | AGCCTTTAAGACGTGGGAGC     | TGTGGAAACGGACCTCTCTG     |
| Spata2l       | CCTGTTCACTGATGGCCTGTG    | GCACATAGCCCCCTGAGAAG     |
| 4933417D19Rik | GCACCACTGATGCCATACC      | GAGCCTATGCCTCCACCTTG     |
| Lrp8          | CATCAAACGGTGCAACCAGG     | TCATCACACACTTTCCCGCC     |
| Lrp8os1       | TCAAACCAAGGATCCCACCC     | CATGAGAGGATCTCGGCACG     |
| 1700047F07Rik | TCGCCTACTCCTGCGTCTAC     | ATCGGTGCGAAGGACATGG      |
| Fam181a       | AAGGTGAAGTGTCTGTCTGTGGTC | CTGGACGCCAGGTTGACAAAATTC |
| Gm29508       | AGACAGACAATCGGATGGG      | AGATGCAGGTCTTCGTTGG      |
| Rbm44         | CAGGTGCTGTGAGAAGACCAC    | GCTGCGCCTAACTCCTTCTTAG   |
| 1700020N18Rik | AGGATCGTGCCTACCTCAAC     | ACAGGAGAAAAGAGCAAGGGTC   |
| Akap1         | CGCTACGTGGACTATGGTGG     | CCAGTCGCACTGTAGCTTGTC    |
| C030037D09Rik | GAAGGCTACTACGTCACTGCAC   | GAGTGAAGGTAGCGCCATGTC    |
| Nkx2-6        | GGAACCCAGTAAAGCCTCGG     | CTCCAAGTCGGGAGATCGTG     |
| 1700092C10Rik | CTGTCCTCCATTGGTTCCAC     | GTCCCTTCTGGGTTTGCTTGG    |
| Lyzl4         | CGGCCTCTTTAGATTCTGTG     | TTTCTTGGCACACTGGATCG     |
| Lyzl4os       | TGGGAAGAGCTGAGAGTAGCC    | ATGAAGCAGAGGAAGGAGAATCA  |
| Surf4         | TGCTTTGGGCTGTTTGAATC     | GGTTGGGACACCAGCAAACA     |

# **Localización celular y sub- celular de lncRNAs en criosecciones testiculares**

## **4.4: CAPÍTULO 4**

#### **4.4.1: Sondas de ARN *Stellaris*® y tipos de microscopía**

A continuación, se seleccionaron algunos lncRNAs de expresión diferencial en las distintas etapas, para intentar caracterizarlos, en cuanto a su localización sub-celular, o interacción con otros transcriptos, como evidencias de posibles funciones.

Se diseñaron sondas de ARN tecnología *Stellaris*® contra algunos lncRNAs escogidos. Estas sondas consisten en hasta 48 oligonucleótidos de unos 20 bases (como mínimo 25 oligonucleótidos) que hibridan contra el ARN blanco, y están directamente conjugados a fluoróforos. Sin embargo, es recomendable que los transcriptos no sean menores a 800-1000 pb, ya que de lo contrario el blanco no es lo suficientemente grande para generar todas las sondas contra él. El programa utilizado para el diseño, incluido en la web de la empresa que las sintetiza, toma en cuenta también el genoma de referencia del organismo para evitar repetidos e hibridaciones inespecíficas. Las sondas *Stellaris*® prometen ser más sensibles y más específicas que las convencionales.

Para la elección de los transcriptos contra los cuales diseñar las sondas, nos basamos en que los mismos fueran de expresión diferencial de alguna/s de las etapas, que fueran dentro de lo posible lo suficientemente largos por lo explicado más arriba, y que sus niveles de expresión fueran relativamente elevados, de modo de aumentar las posibilidades de detección con las sondas. Esto último es muy relevante dado que, como hemos mencionado, la mayor parte de los lncRNAs suele expresarse a niveles bastante inferiores a los de los mRNAs, y por consiguiente pueden encontrarse por debajo de los límites de detección. Las longitudes de los transcriptos, así como los niveles de expresión en cada etapa, se muestran en la Tabla 5 (por las secuencias de los oligonucleótidos diseñados para cada una de las sondas, referirse a la Tabla 2 de Materiales y Métodos).

En particular, el lncRNA *Rbakdn* posee una longitud inferior a la recomendada, pero decidimos correr el riesgo de diseñar sondas contra él de todos modos, ya que resultó ser el lncRNA DE con mayor nivel de expresión en las RS y específico de testículo, de toda la lista.

Además, se seleccionó para la síntesis de sondas, un mRNA solapante con un lncRNA antisentido (*Kcnmb4*; ver más adelante).

**Tabla 5.** Características de los lncRNAs o genes seleccionados y de las sondas de ARN diseñadas.

| LncRNA                  | pb   | FPKM   | FPKM  | FPKM  | FPKM  | Biotipo     | Nº oligos | Marcaje   |
|-------------------------|------|--------|-------|-------|-------|-------------|-----------|-----------|
|                         |      | 2C     | LZ    | PS    | RS    |             |           |           |
| <i>Rbakdn</i>           | 588  | 2,5    | 32,6  | 331,0 | 500,0 | lincRNA     | 25        | Quasar670 |
| <i>1700001L05Rik</i>    | 2270 | 0,4    | 96,5  | 69,9  | 57,7  | lincRNA     | 48        | Quasar670 |
| <i>1700008K24Rik</i>    | 840  | 1,1    | 27,7  | 174,9 | 71,6  | lincRNA     | 29        | Quasar670 |
| <i>Kcnmb4os1</i>        | 824  | 0,9    | 12,7  | 142,2 | 298,4 | antisentido | 26        | Quasar670 |
| <i>Kcnmb4</i>           | 1367 | 0,4    | 2,5   | 28,0  | 25,1  | codificante | 39        | Quasar570 |
| <i>1110020A21Rik</i>    | 853  | 1,2    | 4,6   | 52,6  | 102,4 | lincRNA     | 25        | Quasar670 |
| <i>Malat1</i> (control) | 6938 | 2439,2 | 252,0 | 126,1 | 200,2 | lincRNA     | 48        | Quasar570 |

*A priori*, consideramos como opción para la visualización de resultados de FISH con las sondas *Stellaris*, la microscopía de alta resolución ya que podría ayudarnos a discernir, en el caso de que algunas de las sondas solaparan con señales de anticuerpos o entre ellas, si ambas señales realmente co-localizaban. Dadas las limitaciones en la resolución de la microscopía de fluorescencia confocal convencional (aprox. 200 nm), la localización de una molécula única se vuelve complicada [Schücker *et al.*, 2015]. Tecnologías como la de *dSTORM* (*direct stochastic optical reconstruction microscopy*), pueden alcanzar resoluciones incluso por debajo de los 20 nm [Endesfelder y Heilemann, 2015]. Los inconvenientes con esta tecnología son la preparación más trabajosa de las muestras, y la dificultad en el uso de varios fluoróforos en simultáneo. La microscopía de iluminación estructurada (SIM), por su parte, es un método versátil para obtener imágenes de muestras biológicas con una resolución que, dependiendo de la longitud de onda utilizada, varía entre 100 y 130 nm de resolución lateral y 280 y 350 nm de resolución axial [Demmerle *et al.*, 2017]. En trabajos anteriores en ovocitos de gallina se logró incluso visualizar con altísima precisión el desensamblaje del CS mediante SIM [Sciurano *et al.*, 2019].

Los fluoróforos más recomendados para tecnologías de alta o super-resolución son los Alexa Flúor o ATTO. Para *dSTORM* específicamente, funcionan mejor los Alexa Flúor 647 y 532 combinados [Schücker *et al.*, 2018]. Lamentablemente, las sondas *Stellaris* únicamente se pueden adquirir ya conjugadas a fluoróforos de un tipo diferente que son marca registrada de la empresa que las produce (principalmente Quasar670), por lo que no fueron óptimas para su uso incluso por SIM, porque decaían muy rápidamente debido a que el láser es muy intenso. Otro motivo que nos llevó a descartar esta tecnología para la visualización de los *RNA-FISH* es el hecho de que su utilización es más conveniente para visualizar señales en esparcidos celulares, por ejemplo, por su alta resolución, y no tanto en cortes de tejido. Únicamente logramos ver la señal para *Malat1* (gen control) en las células somáticas (ver a continuación), pero lamentablemente no visualizamos marca para ninguna sonda restante.

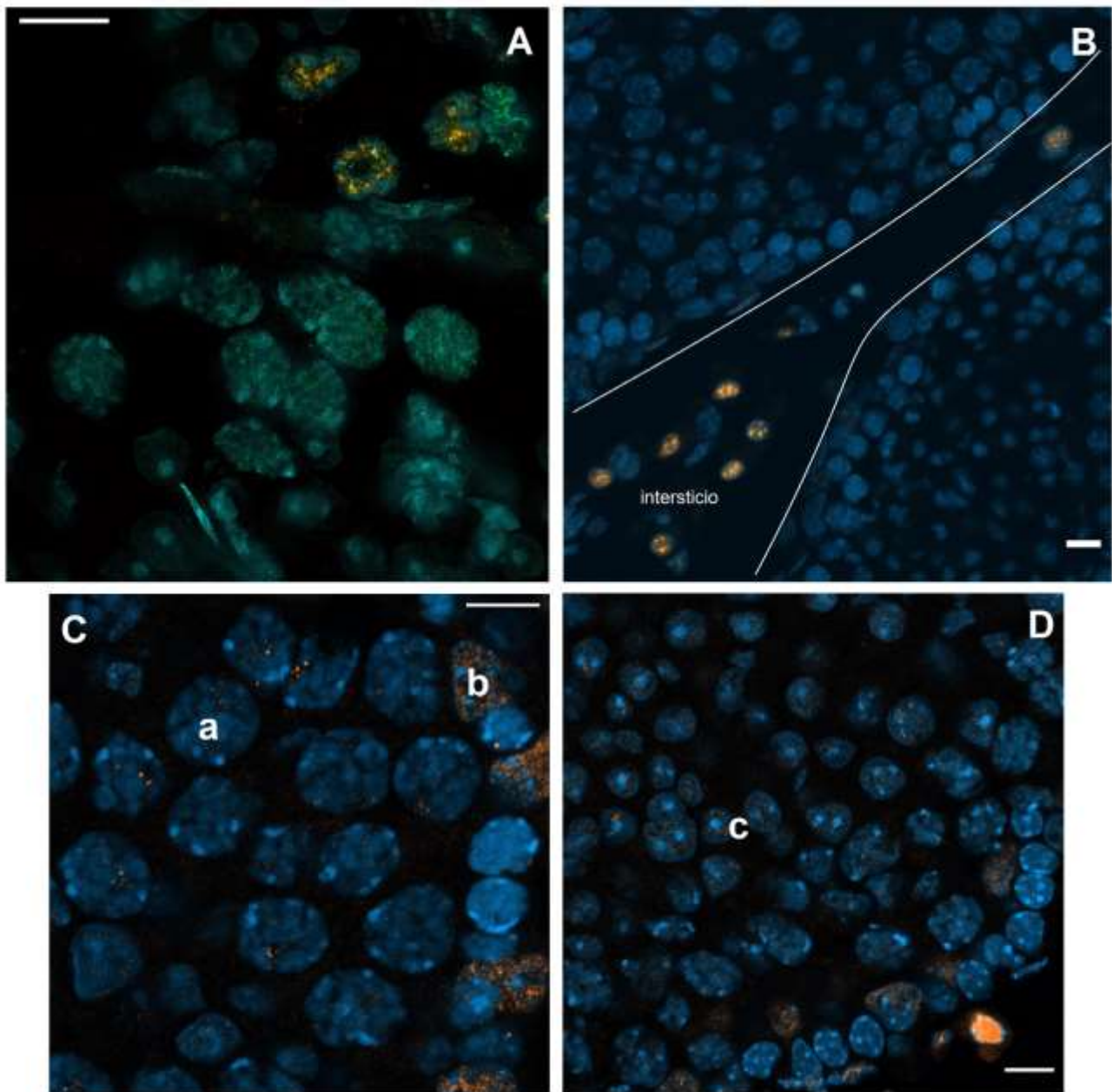
En consecuencia, y luego de varias pruebas, se concluyó que la mejor opción era el uso de microscopía confocal.

#### **4.4.2: *Malat1* como control, y su señal nuclear**

*Malat1* (*metastasis-associated lung adenocarcinoma transcript 1*; también conocido como *Neat2*) es uno de los lncRNAs más estudiados y mejor conservados en distintas especies de mamíferos. Presenta homólogos entre ratón y humano, aunque no se han encontrado homólogos presentes en especies no-mamíferas, y se expresa a niveles elevados en diversos tejidos. Se ha reportado que en células somáticas *Malat1* es retenido en el núcleo (por ejemplo, está presente en el sub-núcleo de fibroblastos de embriones murinos [Liu *et al.*, 2018]), involucrado en la formación de unos dominios nucleares enriquecidos en pre-mRNAs (*nuclear speckles*) [Johnsson *et al.*, 2014]. Se ha observado que las señales de hibridación *in situ* de *Malat1* claramente son intercromatínicas [Tripathi *et al.*, 2010]. Se sugiere que este lncRNA tendría una función en el metabolismo de mRNAs, como procesamiento alternativo y expresión génica, dada su íntima asociación con ciertos sitios de *splicing* [Hutchinson *et al.*, 2007; Nitsche & Stadler, 2017]; además, se lo ha visto desregulado en múltiples cánceres [Johnsson *et al.*, 2014].

La altísima expresión del lincRNA *Malat1* en todas las etapas de la espermatogénesis que estudiamos (ver Tabla 5) nos llevó a escogerlo como control positivo para los ensayos, además de ser una sonda pre-diseñada y verificada por la empresa que sintetiza las sondas *Stellaris*. Hasta donde sabemos, no existen reportes sobre los patrones de expresión de *Malat1* en el testículo. Nuestros ensayos de *RNA-FISH* con *Malat1* mostraron una clara señal nuclear excluyendo la cromatina en las células somáticas de los cortes de testículo, al visualizarse tanto en microscopio de alta resolución (Figura 16A) como en microscopio confocal (Figura 16B, C y D), lo que coincide con lo reportado previamente para otros tipos de células somáticas en otros tejidos.

La co-localización con un anticuerpo anti-SYCP1 (contra elemento transversal del CS), nos permite identificar de modo inequívoco los espermatoцитos I. Como se aprecia en la Figura 16A, la microscopía SIM no permite detectar claramente la señal de *Malat1* en los espermatoцитos (SYCP1-positivos). Sin embargo, por medio de microscopía confocal podemos ver a nivel de las células germinales, espermatoцитos (Fig. 16C) y espermátidas (Fig. 16D) una marca nuclear que excluye la cromatina más débil que en células somáticas, pero con un mismo patrón punteado. Por otra parte, la señal es sumamente intensa en células somáticas del intersticio (Fig. 16B). Además, se observa una señal muy contundente en las células de Sertoli, único tipo de células somáticas presente dentro de los túbulos seminíferos (ver Figura 16C). Estas células se reconocen muy fácilmente, con sus núcleos grandes e irregulares contra la base del túbulo, y sus nucléolos tripartitos intensamente teñidos [Santemma *et al.*, 1992]. La intensa marca en las células somáticas testiculares está de acuerdo con los elevados niveles de expresión detectados en la población 2C mediante *RNAseq* (ver Tabla 5). Este constituye el primer reporte acerca del patrón de expresión de *Malat1* en el testículo, y de la existencia de expresión de *Malat1* en células de la línea germinal.

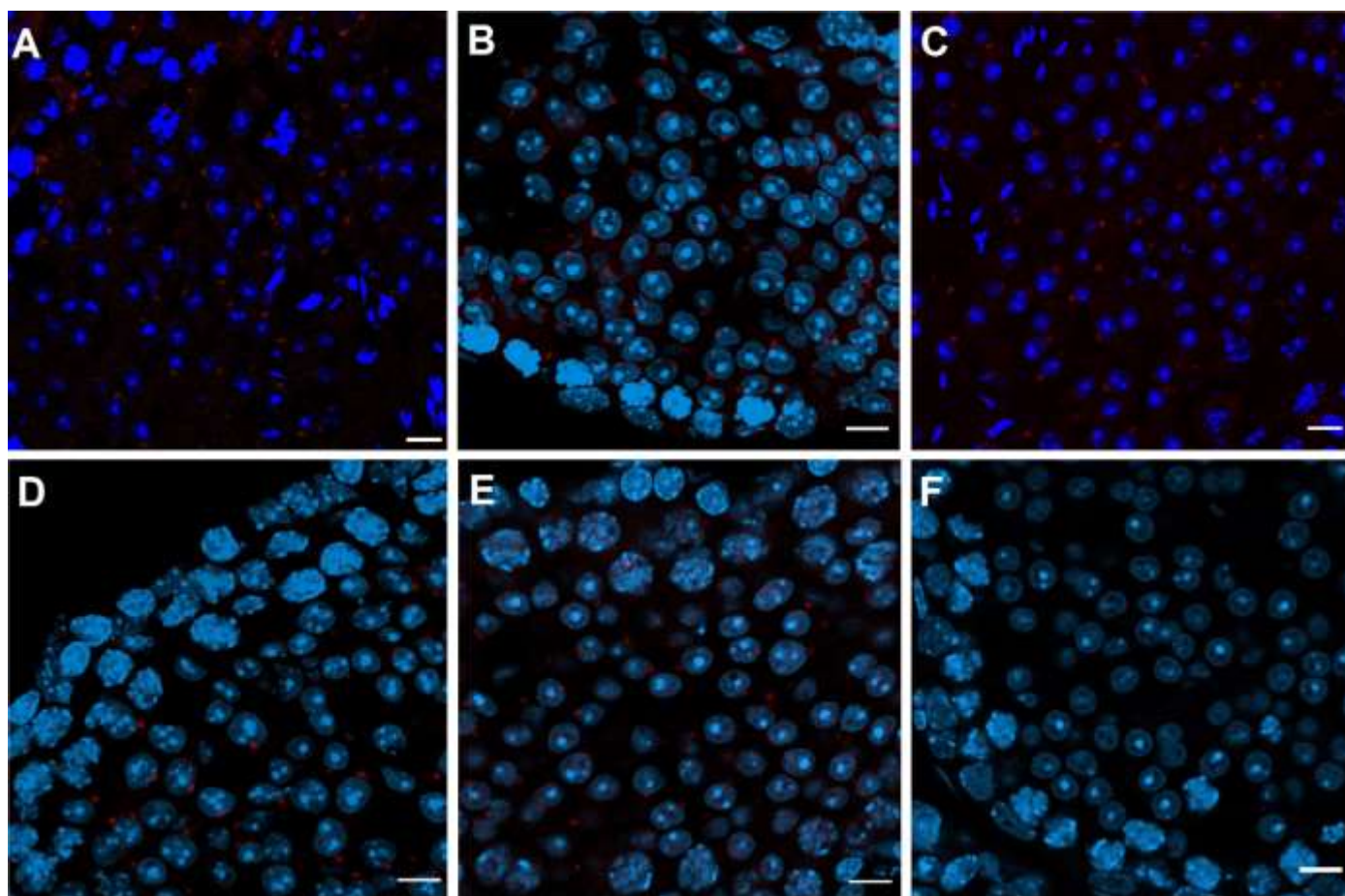


**Figura 16.** Localización de *Malat1*. Inmunofluorescencia de SYCP1 y *RNA-FISH* secuencial de *Malat1* sobre criosección de testículo de ratón vista en microscopía de alta resolución (SIM) (A). *RNA-FISH* de *Malat1* sobre criosecciones de testículo de ratón, vistas al microscopio confocal (B, C, D). En (B) se señala con líneas blancas los bordes de los túbulos. Espermatocitos I (a), células de Sertoli (b), espermátidas redondas (c). Anti-SYCP1 en verde, *Malat1* en naranja, DAPI en azul. Barra: 10  $\mu$ m.

#### 4.4.3: Localización de lncRNAs en el cuerpo cromatoide

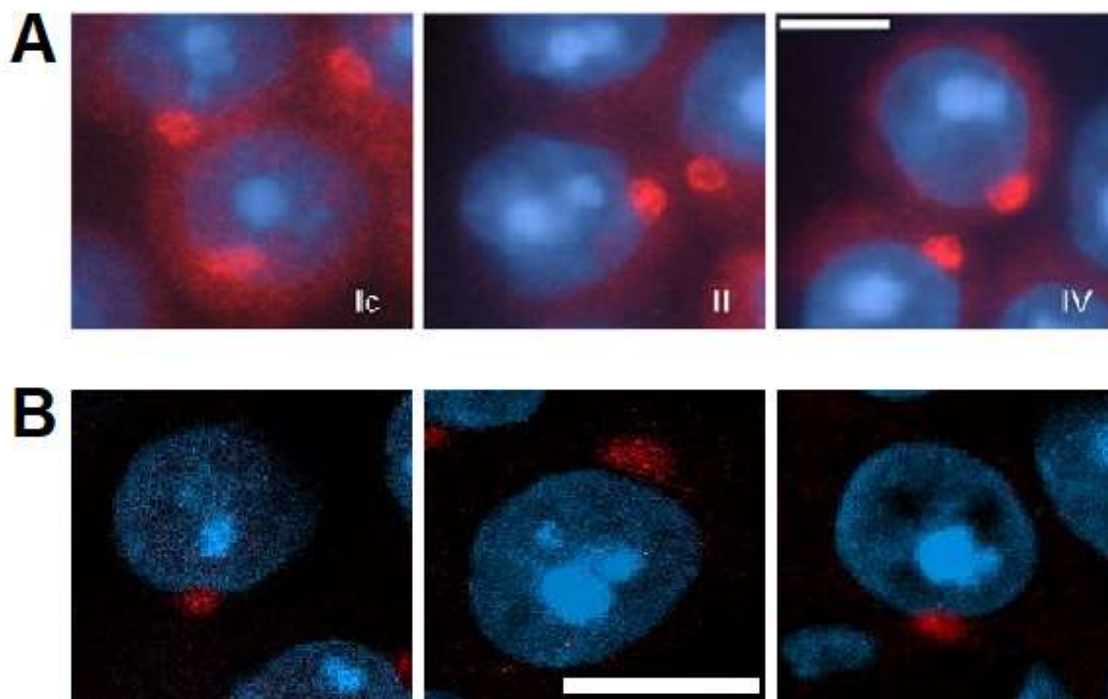
Las hibridaciones *in situ* sobre criosecciones de tejido testicular de ratón de 25 dpp dieron como resultado para los lncRNAs DE en línea germinal masculina testeados en este trabajo, una señal en el citoplasma en forma de punto perinuclear en RS, es decir, las células post-meióticas (Figura 17). Esto fue válido para todos los lncRNAs, sin importar su patrón de expresión: con pico en meiosis (*1700001L05Rik*, *1700008K24Rik*), o que comienzan a expresarse en la entrada en meiosis y continúen aumentando (*Kcnmb4os1*, *Rbakdn*, *1110020A21Rik*). Es importante señalar que esta señal, que no se detecta en los controles negativos (Figura 17F), es además completamente diferente a la observada para *Malat1*, descrita en la sección anterior.

La señal en el punto perinuclear en las espermátidas redondas recuerda de manera notable al cuerpo cromatoide (CC), un organelo perinuclear sin membrana específico de estas células (ver 2.1.3 en Introducción). Kotaja y colegas [2006] demostraron que la proteína MVH (*Mouse Vasa Homolog*, homóloga de la proteína Vasa de *Drosophila*) es un marcador específico de los CCs en ratón [Toyooka *et al.*, 2000], y es además un componente proteico de los gránulos germinales muy conservado en la evolución [Lehtiniemi & Kotaja, 2018]. La misma muestra una señal relativamente baja y difusa en citoplasma de células germinales hasta el estadio de leptoteno. Ya hacia cigoteno la señal se ve como muchos gránulos diminutos en el citoplasma celular. En el correr del paquiteno y diploteno la proteína se concentra en 3 a 5 gránulos localizados en la vecindad del núcleo celular, marcando posiblemente el cemento intermitocondrial, para ya en etapa de espermiogénesis localizarse fundamentalmente en un único gránulo perinuclear y esférico (CC) en las espermátidas redondas (Figura 18) hasta la etapa de elongación nuclear [Toyooka *et al.*, 2000; Noce *et al.*, 2001; Onohara *et al.*, 2010].



**Figura 17.** Localización sub-celular de lncRNAs. *RNA-FISH* de sondas *Stellaris* sobre criosecciones testiculares de ratón, vistas bajo microscopio confocal. DAPI en azul, sondas de ARN en rojo. *1700001L05Rik* (A), *1110020A21Rik* (B), *1700008K24Rik* (C), *Rbakdn* (D), *Kcnmb4os1* (E), control negativo (F). Barra: 10  $\mu$ m.

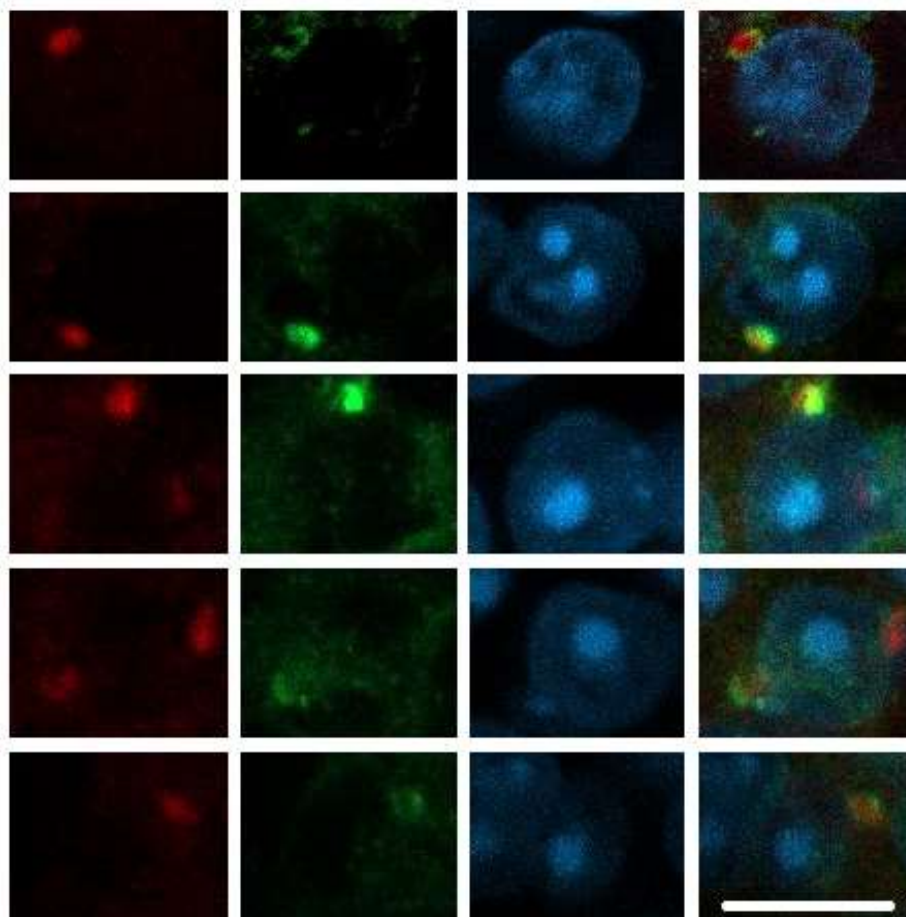
Para confirmar que se tratara de la misma estructura, realizamos una inmunomarcación para MVH con fluorescencia (IF) de manera previa y secuencial con el *RNA-FISH*. En la figura 19 podemos ver cómo la marca con el anticuerpo anti-MVH coincide con la marca de ARN de la sonda contra *Rbakdn*. El resultado para el resto de las sondas (*1700001L05Rik*, *1110020A21Rik*, *1700008K24Rik*, *Rbakdn*, *Kcnmb4os1*) es igual al de esta sonda, coincidiendo en todos los casos con la marca de MVH (datos no mostrados). Este resultado evidencia que varios lncRNAs con expresión diferencial en la espermatogénesis se localizarían en el CC.



**Figura 18.** Cuerpo cromatoide. (A) Localización de MVH en los cuerpos cromatoides, con anticuerpo anti-MVH. DAPI en azul, MVH en rojo. Barra: 5  $\mu$ m. Tomado y modificado de Kotaja *et al.* [2006]. (B) *RNA-FISH* con sondas de ARN *Stellaris* sobre criosecciones testiculares de ratón en nuestros ensayos. Ampliaciones de núcleos de la Figura 17 (de izquierda a derecha: *Kcnmb4os1*, *Rbakdn* y *1110020A21Rik*). DAPI en azul, sondas de ARN en rojo. Barra: 5  $\mu$ m.

#### 4.4.4: Co-localización de un mRNA y su lncRNA antisentido solapante, en el cuerpo cromatoide

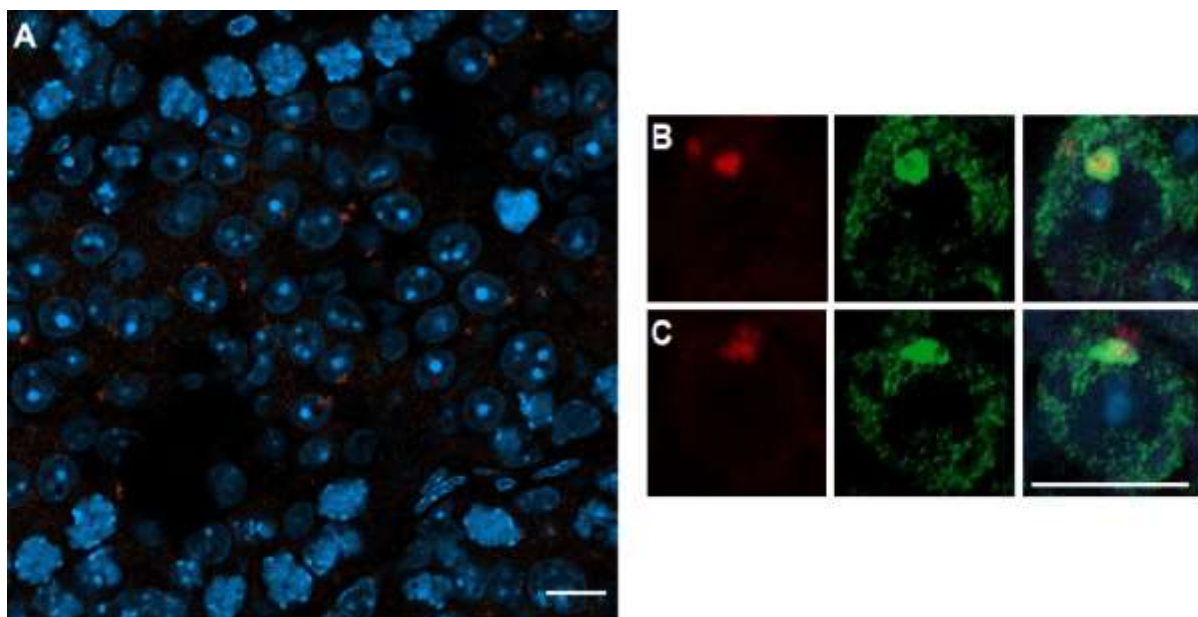
Más allá de los niveles significativamente elevados de co-expresión entre lncRNAs antisentido y sus genes codificantes solapantes, decidimos seleccionar uno de dichos pares para evaluar si co-localizaban, como modo de reforzar la teoría de una posible relación funcional entre ellos. Con este fin, seleccionamos uno de estos pares lncRNA antisentido/gen codificante solapante para el diseño de sondas de ARN *Stellaris* para *RNA-FISH*.



**Figura 19.** LncRNAs y cuerpo cromatoide. IF de MVH y *RNA-FISH* secuenciales sobre criosecciones de tejido testicular de ratón de 25 dpp, vistas al microscopio confocal. DAPI en azul, anti-MVH en verde, sonda de *Rbakdn* en rojo. Barra: 8  $\mu$ m

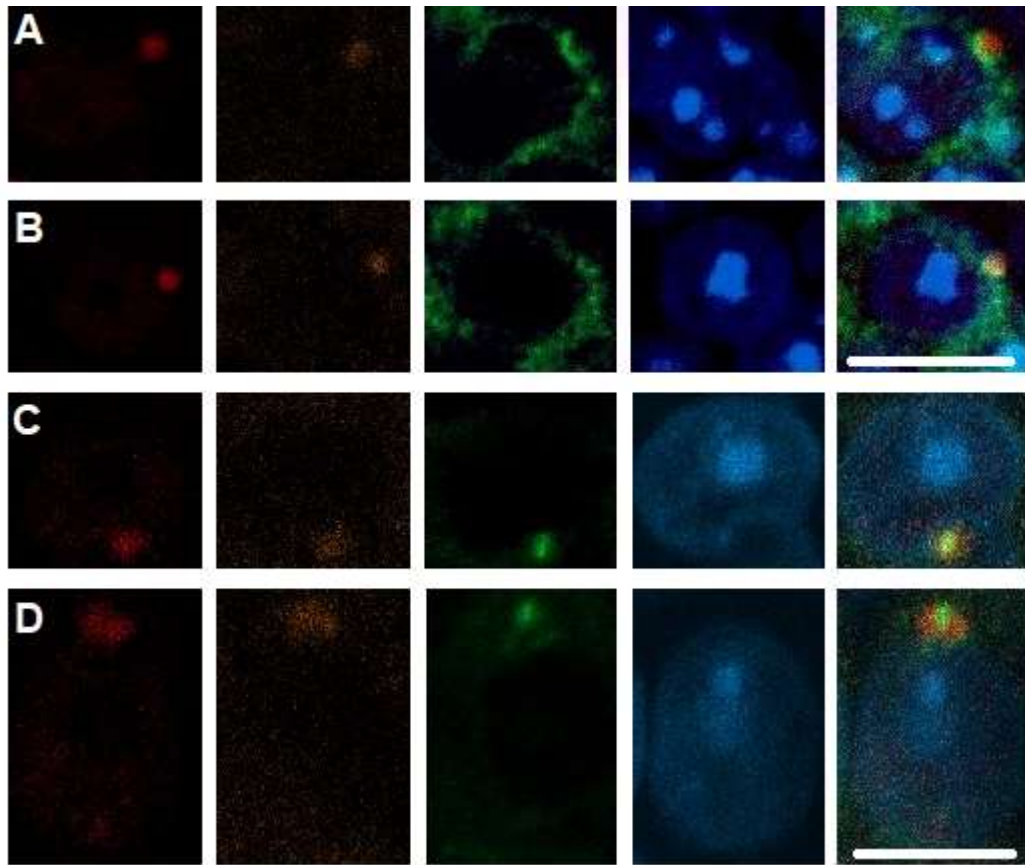
El lncRNA *Kcnmb4os1* es un antisentido al gen *Kcnmb4* y solapa parcialmente con éste (por un esquema de la disposición de ambos genes, ver Figura 22a). Este gen codifica para una unidad regulatoria de un canal de potasio activado por calcio [Weiger *et al.*, 2000]. Se le han identificado roles en distintos sistemas, incluyendo espermatozoides de ratón [Yang *et al.*, 2009] pero hasta el momento no se conoce bien su mecanismo funcional. La expresión de ambos transcritos es diferencial de testículo, cuando comparamos con otros tejidos del ratón (*Kcnmb4*: <https://www.ncbi.nlm.nih.gov/gene/58802>; *Kcnmb4os1*: <https://www.ncbi.nlm.nih.gov/gene/?term=kcnmb4os1>, respectivamente). Además, ambos transcritos, *Kcnmb4* y *Kcnmb4os1*, son diferencialmente

expresados en la espermatogénesis, incrementándose sus niveles de expresión a partir del paquiteno (ver Tabla 5). No solamente que la sonda de ARN diseñada contra el lncRNA *Kcnmb4os1* da una señal en el CC, sino que, llamativamente, la sonda de ARN diseñada contra el mRNA *Kcnmb4* también muestra una localización como marca perinuclear en las espermátidas redondas, además de un punteado disperso por el citoplasma, y coincide con la marca del anticuerpo anti-MVH en la inmunolocalización con fluorescencia (IF) (Figura 20).

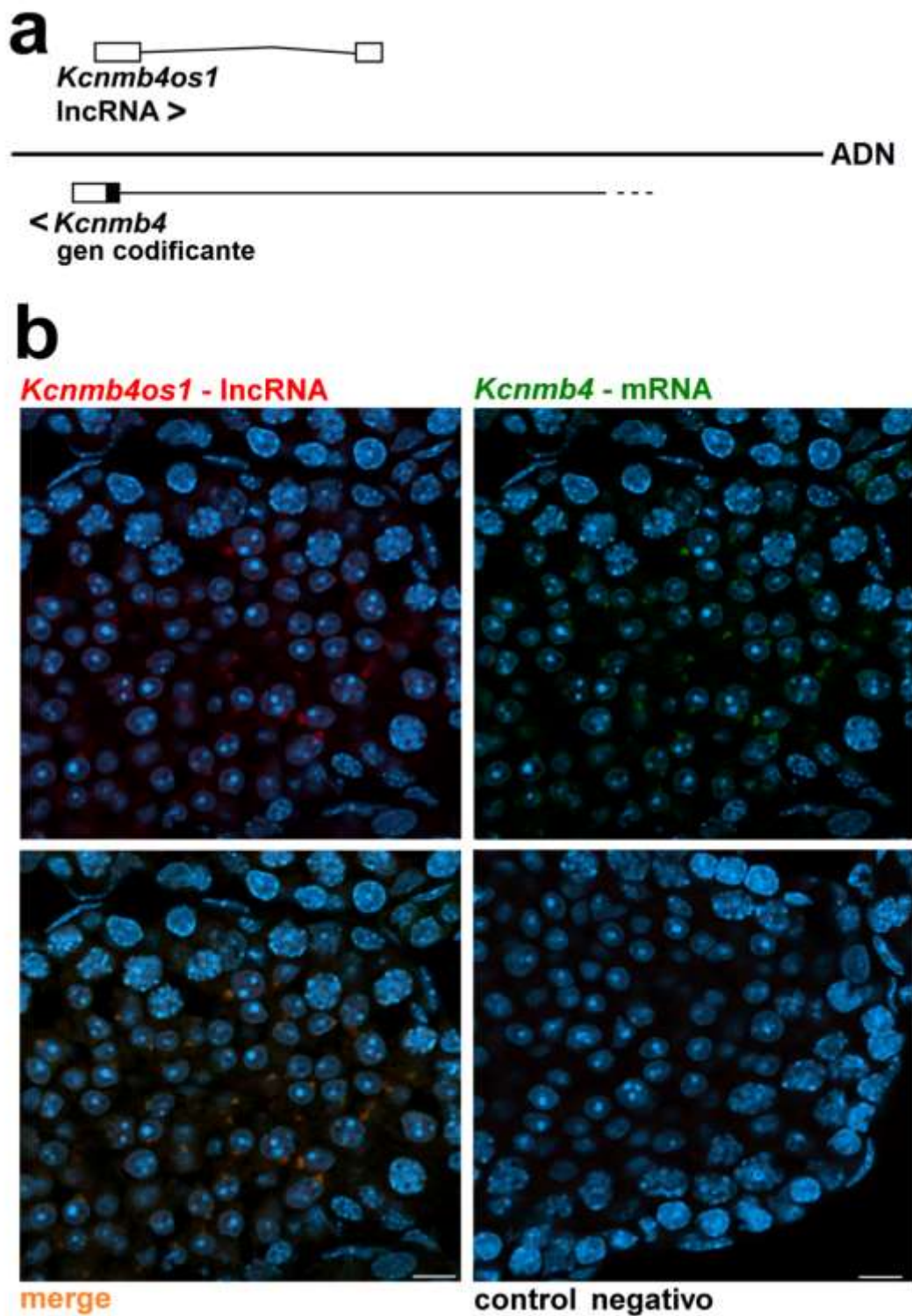


**Figura 20.** *Kcnmb4* en el cuerpo cromatoide. (A) *RNA-FISH* de *Kcnmb4* sobre criosección testicular de ratón, vista al microscopio confocal. (B, C) IF de MVH y *RNA-FISH* secuenciales sobre criosecciones de tejido testicular de ratón, vistas al microscopio confocal. DAPI en azul, anti-MVH en verde, sonda de *Kcnmb4* en naranja (A) y rojo (B, C). Barras: 10  $\mu$ m.

Observando que ambos transcriptos independientemente localizaban sobre el CC, realizamos una co-detección del gen y su lncRNA antisentido mediante IF con *RNA-FISH* sobre criosecciones de testículo. El resultado fue la co-localización de ambos ARNs sobre el CC (verificado con la marcación de MVH), con un coeficiente de correlación de 0,989 (Figuras 21 y 22). Esto indica que al menos en la etapa de inicio de la espermiogénesis, el mRNA y su lncRNA antisentido solapante localizan en el CC.



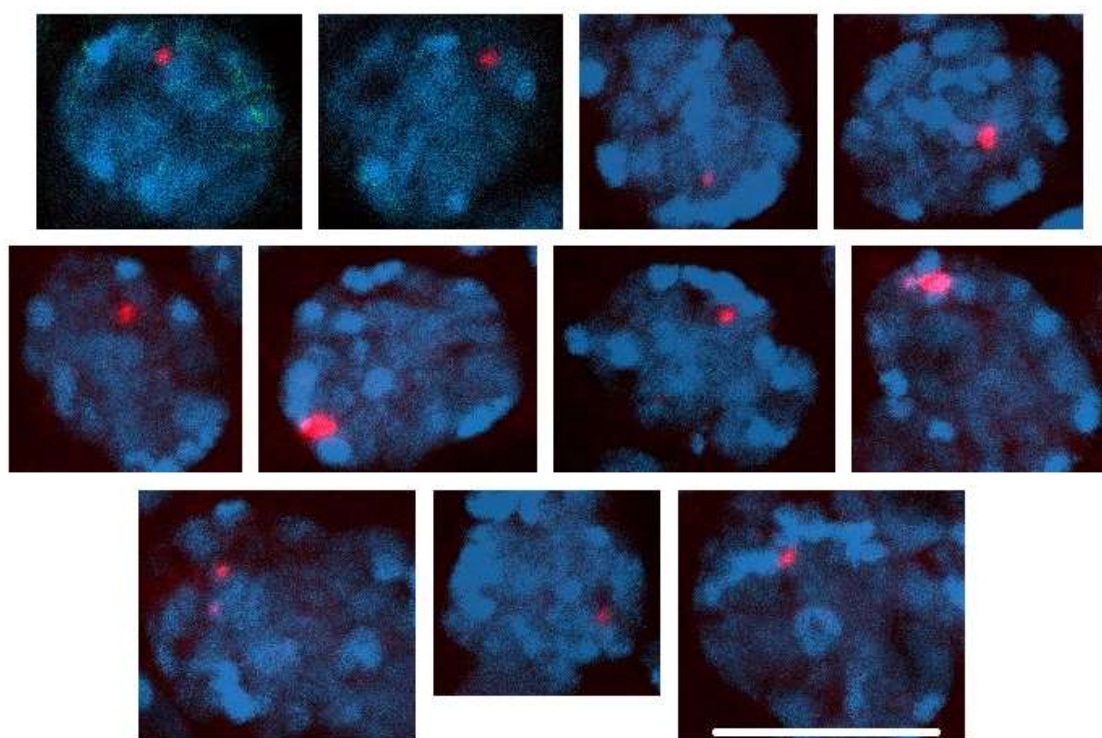
**Figura 21.** Co-localización de *Kcnmb4os1* y *Kcnmb4* en el cuerpo cromatoide. IF y *RNA-FISH* secuenciales de *Kcnmb4* y *Kcnmb4os1* sobre criosecciones de tejido testicular de ratón, vistas al microscopio confocal. DAPI en azul, MVH en verde, sonda de *Kcnmb4* en naranja y *Kcnmb4os1* en rojo. Barra: 9  $\mu$ m (A, B); 5  $\mu$ m (C, D).



**Figura 22.** Figura 6a y b, tomada y modificada del artículo anexo en sección 4.3 [Trovero *et al.*, 2020]. Diagrama de estructura y orientación de los transcritos *Kcnmb4* y *Kcnmb4os1* (a). RNA-FISH de *Kcnmb4* y *Kcnmb4os1* sobre criosecciones de tejido testicular de ratón, vistas al microscopio confocal (b). DAPI en azul, sonda de *Kcnmb4* en verde y *Kcnmb4os1* en rojo. Barra: 10 µm.

#### 4.4.5: Localización del lncRNA antisentido *Kcnmb4os1* en las células en profase meiótica

Además de la localización en el CC, sorprendentemente la sonda del lncRNA antisentido *Kcnmb4os1* muestra una señal en un punto (aunque en algunas células parecen ser dos puntos muy próximos, y en ocasiones incluso se observan dos puntos algo más separados) en el núcleo de los espermatoцитos, principalmente en PS (Figura 23). Es de destacar que a diferencia de la señal en el CC de las espermátidas, que se observó para las distintas sondas, una señal similar en el núcleo de los espermatoцитos no se observó para ninguna otra de las sondas ensayadas.

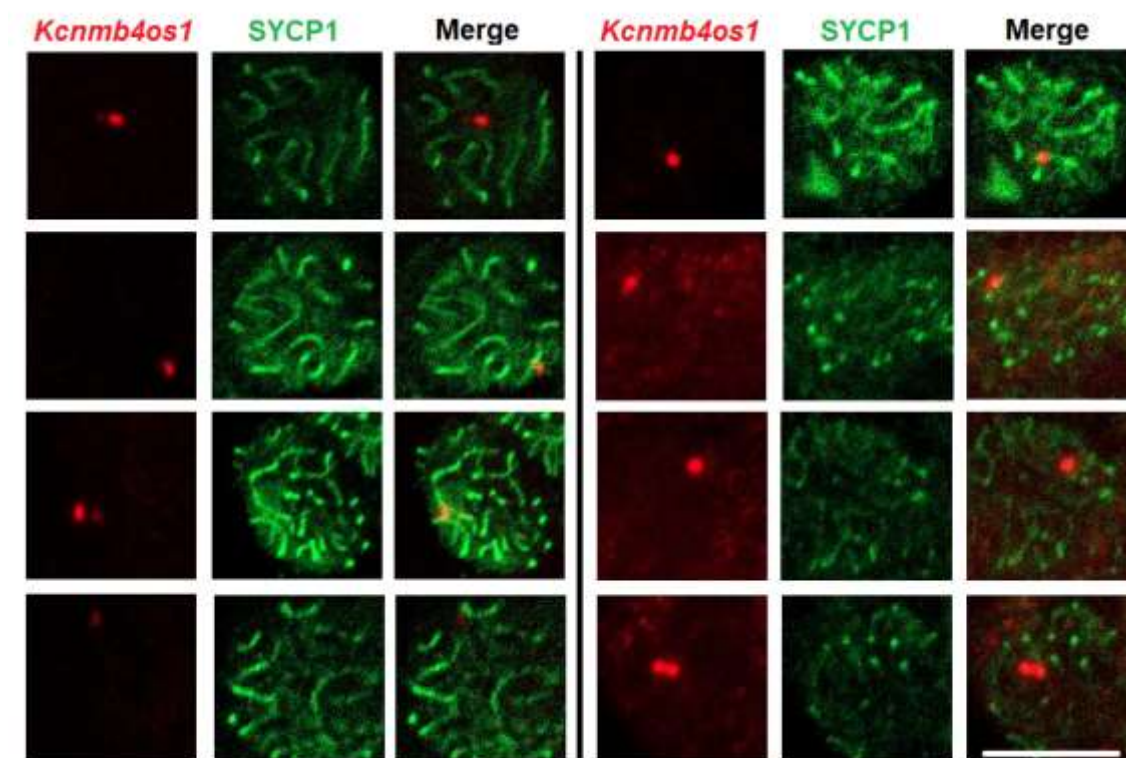


**Figura 23.** *Kcnmb4os1* en el núcleo de espermatoцитos paquiténicos. RNA-FISH de *Kcnmb4os1* sobre criosecciones de tejido testicular de ratón, vistas al microscopio confocal. DAPI en azul, sonda de *Kcnmb4os1* en rojo. Barra: 10  $\mu$ m.

Esta señal producida por la sonda *Kcnmb4os1* en el núcleo de los PS, de algún modo nos hace recordar a las producidas por los lncRNAs identificados por Ding y colaboradores [2012 (a)] en *S. pombe*, y que participan en el

reconocimiento de homologías y apareamiento homólogo durante la profase meiótica en esta levadura (ver sección 2.3.1 y Figura 7 en Introducción).

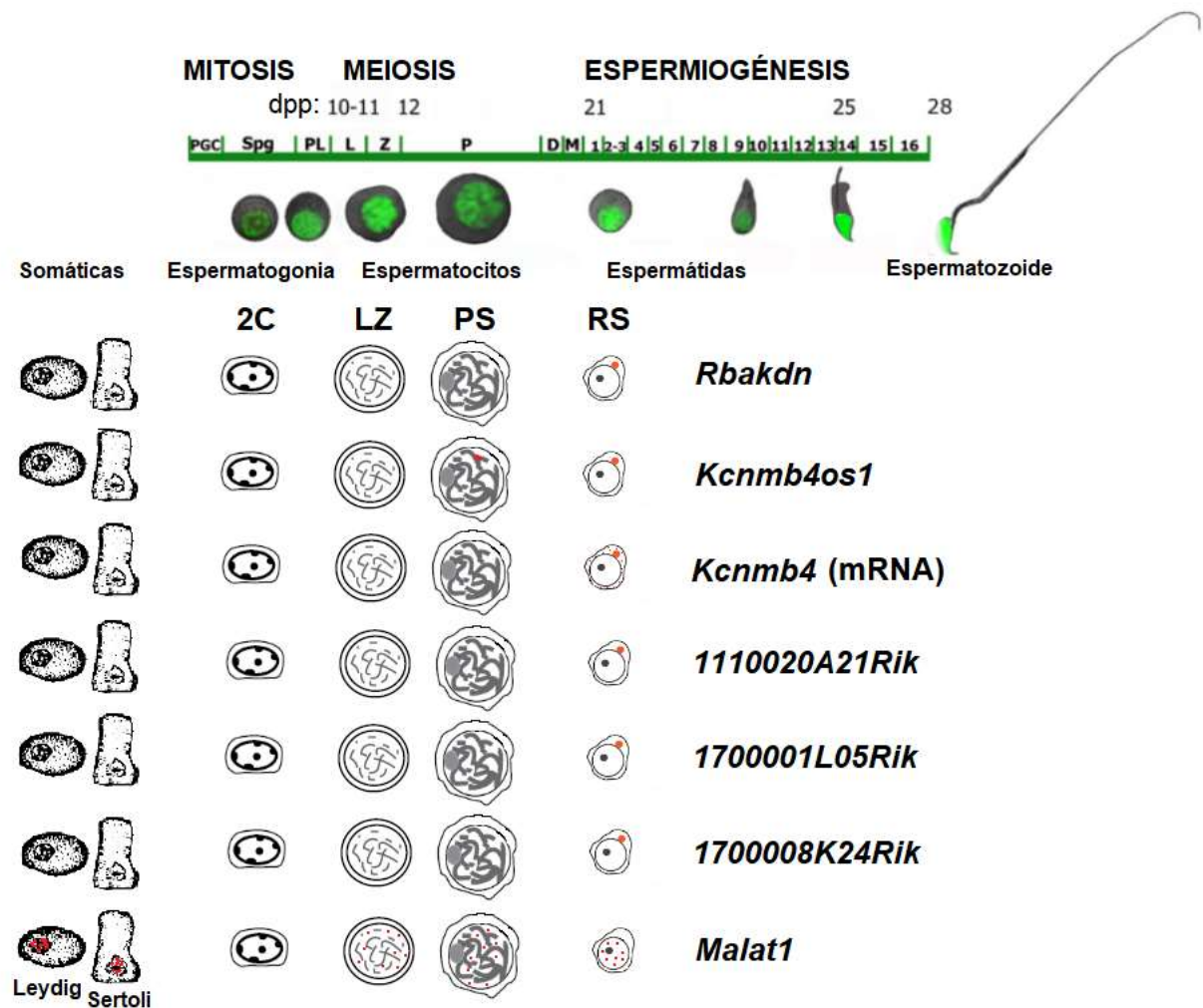
Una duda que nos surgió es si esta señal estaba de alguna forma vinculada a los CSs y, por lo tanto, próxima a los sitios de apareamiento. Para ello realizamos una IF previa al *RNA-FISH* secuencial con un anticuerpo anti-SYCP1, de modo de evaluar si se podría sugerir una posible vinculación de este lncRNA con el proceso de intercambio de material genético que ocurre en el paquiteno (Figura 24).



**Figura 24.** *Kcnmb4os1* y el complejo sinaptonémico. IF y *RNA-FISH* secuenciales de *Kcnmb4os1* sobre criosecciones de tejido testicular de ratón, vistas al microscopio confocal. Anti-SYCP1 en verde, sonda de *Kcnmb4os1* en rojo. Barra: 8  $\mu$ m.

Los resultados obtenidos hasta ahora no logran despejar las dudas sobre la interacción o al menos acercamiento de esta señal del ARN con los CSs, ya que en algunas células la señal de ARN parece estar sobre el CS, y en otras no queda claro.

En la figura 25 se representa la señal observada con las sondas *Stellaris* para los distintos lncRNAs estudiados.



**Figura 25.** Esquema de la espermatogénesis del ratón, donde se representan las señales obtenidas para las distintas sondas, en rojo. Tomado y modificado de da Cruz *et al.* [2016] y de Meikar *et al.* [2011].

# **Cultivos celulares de sobrevida: resultados preliminares**

## **4.5: CAPÍTULO 5**

Nuestro próximo objetivo es la realización de estudios funcionales, como se detalla en Conclusiones y Perspectivas (sección 6), y empleando como modelo el par lncRNA antisentido/mRNA *Kcnmb4os1/Kcnmb4*. Sin embargo, la realización de estos estudios en mamíferos es tremendamente dificultosa, dada la inexistencia de líneas de cultivo establecidas de células de germinales masculinas, y la imposibilidad de mantenerlas en cultivo.

Nos propusimos entonces, intentar poner a punto cultivos de sobrevida de espermatocitos por un lado y espermátidas por otro, de ratón. Se los denomina de sobrevida, ya que las células se mantienen viables a corto plazo (de 24 a 72 horas), sin duplicación celular (ni diferenciación), por lo cual terminan atrofiándose y mueren. Si bien el sistema distaría de ser el ideal, podría sin embargo permitirnos realizar algunos de los estudios que nos hemos propuesto.

La metodología utilizada para la preparación de suspensiones celulares mediante disgregación mecánica y posterior clasificación celular por citometría de flujo, desarrollada por nuestro grupo de investigación (ver 3.2.1), es ideal para la obtención de ácidos nucleicos íntegros y de excelente calidad, por ser rápida y no utilizar enzimas ni luz UV. Sin embargo, cuando buscamos viabilidad celular, esta metodología no resulta óptima, ya que la disgregación mecánica provoca estrés en la célula y hace que su viabilidad se vea comprometida. Con el objetivo de mejorar este parámetro, para desarrollar un cultivo celular de sobrevida de células germinales a partir de células purificadas por citometría, decidimos probar un protocolo de disgregación enzimática [Gaysinskaya & Bortvin, 2015], que si bien es más lento, también es menos agresivo para las células. En la tabla 6 se muestran los valores de viabilidad a tiempo 0 (a la salida del citómetro de flujo), y el tiempo de clasificación para las fracciones 4C (espermatozoides) y RS, a partir de suspensiones testiculares de ratones de 21 y 24-26 dpp, respectivamente.

En definitiva, el nuevo protocolo de disgregación enzimática del tejido resultó óptimo para la obtención de células viables para cultivo celular de sobrevida. Nos permitió obtener fracciones celulares de altísima viabilidad inicial y en tiempos mucho más cortos, lo que es muy beneficioso para los cultivos celulares, ya que se llevan a la estufa en tiempos menores luego de la clasificación, e inician la incubación con una viabilidad superior al 90%, todo lo cual es esperable que mejoraría su viabilidad a las 48 horas.

**Tabla 6.** Características de las fracciones celulares clasificadas por citometría de flujo, partiendo de suspensiones celulares obtenidas por dos métodos diferentes.

| Fracción celular/<br>protocolo | Viabilidad inicial | Tiempo de clasificación de<br>50.000 células |
|--------------------------------|--------------------|----------------------------------------------|
| 4C/ MM                         | 25-50% vivas**     | 120 minutos                                  |
| 4C/ E                          | 90-95% vivas+      | 30 minutos                                   |
| RS/ MM                         | 42-50% vivas**     | 35 minutos                                   |
| RS/ E                          | 90-95% vivas+      | 10 minutos                                   |

\*Refiere a la representación de ese tipo celular en el total de la suspensión testicular.

\*\*Viabilidad estimada mediante tinción con azul de tripán.

+ Viabilidad estimada mediante tinción con VDG/IP.

MM: protocolo de disgregación mecánica con *MediMachine*.

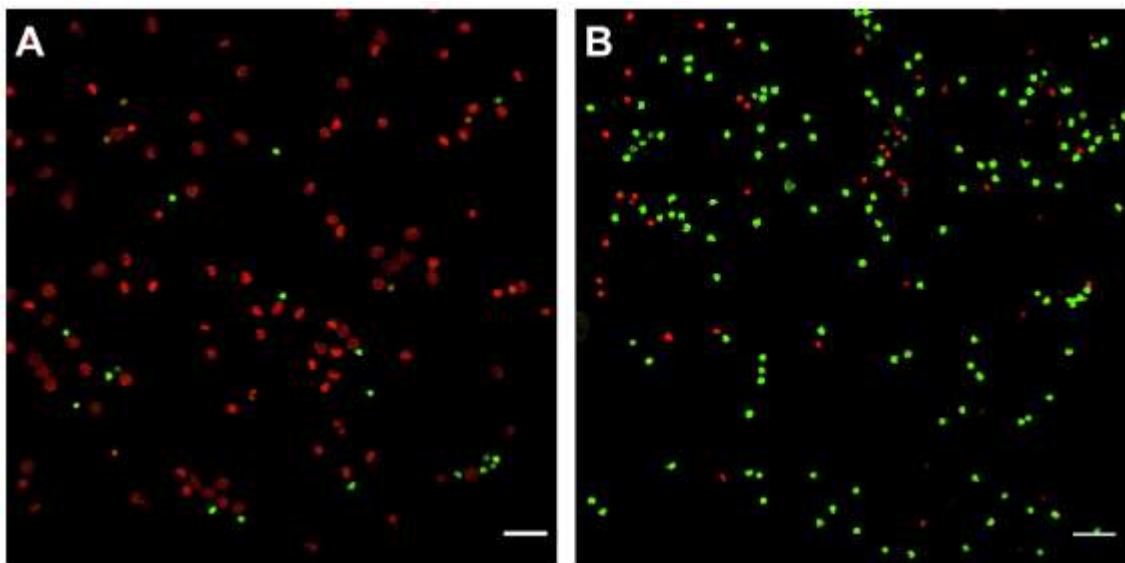
E: protocolo de disgregación enzimática.

Seguidamente, calculamos la viabilidad celular mediante tinción con VDG/IP, a las 24 y 48 horas de incubación. Las células viables están teñidas con VDG y se ven verdes al microscopio confocal. Sin embargo, y dado que en el medio de cultivo hay IP, una vez que mueren, el IP desplaza al VDG del núcleo, y entonces se ven rojas (Figura 26). La incubación en estufa de CO<sub>2</sub> de los cultivos celulares de las fracciones 4C y RS se ensayó en dos soportes: células adheridas en placas de cultivo pre-tratadas con poli-L-lisina, o suspendidas en tubos sin agitación. Los resultados de viabilidad para cada fracción ya sea en placa o en tubo, a los distintos tiempos se resumen en la tabla 7.

**Tabla 7.** Viabilidad de los cultivos celulares de sobrevida a las 24 y 48 horas.

| Fracción celular/ soporte | Viabilidad a 24 horas | Viabilidad a 48 horas |
|---------------------------|-----------------------|-----------------------|
| 4C/ placa                 | 42-62% vivas          | 14-22% vivas          |
| 4C/ tubo                  | 93-96% vivas          | 70-86% vivas          |
| RS/ placa                 | 18-21% vivas          | 3-5% vivas            |
| RS/ tubo                  | 76-77% vivas          | 37-40% vivas          |

Con estos resultados, podemos concluir que las células de la línea germinal se mantienen mejor en suspensión y no adheridas a la superficie. La viabilidad a las 48 horas se ve muy mejorada cuando ambas fracciones son incubadas en tubo. Sin embargo, 4C tiene mejor viabilidad que RS. En suma, estos resultados son prometedores, y consideramos que nos permitirán abordar los siguientes pasos de estudio con perspectivas de éxito.



**Figura 26.** Viabilidad de los cultivos de sobrevida. Ejemplo de imágenes para recuento de viables de la fracción 4C a las 48 horas de incubación, crecidas adheridas en placa (A) o en tubos en suspensión (B). Células viables en verde (VDG) y células muertas en rojo (IP). Barra: 50  $\mu$ m.

## **5. DISCUSIÓN**

### 5.1: Acerca de nuestros abordajes metodológicos

El primer objetivo de esta tesis fue la obtención de fracciones celulares puras de distintos estadios espermatogénicos del ratón. Para ello utilizamos una metodología desarrollada por nuestro grupo de investigación, basada en la disgregación mecánica del tejido testicular, y tinción con el colorante vital de ADN VDG, que es excitado en luz visible, evitando el uso de luz UV. Estas dos características del método permiten una mejor preservación de los ácidos nucleicos, óptimo para nuestro objetivo de extraer ARN de buena calidad a partir de las poblaciones celulares clasificadas. El hecho de clasificar las células mediante citometría de flujo, nos permitió además obtener altos niveles de pureza (~99%). Asimismo, es una metodología rápida cuyo objetivo es conservar la integridad de los ácidos nucleicos. Otros trabajos similares han utilizado testículo entero como muestra [Laiho *et al.*, 2013; Hong *et al.*, 2018], o métodos de enriquecimiento celular como *STA-PUT* o elutriación [Bao *et al.*, 2013; Soumillon *et al.*, 2013; Liang *et al.*, 2014; Lin *et al.*, 2016; Wichman *et al.*, 2017] que son más lentos y obtienen fracciones menos puras.

Recientemente, la secuenciación de células únicas (*single cell sequencing*) ha comenzado a ganar mucha atención, ya que permite analizar los perfiles de miles de células individuales en una población. Diversos trabajos han utilizado esta metodología para estudiar la expresión génica a lo largo de la espermatogénesis, principalmente a nivel de transcritos codificantes [e.g. Suzuki *et al.*, 2019]. Sin embargo, mientras en los estudios basados en aislamiento de poblaciones celulares la asignación de tipo celular se basa habitualmente en el análisis de una alícuota de cada población purificada (observación microscópica, inmunocitoquímica), en la secuenciación de células únicas sin clasificación previa, se basa en la expresión de genes marcadores. Dado el profundo desfase entre transcripción y traducción durante la espermatogénesis, consideramos que este criterio puede conducir a errores con respecto a la asignación de transcritos a los distintos tipos celulares. El único trabajo de secuenciación de célula única con células previamente clasificadas reportado, es el trabajo de Chen y cols. [2018], pero como hemos mencionado anteriormente, sincronizó la espermatogénesis artificialmente con probable alteración de la expresión génica en relación a lo que ocurre naturalmente.

Con el ARN total extraído de las 12 muestras, 4 poblaciones por triplicado (2C, LZ, PS y RS), construimos librerías para secuenciación hebra-específica. La importancia de que las librerías sean direccionales radica en el hecho de poder identificar la expresión propia de los lncRNAs que son antisentido y solapan con otros transcritos, ya sea codificantes de proteínas o ncRNAs. Como vimos en nuestros resultados, y también fue reportado en otros trabajos, existe una alta proporción de lncRNAs antisentido que son expresados en el testículo. Bao y cols. [2013] a partir de un estudio de microarreglos, encontraron que en el testículo total del ratón los lincRNAs y antisentido eran los dos biotipos más expresados. Otro trabajo ha sugerido que 21% de los lncRNAs expresados en la línea germinal del ratón serían antisentido a exones codificantes [Lin *et al.*, 2016]. Nuestros resultados elevan ese porcentaje, indicando que alrededor de un 40% de los lncRNAs expresados en los distintos tipos celulares del testículo serían antisentido. Esto estaría de acuerdo con estudios que mostraron que el 37% de los lncRNAs detectados en espermatozoides de gallo eran antisentido [Liu *et al.*, 2017 (a)]. Song y colaboradores [2018], por su parte, reportaron que de los lncRNAs más expresados en el testículo total del ratón adulto, la mayoría (un 54%) sería antisentido.

Por otra parte, nuestro estudio de expresión de lncRNAs a lo largo de la espermatogénesis fue efectuado mediante *RNAseq*. En otras publicaciones donde también se estudió la expresión de lncRNAs en el testículo utilizaron o bien microarreglos [Bao *et al.*, 2013; Sun *et al.*, 2013; Liang *et al.*, 2014; Hong *et al.*, 2018] cuyas sondas muchas veces tienen la desventaja de ser redundantes además de limitarse a la información contenida en el microarreglo, o emplearon secuenciación de librerías no direccionales [Chalmel *et al.*, 2014; Chen *et al.*, 2018] que no serían las óptimas para el estudio de estos lncRNAs.

Hace no tanto tiempo atrás, uno de los protocolos más clásicos de análisis de *RNAseq* consistía en alinear las lecturas al genoma de referencia con *TopHat*, utilizando el genoma indexado con *Bowtie*. Posteriormente, se procedía a ensamblar los transcritos con las lecturas alineadas mediante *Cufflinks*, y utilizar *Cuffdiff* para hallar los transcritos diferencialmente expresados entre las condiciones a analizar [Trapnell *et al.*, 2012]. En el análisis de los datos de secuenciación de este trabajo utilizamos para uno de los abordajes un protocolo muy similar y derivado del antes mencionado, pero con los paquetes sucesores

más modernos y optimizados: *HISAT*, *StringTie* y *Ballgown*, que tiene desempeño mejorado con respecto a los anteriores [Pertea *et al.*, 2016; Ulitsky, 2016]. Además, como un segundo procesamiento de datos de secuenciación alternativo empleamos el programa comercial CLC *Genomics Workbench*, para luego cruzar los datos y continuar trabajando con los coincidentes.

Utilizamos como criterios arbitrarios de corte para transcriptos expresados una varianza entre muestras  $> 1$  y una expresión mínima de 0,1 FPKM, que significa un nivel de restricción moderado. En otros trabajos similares se han marcado cortes más laxos como  $RPKM > 0$  ó  $FPKM \geq 0,01$  [Soumillon *et al.*, 2013; Ran *et al.*, 2016], o más estrictos como ser  $FPKM \geq 0,5$ ,  $TPM > 1$  ó  $TPM > 3$  [Ball *et al.*, 2016; Weng *et al.*, 2017; Chen *et al.*, 2018]. Si a los lncRNAs diferenciales nos referimos, establecimos como criterio de corte un FDR p-valor  $> 0,05$  y un  $\log_2 FC > |2|$ . La corrección del p-valor estadístico por el FDR (*False Discovery Rate*), es la tasa a la que las características significativas son verdaderamente nulas. Un FDR p-valor de 0,05 significa que de entre todas las condiciones que son consideradas como significativas, en promedio 5% de ellas es realmente nula [Storey & Tibshirani, 2003]. Estos dos cortes, principalmente el del valor de cambio ( $\log_2 FC > |2|$ , o sea diferencias en niveles de expresión de al menos 4x), pueden ser bastante restrictivos, pero decidimos darle preferencia a la confiabilidad y robustez de los datos en detrimento de la cantidad. Asimismo, no trabajamos con transcriptos sino con los genes de lncRNAs anotados en el genoma de la base de datos de *Ensembl*. Paralelamente, a partir de nuestro trabajo hemos identificado más de 60.000 transcriptos de lncRNAs nuevos (incluyendo nuevas variantes de *splicing*), el análisis de los cuales forma parte de otra tesis doctoral de nuestro grupo de investigación, actualmente en curso (datos no mostrados).

En suma, la metodología de clasificación celular empleada en este trabajo nos permitió obtener fracciones altamente puras de las poblaciones celulares a estudiar, sin ninguna estrategia que implique manipulación artificial del sistema (ej. sincronizaciones artificiales), e incluyendo la etapa de profase meiótica temprana, que por la dificultad de su aislamiento no suele incluirse en los estudios. Esto, sumado a la secuenciación específica de hebra, en conjunto con los programas actualizados de análisis y las estrictas condiciones utilizadas, contribuyen a la confiabilidad de los resultados obtenidos.

Con respecto a la anotación de los lncRNAs, la misma varía entre bases de datos y es muy dinámica; debido a su falta de potencial codificante y a sus niveles inferiores de expresión, la identificación y anotación de lncRNAs se vuelve más tediosa que la de genes codificantes. Más aún, dado que la mayoría de los catálogos de lncRNAs disponibles actualmente derivan de ensamblajes de datos de *RNAseq*, los extremos 5' de los transcriptos en general son imprecisos [Hon *et al.*, 2017]. La versión 93 del *Ensembl*, una base de datos del instituto EMBL-EBI que incluye anotaciones curadas de lncRNAs en genomas de diferentes organismos y también información de variantes y datos de regulación [Yates *et al.*, 2020], fue utilizada en este trabajo. Poco más de 8.800 lncRNAs estaban anotados para el genoma del ratón en dicha versión (la versión 100 actual ya suma unos 9.200 lncRNAs, cuya anotación es la misma del GENCODE versión M25). GENCODE es otro ejemplo de repositorio que incluye lncRNAs con anotaciones precisas, parte del proyecto ENCODE [Derrien *et al.*, 2012; ENCODE Project Consortium, 2012]. En los aproximadamente 8.800 transcriptos, están incluidas 7 categorías de lncRNAs (intergénicos, antisentido, sentido exónicos, sentido intrónicos, lncRNAs macro, lncRNAs solapantes en 3' y promotores bidireccionales) dejando afuera otros transcriptos no codificantes que en otras bases de datos no se excluyen, como por ejemplo los pseudogenes o intrones retenidos. Existen actualmente múltiples bases de datos de lncRNAs que aportan distinta información sobre los mismos: estructura secundaria (fRNAdb), promotores e imprintado (LncRBase), interacciones con miRNAs, mRNAs, otros ncRNAs o proteínas (ENCORI, CHIPBase v2.0), polimorfismos (lncRNASNP), asociación a enfermedades (LncRNADisease v2.0), entre muchas otras [Fritah *et al.*, 2014]. C-It-Loci es otra base de datos de *loci* enriquecidos en distintos tejidos para humano, ratón y pez cebra, y estiman unos 2.330 lncRNAs específicos de testículo [Weirick *et al.*, 2015].

Mediante los análisis bioinformáticos con los dos abordajes y los estrictos parámetros de corte antes mencionados, logramos detectar en el testículo un 10% (878 lncRNAs) de los lncRNAs totales anotados en el genoma de *Ensembl* v93.

## 5.2: Algunas consideraciones en relación a la comparación con otros trabajos

Simultáneamente con el desarrollo de esta tesis, se han realizado trabajos de caracterización de lncRNAs en el testículo para varias especies de vertebrados: en humano se encontraron 137 lncRNAs diferenciales y específicos de algún sub-tipo celular [Jan *et al.*, 2017]; en gallo se detectaron casi 2.600 lncRNAs y 124 diferenciales entre espermatozoides con alta y baja movilidad [Liu *et al.*, 2017 (a)]; en cerdo se identificaron unos 750 lncRNAs expresados y 100 diferenciales entre testículo maduro e inmaduro, mientras que en testículo post-natal se observaron alrededor de 770 lncRNAs diferenciales [Ran *et al.*, 2016; Weng *et al.*, 2017]; unos 1.110 lncRNAs fueron detectados como DE entre testículo de oveja inmaduro y maduro [Yang *et al.*, 2018]; en tortuga se reportaron casi 920 lncRNAs testículo-específicos [Ma *et al.*, 2020]; finalmente, de unos 17.400 lncRNAs, fueron identificados 230 lncRNAs diferenciales entre testículos de palomas con alta y baja movilidad espermática [Xu *et al.*, 2020]. En ratón también se han hecho otros trabajos de esta índole que, como ya fuimos mencionando a lo largo de esta sección, difieren de esta tesis en varios aspectos: la obtención de la muestra (ratones de una única edad y no ratones de edades crecientes), la purificación de las fracciones celulares (testículo entero o fracciones enriquecidas por *STA-PUT* o elutriación y no purificación por citometría de flujo), análisis de la expresión de lncRNAs (microarreglos o *RNAseq* no direccional, y no *RNAseq* hebra-específico), los *software* utilizados en los análisis bioinformáticos, etc.

A la hora de elegir qué trabajos re-analizar y comparar con nuestros resultados, nos decidimos por los de Lin *et al.* [2016] y Wichman *et al.* [2017] porque, si bien las fracciones celulares fueron obtenidas por *STA-PUT* en ambos casos y su pureza era inferior, varias de las poblaciones celulares con las que trabajaron son comparables a las nuestras (aunque como hemos mencionado en secciones anteriores, en ninguno de ambos trabajos se contaba con una fracción correspondiente a profase meiótica temprana). Además, se trataba de las únicas publicaciones en que se realizaron *RNAseq* hebra-específicas de testículo de ratón con poblaciones celulares enriquecidas en etapas de la espermatogénesis. Como ya hemos señalado, el re-análisis de los datos de Lin y Wichman lo pudimos realizar únicamente con el programa CLC, ya que el

trabajo de Lin *et al.* solamente contó con dos réplicas por población, y el paquete *Ballgown* de nuestro abordaje HSB es recomendado para más de tres réplicas por muestra. Más allá de esto, las coincidencias con el trabajo de Lin *et al.* [2016] para las fracciones PS y RS presentaron coeficientes de correlación de bajo a moderado estadísticamente significativos, mientras que los coeficientes de correlación para las coincidencias con el artículo de Wichman *et al.* [2017] fueron de moderado a alto con valor estadístico. Esto nos da la pauta que, más allá de las notorias diferencias existentes entre los estudios, el análisis *in silico* realizado puede aplicarse a otros datos, con buenos resultados.

En una nueva comparación, utilizamos nuestros datos analizados únicamente con el abordaje HSB y los alineamos a una anotación de lincRNAs de transcriptos más largos, como lo es la de CLS desarrollada por Lagarde y colegas [Lagarde *et al.*, 2017]. Si bien se detectó un mayor número de lincRNAs expresados en cada etapa, las proporciones de cada una se mantuvieron constantes, y las coincidencias llegaron a un 36% con coeficientes de correlación de alto a muy alto y estadísticamente significativos (como se mencionó en una sección anterior, este es un buen nivel de coincidencia tomando en cuenta que el porcentaje total de coincidencia entre los catálogos de transcriptos CLS y GENCODE para ratón es de 20%). Es preciso mencionar que estas comparaciones fueron difíciles debido a redundancias de los identificadores de genes [como ejemplo, que varios ids de la anotación de CLS correspondían a un mismo gen en la anotación de GENCODE (misma que *Ensembl*), entre otras]. Igualmente, intentamos eliminar todas las redundancias para los análisis. En general, se puede decir que las coincidencias en las comparaciones con otros trabajos y anotaciones genómicas fueron superiores a lo esperado.

Es importante señalar que el re-análisis de nuestros datos con la referencia CLS produjo nueva y valiosa información. En particular, permitió asignar más de 2.000 genes correspondientes a transcriptos largos y completos de lincRNAs, a etapas específicas de la espermatogénesis. En consecuencia, ahora disponemos de una lista de lincRNAs expresados en las distintas etapas de la espermatogénesis y con su longitud completa, lo que constituye un insumo muy útil para estudios futuros (Tabla Suplementaria S3 del artículo de la sección 4.3).

### 5.3: La meiosis y (principalmente) la espermiogénesis como fuentes de lncRNAs

Las espermatídas redondas (RS) fueron la población celular que expresó mayor cantidad de lncRNAs, seguida de las células meióticas (LZ y PS), y en último lugar las células pre-meióticas (2C). En el testículo humano, un estudio reciente del perfil de expresión de lncRNAs también indicó que la mayoría se expresaba en espermiogénesis, tanto para lncRNAs anotados como nuevos transcritos [Rolland *et al.*, 2019]. Song *et al.* [2018] en un estudio *in silico* de lncRNAs presentes en bases de datos, generaron listas de los 15 lincRNAs y 15 lncRNAs antisentido más expresados en testículo adulto de ratón. Catorce de esos 15 lincRNAs y 13 de los 15 antisentido están presentes en nuestra lista de 878 lncRNAs expresados. Además, 9 del total de 30 lncRNAs de las listas de Song y cols. (*Gm11837*, *1700018L02Rik*, *Proscos*, *1700001L05Rik*, *Smkr-ps*, *Malat1*, *Gm2762*, *1700052I22Rik*, *1700031L13Rik*) también figuran en nuestra lista top 10 de lncRNAs DE (Tabla 4). Bao *et al.* [2013], a partir de un estudio de microarreglos en el que emplearon testículo total de ratones de edades crecientes, también detectaron que la mayoría de los lncRNAs mostraban sus niveles más altos en las edades compatibles con la aparición de los espermatoцитos y las espermatídas. Nuestros resultados también son coincidentes con los de Soumillon y colaboradores [2013], que a partir de un trabajo en el que utilizaron poblaciones celulares enriquecidas mediante elutriación, señalaron que en primer lugar las espermatídas y en segundo los espermatoцитos eran las células de línea germinal que expresaban mayor número de lncRNAs. Contrariamente, nuestros resultados se oponen a los del trabajo de Liang *et al.* [2014], que empleando microarreglos estudiaron el perfil de expresión de lncRNAs de 4 poblaciones espermatoгénicas de ratón aisladas mediante *STA-PUT*, y encontraron que el máximo número de lncRNAs expresados se daba en espermatoгонias tipo A, y el mínimo en espermatoцитos paquiténicos.

Si nos referimos ahora a los lncRNAs expresados diferencialmente (DE) entre etapas, nuestros resultados indicaron que también la mayor cantidad de lncRNAs DE se dan en el ingreso a espermiogénesis, seguidos del inicio de la meiosis (tanto para el corte en  $\log_2(FC) \geq |2|$ , como cuando el umbral es llevado

a  $\log_2 (FC) \geq |1|$ , o sea cuando la diferencia en niveles de expresión es de al menos 2x). Laiho y cols. [2013] trabajaron con testículo de ratones en edades crecientes durante la primera onda espermatogénica y encontraron unos 108 lncRNAs reprimidos y 947 lncRNAs sobre-expresados, principalmente en edades coincidentes con la aparición de los espermatoцитos y con el inicio de la espermiogénesis, pero no avanzaron en su caracterización. Oponiéndose a todos estos resultados, Wichman *et al.* [2017] identificaron un mayor número de lncRNAs DE en el pasaje de espermatogonias a paquiteno que de paquiteno a espermátidas (vale destacar que si bien este fue uno de los trabajos que utilizamos para comparar con nuestros datos, la comparación la realizamos contra los datos crudos).

Cuando nos preguntamos cuál o cuáles pueden ser los motivos para que las RS expresen mayor número de lncRNAs incluso diferencialmente, podemos pensar en varias cosas. Por un lado, sabemos que en etapas posteriores de la espermiogénesis tiene lugar una represión transcripcional muy importante debido a la altísima compactación de la cromatina, que puede llevar a que en estas etapas previas se transcriban ARNs de manera amplia y promiscua. Esto sucede tanto con transcriptos testículo-específicos como con ARNs de mantenimiento celular general, aunque a niveles menores, que luego se almacenarán y traducirán en el momento del proceso espermatogénico en que deban actuar [White-Cooper & Davidson, 2011]. Resulta interesante que un análisis multidimensional del transcriptoma de lncRNAs comparativo a lo largo de cinco estadios del desarrollo de cuatro órganos diferentes (incluido testículo) para tres especies (una de ellas, el ratón), reveló que los lncRNAs se expresan a niveles abrumadores en los testículos de adulto y de edad avanzada, consistente con la transcripción permisiva durante la espermatogénesis. A su vez, son menos abundantes en estadios embrionarios y neonatales, coincidiendo con la necesidad de una regulación estricta de la expresión y menor ruido transcripcional en el desarrollo temprano del órgano [Darbellay & Necsulea, 2020]. Por otro lado, pero en la misma línea de pensamiento, Soumillon *et al.* [2013] a partir de estudios epigenéticos a nivel masivo, han sugerido que la amplia transcripción de ARNs en el testículo se debería a un estado permisivo de la cromatina en la línea germinal. Estos mismos autores postulan que ello puede ser resultado de un extensivo remodelado de la cromatina, característica

que no evaluamos en nuestro trabajo. Por otra parte, tampoco podemos descartar que al menos varios o muchos de estos lncRNAs estén cumpliendo con funciones regulatorias o estructurales en algún proceso durante la espermatogénesis. De hecho, se han identificado y caracterizado varios lncRNAs con funciones en la espermatogénesis, como por ejemplo, *H19* [Sahlu *et al.*, 2020], *HongrES2* [Luk *et al.*, 2014], o *Mrhl* [Kataruka *et al.*, 2017] (ver 2.3.1 en Introducción). Asimismo, la espermatogénesis, y en particular la espermiogénesis, se caracterizan por los elevadísimos niveles de regulación postranscripcional. Dado que hay tantos lncRNAs expresados precisamente en las células post-meióticas, no resulta descabellado pensar que muchos lncRNAs podrían estar participando en estos procesos regulatorios. Más aún, la represión traduccional es una estrategia utilizada durante la espermiogénesis para regular la sincronización de la síntesis proteica, cuya producción en etapas incorrectas puede ser altamente perjudicial. Esto se ha demostrado por ejemplo, para el caso de la protamina 1 (*Prm1*), cuyo mRNA es secuestrado durante semanas previo a su traducción en espermátidas elongadas, y cuya traducción prematura es causa conocida de infertilidad [Lee *et al.*, 1995]. Por lo tanto, los lncRNAs podrían también participar en la regulación traduccional. Una característica sobresaliente de la espermatogénesis es la enorme cantidad de proteínas de unión a ARN (RBPs). Muchas de ellas actúan formando partículas de ribonucleoproteínas con mRNAs, y de esa manera los inactivan traduccionalmente. Por lo tanto, los mecanismos de regulación postranscripcional mediados por RBPs son de vital importancia [Sutherland *et al.*, 2015; Ozturk & Uysal, 2018].

En suma, es posible que el estado “permisivo” de la cromatina en las células espermatogénicas haya potenciado la expresión de diversos tipos de ARN, entre ellos una gran cantidad de lncRNAs. Creemos que posiblemente buena parte de ellos correspondan a transcripción “promiscua” o ruido transcripcional, en tanto otros habrían adquirido, a lo largo de la evolución, funciones relevantes para el proceso de desarrollo de las células germinales masculinas que, dada su necesidad de acopio de ARNs, secuestro y traducción retardada de mensajeros [Kleene, 2001], se haya visto ampliamente beneficiado por la regulación mediada por ARNs.

#### 5.4: De la caracterización global del transcriptoma de lncRNAs

Una vez definidos los lncRNAs expresados y DE en las diferentes etapas estudiadas de la espermatogénesis del ratón, caracterizamos el transcriptoma de lncRNAs de forma global.

La mayor parte de los lncRNAs expresados (y DE) en nuestras listas tienen de 1 a 4 exones (teniendo 2 ó 3 exones aproximadamente el 60% del total) y entre 500 y 3000 nt de largo de transcripto (~74%), para todas las etapas sin diferencia. Si bien no teníamos ningún fundamento lógico para sospechar que fuésemos a encontrar diferencias en estas dos características para alguna de las etapas, el análisis lo realizamos teniendo en cuenta los resultados obtenidos para testículo de rata [Chalmel *et al.*, 2014], donde se reportó que los lncRNAs con pico de expresión meiótico eran particularmente más largos. Nuestros resultados no coinciden con este trabajo, pero sí con lo descrito en testículo total de cerdo [Ran *et al.*, 2016] y de gallo [Liu *et al.*, 2017 (a)], y curiosamente también en endometrio de cerdo [Wang *et al.*, 2016]. Asimismo, en la base de datos de lncRNAs de línea germinal, *GermLncRNA*, se reporta que la mayoría contienen 1 ó 2 exones y menos de 1.500 nt de largo [Luk *et al.*, 2015]. Por su parte, Ravasi *et al.* [2006] habían estudiado el genoma de ratón, y su expresión mediante microarreglos, y concluyeron que los lncRNAs tenían exones más largos pero menor cantidad de intrones que los transcriptos codificantes.

Con respecto a la distribución de los lncRNAs expresados a partir de los distintos cromosomas, en términos generales no encontramos diferencias significativas en la distribución de los lncRNAs para ninguno de los autosomas, en los cuales se expresan entre 8 y 16% de los lncRNAs anotados. Para el cromosoma X el porcentaje es levemente menor (7%), sin ser significativo. Como hemos mencionado, aunque no podemos sacar conclusiones ya que es muy bajo el número de lncRNAs que detectamos que se expresa desde el cromosoma X, sin embargo, podemos ver la tendencia al apagado de los mismos en PS y encendido en RS, lo que sugiere un comportamiento para los lncRNAs similar al de los genes codificantes y no al de los miRNAs en relación a la MSCI. Vockel y cols. [2019] plantearon que el cromosoma X está enriquecido en genes expresados en la espermatogénesis temprana. Contrariamente, en nuestras

listas de 2C y LZ, que podrían corresponderse a dicha etapa, no encontramos un gran número de genes de lncRNAs del cromosoma X expresados.

Lo que llamó poderosamente nuestra atención es la ausencia de lncRNAs del cromosoma Y en todas las etapas estudiadas de la espermatogénesis, particularmente teniendo en cuenta que, a pesar de su pequeño tamaño, existe un elevado número de lncRNAs anotados para el cromosoma Y en la base de datos *Ensembl*. Este resultado había sido reportado en un estudio sobre el transcriptoma del cuerpo cromatoide (CC), en el que encontraron que hay ARNs no codificantes enriquecidos en el CC, que están codificados en todos los cromosomas excepto en el cromosoma Y [Meikar *et al.*, 2014]. Nuestros resultados amplían este resultado, refiriéndolo a todos los lncRNAs de la espermatogénesis, y no particularmente a los del CC. Ran y colaboradores [2016] caracterizaron la expresión de lncRNAs en el testículo de cerdo e identificaron más de 750 lncRNAs distribuidos en todos los cromosomas excepto en el Y, concordando con nuestros resultados. Contrariamente, Sun *et al.* [2013] reportaron que la distribución cromosómica de los lncRNAs diferenciales era homogénea entre todos los cromosomas incluido el Y, cuando estudiaron la expresión diferencial de lncRNAs entre testículo de ratón neonatal y adulto. Nuestros resultados no nos permiten atribuir este fenómeno a la MSCI que ocurre durante el paquiteno, ya que la depleción en la expresión de lncRNAs desde el Y se da en todas las etapas estudiadas. Definitivamente, la ausencia de genes de lncRNAs expresados durante la espermatogénesis en el cromosoma Y es un hecho intrigante, que deberá ser estudiado en el futuro.

#### 5.5: ¿Existe alguna relación entre la expresión de lncRNAs y la de genes codificantes?

Los lincRNAs fueron el biotipo más expresado (y DE) en las etapas estudiadas del proceso (50% en promedio), seguidos en segundo lugar por los lncRNAs antisentido (poco más del 42% en promedio), y muy de lejos por los promotores bidireccionales (promedio del 6%); el resto de los biotipos no superó en ningún caso el 1% en ninguna de las etapas. Los valores globales no resultan tan sorprendentes, ya que en el genoma de referencia utilizado (*Ensembl* v93)

lincRNAs y antisentido son los dos biotipos más anotados. Además, los resultados para las dos categorías más expresadas coinciden con los de Lin *et al.* [2016], Song *et al.* [2018], Bao *et al.* [2013] y Sun *et al.* [2013] también en ratón, o con los de Liu *et al.* [2017 (a)] para gallo. Lo que sí resulta llamativo es la alta proporción de lncRNAs antisentido que se expresan, en relación a la baja proporción de genoma codificante. Esto sugeriría que el mecanismo de regulación en la expresión de genes mediante la acción de ARNs antisentido podría estar ampliamente difundido en este sistema.

Cuando evaluamos específicamente los niveles de co-expresión con genes codificantes, sorprendentemente encontramos que de los lncRNAs antisentido solapantes con genes codificantes, casi el 80% tiene una correlación positiva donde ambos son sobreexpresados o reprimidos en la misma etapa, mientras que para apenas un 5% no encontramos correlación. Los coeficientes de correlación para los pares antisentido/gen codificante correlacionables fueron de moderados a altos, con significancia estadística. Lo que es más, para la mayor parte de los genes codificantes solapantes de estos pares, encontramos un reporte de función o expresión específica en el testículo. Esto nos remite a la idea ya discutida de la accesibilidad de la cromatina, y nos lleva a pensar que este estado “permisivo” podría facilitar la transcripción desde la hebra opuesta [Barman *et al.*, 2019]. Por otro lado, como también hemos mencionado, en otros tejidos se ha reportado que lncRNAs antisentido o el evento de su transcripción pueden regular los transcriptos codificantes en la hebra opuesta, ya sea activando o reprimiendo su transcripción [Pelechano & Steinmetz, 2013; Barman *et al.*, 2019]. Según Engreitz *et al.* [2016], la función de los lncRNAs de unir complejos proteicos regulatorios e influir en la expresión local o de genes vecinos, pueden explicar el hecho de que su expresión esté correlacionada con la de estos genes. En este sentido, podemos pensar que para al menos algunos de los pares lncRNA antisentido/gen codificante, la mayoría de los cuales están ambos sobreexpresados en RS en nuestras listas, el lncRNA tenga una regulación positiva sobre el gen codificante. Este tipo de mecanismo de acción fue descrito en otros tejidos [Guil & Esteller, 2012]. Por otra parte, este hecho nos permite plantearnos otra hipótesis: siendo que RS se caracteriza por su amplia regulación postranscripcional, en la que destaca el retardo traduccional como estrategia, a través del secuestro de mRNAs [da Cruz *et al.*, 2016]; y que,

además, la amplia actividad de transcripción en RS va acompañada de una traducción ineficiente para prevenir la producción de proteínas a destiempo [Kleene, 2001; Soumillon *et al.*, 2013], podemos suponer que al menos algunos lncRNAs antisentido podrían actuar sobre su transcrito codificante solapante, ya sea secuestrándolo y evitando su traducción, o bien estabilizándolo para su traducción más tardía. Un ejemplo de esto último, es el caso del lncRNA *BACE1-AS*, que estabiliza al transcrito del gen en la hebra sentido [Faghihi *et al.*, 2010], y que mencionamos en la Introducción como asociado a la enfermedad de Alzheimer (ver 2.2.9). En el trabajo de Sun *et al.* [2013] se vio que muchos de los lncRNAs diferenciales en el testículo neonatal del ratón solapaban o eran vecinos de genes codificantes para factores de transcripción y genes involucrados en espermatogénesis, como *Ovol 1 y 2*, *Sox 3 y 9*, *Wt1*, *Sycp2*, *Prm1 y 2* (protaminas 1 y 2). Sin embargo, no podemos descartar la posibilidad de que, como fue mencionado anteriormente, la expresión desde ambas hebras se produzca debido a un estado de cromatina permisiva y desempaquetada que la favorezca [Barman *et al.*, 2019], y que no esté cumpliendo funciones específicas.

En lo que respecta al análisis de co-expresión con los lincRNAs, la correlación fue menor que para los lncRNAs antisentido, pero aumentaba cuanto menor fuera la distancia entre el gen codificante y el del lincRNA. Esto indicaría que a medida que los genes se acercan, su patrón de expresión se asemeja en mayor grado. Ello nos induce a seguir pensando en la posibilidad de que sea un estado de cromatina laxa en la región que favorezca la expresión de ambos, como en el caso de los antisentido (cuanto más cerca de un gen que se transcribe, más “proclive” se encontraría esa secuencia a ser transcripta). En apoyo a esto, en el análisis de la distribución de las distancias de lincRNAs con respecto a sus genes vecinos, vimos que el porcentaje observado de lincRNAs en las cercanías de genes codificantes es mayor al esperado si la distribución fuera al azar. En el mismo sentido, en el trabajo de Jia *et al.* [2010] se vio que un 35% de los genes de lncRNAs humanos analizados residían a menos de 10 Kb de genes codificantes, y se propuso que por tanto podrían actuar en *cis* regulando la expresión de dichos vecinos. Del mismo modo, Ran y cols. [2016] caracterizaron la expresión de lncRNAs en testículos de cerdo y observaron que un 40% de los lncRNAs expresados estaban a una distancia de 10 Kb con respecto a genes codificantes. Bao y colaboradores [2013], por su parte, estudiaron los clústeres

de mRNAs a una distancia de 30 Kb de lncRNAs en ratón, concluyendo que se encontraban sobreexpresados o reprimidos en los mismos tiempos. De igual manera, el hecho de que un lncRNA se encuentre a distancias mayores con respecto a un gen codificante no significa que no tenga una función o relación con el mismo, ya que podría estar actuando en *trans*, sobre genes alejados en el mismo cromosoma o incluso en otro cromosoma. Un claro ejemplo es el mencionado *HOTAIR* [Rinn *et al.*, 2007]. Tanto el hecho de que los lncRNAs antisentido tengan altísima correlación positiva con sus genes codificantes solapantes, como el resultado de que el patrón de expresión de los lncRNAs sea más correlacionable con sus genes vecinos a medida que la distancia entre ellos se acorta, coincide con lo propuesto por Derrien *et al.* [2012] a partir de un análisis *in silico* de la anotación de humano contenida en la base de datos GENCODE.

Adicionalmente, destacamos que todas las correlaciones entre los patrones de expresión de lncRNAs y genes codificantes solapantes o vecinos que elegimos para su verificación mediante RT-qPCR, pudieron ser exitosamente confirmadas, lo que constituye una evidencia de la robustez de los datos. La confirmación de la co-expresión de estos pares de genes con sus lncRNAs de ninguna manera significa una correlación verdadera en la espermatogénesis, pero es un indicio más de una posible relación funcional entre ellos.

Además de la co-expresión, la co-localización podría también sugerir alguna relación funcional. En tal sentido, hemos estudiado los patrones de localización del mRNA *Kcnmb4* y su lncRNA antisentido *Kcnmb4os1*, parcialmente solapante, cuya expresión se correlacionaba a lo largo de las distintas etapas de la espermatogénesis. Comprobamos que ambos localizaban en una misma estructura celular, el CC, con un índice de co-localización casi perfecto. La co-localización de este lncRNA y el mRNA solapante podría ser indicio de un mecanismo funcional o de regulación entre ellos. A modo de ejemplo, podemos imaginar un mecanismo de reconocimiento entre ambos, basado en la complementariedad de bases [Barman *et al.*, 2019].

Finalmente, si un lncRNA tuviese una función relevante en el proceso espermatogénico junto con su gen vecino o solapante, es posible esperar que esta relación estuviera conservada en otras especies. Dado que, en general, el nivel de conservación de los lncRNAs es bajo, podemos especular con que este

subconjunto de lncRNAs con función sobre genes codificantes estuviera conservado por encima de la media. Estudiamos entonces la conservación en el genoma humano. Encontramos que un 45 a 50% de los pares de lncRNAs antisentido/gen codificante co-expresados en la espermatogénesis del ratón, estaban presentes en humano, siendo éste un porcentaje bastante superior al esperado tomando en cuenta los datos reportados de conservación general de lncRNAs entre ambas especies. Por ejemplo, un trabajo sugirió que un 33% y 17% de los lncRNAs de cerdo coinciden con los de humano y ratón, respectivamente, mientras un 17% coinciden entre las tres especies [Ran *et al.*, 2016]. Por otra parte, se ha visto que para muchos lncRNAs lo importante no es la secuencia en sí, sino el hecho de que se transcriban [Quinn & Chang, 2016], y en consecuencia lo importante sería la conservación de su posición (sintenia). En ese sentido, Cabili y cols. [2011] identificaron lncRNAs ortólogos en otras especies para el 12% de los lncRNAs humanos, que mantenían sintenia pero moderada homología de secuencia. En un estudio evolutivo a gran escala donde se incluyeron 11 especies de tetrápodos, identificaron unos 11.000 lncRNAs específicos de primates, y de ellos alrededor de 2.500 altamente conservados dentro de los primates. La mayoría de los lncRNAs homólogos eran sinténicos, incluso entre especies más alejadas. También se conservaban en gran medida las redes de co-expresión con genes codificantes de proteínas, lo que asocia a los lncRNAs con procesos fundamentales como pueden ser espermatogénesis o desarrollo embrionario [Necsulea *et al.*, 2014]. En definitiva, es posible o bien que la función del lncRNA dependa de la secuencia conservada pero que ésta sea tan divergente que no pueda ser detectada con las herramientas actuales, o que, alternativamente, la función del mismo no dependa de la secuencia específica de bases [Ulitsky & Bartel, 2013]. En vista de esto, evaluamos también la conservación de posición. Para un 59% de los genes codificantes pertenecientes a los pares lncRNA/gen codificante separados entre sí a  $< 30$  Kb en el ratón, encontramos un lncRNA a menos de 30 Kb de ese gen también en humano.

### 5.6: ¿Todos los caminos conducen al cuerpo cromatoide?

Una vez realizados los análisis de expresión diferencial de los lncRNAs, y de haber visto sus correlaciones con genes solapantes o vecinos, decidimos seleccionar algunos de ellos para estudiar su localización sub-celular. De esta manera podríamos generar hipótesis sobre su posible función o mecanismo de acción, de acuerdo a si están acumulados en el núcleo, o citoplasma, o en algún organelo celular, etc.

Sorprendentemente, los ensayos de *RNA-FISH* para todas las sondas testiculares, sin excepción y sin importar el patrón de expresión o biotipo de lncRNA blanco (antisentido o lincRNA), mostraron una marca como un gran punto perinuclear en las células espermiogénicas (RS). Descartamos la posibilidad de que esta señal sea artefactual, ya que la hibridación con el lncRNA ubicuo *Malat1* dio una marca completamente diferente, no mostrando señal en el CC, a pesar de que sus niveles de expresión en RS son relativamente muy elevados (y por el resultado obtenido en los controles negativos). Por supuesto, estamos convencidos de que además de la señal en el CC, es altamente probable que las diversas sondas den un patrón de marca adicional (ej. disperso en el citoplasma), que no llegamos a detectar. Esto parece especialmente probable para el caso de aquellos lncRNAs en que su nivel de expresión fue por lo menos similar en PS que en RS (ver Resultados, Tabla 5), y sin embargo en PS no detectamos señal. En el CC, por el contrario, la señal se haría claramente visible al concentrar gran cantidad de marca en una superficie reducida. Además, lo de los umbrales de detección parece evidente para el caso de *Malat1*, en que más allá de sus niveles de expresión inusualmente elevados en la población 2C (promedialmente, al menos 10 veces por encima de los de las otras sondas), se expresa a niveles relativamente elevados en todas las poblaciones celulares. Sin embargo, y a pesar de que la señal se concentra en el núcleo, su detección resultó dificultosa y es menos evidente en los otros tipos celulares que en la población 2C.

La clara señal en forma de gránulo perinuclear en RS nos llamó particularmente la atención porque parecía localizarse en el CC. Mediante una inmunodetección con fluorescencia (IF) seguida de *RNA-FISH* co-localizamos las sondas con el anticuerpo anti-MVH, y pudimos verificar que la señal de las

sondas coincide con la señal del anticuerpo, corroborando que se trata del CC. Como hemos mencionado (ver sección 2.1.3 en Introducción), trabajos recientes han propuesto que el CC acumula proteínas de unión a ARN, proteínas relacionadas a la regulación del ARN y su procesamiento, ARNs no codificantes cortos y largos, además de algunos mRNAs [Meikar *et al.*, 2014]. En el trabajo de Meikar *et al.* [2014] se vio una tendencia a la acumulación por parte del CC de una gran proporción de transcriptos expresados diferencialmente que estarían involucrados en procesos meióticos y post-meióticos. Actualmente, está creciendo la idea de que este gránulo sin membrana estaría de algún modo vinculado al procesamiento postranscripcional en las espermátidas redondas. Si así fuera, tendría sentido que el CC sea un corpúsculo exclusivo de la espermatogénesis y en particular de la espermiogénesis, dados los niveles inusuales de procesamiento postranscripcional que tienen lugar durante este proceso. El CC podría actuar como depósito de moléculas, ya sea secuestrándolas para su uso en las etapas siguientes en que no hay transcripción y contribuyendo de ese modo a la sincronización apropiada de su traducción, o incluso reteniéndolas para su posterior degradación en caso de no ser necesarias [Toppari *et al.*, 1991; Meikar *et al.*, 2014]. Como se mencionó brevemente en la Introducción, evidencias recientes indican que también participaría en la degradación de transcriptos vía el decaimiento de mRNAs sin sentido (NMD) [Lehtiniemi & Kotaja, 2018], no solamente de transcriptos aberrantes, sino también de transcriptos funcionales, colaborando con la sincronización de los distintos eventos de la espermatogénesis y los productos que participan en ellos. Por otra parte, se ha visto que los CCs tienen movimientos muy característicos: en una primera instancia los CCs se mueven rápidamente en la envoltura nuclear y generan contactos transitorios con el complejo del Golgi. En una segunda etapa, estos movimientos se vuelven más lentos y los CCs se localizan en la parte superior de unos “bolsillos” de la envoltura nuclear, del lado citoplásmico. Se observó una continuidad material entre estas estructuras y los complejos de poro nuclear, internos en el núcleo. Por estos motivos, desde hace tiempo se ha propuesto que los CCs tendrían un rol en el transporte e intercambio de productos génicos entre el núcleo y el citoplasma en las espermátidas redondas, incluyendo ARNs [Parvinen & Parvinen, 1979; Parvinen, 2005].

Las preguntas que surgen entonces son: ¿por qué todos los lncRNAs que estudiamos se dirigen al CC en la post-meiosis? ¿Podrán los lncRNAs estar desempeñando una función, por ejemplo, en el secuestro de mensajeros ya sea para su posterior utilización o degradación? Podemos también preguntarnos si estos lncRNAs que muestran localización en el CC estén allí almacenados para su eliminación porque ya cumplieron su función en etapas previas, o si, contrariamente, fueron transcritos antes de la represión transcripcional y almacenados hasta el momento en que deban actuar. Otra pregunta que surge es si el secuestro de lncRNAs en el CC es un mecanismo ampliado para muchos lncRNAs espermatogénicos, o si por el contrario está restringido sólo a algunos.

Un hecho particularmente interesante fue la co-localización en el CC del lncRNA antisentido *Kcnmb4os1* y su mRNA solapante *Kcnmb4*. *Kcnmb4* es un mRNA que comparte el patrón de expresión con su lncRNA antisentido, y ambos se incrementan en paquitenos. El gen codifica para una unidad reguladora de un canal de potasio activado por calcio, y es sabido que los canales de calcio juegan papeles esenciales en la activación y movilidad espermáticas, en la reacción del acrosoma y la fertilización [Shukla *et al.*, 2012].

Si bien la hibridación no es un ensayo funcional que nos permita asegurarlo, la co-localización del lncRNA *Kcnmb4os1* y el mRNA *Kcnmb4* en el CC puede significar un indicio de mecanismo funcional o de regulación entre un lncRNA antisentido y un gen codificante solapante. Una hipótesis interesante que merece mayor atención, sería que *Kcnmb4os1* interactuara de algún modo con su transcrito sentido, lo dirigiera al CC, y eventualmente pudiera controlar su traducción o degradación mediante la unión entre ambos, u otro mecanismo regulatorio. Si esta hipótesis fuese verdad, tal vez podría ser parte de un mecanismo para modular la función del canal iónico en las células germinales. De todos modos, no podemos descartar la posibilidad de que simplemente ambos transcritos de forma independiente se almacenen en el CC, como vimos para el resto de los lncRNAs estudiados, y fue reportado para múltiples ARNs [Meikar *et al.*, 2014].

### 5.7: ¿Podrán participar lncRNAs en la regulación de procesos meióticos?

Para la sonda de ARN del lncRNA antisentido *Kcnmb4os1* pudimos ver además de la señal en el CC, una marca nuclear a nivel de los espermatoцитos. Esta señal, claramente visible ya que se concentraba en forma de uno o dos puntos brillantes en el núcleo (muchas veces con aspecto de dos puntos fusionados), indefectiblemente nos recordó a la señal observada por Ding y col. [Ding *et al.*, 2012 (b); Ding *et al.*, 2019] para los lncRNAs participantes en el reconocimiento de homologías en los núcleos en cola de caballo de las células en profase meiótica de la levadura *S. pombe* (ver Figura 7, y descripción detallada en Introducción, sección 2.3.1). En esta levadura se observó que dichos lncRNAs se acumulaban en sus propios *loci* cromosómicos formando dos puntos, los cuales convergían en un punto único en concordancia con el apareamiento homólogo [Ding *et al.*, 2012 (b); Ding *et al.*, 2019].

Lo primero que nos preguntamos fue si esa marca nuclear se localizaría sobre su propio *locus*, de manera análoga a lo que ocurre en la levadura. Como primera aproximación, intentamos ver si la marca estaba relacionada a la cromatina y particularmente al CS mediante la inmunomarcación paralela de la proteína de complejo sinaptonémico SYCP1. Si bien los resultados lo sugieren, lamentablemente los ensayos no nos permitieron la suficiente resolución para poder sacar conclusiones definitivas al respecto. La hipótesis de que lncRNAs pudieran estar participando en la regulación de procesos clave de la profase meiótica, como el reconocimiento homólogo u otros, en los espermatoцитos de los mamíferos es sumamente atrayente, ya que no se sabe nada al respecto. Desafortunadamente, los ensayos requeridos para evaluarlo serían bastante más complejos que en las levaduras. No obstante ello, hemos iniciado algunos estudios en esta dirección, los que se detallan en la siguiente sección.

### 5.8: Avances en los cultivos celulares de sobrevida de línea germinal

Para poder determinar las verdaderas funciones o mecanismos de acción de los lncRNAs en el proceso espermatoγénico, si bien los análisis y las hibridaciones realizadas nos dan indicios de posibles funciones, se deben

realizar ensayos funcionales (como pérdida de función, sobreexpresión del lncRNA, etc.). En el caso de los lincRNAs, se podrían construir ratones mutantes *knock-out* (KO) (por ejemplo, por CRISPR/Cas9) y analizar su fenotipo. Sin embargo, no es tan simple la generación de ratones mutantes KO para un lncRNA antisentido, ya que al solapar con otro transcripto, la alteración del antisentido va a afectar indudablemente al transcripto solapante. Más aún, dado que, como hemos mencionado, en muchos casos se ha propuesto que la función del lncRNA está dada simplemente por el hecho de su transcripción, la alteración de su secuencia o eliminación de parte del transcripto, podrían no tener efecto.

Otra opción sería entonces la generación de mutantes *knock-down* (KD), donde por medio de ARNs cortos interferentes se reduzca la expresión del lncRNA sin afectar al transcripto codificante solapante (KD estable). En este sentido, y ante la falta de cultivos celulares eficientes de línea germinal [Reuter *et al.*, 2012], se debería trabajar con ratones mutantes, cuya generación y experimentación son más laboriosas y costosas económicamente. Además, nuestro objetivo no es caracterizar el fenotipo derivado de la eventual alteración en la expresión de la proteína (en nuestro caso, un canal iónico), sino profundizar en los mecanismos moleculares subyacentes a la posible interacción entre el lncRNA y el mRNA.

Nos propusimos entonces desarrollar un método de cultivo celular de sobrevida para línea germinal, partiendo de células purificadas por citometría de flujo, para las fracciones 4C (varios estadios de espermatoцитos I) y RS. Existen escasos antecedentes de este tipo de cultivos. Gerton y Millette [1986] obtuvieron células de PS y RS mediante disgregación enzimática de testículo y *STA-PUT*, cuya viabilidad a las 24 y 48 horas fue de 90% y 85%, respectivamente. Estos autores también indicaron que a mayor densidad celular en el cultivo, menor la viabilidad, ya que no todas las células acceden por igual a los nutrientes y al oxígeno. Mediante disgregación enzimática de biopsias testiculares y posterior *STA-PUT*, Joshi y cols. [1990] colectaron fracciones de células de PS y RS, cuya viabilidad a las 24 horas fue de 73 y 91%, respectivamente. Bellvé [1993] purificó espermatoцитos y espermátidas mediante *STA-PUT*, y las cultivó hasta 48 horas, con una viabilidad del 85% a las 20 horas de incubación. Planteó, además, que para que las células germinales mantengan su viabilidad a mayores intervalos temporales, deberían co-cultivarse con células

de Sertoli, lo cual para nuestros objetivos de pureza no es factible. En un estudio donde se obtuvieron fracciones enriquecidas en células PS por *STA-PUT*, la viabilidad de las mismas fue de 90% a las 24 horas, 70-80% las 48 horas, y decayó abruptamente a 10-30% a las 90 horas de incubación [Handel *et al.*, 1995]. Finalmente, en un trabajo donde purificaron células de RS por elutriación con un 90% de pureza y 100% de viabilidad inicial, éstas se mantuvieron con una buena viabilidad (60% vivas) a las 72 horas de incubación [Dehnugara *et al.*, 2012].

Nos establecimos como meta alcanzar una viabilidad celular alta a las 48 horas de incubación, como tiempo mínimo arbitrario para poder ver un efecto de la interferencia por ARN (RNAi) en la expresión de los lncRNAs a testear (los tiempos de incubación con RNAi dependen del tipo celular, de la concentración y la eficiencia de los RNAi, de la eficiencia del método de transfección; por lo tanto, el ensayo deberá ser puesto a punto previamente). Si bien con nuestra metodología de disgregación mecánica del tejido y posterior clasificación por citometría de flujo [Rodríguez-Casuriaga *et al.*, 2014; Geisinger & Rodríguez-Casuriaga, 2017] los resultados habían sido satisfactorios, notamos una desmejora importante en los ensayos realizados empleando disgregadores (medicons) adquiridos en una segunda compra, con una viabilidad inicial muy baja para ambas fracciones celulares, y tiempos de colecta de células muy prologados, provocando que las células colectadas comenzaran a morir antes de ser incubadas en la temperatura y atmósfera deseadas. La implementación del protocolo de disgregación enzimática [Romrell *et al.*, 1976; Gaysinskaya & Bortvin, 2015] nos permitió aumentar en gran medida la viabilidad inicial y disminuir los tiempos de clasificación. Con una disociación mecánica del tejido testicular de ratón las células tienen una viabilidad inicial del 80%, mientras que con disgregación enzimática (colagenasa, tripsina y DNAsa) se alcanzan fracciones celulares con 98% de viabilidad [Aslam *et al.*, 1998].

Las células de cada población fueron colectadas inicialmente en placas pre-tratadas con poli-L-lisina para incubarlas adheridas a una superficie. Ante la observación de que a las 24 y 48 horas aquellas células que se habían adherido a la placa en su mayoría estaban muertas, mientras que la menor proporción que había quedado suspendida en el medio de cultivo se encontraba en gran parte viable, procedimos a colectar las células en tubos e incubarlas en suspensión.

De esta forma, logramos aumentar la viabilidad a tiempo final de forma significativa.

El desarrollo de este método para cultivos celulares de sobrevida de línea germinal constituye un gran avance, ya que no contábamos hasta el momento con una herramienta de esta índole. Hasta donde conocemos, el trabajo más aproximado es el de Aslam y colaboradores [1998], en el cual realizaron una disgregación enzimática del testículo con posterior clasificación celular en citometría de flujo, alcanzando un 99% de viabilidad inicial, pero no incubaron las células en estufas de cultivo ni midieron la viabilidad a tiempos posteriores. Al mismo tiempo, este protocolo nos será de gran utilidad para realizar los ensayos funcionales que, esperamos, nos permitan contribuir a dilucidar los mecanismos de acción de los lncRNAs en los distintos procesos de la espermatogénesis, así como a entender las relaciones regulatorias entre los lncRNAs y otros transcritos.

## **6. CONCLUSIONES** **Y PERSPECTIVAS**

## 6.1: Conclusiones

Como conclusión general del trabajo, hemos caracterizado el transcriptoma global de lncRNAs a lo largo de distintas etapas de la espermatogénesis del ratón, que incluye pre-meiosis, meiosis y post-meiosis. La enorme mayoría de los lncRNAs se expresan en la etapa post-meiótica, y esto se cumple también para los lncRNAs diferencialmente expresados. Describimos las características generales de los mismos: son transcritos cortos, de bajo número de exones, la extensa mayoría son lincRNAs o antisentido, se expresan desde todos los cromosomas exceptuando el Y, y no muestran diferencias en estas características para ninguna de las etapas estudiadas. Asimismo, muestran una conservación relativamente baja: alrededor del 50% de lncRNAs antisentido y 59% de los lincRNAs están presentes, o se encontró un lncRNA en posición sinténica, en humanos. Cuando nos referimos a la co-expresión con genes codificantes vecinos o solapantes, un 95% de los lncRNAs antisentido pueden correlacionarse con su gen codificante solapante; por otra parte, si bien la correlación fue menor para los lincRNAs y sus genes codificantes vecinos a < 300 Kb (57% de los pares correlacionables), la misma aumentó conforme la distancia entre el lincRNA y su gen vecino se reducía (66% a distancia < 30 Kb). A partir de estos resultados nos planteamos dos hipótesis: una primera, en que el estado permisivo de la cromatina laxa alrededor de un gen que está siendo transcrito favorezca la transcripción desde la hebra opuesta o en regiones cercanas; y una segunda, en la que al menos parte de los lncRNAs co-expresados con genes codificantes, tengan una función regulatoria sobre los mismos, mediante un mecanismo, en la mayor parte de los casos, aún desconocido. Ambas hipótesis no son excluyentes, dado que es posible que el estado permisivo de la cromatina haya favorecido la expresión de gran cantidad de lncRNAs, algunos de los cuales habrían adquirido una función (por ejemplo debido a su secuencia o estructura, capacidad de reclutar factores, etc.).

Intentando aportar evidencias para la dilucidación de sus posibles mecanismos funcionales, todos los lncRNAs localizaron en el CC, un organelo perinuclear no membranoso típico de células post-meióticas, que se ha postulado podría servir de almacenamiento para proteínas, ARNs codificantes y no codificantes. En el mismo sentido, un par de ARNs, el lncRNA antisentido

*Kcnmb4os1* y su mRNA solapante *Kcnmb4*, co-localizaron en el CC. Este almacenamiento de moléculas podría deberse a que ya hayan ejercido sus funciones y se secuestren para su degradación, o bien a transcritos que se produjeron previo a la represión transcripcional de la espermiogénesis y aguarden el momento preciso para actuar. Indefectiblemente, nos preguntamos si la co-localización de ambos ARNs parcialmente complementarios, en esta estructura, podrá ser la “punta del *iceberg*” de un mecanismo de direccionamiento al CC.

Asimismo, el lncRNA *Kcnmb4os1* se localizó en un punto en el núcleo de células meióticas, posiblemente relacionado a la cromatina, lo que podría indicar una función regulatoria o estructural en este proceso, por ejemplo a nivel del reconocimiento y apareamiento de cromosomas homólogos. De todas maneras, se necesitan ensayos funcionales para revelar los mecanismos de acción de los lncRNAs, y en ese sentido hemos iniciado el desarrollo y puesta a punto de un cultivo celular de sobrevida para células germinales, que nos permita realizar estos estudios funcionales *in vitro*.

## 6.2: Perspectivas

Como perspectivas de esta tesis, y en la línea de lo que venimos y continuaremos trabajando, nos planteamos:

-Determinar si la señal del lncRNA *Kcnmb4os1* observada en los núcleos de las células meióticas se localiza sobre su propio *locus*. Para ello diseñamos una sonda única de ADN mediante PCR con oligonucleótidos marcados con digoxigenina, y realizamos hibridación *in situ* de ADN. Desafortunadamente, al ser un *locus* simple, y a pesar de amplificar la señal con tiramida, no logramos llegar al límite de detección de la técnica. Es así que buscaremos una sonda alternativa, y ajustaremos los protocolos para realizar hibridación *in situ* de ARN y ADN en forma secuencial, para intentar confirmar si la señal del lncRNA se localiza sobre su propio *locus* en el genoma. De confirmarse esto, constituiría un argumento más en favor de su posible participación en la búsqueda de homologías, de modo análogo a lo reportado para la levadura *S. pombe*.

-Continuando con la señal del lncRNA *Kcnmb4os1* en los núcleos meióticos, nos planteamos como perspectiva su detección en esparcidos de espermatoцитos de ratón. Esta técnica incluye un tratamiento hipotónico que provoca cierta expansión del contenido nuclear, facilitando la visualización de las marcas empleadas, y por lo tanto podría proporcionarnos una evidencia más clara de la localización de la señal. En este sentido, destacamos que estos ensayos ya están siendo llevados a cabo por un estudiante de grado del grupo en el marco de su tesina de Licenciatura. Adicionalmente, existe la posibilidad de realizar estos estudios no sólo en especímenes tipo salvaje, sino también en una línea de ratones mutantes (generada recientemente en nuestro grupo de investigación en el marco de otra tesis de maestría) cuya profase meiótica está detenida en estadios tempranos, es decir, presenta arresto pre-paquiténico. En estos mutantes los cromosomas meióticos se hallan perfectamente alineados pero no pueden aparearse, por lo cual resulta interesante ver la ubicación detallada de este lncRNA en dicha situación. En caso de que este lncRNA pudiera estar vinculado al reconocimiento de homologías y apareamiento meiótico, en los espermatoцитos de los individuos mutantes sería esperable observar en todos los casos dos puntos, que no llegarían a confluir en un punto único.

-Profundizar en el estudio de la posible interacción entre lncRNAs y mRNAs. En particular, nos interesa estudiar si hay una co-regulación entre la expresión del lncRNA antisentido *Kcnmb4os1* y su gen codificante solapante *Kcnmb4*. Dada la imposibilidad de generar mutantes *knock-out* de los lncRNAs antisentido, porque afectarían también a su gen sentido, como alternativa nos proponemos hacer interferencia de ARN con ARNs pequeños interferentes (siRNAs), tanto para el gen codificante como para el lncRNA, y evaluar las consecuencias de la pérdida de función. Para poder cumplir con esto, y debido a la ausencia de sistemas de cultivo eficientes de línea germinal que nos permitan reproducir la espermatogénesis *in vitro*, estamos abocados a la puesta a punto de los cultivos de sobrevida de línea germinal, en particular de fracciones de espermatoцитos I (4C) y espermátidas redondas (RS), partiendo de células purificadas por citometría de flujo. Una vez puestos a punto, pretendemos hacer la interferencia, y:

1. Evaluar mediante RT-qPCR al silenciar el lncRNA, si se afecta la expresión del mRNA, y viceversa, en la fracción 4C (que es el tipo celular en que se activa su expresión). Esto nos permitirá estudiar si existe un efecto regulatorio de la expresión de uno de los ARNs sobre el otro.

2. En el caso de la fracción RS, intentaremos hacer RNA-FISH en las células interferidas, para estudiar si al silenciar el lncRNA, el gen de todos modos se dirige al CC, o no, y viceversa. Así, podríamos brindar evidencias acerca de un posible mecanismo de direccionamiento/secuestro de ARNs.

-Destacamos además que este trabajo de tesis ha producido un gran caudal de datos de secuenciación masiva, que están siendo objeto de estudio de otra tesis doctoral del grupo de investigación. Durante la misma, se está realizando el análisis de nuevos transcritos no anotados y variantes de *splicing* a lo largo de la espermatogénesis del ratón.

## **7. BIBLIOGRAFÍA**

Alzheimer M, Baier A, *et al.* 2010. Synaptonemal complex protein SYCP3 exists in two isoforms showing different conservation in mammalian evolution. *Cytogenet Genome Res*, 128: 162-168. DOI: 10.1159/000303341.

Amaral PP, Dinger ME, *et al.* 2008. The eukaryotic genome as an RNA machine. *Science*, 319(5871): 1787-1789. DOI: 10.1126/science.1155472.

Anguera MC, Ma W, *et al.* 2011. Tsx produces a long noncoding RNA and has general functions in the germline, stem cells, and brain. *PLoS Genet*, 7(9): e1002248. DOI: 10.1371/journal.pgen.1002248.

Aslam I, Robins A, *et al.* 1998. Isolation, purification and assessment of viability of spermatogenic cells from testicular biopsies of azoospermic men. *Hum Reprod*, 13(3): 639-45. DOI: 10.1093/humrep/13.3.639.

Atkinson SR, Marguerat S, *et al.* 2012. Exploring long non-coding RNAs through sequencing. *Semin Cell Dev Biol*, 23(2): 200-205. DOI: 10.1016/j.semcdb.2011.12.003.

Ball RL, Fujiwara Y, *et al.* 2016. Regulatory complexity revealed by integrated cytological and RNA-seq analyses of meiotic substages in mouse spermatocytes. *BMC Genomics*, 17: 628-644. DOI: 10.1186/s12864-016-2865-1.

Bao J, Wu J, *et al.* 2013. Expression profiling reveals developmentally regulated lncRNAs repertoire in the mouse male germline. *Biol Reprod*, 89(5): 107. DOI: 10.1095/biolreprod.113.113308.

Barman P, Reddy D, *et al.* 2019. Mechanisms of antisense transcription initiation with implications in gene expression, genomic integrity and disease pathogenesis. *Non-coding RNA*, 5(1): 11-33. DOI: 10.3390/ncrna5010011.

Bellvé AR, Cavicchia JC, *et al.* 1977. Spermatogenic cells of the prepuberal mouse. Isolation and morphological characterization. *J Cell Biol*, 74(1): 68-85. DOI: 10.1083/jcb.74.1.68.

Bellvé AR. 1993. Purification, culture, and fractionation of spermatogenic cells. *Methods Enzymol*, 225: 84-113. DOI: 10.1016/0076-6879(93)25009-q.

Benjamini Y & Hochberg Y. 1995. Controlling de false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*, 57(1): 289-300. DOI: 10.2307/2346101.

Bettegowda A & Wilkinson MF. 2010. Transcription and post-transcriptional regulation of spermatogenesis. *Phil Trans R Soc B*, 365: 1637-1651. DOI: 10.1098/rstb.2009.0196.

Braun RE. 2001. Packaging paternal chromosomes with protamine. *Nat Genet*, 28(1):10-12. DOI: 10.1038/ng0501-10.

Breschi A, Gingeras TR, *et al.* 2017. Comparative transcriptomics in human and mouse. *Nat Rev Genet*, 18: 425-440. DOI: 10.1038/nrg.2017.19.

Bo H, Liu Z, *et al.* 2020. Long noncoding RNAs expression profile and long noncoding RNA-mediated competing endogenous RNA network in nonobstructive azoospermia patients. *Epigenomics*, 12(8): 673-684. DOI: 10.2217/epi-2020-0008.

Bolcun-Filas E & Handel MA. 2018. Meiosis: the chromosomal foundation of reproduction. *Biol Reprod*, 99: 112-126. DOI: 10.1093/biolre/iroy021.

Bollschweiler D, Radu L, *et al.* 2019. Molecular architecture of the SYCP3 fibre and its interaction with DNA. *Open Biol*, 9(10): 190094. DOI: 10.1098/rsob.190094.

Cabili MN, Trapnell C, *et al.* 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev*, 25: 1915-1927. DOI: 10.1101/gad.17446611.

Cao J. 2014. The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online*, 16: 11. DOI: 10.1186/1480-9222-16-11.

Chalmel F, Lardenois A, *et al.* 2014. High-resolution profiling of novel transcribed regions during rat spermatogenesis. *Biol Reprod*, 91: 5. DOI: 10.1095/biolreprod.114.118166.

Chen Y, Zheng Y, *et al.* 2018. Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res*, 28(9): 879-896. DOI: 10.1038/s41422-018-0074-y.

Chuma S, Hosokawa M, *et al.* 2009. Ultrastructural characterization of spermatogenesis and its evolutionary conservation in the germline: Germinal granules in mammals. *Molecular and Cellular Endocrinology*, 306(1-2): 17-23. DOI: 10.1016/j.mce.2008.11.009.

Cloutier JM & Turner JM. 2010. Meiotic sex chromosome inactivation. *Curr Biol*, 20(22): R962-R963. DOI: 10.1016/j.cub.2010.09.041.

Cohen PE, Pollack SE, *et al.* 2006. Genetic analysis of chromosome pairing, recombination, and cell cycle control during first meiotic prophase in mammals. *Endocrine reviews*, 27(4): 398-426. DOI: 10.1210/er.2005-0017.

Costa Y, Speed R, *et al.* 2005. Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis. *J Cell Science*, 118: 2755-2762. DOI: 10.1242/jcs.02402.

da Cruz I, Rodríguez-Casuriaga R, *et al.* 2016. Transcriptome analysis of highly purified mouse spermatogenic cell populations: gene expression signatures switch from meiotic-to-postmeiotic-related processes at pachytene stage. *BMC Genomics*, 17: 294. DOI: 10.1186/s12864-016-2618-1.

Darbellay F & Necsulea A. 2020. Comparative Transcriptomics Analyses across Species, Organs, and Developmental Stages Reveal Functionally Constrained lncRNAs. *Mol Biol Evol*, 37(1): 240-259. DOI: 10.1093/molbev/msz212.

Das MK, Furu K, *et al.* 2018. Knockdown of SPRY4 and SPRY4-IT1 inhibits cell growth and phosphorylation of Akt in human testicular germ cell tumours. *Sci Rep*, 8(1): 2462. DOI: 10.1038/s41598-018-20846-8.

Dehnugara T, Dhar S, *et al.* 2012. An in vitro, short-term culture method for mammalian haploid round spermatids amenable for molecular manipulation. *Mol Reprod Dev*, 79(1): 19-30. DOI: 10.1002/mrd.21396.

Demmerle J, Innocent C, *et al.* 2017. Strategic and practical guidelines for successful structured illumination microscopy. *Nat Protoc*, 12(5): 988-1010. DOI: 10.1038/nprot.2017.019.

Derrien T, Johnson R, *et al.* 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*, 22(9): 1775-1789. DOI: 10.1101/gr.132159.111.

De Vries FA, de Boer E, *et al.* 2005. Mouse Sycp1 functions in synaptonemal complex assembly, meiotic recombination, and XY body formation. *Genes Dev*, 19: 1376-1389. DOI: 10.1101/gad.329705.

Diederichs S. 2014. The four dimensions of noncoding RNA conservation. *Trends Genet*, 30(4): 121-123. DOI: 10.1016/j.tig.2014.01.004.

Ding DQ, Haraguchi T, *et al.* 2012 (a). Chromosomally-retained RNA mediates homologous pairing. *Nucleus*, 3(6): 516-519. DOI:10.4161/nucl.22732.

Ding DQ, Okamasa K, *et al.* 2012 (b). Meiosis-specific noncoding RNA mediates robust pairing of homologous chromosomes in meiosis. *Science*, 336(6082): 732-736. DOI: 10.1126/science.1219518.

Ding DQ, Haraguchi T, *et al.* 2013. The role of chromosomal retention of noncoding RNA in meiosis. *Chromosome Res*, 21(6-7): 665-672. DOI: 10.1007/s10577-013-9389-1.

Ding DQ, Okamasa K, *et al.* 2019. Chromosome-associated RNA-protein complexes promote pairing of homologous chromosomes during meiosis in *Schizosaccharomyces pombe*. *Nat Commun*, 10(1): 5598. DOI: 10.1038/s41467-019-13609-0.

Dinger ME, Amaral PP, *et al.* 2008. Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res*, 18(9): 1433-1445. DOI: 10.1101/gr.078378.108.

Djebali S, Davis CA, *et al.* 2012. Landscape of transcription in human cells. *Nature*, 489: 101-108. DOI: 10.1038/nature11233.

Dong FN, Amiri-Yekta A, *et al.* 2018. Absence of CFAP69 Causes Male Infertility due to Multiple Morphological Abnormalities of the Flagella in Human and Mouse. *Am J Hum Genet*, 102(4): 636-648. DOI: 10.1016/j.ajhg.2018.03.007.

Dunce JM, Dunne OM, *et al.* 2018. Structural basis of meiotic chromosome synapsis through SYCP1 self-assembly. *Nat Struct Mol Biol*, 25(7): 557-569. DOI: 10.1038/s41594-018-0078-9.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414): 57-74. DOI: 10.1038/nature11247.

Endesfelder U & Heilemann M. 2015. Direct stochastic optical reconstruction microscopy (dSTORM). *Methods Mol Biol*, 1251: 263-276. DOI: 10.1007/978-1-4939-2080-8\_14.

Engreitz JM, Haines JE, *et al.* 2016. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature*, 539(7629): 452-455. DOI: 10.1038/nature20149.

Erickson RP. 1990. Post-meiotic gene expression. *Trends Genet*, 6: 264-268. DOI: 10.1016/0168-9525(90)90209-O.

Esponda P & Giménez-Martín G. 1972. The attachment of the synaptonemal complex to the nuclear envelope. *Chromosoma*, 38: 405-417. DOI: 10.1007/BF00320159.

Faghihi MA, Zhang M, *et al.* 2010. Evidence for natural antisense transcript-mediated inhibition of microRNA function. *Genome Biol*, 11(5): R56. DOI: 10.1186/gb-2010-11-5-r56.

Fatica A & Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet*, 15(1): 7-21. DOI: 10.1038/nrg3606.

Faulkner GJ & Carninci P. 2009. Altruistic functions for selfish DNA. *Cell Cycle*, 8: 2895–2900. DOI: 10.4161/cc.8.18.9536.

Firth MC, Bailey TL, *et al.* 2006. Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA Biol*, 3(1): 40-48. DOI: 10.4161/rna.3.1.2789.

Fraune J, Schramm S, *et al.* 2012. The mammalian synaptonemal complex: protein components, assembly and role in meiotic recombination. *Exp Cell Res*, 318: 1340-1346. DOI: 10.1016/j.yexcr.2012.02.018.

Fraune J, Wiesner M, *et al.* 2014. The synaptonemal complex of basal metazoan Hydra: More similarities to vertebrate than invertebrate meiosis model organisms. *J Genet Genomics*, 41: 107-115. DOI: 10.1016/j.jgg.2014.01.009.

Frazee AC, Pertea G, *et al.* 2015. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat Biotechnol*, 33: 243-266. DOI: 10.1038/nbt.3172.

Fritah S, Niclou SP, *et al.* 2014. Databases for lncRNAs: a comparative evaluation of emerging tools. *RNA*, 20(11): 1655-1665. DOI: 10.1261/rna.044040.113.

Froberg JE, Yang L, *et al.* 2013. Guided by RNAs: X-inactivation as a model for lncRNA function. *J Mol Biol*, 425(19): 3698-3706. DOI: 10.1016/j.jmb.2013.06.03.

Gabory A, Jammes H, *et al.* 2010. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays*, 32: 473-480. DOI: 10.1002/bies.200900170.

Gan H, Cai T, *et al.* 2013. Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Mol Cell Proteomics*, 12: 1144-1157. DOI: 10.1074/mcp.M112.020123.

Gaysinskaya V & Bortvin A. 2015. Flow cytometry of murine spermatocytes. *Curr Protoc Cytom*, 72: 7.44.1-7.44.24. DOI: 10.1002/0471142956.cy0744s72.

Ge P, Zhang J, *et al.* 2019. CircRNA expression profile and functional analysis in testicular tissue of patients with non-obstructive azoospermia. *Reprod Biol Endocrinol*, 17(1): 100. DOI: 10.1186/s12958-019-0541-4.

Geisinger A. 2008. Spermatogenesis In Mammals: A Very Peculiar Cell Differentiation Process. En: Ivanova LB (ed) *Cell Differentiation Research Developments*, 97-123. Nova Publishers, New York. ISBN: 978-160021938-2.

Geisinger A & Rodríguez-Casuriaga R. 2010. Flow cytometry for gene expression studies in Mammalian spermatogenesis. *Cytogenet Genome Res*, 128(1-3): 46-56. DOI: 10.1159/000291489.

Geisinger A & Rodríguez-Casuriaga R. 2017. Flow cytometry for the isolation and characterization of rodent meiocytes. *Methods Molec Biol*, 1471: 217-230. DOI: 10.1007/978-1-4939-6340-9\_11.

Gerton GL & Millette CF. 1986. Stage-specific synthesis and fucosylation of plasma membrane proteins by mouse pachytene spermatocytes and round spermatids in culture. *Biol Reprod*, 35(4): 1025-35. DOI: 10.1095/biolreprod35.4.1025.

Gerton JL & Hawley RS. 2005. Homologous chromosome interactions in meiosis: diversity amidst conservation. *Nat Rev Genet*, 6(6): 477-487. DOI: 10.1038/nrg1614.

Getun IV, Torres B, *et al.* 2011. Flow cytometry purification of mouse meiotic cells. *J Vis Exp*. 50: 2602. DOI: 10.3791/2602.

Girard A, Sachidanandam R, *et al.* 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, 442(7099): 199-202. DOI: 10.1038/nature04917.

Gómez-H L, Felipe-Medina N, *et al.* 2016. C14ORF39/SIX6OS1 is a constituent of the synaptonemal complex and is essential for mouse fertility. *Nat Commun*, 7: 13298. DOI: 10.1038/ncomms13298.

Guil S & Esteller M. 2012. Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol*, 19(11): 1068-1075. DOI: 10.1038/nsmb.2428.

Guo B, Zhu X, *et al.* 2020. The Roles of LncRNAs in Osteogenesis, Adipogenesis and Osteoporosis. *Curr Pharm Des*, (publicado *online*, en prensa). DOI: 10.2174/1381612826666200707130246.

Gupta RA, Shah N, *et al.* 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, 464: 1071-1076. DOI: 10.1038/nature08975.

Guttman M, Amit I, *et al.* 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, 458(7235): 223-227. DOI: 10.1038/nature07672.

Hamer G, Gell K, *et al.* 2006. Characterization of a novel meiosis-specific protein within the central element of the synaptonemal complex. *J Cell Science*, 119: 4025-4032. DOI: 10.1242/jcs.03182.

Han Z, Xue W, *et al.* 2020. Genome-wide identification and analysis of the eQTL lncRNAs in multiple sclerosis based on RNA-seq data. *Brief Bioinform*, 21(3): 1023-1037. DOI: 10.1093/bib/bbz036.

Handel MA, Caldwell KA, *et al.* 1995 Culture of pachytene spermatocytes for analysis of meiosis. *Dev Genet*, 16(2): 128-39. DOI: 10.1002/dvg.1020160206.

Handel MA & Schimenti JC. 2010. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nat Rev Genet*, 11(2): 124-136. DOI: 10.1038/nrg2723.

Hao SL, Ni FD, *et al.* 2019. The dynamics and regulation of chromatin remodeling during spermiogenesis. *Gene*, 706: 201-210. DOI: 10.1016/j.gene.2019.05.027.

Hentze MW & Preiss T. 2013. Circular RNAs: splicing's enigma variations. *EMBO J*, 32(7): 923-925. DOI: 10.1038/emboj.2013.53.

Hernández-Hernández A, Rincón-Arango H, *et al.* 2008. Differential distribution and association of repeat DNA sequences in the lateral element of the synaptonemal complex in rat spermatocytes. *Chromosoma*, 117(1): 77-87. DOI: 10.1007/s00412-007-0128-2.

Hezroni H, Koppstein D, *et al.* 2015. Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep*, 11(7): 1110-1122. DOI: 10.1016/j.celrep.2015.04.023.

Hogarth CA, Arnold S, *et al.* 2015. Processive pulses of retinoic acid propel asynchronous and continuous murine sperm production. *Biol Reprod*, 92(2): 37. DOI: 10.1095/biolreprod.114.126326.

Homolka D, Ivanek R, *et al.* 2011. Differential expression of non-coding RNAs and continuous evolution of the X chromosome in testicular transcriptome of two mouse species. *PLoS One*, 6(2): e17198. DOI: 10.1371/journal.pone.0017198.

Hon CC, Ramilowski JA, *et al.* 2017. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, 543(7644): 199-204. DOI: 10.1038/nature21374.

Hong SH, Kwon JT, *et al.* 2018. Profiling of testis-specific long noncoding RNAs in mice. *BMC Genomics*, 19(1): 539. DOI: 10.1186/s12864-018-4931-3.

Honson DD & Macfarlan TS. 2018. A lncRNA-like role for LINE1s in development. *Dev Cell*, 46(2): 132-134. DOI: 10.1016/j.devcel.2018.06.022.

Hu K, Zhang J, *et al.* 2017. LncRNA AK015322 promotes proliferation of spermatogonial stem cell C18-4 by acting as a decoy for microRNA-19b-3p. *In Vitro Cell Dev Biol Anim*, 53(3): 277-284. DOI: 10.1007/s11626-016-0102-5.

Hung T & Chang HY. 2010. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol*, 7(5): 582–585. DOI: 10.4161/rna.7.5.13216.

Hutchinson JN, Ensminger AW, *et al.* 2007. A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC Genomics*, 8: 39. DOI: 10.1186/1471-2164-8-39.

Ilik I & Akhtar A. 2009. roX RNAs: non-coding regulators of the male X chromosome in flies. *RNA Biol*, 6(2): 113-121. DOI: 10.4161/rna.6.2.8060.

Ishiguro KI, Matsuura K, *et al.* 2020. MEIOSIN Directs the Switch from Mitosis to Meiosis in Mammalian Germ Cells. *Dev Cell*, 52(4): 429-445.e10. DOI: 10.1016/j.devcel.2020.01.010.

Jan SZ, Vormer TL, *et al.* 2017. Unraveling transcriptome dynamics in human spermatogenesis. *Development*, 144(20): 3659-3673. DOI: 10.1242/dev.152413.

Jarroux J, Morillon A, *et al.* 2017. History, Discovery, and Classification of lncRNAs. *Adv Exp Med Biol*, 1008: 1-46. DOI: 10.1007/978-981-10-5203-3\_1.

Jia H, Osak M, *et al.* 2010. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. *RNA*, 16(8): 1478-1487. DOI: 10.1261/rna.1951310.

Johnsson P, Lipovich L, *et al.* 2014. Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim Biophys Acta*, 1840(3): 1063-1071. DOI: 10.1016/j.bbagen.2013.10.035.

Joshi MS, Anakwe OO, *et al.* 1990. Preparation and short-term culture of enriched populations of guinea pig spermatocytes and spermatids. *J Androl*, 11(2): 120-30. DOI: 10.1002/j.1939-4640.1990.tb01592.x.

Kanduri C. 2016. Long noncoding RNAs: Lessons from genomic imprinting. *Biochim Biophys Acta*, 1859(1): 102-111. DOI: 10.1016/j.bbagem.2015.05.006.

Kapranov P, Cheng J, *et al.* 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, 316: 1484-1488. DOI: 10.1126/science.1138341.

Kataruka S, Akhade VS, *et al.* 2017. Mrhl Long Noncoding RNA Mediates Meiotic Commitment of Mouse Spermatogonial Cells by Regulating Sox8 Expression. *Mol Cell Biol*, 37(14): e00632-16. DOI: 10.1128/MCB.00632-16.

Kazemian M, Ren M, *et al.* 2015. Comprehensive assembly of novel transcripts from unmapped human RNA-Seq data and their association with cancer. *Mol Syst Biol*, 11(8): 826. DOI:10.15252/msb.156172.

Khalil AM, Guttman M, *et al.* 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA*, 106(28): 11667-11672. DOI: 10.1073/pnas.0904715106.

Kim D, Langmead B, *et al.* 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*, 12: 357-360. DOI: 10.1038/nmeth.3317.

Kleckner N. 2006. Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma*, 115(3): 175-194. DOI: 10.1007/s00412-006-0055-7.

Kleene KC. 2001. A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mech Dev*, 106: 3-23. DOI: 10.1016/s0925-4773(01)00413-0.

Kopp F & Mendell JT. 2018. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*, 172(3): 393-407. DOI: 10.1016/j.cell.2018.01.011.

Kotaja N, Lin H, *et al.* 2006. Interplay of PIWI/Argonaute protein MIWI and kinesine KIF17b in chromatoid bodies of male germ cells. *Journal of Cell Science*, 119: 2819-2825. DOI: 10.1242/jcs.03022.

Kotaja N & Sassone-Corsi P. 2007. The chromatoid body: a germ-cell specific RNA-processing centre. *Nat Rev Mol Cell Biol*, 8: 85-90. DOI: 10.1038/nrm2081.

Koziol MJ & Rinn JL. 2010. RNA traffic control of chromatin complexes. *Curr Opin Genet Dev*, 20(2): 142-148. DOI:10.1016/j.gde.2010.03.003.

Kreutzer FP, Fiedler J, *et al.* 2020. Non-coding RNAs: key players in cardiac disease. *J Physiol*, 598(14): 2995-3003. DOI: 10.1113/JP278131.

Kurihara M, Otsuka K, *et al.* 2017. A Testis-Specific Long Non-Coding RNA, lncRNA-Tcam1, Regulates Immune-Related Genes in Mouse Male Germ Cells. *Front Endocrinol*, 8: 299. DOI: 10.3389/fendo.2017.00299.

La Salle S, Sun F, *et al.* 2009. Isolation and short-term culture of mouse spermatocytes for analysis of meiosis. *Methods Mol Biol*, 558: 279-297. DOI: 10.1007/978-1-60761-103-5\_17.

Lagarde J, Uszczynska-Ratajczak B, *et al.* 2017. High-throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet*, 49: 1731-1740. DOI: 10.1038/ng.3988.

Laiho A, Kotaja N, *et al.* 2013. Transcriptome profiling of the murine testis during the first wave of spermatogenesis. *PLoS One*, 8(4): e61558. DOI: 10.1371/journal.pone.0061558.

Latos PA, Pauler FM, *et al.* 2012. Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science*, 338(6113): 1469-1472. DOI: 10.1126/science.1228110.

Lee K, Haugen HS, *et al.* 1995. Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *PNAS USA*, 92(26): 12451-12455. DOI: 10.1073/pnas.92.26.12451.

Lehtiniemi T & Kotaja N. 2018. Germ granule-mediated RNA regulation in male germ cells. *Reproduction*, 155(2): R77-R91. DOI: 10.1530/REP-17-0356.

Li L, Wang M, et al. 2016. A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells. *Cell Death Dis*, 7(3):e2140. DOI: 10.1038/cddis.2016.24.

Li LJ, Leng RX, et al. 2017. Translation of noncoding RNAs: Focus on lncRNAs, pri-miRNAs, and circRNAs. *Exp Cell Res*, 361(1): 1-8. DOI: 10.1016/j.yexcr.2017.10.010.

Li W, Ning JZ, et al. 2018. MALAT1 Promotes Cell Apoptosis and Suppresses Cell Proliferation in Testicular Ischemia-Reperfusion Injury by Sponging MiR-214 to Modulate TRPV4 Expression. *Cell Physiol Biochem*, 46(2): 802-814. DOI: 10.1159/000488738.

Li H, Yang C, et al. 2020. Identification of potential key mRNAs and lncRNAs for psoriasis by bioinformatic analysis using weighted gene co-expression network analysis. *Mol Genet Genomics*, 295(3): 741-749. DOI: 10.1007/s00438-020-01654-0.

Liang M, Li W, et al. 2014. Sequential expression of long noncoding RNA as mRNA gene expression in specific stages of mouse spermatogenesis. *Sci Rep*, 4: 5966. DOI: 10.1038/srep05966.

Liang M, Hu K, et al. 2019. Upregulated lncRNA Gm2044 inhibits male germ cell development by acting as miR-202 host gene. *Anim Cells Syst (Seoul)*, 23(2): 128-134. DOI: 10.1080/19768354.2019.1591506.

Liang M, Wang H, et al. 2020. lncRNA-Gm2044 is transcriptionally activated by A-MYB and regulates Sycp1 expression as a miR-335-3p sponge in mouse spermatocyte-derived GC-2spd(ts) cells. *Differentiation*, 114: 49-57. DOI: 10.1016/j.diff.2020.05.004.

Lin X, Han M, et al. 2016. Expression dynamics, relationships, and transcriptional regulations of diverse transcripts in mouse spermatogenic cells. *RNA Biol*, 13: 1011-1024. DOI: 10.1080/15476286.2016.1218588.

Liu Y, Sun Y, *et al.* 2017 (a). Analyses of long non-coding RNA and mRNA profiling using RNA sequencing in chicken testis with extreme sperm motility. *Sci Rep*, 7: 9055. DOI: 10.1038/s41598-017-08738-9.

Liu Y, Zheng L, *et al.* 2017 (b). Emerging roles and mechanisms of long noncoding RNAs in atherosclerosis. *Int J Cardiol*, 228: 570-582. DOI: 10.1016/j.ijcard.2016.11.182.

Liu KS, Li TP, *et al.* 2018. Advances of Long Noncoding RNAs-mediated Regulation in Reproduction. *Chin Med J (Engl)*, 131(2): 226-234. DOI: 10.4103/0366-6999.222337.

Liu X, Li Y, *et al.* 2020 (a). Long non-coding RNA TTN-AS1 promotes tumorigenesis of ovarian cancer through modulating the miR-139-5p/ROCK2 axis. *Biomed Pharmacother*, 125: 109882. DOI: 10.1016/j.biopha.2020.109882.

Liu Q, Wang Z, *et al.* 2020 (b). A Novel Prognostic Signature of mRNA-lncRNA in Breast Cancer. *DNA Cell Biol*, 39(4): 671-682. DOI: 10.1089/dna.2019.5223.

Livak KJ & Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. *Methods*, 25: 402-408. DOI: 10.1006/meth.2001.1262.

Loda A & Heard E. 2019. Xist RNA in action: Past, present, and future. *PLoS Genet*, 15(9): e1008333. DOI: 10.1371/journal.pgen.1008333.

Lu H, Xu D, *et al.* 2020. RNA-sequencing and bioinformatics analysis of long noncoding RNAs and mRNAs in the asthenozoospermia. *Biosci Rep*, 40(7): BSR20194041. DOI: 10.1042/BSR20194041.

Luangpraseuth-Prosper A, Lesueur E, *et al.* 2015. TOPAZ1, a germ cell specific factor, is essential for male meiotic progression. *Dev Biol*, 406(2): 158-171. DOI: 10.1016/j.ydbio.2015.09.002.

Luk AC, Chan WY, *et al.* 2014. Long noncoding RNAs in spermatogenesis: insights from recent high-throughput transcriptome studies. *Reproduction*, 147(5): R131-R141. DOI: 10.1530/REP-13-0594.

Luk AC, Gao H, *et al.* 2015. GermlncRNA: a unique catalogue of long non-coding RNAs and associated regulations in male germ cell development. *Database (Oxford)*, 2015: bav044. DOI: 10.1093/database/bav044.

Lü M, Tian H, *et al.* 2015. Downregulation of miR-320a/383-sponge-like long non-coding RNA NLC1-C (narcolepsy candidate-region 1 genes) is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation. *Cell Death Dis*, 6(11): e1960. DOI: 10.1038/cddis.2015.267.

Ma L, Bajic VB, *et al.* 2013. On the classification of long non-coding RNAs. *RNA Biol*, 10(6): 925-933. DOI: 10.4161/rna.24604.

Ma X, Cen S, *et al.* 2020. Genome-wide identification and comparison of differentially expressed profiles of miRNAs and lncRNAs with associated ceRNA networks in the gonads of Chinese soft-shelled turtle, *Pelodiscus sinensis*. *BMC Genomics*, 21(1): 443. DOI: 10.1186/s12864-020-06826-1.

Marchese FP & Huarte M. 2014. Long non-coding RNAs and chromatin modifiers: their place in the epigenetic code. *Epigenetics*, 9(1): 21-26. DOI: 10.4161/epi.27472.

Marques AC & Ponting CP. 2014. Intergenic lncRNAs and the evolution of gene expression. *Curr Opin Genet Dev*, 27: 48-53. DOI: 10.1016/j.gde.2014.03.009.

McSwiggin HM & O'Doherty AM. 2018. Epigenetic reprogramming during spermatogenesis and male factor infertility. *Reproduction*, 156(2): R9-R21. DOI: 10.1530/REP-18-0009.

Melé M, Mattioli K, *et al.* 2017. Chromatin environment, transcriptional regulation, and splicing distinguish lincRNAs and mRNAs. *Genome Res*, 27(1): 27-37. DOI: 10.1101/gr.214205.116.

Meikar O, Da Ros M, *et al.* 2011. Chromatoid body and small RNAs in male germ cells. *Reproduction*, 142(2): 195-209. DOI: 10.1530/REP-11-0057.

Meikar O, Vagin VV, *et al.* 2014. An atlas of chromatoid body components. *RNA*, 20: 483-495. DOI: 10.1261/rna.043729.113.

Meistrich ML. 1977. Separation of spermatogenic cells and nuclei from rodent testes. *Methods Cell Biol*, 15:15-54. DOI: 10.1016/s0091-679x(08)60207-1.

Memczak S, Jens M, *et al.* 2013. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441): 333-338. DOI: 10.1038/nature11928.

Mercer TR, Dinger ME, *et al.* 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet*, 10(3): 155-159. DOI: 10.1038/nrg2521.

Moens PB, Marcon E, *et al.* 2007. Initiation and resolution of interhomolog connections: crossover and non-crossover sites along mouse synaptonemal complexes. *J Cell Sci*, 120(6): 1017-1027. DOI: 10.1242/jcs.03394.

Monesi V. 1964. Ribonucleic acid synthesis during mitosis and meiosis in the mouse testis. *J Cell Biol*, 22(3): 521-532. DOI:10.1083/jcb.22.3.521.

Morlando M & Fatica A. 2018. Alteration of Epigenetic Regulation by Long Noncoding RNAs in Cancer. *Int J Mol Sci*, 19(2): 570. DOI: 10.3390/ijms19020570.

Nakagawa S & Kageyama Y. 2014. Nuclear lncRNAs as epigenetic regulators-beyond skepticism. *Biochim Biophys Acta*, 1839(3): 215-222. DOI: 10.1016/j.bbagr.2013.10.009.

Nakajima R, Sato T, *et al.* 2017. A noncoding RNA containing a SINE-B1 motif associates with meiotic metaphase chromatin and has an indispensable function during spermatogenesis. *PLoS One*, 12(6): e0179585. DOI: 10.1371/journal.pone.0179585.

Necsulea A, Soumillon M, *et al.* 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, 505(7485): 635-640. DOI: 10.1038/nature12943.

Niazi F & Valadkhan S. 2012. Computational analysis of functional long noncoding RNAs reveals lack of peptide-coding capacity and parallels with 3'UTRs. *RNA*, 18: 825-843. DOI: 0.1261/rna.029520.111.

Nitsche A & Stadler PF. 2017. Evolutionary clues in lncRNAs. *Wiley Interdiscip Rev RNA*, 8(1): 10.1002/wrna.1376. DOI: 10.1002/wrna.1376.

Noce T, Okamoto-Ito S, *et al.* 2001. Vasa homolog genes in mammalian germ cell development. *Cell Struct Funct*, 26(3): 131-136. DOI: 10.1247/csf.26.131.

Okamoto K. 2012. Epigenetics: a way to understand the origin and biology of testicular germ cell tumors. *Int J Urol*, 19(6): 504-511. DOI: 10.1111/j.1442-2042.2012.02986.x.

Onohara Y, Fujiwara T, *et al.* 2010. Localization of mouse vasa homolog protein in chromatoid body and related nuage structures of mammalian spermatogenic cells during spermatogenesis. *Histochem Cell Biol*, 133(6): 627-639. DOI: 10.1007/s00418-010-0699-5.

Orjalo AV & Johansson HE. 2016. Stellaris® RNA Fluorescence In Situ Hybridization for the Simultaneous Detection of Immature and Mature Long Noncoding RNAs in Adherent Cells. En: Feng Y, Zhang L (eds) *Long Non-Coding RNAs. Methods in Molecular Biology*, 1402: 119-134. Humana Press, New York, NY. DOI: 10.1007/978-1-4939-3378-5\_10.

Orom UA, Derrien T, *et al.* 2010. Long noncoding RNAs with enhancer-like function in human cells. *Cell*, 143(1): 46-58. DOI: 10.1016/j.cell.2010.09.001.

Ozturk S & Uysal F. 2018. Potential roles of the poly(A)-binding proteins in translational regulation during spermatogenesis. *J Reprod Dev*, 64(4): 289-296. DOI: 10.1262/jrd.2018-026.

Page SL & Hawley RS. 2004. The genetics and molecular biology of the synaptonemal complex. *Annu Rev Cell Dev Biol*, 20: 525-558. DOI: 10.1146/annurev.cellbio.19.111301.155141.

Parvinen M & Parvinen LM. 1979. Active movements of the chromatoid body. A possible transport mechanism for haploid gene products. *J Cell Biol*, 80(3): 621-628. DOI: 10.1083/jcb.80.3.621.

Parvinen M. 2005. The chromatoid body in spermatogenesis. *Int Journal of Andrology*, 28: 189-201. DOI: 10.1111/j.1365-2605.2005.00542.x.

Pauli A, Rinn JL, *et al.* 2011. Non-coding RNAs as regulators of embryogenesis. *Nat Rev Genet*, 12(2):136-149. DOI: 10.1038/nrg2904.

Pauli A, Valen E, *et al.* 2012. Systematic identification of long noncoding RNAs expressed during zebrafish embryogenesis. *Genome Res*, 22: 577-591. DOI: 10.1101/gr.133009.111.

Pelechano V & Steinmetz LM. 2013. Gene regulation by antisense transcription. *Nat Rev Genet*, 14(12): 880-893. DOI: 10.1038/nrg3594.

Pertea M, Pertea GM, *et al.* 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*, 33: 290-295. DOI: 10.1038/nbt.3122.

Pertea M, Kim D, *et al.* 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*, 11: 1650-1667. DOI: 10.1038/nprot.2016.095.

Pertea M, Shumate A, *et al.* 2018. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol*, 19: 208. DOI: 10.1186/s13059-018-1590-2.

Petersen C & Soder O. 2006. The sertoli cell--a hormonal target and 'super' nurse for germ cells that determines testicular size. *Horm Res*, 66(4): 153-161. DOI: 10.1159/000094142.

Ponjavic J, Ponting CP, *et al.* 2007. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res*, 17(5): 556-565. DOI: 10.1101/gr.6036807.

Ponjavic J, Oliver PL, *et al.* 2009. Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet*, 5(8): e1000617. DOI: 10.1371/journal.pgen.1000617.

Pontier DB & Gribnau J. 2011. Xist regulation and function eXplored. *Hum Genet*, 130: 223-236. DOI: 10.1007/s00439-011-1008-7.

Ponting CP, Oliver PL, *et al.* 2009. Evolution and functions of long noncoding RNAs. *Cell*, 136: 629-641. DOI: 10.1016/j.cell.2009.02.006.

Quan G & Li J. 2018. Circular RNAs: biogenesis, expression and their potential roles in reproduction. *J Ovarian Res*, 11(1):9. DOI: 10.1186/s13048-018-0381-4.

Quinn JJ & Chang HY. 2016. Unique features of long non-coding RNA biogenesis and function. *Nat Rev Genet*, 17(1): 47-62. DOI:10.1038/nrg.2015.10.

Ran M, Chen B, *et al.* 2016. Systematic identification of long noncoding RNAs in immature and mature porcine testes. *Biol Reprod*, 94: 77. DOI: 10.1095/biolreprod.115.136911.

Ravasi T, Suzuki H, *et al.* 2006. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* 16(1): 11-19. DOI: 10.1101/gr.4200206.

Reuter K, Schlatt S, *et al.* 2012. Fact or fiction: In vitro spermatogenesis. *Spermatogenesis*, 2(4): 245-252. DOI: 10.4161/spmg.21983.

Rinn JL, Kertesz M, *et al.* 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, 129: 1311-1323. DOI: 10.1016/j.cell.2007.05.022.

Rinn JL & Chang HY. 2020. Long Noncoding RNAs: Molecular Modalities to Organismal Functions. *Annu Rev Biochem*, 89: 283-308. DOI: 10.1146/annurev-biochem-062917-012708.

Rivas E, Clements J, *et al.* 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods*, 14(1): 45-48. DOI: 10.1038/nmeth.4066.

Rodríguez-Casuriaga R, Geisinger A, *et al.* 2011. High-purity flow sorting of early meiocytes based on DNA analysis of guinea pig spermatogenic cells. *Cytometry A*, 79: 625-634. DOI: 10.1002/cyto.a.21067.

Rodríguez-Casuriaga R, Folle GA, *et al.* 2013. Simple and efficient technique for the preparation of testicular cell suspensions. *J Vis Exp*, 78: e50102. DOI: 10.3791/50102.

Rodríguez-Casuriaga R, Santiñaque FF, *et al.* 2014. Rapid preparation of rodent testicular cell suspensions and spermatogenic stages purification by flow

cytometry using a novel blue-laser-excitable vital dye. *MethodsX*, 1: 239-243. DOI: 10.1016/j.mex.2014.10.002.

Roeder GS & Bailis JM. 2000. The pachytene checkpoint. *Trends in genetics*, 16(9): 395-403. DOI: 10.1016/s0168-9525(00)02080-1.

Rolland AD, Evrard B, *et al.* 2019. RNA profiling of human testicular cells identifies syntenic lncRNAs associated with spermatogenesis. *Hum Reprod*, 34(7): 1278-1290. DOI: 10.1093/humrep/dez063.

Romrell LJ, Bellvé AR, *et al.* 1976. Separation of mouse spermatogenic cells by sedimentation velocity. A morphological characterization. *Dev Biol*, 49(1): 119-131. DOI: 10.1016/0012-1606(76)90262-1.

Sahlu BW, Zhao S, *et al.* 2020. Long noncoding RNAs: new insights in modulating mammalian spermatogenesis. *J Anim Sci Biotechnol*, 11: 16. DOI: 10.1186/s40104-019-0424-8.

Salemi M, Cannarella R, *et al.* 2019. Evidence for long noncoding RNA GAS5 up-regulation in patients with Klinefelter syndrome. *BMC Med Genet*, 20(1): 4. DOI: 10.1186/s12881-018-0744-0.

Salmena L, Poliseno L, *et al.* 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, 146(3): 353-358. DOI: 10.1016/j.cell.2011.07.014.

Santemma V, Rosati P, *et al.* 1992. Human Sertoli cells in vitro: morphological features and androgen-binding protein secretion. *J Steroid Biochem Mol Biol*, 43(5): 423-429. DOI: 10.1016/0960-0760(92)90080-3.

Sarropoulos I, Marin R, *et al.* 2019. Developmental dynamics of lncRNAs across mammalian organs and species. *Nature*, 571(7766): 510-514. DOI: 10.1038/s41586-019-1341-x.

Sciurano RB, Pigozzi MI, *et al.* 2019. Disassembly of the synaptonemal complex in chicken oocytes analyzed by super-resolution microscopy. *Chromosoma*, 128(3): 443-451. DOI: 10.1007/s00412-019-00693-w.

Scherthan H. 2007. Telomere attachment and clustering during meiosis. *Cell Mol Life Sci*, 64: 117-124. DOI: 10.1007/s00018-006-6463-2.

Schindelin J, Arganda-Carreras I, *et al.* 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9: 676-682. DOI: 10.1038/nmeth.2019.

Schramm S, Fraune J, *et al.* 2011. A novel mouse synaptonemal complex protein is essential for loading of central element proteins, recombination, and fertility. *PLoS Genet*, 7: e1002088. DOI: 10.1371/journal.pgen.1002088.

Schücker K, Holm T, *et al.* 2015. Elucidation of synaptonemal complex organization by super-resolution imaging with isotropic resolution. *Proc Natl Acad Sci USA*, 112(7): 2029-2033. DOI: 10.1073/pnas.1414814112.

Schücker K, Sauer M, *et al.* 2018. Superresolution imaging of the synaptonemal complex. *Methods Cell Biol*, 145: 335-346. DOI: 10.1016/bs.mcb.2018.03.033.

Shafiq S, Li J, *et al.* 2016. Functions of plants long non-coding RNAs. *Biochim Biophys Acta*, 1859(1): 155-162. DOI: 10.1016/j.bbagr.2015.06.009.

Sharlip ID, Jarow JP, *et al.* 2002. Best practice policies for male infertility. *Fertil Steril*, 77(5): 873-882. DOI: 10.1016/s0015-0282(02)03105-9.

Sharma H & Carninci P. 2020. The Secret Life of lncRNAs: Conserved, yet Not Conserved. *Cell*, 181(3): 512-514. DOI: 10.1016/j.cell.2020.04.012.

Shichino Y, Yamashita A, *et al.* 2014. Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. *Open Biol*, 4(6): 140022. DOI: 10.1098/rsob.140022.

Shukla KK, Mahdi AA, *et al.* 2012. Ion channels in sperm physiology and male fertility and infertility. *J Androl*, 33(5): 777-788. DOI: 10.2164/jandrol.111.015552.

Song X, Kyi-Tha-Thu C, *et al.* 2018. 1700108J01Rik and 1700101O22Rik are mouse testis-specific long non-coding RNAs. *Histochem Cell Biol*, 149(5): 517-527. DOI: 10.1007/s00418-018-1642-4.

Sosa E, Flores L, *et al.* 2015. Escape of X-linked miRNA genes from meiotic sex chromosome inactivation. *Development*, 142(21): 3791-3800. DOI: 10.1242/dev.127191.

Soumillon M, Necsulea A, *et al.* 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep*, 3: 2179-2190. DOI: 10.1016/j.celrep.2013.05.031.

Spadaro PA, Flavell CR, *et al.* 2015. Long Noncoding RNA-Directed Epigenetic Regulation of Gene Expression Is Associated With Anxiety-like Behavior in Mice. *Biol Psychiatry*, 78(12): 848-859. DOI: 10.1016/j.biopsych.2015.02.004.

Storey JD & Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*, 100(16): 9440-9445. DOI: 10.1073/pnas.1530509100.

Sun J, Lin Y, *et al.* 2013. Long non-coding RNA expression profiling of mouse testis during postnatal development. *PLoS One*, 8(10): e75750. DOI: 10.1371/journal.pone.0075750.

Sutherland JM, Siddall NA, *et al.* 2015. RNA binding proteins in spermatogenesis: an in depth focus on the Musashi family. *Asian J Androl*, 17(4): 529-536. DOI: 10.4103/1008-682X.151397.

Suzuki S, Diaz VD, *et al.* 2019. What has single-cell RNA-seq taught us about mammalian spermatogenesis? *Biol Reprod*, 101(3): 617-634. DOI: 10.1093/biolre/ioz088.

Syrjänen JL, Pellegrini L, *et al.* 2014. A molecular model for the role of SYCP3 in meiotic chromosome organisation. *Elife*, 3: e02963. DOI: 10.7554/eLife.02963.

Taheri M, Eghtedarian R, *et al.* 2020 (a). Exploring the Role of Non-Coding RNAs in the Pathophysiology of Systemic Lupus Erythematosus. *Biomolecules*, 10(6): 937. DOI: 10.3390/biom10060937.

Taheri M, Eghtedarian R, *et al.* 2020 (b). Dysregulation of non-coding RNAs in Rheumatoid arthritis. *Biomedicine & Pharmacotherapy*, 130: 110617. DOI: 10.1016/j.biopha.2020.110617.

Tanaka H & Baba T. 2005. Gene expression in spermiogenesis. *Cell Mol Life Sci*, 62(3): 344-354. DOI: 10.1007/s00018-004-4394-y.

Toppari J, Kangasniemi M, *et al.* 1991. Stage- and cell-specific gene expression and hormone regulation of the seminiferous epithelium. *J Electron Microsc Tech*, 19(2): 203-214. DOI: 10.1002/jemt.1060190207.

Toyooka Y, Tsunekawa N, *et al.* 2000. Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mech Dev*, 93(1-2): 139-149. DOI: 10.1016/s0925-4773(00)00283-5.

Trapnell C, Roberts A, *et al.* 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*, 7(3): 562-578. DOI: 10.1038/nprot.2012.016.

Tripathi V, Ellis JD, *et al.* 2010. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol Cell*, 39(6): 925-938. DOI: 10.1016/j.molcel.2010.08.011.

Tripathi V, Song DY, *et al.* 2012. SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Mol Biol Cell*, 23(18): 3694-3706. DOI: 10.1091/mbc.E12-03-0206.

Trovero MF & Geisinger A. 2019. Los ARNs no codificantes largos y su vinculación con las patologías testiculares. *AnFaMed*, 6 (1): 10-27. DOI: 10.25184/anfamed2019v6n1a8.

Trovero MF, Rodríguez-Casuriaga R, *et al.* 2020. Revealing stage-specific expression patterns of long noncoding RNAs along mouse spermatogenesis. *RNA Biology*, 17 (3): 350-365. DOI: 10.1080/15476286.2019.1700332.

Tsagakis I, Douka K, *et al.* 2020. Long non-coding RNAs in development and disease: conservation to mechanisms. *J Pathol*, 250(5): 480-495. DOI: 10.1002/path.5405.

Ulitsky I, Shkumatava A, *et al.* 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell*, 147(7): 1537-1550. DOI: 10.1016/j.cell.2011.11.055.

Ulitsky I & Bartel DP. 2013. LincRNAs: genomics, evolution, and mechanisms. *Cell*, 154: 26-46. DOI: 10.1016/j.cell.2013.06.020.

Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet*, 17(10): 601-614. DOI: 10.1038/nrg.2016.85.

Valadkhan S & Nilsen TW. 2010. Reprogramming of the non-coding transcriptome during brain development. *J Biol*, 9(1): 5. DOI: 10.1186/jbiol197.

Vockel M, Riera-Escamilla A, *et al.* 2019. The X chromosome and male infertility. *Hum Genet*, (publicado online, en prensa). DOI: 10.1007/s00439-019-02101-w.

Wang Y, Xue S, *et al.* 2016. Analyses of Long Non-Coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Sci Rep*, 6: 20238. DOI: 10.1038/srep20238.

Wang HV & Chekanova JA. 2017. Long Noncoding RNAs in Plants. *Adv Exp Med Biol*, 1008: 133-154. DOI: 10.1007/978-981-10-5203-3\_5.

Washietl S, Kellis M, *et al.* 2014. Evolutionary dynamics and tissue specificity of human long noncoding RNAs in six mammals. *Genome Res*, 24(4): 616-628. DOI: 10.1101/gr.165035.113.

Weiger TM, Holmqvist MH, *et al.* 2000. A novel nervous system beta subunit that downregulates human large conductance calcium-dependent potassium channels. *J Neurosci*, 20(10): 3563-3570. DOI: 10.1523/JNEUROSCI.20-10-03563.2000.

Weirick T, John D, *et al.* 2015. C-It-Loci: a knowledge database for tissue-enriched loci. *Bioinformatics*, 31(21): 3537-3543. DOI: 10.1093/bioinformatics/btv410.

Weng B, Ran M, *et al.* 2017. Genome-wide analysis of long non-coding RNAs and their role in postnatal porcine testis development. *Genomics*, 109(5-6): 446-456. DOI: 10.1016/j.ygeno.2017.07.001.

White-Cooper H & Davidson I. 2011. Unique aspects of transcription regulation in male germ cells. *Cold Spring Harb Perspect Biol*, 3(7): a002626. DOI: 10.1101/cshperspect.a002626.

Wichman L, Somasundaram S, *et al.* 2017. Dynamic expression of long noncoding RNAs reveals their potential roles in spermatogenesis and fertility. *Biol Reprod*, 97: 313-323. DOI: 10.1093/biolre/iox084.

Winge SB, Dalgaard MD, *et al.* 2018. Transcriptome profiling of fetal Klinefelter testis tissue reveals a possible involvement of long non-coding RNAs in gonocyte maturation. *Hum Mol Genet*, 27(3): 430-439. DOI: 10.1093/hmg/ddx411.

Winkel K, Alsheimer M, *et al.* 2009. Protein SYCP2 provides a link between transverse filaments and lateral elements of mammalian synaptonemal complexes. *Chromosoma*, 118: 259-267. DOI: 10.1007/s00412-008-0194-0.

Wu T & Du Y. 2017. LncRNAs: From Basic Research to Medical Application. *Int J Biol Sci*, 13(3): 295-307. DOI: 10.7150/ijbs.16968.

Xu X, Tan Y, *et al.* 2020. Analysis of Long Noncoding RNA and mRNA Expression Profiles of Testes with High and Low Sperm Motility in Domestic Pigeons (*Columba livia*). *Genes (Basel)*, 11(4): 349. DOI: 10.3390/genes11040349.

Yang CT, Zeng XH, *et al.* 2009. Interactions between beta subunits of the KCNMB family and Slo3: beta4 selectively modulates Slo3 expression and function. *PLoS One*, 4(7): e6135. DOI: 10.1371/journal.pone.0006135.

Yang H, Wang F, *et al.* 2018. Comprehensive analysis of long noncoding RNA and mRNA expression patterns in sheep testicular maturation. *Biol Reprod*, 99(3): 650-661. DOI: 10.1093/biolre/iy088.

Yates AD, Achuthan P, *et al.* 2020 Ensembl 2020. *Nucleic Acids Res*, 48(D1): D682-D688. DOI: 10.1093/nar/gkz966.

Yoneda R, Satoh Y, *et al.* 2016. A genomic region transcribed into a long noncoding RNA interacts with the Prss42/Tessp-2 promoter in spermatocytes during mouse spermatogenesis, and its flanking sequences can function as enhancers. *Mol Reprod Dev*, 83(6): 541-557. DOI: 10.1002/mrd.22650.

Zhao Y, Sun H, *et al.* 2016. Long noncoding RNAs in DNA methylation: new players stepping into the old game. *Cell Biosci*, 6: 45. DOI: 10.1186/s13578-016-0109-3.

Zheng LL, Li JH, *et al.* 2016. deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res*, 44(D1): D196-D202. DOI: 10.1093/nar/gkv1273.

Zhou F, Chen W, *et al.* 2019. Regulation of long non-coding RNAs and circular RNAs in spermatogonial stem cells. *Reproduction*, 158(1): R15-R25. DOI: 10.1530/REP-18-0517.

Znaor A, Lortet-Tieulent J, *et al.* 2014. International variations and trends in testicular cancer incidence and mortality. *Eur Urol*, 65(6): 1095-1106. DOI: 10.1016/j.eururo.2013.11.004.

Zorrilla M & Yatsenko AN. 2013. The Genetics of Infertility: Current Status of the Field. *Curr Genet Med Rep*, 1(4). DOI: 10.1007/s40142-013-0027-1.

## **8. ANEXOS**

## 8.1: Anexo 1: Soluciones, tampones y medios

### **PBS 1X**

NaCl – 8,2 g/l

Na<sub>2</sub>HPO<sub>4</sub> – 1,7 g/l

KH<sub>2</sub>PO<sub>4</sub> – 0,4 g/l

-Ajustar a pH 7,4 con HCl.

-Esterilizar por autoclave.

### **PHEM 1X**

PIPES – 60 mM

HEPES – 25 mM

EGTA – 10 mM

MgCl<sub>2</sub> – 2 mM

-Ajustar a pH 7,4 con NaOH.

### **Tampón TE**

Tris-HCl – 10 mM

EDTA – 1 mM

-Ajustar a pH 8,0.

### **Solución DMEM/colagenasa/DNAsa**

Medio DMEM

Colagenasa - 200 U/ml

DNAsa I filtrada– 5 µg/ml

CaCl<sub>2</sub> – 1,5 mM

### **Solución DMEM/colagenasa/DNAsa/tripsina**

Medio DMEM

Colagenasa - 200 U/ml

DNAsa I filtrada – 5 µg/ml

CaCl<sub>2</sub> – 1,5 mM

Tripsina filtrada – 0,025%

**Medio para cultivos de sobrevida**

MEM alfa (Capricorn Scientific GmbH) – 50 ml

DL- lactato de sodio 60% (m/m) (L4263, Sigma-Aldrich) – 145 µl

Suero fetal bovino – 2,5 ml

HEPES – 0,295 g

-Ajustar a pH 7-7,2 con NaOH.

8.2: Anexo 2: Tabla de lncRNAs diferenciales con FDR p-valor  $\leq 0,05$  y  $|2| > \log_2$   
 $FC \geq |1|$ .

| Gen                  | Cromosoma | Biotipos                       | ID                 | log2 FC | FDR p-val |
|----------------------|-----------|--------------------------------|--------------------|---------|-----------|
| LZ vs 2C             |           |                                |                    |         |           |
| <i>Gm42882</i>       | 5         | antisense                      | ENSMUSG00000106969 | 1,998   | 0,003     |
| <i>Gm45441</i>       | 7         | antisense                      | ENSMUSG00000110020 | 1,996   | 0,016     |
| <i>4930509G22Rik</i> | 16        | lincRNA                        | ENSMUSG00000096959 | 1,984   | 0,003     |
| <i>Tnfsf13os</i>     | 11        | antisense                      | ENSMUSG00000085890 | 1,970   | 0,008     |
| <i>Gm15559</i>       | 5         | antisense                      | ENSMUSG00000086401 | 1,908   | 0,002     |
| <i>Gm31072</i>       | 14        | lincRNA                        | ENSMUSG00000114979 | 1,871   | 0,020     |
| <i>Gm31126</i>       | 13        | lincRNA                        | ENSMUSG00000114035 | 1,857   | 0,034     |
| <i>CT030190.1</i>    | 16        | antisense                      | ENSMUSG00000116589 | 1,822   | 0,001     |
| <i>4933433N18Rik</i> | 13        | lincRNA                        | ENSMUSG00000113227 | 1,815   | 0,015     |
| <i>Gm21388</i>       | 3         | lincRNA                        | ENSMUSG00000114832 | 1,810   | 0,043     |
| <i>1700064M15Rik</i> | 12        | lincRNA                        | ENSMUSG00000100211 | 1,800   | 0,037     |
| <i>1700113A16Rik</i> | 3         | antisense                      | ENSMUSG00000096935 | 1,789   | 0,008     |
| <i>1700100I10Rik</i> | 14        | lincRNA                        | ENSMUSG00000115372 | 1,787   | 0,006     |
| <i>4933427D06Rik</i> | 6         | lincRNA                        | ENSMUSG00000055403 | 1,768   | 0,009     |
| <i>D130020L05Rik</i> | 12        | antisense                      | ENSMUSG00000097121 | 1,750   | 0,016     |
| <i>D930016D06Rik</i> | 5         | lincRNA                        | ENSMUSG00000097392 | 1,746   | 0,002     |
| <i>9230105E05Rik</i> | 10        | antisense                      | ENSMUSG00000090191 | 1,732   | 0,017     |
| <i>BC048644</i>      | 8         | bidirectional_promoter_lincRNA | ENSMUSG00000074052 | 1,710   | 0,015     |
| <i>Gm44009</i>       | 6         | lincRNA                        | ENSMUSG00000107503 | 1,683   | 0,019     |
| <i>4930488L21Rik</i> | 8         | antisense                      | ENSMUSG00000097520 | 1,681   | 0,003     |
| <i>D130017N08Rik</i> | 5         | lincRNA                        | ENSMUSG00000097287 | 1,661   | 0,042     |
| <i>Gm31373</i>       | 1         | lincRNA                        | ENSMUSG00000104487 | 1,641   | 0,016     |
| <i>AC158554.1</i>    | 15        | lincRNA                        | ENSMUSG00000116295 | 1,637   | 0,016     |
| <i>Xist</i>          | X         | lincRNA                        | ENSMUSG00000086503 | 1,634   | 0,005     |
| <i>AY702103</i>      | 17        | lincRNA                        | ENSMUSG00000067103 | 1,629   | 0,014     |
| <i>Gm20512</i>       | 7         | antisense                      | ENSMUSG00000092381 | 1,625   | 0,012     |
| <i>Gm31138</i>       | 9         | lincRNA                        | ENSMUSG00000110491 | 1,538   | 0,002     |
| <i>1700048O20Rik</i> | 9         | lincRNA                        | ENSMUSG00000043773 | 1,538   | 0,004     |
| <i>4921515L22Rik</i> | 10        | antisense                      | ENSMUSG00000112406 | 1,499   | 0,008     |
| <i>Gm11635</i>       | 11        | lincRNA                        | ENSMUSG00000086617 | 1,485   | 0,019     |
| <i>4930412M03Rik</i> | 11        | lincRNA                        | ENSMUSG00000051008 | 1,485   | 0,006     |
| <i>1700003I22Rik</i> | 1         | lincRNA                        | ENSMUSG00000100372 | 1,483   | 0,004     |
| <i>Gm6288</i>        | 6         | antisense                      | ENSMUSG00000072660 | 1,482   | 0,011     |
| <i>Arf4os</i>        | 14        | antisense                      | ENSMUSG00000090861 | 1,481   | 0,017     |
| <i>4930471C06Rik</i> | 4         | lincRNA                        | ENSMUSG00000086138 | 1,464   | 0,009     |
| <i>Gm47077</i>       | 9         | antisense                      | ENSMUSG00000111654 | 1,461   | 0,004     |
| <i>Gm10634</i>       | 9         | antisense                      | ENSMUSG00000074149 | 1,460   | 0,018     |
| <i>Gm6410</i>        | 8         | lincRNA                        | ENSMUSG00000097308 | 1,430   | 0,007     |
| <i>Gm33301</i>       | 15        | lincRNA                        | ENSMUSG00000115367 | 1,398   | 0,040     |
| <i>4930449C09Rik</i> | 15        | lincRNA                        | ENSMUSG00000115676 | 1,395   | 0,002     |

|                      |    |                                |                     |        |       |
|----------------------|----|--------------------------------|---------------------|--------|-------|
| <i>Gm13201</i>       | 4  | antisense                      | ENSMUSG00000078494  | 1,386  | 0,039 |
| <i>2610318M16Rik</i> | 14 | antisense                      | ENSMUSG000000114532 | 1,372  | 0,005 |
| <i>Gm1698</i>        | 8  | lincRNA                        | ENSMUSG000000074404 | 1,355  | 0,048 |
| <i>Gm10069</i>       | 6  | antisense                      | ENSMUSG000000059659 | 1,317  | 0,033 |
| <i>1700108N11Rik</i> | 2  | antisense                      | ENSMUSG000000084837 | 1,314  | 0,039 |
| <i>Gm36423</i>       | 13 | antisense                      | ENSMUSG000000114138 | 1,314  | 0,002 |
| <i>Gm47078</i>       | 10 | lincRNA                        | ENSMUSG000000112275 | 1,314  | 0,017 |
| <i>1700034P13Rik</i> | 1  | antisense                      | ENSMUSG000000097893 | 1,298  | 0,029 |
| <i>Gm31181</i>       | 10 | lincRNA                        | ENSMUSG000000112202 | 1,274  | 0,037 |
| <i>1700026N04Rik</i> | 13 | bidirectional_promoter_lincRNA | ENSMUSG000000114681 | 1,272  | 0,007 |
| <i>1700080G11Rik</i> | 4  | lincRNA                        | ENSMUSG000000086771 | 1,269  | 0,002 |
| <i>1700019G24Rik</i> | 6  | antisense                      | ENSMUSG000000085416 | 1,261  | 0,004 |
| <i>1700030M09Rik</i> | 8  | bidirectional_promoter_lincRNA | ENSMUSG000000097424 | 1,258  | 0,005 |
| <i>1110020A21Rik</i> | 17 | lincRNA                        | ENSMUSG000000097047 | 1,247  | 0,006 |
| <i>1700019C18Rik</i> | 13 | lincRNA                        | ENSMUSG000000113770 | 1,241  | 0,010 |
| <i>Gm17227</i>       | 19 | antisense                      | ENSMUSG000000091558 | 1,238  | 0,006 |
| <i>Gm21917</i>       | 15 | antisense                      | ENSMUSG000000096111 | 1,237  | 0,013 |
| <i>4930522O17Rik</i> | 4  | antisense                      | ENSMUSG000000086947 | 1,234  | 0,013 |
| <i>1700102H20Rik</i> | 17 | lincRNA                        | ENSMUSG000000116852 | 1,229  | 0,029 |
| <i>1700095J03Rik</i> | 7  | antisense                      | ENSMUSG000000085407 | 1,220  | 0,006 |
| <i>Gm35584</i>       | 3  | lincRNA                        | ENSMUSG000000102936 | 1,213  | 0,049 |
| <i>2700046G09Rik</i> | 19 | lincRNA                        | ENSMUSG000000097787 | 1,212  | 0,048 |
| <i>2810029C07Rik</i> | 12 | bidirectional_promoter_lincRNA | ENSMUSG000000096995 | 1,204  | 0,010 |
| <i>2700012I20Rik</i> | 9  | lincRNA                        | ENSMUSG000000097334 | 1,192  | 0,024 |
| <i>1700026J12Rik</i> | 16 | lincRNA                        | ENSMUSG000000073976 | 1,168  | 0,027 |
| <i>Astx5</i>         | X  | lincRNA                        | ENSMUSG000000095521 | 1,166  | 0,029 |
| <i>1700021F13Rik</i> | 5  | lincRNA                        | ENSMUSG000000107053 | 1,159  | 0,012 |
| <i>4930558J18Rik</i> | 1  | lincRNA                        | ENSMUSG000000097519 | 1,155  | 0,005 |
| <i>1700022N22Rik</i> | 17 | lincRNA                        | ENSMUSG000000097523 | 1,150  | 0,020 |
| <i>4930555A03Rik</i> | 3  | lincRNA                        | ENSMUSG000000106157 | 1,148  | 0,021 |
| <i>4930455D15Rik</i> | 18 | lincRNA                        | ENSMUSG000000097553 | 1,141  | 0,006 |
| <i>Gm26787</i>       | 9  | antisense                      | ENSMUSG000000096993 | 1,135  | 0,003 |
| <i>1700023F02Rik</i> | 10 | lincRNA                        | ENSMUSG000000100000 | 1,127  | 0,006 |
| <i>0610040B10Rik</i> | 5  | antisense                      | ENSMUSG000000089889 | 1,121  | 0,007 |
| <i>Gm26663</i>       | 4  | lincRNA                        | ENSMUSG000000097019 | 1,091  | 0,005 |
| <i>4930505N22Rik</i> | 19 | lincRNA                        | ENSMUSG000000097690 | 1,089  | 0,007 |
| <i>1700040D17Rik</i> | 3  | lincRNA                        | ENSMUSG000000097515 | 1,081  | 0,019 |
| <i>1700045H11Rik</i> | 4  | antisense                      | ENSMUSG000000078492 | 1,076  | 0,035 |
| <i>4921523L03Rik</i> | 9  | antisense                      | ENSMUSG000000093749 | 1,068  | 0,002 |
| <i>1700116B05Rik</i> | 16 | lincRNA                        | ENSMUSG000000092276 | 1,041  | 0,004 |
| <i>4930594O21Rik</i> | 3  | lincRNA                        | ENSMUSG000000085677 | 1,033  | 0,003 |
| <i>Aknaos</i>        | 4  | antisense                      | ENSMUSG000000086953 | 1,025  | 0,019 |
| <i>Neat1</i>         | 19 | lincRNA                        | ENSMUSG000000092274 | -1,056 | 0,010 |
| <i>Gm3892</i>        | 4  | antisense                      | ENSMUSG000000095348 | -1,355 | 0,003 |
| <i>Gm38186</i>       | 3  | antisense                      | ENSMUSG000000103668 | -1,496 | 0,030 |
| <i>Gm10593</i>       | 4  | antisense                      | ENSMUSG000000095258 | -1,514 | 0,002 |

|                      |    |                |                    |        |       |
|----------------------|----|----------------|--------------------|--------|-------|
| <i>Gm14230</i>       | 2  | lincRNA        | ENSMUSG00000087179 | -1,574 | 0,037 |
| <i>Snhg20</i>        | 11 | lincRNA        | ENSMUSG00000086859 | -1,620 | 0,026 |
| <i>Fam120aos</i>     | 13 | lincRNA        | ENSMUSG00000097059 | -1,649 | 0,012 |
| <i>Epb41l4aos</i>    | 18 | lincRNA        | ENSMUSG00000087590 | -1,711 | 0,018 |
| <i>1810058l24Rik</i> | 6  | lincRNA        | ENSMUSG00000073155 | -1,714 | 0,007 |
| <i>5031434C07Rik</i> | 6  | antisense      | ENSMUSG00000044574 | -1,768 | 0,020 |
| <i>Gm47920</i>       | 14 | sense_intronic | ENSMUSG00000114709 | -1,795 | 0,002 |
| <i>Gm5084</i>        | 13 | lincRNA        | ENSMUSG00000050765 | -1,851 | 0,005 |
| <i>Airn</i>          | 17 | antisense      | ENSMUSG00000078247 | -1,954 | 0,002 |
| <i>Gm15545</i>       | 7  | antisense      | ENSMUSG00000087138 | -1,997 | 0,034 |

#### PS vs LZ

|                      |    |                               |                    |       |       |
|----------------------|----|-------------------------------|--------------------|-------|-------|
| <i>Gm28402</i>       | 6  | antisense                     | ENSMUSG00000101651 | 1,974 | 0,041 |
| <i>Gm15343</i>       | 10 | antisense                     | ENSMUSG00000085468 | 1,957 | 0,040 |
| <i>4930500F10Rik</i> | 9  | antisense                     | ENSMUSG00000097592 | 1,949 | 0,028 |
| <i>1700030A11Rik</i> | 17 | antisense                     | ENSMUSG00000085333 | 1,942 | 0,038 |
| <i>Gm31267</i>       | 10 | lincRNA                       | ENSMUSG00000112016 | 1,930 | 0,040 |
| <i>4930404K13Rik</i> | 14 | sense_intronic                | ENSMUSG00000115639 | 1,893 | 0,045 |
| <i>Fam219aos</i>     | 4  | antisense                     | ENSMUSG00000036162 | 1,892 | 0,026 |
| <i>4930448A20Rik</i> | 7  | lincRNA                       | ENSMUSG00000108382 | 1,862 | 0,038 |
| <i>4930589P08Rik</i> | 4  | antisense                     | ENSMUSG00000085593 | 1,854 | 0,041 |
| <i>Gm43391</i>       | 5  | antisense                     | ENSMUSG00000106004 | 1,835 | 0,042 |
| <i>1700100l10Rik</i> | 14 | lincRNA                       | ENSMUSG00000115372 | 1,832 | 0,044 |
| <i>4930505N22Rik</i> | 19 | lincRNA                       | ENSMUSG00000097690 | 1,786 | 0,039 |
| <i>Gm31138</i>       | 9  | lincRNA                       | ENSMUSG00000110491 | 1,757 | 0,045 |
| <i>Gm26708</i>       | 7  | bidirectional_promoter_lncRNA | ENSMUSG00000097016 | 1,752 | 0,034 |
| <i>4933416M07Rik</i> | 8  | antisense                     | ENSMUSG00000097262 | 1,717 | 0,049 |
| <i>4930402F11Rik</i> | 7  | lincRNA                       | ENSMUSG00000108885 | 1,623 | 0,044 |
| <i>Gm44696</i>       | 6  | lincRNA                       | ENSMUSG00000107875 | 1,590 | 0,026 |
| <i>4933406K04Rik</i> | 12 | lincRNA                       | ENSMUSG00000085820 | 1,559 | 0,045 |
| <i>Gm6729</i>        | 10 | lincRNA                       | ENSMUSG00000112586 | 1,501 | 0,026 |

#### RS vs PS

|                      |    |           |                    |       |       |
|----------------------|----|-----------|--------------------|-------|-------|
| <i>4930513D17Rik</i> | 5  | lincRNA   | ENSMUSG00000106741 | 1,999 | 0,003 |
| <i>Gm37821</i>       | 1  | lincRNA   | ENSMUSG00000104188 | 1,988 | 0,007 |
| <i>1700034K08Rik</i> | 9  | lincRNA   | ENSMUSG00000100294 | 1,980 | 0,007 |
| <i>4930524O07Rik</i> | 9  | antisense | ENSMUSG00000090207 | 1,973 | 0,026 |
| <i>4933424L21Rik</i> | 11 | antisense | ENSMUSG00000085719 | 1,970 | 0,006 |
| <i>Gm26725</i>       | 5  | antisense | ENSMUSG00000097092 | 1,959 | 0,010 |
| <i>2610318M16Rik</i> | 14 | antisense | ENSMUSG00000114532 | 1,944 | 0,012 |
| <i>Gm12843</i>       | 4  | antisense | ENSMUSG00000085749 | 1,937 | 0,011 |
| <i>Gm40155</i>       | 3  | antisense | ENSMUSG00000106629 | 1,929 | 0,003 |
| <i>Gm47381</i>       | 13 | lincRNA   | ENSMUSG00000114688 | 1,929 | 0,012 |
| <i>4930579C12Rik</i> | 9  | lincRNA   | ENSMUSG00000074146 | 1,923 | 0,012 |
| <i>Gm14424</i>       | 2  | lincRNA   | ENSMUSG00000085015 | 1,914 | 0,048 |
| <i>4930583P06Rik</i> | 2  | lincRNA   | ENSMUSG00000086954 | 1,904 | 0,022 |
| <i>Gm36423</i>       | 13 | antisense | ENSMUSG00000114138 | 1,889 | 0,013 |
| <i>1700020G17Rik</i> | 10 | antisense | ENSMUSG00000101801 | 1,880 | 0,001 |

|                      |    |                                |                    |        |       |
|----------------------|----|--------------------------------|--------------------|--------|-------|
| <i>Gm41279</i>       | 15 | lincRNA                        | ENSMUSG00000115335 | 1,876  | 0,010 |
| <i>4930535O05Rik</i> | 8  | lincRNA                        | ENSMUSG00000098360 | 1,859  | 0,001 |
| <i>Gm44933</i>       | 7  | lincRNA                        | ENSMUSG00000108864 | 1,851  | 0,002 |
| <i>Gm15904</i>       | 4  | antisense                      | ENSMUSG00000085961 | 1,847  | 0,004 |
| <i>4930533I22Rik</i> | 6  | antisense                      | ENSMUSG00000093535 | 1,845  | 0,005 |
| <i>Gm33104</i>       | 10 | lincRNA                        | ENSMUSG00000110902 | 1,837  | 0,009 |
| <i>Gm2511</i>        | 7  | lincRNA                        | ENSMUSG00000108695 | 1,835  | 0,035 |
| <i>AC164425.1</i>    | 16 | lincRNA                        | ENSMUSG00000116799 | 1,834  | 0,037 |
| <i>4930455M05Rik</i> | 13 | lincRNA                        | ENSMUSG00000114787 | 1,819  | 0,002 |
| <i>Mir670hg</i>      | 2  | lincRNA                        | ENSMUSG00000075020 | 1,772  | 0,012 |
| <i>Far1os</i>        | 7  | antisense                      | ENSMUSG00000084984 | 1,758  | 0,027 |
| <i>4921522P10Rik</i> | 8  | lincRNA                        | ENSMUSG00000040467 | 1,752  | 0,003 |
| <i>Gm568</i>         | 4  | lincRNA                        | ENSMUSG00000048824 | 1,746  | 0,022 |
| <i>AC127568.1</i>    | 10 | lincRNA                        | ENSMUSG00000116971 | 1,745  | 0,002 |
| <i>1700054A03Rik</i> | 19 | lincRNA                        | ENSMUSG00000099988 | 1,739  | 0,012 |
| <i>2700012I20Rik</i> | 9  | lincRNA                        | ENSMUSG00000097334 | 1,718  | 0,043 |
| <i>4930432O09Rik</i> | 10 | lincRNA                        | ENSMUSG00000112846 | 1,713  | 0,028 |
| <i>1700116B05Rik</i> | 16 | lincRNA                        | ENSMUSG00000092276 | 1,708  | 0,042 |
| <i>Gm10138</i>       | 1  | antisense                      | ENSMUSG00000100954 | 1,697  | 0,032 |
| <i>Gm15343</i>       | 10 | antisense                      | ENSMUSG00000085468 | 1,684  | 0,047 |
| <i>4932416K20Rik</i> | 8  | lincRNA                        | ENSMUSG00000069925 | 1,666  | 0,010 |
| <i>1700019G24Rik</i> | 6  | antisense                      | ENSMUSG00000085416 | 1,666  | 0,013 |
| <i>Gm7008</i>        | 12 | antisense                      | ENSMUSG00000035983 | 1,654  | 0,018 |
| <i>Haglrl</i>        | 2  | bidirectional_promoter_lincRNA | ENSMUSG00000075277 | 1,636  | 0,005 |
| <i>Gm47900</i>       | 10 | lincRNA                        | ENSMUSG00000112379 | 1,621  | 0,003 |
| <i>4921504A21Rik</i> | 5  | antisense                      | ENSMUSG00000097626 | 1,552  | 0,024 |
| <i>Kcnab3os</i>      | 11 | antisense                      | ENSMUSG00000086900 | 1,539  | 0,015 |
| <i>Gm13001</i>       | 4  | lincRNA                        | ENSMUSG00000087417 | 1,522  | 0,032 |
| <i>2810029C07Rik</i> | 12 | bidirectional_promoter_lincRNA | ENSMUSG00000096995 | 1,502  | 0,013 |
| <i>CT025155.1</i>    | 16 | lincRNA                        | ENSMUSG00000116871 | 1,484  | 0,029 |
| <i>4930556J02Rik</i> | 14 | lincRNA                        | ENSMUSG00000115656 | 1,477  | 0,041 |
| <i>Arl4aos</i>       | 12 | antisense                      | ENSMUSG00000086315 | 1,467  | 0,029 |
| <i>Gm31314</i>       | 5  | lincRNA                        | ENSMUSG00000106938 | 1,457  | 0,014 |
| <i>Gm26708</i>       | 7  | bidirectional_promoter_lincRNA | ENSMUSG00000097016 | 1,447  | 0,031 |
| <i>AC169509.2</i>    | 16 | sense_intronic                 | ENSMUSG00000115869 | 1,392  | 0,019 |
| <i>4930403D09Rik</i> | 11 | antisense                      | ENSMUSG00000087122 | 1,113  | 0,022 |
| <i>2010007H06Rik</i> | 9  | antisense                      | ENSMUSG00000044694 | 1,047  | 0,009 |
| <i>4930519H02Rik</i> | 5  | lincRNA                        | ENSMUSG00000105185 | 1,024  | 0,007 |
| <i>Gm31105</i>       | 8  | antisense                      | ENSMUSG00000109587 | -1,049 | 0,023 |
| <i>Gm44696</i>       | 6  | lincRNA                        | ENSMUSG00000107875 | -1,107 | 0,021 |
| <i>Gm17552</i>       | 19 | lincRNA                        | ENSMUSG00000097233 | -1,198 | 0,003 |
| <i>Gm31904</i>       | 10 | lincRNA                        | ENSMUSG00000112219 | -1,217 | 0,021 |
| <i>Gm14152</i>       | 2  | lincRNA                        | ENSMUSG00000087495 | -1,223 | 0,038 |
| <i>1810058I24Rik</i> | 6  | lincRNA                        | ENSMUSG00000073155 | -1,234 | 0,046 |
| <i>4930442J19Rik</i> | 2  | lincRNA                        | ENSMUSG00000087609 | -1,249 | 0,019 |
| <i>Gm5922</i>        | 9  | lincRNA                        | ENSMUSG00000111028 | -1,303 | 0,012 |

|                      |    |                               |                     |        |       |
|----------------------|----|-------------------------------|---------------------|--------|-------|
| <i>AC163291.3</i>    | 15 | bidirectional_promoter_lncRNA | ENSMUSG000000116402 | -1,304 | 0,033 |
| <i>Gm6288</i>        | 6  | antisense                     | ENSMUSG000000072660 | -1,318 | 0,012 |
| <i>1700040D17Rik</i> | 3  | lincRNA                       | ENSMUSG000000097515 | -1,338 | 0,050 |
| <i>Gm45441</i>       | 7  | antisense                     | ENSMUSG000000110020 | -1,365 | 0,035 |
| <i>Gm15966</i>       | 16 | lincRNA                       | ENSMUSG000000085865 | -1,370 | 0,039 |
| <i>4933433N18Rik</i> | 13 | lincRNA                       | ENSMUSG000000113227 | -1,371 | 0,028 |
| <i>E130317F20Rik</i> | 10 | bidirectional_promoter_lncRNA | ENSMUSG000000091994 | -1,442 | 0,031 |
| <i>Gm13817</i>       | 2  | lincRNA                       | ENSMUSG000000085293 | -1,462 | 0,002 |
| <i>BC048644</i>      | 8  | bidirectional_promoter_lncRNA | ENSMUSG000000074052 | -1,479 | 0,010 |
| <i>Gm20621</i>       | 15 | sense_overlapping             | ENSMUSG000000093394 | -1,490 | 0,007 |
| <i>AC116487.2</i>    | 15 | lincRNA                       | ENSMUSG000000116288 | -1,540 | 0,017 |
| <i>Gm17232</i>       | 14 | antisense                     | ENSMUSG000000090374 | -1,595 | 0,007 |
| <i>4932702P03Rik</i> | 13 | antisense                     | ENSMUSG000000086796 | -1,615 | 0,003 |
| <i>Gm15559</i>       | 5  | antisense                     | ENSMUSG000000086401 | -1,714 | 0,002 |
| <i>Gm41335</i>       | 15 | lincRNA                       | ENSMUSG000000115520 | -1,736 | 0,018 |
| <i>4930524O08Rik</i> | 9  | lincRNA                       | ENSMUSG000000098627 | -1,743 | 0,016 |
| <i>Gm16973</i>       | 14 | lincRNA                       | ENSMUSG000000097743 | -1,775 | 0,032 |
| <i>Gm14167</i>       | 2  | antisense                     | ENSMUSG000000087433 | -1,778 | 0,024 |
| <i>Gm31181</i>       | 10 | lincRNA                       | ENSMUSG000000112202 | -1,793 | 0,015 |
| <i>Gm14144</i>       | 2  | lincRNA                       | ENSMUSG000000087337 | -1,797 | 0,007 |
| <i>Cep83os</i>       | 10 | bidirectional_promoter_lncRNA | ENSMUSG000000097164 | -1,894 | 0,013 |
| <i>1700029M20Rik</i> | 4  | lincRNA                       | ENSMUSG000000086788 | -1,905 | 0,007 |
| <i>4930463O16Rik</i> | 10 | lincRNA                       | ENSMUSG000000020033 | -1,925 | 0,004 |
| <i>Gm32031</i>       | 7  | bidirectional_promoter_lncRNA | ENSMUSG000000109118 | -1,959 | 0,019 |

