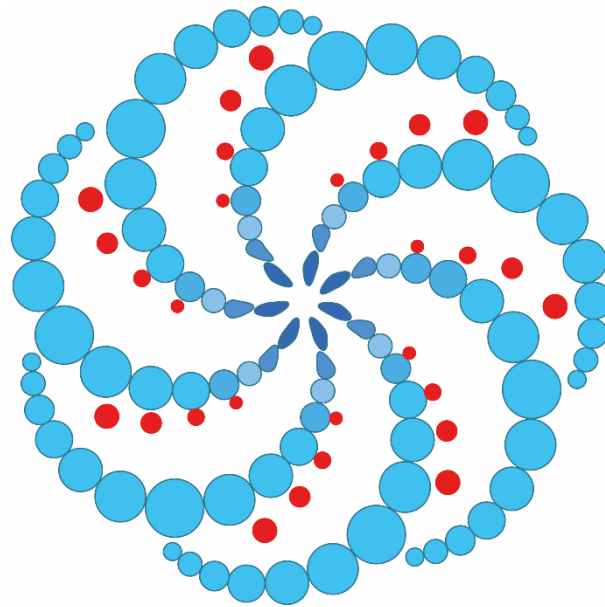


## TESINA DE GRADO EN CIENCIAS BIOLÓGICAS

# Localización sub-celular de ARNs no codificantes largos, candidatos a cumplir funciones estructurales y/o regulatorias durante la espermatogénesis



Mateo Alberto François López  
Orientadora: María Fernanda Trovero. Co-orientadora: Adriana Geisinger  
Laboratorio de Biología Molecular de la Reproducción  
(Departamento de Biología Molecular),  
Instituto de Investigaciones Biológicas Clemente Estable

10 de octubre del 2022

## **Agradecimientos:**

Agradezco especialmente a mi orientadora Fernanda Trovero por aceptarme como tesista permitiéndome aportar y colaborar en la línea de investigación que con tanto esfuerzo y esmero desarrolló durante su tesis de doctorado, y por los valiosos consejos que me transmitió durante todo el desarrollo de este trabajo. De igual forma, agradezco a mi co-orientadora Adriana Geisinger por su apoyo incondicional, por su preocupación constante en mi bienestar y por buscar los mejores caminos para realizar este proyecto. Agradezco a ambas por incluir parte de los resultados de esta tesis en una publicación en una revista internacional arbitrada de alto impacto, siendo esta mi primera publicación científica.

Quiero agradecerle también a la Dra. Rosana Rodríguez-Casuriaga, por siempre estar a entera disposición al momento de realizar nuevas técnicas, aclarar dudas y por ayudarnos durante la visualización de resultados al microscopio.

Agradezco al Dr. Ricardo Benavente, quien, a pesar de estar en el extranjero, colaboró durante el análisis de resultados y brindó materiales indispensables para este trabajo.

También quiero agradecerle a Adrián Ortiz, compañero de carrera, colega y amigo de la vida, que gracias a él se me presentó la oportunidad de integrarme a este grupo de trabajo y a esta línea de investigación.

Agradezco a todos los compañeros de laboratorio del Departamento de Biología Molecular del IIBCE, por siempre ayudarme en lo que necesitara, y por mantener un clima de trabajo familiar y colaborativo que hicieron de mis días de mesada una experiencia mucho más amena.

Por último, agradecerle a mi familia por siempre impulsarme y darme la oportunidad de tener una educación universitaria, que sin su apoyo hubiese sido imposible.

A todos Gracias.

## **Resumen:**

La espermatogénesis es el proceso por el cual las células de la línea germinal experimentan importantes cambios morfológicos y genéticos que conllevan a la formación de un espermatozoide maduro. Durante la espermatogénesis, las células están ligadas a una importante regulación postranscripcional, en la cual los ARNs no codificantes (ncRNAs), y en particular los ARNs no codificantes largos (lncRNAs), han comenzado a destacar por su potencial rol como reguladores de distintos procesos biológicos. Dado que el testículo es el tejido que presenta mayor expresión y diversidad de este tipo de transcriptos y que los mismos presentan patrones de expresión muy restringidos en el tiempo y tipo celular, se sugiere que por lo menos una parte de estos transcriptos debería tener un rol importante durante el desarrollo testicular y en la espermatogénesis.

Con el objetivo de aportar nuevos conocimientos sobre las bases moleculares de la reproducción masculina, en lo relativo al potencial rol de secuencias no codificantes, se seleccionaron tres lncRNAs con distintas características para el estudio de su localización sub-celular. Como resultados obtuvimos que los tres lncRNAs seleccionados (*Rbakdn*, *1110020A21Rik* y *Kcnmb4os1*) se localizaban sobre el cuerpo cromatoide (CC) de espermátidas redondas. El CC es un organelo sin membrana específico de células posmeióticas, para el cual se propone un importante rol en la regulación postranscripcional de dichas células. Adicionalmente, se determinó que el lncRNA antisentido solapante *Kcnmb4os1* se localiza en el CC con su respectivo mRNA antisentido *Kcnmb4*. Dado que ambos transcriptos poseen regiones complementarias cabe la posibilidad de que *Kcnmb4os1* secuestre a *Kcnmb4* y lo redirija al CC, tratándose de un novedoso mecanismo de regulación postranscripcional. Por último, el lncRNA *Kcnmb4os1* también mostró una señal puntual dentro del núcleo de los espermátocitos, localizándose muy cerca de uno de los extremos de un cromosoma de tamaño mediano. Podría tratarse del mismo cromosoma donde este lncRNA está codificado, y acumularse sobre su propio *locus*, similar a lo que se ha reportado para varios lncRNAs en la levadura *Schizosaccharomyces pombe*, donde esos lncRNAs median la recombinación homóloga.

En este estudio logramos estudiar la localización sub-celular de los tres lncRNAs seleccionados en células espermátogénicas mediante distintas metodologías, y así elaborar hipótesis sobre sus posibles mecanismos de acción y procesos en los que estén involucrados. Sin embargo, aún quedan muchos interrogantes por responder, para los cuales se plantean estudios a futuro, como ensayos funcionales mediante el silenciamiento *in vivo* y *chromosome painting* que logren ampliar nuestro conocimiento sobre los lncRNAs.

# Índice

|                                                                                         |           |
|-----------------------------------------------------------------------------------------|-----------|
| <b>1. Introducción.....</b>                                                             | <b>1</b>  |
| <b>1.1. Testículo y espermatogénesis .....</b>                                          | <b>1</b>  |
| <b>1.1.1 Estructura y función del testículo.....</b>                                    | <b>1</b>  |
| <b>1.1.2 La espermatogénesis: formación de los gametos masculinos .....</b>             | <b>4</b>  |
| <b>1.1.3 La meiosis .....</b>                                                           | <b>6</b>  |
| <b>1.1.4 Espermioogénesis.....</b>                                                      | <b>9</b>  |
| <b>1.2 ARNs no codificantes .....</b>                                                   | <b>12</b> |
| <b>1.2.1 Funciones de los lncRNAs .....</b>                                             | <b>15</b> |
| <b>1.2.2 lncRNAs durante la espermatogénesis.....</b>                                   | <b>17</b> |
| <b>1.3 Antecedentes del grupo de trabajo.....</b>                                       | <b>22</b> |
| <b>2. Hipótesis y objetivos .....</b>                                                   | <b>22</b> |
| <b>2.1 Hipótesis .....</b>                                                              | <b>22</b> |
| <b>2.2 Objetivo general.....</b>                                                        | <b>22</b> |
| <b>2.3 Objetivos específicos .....</b>                                                  | <b>23</b> |
| <b>3. Materiales y Métodos .....</b>                                                    | <b>23</b> |
| <b>3.1 Selección de lncRNAs candidatos.....</b>                                         | <b>23</b> |
| <b>3.2 Análisis de expresión mediante RT-qPCR.....</b>                                  | <b>24</b> |
| <b>3.2.1 Diseño de cebadores.....</b>                                                   | <b>24</b> |
| <b>3.2.2 RT-qPCR .....</b>                                                              | <b>25</b> |
| <b>3.2.3 Eficiencia de cebadores .....</b>                                              | <b>26</b> |
| <b>3.2.4 Análisis de expresión de los transcritos estudiados.....</b>                   | <b>26</b> |
| <b>3.3 Localización sub-celular.....</b>                                                | <b>27</b> |
| <b>3.3.1 Diseño de sondas de ARN Stellaris®.....</b>                                    | <b>27</b> |
| <b>3.3.2 Animales y eutanasia .....</b>                                                 | <b>29</b> |
| <b>3.3.3 Preparación de tejido congelado y criosecciones de tejido testicular .....</b> | <b>30</b> |
| <b>3.3.4 Preparación de esparcidos (<i>spreading</i>) de suspensión celular .....</b>   | <b>30</b> |
| <b>3.3.5 Preparación de buffers para RNA-FISH e inmunodetección .....</b>               | <b>31</b> |
| <b>3.3.6 Anticuerpos.....</b>                                                           | <b>32</b> |
| <b>3.3.7 Protocolo RNA-FISH para tejido congelado.....</b>                              | <b>32</b> |
| <b>3.3.8 Protocolo RNA-FISH e inmunodetección para tejido congelado .....</b>           | <b>33</b> |

|                                                                                                                               |    |
|-------------------------------------------------------------------------------------------------------------------------------|----|
| 3.3.9 Protocolo RNA-FISH e inmunodetección para <i>spreading</i> de suspensión celular .....                                  | 34 |
| 3.3.10 Visualización de los preparados .....                                                                                  | 34 |
| 4. Resultados .....                                                                                                           | 35 |
| 4.1 IncRNAs seleccionados .....                                                                                               | 35 |
| 4.2 Diseño de cebadores y eficiencia de los mismos .....                                                                      | 36 |
| 4.3 Validación de patrones de expresión mediante RT-qPCR .....                                                                | 38 |
| 4.4 Validación de <i>RNA-FISH</i> con <i>Malat1</i> como control positivo .....                                               | 40 |
| 4.5 Localización sub-celular de los IncRNAs estudiados: ¿cuerpo cromatoide? ..                                                | 42 |
| 4.6 Co-localización en el cuerpo cromatoide del mRNA <i>Kcnmb4</i> con su IncRNA antisentido solapante <i>Kcnmb4os1</i> ..... | 45 |
| 4.7 El IncRNA <i>Kcnmb4os1</i> se localiza dentro del núcleo de los espermatoцитos                                            | 48 |
| 5. Discusión .....                                                                                                            | 52 |
| 5.1 Los IncRNAs durante la espermatogénesis y su relación con el cuerpo cromatoide .....                                      | 52 |
| 5.2 Los IncRNAs antisentido solapantes y los transcritos codificantes .....                                                   | 55 |
| 5.3 <i>Kcnmb4os1</i> se localiza dentro del núcleo de los espermatoцитos .....                                                | 57 |
| 5.4 ¿LncRNAs “ <i>moonlighting</i> ” ? .....                                                                                  | 59 |
| 6. Conclusiones y perspectivas .....                                                                                          | 60 |
| 7. Bibliografía .....                                                                                                         | 61 |

# **1. Introducción:**

## **1.1. Testículo y espermatogénesis**

### **1.1.1 Estructura y función del testículo:**

Los testículos son los órganos sexuales, en donde se lleva a cabo la formación de los gametos masculinos (espermatozoides) en el proceso llamado espermatogénesis. Durante la espermatogénesis las células de la línea germinal, las cuales se diferencian muy tempranamente en el desarrollo embrionario, experimentan cambios morfológicos y genéticos que conllevan a la formación de los gametos masculinos. Estas células serán todo lo que un organismo podrá aportar a la siguiente generación, por lo que la correcta formación de las mismas es de suma importancia para la conservación y prolongación de una especie. Cada testículo está constituido por una cápsula que contiene en su interior numerosos túbulos seminíferos, separados uno del otro por tejido intersticial. Esta cápsula, que envuelve los túbulos, se encuentra formada por tres capas de tejido: una capa más externa que es continua con el peritoneo visceral, denominada túnica vaginalis; una capa intermedia, formada en su mayor parte por tejido conectivo fibroelástico denso y células del músculo liso, llamada túnica albugínea; y una capa más interna, formada por tejido conectivo laxo, muy rica en vasos sanguíneos, denominada túnica vascular. Estas tres capas, además de contener a los túbulos seminíferos, también regulan el flujo sanguíneo que llega al testículo y, en algunas especies, son capaces de contraerse de forma rítmica o bajo ciertos estímulos, con la función de colaborar en la liberación del esperma desde los túbulos seminíferos (Creasy & Chapin, 2013).

Dentro cada túbulo seminífero es donde se lleva a cabo la formación de los gametos masculinos. La disposición interna del testículo y el número de túbulos seminíferos, varían según la especie. En el ratón, los túbulos seminíferos se disponen en forma de bucles contorneados cuyos extremos se encuentran unidos a un reservorio en forma de saco denominado "*rete testis*", sobre uno de los polos del testículo. Si bien cada bucle tiene numerosas circunvalaciones, la mayoría se encuentran unos paralelos a los otros, y se extienden a lo largo del eje del testículo (Figura 1A). Por lo tanto, si cortamos una sección histológica transversal del testículo, observaríamos a los túbulos seminíferos en forma de círculos o anillos (Figura 1B), mientras que si realizásemos un corte que siga la dirección del eje testicular, tendríamos secciones longitudinales de los túbulos (Paz Rodríguez, 2016). La gran capacidad de los túbulos seminíferos de agruparse y contornearse dentro

del testículo, permite que en los ratones se puedan tener 12 metros de túbulos seminíferos en un gramo de tejido testicular, y hasta 24 metros por gramo en los seres humanos (Creasy & Chapin, 2013).

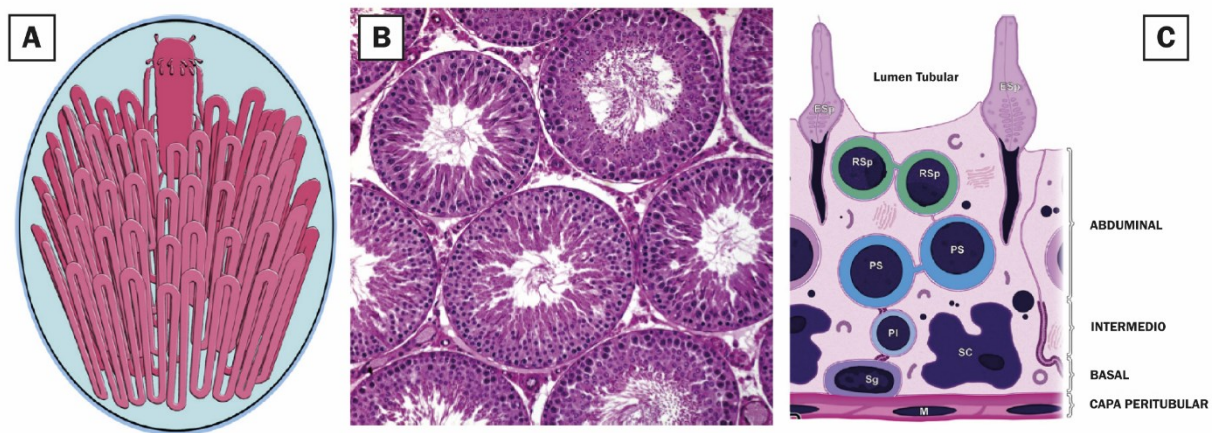
Los túbulos seminíferos se encuentran separados por tejido intersticial, el cual está formado fundamentalmente por células de tejido conectivo como fibroblastos, vasos sanguíneos, canales linfáticos, células del sistema inmune residentes de tejido, así como también las células de Leydig, fundamentales para la producción y secreción de testosterona (Kerr *et al.*, 2006).

La parte más externa de los túbulos seminíferos está conformada por una capa peritubular que consta de la llamada lámina propia, y de células peritubulares denominadas mioideas. La lámina propia actúa en su cara interna, como membrana basal a un epitelio pluriestratificado muy complejo formado por las células germinales y células somáticas denominadas células de Sertoli, encargadas del apoyo metabólico y estructural de las primeras (Figura 1C). Por su parte, las células peritubulares tienen múltiples funciones, pero principalmente se encargan de regular de forma paracrina a las células de Leydig y de Sertoli, así como de brindar las propiedades contráctiles a los túbulos seminíferos (Creasy & Chapin, 2018).

La correcta funcionalidad de las células de Sertoli es fundamental para la integridad de los túbulos seminíferos y para el correcto proceso de la gametogénesis. Estas son las únicas células somáticas dentro de los túbulos seminíferos y, como mencionamos anteriormente, son las encargadas de ofrecer sustento estructural y metabólico a las células de la línea germinal. Las células de Sertoli mantienen una forma de columna irregular; en su parte basal se encuentran ancladas a la lámina propia mediante hemidesmosomas y en su parte basolateral se encuentran unidas entre sí mediante uniones oclusivas especializadas, las cuales son el componente más fuerte de la barrera hematotesticular. Esta barrera hematotesticular mantiene el interior de los túbulos impermeable a muchas moléculas y compuestos que circulan por la sangre, y que podrían dañar o alterar a las células germinales, así como también las mantiene aisladas del propio sistema inmune del organismo, que podría reconocer a las células germinales como células extrañas (Creasy & Chapin, 2018). Además, las células de Sertoli son las encargadas de crear el microambiente necesario para el desarrollo y formación de los gametos mediante una estricta regulación endocrina y paracrina (Petersen & Soder, 2006), generan el movimiento de las células germinales desde la parte basal del tubo seminífero hasta su liberación en el lumen, colaboran en el suministro constante de nutrientes y oxígeno a las células

germinales, y se encargan de procesar y fagocitar los desechos de las células germinales (Creasy & Chapin, 2018).

Las células de la línea germinal se encuentran incrustadas en invaginaciones laterales o apicales de las células de Sertoli en una íntima relación de uniones y señalización. Las células germinales se van desarrollando y diferenciando hacia el interior del túbulo, donde la parte apical de las células de Sertoli determina el comienzo del lumen, es decir, el espacio dentro del túbulo seminífero al cual serán liberados los espermatozoides (Creasy & Chapin, 2018).



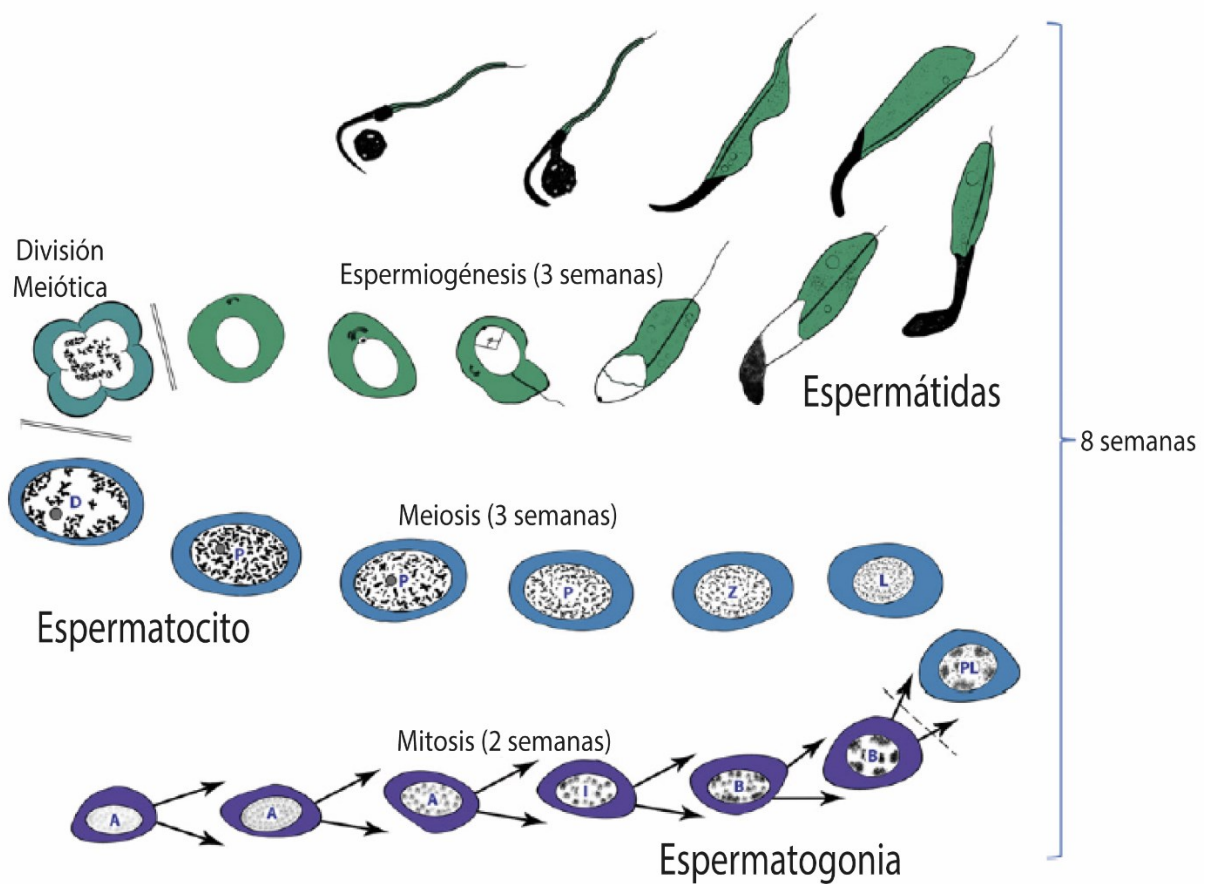
**Figura 1:** Estructura general del testículo. En (A) se muestran la estructura del testículo y disposición de los túbulos seminíferos en testículo de rata. En (B) se observa un corte histológico de testículo de rata, donde pueden apreciarse las secciones transversales de los túbulos seminíferos, separados por tejido intersticial. La figura (C) muestra una representación del epitelio pluriestratificado dentro de cada túbulo seminífero. En la parte basal observamos a las células mioides (M) junto a la lámina propia y a los núcleos de las células de Sertoli (SC) entre las cuales se encuentran invaginadas las células germinales como: espermatogonias tipo A (Sg), espermatocitos en preleptoteno (PI), espermatocitos en paquitenio (PS), espermátidas redondas (RSp) y espermátidas elongadas (ESp). La maduración de las células germinales se da desde la parte basal del túbulo seminífero hacia el lumen del túbulo, donde son liberados los espermatozoides. Imágenes tomadas y adaptadas de Creasy & Chapin, 2018.

### **1.1.2 La espermatogénesis: formación de los gametos masculinos**

La espermatogénesis comienza a partir de la maduración sexual del individuo, y se continúa durante toda la vida adulta. Se inicia a partir de células madre espermatogoniales, las que pasan por una serie de etapas que involucran rondas de mitosis, meiosis y posteriores rondas de diferenciación (espermiogénesis), hasta llegar a formar el gameto masculino: el espermatozoide (Figura 2). Estas etapas de diferenciación pueden dividirse a grandes rasgos en tres procesos que se dan de forma consecutiva a partir de una célula germinal, pero que ocurren en simultáneo a nivel del túbulo seminífero (Creasy & Chapin, 2013).

En primer lugar, se da la proliferación de las espermatogonias mediante mitosis; estas células yacen sobre la membrana basal, pero no están unidas a ésta. Las células madre de espermatogonias surgen durante las primeras etapas del desarrollo y se auto-mantienen mediante duplicación, conservándose en un número relativamente fijo durante gran parte de la vida del organismo (Oatley & Brinster, 2012). En esta primera etapa podemos distinguir tres tipos fundamentales de espermatogonias: las células madre de espermatogonias, las espermatogonias proliferativas y las espermatogonias a diferenciarse. Las primeras dos son las encargadas de mantener una población constante de células, mientras que el tercer grupo continúa por un camino de diferenciación hacia espermatogonia tipo A. Las espermatogonias tipo A pueden dividirse en cuatro sub grupos que van desde las espermatogonias A1 hasta las A4, cada una descendiente de la anterior. Este último grupo A4 finalmente se dividirá para formar espermatogonias intermedias, y posteriormente espermatogonias tipo B. A lo largo de todas estas divisiones mitóticas, la etapa de citocinesis se da de forma incompleta, lo que ocasiona que las células descendientes permanezcan formando un sincitio continuo, lo cual podría estar ayudando en la sincronización de la diferenciación y división de estas poblaciones celulares (Creasy & Chapin, 2013).

Los distintos tipos de espermatogonia pueden diferenciarse entre sí a lo largo de estas divisiones ya que varían levemente la distribución de su heterocromatina en el proceso, por ejemplo, si bien las espermatogonias tipo A son de tamaño y forma similar a las de tipo B, las primeras muestran núcleos más claros, con su heterocromatina agrupada más suavemente en forma de nube, mientras que las espermatogonias tipo B presentan núcleos más intensos, con su heterocromatina compactada densamente sobre la periferia del núcleo (Creasy & Chapin, 2013). Las espermatogonias tipo B darán paso a la segunda etapa de la espermatogénesis: la meiosis.



**Figura 2:** Ilustración de la espermatogénesis. Se muestran las distintas etapas y los tipos celulares involucrados durante la espermatogénesis, donde podemos distinguir: en primer lugar, la etapa proliferativa de las espermatogonias mediante mitosis; luego la etapa meiótica de los espermatocitos; y finalmente la etapa de diferenciación de las espermátidas, la espermiogénesis, hasta llegar a un espermatozoide maduro. Imagen obtenida y modificada de Creasy & Chapin, 2013.

### **1.1.3 La meiosis:**

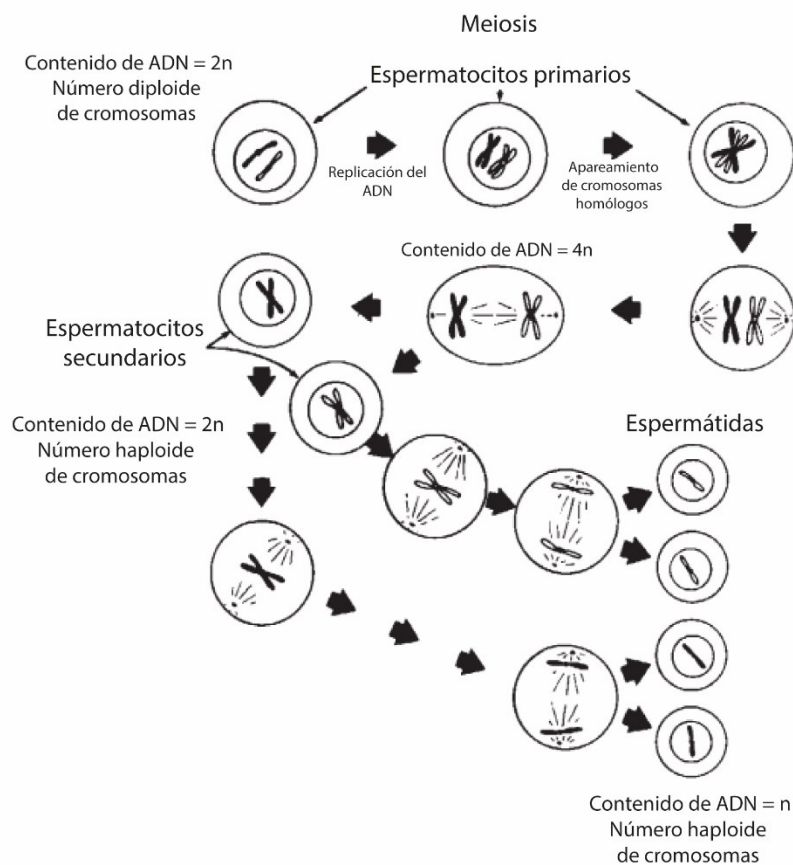
La meiosis es la segunda etapa dentro de la espermatogénesis, y es un tipo de división celular especializada presente en todos los organismos con reproducción sexual, la cual en vertebrados ocurre únicamente en las células de la línea germinal. En el testículo la meiosis comienza a partir de la diferenciación de las espermatogonias tipo B en espermatoцитos. Estos espermatoцитos sufrirán dos divisiones meióticas (meiosis I y II) con una única previa duplicación de su material genético. La primera división meiótica es reduccional ya que el contenido de ADN de los espermatoцитos, de  $4C$  ( $2N$ ) pasará a ser de  $2C$  ( $1N$ ) tras la segregación de los cromosomas homólogos. La segunda división es ecuacional y separa a las cromátidas hermanas entre sí, resultando en células de contenido de ADN  $1C$  ( $1N$ ). Al finalizar ambas divisiones, se obtienen cuatro células haploides con un único juego de cromosomas (Figura 3) (Handel & Schimenti, 2010).

Cada una de las dos divisiones meióticas se divide de igual forma que la mitosis en profase, metafase, anafase y telofase, distinguiendo entre I y II para cada una de estas etapas, según sea la división de la que se trate. Sin embargo, durante la meiosis ocurren procesos muy distintos a los de la mitosis. En particular durante la meiosis I, que se da en los espermatoцитos primarios ( $2N$ ,  $ADN=4C$ ), se destaca la profase I, la cual es más larga que las otras etapas y se subdivide en cinco estadios: leptoteno, cigoteno, paquiteno, diploteno y diacinesis. Durante el primero de estos estadios, el leptoteno, se comienzan a organizar los llamados elementos axiales (que originarán los elementos laterales del complejo sinaptonémico, ver a continuación) a lo largo de cada cromosoma, a la vez que los telómeros se unen a la envoltura nuclear, donde se agrupan en un sector limitado de la envoltura nuclear y comienzan a formar una conformación cromosómica denominada *bouquet*, ya que resulta similar a un ramo de flores (Scherthan, 2007).

Durante el cigoteno comienza el apareamiento de los cromosomas homólogos, al mismo tiempo que los elementos axiales comienzan a constituir los elementos laterales (ELs) de la estructura proteica de los complejos sinaptonémicos (CSs), responsables del correcto apareamiento y recombinación de los cromosomas homólogos (Cohen *et al.*, 2006). Durante estas dos primeras etapas se pueden llegar a identificar los sitios donde se dará el intercambio de hebras, denominados nódulos de recombinación tempranos (Zickler & Kleckner, 1999).

Respecto del CS, este es un complejo proteico altamente conservado en la evolución, que se encuentra presente desde levaduras hasta mamíferos. Su estructura está conformada por dos ELs, un elemento central (EC), y filamentos transversos, encargados

de mantener unidos entre sí a los ELs y el EC (Page & Hawley, 2004). Hasta el momento, en los mamíferos se conocen ocho proteínas que forman parte del CS: en sus ELs encontramos a SYCP2 y SYCP3, en los filamentos transversales a SCYP1, y en el EC a las proteínas SYCE1, SYCE2, SYCE3, TEX12 y C14ORF39/SIX6OS1 (Fraune *et al.*, 2012; Gómez-H *et al.*, 2016). El CS participa en el proceso de apareamiento y recombinación, pero hasta ahora no se conoce el mecanismo exacto que utiliza para llevarlo a cabo; por lo pronto se ha demostrado que algunas de las proteínas que lo componen, como SYCP3, son capaces de unir ADN, mientras que otras como SCYP2 y SYCP1 presentan dominios de unión al ADN (Peltari *et al.*, 2001; Winkel *et al.*, 2009; Syrjänen *et al.*, 2014; Bollschweiler *et al.*, 2019). También se ha propuesto la posibilidad de que en este proceso participen ARNs no codificantes, pero aún no se sabe con certeza.



**Figura 3:** Diagrama ilustrativo de la etapa meiótica durante la espermatogénesis, enfocándose en la primera y segunda división meióticas. Durante la primera división meiótica se da la separación de los cromosomas homólogos, mientras que en la segunda división las cromátidas hermanas se separan para dar lugar a la formación de cuatro espermátidas. Imagen tomada y modificada de Kerr *et al.*, 2006.

Hacia el estadio de paquitenio los cromosomas homólogos se encuentran totalmente apareados y el CS totalmente formado, dando lugar entonces al proceso de recombinación homóloga, también llamado *crossing-over*, en donde se produce un intercambio recíproco del material genético entre cromosomas homólogos. Los sitios donde efectivamente se dio el intercambio de material genético pueden ser identificados como nódulos de recombinación tardíos, los cuales son menores en número respecto de los nódulos de recombinación tempranos, ya que no todos son resueltos como *cross-overs*. Durante el estadio de diploteno, el CS comienza a desensamblarse; ante la pérdida de unión brindada en parte por el CSs, los cromosomas quedan unidos únicamente por los sitios donde efectivamente hubo un evento de recombinación, comúnmente conocidos como quiasmas (Cohen *et al.*, 2006; Bolcun-Filas & Handel, 2018). Finalmente, durante la diacinesis, última etapa de la profase meiótica I, los cromosomas comienzan a migrar hacia la periferia del núcleo, al mismo tiempo que se disuelven los quiasmas, haciendo que los cromosomas queden unidos únicamente por sus extremos. En este estadio comienza la formación del huso y se desensambla la envoltura nuclear, preparándose para la entrada en la metafase I. Las siguientes etapas de la meiosis estarán marcadas por la separación de los cromosomas homólogos en dos células hijas denominados espermatocitos secundarios ( $1N$ ,  $ADN=2C$ ). Finalmente, la segunda división meiótica ocasionará la separación de las cromátidas hermanas teniendo como resultado cuatro células haploides ( $1N$ ,  $ADN=C$ ), conocidas como espermátidas redondas (Kerr *et al.*, 2006; Bolcun-Filas & Handel, 2018).

El proceso de apareamiento y recombinación de cromosomas homólogos que sucede durante la profase I de la meiosis es una de las fuentes principales de biodiversidad a nivel genético en las especies con reproducción sexual, y tiene un papel crucial para la segregación cromosómica posterior. Siendo un mecanismo tan importante, las modificaciones genéticas que pudiesen alterar estos procesos suelen tener fuertes consecuencias sobre la fertilidad del organismo o sobre su descendencia. Para esto, existen puntos de control (*check-points*) a lo largo de la meiosis, fundamentalmente en la transición paquitenio/diploteno, en los cuales las alteraciones importantes son detectadas, generalmente desencadenando el arresto meiótico y la apoptosis, y provocando, en el peor de los casos, infertilidad en el organismo (Handel & Schimenti, 2010; Cohen *et al.*, 2006).

#### **1.1.4 Espermiogénesis:**

Como resultado de la última división meiótica, surgen las espermatídas redondas. Estas células entrarán en la tercera y última etapa de la espermatogénesis, donde sufrirán diversos cambios morfológicos y genéticos, en el proceso denominado espermiogénesis. La espermiogénesis no involucra ningún tipo de división celular; esencialmente ésta es un proceso de diferenciación terminal, en donde células con una forma convencional se convierten en células altamente especializadas en el movimiento y la fecundación (Kerr *et al.*, 2006).

Las espermatídas recién formadas madurarán a espermatídas elongadas y finalmente a espermatozoides, los cuales serán liberados a la luz del túbulo seminífero. Sin embargo, para que una espermatída llegue a ser un espermatozoide deberá modificar profundamente su forma. En primer lugar, se formará la vesícula acrosómica a partir del aparato de Golgi, y se generará una caperuza sobre la superficie nuclear. El acrosoma queda posicionado entre la membrana celular y el núcleo de la espermatída. Paso seguido, el núcleo cambia de posición mediante rotación, lo que genera que el acrosoma quede mirando hacia el lado de la membrana basal y deja al centriolo hacia el lado del lumen tubular. Esta rotación es necesaria ya que a partir del centriolo comenzará a formarse el flagelo, que se extenderá hacia la luz del lumen (Kerr *et al.*, 2006).

De forma consecutiva, el núcleo comienza a cambiar de forma al elongarse y compactarse al condensar su cromatina. Dicha compactación es producto de reemplazar sus histonas por otras proteínas: en primer lugar, por proteínas de transición y finalmente por protaminas, las cuales compactan enormemente su ADN, resultando en un silenciamiento transcripcional generalizado (Kleene, 2001). Finalmente, gran parte del citoplasma es eliminado en conjunto con muchos de sus organelos en un cuerpo residual que es fagocitado por las células de Sertoli, mientras que las mitocondrias migran hacia la base del flagelo, donde formarán un anillo con el fin de colaborar energéticamente en el movimiento del mismo. Tras todos estos pasos de diferenciación celular, se forman los espermatozoides maduros, los cuales son finalmente liberados al lumen del túbulo seminífero (Kerr *et al.*, 2006) y migran hacia el epidídimo para su capacitación (adquisición de capacidad fecundante). Estos importantes cambios de forma y compactación ayudarán al espermatozoide a cumplir su cometido, que es el de llegar hasta un gameto femenino y fecundarlo (Creasy & Chapin, 2018).

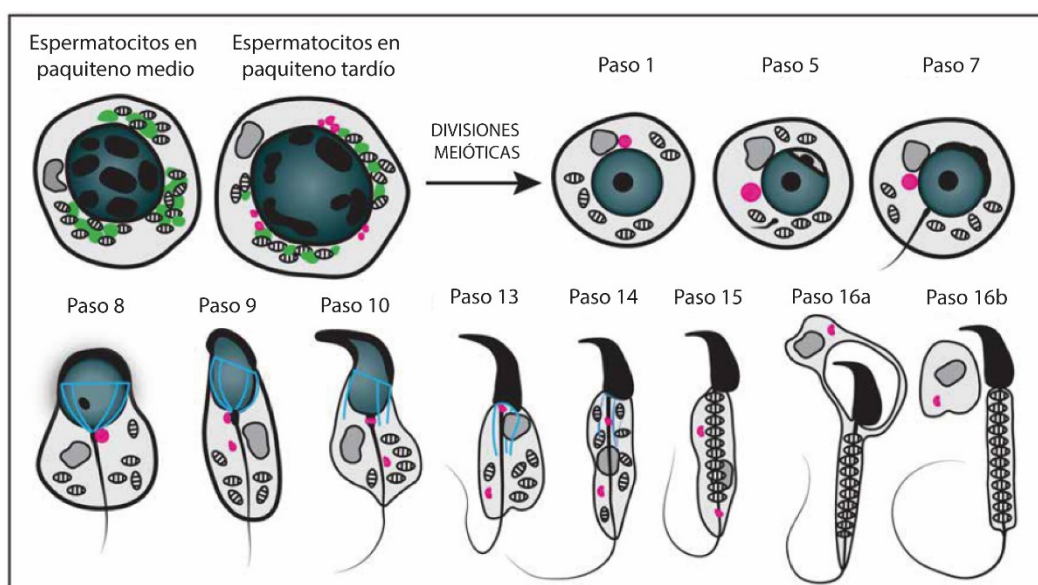
De entre los eventos mencionados, un suceso que nos interesa destacar es la fuerte compactación de la cromatina en las espermatídas tardías. La compactación del ADN, al

sustituir histonas por protaminas, tiene la principal ventaja de volver a los espermatozoides células más hidrodinámicas, además de proteger el material genético de agresiones químicas que pudiese sufrir durante el trayecto que deben recorrer estas células para llevar a cabo la fecundación. Sin embargo, tiene una desventaja no menor, que es el total silenciamiento transcripcional, consecuencia de la imposibilidad de la maquinaria transcripcional de acceder al material genético. Por lo tanto, los ARNs mensajeros (mRNAs) de las proteínas necesarias para las etapas avanzadas de la espermiogénesis y para los espermatozoides deben ser sintetizados en etapas previas, y su traducción reprimida hasta el momento que sea necesario para no alterar el normal desarrollo espermático (Eddy, 2002; Gan *et al.*, 2013). Un ejemplo de esto se da con los mRNAs que codifican para las propias protaminas; los mismos son sintetizados en espermátidas redondas, almacenados y reprimidos hasta ser traducidos más tarde en las espermátidas elongadas (Kleene, 1993), y su expresión anticipada está indicada como causa de infertilidad (Lee *et al.*, 1995).

Por lo tanto, durante la espermiogénesis, las espermátidas cuentan con algunas estrategias de regulación postranscripcional, con el fin de almacenar mRNAs necesarios en etapas posteriores. Pueden, por ejemplo, retener a los mRNAs dentro de ribonucleoproteínas, unir represores a las regiones UTR de los transcritos e impedir la unión de la maquinaria de traducción, regular el largo de su cola poliadenilada en el extremo 3', así como también pueden ser retenidos en organelos sin membrana (Kleene, 2001). Estos organelos sin membrana, conocidos como gránulos germinales, están formados por diversos tipos de ARNs y proteínas de unión al ARN, y presentes en la vasta mayoría de las células germinales de metazoarios con reproducción sexual. Es sabido de su participación en regular la localización, estabilidad y traducción de ARNs, así como también de mediar la producción y función de muchos ARNs no codificantes cortos (Voronina *et al.*, 2011; Lehtiniemi & Kotaja, 2018). Si bien se han descrito diferentes tipos de gránulos germinales en distintos organismos, muchos de éstos comparten componentes homólogos indicando que sus funciones son compartidas y mantenidas en la escala zoológica (Chuma *et al.*, 2009).

En los testículos de los mamíferos, los dos gránulos germinales más estudiados son el cemento intermitocondrial (IMC) y el denominado cuerpo cromatoide (CC). El IMC es encontrado formando clústeres alrededor de las mitocondrias de muchos tipos celulares durante la diferenciación de las células germinales, pero es particularmente evidente en los espermátocitos en paquíteno, donde coexiste con otros gránulos más pequeños que se creen los precursores del CC (Lehtiniemi & Kotaja, 2018). Algunos autores han sugerido

que el propio IMC podría ser precursor del CC, pero todavía no está claro (Chuma *et al.*, 2006). Por su parte, el CC es un organelo sin membrana que se conoce hace ya más de 100 años, fundamentalmente por su gran tamaño que le permite ser visualizado a través de un microscopio óptico (Brunn, 1876). Es detectado en las espermatídias redondas localizado en posición perinuclear. A lo largo de la espermiogénesis el CC cambia de posición, quedando sobre la base del flagelo en las espermatídias redondas tardías. Hacia el final de esta fase, el CC reduce significativamente su volumen y se divide en dos, de tal forma que durante la etapa de espermatídida elongada una de estas mitades será descartada dentro del cuerpo residual, mientras que la otra formará un anillo alrededor de la base del flagelo, que se cree estar involucrado con la formación del anillo de mitocondrias en las posteriores etapas de la espermiogénesis (Figura 4) (Lehtiniemi & Kotaja, 2018).



**Figura 4:** Diagrama de la diferenciación de la línea germinal durante la espermatogénesis, desde espermatocitos en paquíteno hasta espermatozoides maduros. Se muestra un modelo sobre la disposición del IMC (verde), presente en espermatocitos en paquíteno, posicionado entre las mitocondrias alrededor del núcleo celular. Los precursores del CC (rosado) aparecen en espermatocitos tardíos y coexisten con el IMC. Tras la meiosis el CC se condensa en un único gránulo presente en espermatídias redondas (pasos 1, 5 y 7). Durante la elongación de las espermatídias el CC se posiciona sobre la base del flagelo (paso 8), y finalmente se divide en dos mitades, una de las cuales será descartada y la otra formará un anillo en la base del flagelo (pasos 9 al 16). En azul se muestra la estructura transitoria que forman los microtúbulos durante estas etapas, y en gris se muestran las mitocondrias y el aparato de Golgi. Tomado y modificado de Lehtiniemi & Kotaja, 2018.

Si bien la función exacta de estos gránulos germinales aún no ha sido dilucidada en su totalidad, el IMC y CC comparten componentes estructurales similares por lo que se estima que podrían compartir o colaborar entre ambos en funciones similares. Se presume que, en el caso del IMC, éste podría principalmente actuar como un sitio de procesamiento primario para ARNs de unión a proteínas PIWI (piRNAs) durante la meiosis, que luego serían transferidos al CC. Para el caso del CC, éste no sólo estaría involucrado en el procesamiento de piRNAs, sino que también participaría en la regulación, almacenamiento y degradación de muchos otros ARNs involucrados en una gran variedad de procesos biológicos. Si bien el rol específico que cumple el CC es desconocido hasta el momento, en la última década han surgido investigaciones que han colaborado en ir develando el misterio alrededor de este organelo sin membrana. En particular, se ha demostrado que el CC retiene una considerable porción de los mRNAs transcriptos en espermátidas redondas, así como proteínas de unión al ARN y componentes de la maquinaria de decaimiento de mRNAs aberrantes (*non-sense mediated decay*, NMD) (Meikar *et al.*, 2014; Lehtiniemi & Kotaja, 2018).

## **1.2 ARNs no codificantes:**

La perspectiva con la que se observa el genoma ha cambiado a lo largo del tiempo; tradicionalmente se consideraba que las proteínas y los genes que las codifican eran quienes determinaban la mayor parte de las funciones en la célula. Sin embargo, actualmente este dogma ha cambiado gracias a la aparición de nuevas tecnologías como las de secuenciación masiva (NGS) y estudios de transcriptómica como el RNA-seq. Estas tecnologías han demostrado que, contrario a lo que se creía, la mayor parte del genoma es transcripto y que los genes codificantes sólo representarían una pequeña parte de la expresión de una célula (Luk *et al.*, 2014; Amodio *et al.*, 2018). La proporción de genoma que no codifica proteínas varía enormemente entre distintos organismos, y en algunos casos puede ser muy alta, como por ejemplo en los seres humanos, donde representa hasta un 98,5% del genoma (Amaral *et al.*, 2018). Durante mucho tiempo se consideró a esta parte no codificante del genoma como basura transcripcional sin funciones importantes, hasta que en las últimas décadas comenzaron a surgir investigaciones demostrando la funcionalidad e importancia biológica de estas secuencias no codificantes, siendo hoy en día un área de fuerte investigación (Amodio *et al.*, 2018). Estas secuencias que son transcriptas, pero no traducidas a proteínas, son denominadas ARNs no codificantes (ncRNAs). Los ncRNAs pueden clasificarse de acuerdo a su tamaño en ncRNAs cortos

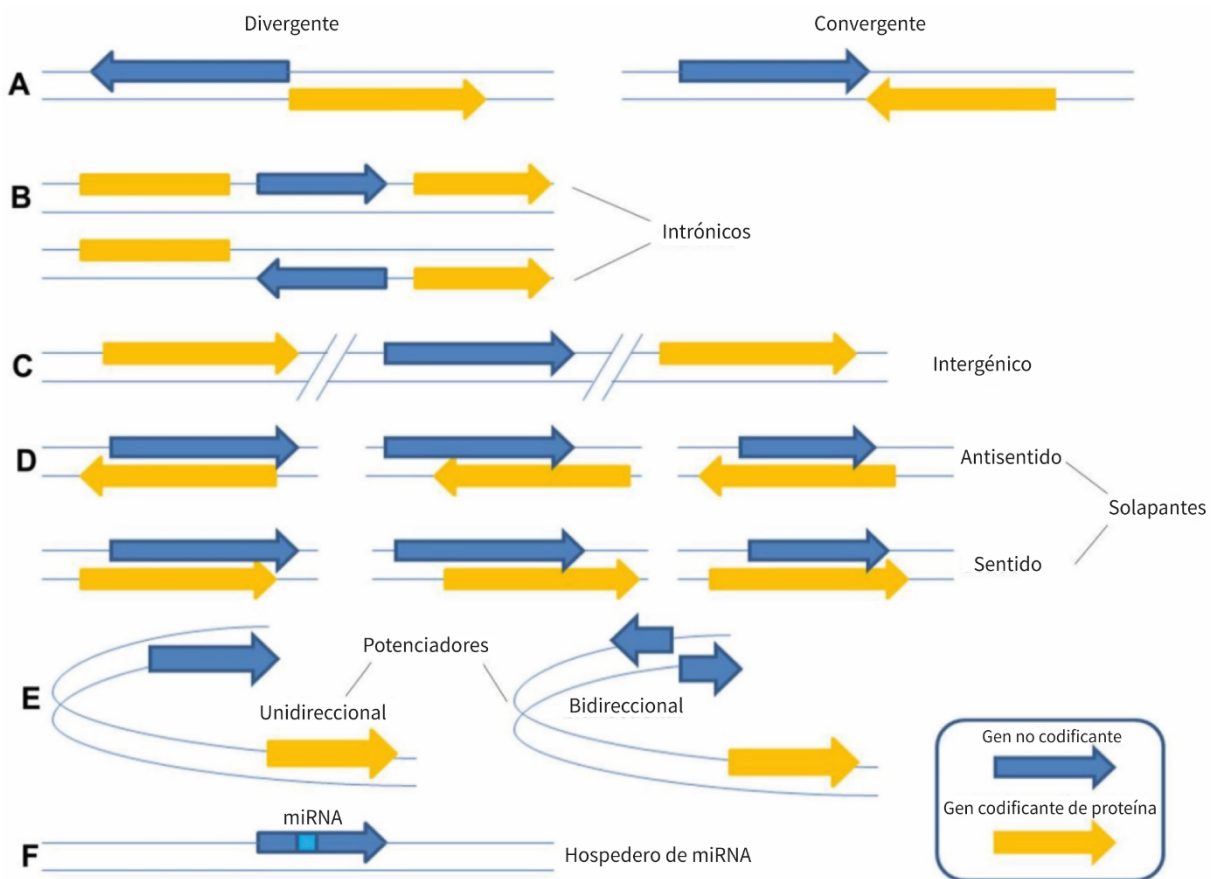
(sncRNAs), con una longitud menor a 200 pb y en ncRNAs largos (lncRNAs), definidos arbitrariamente como aquellos que poseen más de 200 pb (Luk *et al.*, 2014).

Por su parte, existen varios tipos de ARNs no codificantes cortos como son los ARNs ribosomales 5S (rRNAs), los piRNAs, micro ARNs (miRNAs) o ARNs de interferencia (siRNAs), los cuales han sido intensamente estudiados en las últimas décadas y han sido asociados con distintas vías de silenciamiento de genes específicos, así como a la protección del genoma celular contra virus, repetidos móviles de ADN y transposones, entre otros (Chen *et al.*, 2007; Gomes *et al.*, 2013).

En el caso de los ARNs no codificantes largos, estos han sido históricamente menos estudiados que los sncRNAs, pero en las últimas décadas han llamado el interés de muchos investigadores por su potencial rol en distintos procesos biológicos. Los lncRNAs comparten una serie de características dentro de los vertebrados. En cuanto a su estructura, los lncRNAs son sintetizados en su gran mayoría por la ARN polimerasa II (enzima utilizada en la síntesis de ARNs mensajeros) y como consecuencia en general presentan un casquete de 7-metilguanosina en su extremo 5' y una cola poliadenilada en su extremo 3'. Además, poseen pocos exones, tienen un bajo contenido de citosina y guanina, y sus secuencias están muy poco conservadas en comparación a los ARNs codificantes de proteínas. Asimismo, estos transcriptos en general son expresados en menor nivel y por tanto son mucho menos abundantes que los ARNs codificantes. Sin embargo, los lncRNAs poseen un patrón de expresión altamente acotado en tiempo y espacio, es decir, que éstos se expresan en determinados momentos y en determinados tipos celulares, lo que, en comparación a los ARNs codificantes, hace a los lncRNAs altamente específicos de tejido, de tipo celular y de un momento particular del desarrollo (Dinger *et al.*, 2008; Guttman *et al.*, 2010; Cabili *et al.*, 2011; Pauli *et al.*, 2012).

En comparación a los sncRNA, la definición de lncRNAs es más compleja y abarca muchos tipos de transcriptos distintos. Los lncRNAs son difíciles de clasificar debido a que su procesamiento postranscripcional puede ocurrir de múltiples formas, así como también presentar un amplio espectro de funciones celulares distintas (Wu & Du, 2017). Como consecuencia, los lncRNAs son tradicionalmente clasificados según su localización en el genoma respecto de los genes codificantes más cercanos. Es así como se los divide fundamentalmente en ocho categorías detalladas a continuación: divergentes, los cuales se originan desde la hebra opuesta a la región promotora del gen codificante y son transcriptos en dirección opuesta al mismo; convergentes, que se originan sobre la hebra opuesta, pero son transcritos en dirección tal que enfrenta al gen codificante (Figura 5A);

intrónicos, los cuales son sintetizados desde la secuencia de un intrón dentro de un gen codificante (Figura 5B); intergénicos, que se originan a partir de distancias lejanas de otros genes codificantes, generalmente a más de 10 kb del gen más cercano (Figura 5C); sentido solapantes, cuya secuencia esta solapada con un gen codificante y va en la misma dirección que éste (Figura 5D); antisentido solapantes, que de igual forma solapan con un gen, pero estos lo hacen en dirección opuesta al mismo; potenciadores (*enhancers*), que pueden ser transcritos de forma unidireccional o bidireccional (Figura 5E); y, por último, hospederos de miRNAs, que poseen dentro de su secuencia, un miRNA que puede ser procesado en etapas siguientes a su transcripción (Figura 5F) (Wu & Du, 2017).



**Figura 5:** Diagrama de los distintos tipos de lncRNAs según su localización en el genoma respecto de genes codificantes. Se muestran la secuencia y dirección de los lncRNAs en azul, y en amarillo para los genes codificantes. Podemos observar ocho tipos distintos de lncRNAs: divergentes y convergentes (A), intrónicos (B), intergénicos (lincRNA) (C), solapante sentido o antisentido (D), potenciadores (eRNAs), unidireccionales o bidireccionales (E) y finalmente, hospederos de miRNAs (miRNA host) (F). Fuente: tomada y modificada de Wu & Du, 2017.

### **1.2.1 Funciones de los lncRNAs:**

Los lncRNAs pueden presentar diversas funciones en la célula, y es en parte debido a ello que su estudio ha captado la atención e interés de muchos investigadores en los últimos tiempos, principalmente cuando se los empezó a asociar a varios procesos biológicos importantes y a distintas formas en que pueden desempeñar su función.

Algunos lncRNAs tienen funciones como reguladores epigenéticos, a través de la modificación del estado de la cromatina. Uno de los lncRNAs más estudiados en esta área es *Xist* (*X-inactive specific transcript*), que es el encargado del silenciamiento de un cromosoma X entero en hembras. Como consecuencia de poseer dos cromosomas iguales, las hembras de mamífero podrían tener un desbalance de expresión en comparación a los machos, ya que tendrían duplicada la cantidad de transcriptos y proteínas provenientes de estos dos cromosomas. Este problema es solucionado a través de la inactivación de uno de los dos cromosomas de forma total por el lncRNA *Xist*, el cual se transcribe desde el cromosoma a inactivarse y se localiza alrededor del cromosoma cubriéndolo entero, reclutando proteínas remodeladoras de la cromatina e impidiendo así la entrada de la maquinaria transcripcional (Pontier, 2011; Trovero & Geisinger, 2019). Otro lncRNA ampliamente estudiado es el denominado *HOTAIR* (*HOX antisense intergenic RNA*), el cual puede controlar el estado de la cromatina de forma *cis* o *trans*, regulando la expresión genética al interactuar con complejos modificadores de la cromatina, y generando silenciamiento génico (Tsai *et al.*, 2010; Gomes *et al.*, 2013).

Por otra parte, se han descrito varios lncRNAs actuando como reguladores transcripcionales y postranscripcionales. Podemos por ejemplo nombrar a *PANDA* (*p53-associated ncRNA DNA damage-activated*) el cual es inducido ante un daño al ADN, de forma dependiente de p53 (conocido inductor de la apoptosis). Este lncRNA se encarga de regular y limitar la expresión de genes pro-apoptóticos, a través de la interacción de factores de transcripción específicos. Además, algunos lncRNAs pueden actuar como co-reguladores, modulando la actividad de factores de transcripción; por ejemplo, el lncRNA *Evf-2* actúa como co-activador del factor de transcripción *homeobox* Dlx2, el cual es muy importante durante el desarrollo neuronal embrionario (Gomes *et al.*, 2013).

Como reguladores postranscripcionales, algunos lncRNA antisentido pueden unirse a mRNAs enmascarando los sitios de corte de *splicing* y así alterar el balance entre las distintas variantes de un mismo transcripto codificante. Otros pueden unirse a las regiones 3' UTR de mensajeros y reclutar proteínas vinculadas a su degradación (Gomes *et al.*,

2013). Por el contrario, el lncRNA *Uchl1* está encargado de regular positivamente la traducción de un mensajero específico, el de la proteína neuronal UCHL1 (*ubiquitin carboxy terminal hydrolase L1*) involucrada en la respuesta al estrés celular, mediante la facilitación de su unión a polirribosomas (Carrieri *et al.*, 2012; Gomes *et al.*, 2013).

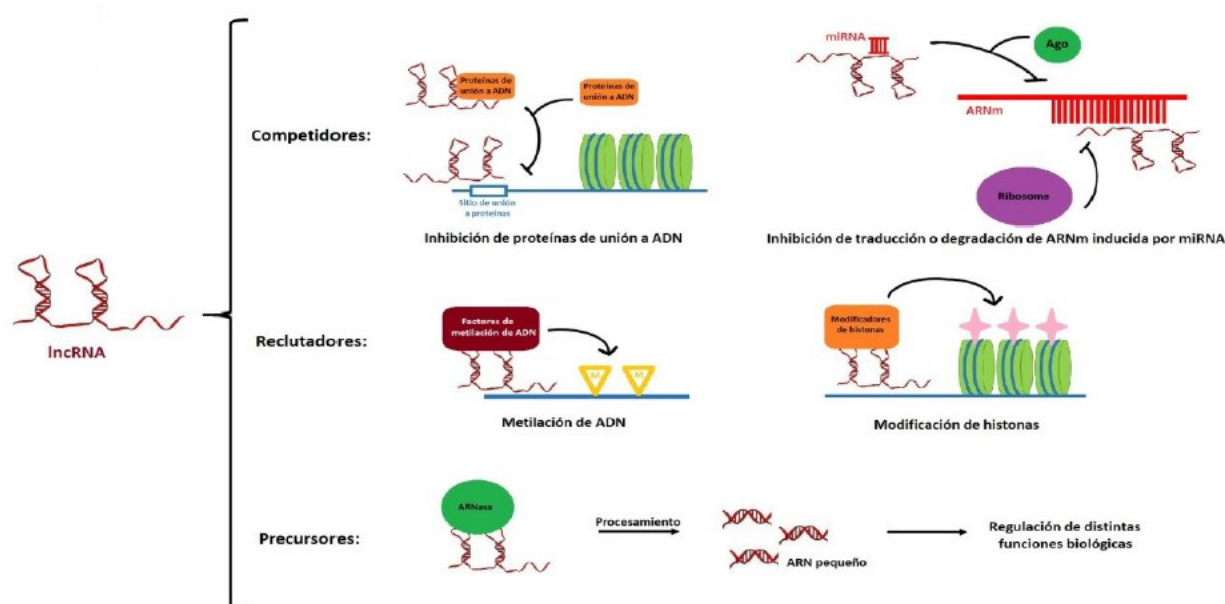
Algunos lncRNAs también pueden actuar como precursores de otros ncRNAs cortos que pueden cumplir su función en muchos otros procesos biológicos. Tal es el caso del lncRNA *H19*, el cual es de expresión únicamente materna, encontrándose imprimentado (silenciado) en el cromosoma paterno, y está involucrado en la regulación del peso corporal y la proliferación celular. Este lncRNA actúa como precursor de un miRNA llamado *miR-675*; si bien aún se conoce muy poco acerca del mismo, este podría estar relacionado con la regulación génica (Wei *et al.*, 2017; Trovero & Geisinger, 2019)

En definitiva, si bien aún se desconoce la forma en que actúan muchos lncRNAs, en términos generales estos pueden actuar de tres formas distintas (Figura 6):

- 1) Competidores: pueden unirse directamente a proteínas de unión al ADN (ej: factores de transcripción) e inhibir su función, o unirse de forma sitio-específica al ADN e impedir la unión de proteínas y otros factores, así como también pueden unirse a mRNAs de forma antisentido y complementaria e impedir la unión de otros factores como miRNAs, induciendo su degradación o cambios postranscripcionales.
- 2) Reclutadores o activadores: los lncRNAs pueden activar epigenéticamente determinados sitios del genoma al unirse a éstos, y reclutar factores modificadores de histonas, así como también factores que metilen ADN en sitios específicos.
- 3) Precursores: estos lncRNAs pueden ser procesados por una ARNasa para formar ARNs cortos que pueden cumplir diversas funciones, como el silenciamiento génico (Luk *et al.*, 2014).

Se ha demostrado que los lncRNAs tienen roles importantes en diversos procesos biológicos durante el desarrollo, como por ejemplo la proliferación, diferenciación y apoptosis celulares (Hong *et al.*, 2018; Kopp & Mendell, 2018). El gran interés alrededor de estos transcritos en las últimas décadas también se debe en parte a que muchos lncRNAs se han visto asociados a múltiples enfermedades humanas. En general, la desregulación en la expresión de muchos lncRNAs ha sido asociada con distintos tipos de cáncer; tal es el caso de algunos que han sido propuestos como biomarcadores para el diagnóstico de esta enfermedad. Por ejemplo, la pérdida de función del lncRNA *Xist*, provoca que uno de los cromosomas X no sea inactivado en hembras, lo que desencadena la activación de

genes inhibidores de la apoptosis codificados en este cromosoma. Dicha pérdida de *Xist* se ha visto presente en distintos cánceres de ovario, resistentes a los tratamientos que buscan inducir la apoptosis en células cancerígenas (Engreitz *et al.*, 2013). Otro ejemplo mencionado anteriormente es el lncRNA *HOTAIR*, el cual se ha visto desregulado en células tumorales y su expresión aberrante está asociada a células cancerígenas invasivas y metastáticas, durante cáncer de mama (Gupta *et al.*, 2010; Mozdarani *et al.*, 2020). De igual forma sucede para *H19*, donde la pérdida del *imprinting* genera la sobreexpresión de este lncRNA en carcinomas de ovario, induciendo la expresión de genes pro-metastáticos (Hosseini *et al.*, 2017; Liu *et al.*, 2018).



**Figura 6:** Diagrama de modos de acción generales de los lncRNAs. Los lncRNAs pueden actuar como competidores, como reclutadores, o también pueden actuar como precursores de miRNAs involucrados en diversos procesos biológicos. Fuente: Trovero & Geisinger, 2019.

### **1.2.2 LncRNAs durante la espermatogénesis:**

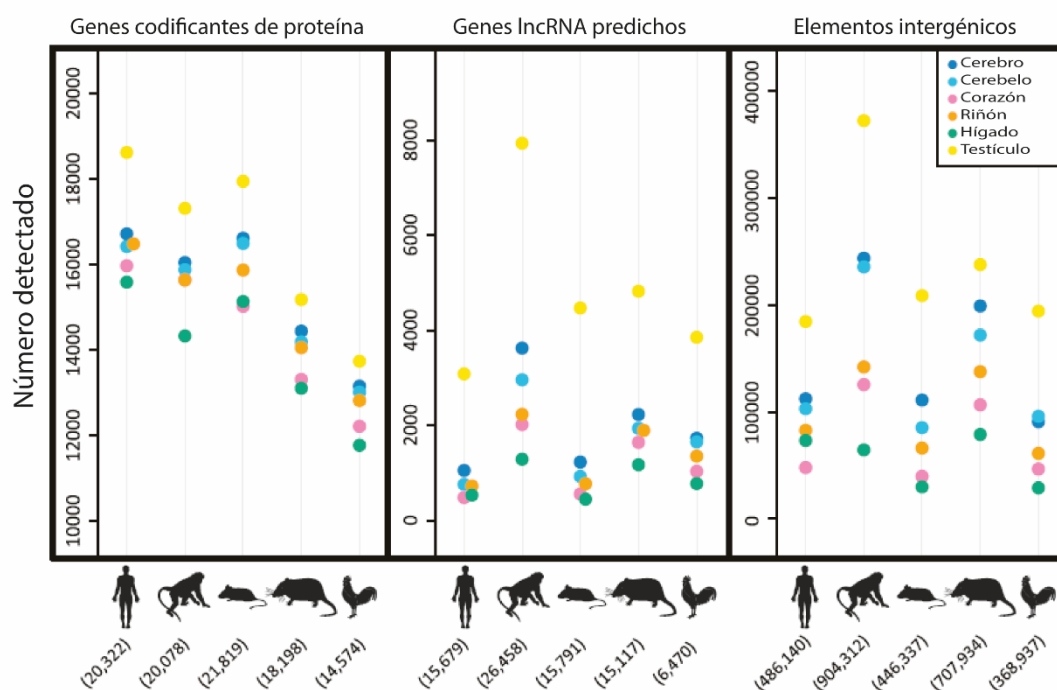
La espermatogénesis es un proceso altamente regulado, implicando importantes cambios morfológicos y genéticos en las células germinales. Particularmente, se ha demostrado tanto mediante microarreglos (Hong *et al.*, 2018) como mediante RNAseq (Soumillon *et al.*, 2013) que, en vertebrados, el testículo posee mayor complejidad de

transcriptoma respecto de otros tejidos como cerebro, corazón, riñón e hígado. Asimismo, el testículo presenta el máximo número de transcritos totales expresados, tanto de lncRNAs como de mRNAs, así como también de transcritos provenientes de regiones intergénicas como pseudogenes o elementos transponibles (Figura 7) (Soumillon *et al.*, 2013). Además, el testículo es el tejido que posee el mayor número de lncRNAs y mRNAs tejido-específicos, es decir, que se expresan únicamente en el testículo. Por lo tanto, el testículo es caracterizado por su alta complejidad y diversidad a nivel transcriptómico, donde sucede que los lncRNAs superan en diversidad a los mRNAs (Hong *et al.*, 2018).

Gracias al advenimiento de tecnologías como el RNA-seq y al refinamiento en la precisión en métodos de clasificación y separación celular, fue posible ver la expresión génica más en detalle de los distintos tipos celulares que conforman el testículo y descubrir nuevos transcritos cuya existencia era desconocida hasta entonces. En particular, se observó que las espermátidas redondas y los espermatoцитos son las células que contribuyen en mayor medida a la gran diversidad y complejidad del transcriptoma de testículo, en primer y segundo lugar respectivamente. Si bien las espermátidas representan el 45% de las células del testículo y los espermatoцитos sólo el 5%, son estos últimos quienes fabrican más cantidad de ARNs por célula. El tamaño del transcriptoma de un espermatoцитo, en cuanto a cantidad se refiere, puede superar hasta seis veces al de una espermátida. Sin embargo, al representar un porcentaje tan bajo del total de células del testículo, su importante aporte al transcriptoma de este tejido queda opacado por la abundancia de las espermátidas redondas (Soumillon *et al.*, 2013).

Como consecuencia de la alta complejidad que presentan los lncRNAs, sus patrones de expresión tan restringidos, y su gran abundancia en el testículo, se sugiere que por lo menos una parte de estos transcritos debería tener un rol importante durante el desarrollo testicular y en la espermatogénesis (Sun & Wu, 2015). En la actualidad se conocen varios lncRNAs que cumplen diferentes funciones durante el desarrollo de las células germinales en el testículo. Por ejemplo, el denominado *Mrhl* (*Meiotic recombination hot spot locus*) es un lncRNA testicular vinculado a la regulación negativa de la vía de señalización Wnt, la cual es crucial durante varias etapas de la espermatogénesis. La disrupción de este lncRNA genera morfologías aberrantes a nivel de tejido y la ausencia total de células posmeióticas (Akhade *et al.*, 2014; Kerr *et al.*, 2014). Otro ejemplo es el de *HongrES2*, un lncRNA hospedero de un miRNA denominado *mi-HongrES2*. Este miRNA está encargado de inhibir la expresión de una proteína específica del epidídimo de la rata (CES7), la cual estaría involucrada en la capacitación espermática (Ni *et al.*, 2011). Además, se han identificado

varios otros vinculados a distintos procesos y momentos de la espermatogénesis, por ejemplo, en la autorrenovación de las espermatogonias o en la proliferación celular, mientras que la sobreexpresión de algunos otros ha sido vinculada a distintas patologías testiculares causantes de infertilidad, como astenozoospermia (espermatozoides con baja movilidad) o azoospermia no obstructiva (falta de producción de espermatozoides), así como también han sido asociados a cáncer testicular (Bhartiya *et al.*, 2012; Luk *et al.*, 2014; Trovero & Geisinger, 2019). La infertilidad afecta aproximadamente a una de cada 6 ó 7 parejas, y la mitad de estos casos se debe a factores masculinos. A pesar de esto, muchas veces se desconocen cuáles son las causas y los mecanismos moleculares que la generan (Trovero & Geisinger, 2019). Por lo tanto, en este sentido es pertinente profundizar en el estudio de los lncRNAs con el fin de conocer cuáles son los mecanismos de acción de los mismos en el proceso espermatogénico.

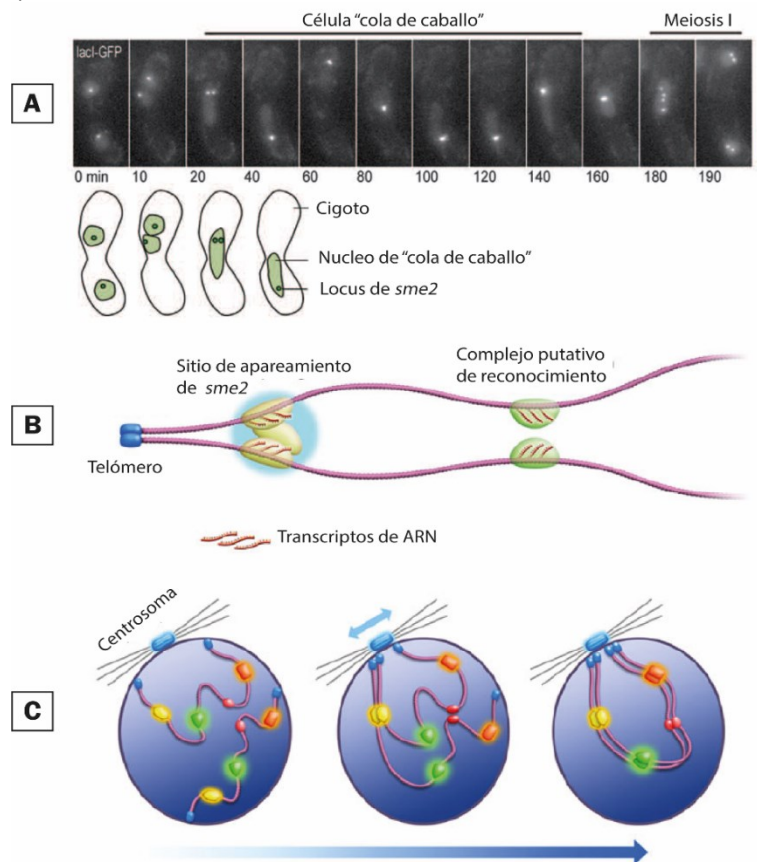


**Figura 7:** Complejidad del transcriptoma en los vertebrados. Se muestran el número detectado de genes codificantes, lncRNAs predichos y elementos intergénicos (pseudogenes y elementos transponibles) para seis órganos: cerebro (Azul), cerebelo (celeste), corazón (rosa), riñón (naranja), hígado (verde) y testículo (amarillo), para cinco organismos distintos: humano, mono, ratón, zarigüeya y gallo. Fuente: modificado de Soumillon *et al.*, 2013.

Dado que las espermatídas son el tipo celular que representa el mayor aporte de lncRNAs al testículo (Soumillon *et al.*, 2013; Trovero *et al.*, 2020), resulta interesante profundizar en las posibles funciones que éstos pudiesen tener en este tipo celular, por ejemplo, en la regulación de la expresión transcripcional y postranscripcional. Por otro lado, siendo los espermatoцитos el segundo tipo celular en importancia de aporte a la alta complejidad del transcriptoma testicular, cabe la posibilidad de que durante el proceso meiótico que estas células llevan a cabo existan lncRNAs vinculados al mismo, por ejemplo, durante el apareamiento de cromosomas homólogos o en la recombinación del material genético durante la profase meiótica, mediado por los CSs.

En ese sentido, existen estudios realizados sobre la levadura *Schizosaccharomyces pombe*, en los cuales se ha demostrado la participación de un lncRNA mediando el proceso de apareamiento entre cromosomas homólogos (Ding *et al.*, 2012 [A]). En esta levadura, la entrada a la meiosis está regulada por Mei2, una proteína de unión al ADN que es fundamental para el inicio de la meiosis. Durante la profase meiótica, esta proteína se acumula formando un punto alrededor del *locus sme2* en el cromosoma II de *S. pombe*. Este *locus* codifica para un ncRNA poliadenilado denominado *meiRNA* (Ding *et al.*, 2012 [B]). La acumulación de la proteína de entrada meiótica Mei2 en el *locus sme2* está mediada por la región 5' de *meiRNA*, a la cual esta proteína se une. Al mismo tiempo la proteína Mmi1, caracterizada por estar a cargo de una vía de degradación de transcritos pro-meióticos, y, por ende, un inhibidor de la meiosis, es secuestrada en estos puntos alrededor de *sme2* (Ding *et al.*, 2012 [A]). Esta estructura en forma de punto alrededor de *sme2*, formada por *meiRNA*, Mei2 y Mmi1, colaboraría en la entrada de la meiosis, al secuestrar a Mmi1 e impedir que esta proteína degrade otros transcritos pro-meióticos, pero también estaría involucrada en el apareamiento de los cromosomas homólogos. El *locus* de *sme2* muestra una frecuencia de apareamiento con su homólogo mucho más alta que otras regiones durante la profase meiótica, y se ha demostrado que este robusto apareamiento requiere de la transcripción de *meiRNA* (Ding *et al.*, 2013). La importancia de este *locus* y de su transcripción para que se lleve a cabo el apareamiento en esta región se ha demostrado experimentalmente al deletar el *locus sme2*, teniendo como consecuencia la ausencia de apareamiento entre cromosomas homólogos en ese punto. A su vez, al traslocar esta secuencia a otra localización en el cromosoma, se observó que ocurría el apareamiento de homólogos en esa región (Ding *et al.*, 2012 [B]). A raíz de estos estudios se ha propuesto que *meiRNA* mediaría de forma directa o indirecta el apareamiento sobre el *locus sme2*, y se sugiere que los cromosomas homólogos podrían reconocerse entre sí

sin la necesidad de una interacción directa entre sus secuencias de ADN, sino que podrían hacerlo por medio de ARNs, que actúen como sitios de reconocimiento de homología o reclutando proteínas de unión al ADN específicas para que se lleve a cabo el apareamiento (Figura 8) (Ding *et al.*, 2013). Más recientemente, el mismo grupo ha descrito otros dos lncRNAs que actuarían de modo semejante, en otros dos *loci* de *S. pombe* (Ding *et al.*, 2019). Algunos componentes proteicos de *S.pombe* participantes en estos procesos se encuentran conservados en eucariotas superiores (Ding *et al.*, 2019), por lo que se plantea que el mecanismo de reconocimiento y apareamiento de cromosomas homólogos en estos organismos podría darse de forma similar, involucrando la participación de lncRNAs (Trovero, 2020).



**Figura 8:** Apareamiento de cromosomas homólogos en *Schizosaccharomyces pombe* a nivel del locus *sme2*. En (A) se muestra una proyección tridimensional de un *stack* de células de levadura viva durante un tramo temporal de 190 minutos; barra de escala 5 μm. Se observa la señal de *sme2* confluyendo sobre un mismo punto durante el apareamiento de homólogos. También se muestra un esquema gráfico de esta célula. En (B, C) se muestra el diagrama del modelo planteado de reconocimiento y apareamiento mediado por ARNs y proteínas. Tomado y modificado de: (A) Ding *et al.*, 2012 [B]; (B, C) Ding *et al.*, 2013.

### **1.3 Antecedentes del grupo de trabajo:**

Este proyecto se encuentra enmarcado en una de las líneas de trabajo de un grupo multidisciplinario que desarrolla actividades en conjunto, contando con expertos en biología reproductiva, citometristas y bioinformáticos. En particular se ha trabajado con la tecnología "*fluorescence-activated cell sorting*" (FACS) para análisis de espermatogénesis de roedores por citometría de flujo. Gracias a esta tecnología, se ha logrado desarrollar un método para purificar células estadio-específicas de la espermatogénesis con un altísimo grado de pureza (Rodríguez-Casuriaga *et al.*, 2009; Rodríguez-Casuriaga *et al.*, 2011; Rodríguez-Casuriaga *et al.*, 2014). Mediante esta metodología se aislaron distintas poblaciones celulares del testículo de ratón: 2C (consta de células espermatogonias y células somáticas, conformando una población heterogénea de células con un contenido de ADN 2C), LZ (espermátocitos en leptoteno y cigoteno [*zygotene*]), PS (por *pachytene spermatocytes*; espermátocitos en paquíteno) y RS (por *round spermatids*; espermátidas redondas). A partir de estas cuatro poblaciones altamente puras, parte de una tesis de doctorado del grupo (Trovero, 2020) se enfocó en obtener el ARN total con el cual generar genotecas hebra-específicas (fundamental para la identificación de transcritos antisentido), las cuales fueron secuenciadas y mediante análisis bioinformáticos se estudió la expresión de lncRNAs. De esta forma, se generaron listas de lncRNAs expresados y diferencialmente expresados en las distintas etapas de la espermatogénesis, las cuales fueron utilizadas en este trabajo para seleccionar genes de lncRNAs candidatos a ser estudiados, mediante la síntesis de sondas de ARN específicas, en la búsqueda de determinar su localización sub-celular, y así ayudar a dilucidar sus posibles funciones durante la espermatogénesis.

## **2. Hipótesis y objetivos:**

### **2.1 Hipótesis:**

Al menos una parte de los lncRNAs de alta expresión en células espermatogénicas, cumplen un rol estructural y/o regulatorio durante la espermatogénesis.

### **2.2 Objetivo general:**

Aportar nuevos conocimientos sobre las bases moleculares de la reproducción masculina, en particular en lo relativo al potencial rol de secuencias no codificantes y a los posibles mecanismos subyacentes que éstas puedan cumplir durante la espermatogénesis.

### **2.3 Objetivos específicos:**

- 1) Seleccionar un grupo de secuencias no codificantes candidatas a desempeñar funciones estructurales y/o regulatorias, en base a las listas generadas de transcritos expresados en las poblaciones del testículo, y confirmar su patrón de expresión mediante RT-qPCR.
- 2) Generar sondas de ARN específicas contra las secuencias anteriormente seleccionadas, con el fin de estudiar su localización sub-celular mediante hibridación *in situ* y microscopía láser confocal, de tal forma de aportar información que pueda contribuir a elaborar hipótesis sobre sus posibles mecanismos de acción y procesos en los que estén involucradas.
- 3) Estudiar la posible co-localización y disposición de nuestras sondas respecto a anticuerpos contra proteínas de interés marcadoras de estructuras específicas, sobre cortes de tejido testicular congelado y esparcidos (*spreading*) de suspensiones celulares de testículo.

## **3. Materiales y Métodos:**

### **3.1 Selección de lncRNAs candidatos:**

Las listas de transcritos no codificantes generadas anteriormente por el equipo de investigación fueron utilizadas para seleccionar las secuencias estudiadas. Del total de transcritos se seleccionaron cuatro lncRNAs que cumplieran con la mayor cantidad de características deseadas. En particular, se buscó que los transcritos fueran específicos de testículo (en base a la información disponible en bases de datos) y que presentasen un patrón de expresión diferencial a lo largo de las distintas etapas celulares estudiadas, como por ejemplo una expresión baja seguida de una sobreexpresión en la etapa siguiente o viceversa. Se buscó también que tuvieran un nivel de expresión alto en la/s etapa/s de interés (PS y/o RS). Si bien el nivel de expresión de este tipo de transcritos suele ser bastante inferior al de los mRNAs, niveles muy bajos de expresión podrían imposibilitar su posterior visualización mediante microscopia confocal, como consecuencia de emitir una señal de fluorescencia que se encontrase por debajo del límite de detección del equipo.

También se buscó, con especial atención, un patrón de co-expresión o expresión antagonista de un lncRNA antisentido con su respectivo transcrito codificante contra el cual solapara.

Adicionalmente, el largo de los transcritos fue un aspecto importante a resaltar. En la medida de lo posible se buscaron lncRNAs que fueran mayores a 800 pares de bases,

un parámetro recomendado por la empresa fabricante para garantizar su buena señalización, sensibilidad y especificidad, ya que de lo contrario podría no llegarse a diseñar el número mínimo de sondas de ARN contra el transcripto y así dificultar su posterior visualización. En este mismo sentido, el número de sondas totales contra cada transcripto que fuese posible diseñar también fue un factor a tener en cuenta al momento de la selección de los mismos. En particular y por recomendación del fabricante se optó por aquellos transcriptos sobre los cuales se pudiese diseñar un mínimo de 25 sondas, e idealmente alcanzar el máximo de 48 sondas (ver más abajo).

### **3.2 Análisis de expresión mediante RT-qPCR:**

#### **3.2.1 Diseño de cebadores:**

Los cebadores fueron diseñados haciendo uso de la herramienta bioinformática *Primer-BLAST* del National Center for Biotechnology Information (NCBI), teniendo como referencia el transcriptoma del ratón (*Mus musculus*). Se diseñaron estos cebadores en las regiones de unión de los exones del transcripto de interés, de tal forma de evitar la posible amplificación de ADN genómico contaminante. Además, se buscó que los cebadores diseñados no solaparan con otras regiones del genoma. Otras consideraciones que se tuvieron en cuenta a la hora de su diseño fueron: la temperatura de *melting* de cada cebador (cercana a los 60°C y similares entre cada par), el tamaño de los amplicones (idealmente entre 150 y 250 pares de bases) y el contenido en GC (entre 50 y 60%). También se buscó evitar la formación estructuras secundarias como homodímeros, heterodímeros, u horquillas (*hairpins*).

El análisis de estas características para cada cebador se llevó a cabo mediante el uso de la herramienta informática de libre acceso *OlygoAnalyzer* de IDT (Integrated DNA Technologies Inc, USA).

Se seleccionó el diseño de los juegos de cebadores que presentaron las mejores características, y los mismos fueron sintetizados por MacroGen (Seúl, Corea del Sur). Para el caso de *Surf4* se utilizó un juego de cebadores diseñados y probados por el equipo de trabajo en investigaciones anteriores. El diseño de los cebadores utilizados puede verse en la Tabla 1.

**Tabla 1.** Lista de cebadores utilizados en los ensayos de RT-qPCR

| Nombre del Cebador           | Secuencia del cebador 5'-3' | Largo del amplicón (pb) |
|------------------------------|-----------------------------|-------------------------|
| <i>Rbakdn_Directo</i>        | CAACAGCCCATCCTTTTGCAG       | 242                     |
| <i>Rbakdn_Reverso</i>        | CTGGACCACATAGAAGCCACC       |                         |
| <i>1110020A21Rik_Directo</i> | AGAGAAGTGCCCCTAGACCC        | 159                     |
| <i>1110020A21Rik_Reverso</i> | GGTGTCCCAAGCAAAGGTGTG       |                         |
| <i>Kcnmb4os1_Directo</i>     | CAGACGACAAGTGCTCTCCG        | 215                     |
| <i>Kcnmb4os1_Reverso</i>     | GGTGTGGTGGTTCCAGACTC        |                         |
| <i>Kcnmb4_Directo</i>        | CAGATCGGCGAGGTGTTCG         | 174                     |
| <i>Kcnmb4_Reverso</i>        | ACAGGGCGGGATATAGGAGC        |                         |
| <i>Surf4_Directo</i>         | TGCTTTGGGCTGTTTGGAATC       | 162                     |
| <i>Surf4_Reverso</i>         | GGTTGGGACACCAGCAAACA        |                         |

### **3.2.2 RT-qPCR:**

Los patrones de expresión de los lncRNAs seleccionados fueron confirmados por RT-qPCR, utilizando el kit *Power SYBR Green Cell-to-Ct* (Ambion, Life Technologies), siguiendo las instrucciones del fabricante. Se estudió su expresión en cuatro poblaciones celulares distintas del testículo: 2C, LZ, PS y RS, según lo consignado más arriba (ver “Antecedentes del grupo de trabajo”). Estas poblaciones celulares ya habían sido purificadas por citometría de flujo a partir de ratones de 12-24 días posparto (dpp) y almacenadas a -80°C en estudios anteriores.

El ensayo de RT-qPCR constó de tres partes consecutivas, las cuales se detallan a continuación:

- Siguiendo las instrucciones del kit se realizó la lisis de las poblaciones celulares (3000 células por lisis) en un volumen de 50 µl (0,5 µl DNasa I y 49,5 µl *Lysis Solution*). Se incubó durante 5 minutos y se detuvo la reacción.
- Para cada reacción de transcripción reversa se utilizaron 22,5 µl de lisado celular en un volumen final de 50 µl (25 µl SYBR RT buffer; 2,5 µl RT Enzyme Mix y 22,5 µl de lisado). Se incubaron durante 1 hora a 37°C.
- La etapa de Real Time PCR se llevó a cabo utilizando el equipo de qPCR *CFX96 Touch Real-Time PCR Detection System 1* (BioRad, Hercules, CA, USA). Para cada

lncRNA estudiado se utilizaron 2 µl de ADN copia (ADNc), en un volumen de reacción final de 20 µl (10 µl *Power SYBR Green PCR Master Mix*; 0,8 µl *primer* directo 10 µM; 0,8 µl *primer* reverso 10 µM; 2 µl de ADNc y 6,4 µl agua libre de nucleasas). Se utilizó el mismo programa de amplificación para todos los lncRNAs estudiados (Tabla 2).

Tabla 2. Programa utilizado para los ensayos de qPCR.

|                                            | Etapa | Repeticiones                                                     | Temperatura | Tiempo      |
|--------------------------------------------|-------|------------------------------------------------------------------|-------------|-------------|
| Activación enzimática                      | 1     | 1                                                                | 95°C        | 10 minutos  |
| PCR (ciclado)                              | 2     | 40                                                               | 95°C        | 15 segundos |
|                                            |       |                                                                  | 60°C        | 1 minuto    |
| Curva de disociación<br>( <i>Melting</i> ) | 3     | <i>Setting</i> por defecto (65 segundos, 95°C – 0,5°C por ciclo) |             |             |

### **3.2.3 Eficiencia de cebadores:**

Para estudiar la eficiencia de los distintos pares de cebadores se realizaron qPCR de diluciones seriadas de ADNc (1:1, 1:2, 1:4, 1:8) obtenido tras la lisis y transcripción reversa de poblaciones celulares aisladas previamente mediante citometría de flujo. Posteriormente se graficó el Ct en función del logaritmo de la dilución y se calculó la pendiente, la cual fue utilizada para el cálculo de eficiencia según la fórmula:  $E = (10^{(-\frac{1}{m})}) - 1$ , donde E es la eficiencia del par de cebadores y m la pendiente. Se utilizaron aquellos pares de cebadores que tuviesen valores de eficiencia entre 90-100%.

### **3.2.4 Análisis de expresión de los transcriptos estudiados:**

Con el fin de confirmar los patrones de expresión de los lncRNAs seleccionados se realizaron ensayos de RT-qPCR para cada transcripto como se describe anteriormente, estudiando su expresión en las cuatro poblaciones celulares mencionadas (2C, LZ, PS y RS). Se realizaron tres replicas biológicas para cada transcripto y cada población celular, y se realizó una cuantificación relativa mediante el cálculo del  $2^{-\Delta\Delta Ct}$  de cada triplicado. Se utilizó como gen normalizador al mRNA del gen *Surf4*, comprobado previamente como un buen gen normalizador (Trovero, 2020), ya que su expresión no se ve significativamente

alterada en las distintas poblaciones celulares de testículo estudiadas, y su nivel de expresión es cercano al esperado para nuestros transcritos. Se promedió el  $2^{-\Delta\Delta Ct}$  de los triplicados y también se calculó el desvío estándar en cada caso. Estos datos fueron visualizados mediante el *software CFX Maestro* propio del equipo utilizado, y se analizaron con el procesador de bases de datos Excel.

### **3.3 Localización sub-celular:**

#### **3.3.1 Diseño de sondas de ARN Stellaris®:**

Se diseñaron sondas de ARN con tecnología Stellaris® (Biosearch Technologies) contra los lncRNAs seleccionados, para hibridación *in situ* de ARN fluorescente (RNA-FISH). Esta tecnología consta del diseño de hasta 48 sondas de ARN, de entre 18 y 22 bases, conjugadas a fluoróforos, que en su conjunto hibridan de forma específica contra el transcripto objetivo para el cual fueron diseñadas. Esta tecnología presenta algunas ventajas, por ejemplo, que el alto número de sondas para cada transcripto estudiado asegura una muy buena sensibilidad, especificidad, y permite discriminar falsos negativos y falsos positivos. Las señales positivas agrupan la fluorescencia concentrada de un gran número de sondas (al menos 25), mientras que las señales “*off-target*” serán producto de la unión de sondas individuales que emiten una fluorescencia débil y difusa. Además, el pequeño tamaño de estas sondas permite que las mismas penetren fácilmente dentro del tejido e hibriden con su blanco.

Se diseñaron sondas contra los transcritos *1110020A21Rik*, *Rbakdn*, *Kcnmb4os1* conjugadas al fluoróforo Quasar670 y sondas contra el mRNA *Kcnmb4* conjugada al fluoróforo Quasar570 (los fluoróforos Quasar son los utilizados por la empresa que fabrica las sondas). Como control positivo se obtuvieron sondas prediseñadas y validadas por la empresa contra el transcripto *Malat1*, también conjugadas al fluoróforo Quasar570. La selección de *Malat1* como control positivo para los ensayos de RNA-FISH de este trabajo, se debe fundamentalmente a cuatro ventajas: en primer lugar, a su altísima expresión en diversos tejidos, así como también a lo largo de las distintas etapas celulares de la espermatogénesis y especialmente en las células somáticas del testículo (según las listas generadas por nuestro grupo; Trovero, 2020); en segundo lugar, debido a su tamaño (permite hibridar un alto número de sondas cortas); y en tercer lugar, debido a la vasta cantidad de bibliografía sobre este transcripto.

El diseño de estas sondas se realizó con la ayuda del diseñador *Stellaris Probe Designer* de Biosearch Technologies (<https://www.biosearchtech.com/stellaris-designer>). Para esto se tuvo en cuenta las recomendaciones del fabricante, y una serie de parámetros como la longitud de las sondas, el número de sondas a diseñar para cada transcripto, la distancia entre ellas, así como la ausencia de complementariedad de cada una de ellas en otras regiones del genoma, entre otros (la secuencia de las sondas diseñadas se muestra en la Tabla 3). Una vez diseñadas, su síntesis fue encargada a la misma empresa especializada, Biosearch Technologies (UK).

Tras su llegada a nuestro laboratorio y siguiendo las recomendaciones del fabricante, las sondas liofilizadas fueron resuspendidas en buffer TE (Tris-HCL 10 mM, EDTA 1 mM, pH 8,0), obteniendo una concentración final de 12,5  $\mu$ M. Este *stock* fue protegido de la luz y almacenado a -20 °C hasta su uso.

**Tabla 3:** secuencias de las sondas Stellaris® diseñadas para cada transcripto, su número y el fluoróforo al cual fueron conjugadas.

| Transcripto          | Secuencia de sondas                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Nº Sondas | Fluoróforo |
|----------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|------------|
| <i>Rbakdn</i>        | CAGGTCAGTTGTGGCTCT; TCCCAAGAAGTGTGAGG;<br>CTGGTCAACGGTCTCCAG; CTCTGTGTGAGCCCTCAG;<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; CTTAGCAGCTGAAGGGC;<br>GGATAATGGTGTCCACCC; GAGTCCACATATGTCCC;<br>AATGGTGCTGCAGGTGGG.                                                       | 25        | Quasar670  |
| <i>1110020A21Rik</i> | ACTCTACCTATGGCTACTAG; GCGTAATTTGACTGACGGA;<br>AAGGATATCGGTCTTCGAGT; CGCCATTTGAACCAATCAGA;<br>AAGCCAATAGTCGGAGGATC; TGGTTTCTTGACGGACAAGT;<br>CCGAAGTGACGGAAGTTGAA; CCTTTTAGGATACTTTTGCG;<br>CTTGGCCATCTACGAATACT; AGAGGTCAGAACTGCAGGGT;<br>TGTTCTGTTCTCCGCAAGTG; TGCCCGTTGGTTTCTTTATG;<br>GGACCAATTTCCACCTATAA; CTTCTCTGGCCAATCAGAAG;<br>TTCTTCCAATGAGGGTCTAG; CACAGCTCTATGCCAAAGAG;<br>TAAGGTGTCCCAAGCAAAGG; CCCGTAATTTCAATCTTCTAA;<br>CCTTATTTTGGAAGGTGAGG; CCCTTATTTGTTGATCTTTA;<br>TCGAGGCAAGGGAATCCAAG; ATCAAGCAGTGCGAGACAGG;<br>CATATTGCATTTGAGTGTCT; GCCTAGTTACATCACTGAAT;<br>GTGAGTTGTATACAATCCGC. | 25        | Quasar670  |

|                  |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |    |           |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|-----------|
| <i>Kcnmb4os1</i> | CGCAGGAGATGGAAGGGG; CAGAGCCCACATTTGGTG;<br>AAGCCGAGACCAGAGGTT; GTCAAGGCGGAAGCCATG;<br>ATCTGTGCCAAGAGCCTG; TTCTCATCGTGGTCCTGA;<br>GTGGTGGCGTTTGTGGTG; GATTGCTCTCCTGCACTG;<br>TACAGCGCACACACGACG; AGACCAGAGGACGTTCTC;<br>CTGTCTCCACTCTCTTG; ACTCACTGGTCTTCCTTT;<br>GAATGTGCTGGGCTGAGA; CTTGTCGTCTGCAGGTTG;<br>ATTATTCCCTACGCCGGA; GTCATCACTTTCAAGGGT;<br>GGACGGTGTCTGCAATCA; ATGGGTGTGGAGCAGTCT;<br>AGAGCTCACGGCTTTTGG; ATGGCATTCAGTGCCATC;<br>GCTGGGTGTTCTCTTATT; CCTCTTAGGAGCTCTTTC;<br>GTGTGGTGGTTCCAGACT; TCGCTTCCTTCTCTCCTG;<br>TGGATCCAACAGTTTCCG; AAGTGTGCCCCAAGATCC.                                                                                                                                                                                                                                                                                       | 26 | Quasar670 |
| <i>Kcnmb4</i>    | CGTAAGACACCCTGAGCT; TCTTCGGCTTCCGTGTAC;<br>AACAAGCCGAGCCGGATG; GGATGCCGGAGACGATGA;<br>GCCGAAGATGAAGAGCGA; AAGGCGGGACTGAGCCAG;<br>CGTGGCTTGCAATCCTG; GACAGCACGGTGCACTT;<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>GTAGGACACGATGAGA; CAGGCTCTTGGCACAGAT;<br>GCGTTTCTTCATGGCTTC; AACCTCTGGTCTCGGCTT;<br>CAGGCGAGTGCCGACGAG; ACCAAATGTGGGCTCTGC;<br>GTTTCTGCTGCCGACGAG; CAGTCTTCCAAGTGTGCC;<br>TACAGGCAGTTAAGTCCC; GTTGCTACAGCACATGGA;<br>AGAGCTTGACCCTTTAGC; TAGTCACTCCATAGAGCT;<br>ACCGACTTCTTTGAGGGT. | 39 | Quasar570 |

### **3.3.2 Animales y eutanasia:**

Se utilizaron ratones macho CD-1 (*Mus musculus*) de 25 dpp obtenidos de las instalaciones del Instituto de Higiene de la Facultad de Medicina (Udelar, Montevideo, Uruguay). A esta edad los individuos se encuentran cursando la primera onda espermatogénica, disponen de células meióticas en sus diferentes estadios, y espermátidas redondas y en elongación. Sin embargo, aún no presentan espermatozoides, que en los animales adultos constituyen una muy alta proporción dentro de los túbulos seminíferos. La ventaja de usar individuos de esa edad es, por lo tanto, que las células meióticas y en estadios iniciales de espermiogénesis se encuentran proporcionalmente más representadas que en animales adultos. 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). El procesamiento y eutanasia de los animales se realizó siguiendo las recomendaciones de la Comisión Nacional de Experimentación Animal (CNEA), mediante dislocación cervical. Inmediatamente después, los testículos fueron diseccionados y preparados para su estudio.

### **3.3.3 Preparación de tejido congelado y criosecciones de tejido testicular:**

La preparación de tejido congelado se llevó a cabo inmediatamente después de la eutanasia de los ratones, y se prosiguió de la siguiente forma:

- a) Los testículos diseccionados fueron colocados en PBS 1X y cortados al medio cuidadosamente, sin quitarles la túnica albugínea.
- b) Luego se los fijó en paraformaldehído 4% diluido en buffer PHEM 1X (PIPES 60mM, HEPES 25 mM, EGTA 10 mM, MgCl<sub>2</sub> 2 mM) durante una hora.
- c) Las mitades de testículo fueron lavadas durante 10 minutos en Buffer PHEM 1X, a temperatura ambiente.
- d) Posteriormente se incubaron en sacarosa al 15% en PHEM 1X, a temperatura ambiente hasta que el tejido se hubiera hundido en la solución.
- e) Se realizó una segunda incubación en sacarosa al 30% en PHEM 1X a 4°C durante toda la noche.
- f) Una vez terminada esta incubación, los trozos de testículo fueron colocados en un soporte adecuado, con la cara cortada hacia abajo, y se los cubrió por completo con medio de congelación, para luego proceder a un congelamiento rápido en nitrógeno líquido. Estos bloques fueron almacenados a -80°C hasta su momento de uso.

Los bloques que contenían las mitades de testículo diseccionado fueron cortados mediante el uso de un criostato SLEE (Mainz, Alemania) modelo MEV, realizando cortes de 8-10 µm de espesor y colocados sobre portaobjetos que fueron tratados previamente con poli-L-lisina (Sigma-Aldrich) en dilución 1/10 según los procedimientos habituales. Estos portaobjetos conteniendo los cortes fueron almacenados a -20°C hasta su uso.

### **3.3.4 Preparación de esparcidos (*spreading*) de suspensión celular:**

Previo al inicio de la técnica, los portaobjetos de vidrio utilizados fueron sumergidos durante 15 minutos en etanol 95% y hervidos durante 10 minutos en agua ultrapura. Una vez terminado este paso, se continuó de la siguiente forma:

- a) Los testículos diseccionados fueron colocados en medio de cultivo celular DMEM y se les retiró la túnica albugínea. Los mismos fueron triturados cuidadosamente con un bisturí estéril y se disgregó el tejido mediante pipeteo.
- b) Luego, se filtró la suspensión eliminando así los trozos de tejido remanentes de mayor tamaño. Se buscó obtener una suspensión celular de  $3 \times 10^7$  células/ml de

medio, ya que mayores concentraciones celulares pueden inhibir la técnica de *spreading*.

- c) Posteriormente se mezclaron cantidades iguales (1:1) de suspensión celular y buffer hipotónico (ver más abajo), y se incubó durante 30 minutos a temperatura ambiente.
- d) Pasado el tiempo de incubación, se agregó un volumen de sacarosa 100 mM y se incubó nuevamente durante 1 hora.
- e) Sobre un portaobjetos remojado en PFA 1% (diluido en agua tratada con DMPC y Tritón 0,15%) se agregaron 100 µl de la solución de suspensión celular, y se los dejó secar a 4°C durante toda una noche.
- f) Por último, los portaobjetos fueron dejados a temperatura ambiente hasta que secaron por completo. Una vez secos, fueron guardados a -80°C hasta su uso.

El buffer hipotónico utilizado para este ensayo fue preparado con 10 ml de agua tratada con DMPC, citrato de sodio 20 mM, ácido etilendiaminotetraacético (EDTA) 6 mM y sacarosa 50 mM. Al momento de su uso, se le agregó al tubo 0,3 µl de Tris-HCl 1M (pH 8) y 0,5 µl de Ditiotreitól (DTT) 100 mM, por cada ml de buffer.

### **3.3.5 Preparación de buffers para RNA-FISH e inmunodetección:**

Para la realización de este ensayo se requiere de la preparación de un buffer de fijación, un buffer de hibridación y buffers de lavado A y B.

El buffer de fijación se preparó con un 3.7% de formaldehído junto a 10% de PBS 10X, en agua ultra pura libre de ARNasas.

El buffer de hibridación es comercial, y lo provee el mismo fabricante (Biosearch Technologies Cat# SMF-HB1-10), que debe completarse con un 10% de formamida desionizada.

Los buffers de lavado A y B también son buffers comerciales. En el caso del buffer A, se procedió teniendo en cuenta las indicaciones sugeridas por la empresa, y mezclando 20% *Stellaris RNA FISH Wash Buffer A* (Biosearch Technologies Cat# SMF-WA1-60), y 10% de formamida desionizada, en agua ultrapura libre de ARNasas. Este buffer se preparó fresco en el momento de uso. En el caso del buffer de lavado B, se agregaron por única vez 88 ml de agua ultrapura al recipiente de *Wash Buffer B* (Biosearch Technologies Cat# SMF-WB1-20), quedando listo para su uso en todos los ensayos siguientes, y almacenándose a 4°C hasta su uso.

### **3.3.6 Anticuerpos:**

Los anticuerpos primarios empleados en este trabajo fueron policlonales anti-MVH de conejo (ab13840, Abcam) en una dilución de 1:2000 y anti-SYCP3 de conejo (cedido por el Dr. Ricardo Benavente, Universidad de Würzburg, Alemania), en una dilución 1:200.

Como anticuerpo secundario, se usó un anti-conejo conjugado a Alexa Flúor 488 (A-11034, Invitrogen), en una dilución 1:1000.

### **3.3.7 Protocolo RNA-FISH para tejido congelado:**

La técnica de hibridación *in situ* se llevó a cabo siguiendo los pasos indicados por los fabricantes de las sondas Stellaris.

- a) Se dejó a los portaobjetos con las secciones de tejido congelado llegar a temperatura ambiente.
- b) Los mismos fueron cubiertos con 1 ml de buffer de fijación durante 10 minutos a temperatura ambiente.
- c) Pasado el tiempo de incubación, se realizaron dos lavados de 5 minutos en PBS 1X.
- d) Luego, los portaobjetos fueron sumergidos para su permeabilización en etanol al 70% durante por lo menos 1 hora.
- e) Terminada la permeabilización, los portaobjetos fueron cubiertos con 1 ml del buffer de lavado A durante 5 minutos.
- f) El excedente de buffer de lavado A fue volcado, y los portaobjetos fueron colocados dentro de una cámara húmeda cerrada.
- g) Se cubrió cada portaobjetos con 200 µl de buffer de hibridación conteniendo la sonda de ARN Stellaris correspondiente en la concentración adecuada. Sobre este portaobjetos se colocó un trozo de parafilm™ de tal forma de distribuir completamente el buffer por arriba del tejido, y evitar la evaporación.
- h) La cámara húmeda se incubó en oscuridad a 37°C durante 16 horas como máximo.
- i) Pasado este tiempo de incubación, se retiró cuidadosamente el parafilm™ que cubría los portaobjetos y se cubrió con 1 ml de buffer de lavado A, incubándolos en cámara húmeda y oscura durante 30 minutos a 37°C.
- j) Paso seguido, se volcó el excedente de buffer y se cubrió con 500 µl de buffer de lavado A conteniendo 0,5 µl de DAPI 1 mg/ml, de tal forma de contrastar los núcleos

celulares. Nuevamente se incubaron los portaobjetos en cámara húmeda y oscura, a 37°C durante 30 minutos.

- k) El excedente de buffer de tinción con DAPI fue volcado y se realizó un lavado de 5 minutos con 1ml de buffer de lavado B.
- l) Una vez realizado el lavado, se removió lo más posible del Buffer B alrededor del tejido, y se cubrió con una pequeña gota de líquido de montaje *ProLong™ Diamond Antifade* (Thermo Fisher Scientific) para proteger las sondas de la luz.
- m) Se colocó un cubreobjetos limpio y se selló el perímetro del preparado con esmalte para uñas genérico.
- n) Una vez secos, los preparados fueron guardados a -20°C hasta su visualización en un microscopio confocal.

### **3.3.8 Protocolo RNA-FISH e inmunodetección para tejido congelado:**

- a) Se siguieron de manera idéntica los pasos a) a d) del protocolo de RNA-FISH para tejido congelado detallado más arriba.
- b) A continuación del etanol 70%, se cubrió a los portaobjetos con 1 ml de 0,1% de Triton X-100 en PBS 1X durante 5 minutos a temperatura ambiente, y luego se realizó un lavado de 5 minutos con 1 ml de PBS 1X.
- c) Se cubrió a los portaobjetos con 500 µl de PBS 1X conteniendo la dilución correspondiente de anticuerpo primario, y se lo dejó incubar durante 1 hora a temperatura ambiente.
- d) Pasado el tiempo de incubación, se lavó los portaobjetos con 1 ml de PBS 1X tres veces durante 10 minutos.
- e) Luego se procedió a la incubación con el anticuerpo secundario en su dilución correspondiente en PBS 1X durante 1 hora a temperatura ambiente.
- f) Nuevamente se repitieron los tres lavados de 1 ml de PBS 1X durante 10 minutos cada uno.
- g) A continuación, se cubrió a los portaobjetos con 1 ml de buffer de fijación y se los incubó durante 10 minutos a temperatura ambiente.
- h) Luego se realizaron dos lavados de 5 minutos con PBS 1X.
- i) Los portaobjetos fueron luego cubiertos con 1 ml del buffer de lavado A durante 5 minutos.
- j) Se eliminó el excedente de buffer A, los portaobjetos fueron colocados dentro de cámara húmeda, y se prosiguió con el protocolo de hibridación de sondas Stellaris

de manera idéntica que en el protocolo RNA-FISH para tejido congelado (pasos g) a n) del protocolo anteriormente descrito).

### **3.3.9 Protocolo RNA-FISH e inmunodetección para *spreading* de suspensión celular:**

Para el caso del ensayo de RNA-FISH sobre esparcidos de suspensiones celulares se prosiguió siguiendo de manera muy similar al protocolo mostrado anteriormente (Protocolo RNA-FISH e inmunodetección para tejido congelado), pero evitando los pasos que son específicos para tejido. El protocolo utilizado fue el siguiente:

- a) Se dejó a los portaobjetos con los esparcidos de suspensión celular llegar a temperatura ambiente.
- b) Se realizó un lavado con 1 ml de PBS 1X durante 5 minutos.
- c) Los portaobjetos fueron cubiertos con 1 ml de buffer de fijación durante 10 minutos a temperatura ambiente.
- d) Pasado el tiempo de incubación se realizaron dos lavados de 5 minutos en PBS 1X.
- e) Se prosiguió con los pasos c) al j).

### **3.3.10 Visualización de los preparados:**

Todos los preparados fueron visualizados en un microscopio confocal ZEISS LSM 800 (Carl Zeiss Microscopy, Alemania), con o sin módulo de súper-resolución *Ayriscan*, dependiendo del caso. Las imágenes obtenidas fueron analizadas con el programa *Fiji ImageJ* (Schindelin *et al.*, 2012). Para evaluar la co-localización entre los ARNs *Kcnmb4os1* y *Kcnmb4*, se utilizó el *plugin* “*Colocalization Finder*” del *ImageJ*.

## **4. Resultados:**

### **4.1 lncRNAs seleccionados:**

En primera instancia se seleccionaron algunos lncRNAs a partir de las listas de transcritos, las cuales fueron generadas en estudios anteriores por el equipo de trabajo. La selección de estos lncRNAs se realizó siguiendo los parámetros indicados en Materiales y Métodos. De esta forma, los lncRNAs que más se ajustaron a lo buscado fueron: *Rbakdn*, *1110020A21Rik* y *Kcnmb4os1*. Estos tres lncRNAs son altamente específicos del testículo, tejido donde se expresan en altos niveles en comparación a otros (Figura 9), y de las células espermatogénicas (Tabla 4). Los tres transcritos aumentaron significativamente durante la profase meiótica, incrementando su nivel de expresión en LZ con respecto a la población 2C. Su nivel de expresión aumentó significativamente durante la profase meiótica (PS; un incremento de más de 10 veces en todos los casos), y continuó incrementándose durante la espermiogénesis (RS).

En el caso de *Rbakdn*, se trata de un transcrito no codificante del tipo lincRNA que, si bien posee un largo inferior a lo recomendado, tuvo tres puntos a favor para ser seleccionado. El primero fue que a pesar de su corta secuencia se pudo diseñar el número mínimo necesario de sondas de ARN; el segundo punto es que es altamente específico del testículo y conservado entre ratón y humano, y, por último, su nivel de expresión, lo cual fue el aspecto definitivo para su selección: *Rbakdn* mostró ser el transcrito que presentó el mayor nivel de expresión en la etapa RS del total de la lista, superando por varias cifras al segundo transcrito más expresado (también fue uno de los transcritos con mayor nivel de expresión en PS). El lncRNA *1110020A21Rik* es un transcrito no codificante también del tipo lincRNA, que cumple con las características deseadas de largo, especificidad de testículo y número de sondas, cuya mayor expresión también se da en la etapa RS. En el caso de *Kcnmb4os1*, el mismo es un lncRNA antisentido, con muy buen nivel de expresión en PS, y expresándose mayoritariamente en la etapa RS. Al tratarse de un lncRNA antisentido, también se diseñaron sondas de ARN contra el mRNA sobre el cual solapa, llamado *Kcnmb4*. Este mRNA, específico del testículo (Figura 9), comienza a expresarse mayoritariamente en PS y se mantiene durante RS (aunque en este caso particular, su nivel de expresión es menor que el del lncRNA), por lo que postulamos que podría existir alguna interacción con su lncRNA antisentido en estas etapas. Como control positivo para el ensayo de RNA-FISH, se obtuvieron sondas prediseñadas por la empresa contra el transcrito *Malat1*. Este transcrito, del tipo lincRNA, tiene un largo bastante mayor a los

transcriptos previamente seleccionados, por lo que pueden diseñarse 48 sondas sin inconvenientes. Si bien *Malat1* se expresa a elevados niveles en varios tejidos, en el testículo se expresa mayoritariamente en células intersticiales (pertenecientes a la población 2C), y se mantiene en niveles relativamente altos en el resto de las etapas celulares de la espermatogénesis.

Las características principales de estos transcriptos se muestran en la tabla 4.

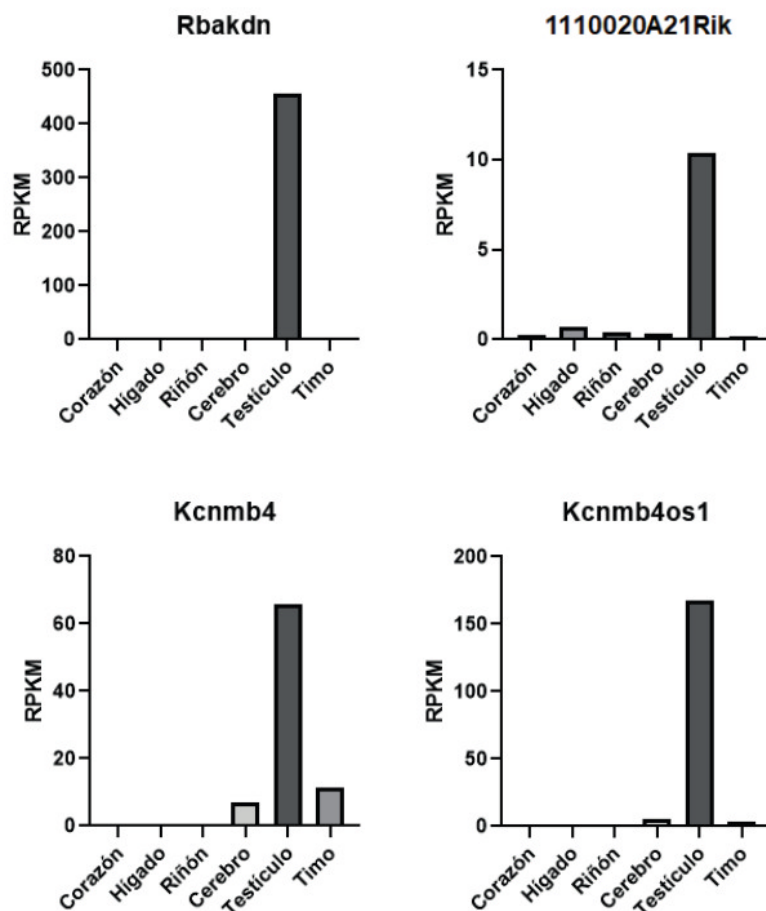
**Tabla 4:** Características principales de los transcriptos seleccionados para su estudio, y el número de sondas de tecnología Stellaris® que se fabricaron. FPKM refiere a los niveles de expresión en cada etapa.

| Transcripto          | Localización cromosómica | Biotipo     | Largo (pb) | 2C (FPKM) | LZ (FPKM) | PS (FPKM) | RS (FPKM) | N° de sondas |
|----------------------|--------------------------|-------------|------------|-----------|-----------|-----------|-----------|--------------|
| <i>Rbakdn</i>        | 5 G2; 5                  | lincRNA     | 588        | 2,5       | 32,7      | 332,1     | 501,0     | 25           |
| <i>1110020A21Rik</i> | 17; 17 E4                | lincRNA     | 853        | 1,2       | 4,6       | 52,6      | 102,4     | 25           |
| <i>Kcnmb4os1</i>     | 10; 10 D2                | antisentido | 824        | 0,9       | 12,7      | 142,2     | 298,4     | 26           |
| <i>Kcnmb4</i>        | 10; 10 D2                | mRNA        | 1367       | 0,4       | 2,5       | 28,0      | 25,1      | 39           |
| <i>Malat1</i>        | 19; 19 A                 | lincRNA     | 6938       | 2439,2    | 252,0     | 126,0     | 200,1     | 48           |

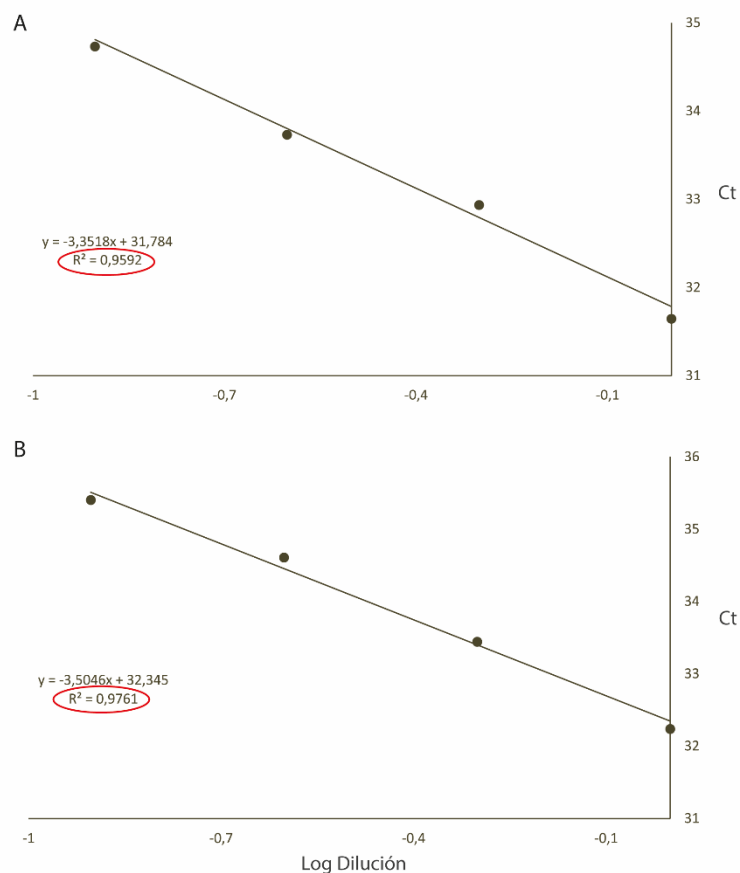
## **4.2 Diseño de cebadores y eficiencia de los mismos:**

Se diseñaron distintos juegos de cebadores contra los transcriptos seleccionados, teniendo como finalidad confirmar mediante RT-qPCR los patrones de expresión de los mismos. Por lo tanto, en su diseño se priorizó conseguir las características óptimas en cuanto a complementariedad, largo, TM, contenido GC, entre otras (el diseño de todos los cebadores utilizados puede verse en la Tabla 3, Materiales y Métodos).

Antes de llevar a cabo el ensayo de RT-qPCR definitivo, se evaluó la eficiencia de los pares de cebadores a utilizar. Como resultado se obtuvo que los cebadores diseñados presentaron una buena eficiencia, con valores entre 90-100% y se comprobó la ausencia de amplificaciones inespecíficas a través de las curvas de *melting*. A modo representativo se muestran las curvas de eficiencia de los cebadores diseñados para los transcriptos *Kcnmb4os1* y *Kcnmb4* (Figura 10). En el Anexo, la figura suplementaria 1 muestra las curvas de amplificación y de *melting* para estos mismos transcriptos.



**Figura 9:** Expresión de los cuatro transcritos seleccionados en distintos tejidos. Tanto los tres lncRNAs (*Rbakdn*, *1110020A21Rik* y *Kcnmbos1*) como el mRNA *Kcnmb4*, son altamente específicos del testículo, tejido donde presentan mayor número de *reads* por Kilobase de transcripto por millón de secuencias mapeadas (RPKM), en comparación a otros tejidos como corazón, hígado, riñón, cerebro y timo. Estos datos fueron extraídos a partir de las bases de datos de NCBI (cuyos “gene id” son: *Rbakdn*: ENSMUSG00000104529, *1110020A21Rik*: ENSMUSG00000097047, *Kcnmb4os1*: ENSMUSG00000020164, *Kcnmb4*: ENSMUSG00000054934).

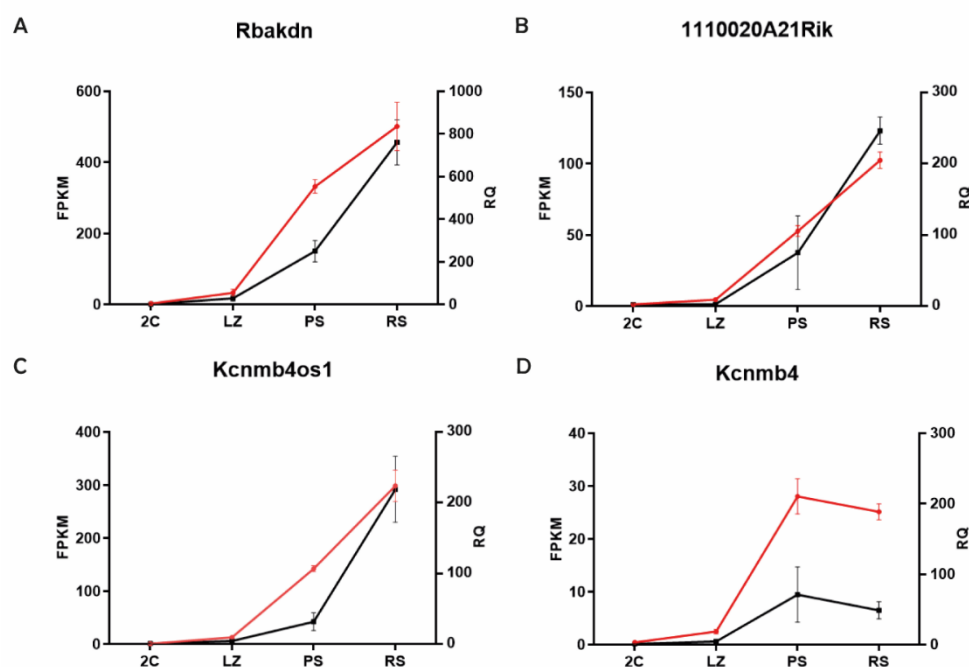


**Figura 10:** Curvas de eficiencia para los pares de cebadores utilizados en la amplificación de los transcriptos *Kcnmb4os1* (A) y *Kcnmb4* (B). En el eje X se encuentra graficado el logaritmo de la dilución de ADNc utilizada en este ensayo de RT-qPCR, mientras que en el eje Y se muestra el promedio de Ct al cual levantó cada triplicado de estas diluciones. Se puede apreciar la ecuación perteneciente a cada curva, donde se muestra el valor  $R^2$  (rojo), el cual multiplicado por 100, corresponde a la eficiencia de los cebadores estudiados. En este caso la eficiencia para los cebadores de *Kcnmb4os1* fue de 95,9% y para los cebadores de *Kcnmb4* fue de 97,6%.

#### **4.3 Validación de patrones de expresión mediante RT-qPCR:**

Previo al diseño de las sondas Stellaris, se buscó validar mediante RT-qPCR los patrones que se observaron en estudios anteriores obtenidos a través de la metodología de RNAseq. De esta forma, no solamente verificaríamos que los mismos se expresan en las distintas poblaciones celulares estudiadas (2C, LZ, PS y RS) y que mantuviesen el patrón observado previamente, sino que también podríamos verificar la correlación de expresión entre el lncRNA antisentido *Kcnmb4os1* y su mRNA *Kcnmb4*.

Como se mencionó anteriormente, para estos ensayos de RT-qPCR de expresión relativa se utilizó como gen normalizador a *Surf4* (ver Materiales y Métodos), ya que el mismo no presentaba una gran variación a lo largo de las distintas poblaciones celulares estudiadas, su nivel de expresión no era muy distinto al de nuestros transcriptos, y había sido utilizado exitosamente en instancias anteriores por el equipo de trabajo con la misma finalidad. De esta forma, se obtuvo que los patrones de expresión de los transcriptos estudiados eran consistentes con aquellos observados en estudios anteriores mediante RNAseq (Figura 11). Como era esperado, los cuatro transcriptos analizados en este ensayo presentaron una mayor expresión a partir de la etapa PS aumentado aún más, o manteniéndose, hacia la etapa RS. Para el caso del lncRNA antisentido y su mRNA, se confirmó su patrón de co-expresión aumentando en la transición de LZ a PS, continuando el aumento hacia RS en el caso del lncRNA, y manteniéndose en el caso del mRNA (en este caso se observó una leve disminución, aunque la misma no fue significativa).



**Figura 11:** Validación de los patrones de expresión de los lncRNAs: *Rbakdn* (A), *1110020A21Rik* (B), *Kcnmb4os1* (C) y el mRNA: *Kcnmb4* (D). En el eje X se observan las distintas poblaciones celulares estudiadas (2C, LZ, PS y RS). En el eje Y izquierdo, se muestran los valores de Fragmentos de transcrito por Millón de Secuencias Mapeadas (FPKM) obtenidos previamente por RNAseq (rojo). En el eje Y derecho se muestran los valores de cuantificación relativa (RQ) obtenidos a partir del ensayo de RT-qPCR (negro), donde se promedió el cálculo  $2^{-\Delta\Delta Ct}$  de tres réplicas biológicas para cada población celular. Estos gráficos muestran que los patrones de expresión obtenidos mediante RT-qPCR son consistentes con aquellos obtenidos mediante RNAseq en estudios anteriores.

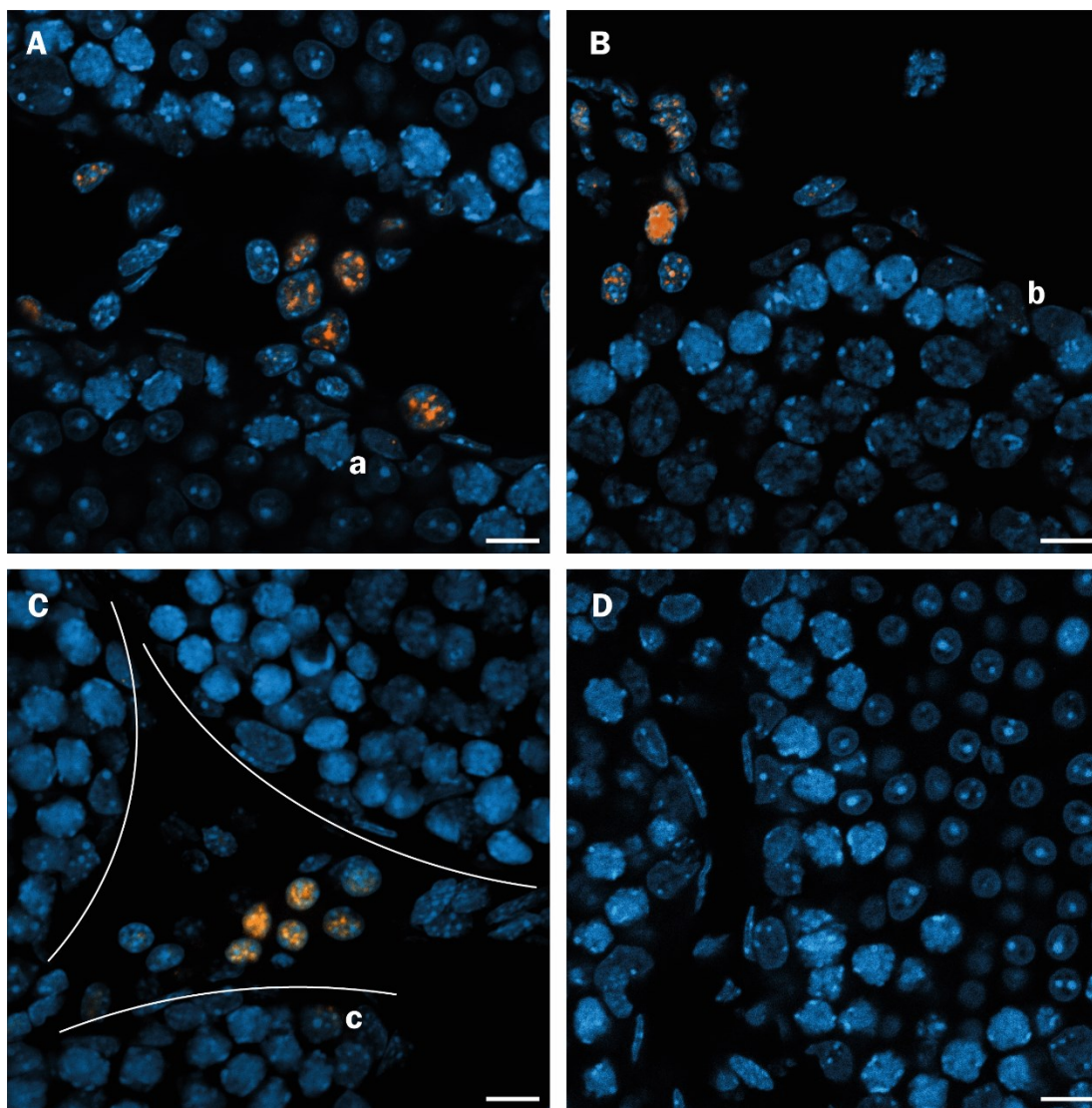
#### **4.4 Validación de RNA-FISH con *Malat1* como control positivo:**

El lncRNA *Malat1* (*metastasis-associated lung adenocarcinoma transcript 1*) conocido también como *Neat2* (*nuclear-enriched abundant transcript 2*), *HCN*, *LINC00047*, *NCRN00047*, o *PRO2853* (Hutchinson *et al.*, 2007), es uno de los lncRNAs más estudiados hasta el momento, y fue identificado inicialmente como un gen que se encontraba sobreexpresado en células metastásicas de cáncer de pulmón (Ji *et al.*, 2003). En humanos, el gen de este lncRNA se encuentra localizado en el cromosoma 11q13, con un largo de 7 kb, mientras que en el ratón se encuentra en el cromosoma 19qA, teniendo aproximadamente 6,7 kb de largo (Zhang *et al.* 2017).

En comparación a genes codificantes, los ncRNAs suelen tener menos restricciones evolutivas a la hora de modificar sus secuencias primarias que los lncRNAs, así como también en sus estructuras secundarias, por lo que las secuencias y estructuras secundarias predichas para muchos lncRNAs están pobremente conservadas (Ulitsky & Bartel, 2013). Sin embargo, *Malat1* es una excepción, donde la secuencia primaria, así como su estructura secundaria se encuentran altamente conservadas (Nitsche & Stadler, 2017).

El transcripto maduro de *Malat1* reside dentro del núcleo celular, co-localizando con los *speckles* nucleares (también llamados clústeres de gránulos intercromatínicos), unos dominios nucleares no membranosos localizados en regiones intercromatínicas, y enriquecidos en factores de splicing (Tripathi *et al.*, 2010) con los que *Malat1* interaccionaría (Nakagawa *et al.*, 2012).

Si bien *Malat1* se encuentra sobreexpresado en múltiples células tumorales (Michalik *et al.*, 2014), por lo que ha sido históricamente asociado al cáncer (Amodio *et al.*, 2018), su expresión en células normales de tejido adulto de ratón varía sustancialmente entre los distintos tipos de tejido, siendo el cerebro el tejido donde este lncRNA se expresa en niveles más altos, tejido donde se lo ha visto involucrado en la regulación de genes relacionados a la formación de sinapsis y desarrollo de dendritas (Nakagawa *et al.*, 2012; Zhang *et al.*, 2017). Sin embargo, esencialmente todas las células de un individuo adulto de ratón tienen cierta expresión de *Malat1*, en mayor o en menor medida (Nakagawa *et al.*, 2012). En el testículo, los estudios previos de RNAseq realizados por el equipo de investigación demostraron una alta expresión de *Malat1* en todas las etapas de la espermatogénesis, pero muy particularmente abundante en células de contenido 2C, que incluye células pre-meióticas y células somáticas (Trovero, 2020).



**Figura 12:** Localización sub-celular de *Malat1* mediante RNA-FISH en crio-secciones de testículo de ratón de 25 dpp, visualizado a través de microscopía confocal. (A, B, C) RNA-FISH con sondas Stellaris® para el lncRNA *Malat1*. Observamos la señal de *Malat1* en naranja brillante y la señal de DAPI en azul. En (C) se especifica con una línea blanca los bordes de los túbulos seminíferos. En (a, b, c) puede verse una muy tenue señal en los núcleos de células de Sertoli. (D) Control negativo (sin sonda). Barras de escala: 10  $\mu$ m

Los resultados de los ensayos de RNA-FISH con sondas contra *Malat1* pueden verse en la Figura 12. Obtuvimos una clara señal nuclear, la cual es intercromatínica en células somáticas, tal como era esperado según los estudios de Tripathi *et al.* (2010) para otros tipos de células somáticas. Puede apreciarse una señal muy intensa de color naranja

brillante de las sondas de *Malat1* en los núcleos de células intersticiales (Figura 12A, B, C), aunque también puede notarse una señal mucho más débil dentro de los núcleos de células somáticas intratubulares (a, b, c), que por su aspecto probablemente corresponden a las células de Sertoli. Dichos resultados concuerdan con lo esperado según los niveles de expresión de RNAseq para células de contenido 2C obtenidos anteriormente por el equipo de investigación (Tabla 4), por lo que se evidencia la validez de la técnica. Cabe señalar que la señal en las células de Sertoli era bastante más intensa, pero la misma se observa débil debido a que se disminuyó considerablemente el valor del fotomultiplicador (PMT) para la captura, porque de lo contrario la señal en las células intersticiales aparecía demasiado saturada. En ese sentido, destacamos que también se detectó una buena señal, aunque bastante más tenue, en las células de la línea germinal, coincidiendo con lo esperado de acuerdo con los datos de transcriptómica (Trovero, 2020). No obstante, dicha señal no se aprecia en las imágenes mostradas en la Figura 12, nuevamente debido a la disminución del valor del PMT.

#### **4.5 Localización sub-celular de los lncRNAs estudiados: ¿cuerpo cromatoide?**

Todos los lncRNAs estudiados en este trabajo mostraron una señal en los ensayos de RNA-FISH, sobre secciones de tejido testicular congelado de ratones con una edad de 25 dpp. En todos los casos se visualizó una señal citoplasmática, la cual se encontraba formando un punto en posición perinuclear en células posmeióticas (Figura 13), más específicamente en las espermátidas redondas, identificables por la forma de su núcleo y la compactación de su cromatina (Romrell *et al.*, 1976). Cabe destacar que esta señal estaba ausente tanto en el control positivo con *Malat1* (Figura 12), así como en el control negativo de este ensayo (Figura 13D), y que fue detectada para los tres lncRNAs estudiados: *Rbakdn*, *1110020A21Rik* y *Kcnmb4os1*.

Esta señal de punto, en posición perinuclear y en este tipo celular, es muy similar a la observada en los estudios de Kotaja *et al.* (2006), Meikar *et al.* (2011) y Lehtiniemi & Kotaja (2018), donde la misma corresponde al cuerpo cromatoide (CC). Como mencionamos anteriormente, el CC es un organelo perinuclear sin membrana, único de las células posmeióticas (ver Introducción).

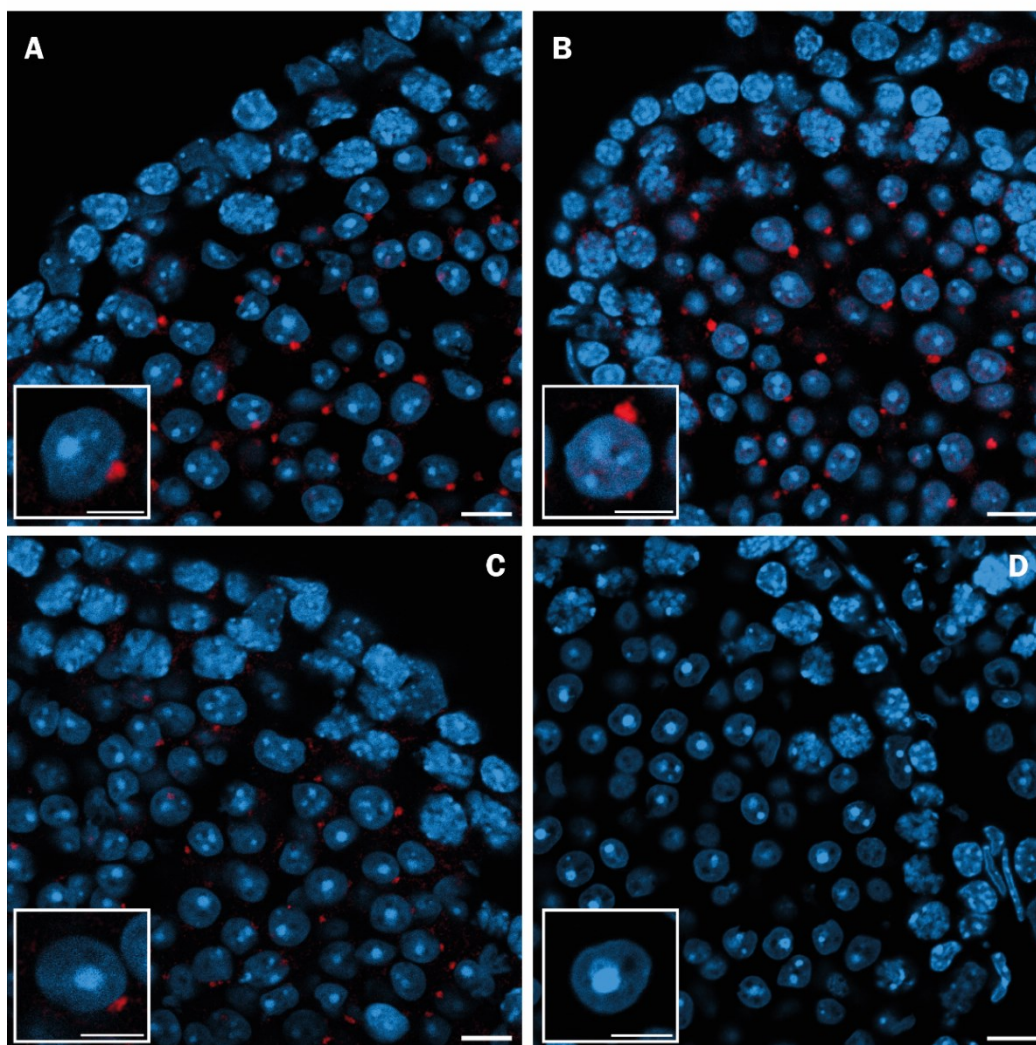
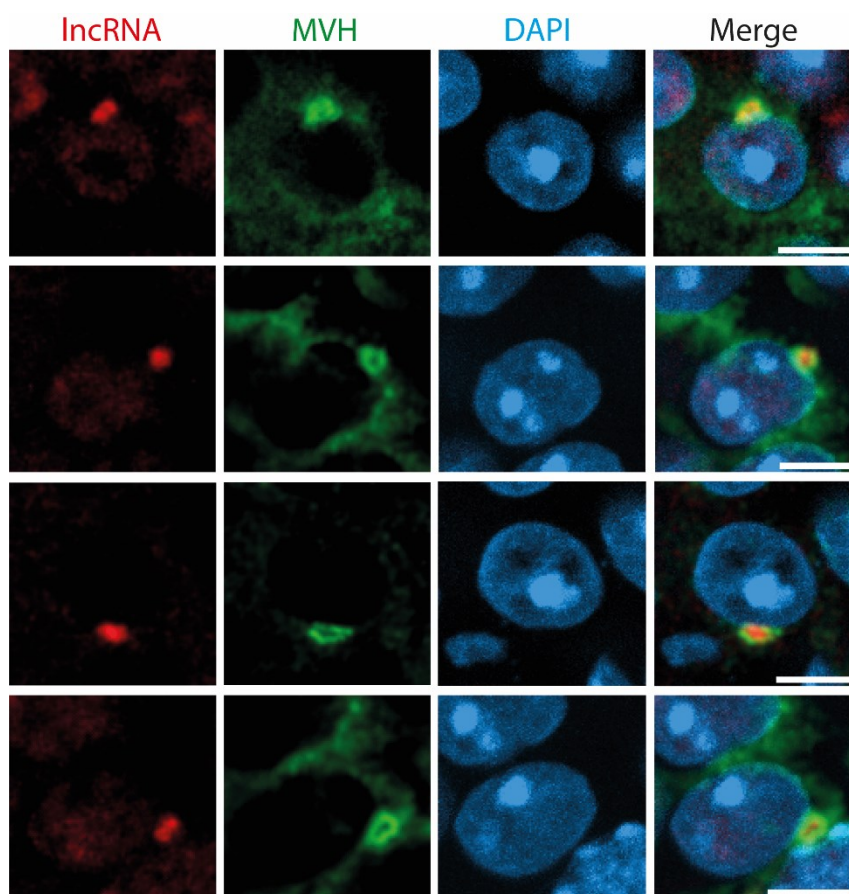


Figura 13: Localización sub-celular de los lncRNAs estudiados en ensayos de RNA-FISH, utilizando sondas Stellaris®, sobre criosecciones testiculares de ratón de 25 dpp. Se visualizan los resultados a través de microscopía confocal, donde las sondas para cada lncRNA se observan en rojo, y DAPI en azul. En (A) vemos a *Rbakdn*, en (B) *Kcnmb4os1*, en (C) *1110020A21Rik* y en (D) control negativo. Los recuadros muestran ampliaciones de un núcleo. Barras de escala: 10 µm (imágenes) y 5 µm (recuadros).

Con la finalidad de evidenciar que la señal observada en nuestros ensayos se trataba efectivamente del CC, procuramos utilizar un marcador específico de estos gránulos en conjunto con nuestras sondas de RNA-FISH. Un candidato óptimo para dicho fin era la proteína MVH. Estudios anteriores han reportado que MVH (por su nombre en inglés: *Mouse Vasa Homologue*), una helicasa de ARN, es una de las proteínas más abundantes del CC, contribuyendo junto con la proteína MIWI a aproximadamente dos tercios de la estructura proteica del CC (Meikar *et al.*, 2014). MVH no solamente es uno de los

componentes proteicos más abundantes del CC, sino también se ha demostrado que se expresa exclusivamente en células germinales primordiales tras la colonización de las gónadas y en células germinales hasta etapas posmeióticas (Toyooka *et al.*, 2000).

Utilizando un anticuerpo anti-MVH en ensayos de inmunodetección por fluorescencia en conjunto con RNA-FISH con nuestras sondas de ARN, obtuvimos que ambas señales co-localizan sobre los mismos puntos. Podemos apreciar en la Figura 14 cómo la señal de las sondas contra *Rbakdn* y *111020A21Rik* en rojo coinciden con la señal del anticuerpo anti-MVH observada en verde. El resultado de estos ensayos para las sondas contra *Kcnmb4os1* es idéntico y se verá más adelante. Con estos resultados podemos concluir que los lncRNAs estudiados en este trabajo se encuentran contenidos dentro de la estructura del CC en las espermátidas redondas.



**Figura 14:** Co-localización de lncRNAs con el CC. Se observa el resultado tras realizar un ensayo de inmunodetección con un anticuerpo anti-MVH y RNA-FISH en criosecciones de testículo de ratón de 25 dpp, observadas a través de microscopio confocal. En rojo se visualizan las sondas contra los lncRNAs: *Rbakdn* (primeras dos filas) y *111020A21Rik* (tercera y cuarta fila), mientras que en verde se observa anti-MVH, y DAPI en azul. Barras de escala: 5  $\mu$ m.

#### **4.6 Co-localización en el cuerpo cromatoide del mRNA *Kcnmb4* con su lncRNA antisentido solapante *Kcnmb4os1*:**

De los transcritos seleccionados para estudiar su localización, *Kcnmb4os1* era un lncRNA antisentido cuya secuencia en el extremo 5' solapa parcialmente (253 pb) con el extremo 3' de su gen codificante *Kcnmb4* (Figura 15). *Kcnmb4* (*Potassium calcium-activated channel subfamily M regulatory beta subunit 4*) codifica para la unidad reguladora  $\beta 4$  de un canal de potasio activado por calcio, de la familia de canales Maxik (Meera *et al.*, 2000). Este tipo de canales son caracterizados por una gran conductancia al potasio y por ser muy sensibles al voltaje y a  $\text{Ca}^{2+}$ , estando formados por cuatro subunidades  $\alpha$  transmembrana, las cuales pueden estar complementadas por subunidades  $\beta$ , como KCNMB4, que pueden modificar la conducta del canal, por ejemplo, alterando su cinética, su sensibilidad al  $\text{Ca}^{2+}$  o al voltaje, así como su respuesta a algunos agentes farmacológicos (Toro *et al.*, 2013). Si bien no se conoce aún la función específica de la subunidad codificada por *Kcnmb4*, por ejemplo, se ha sugerido que en el cerebro de ratón esta subunidad puede modular la excitabilidad neuronal y la liberación de neurotransmisores, provocando una desaceleración de la cinética de activación y un cambio de la dependencia del voltaje de los canales que complementa (Weiger *et al.*, 2000). En el testículo, también se ha propuesto que la proteína KCNMB4 podría estar involucrada en modular la función de canales específicos de potasio de este tejido (Yang *et al.*, 2009).

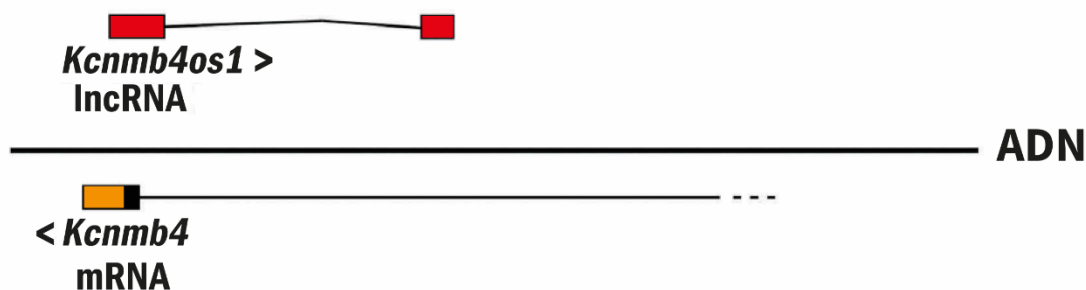
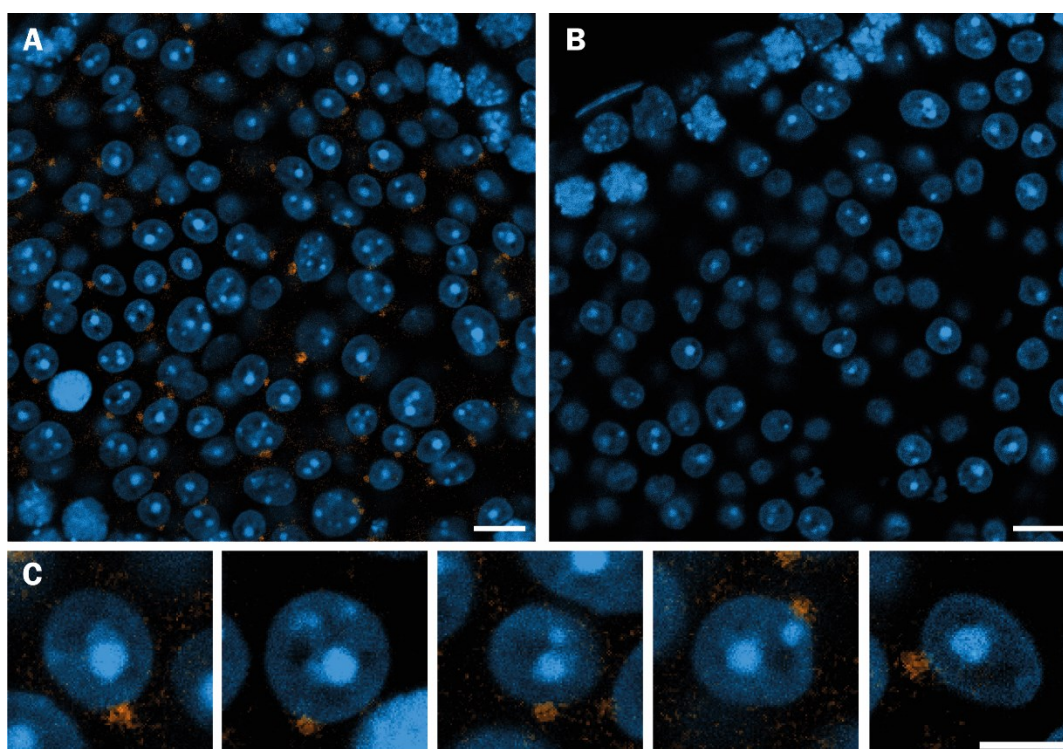


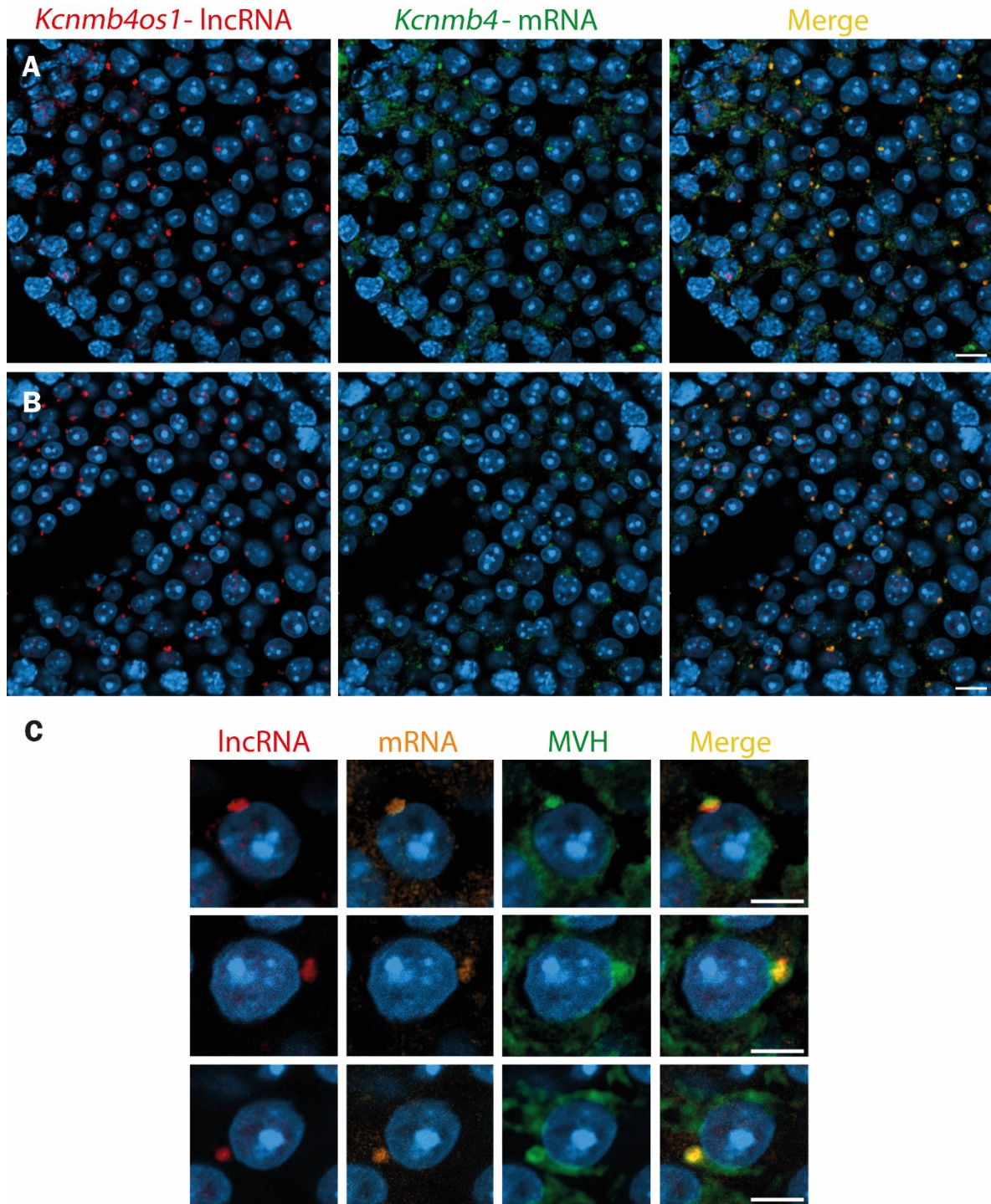
Figura 15: Diagrama de estructura y dirección del mRNA *Kcnmb4* y su lncRNA antisentido solapante *Kcnmb4os1*. Se muestra cómo ambos transcritos solapan parcialmente entre sí. Tomado y modificado de Trovero *et al.*, 2020.

Como vimos previamente, el mRNA que codifica a esta proteína está sobreexpresado en el tejido testicular (Figura 9). Al realizar un ensayo de RNA-FISH con las sondas diseñadas contra este mensajero, observamos una señal muy similar a la vista anteriormente para los lncRNAs estudiados, donde la señal para *Kcnmb4* parece agruparse en un punto en posición perinuclear de las espermátidas (Figura 16).

Con la finalidad de demostrar que la señal emitida por las sondas contra este mRNA también provenía del CC, se realizó un ensayo de RNA-FISH utilizando las sondas contra el mensajero de *Kcnmb4* y contra el lncRNA antisentido solapante *Kcnmb4os1*, y se verificó mediante co-inmunodetección con el anticuerpo anti-MVH. Ambas señales, tanto de *Kcnmb4* como de *Kcnmb4os1* colocalizan sobre el CC (Figura 17), con un coeficiente de correlación de Pearson de 0,974. Para reforzar dicho resultado, la Figura Suplementaria 2 del Anexo, muestra un diagrama de correlación para la señal emitida por ambas sondas sobre el tejido. Por lo tanto, de este ensayo podemos concluir que tanto el mRNA *Kcnmb4* como su lncRNA antisentido solapante *Kcnmb4os1* se encuentran dentro de la estructura del CC.



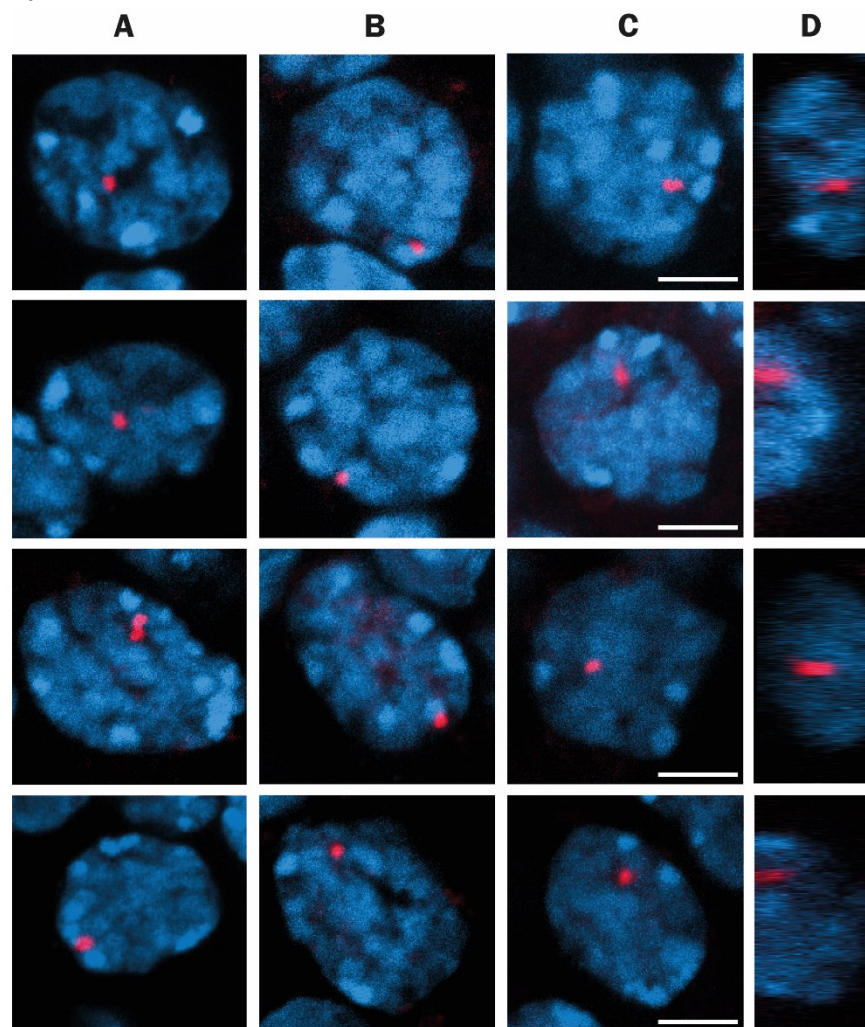
**Figura 16:** Localización sub-celular del mRNA *Kcnmb4* sobre criosecciones de testículo de ratón de 25 dpp, visualizadas en microscopio confocal. En (A) podemos observar la señal emitida por las sondas de *Kcnmb4* en naranja y DAPI en azul, a una escala más amplia del túbulo seminífero mientras que en (C) se aprecian ampliaciones de espermátidas con dicha señal. En (B) se muestra el control negativo. Barras de escala: 10  $\mu$ m (A, B) y 5  $\mu$ m (C).



**Figura 17:** Co-localización del mRNA *Kcnmb4* y el lncRNA *Kcnmb4os1* en el CC mediante RNA-FISH e inmunodetección sobre criosecciones de tejido testicular de ratón de 25 dpp. En (A y B) podemos observar el ensayo de RNA-FISH utilizando las sondas contra *Kcnmb4os1* en rojo y *Kcnmb4* en verde, visualizando la superposición de ambas señales en amarillo. En (C) se muestra el ensayo de inmunodetección con anti-MVH en verde, seguido de RNA-FISH con ambas sondas: *Kcnmb4os1* en rojo y *Kcnmb4* en naranja. DAPI en azul para todas las imágenes. Barras de escala: 10  $\mu$ m (A, B) y 5  $\mu$ m (C).

#### **4.7 El lncRNA *Kcnmb4os1* se localiza dentro del núcleo de los espermatocitos:**

Durante los ensayos de RNA-FISH con sondas contra el lncRNA *Kcnmb4os1*, notamos que, además de presentarse la notoria señal en el CC, también se apreciaba una señal dentro del núcleo de algunas células de la profase meiótica (Figura 18). Esta señal, que no se observó para ninguna de las otras sondas, y que no estaba presente en ninguno de los controles positivo o negativo, se apreciaba como un punto que por momentos pareciesen ser dos puntos muy próximos uno del otro (ver, por ejemplo, Figura 18A, tercera fila), dentro de los núcleos de los espermatocitos en paquiteno, identificables por la forma de su núcleo y las características de su cromatina.



**Figura 18:** Localización del lncRNA *Kcnmb4os1* en el núcleo de espermatocitos en paquiteno, sobre criosecciones de ratón de 25 dpp, observado a través de microscopio confocal. En las columnas (A, B y C) se muestran distintos ejemplos, con la señal de *Kcnmb4os1* en rojo. En (D) se observa la vista ortogonal de los núcleos de (C) con el fin de demostrar que esta señal se halla dentro del núcleo. DAPI en azul. Barras de escala: 5  $\mu$ m.

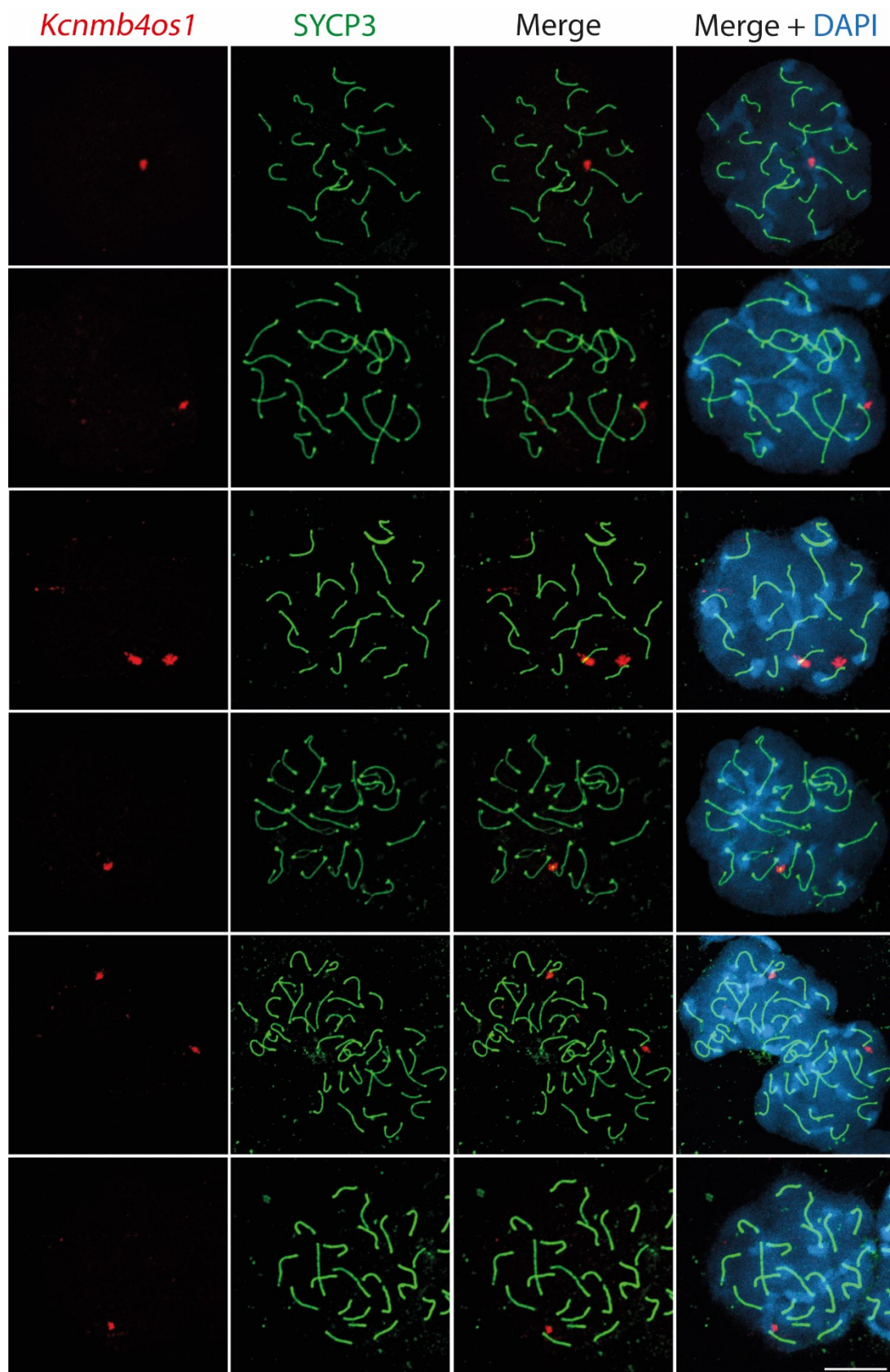
Estos resultados nos recordaron de manera notable a los obtenidos por Ding *et al.*, 2012, en relación a un lncRNA participante en el reconocimiento de homologías durante la profase meiótica de la levadura *S. pombe* (ver Introducción, Figura 8 y texto relacionado).

Por lo tanto, inspirados por la posibilidad de que la señal observada en nuestros experimentos pudiese ser parte de un mecanismo similar, se buscó evidenciar si existía una interacción entre *Kcnmb4os1* y el complejo sinaptonémico. Dado que no es tan sencillo discernirlo en cortes, decidimos optar por otro tipo de preparados, donde pudiese verse más claramente esta posible interacción. Para esto utilizamos la técnica de esparcido de suspensión celular, la cual consiste en un tratamiento hipotónico (ver Materiales y Métodos) donde eliminamos el citoplasma, quedándonos únicamente con los núcleos, cuyo contenido es esparcido sobre el portaobjetos de la muestra. De esta forma es posible observar más al detalle el contenido nuclear de las células en cuestión.

En la figura 19 se muestra el resultado de realizar dicho ensayo de RNA-FISH sobre esparcidos, en simultáneo con un anticuerpo contra el componente de los elementos laterales del complejo sinaptonémico SYCP3 (Yuan *et al.*, 2000). Como se puede apreciar detectamos una notoria señal de las sondas contra *Kcnmb4os1*, la que pareciera estar en contacto o localizada muy cerca de uno de los extremos de un complejo sinaptonémico de uno (o excepcionalmente dos) cromosomas (es importante tener en cuenta que, eventualmente, el propio tratamiento hipotónico y “estallado” de los núcleos podría despegar, e incluso llegar a desplazar, la señal de su localización exacta).

La figura 20 muestra una ampliación de algunas imágenes de FISH/inmunolocalización sobre esparcidos, de modo de permitir observar la señal, así como su localización, con mayor grado de detalle. En las imágenes A a C se observan imágenes de complejos sinaptonémicos aún no completamente ensamblados (los recuadros interiores, sin la sonda, permiten apreciar regiones aún de hebra simple, formando horquillas). En estas imágenes, parece apreciarse una señal doble, localizada sobre, o bordeando, las zonas sin aparear.

Si bien con los resultados obtenidos hasta el momento no podemos aseverar rotundamente que exista una interacción entre la señal del lncRNA y el complejo sinaptonémico o la cromatina, sí podemos plantear que existe una fuerte posibilidad de que esta interacción sea real, así como también que la señal se localice sobre su propio *locus*, tal como sucede en *S. pombe*.



**Figura 19:** Ensayo de inmunodetección con anti-SYCP3 y RNA-FISH con sondas de *Kcnmb4os1*, en esparcidos de suspensiones celulares de tejido testicular de ratón, vistas al microscopio confocal. Se muestra la señal del lncRNA en rojo, SYCP3 en verde y DAPI en azul. Barra de escala: 10  $\mu$ m

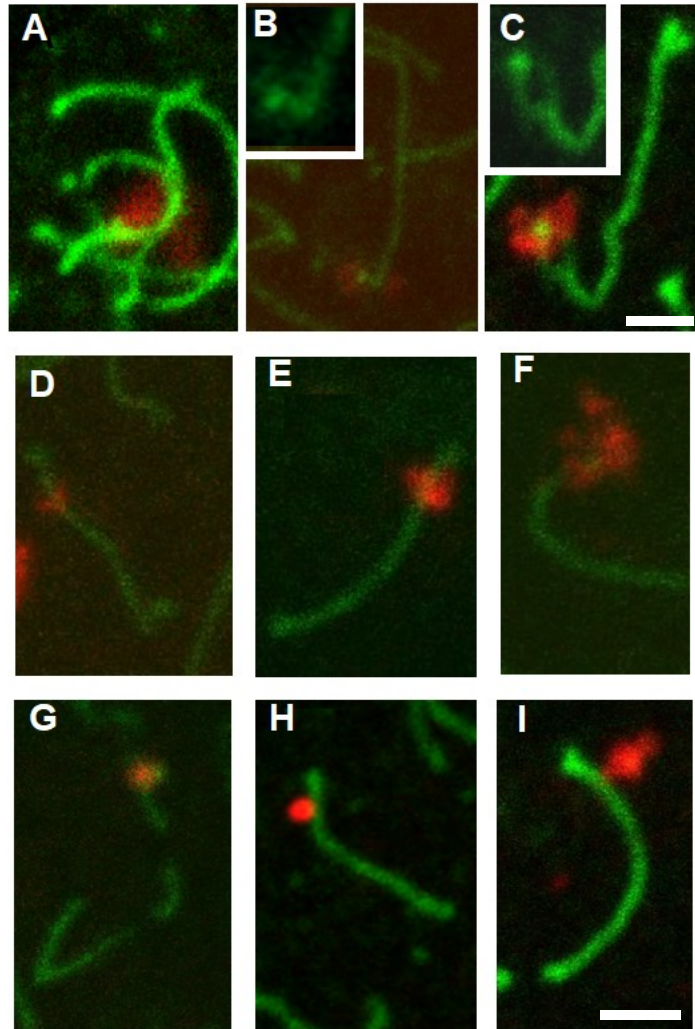


Figura 20: Ampliaciones de imágenes de inmunofluorescencia sobre complejos sinaptonémicos con anticuerpo anti-SYCP3 (verde) y RNA-FISH secuencial de *IncRNA-1* (rojo) sobre esparcidos de espermatocitos de ratón. Se observan ejemplos de la señal en complejos sinaptonémicos con distintos grados de ensamblado. Barras de escala: 2  $\mu$ m

## **5. Discusión:**

Este trabajo tuvo como objetivo aportar nuevos conocimientos sobre las bases moleculares de la reproducción masculina, en particular en lo relativo al potencial rol de secuencias no codificantes largas y a los posibles mecanismos subyacentes en los que éstas puedan participar durante la espermatogénesis. Nuestra hipótesis es que la alta expresión de lncRNAs estaría indicando que al menos una parte de estos transcriptos cumplen un rol estructural y/o regulatorio durante la espermatogénesis. A continuación, se discutirán los resultados obtenidos durante el transcurso de esta investigación.

### **5.1 Los lncRNAs durante la espermatogénesis y su relación con el cuerpo cromatoide:**

El testículo tiene la particularidad de ser el tejido que presenta la mayor diversidad y complejidad a nivel transcriptómico en relación a otros órganos. Esta característica no sólo abarca a los transcriptos codificantes, sino que también incluye la abrumadora expresión de otro tipo de transcriptos, como pseudogenes o elementos transponibles provenientes de regiones intergénicas, así como también lncRNAs (Soumillon *et al.*, 2013). Como hemos mencionado previamente, de la totalidad de células que componen el tejido testicular, las espermátidas redondas y los espermátocitos son los tipos celulares que presentan el mayor nivel de expresión y complejidad de transcriptoma, en primer y segundo lugar respectivamente. En este mismo orden, son estos dos tipos celulares los que muestran el máximo número de lncRNAs expresados y diferencialmente expresados en todo el tejido (Trovero *et al.*, 2020).

La razón por la cual estos tipos celulares presentan esta alta expresión transcriptómica todavía no ha sido dilucidada; sin embargo, algunos investigadores han planteado distintas causas que la originen. En primer lugar, una posibilidad es que esta alta expresión esté vinculada a la importante represión transcripcional que sufren las células germinales en etapas avanzadas de la espermiogénesis. Como se mencionó anteriormente, durante la espermiogénesis ocurre el intercambio de las histonas, por proteínas de transición y finalmente por protaminas, lo que tiene como consecuencia una dramática compactación de la cromatina, ocasionando un silenciamiento generalizado del genoma. Este hecho implica que estas células deban sintetizar sus mensajeros específicos en etapas anteriores, y almacenarlos hasta el momento de su traducción en etapas avanzadas de la

espermiogénesis donde sean requeridas (White-Cooper & Davidson, 2011; Trovero, 2020). Este es un mecanismo demostrado para algunos genes codificantes, como es el caso de las protaminas, cuyos mRNAs son sintetizados en espermatidas redondas, pero no son traducidos hasta la etapa de elongación nuclear en espermatidas elongadas (Kleene, 1993). Este cambio en la remodelación de la cromatina que sufren las espermatidas en etapas avanzadas conlleva la necesidad de utilizar diversos mecanismos de regulación genética, dentro de los cuales algunos lncRNAs podrían estar involucrados (Geisinger *et al.*, 2021).

En segundo lugar, otra posibilidad planteada es que este elevado nivel de transcripción se deba a un estado más “permisivo” de la cromatina, como consecuencia del proceso de remodelado de la misma, en particular durante el remplazo de las histonas. En este contexto, la cromatina quedaría más expuesta a la entrada de la maquinaria transcripcional, teniendo como resultado una transcripción “promiscua” y generando un “ruido” transcripcional (Soumillon *et al.*, 2013). Si bien es probable que parte de esta extensa transcripción sea solamente ruido, es muy posible que muchos transcriptos no codificantes expresados durante la espermatogénesis posean una función importante durante este proceso. En este mismo sentido, anteriormente se mencionaron algunos ejemplos de lncRNAs involucrados en distintas etapas de la espermatogénesis, como son *Mrhl* y *HongrES2*, cuya expresión aberrante o delección ocasiona distintas patologías testiculares (Ni *et al.*, 2011; Akhade *et al.*, 2014; Kerr *et al.*, 2014). Además de estos ejemplos existen otros lncRNAs importantes durante la espermatogénesis, tales como *Tsx* (*Testis specific X-linked*) expresado fundamentalmente en profase meiótica e involucrado en la inactivación del cromosoma X, así como también los lncRNAs *Drm* (*dmrt1-related gene*), *AK015322*, *LncRNA033862*, *Spga-lncRNA1*, *Spga-lncRNA2*, para los cuales se ha demostrado funcionalmente su importancia en procesos como la proliferación y mantenimiento de espermatogonias y en la regulación de la entrada a la meiosis (Zhang *et al.*, 2010; Li *et al.*, 2016; Hu *et al.*, 2017; Joshi & Rajender, 2020).

De hecho, en investigaciones recientes (publicadas posteriormente a la realización de nuestros experimentos), uno de los lncRNAs aquí estudiados, ha demostrado ser muy importante durante la espermatogénesis en el ratón. Liu *et al.* (2021) demostraron que el lncRNA *Rbakdn* es un regulador vital de la meiosis. Estos investigadores reportaron que *Rbakdn* puede unirse a la proteína Ptbp2, un factor regulador de *splicing* alternativo esencial durante la espermatogénesis, capaz de regular postranscripcionalmente el *splicing* de más de 200 genes en el desarrollo de las células germinales. Ellos observaron tanto *in vitro* como *in vivo*, que el *knock-down* de *Rbakdn* reducía significativamente la expresión de esta

proteína, y plantean que este lncRNA podría estar involucrado con la estabilidad de Ptbp2, protegiéndola al inhibir su degradación mediada por ubiquitinación. Por lo tanto, *Rbakdn* tendría un rol fundamental en el proceso meiótico, mediante la regulación de este factor de transcripción muy importante durante la meiosis (Liu *et al.*, 2021).

En definitiva, si bien el estado permisivo de la cromatina durante la espermatogénesis podría generar la alta expresión de transcritos observada en estos tipos celulares, muchos de los cuales, incluyendo lncRNAs, podrían ser simplemente ruido transcripcional, muchos otros podrían haber adquirido funciones importantes en el transcurso de la evolución, e involucrarse en los complejos mecanismos postranscripcionales que las células germinales presentan durante la espermatogénesis (Trovero, 2020).

Relacionado a distintos procesos de regulación postranscripcional específicos de la espermatogénesis surge el CC, como un organelo sin membrana que parece tener cada vez más relevancia. Como se mencionó anteriormente, en los últimos años han surgido investigaciones en las cuales se asocia al CC al procesamiento y regulación de ARNs, ya que el mismo acumularía múltiples tipos de ARNs como piRNAs, mRNAs y lncRNAs, además de poseer en su estructura múltiples proteínas vinculadas a distintos mecanismos de procesamiento y regulación de ARN (Meikar *et al.*, 2014), y componentes de la maquinaria de NMD (Lehtiniemi & Kotaja, 2018).

Durante nuestros ensayos de RNA-FISH, todos los lncRNAs estudiados fueron localizados en el CC de las espermátidas redondas, independientemente de sus patrones de expresión o tipo de transcrito. Dicha señal no fue detectada en los controles positivos ni negativos, y se validó que correspondiera efectivamente al CC mediante colocación de la señal con un anticuerpo anti-MVH, uno de los componentes proteicos más abundantes del CC (Meikar *et al.*, 2014). Si bien la señal de localización perinuclear de estos lncRNAs en el CC es muy fuerte, no podemos descartar la posibilidad de que también se los encuentre distribuidos en parte en el citoplasma celular, pero que no sean claramente detectables a causa del umbral de detección utilizado (sobre el CC las señales de múltiples sondas confluyen sobre un punto haciéndolas visibles, mientras que la difusión en el citoplasma dificultaría su detección). Esto sucede para las sondas contra *Malat1*, el cual si bien es expresado en altos niveles durante distintas etapas espermatogénicas (incluso más alto que algunos de los lncRNAs estudiados), al encontrarse dispersa, su señal aparece más tenue que la señal de los lncRNAs concentrada en el CC (con excepción de la señal

de *Malat1*, concentrada y extremadamente alta, dentro de los núcleos de las células intersticiales).

Aunque los tres lncRNAs aquí estudiados representan una fracción ínfima de la totalidad de los transcritos de esta clase que son expresados durante la espermatogénesis, el hecho de que los tres fuesen a parar al CC podría ser una muestra de la importancia de este organelo en el desarrollo espermatogénico. Más aún, en el marco de una tesis doctoral del grupo se estudiaron otros dos lncRNAs, los cuales también fueron localizados en el CC (Trovero, 2020). Por lo tanto, el CC podría estar actuando como un lugar de almacenamiento de transcritos y proteínas de unión al ARN que pudiesen ser necesarios en etapas más avanzadas de la espermatogénesis, donde el silenciamiento transcripcional impidiese la transcripción de nuevos ARNs, pero también podría estar secuestrando transcritos para mediar su degradación (Meikar *et al.*, 2014).

Es muy posible que el CC esté llevando a cabo alguna de estas dos funciones, así como otras que aún desconocemos. Respecto de los transcritos aquí estudiados queda preguntarnos cuál es la razón por la cual éstos se encuentran en el CC. En este sentido, los estudios realizados sobre la funcionalidad del lncRNA *Rbakdn*, podrían brindarnos una leve pista. Siendo que *Rbakdn* es fundamental para el desenlace de la meiosis en los espermátocitos, es posible que en las espermátidas redondas este transcrito sea enviado al CC, mediante alguna vía de señalización, para ser degradado, habiendo ya cumplido su función principal. Alternativamente, podría tratarse de un lncRNA “moonlighting” (ver más abajo). Para el caso de los lncRNAs *Kcnmb4os1* y *1110020A21Rik*, la causa de su localización en el CC es todavía incierta. De estos últimos lncRNAs, cabe destacar la colocación de *Kcnmb4os1* con su mensajero *Kcnmb4*, la cual se discutirá en la siguiente sección.

## **5.2 Los lncRNAs antisentido solapantes y los transcritos codificantes:**

Los estudios realizados por Trovero (2020) mostraron que los lncRNAs antisentido son el segundo biotipo de lncRNA más expresado durante la espermatogénesis (aproximadamente representando un 40% del total de transcritos), y en la mayoría de los casos resultaron estar regulados positivamente con sus genes codificantes respectivos, donde tanto lncRNA como mRNA son sobreexpresados o reprimidos en las mismas etapas (Trovero *et al.*, 2020). En sus estudios, los mRNAs expresados junto con sus lncRNAs solapantes antisentido durante distintas etapas celulares, y para los cuales existiera información disponible en las bases de datos, mostraron, para el gen codificante, reportes

de vinculación a alguna función específica del testículo (en análisis de ontología de genes) o expresión específica de testículo. Se plantea entonces la posibilidad de que la expresión de estos lncRNAs sea ocasionada por el anteriormente mencionado estado “permisivo” de la cromatina, que permitiese la transcripción de las hebras opuestas al mRNA, sintetizando muchos lncRNAs antisentido solapantes (Barman *et al.*, 2019; Trovero, 2020). Asimismo, en apoyo a esta teoría, se observó que la expresión de lncRNAs intergénicos (lincRNAs) se asemejaba a la de sus genes codificantes vecinos a medida que éstos se acercaban entre sí, lo que sería lógico bajo la premisa de una cromatina más permisiva. Sin embargo, este hecho no significa que la expresión de estos transcriptos no codificantes sea sólo “ruido” transcripcional, sino que algunos de estos lncRNAs podrían haber adquirido importantes funciones regulatorias (Trovero, 2020). Como se ha demostrado para otros tejidos, la expresión de ciertos lncRNAs solapantes puede regular positiva o negativamente a su gen codificante en la hebra opuesta (Barman *et al.*, 2019).

En el transcurso de esta investigación se buscó seleccionar un lncRNA que fuese antisentido solapante a un gen codificante, bajo la sospecha de que pudiese actuar sobre su mRNA solapante regulando su expresión. Teniendo en cuenta la represión transcripcional durante etapas avanzadas de la espermiogénesis, algunos lncRNAs podrían secuestrar a su transcripto codificante producido desde la hebra opuesta y evitar así su traducción, o evitar su degradación con la función de reservarlo hasta que sea necesario en etapas más avanzadas. En este sentido se seleccionó al lncRNA antisentido *Kcnmb4os1* y a su mensajero *Kcnmb4*, los cuales solapan parcialmente en su secuencia y su expresión es similar a lo largo de las distintas etapas espermatogénicas, incrementándose a partir del paquiteno.

Durante los ensayos de RNA-FISH realizados en este estudio las sondas contra ambos transcriptos colocalizaron sobre la estructura del CC, lo que podría estar indicando la existencia de un mecanismo de regulación entre este lncRNA y su transcripto codificante antisentido, o por lo menos una relación funcional entre ambos. Dado que ambos transcriptos solapan parcialmente sobre 253 pb del extremo 5' de *Kcnmb4os1* y 3' de *Kcnmb4*, es posible que exista una interacción entre ambos mediante complementariedad de bases. Los ensayos realizados en el transcurso de esta investigación no son suficientes para determinar con precisión la existencia de esta interacción, ni podemos determinar el modo en que ambos transcriptos van a parar al CC, ya que las vías de señalización que dirigen los transcriptos al CC no han sido dilucidadas aún. Si bien ambos transcriptos podrían ir a parar al CC de forma independiente uno del otro, tal como sucede con otros

ARNs (Meikar *et al.*, 2014), la hipótesis de que *Kcnmb4os1* se una al mRNA *Kcnmb4*, a nivel de su secuencia complementaria y pueda así dirigirlo hacia el CC, para allí controlar su traducción, almacenamiento o degradación, nos resulta sumamente atractiva. Dado que *Kcnmb4* codifica para la unidad regulatoria  $\beta 4$  de un canal de potasio activado por calcio de la familia de canales Maxik (Meera *et al.*, 2000), es posible que su almacenamiento en el CC tenga una implicancia funcional, ya que las vías de señalización que son activadas por  $Ca^{2+}$ , así como los canales  $Ca^{2+}$  dependientes, juegan un rol fundamental durante la quimiotaxis, la hiperactivación y la reacción acrosómica del espermatozoide (Schaefer *et al.*, 1998; Oren-Benaroya *et al.*, 2008; Suarez, 2008; Shukla *et al.*, 2012).

### **5.3 *Kcnmb4os1* se localiza dentro del núcleo de los espermaticitos:**

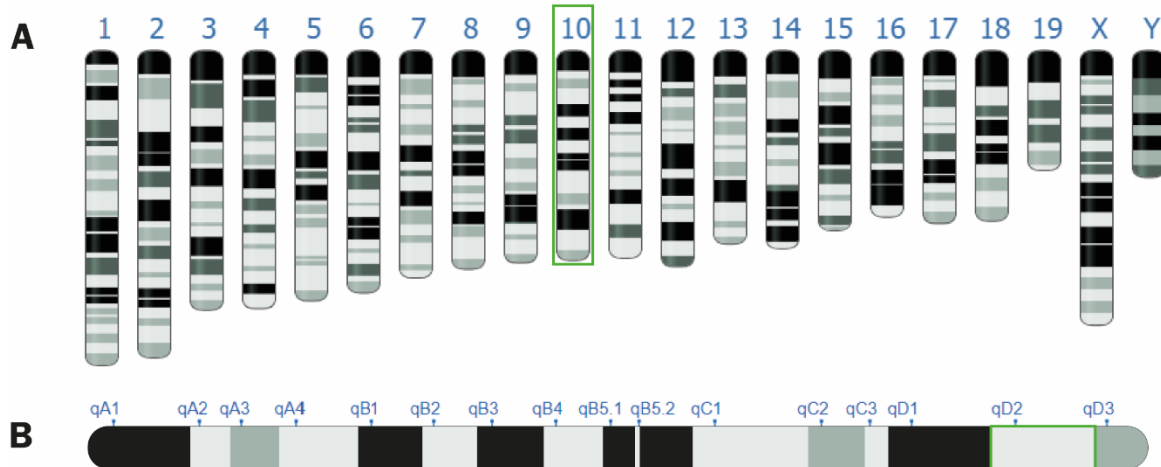
Además de la clara señal que *Kcnmb4os1* mostró sobre el CC en cortes de tejido testicular, este lncRNA también presentó una señal puntual (en algunos casos de dos puntos muy cercanos entre sí) en el interior del núcleo celular de los espermaticitos. Dicha señal no fue observada para ninguno de los otros lncRNAs, ni en el control positivo o negativo.

La presencia de una señal puntual, y por momentos muy similar a dos puntos cercanos durante el proceso meiótico, nos recordó a lo observado en los estudios sobre la levadura *S. pombe*, en donde, como hemos mencionado, un lncRNA actúa como mediador del proceso de apareamiento entre cromosomas homólogos. Como se explicó anteriormente, en esta levadura un lncRNA se acumula sobre su propio *locus* y media el proceso de apareamiento cromosómico. En el mecanismo planteado por Ding *et al.* los cromosomas homólogos podrían reconocerse entre sí sin la necesidad de una interacción directa entre las moléculas de ADN, sino que lo harían mediante ARNs que actúen como sitios de reconocimiento de homología, o como reclutadores de proteínas de unión al ADN específicas (Ding *et al.*, 2012 [B]; Ding *et al.*, 2013).

Teniendo en cuenta estos antecedentes, nos preguntamos si la señal intranuclear de *Kcnmb4os1* aquí observada, podría ser parte de un mecanismo similar en mamíferos. Con este fin, se realizaron en primera instancia ensayos secuenciales de RNA-FISH con inmunolocalización de CSs sobre criosecciones testiculares de ratón, y en segunda instancia sobre esparcidos de suspensiones celulares, para una mayor definición. Estos ensayos mostraron una señal de *Kcnmb4os1*, en algunos casos como un único punto bien definido, o como dos puntos separados, por lo general asociada a un par cromosómico, que

pareciera estar en contacto o localizada muy cerca de uno de los extremos del CS. Es de rigor decir que, en los esparcidos, la señal puntual mencionada no se observó en el 100% de los núcleos, sino que en algunos preparados la señal de *Kcnmb4os1* se presentó disgregada y no definida, lo que creemos que puede deberse a la propia técnica utilizada. Como se mencionó anteriormente, los esparcidos de suspensiones celulares consisten en eliminar el citoplasma celular mediante un tratamiento hipotónico, quedándose con los núcleos celulares, los cuales son esparcidos sobre el portaobjetos; en el proceso, la envoltura nuclear también es eliminada. Por lo tanto, en tan violento proceso es posible que la integridad de algunas estructuras intranucleares se vea muy afectada. En nuestro caso no conocemos qué tan resistente pueda ser la estructura conformada por *Kcnmb4os1*, por lo que la misma podría desarmarse durante la ruptura de la envoltura nuclear. Dado que en los ensayos sobre criosecciones testiculares únicamente se observaron señales puntuales bien definidas, es muy probable que la señal disgregada de algunos esparcidos sea debida a la propia técnica, y no a la real naturaleza de la estructura observada.

Más allá de la posible ruptura, en algunos casos, de este cúmulo formado por *Kcnmb4os1*, podemos observar que en la gran mayoría de los núcleos la señal se localiza llamativamente muy cerca de uno de los extremos de un cromosoma de tamaño mediano, que, comparando entre las distintas imágenes, pareciera tratarse del mismo cromosoma. Dicha observación resulta aún más llamativa al observar la Figura 21, en la que se muestra la localización cromosómica del gen de *Kcnmb4os1*, el cual se encuentra en el cromosoma número 10 en la posición qD2. Se trata de un cromosoma de tamaño mediano, y la posición D2 se haya muy cercana a los telómeros del mismo. Si bien con los resultados obtenidos hasta el momento no podemos evidenciar una interacción entre la señal del lncRNA y el CS o la cromatina, sí podemos plantear una fuerte posibilidad de que esa interacción exista, así como también de que esta señal se localice sobre su propio *locus*, tal como sucede en levaduras. Este lncRNA podría estar desempeñando alguna función regulatoria vinculada a la profase meiótica, como el reconocimiento de cromosomas homólogos u otros procesos, tales como la regulación de la expresión génica al reclutar proteínas modificadoras de la cromatina, como se ha visto en otros tejidos con otros lncRNAs vinculados a canales de potasio (Mohammad *et al.*, 2010; Pandey *et al.*, 2008).



**Figura 21:** Localización cromosómica del lncRNA *Kcnmb4os1*. El gen de *Kcnmb4os1* se encuentra localizado en el cromosoma 10 del ratón (A) y, dentro de éste, se encuentra en la región D2 (B), posición muy cercana a uno de sus telómeros. Imágenes obtenidas y adaptadas a partir de la web de NCBI (<https://www.ncbi.nlm.nih.gov/genome/gdv/browser/gene/?id=67342>).

#### **5.4 ¿LncRNAs “moonlighting” ?:**

Como consecuencia de las dos localizaciones observadas para *Kcnmb4os1*, intranuclear en forma de punto y citoplasmática en el CC, nos preguntamos si se podrá tratar de un lncRNA “moonlighting”. Este término es utilizado para proteínas, refiriendo a un fenómeno muy común en la naturaleza en donde una misma proteína, con una sola cadena polipeptídica, puede desempeñar distintas funciones en distintos compartimentos celulares e incluso en distintos tipos celulares. Las proteínas que generalmente presentan esta característica son enzimas, receptores celulares, canales iónicos y proteínas chaperonas, las cuales a lo largo de la evolución adquirieron funciones extra a las preexistentes (Boukouris *et al.*, 2016). Si bien existe multiplicidad de estudios respecto de proteínas moonlighting, el estudio de esta propiedad para lncRNAs ha sido casi inexistente, incluso teniendo en cuenta la gran relevancia que este tipo de transcritos han adquirido en la última década. Dada la gran variedad de procesos biológicos a los cuales los lncRNAs están asociados y a la gran diversidad de patrones de distribución de los mismos, varios de estos ya han sido señalados como posibles candidatos de lncRNAs moonlighting mediante estudios bioinformáticos (Cheng *et al.*, 2018).

En el caso de *Kcnmb4os1*, si bien son necesarios estudios en mayor profundidad, podríamos plantear que su acumulación intranuclear pudiese estar vinculada al reconocimiento de cromosomas homólogos o a la regulación de la expresión génica, mientras que en su localización citoplasmática su función pudiese ser la de secuestrar y dirigir su mRNA antisentido solapante *Kcnmb4* al CC para retardar su traducción o para su degradación, como se discutió anteriormente. En tal caso, *Kcnmb4os1* podría estar dentro de los primeros lncRNAs descritos con estas características de acción *moonlighting*.

Por otra parte, como mencionáramos más arriba, se ha reportado un patrón de localización y función para *Rbakdn* en las células en profase meiótica. Nosotros, sin embargo, lo localizamos en el CC de las espermátidas redondas. Como hemos expresado, es posible que su localización en el CC no obedezca a una función, sino que simplemente se almacene allí luego de cumplir su rol. No obstante, tampoco podemos descartar la posibilidad de que este lncRNA desempeñe alguna función, diferente de la ejercida en las células meióticas, en relación con el CC de las espermátidas.

## **6. Conclusiones y perspectivas:**

Como conclusión general de este trabajo, podemos decir que se cumplieron los objetivos planteados; en este sentido, se logró generar sondas de ARN específicas contra tres lncRNAs que presentaron patrones de expresión interesantes durante las distintas etapas de la espermatogénesis. De esta forma, fue posible estudiar su localización sub-celular en las células espermátogénicas y así elaborar hipótesis sobre sus posibles mecanismos de acción y procesos en los que estén involucrados. Para los tres lncRNAs aquí estudiados (*Rbakdn*, *1110020A21Rik* y *Kcnmb4os1*), su localización sub-celular coincidió en el cuerpo cromatoide (CC) de las espermátidas redondas, un organelo sin membrana para el cual se está proponiendo un importante rol en la regulación postranscripcional de las células espermátogénicas. La hipótesis imperante es que el CC acumularía transcritos, para ser utilizados en etapas avanzadas de la espermiogénesis donde la represión transcripcional impidiese la fabricación de nuevos ARNs, así como también actuaría como sitio de degradación de transcritos.

Adicionalmente, se determinó que el lncRNA antisentido solapante *Kcnmb4os1* se localiza en el CC con su respectivo mRNA *Kcnmb4*, codificante para una unidad reguladora de un canal de potasio activado por calcio. Si bien ambos podrían ir a parar a esta estructura de forma independiente, también cabe la posibilidad de que *Kcnmb4os1* secuestre a

*Kcnmb4* (por ejemplo, mediante su región complementaria) y de alguna forma lo redirija al CC, como un novedoso mecanismo de regulación postranscripcional.

El lncRNA *Kcnmb4os1* también presentó una señal puntual dentro del núcleo de los espermatoцитos. En ensayos de co-localización con CS la señal se localiza muy cerca de uno de los extremos de un cromosoma de tamaño mediano, el cual posiblemente pudiese tratarse del mismo cromosoma desde donde este lncRNA es transcrito, por lo que eventualmente podría estar acumulándose sobre su propio *locus*. En este sentido, sería ideal co-localizar la señal con sondas de DNA-FISH para el *locus* en cuestión, pero lamentablemente no hemos encontrado opciones comerciales que sinteticen sondas de este tipo, y los distintos proveedores a los que hemos consultado son muy escépticos acerca de la eventual sensibilidad de la técnica para detectar señales de secuencia única sobre cromosomas. Por lo tanto, como perspectiva a futuro, se plantea realizar ensayos de tinción cromosómica (*chromosome painting*), de tal forma de dilucidar si efectivamente *Kcnmb4os1* se acumula sobre su propio *locus* o lo hace en otro cromosoma. Estos ensayos ya están en curso, pero hasta el momento no se han obtenido resultados exitosos por falla de la sonda adquirida de *chromosome painting*. Por otra parte, y dado que *Kcnmb4os1* es un lncRNA muy prometedor, se comenzaron a realizar ensayos funcionales mediante el silenciamiento *in vivo* de este transcrito, de tal forma de estudiar su rol durante la espermatogénesis.

Por último, cabe destacar que parte de los resultados aquí obtenidos fueron incluidos en una publicación en una revista internacional arbitrada de alto impacto, la cual se adjunta en este trabajo.

## **7. Bibliografía:**

Akhade, V. S., Arun, G., Donakonda, S., & Satyanarayana Rao, M. R. (2014). Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells. *RNA biology*, 11(10), 1262-1279.

Amodio, N., Raimondi, L., Juli, G., Stamato, M. A., Caracciolo, D., Tagliaferri, P., & Tassone, P. (2018). MALAT1: a druggable long non-coding RNA for targeted anti-cancer approaches. *Journal of hematology & oncology*, 11(1), 1-19.

Bolcun-Filas, E., & Handel, M. A. (2018). Meiosis: the chromosomal foundation of reproduction. *Biology of reproduction*, 99(1), 112-126.

Boukouris, A. E., Zervopoulos, S. D., & Michelakis, E. D. (2016). Metabolic enzymes moonlighting in the nucleus: metabolic regulation of gene transcription. *Trends in biochemical sciences*, 41(8), 712-730.

Brunn, A. V. (1876). Beiträge zur Entwicklungsgeschichte der Samenkörper. *Archiv für mikroskopische Anatomie*, 12(1), 528-535.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Gustincich, S., & *et al.* (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*, 491(7424), 454-457.

Chen, Y., Pane, A., & Schüpbach, T. (2007). Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in *Drosophila*. *Current Biology*, 17(7), 637-642.

Cheng, L., & Leung, K. S. (2018). Identification and characterization of moonlighting long non-coding RNAs based on RNA and protein interactome. *Bioinformatics*, 34(20), 3519-3528.

Chuma, S., Hosokawa, M., Tanaka, T., & Nakatsuji, N. (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.

Cohen, P. E., Pollack, S. E., & Pollard, J. W. (2006). Genetic analysis of chromosome pairing, recombination, and cell cycle control during first meiotic prophase in mammals. *Endocrine reviews*, 27(4), 398-426.

Creasy, D. M., & Chapin, R. E. (2013). Male reproductive system. *Haschek and Rousseux's handbook of Toxicologic Pathology*, 2493-2598.

Ding, D. Q., Haraguchi, T., & Hiraoka, Y. (2012). (A). Chromosomally-retained RNA mediates homologous pairing. *Nucleus*, 3(6), 516-519.

Ding, D. Q., Haraguchi, T., & Hiraoka, Y. (2013). The role of chromosomal retention of noncoding RNA in meiosis. *Chromosome research*, 21(6), 665-672.

Ding, D. Q., Okamasa, K., Katou, Y., Oya, E., Nakayama, J. I., Chikashige, Y., Hiraoka, Y., & *et al.* (2019). Chromosome-associated RNA-protein complexes promote pairing of homologous chromosomes during meiosis in *Schizosaccharomyces pombe*. *Nature communications*, 10(1), 1-12.

Ding, D. Q., Okamasa, K., Yamane, M., Tsutsumi, C., Haraguchi, T., Yamamoto, M., & Hiraoka, Y. (2012). (B). Meiosis-specific noncoding RNA mediates robust pairing of homologous chromosomes in meiosis. *Science*, 336(6082), 732-736.

Eddy, E. M. (2002). Male germ cell gene expression. *Recent progress in hormone research*, 57(1), 103-128.

Engreitz, J. M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Guttman, M., & *et al.* (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*, 341(6147), 1237973.

Fraune, J., Schramm, S., Alsheimer, M., & Benavente, R. (2012). The mammalian synaptonemal complex: protein components, assembly and role in meiotic recombination. *Experimental cell research*, 318(12), 1340-1346.

Gan, H., Cai, T., Lin, X., Wu, Y., Wang, X., Yang, F., & Han, C. (2013). Integrative proteomic and transcriptomic analyses reveal multiple post-transcriptional regulatory mechanisms of mouse spermatogenesis. *Molecular & Cellular Proteomics*, 12(5), 1144-1157.

Geisinger, A., Rodríguez-Casuriaga, R., & Benavente, R. (2021). Transcriptomics of meiosis in the male mouse. *Frontiers in Cell and Developmental Biology*, 9, 626020.

Gómez-H, L., Felipe-Medina, N., Sánchez-Martín, M., Davies, O. R., Ramos, I., García-Tuñón, I., Pendas, A. M., & *et al.* (2016). C14ORF39/SIX6OS1 is a constituent of the synaptonemal complex and is essential for mouse fertility. *Nature communications*, 7(1), 1-16.

Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., Chang, H. Y., & *et al.* (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *nature*, 464(7291), 1071-1076.

Handel, M. A., & Schimenti, J. C. (2010). Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. *Nature Reviews Genetics*, 11(2), 124-136.

Hosseini, E. S., Meryet-Figuire, M., Sabzalipoor, H., Kashani, H. H., Nikzad, H., & Asemi, Z. (2017). Dysregulated expression of long noncoding RNAs in gynecologic cancers. *Molecular cancer*, 16(1), 1-13.

Hu, K., Zhang, J., & Liang, M. (2017). LncRNA AK015322 promotes proliferation of spermatogonial stem cell C18-4 by acting as a decoy for microRNA-19b-3p. *In Vitro Cellular & Developmental Biology-Animal*, 53(3), 277-284.

Hutchinson, J. N., Ensminger, A. W., Clemson, C. M., Lynch, C. R., Lawrence, J. B., & Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. *BMC genomics*, 8(1), 1-16.

Ji, P., Diederichs, S., Wang, W., Böing, S., Metzger, R., Schneider, P. M., Müller-Tidow, C., & *et al.* (2003). MALAT-1, a novel noncoding RNA, and thymosin  $\beta$ 4 predict metastasis and survival in early-stage non-small cell lung cancer. *Oncogene*, 22(39), 8031-8041.

Joshi, M., & Rajender, S. (2020). Long non-coding RNAs (lncRNAs) in spermatogenesis and male infertility. *Reproductive Biology and Endocrinology*, 18(1), 1-18.

Kerr, G. E., Young, J. C., Horvay, K., Abud, H. E., & Loveland, K. L. (2014). Regulated Wnt/beta-catenin signaling sustains adult spermatogenesis in mice. *Biology of reproduction*, 90(1), 3-1.

Kerr, J. B., Loveland, K. L., O'Bryan, M. K., & Kretser, D. M. (2006). Cytology of the testis and intrinsic control mechanism. *Knobil and Neill's physiology of reproduction*, 827-947.

Kleene, K. C. (1993). Multiple controls over the efficiency of translation of the mRNAs encoding transition proteins, protamines, and the mitochondrial capsule selenoprotein in late spermatids in mice. *Developmental biology*, 159(2), 720-731.

Kleene, K. C. (2001). A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mechanisms of development*, 106(1-2), 3-23.

Kotaja, N., Lin, H., Parvinen, M., & Sassone-Corsi, P. (2006). Interplay of PIWI/Argonaute protein MIWI and kinesin KIF17b in chromatoid bodies of male germ cells. *Journal of cell science*, 119(13), 2819-2825.

Lee, K., Haugen, H. S., Clegg, C. H., and Braun, R. E. (1995). Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. *Proc. Natl. Acad. Sci. U.S.A.* 92, 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.

Li, L., Wang, M., Wu, X., Geng, L., Xue, Y., Wei, X., & Jia, Y. (2016). A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells. *Cell death & disease*, 7(3), e2140-e2140.

Liu, K. S., Li, T. P., Ton, H., Mao, X. D., & Chen, Y. J. (2018). Advances of long noncoding RNAs-mediated regulation in reproduction. *Chinese Medical Journal*, 131(02), 226-234.

Liu, W., Zhao, Y., Liu, X., Zhang, X., Ding, J., Li, Y., Lu, Z., & *et al.* (2021). A Novel Meiosis-Related lncRNA, Rbkdn, Contributes to Spermatogenesis by Stabilizing Ptp2. *Frontiers in genetics*, 1963.

Meera, P., Wallner, M., & Toro, L. (2000). A neuronal  $\beta$  subunit (KCNMB4) makes the large conductance, voltage- and  $\text{Ca}^{2+}$ -activated  $\text{K}^{+}$  channel resistant to charybdotoxin and iberiotoxin. *Proceedings of the National Academy of Sciences*, 97(10), 5562-5567.

Meikar, O., Da Ros, M., Korhonen, H., & Kotaja, N. (2011). Chromatoid body and small RNAs in male germ cells. *Reproduction*, 142(2), 195-209.

Meikar, O., Vagin, V. V., Chalmel, F., Sõstar, K., Lardenois, A., Hammell, M., Kotaja, N., & *et al.* (2014). An atlas of chromatoid body components. *Rna*, 20(4), 483-495.

Michalik, K. M., You, X., Manavski, Y., Doddaballapur, A., Zörnig, M., Braun, T., Dimmeler, S., & *et al.* (2014). Long noncoding RNA MALAT1 regulates endothelial cell function and vessel growth. *Circulation research*, 114(9), 1389-1397.

Mohammad, F., Mondal, T., Guseva, N., Pandey, G. K., & Kanduri, C. (2010). Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. *Development*, 137(15), 2493-2499.

Mozdarani, H., Ezzatizadeh, V., & Rahbar Parvaneh, R. (2020). The emerging role of the long non-coding RNA HOTAIR in breast cancer development and treatment. *Journal of Translational Medicine*, 18(1), 1-15.

Nakagawa, S., Ip, J. Y., Shioi, G., Tripathi, V., Zong, X., Hirose, T., & Prasanth, K. V. (2012). Malat1 is not an essential component of nuclear speckles in mice. *RNA*, 18(8), 1487-1499.

Ni, M. J., Hu, Z. H., Liu, Q., Liu, M. F., Lu, M. H., Zhang, J. S., Zhang, Y. L., & *et al.* (2011). Identification and characterization of a novel non-coding RNA involved in sperm maturation. *PLoS One*, 6(10), e26053.

Nitsche, A., & Stadler, P. F. (2017). Evolutionary clues in lncRNAs. *Wiley Interdisciplinary Reviews: RNA*, 8(1), e1376.

Oren-Benaroya, R., Orvieto, R., Gakamsky, A., Pinchasov, M., & Eisenbach, M. (2008). The sperm chemoattractant secreted from human cumulus cells is progesterone. *Human reproduction*, 23(10), 2339-2345.

Page, S. L., & Hawley, R. S. (2004). The genetics and molecular biology of the synaptonemal complex. *Annu. Rev. Cell Dev. Biol.*, 20, 525-558.

Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Kanduri, C., & *et al.* (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular cell*, 32(2), 232-246.

Paz Rodríguez, J. I. (2016). Variaciones morfológicas en el testículo de Ratones Adultos Knock Out para la proteína de sustrato de receptor insulínico IRS2. *Tesis de doctorado en Biología Celular. Universidad de Salamanca. España.*

Pelttari, J., Hoja, M. R., Yuan, L., Liu, J. G., Brundell, E., Moens, P., Höög, C., & *et al.* (2001). A meiotic chromosomal core consisting of cohesin complex proteins recruits DNA recombination proteins and promotes synapsis in the absence of an axial element in mammalian meiotic cells. *Molecular and cellular biology*, 21(16), 5667-5677.

Petersen, C., & Söder, O. (2006). The sertoli cell—a hormonal target and ‘super’nurse for germ cells that determines testicular size. *Hormone Research in Paediatrics*, 66(4), 153-161.

Pontier, D. B., & Gribnau, J. (2011). Xist regulation and function explored. *Human genetics*, 130(2), 223-236.

Rodríguez-Casuriaga, R., Geisinger, A., López-Carro, B., Porro, V., Wettstein, R., & Folle, G. A. (2009). Ultra-fast and optimized method for the preparation of rodent testicular cells for flow cytometric analysis. *Biological procedures online*, 11(1), 184-195.

Rodríguez-Casuriaga, R., Geisinger, A., Santinaque, F. F., López-Carro, B., & Folle, G. A. (2011). High-purity flow sorting of early meiocytes based on DNA analysis of guinea pig spermatogenic cells. *Cytometry Part A*, 79(8), 625-634.

Rodríguez-Casuriaga, R., Santinaque, F. F., Folle, G. A., Souza, E., Lopez-Carro, B., & Geisinger, A. (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.

Romrell, L. J., Bellvé, A. R., & Fawcett, D. W. (1976). Separation of mouse spermatogenic cells by sedimentation velocity: a morphological characterization. *Developmental biology*, 49(1), 119-131.

Schaefer, M., Hofmann, T., Schultz, G., & Gudermann, T. (1998). A new prostaglandin E receptor mediates calcium influx and acrosome reaction in human spermatozoa. *Proceedings of the National Academy of Sciences*, 95(6), 3008-3013.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Cardona, A., & et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7), 676-682.

Shukla, K. K., Mahdi, A. A., & Rajender, S. (2012). Ion channels in sperm physiology and male fertility and infertility. *Journal of andrology*, 33(5), 777-788.

Suarez, S. S. (2008). Control of hyperactivation in sperm. *Human reproduction update*, 14(6), 647-657.

Syrjänen, J. L., Pellegrini, L., & Davies, O. R. (2014). A molecular model for the role of SYCP3 in meiotic chromosome organisation. *Elife*, 3, e02963.

Toro, L., Li, M., Zhang, Z., Singh, H., Wu, Y., & Stefani, E. (2013). MaxiK channel and cell signalling. *Pflügers Archiv-European Journal of Physiology*, 466(5), 875-886.

- Toyooka, Y., Tsunekawa, N., Takahashi, Y., Matsui, Y., Satoh, M., & Noce, T. (2000). Expression and intracellular localization of mouse Vasa-homologue protein during germ cell development. *Mechanisms of development*, 93(1-2), 139-149.
- Tripathi, V., Ellis, J. D., Shen, Z., Song, D. Y., Pan, Q., Watt, A. T., Prasanth, K. V., & *et al.* (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular cell*, 39(6), 925-938.
- Trovero, M. F. (2020). Rol de los ARNs no codificantes largos en la espermatogénesis. *Tesis de doctorado en ciencias biológicas, Biología Celular y Molecular*. Universidad de la República del Uruguay. Facultad de Ciencias. PEDECIBA.
- Trovero, M. F., & Geisinger, A. (2019). Los ARNs no codificantes largos y su vinculación con las patologías testiculares. In *Anales de la Facultad de Medicina* (Vol. 6, No. 1, pp. 12-47).
- Trovero, M. F., Rodríguez-Casuriaga, R., Romeo, C., Santiñaque, F. F., François, M., Folle, G. A., Geisinger, A., & *et al.* (2020). Revealing stage-specific expression patterns of long noncoding RNAs along mouse spermatogenesis. *RNA biology*, 17(3), 350-365.
- Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Chang, H. Y., & *et al.* (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, 329(5992), 689-693.
- Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell*, 154(1), 26-46.
- Voronina, E., Seydoux, G., Sassone-Corsi, P., & Nagamori, I. (2011). RNA granules in germ cells. *Cold Spring Harbor perspectives in biology*, 3(12), a002774.
- Wei, J. W., Huang, K., Yang, C., & Kang, C. S. (2017). Non-coding RNAs as regulators in epigenetics. *Oncology reports*, 37(1), 3-9.
- Weiger, T. M., Holmqvist, M. H., Levitan, I. B., Clark, F. T., Sprague, S., Huang, W. J., Curtis, R., & *et al.* (2000). A novel nervous system  $\beta$  subunit that downregulates human large conductance calcium-dependent potassium channels. *Journal of Neuroscience*, 20(10), 3563-3570.
- White-Cooper, H., & Davidson, I. (2011). Unique aspects of transcription regulation in male germ cells. *Cold Spring Harbor perspectives in biology*, 3(7), a002626.
- Winkel, K., Alsheimer, M., Öllinger, R., & Benavente, R. (2009). Protein SYCP2 provides a link between transverse filaments and lateral elements of mammalian synaptonemal complexes. *Chromosoma*, 118(2), 259-267.

Yang, C. T., Zeng, X. H., Xia, X. M., & Lingle, C. J. (2009). Interactions between  $\beta$  subunits of the KCNMB family and Slo3:  $\beta 4$  selectively modulates Slo3 expression and function. *PLoS One*, 4(7), e6135.

Yuan, L., Liu, J. G., Zhao, J., Brundell, E., Daneholt, B., & Höög, C. (2000). The murine SCP3 gene is required for synaptonemal complex assembly, chromosome synapsis, and male fertility. *Molecular cell*, 5(1), 73-83.

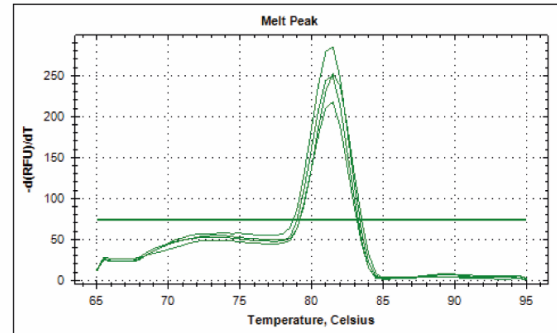
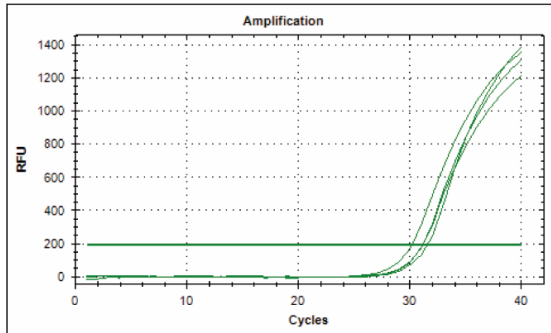
Zhang, L., Lu, H., Xin, D., Cheng, H., & Zhou, R. (2010). A novel ncRNA gene from mouse chromosome 5 trans-splices with Dmrt1 on chromosome 19. *Biochemical and biophysical research communications*, 400(4), 696-700.

Zhang, X., Hamblin, M. H., & Yin, K. J. (2017). The long noncoding RNA Malat1: Its physiological and pathophysiological functions. *RNA biology*, 14(12), 1705-1714.

Zickler, D., & Kleckner, N. (1999). Meiotic chromosomes: integrating structure and function. *Annual review of genetics*, 33, 603.

## 8.Anexo:

lncRNA Kcnmb4os1



mRNA Kcnmb4

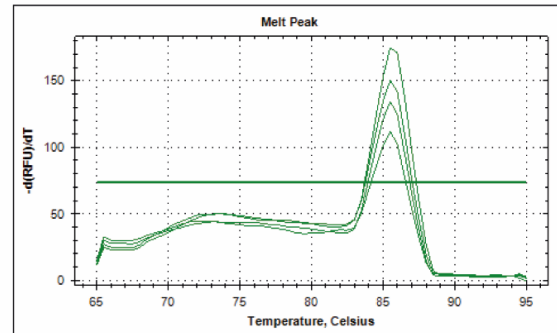
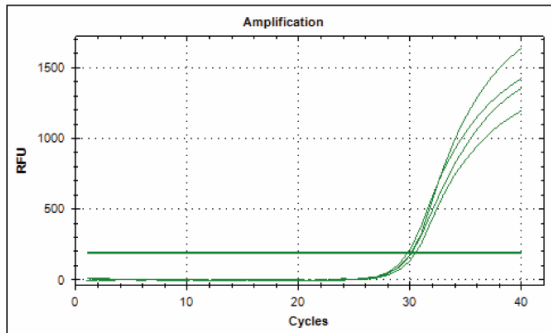


Figura suplementaria 1: A modo representativo se muestran las curvas de amplificación (izquierda) y curvas de *melting* (derecha), obtenidas tras el análisis por RT-qPCR del lncRNA antisentido *Kcnmb4os1* y el mRNA *kcnmb4*. Cada curva mostrada representa una réplica biológica distinta.

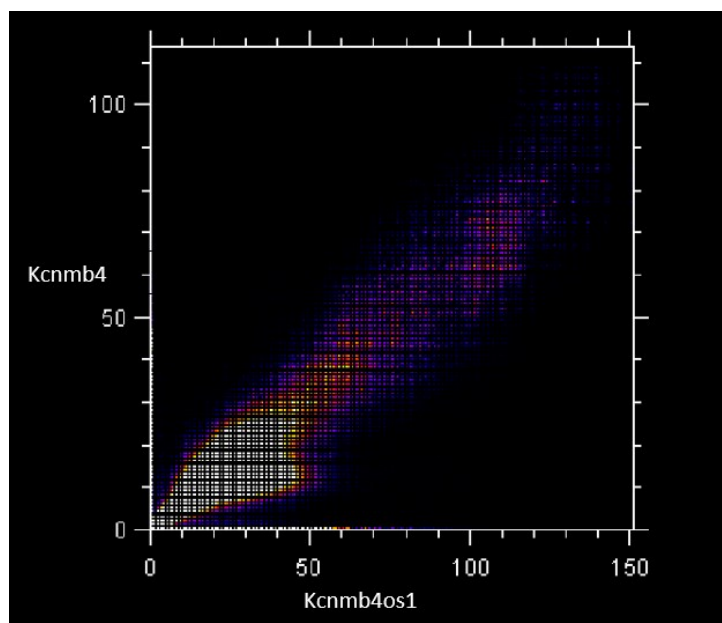
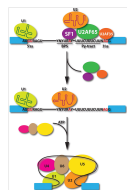


Figura suplementaria 2: Diagrama de correlación para las señales emitidas por ambas sondas contra el mRNA *Kcnmb4* y el lncRNA *Kcnmb4os1*. Podemos observar que ambas señales aumentan proporcionalmente a lo largo de un eje diagonal ascendente, lo que indica un alto índice de correlación positiva entre ambas señales.



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

María F. Trovero, Rosana Rodríguez-Casuriaga, Carlos Romeo, Federico F. Santiñaque, Mateo François, Gustavo A. Folle, Ricardo Benavente, José R. Sotelo-Silveira & Adriana Geisinger

To cite this article: María F. Trovero, Rosana Rodríguez-Casuriaga, Carlos Romeo, Federico F. Santiñaque, Mateo François, Gustavo A. Folle, Ricardo Benavente, José R. Sotelo-Silveira & Adriana Geisinger (2019): Revealing stage-specific expression patterns of long noncoding RNAs along mouse spermatogenesis, RNA Biology, DOI: [10.1080/15476286.2019.1700332](https://doi.org/10.1080/15476286.2019.1700332)

To link to this article: <https://doi.org/10.1080/15476286.2019.1700332>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



View supplementary material [↗](#)



Published online: 23 Dec 2019.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

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.

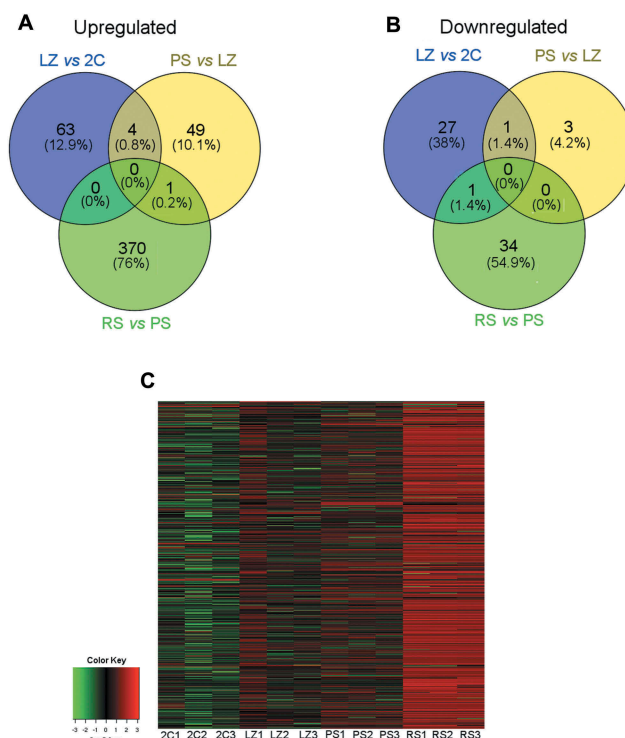
### Round spermatids contain the highest number of lncRNAs among different testicular cell populations

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.

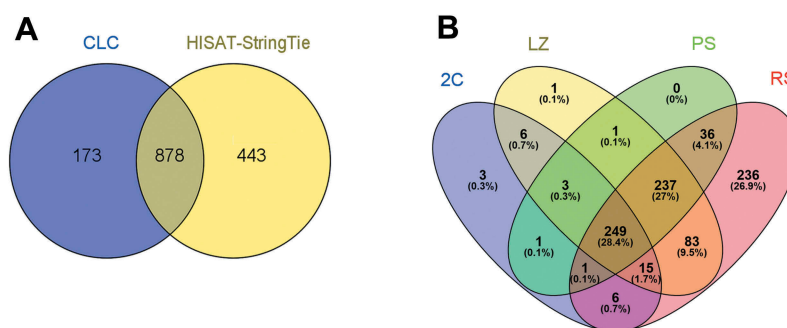
Among the DE lncRNAs, the number of upregulated genes (Fig. 2A) was notoriously higher than that of downregulated genes (Fig. 2B), for all the pairwise comparisons. Despite this, RS exhibited the highest number of upregulated and downregulated DE lncRNA genes as clearly depicted in the heat-maps, where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

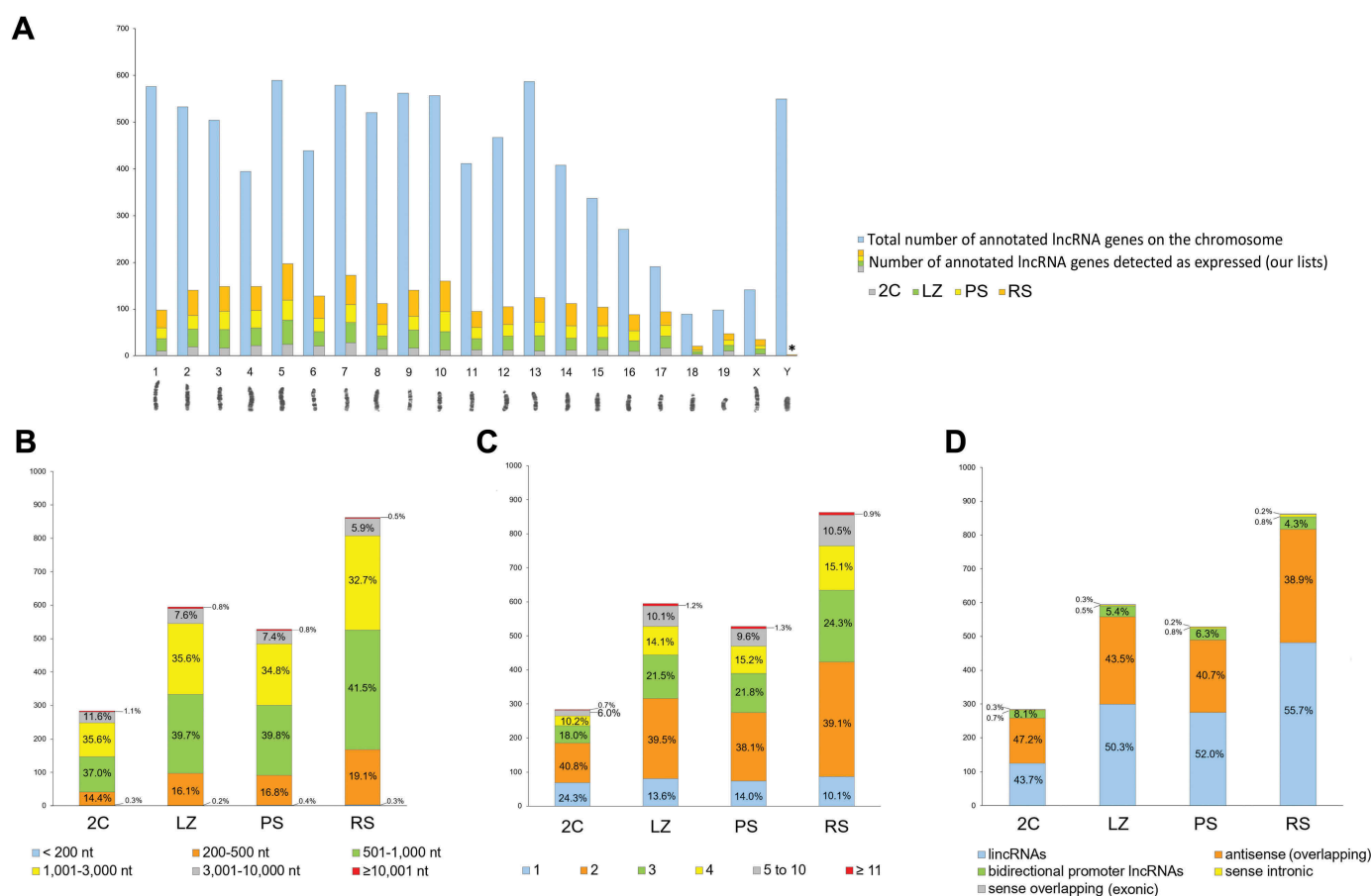
### Characterization of lncRNAs in the different testicular cell populations

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].

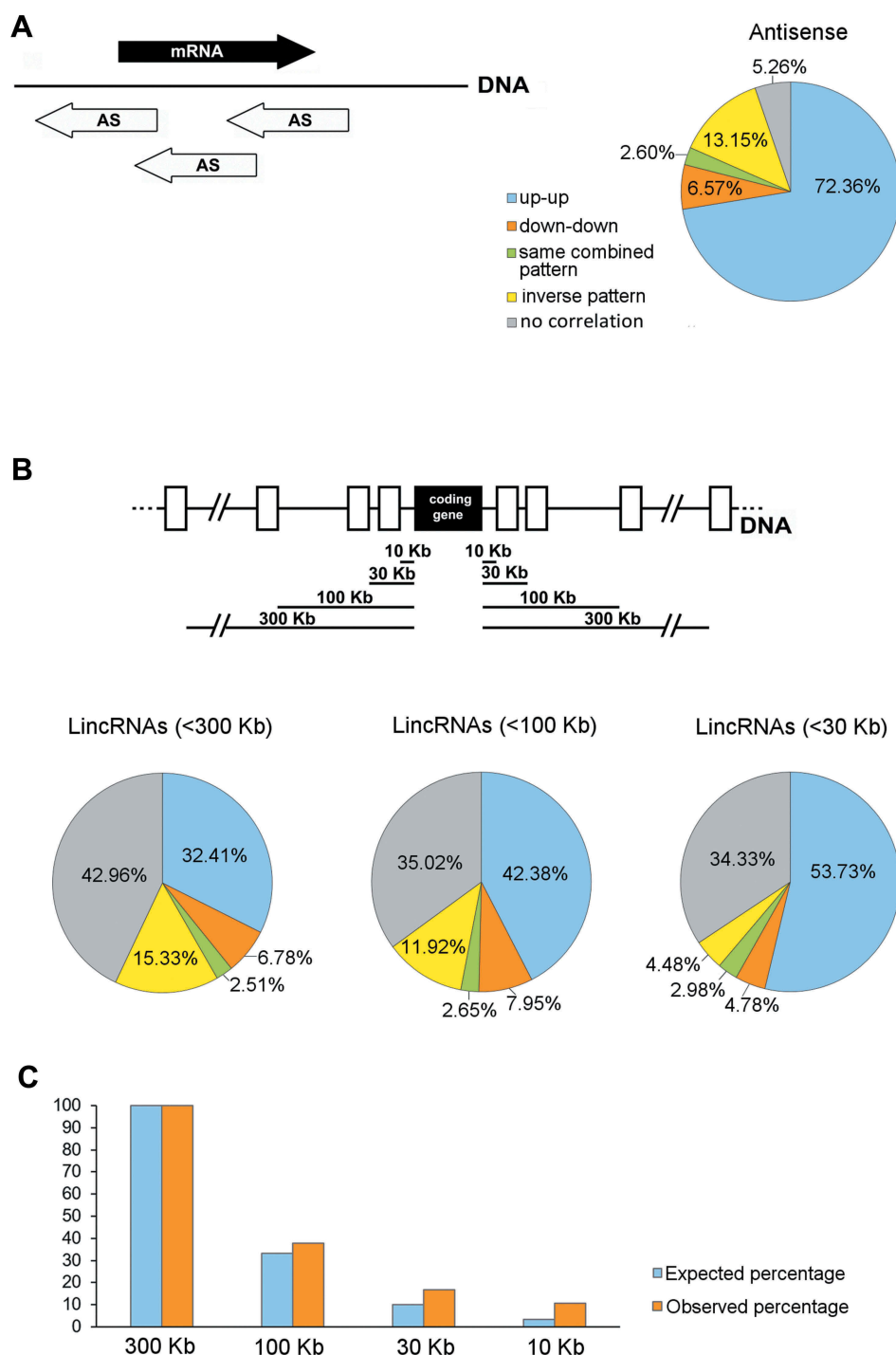
The high representation of AS lncRNAs in our lists, prompted us to analyse their co-expression with overlapping protein-coding genes. We found that for 85.5% of the expressed AS lncRNAs, the host protein-coding gene also appeared as expressed in our lists (Supplementary Table S4). Moreover, 81.5% of 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 lncRNA 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) lncRNAs and their host protein-coding genes. The schematic diagram shows different positions of overlapping AS lncRNAs (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 neighbour 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].

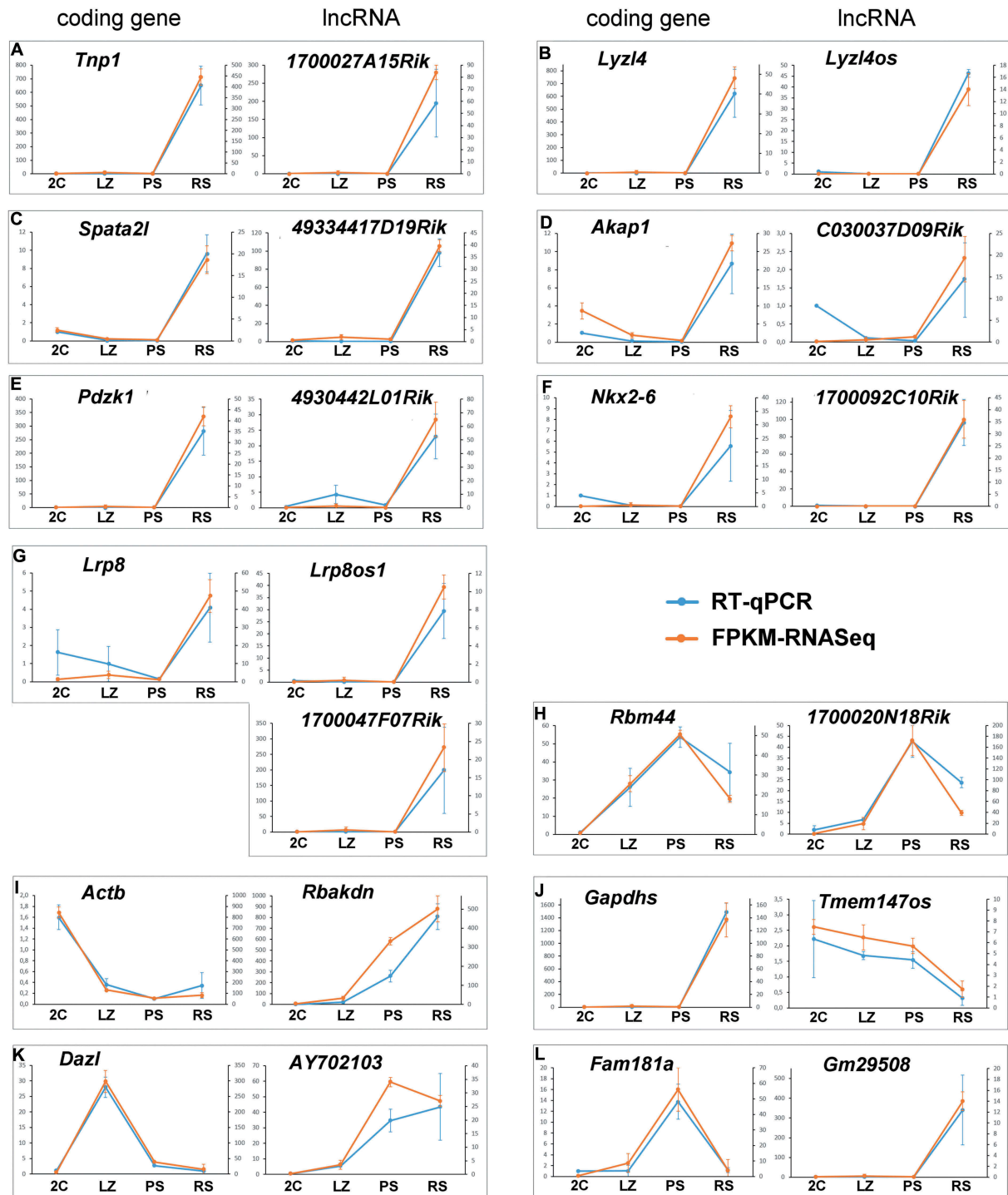
We then selected our co-expression lists of coding genes within <30 Kb from lincRNAs in mouse, and searched for the existence of homologous lincRNAs at syntenic positions in the human genome, as it has been suggested that the maintenance of the genomic position of lincRNAs relative to protein-coding 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]. LYSL4 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.

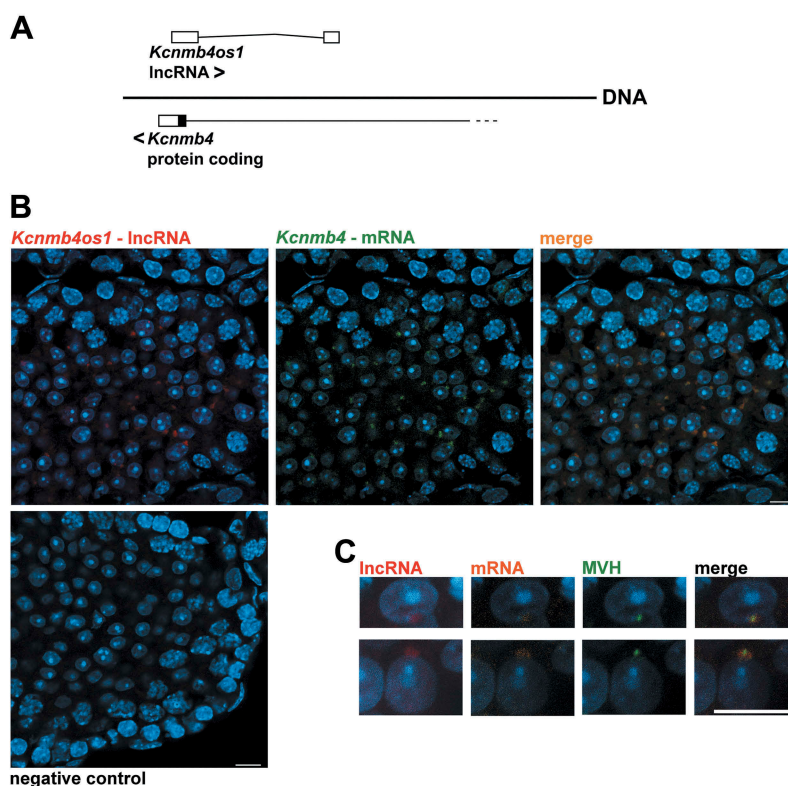
### An antisense lncRNA and its overlapping mRNA co-localize in the chromatoid body of round spermatids

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 FACSVantage 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 lncRNAs, 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 lncRNAs 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 lncRNAs.

### 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 lncRNAs). 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 lncRNAs 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 (Srps1)* is differentially expressed during testis development of the rat. *Gene Expr Patterns.* **2010**;10:1–8.