PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

FACULTAD DE CIENCIAS BIOLÓGICAS

PROGRAMA DE DOCTORADO EN CIENCIAS BIOLÓGICAS

MENCIÓN BIOLOGÍA CELULAR Y MOLECULAR

ESTUDIO DE LA BIOGÉNESIS Y EXPRESIÓN DE mirrorRNAs EN EL TRANSCRIPTOMA DE MAMÍFEROS

TESIS PRESENTADA A LA PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE COMO PARTE DE LOS REQUISITOS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS BIOLÓGICAS CON MENCIÓN EN BIOLOGÍA CELULAR Y MOLECULAR

POR

ROBERTO ANDRÉS MUNITA ROBERT

DIRECTOR DE TESIS: DRA. KATIA GYSLING

COMISIÓN DE TESIS: DR. OMAR ORELLANA

DR. XAVIER JORDANA

DR. FRANCISCO MELO

AGRADECIMIENTOS

En primer lugar quiero agradecer a mi tutora la Dra. Katia Gysling por todos estos años de apoyo y guía, por haber confiado en mí y permitirme desarrollar mi trabajo. Sólo puedo expresar palabras de agradecimiento y admiración, ya que no sólo es una gran tutora y científica, sino que además una gran persona.

A mis compañeros de laboratorio, ya que durante estos años hemos compartido muchas horas de trabajo, de discusión de ideas, de apoyo cuando las cosas no funcionan, en definitiva de amistad. Gracias a Elías, Vero, Raquel, Geo, Carla, Javier, Cristian, Hector, Jaime, Paula y Franini. Guille y Cledi, integrantes del "Team RNA", gracias por su amistad, cariño y paciencia, creo que haber trabajado con ustedes me enseñó muchas cosas y siempre será algo que recordaré con inmenso cariño.

A todo el laboratorio de Farmacología-Bioquímica y de manera muy especial a Don Hector, Sra. Lucy y Sra. Sole, gracias por su constante ayuda.

A las fuentes de financiamiento que permitieron que pudiera desarrollar esta tesis. A CONICYT, FONDECYT, Iniciativa Milenio y a la Facultad de Ciencias Biológicas.

A mis "amigos científicos": Pipe, Tefa, Pancha, Negro y Pancho por todos estos años de amistad. Ustedes que hicieron que este proceso fuera más feliz.

En especial quiero agradecer a mi mamá a mis hermanos y a toda mi familia por su cariño incondicional, gracias por su continuo apoyo a este camino que elegí, ya que sin ustedes esto no podría haberse logrado. Finalmente quiero agradecer a mi señora, "Santa" Cote, por toda su comprensión, paciencia, compañía y amor durante este largo camino que recorrimos juntos, esta tesis es para ti.

ÍNDICE DE MATERIAS

AGRADECIMIENTOS…… ii

ÍNDICE DE MATERIAS…… iii

ÍNDICE DE FIGURAS ……..…… vii

ÍNDICE DE TABLAS. ……..…… viii

ABREVIATURAS…… x

RESUMEN…… xii

ABSTRACT…… xiv

INTRODUCCIÓN…… 1

1. Visión proteo-céntrica de la expresión génica…… 1
2. Menos genes codificantes de proteínas que los esperados…… 2
3. Amplia transcripción del genoma de los eucariontes…… 3
4. Existen múltiples tipos de ncRNAs…… 5
5. Definición y características de los transcritos antisentido naturales…… 6
6. Amplia expresión de los NATs en el genoma de mamíferos…… 6
7. Mecanismos y funciones descritas para los NATs…… 7
   - 7.1. Interferencia Transcripcional…… 8
   - 7.2. Modificaciones de la cromatina…… 8
   - 7.3. *Imprinting* genómico…… 9
   - 7.4. Inactivación del cromosoma X…… 9
   - 7.5. Splicing alternativo…… 10
   - 7.6. *Editing* de RNA…… 10
   - 7.7. Cambios en la estabilidad del RNA…… 11
   - 7.8. Enmascaramiento de sitios de miRNAs…… 11
   - 7.9. Regulación de la traducción…… 12
   - 7.10. Formación de siRNAs endógenos …… 12
8. Clasificaciones de los NATs…… 12
9. mirrorRNAs…… 13
10. Controversia sobre la existencia de los mirrorRNAs…… 15
11. Posibles mecanismos de biogénesis de los mirrorRNAs…… 16
12. Splicing de intrones complementarios a los canónicos…… 18

13. Actividad RdRP en células de mamífero.....20
14. Transcripción antisentido de pseudogenes procesados.....23

MATERIALES..... 25

1. Material Biológico..... 25
   1.1. Bacterias..... 25
   1.2. Células Eucariontes..... 13
   1.3. Plásmidos..... 26
   1.4. RNA..... 26
2. Reactivos e insumos de biología molecular..... 26
3. Oligonucleótidos..... 27
4. Medios de Cultivo..... 31
   4.1. Cultivo bacteriano, medio LB..... 31
   4.2. Cultivo de células eucariontes..... 31
      4.2.1. Dulbecco Modified Eagle's Medium (DMEM base)..... 31
      4.2.2. Medio de crecimiento para las células HEK293, HEK293T, HeLa y COS7..... 31
      4.2.3. Medio de crecimiento para células PC12..... 32
5. Material Fungible..... 32

MÉTODOS..... 33

1. Análisis bioinformáticos..... 33
   1.1. Búsqueda de mirrorRNAs en ESTs y cDNAs humanos..... 33
      1.1.1. Búsqueda de intrones mirror..... 33
      1.1.2. Identificación de ESTs y cDNAs poliadenilados..... 33
      1.1.3. Identificación de genes con ESTs y cDNAs mirror..... 34
   1.2. Búsqueda de mirrorRNAs en datos de RNA-Seq..... 34
      1.2.1. Generación de una biblioteca de SJs..... 34
      1.2.2. Alineamiento de datos de RNA-Seq a la biblioteca de SJs, pseudogenes procesados y al genoma humano..... 34
   1.3. Análisis de *Gene Ontology*..... 35
   1.4. Búsqueda de siRNAs endógenos derivados de mirrorRNAs..... 35
   1.5. Búsqueda de *antisense termini-associated short RNAs* (aTASRs)..... 36
2. PCR..... 37
   2.1. Amplificación de DNA por PCR para clonamientos..... 37
   2.2. PCR cualitativo..... 37
3. Electroforesis de DNA..... 37
   3.1. Geles de agarosa..... 37
   3.2. Visualización de DNA..... 38
4. Purificación de fragmentos de DNA..... 38

5. Determinación de la concentración del DNA/RNA..... 38
6. Tratamientos enzimáticos del DNA..... 39
6.1. Digestiones con enzimas de restricción..... 39
6.2. Ligación de fragmentos de DNA con DNA ligasa del bacteriófago T4..... 39
7. Crecimiento y cultivo de *Escherichia coli*..... 39
8. Obtención de plásmidos recombinantes..... 40
8.1. Transformación de células competentes de *E. coli* DH5α..... 40
8.2. Crecimiento de las cepas bacterianas para amplificación de plásmido..... 40
8.3. Selección de clones positivos mediante PCR en cultivo..... 40
8.4. Purificación de plásmidos a pequeña escala (Miniprep)..... 41
8.5. Generación de un vector de expresión con el gen U83..... 41
9. Extracción de RNA desde células en cultivo..... 41
10. Remoción del DNA genómico..... 42
11. Síntesis de cDNA..... 42
12. Cultivo de células eucariontes..... 42
12.1. Mantención y condiciones de crecimiento de células eucariontes..... 42
12.2. Descongelamiento de líneas celulares..... 43
12.3. Congelamiento de líneas celulares..... 43
12.4. Transfección de células eucariontes..... 43
13. 3' RACE..... 44
14. Poliadenilación de RNA *in vitro*..... 44
15. 5’ RLM RACE..... 45
16. RPA-RT-PCR..... 45

RESULTADOS..... 46

1. Búsqueda y caracterización de mirrorRNAs en el transcriptoma humano..... 46
1.1.1. Búsqueda de mirrorRNAs en ESTs y cDNAs de humano 46
1.1.2. Señales de poliadenilación en mirrorRNAs..... 49
1.1.3. Los mirrorRNAs pueden provenir de pseudogenes procesados presentes en el genoma humano..... 49
1.2. Búsqueda de mirrorRNAs en datos de RNA-seq hebra específica de humano..... 54
1.2.1. Búsqueda de mirrorRNAs en datos de RNA-seq de IBM 2.0..... 54
1.2.2. Búsqueda de mirrorRNAs en datos de RNA-seq del proyecto ENCODE..... 61
1.3. Comparación de datos de cDNAs/ESTs y de RNA-Seq en la búsqueda de mirrorRNAs..... 63
1.4. Características de los mirrorRNAs..... 63
1.4.1. Análisis de *Gene Ontology*..... 63
1.4.2. mirrorRNAs y pseudogenes procesados..... 65
1.5. Búsqueda de RNAs pequeños derivados de mirrorRNAs..... 65
2. Validación experimental de los mirrorRNAs..... 69
2.1. 3’ RACE..... 69

2.1.1. Detección de pseudo-mirrorRNAs mediante 3' RACE..... 74
2.1.2. 3' RACE a RNA poliadenilado *in vitro*..... 74
2.2. 5' RACE..... 77
2.2.1. 5' RACE utilizando TdT y dCTP 77
2.2.2. 5'-RLM-RACE 77
2.3. Validar mirrorRNAs por ensayo de protección de ribonucleasas (RPA)..... 81
3. Biogénesis de los mirrorRNAs..... 83
3.1. Búsqueda de intrones CT-AC..... 83
3.1.1. Búsqueda de intrones mirror con intrones GT-AG 83
3.1.2. Búsqueda de intrones CT-AC en el transcriptoma de ratón y humano..... 87
3.1.3. Validación experimental de los intrones no canónicos CT-AC..... 94
3.1.4. Falseamiento de los intrones CT-AC descritos en la literatura..... 97
3.2. Actividad RdRP..... 103
3.3. mirrorRNAs derivados desde pseudogenes procesados..... 107

DISCUSIÓN..... 111
1. Predicción bioinformática de los mirrorRNAs..... 111
2. Validación experimental de los mirrorRNAs..... 114
3. Características de los mirrorRNAs..... 118
4. Mecanismo de biogénesis de los mirrorRNAs..... 119
4.1. Splicing de los mirrorRNAs en sitios CT-AC..... 120
4.2. Actividad RdRP en células humanas..... 122
4.3. Transcripción antisentido de pseudogenes procesados..... 124

CONCLUSIONES..... 128

BIBLIOGRAFÍA..... 129

INDICE DE FIGURAS

Figura 1: Representación esquemática de los tipos de NATs..... 14

Figura 2: Esquema de los mecanismos propuestos para explicar la biogénesis de los mirrorRNAs..... 17

Figura 3: Estrategia utilizada para buscar mirrorRNAs en datos de ESTs y cDNAs de humano..... 47

Figura 4: Ejemplos de ESTs y cDNAs mirror alineados al genoma humano..... 52

Figura 5: Identificación de un NAT que proviene de un pseudogen procesado del gen DGKZ..... 53

Figura 6: Estrategia utilizada para buscar mirrorRNAs en datos de RNA-Seq hebra específica..... 56

Figura 7: Diagrama de Venn que muestra el número de genes que presentan evidencia de poseer mirrorRNAs o pseudo-mirrorRNAs en datos de RNA-Seq de IBM 2.0 generados con RNA poli(A) + y RNA total..... 59

Figura 8: Ejemplos de mirrorRNAs y pseudo-mirrorRNAs que poseen lecturas de RNA-Seq antisentido a varios SJs..... 60

Figura 9: Diagramas de Venn con la comparación de los mirrorRNAs y pseudo-mirrorRNAs detectados en los datos de cDNAS/ESTs, IBM 2.0 y ENCODE..... 64

Figura 10: Ejemplos de endo-siRNAs identificados utilizando datos de RNA-Seq de RNAs pequeños del proyecto ENCODE..... 68

Figura 11: Esquema del protocolo del 3' RACE..... 70

Figura 12: Los mirrorRNAs de CWC15 y RPS23 fueron validados mediante 3' RACE..... 72

Figura 13: Mediante 3' RACE se identificaron mirrorRNAs parcialmente complementarios a mRNAs sentido para los genes CALR y DDX5..... 73

Figura 14: Mediante 3' RACE se validó la existencia de un mirrorRNA y un pseudo-mirrorRNA de KRT8..... 75

Figura 15: El mirrorRNA de UBR5 en RNA Universal de Humano solo se pudo amplificar mediante 3' RACE al poliadenilar in vitro el RNA..... 76

Figura 16: Esquema que muestra los dos protocolos de 5' RACE utilizados..... 78

Figura 17: Validación experimental de los mirrorRNAs de los genes UBR5 y KRT8 mediante 5' RACE clásico..... 79

Figura 18: Validación experimental de los mirrorRNAs de los genes EIF3D, DDX5, CWC15 y CALR mediante 5'-RLM RACE…….. 80

Figura 19: Validación de mirrorRNAs mediante RPA-RT-PCR…….. 82

Figura 20: Ejemplos de cDNAs con intrones mirror e intrones GT-AG…….. 86

Figura 21: Secuencias directas repetidas en los SJs dificulta el correcto alineamiento de ESTs y cDNAs al genoma…….. 90

Figura 22: Los intrones CT-AC predichos en genes de ratón y sujetos a verificación experimental no pudieron ser validados …….. 96

Figura 23: El splicing de un intrón CT-AC del gen U83 de HHV-6 es un artefacto…….. 100

Figura 24: La maquinaria de splicing celular no es capaz de reconocer y procesar las secuencias complementarias y reversas a los sitios de splicing del gen de Ucn de rata …….. 102

Figura 25: Ejemplo de un mirrorRNA con cola 3' poli(A) y extremo 5' poli(T)…….. 105

Figura 26: Identificación de mirrorRNAs en genes humanos que poseen pseudogenes procesados polimórficos …….. 110

INDICE DE TABLAS

Tabla I: Oligonucleótidos utilizados en la presente tesis 28

Tabla II: Búsqueda de mirrorRNAs en datos de cDNAs y de ESTs de humano..... 48

Tabla III: Genes humanos que poseen evidencia de poseer mirrorRNAs a partir de datos de ESTs y cDNAs..... 50

Tabla IV: Búsqueda de mirrorRNAs en datos de RNA-seq de IBM 2.0..... 55

Tabla V: Búsqueda de mirrorRNAs en datos de RNA-seq de ENCODE..... 62

Tabla VI: Búsqueda de siRNAs derivados de mirrorRNAs..... 67

Tabla VII: Secuencias con intrones mirror e intrones GT-AG..... 85

Tabla VIII: Dinucleótidos de los intrones predichos por BLAT a partir de datos de ESTs y cDNAs de humano y ratón..... 89

Tabla IX: Dinucleótidos de los intrones predichos por GMAP a partir de datos de ESTs y cDNAs de humano y ratón..... 91

Tabla X: Dinucleótidos de los intrones predichos por BLAT y GMAP a partir de datos de ESTs y cDNAs de humano y ratón..... 93

Tabla XI: Intrones CT-AC elegidos para ser validados experimentalmente mediante RT-PCR..... 95

Tabla XII: Secuencias de mirrorRNAs con extremos 3' poli(A) y 5' poli(T)..... 104

Tabla XIII: Búsqueda de RNAs con extremo 5' poli(T) no genómico en datos de RNA-Seq de IBM 2.0 106

Tabla XIV: Búsqueda de RNAs con extremo 5' poli(T) no genómico en datos de RNA-Seq de RNAs pequeños de ENCODE..... 108

ABREVIATURAS

3’ RACE	: *3’ Rapid Amplification of cDNA Ends*
5’ RACE	: *5’ Rapid Amplification of cDNA Ends*
5’ RLM-RACE	: *5’ RNA Ligase Mediated Rapid Amplification of cDNA Ends*
µg	: Microgramo
µL	: Microlitro
ActD	: Actinomicina D
aTASR	: *Antisense termini-associated short RNAs*
CAGE	: *cap-analysis of gene expression*
cDNA	: *Complementary DNA*
CIP	: *Calf Intestinal Phosphatase*
circRNAs	: Circular RNAs
Cols.	: Colaboradores
crasiRNAs	: *Centromere repeats-associated short interacting RNAs*
DNA	: *Deoxyribonucleic acid*
DdDP	: *DNA Dependent DNA Polymerase*
DMSO	: Dimetilsulfoxido
dCTP	: Desoxicitidina trifosfato
dNTP	: Desoxirribonucleótido trifosfato
dsRNA	: *double stranded RNA*
EDTA	: *Ethylenediaminetetraacetic acid*
ENCODE	: *The Encyclopedia of DNA Elements*
Endo-siRNAs	: *Endogenous small interfering RNAs*
eRNAs	: Enhancer RNAs
EST	: *Expressed sequence tag*
FANTOM	: *Functional Annotation Of Mammalian genome*
Fig.	: Figura
g	: Gramo
Gb	: Gigabase
GO	: *Gene ontology*
GSP	: *Gene specific primer*
HDV	: *Hepatitis delta virus*
HHV-6	: *Human herpes virus 6*
IBM 2.0	: *Illumina Body Map 2.0*
Kb	: Kilobase
L	: Litro
LB	: Medio Luria-Bertani
lncRNA	: *Long non-coding RNA*
lincRNA	: *Long intergenic non-coding RNAs*
LINE	: *Long interspersed elements*
Min	: Minuto
miRNA	: *microRNA*
mL	: Mililitro

mM	: Milimolar
mRNA	: *messenger RNA*
NAT	: *Natural antisense transcript*
ncRNA	: *non-coding RNA*
nt	: Nucleótido
ng	: Nanogramo
PASRs	: *promoter-associated sRNAs*
PALRs	: *promoter-associated lRNAs*
pb	: Pares de bases
PBS	: *Phosphate-Buffered Saline*
PCR	: *Polymerase chain reaction*
PET	: *paired end tags*
piRNA	: *Piwi-interacting RNA*
pmol	: Picomol
PRC2	: *Polycomb repressive complex 2*
PROMTs	: *Promoter upstream transcripts*
qPCR	: *Quantitative polymerase chain reaction*
RdRP	: *RNA-dependent RNA polymerase*
RefSeq	: *The Reference Sequence Database*
RMRP	: *RNA component of mitochondrial RNA processing endoribonuclease*
RNA	: *Ribonucleic acid*
RNasa	: Ribonucleasa
RNA-Seq	: Secuenciación masiva de cDNA
RPA	: *RNase Protection Assay*
rpm	: Revoluciones por minuto
RT	: *Reverse transcription*
sdRNAs	: *sno-derived RNAs*
seg	: Segundo
siRNAs	: *Small interfering RNAs*
SJ	: *Splice Junction*
spliRNAs	: *splice-site RNAs*
T-UCRs	: *Transcribed-Ultraconserved Regions*
TAP	: *Tobacco Acid Pyrophosphatase*
TASRs	: *Termini-associated sRNAs*
TdT	: *Terminal transferase*
TERRA	: *Telomeric repeat-containing RNA*
tiRNAs	: *Transcription initiation RNAs*
tRFs	: *tRNA-derived RNA fragments*
TSSa-RNAs	: *Transcription start site-associated RNAs*
UTR	: *Untranslated region*

RESUMEN

Con el advenimiento de nuevas tecnologías se ha descubierto la gran complejidad del transcriptoma de los mamíferos. Se ha reportado la existencia de un tipo especial de transcritos antisentido naturales (NATs), que son perfectamente complementarios a mRNAs sentido a lo largo de varios exones, incluyendo los sitios de unión entre exones. En esta tesis estos transcritos antisentido naturales son llamados mirrorRNAs. Hasta el momento aun existe controversia si los mirrorRNAs son RNAs no codificantes reales o artefactos experimentales. No se han hecho análisis sistemáticos que permitan conocer la abundancia y diversidad de los mirrorRNAs en el transcriptoma humano y tampoco se ha estudiado el mecanismo mediante el cual los mirrorRNAs son generados. En la literatura se han propuesto dos hipótesis para explicar su biogénesis. La primera hipótesis es que estos transcritos son generados por la transcripción bidireccional de un *locus* y que posteriormente los transcritos antisentido sufren splicing en los sitios no consenso CT-AC, que son los sitios complementarios a los sitios consenso GT-AG. La segunda hipótesis que se ha planteado es que los mirrorRNAs se producen por una actividad RNA polimerasa dependiente de RNA (RdRP) que utiliza como molde un mRNA maduro, generando un transcrito antisentido perfectamente complementario. En esta tesis planteamos una tercera hipótesis: Los mirrorRNAs provienen de la transcripción antisentido de pseudogenes procesados presentes en el genoma.

En la presente tesis se realizó un análisis sistemático de la existencia de mirrorRNAs en el transcriptoma humano, utilizando datos de ESTs, cDNAs y de RNA-Seq hebra específica. Además, se validaron experimentalmente mirrorRNAs predichos utilizando 3’ RACE, 5’ RACE y RPA-RT-PCR. Mediante estas aproximaciones se determinó que existen cientos de genes con evidencia bioinformática de que poseen mirrorRNAs. Adicionalmente, se determinó que los mirrorRNAs presentan bajos niveles de expresión y que parte de ellos no estarían poliadenilados. Dentro de la lista de genes predichos para poseer un mirrorRNA se identificó un enriquecimiento de genes que participan en traducción de proteínas y de genes que poseen pseudogenes procesados. Luego se analizó si los mirrorRNAs podrían generar endo-siRNAs con sus contrapartes sentido. Utilizando datos de RNA-Seq generados por ENCODE, no se identificaron endo-siRNAs derivados de mirrorRNAs. En la búsqueda de mirrorRNAs se identificaron múltiples NATs provenientes de la transcripción antisentido de pseudogenes procesados.

En relación a la biogénesis de los mirrorRNAs se estudió si la maquinaria de splicing celular puede procesar intrones en los sitios complementarios a los consenso. Utilizando datos de ESTs y cDNAs de ratón y humano se realizó una búsqueda sistemática de posibles intrones CT-AC. Posteriormente se realizaron validaciones experimentales de múltiples intrones candidatos. No se encontró ningún intrón que se procesara en sitios con secuencias complementarias a las consenso. Se encontró que múltiples artefactos experimentales y de alineamiento pueden llevar a la falsa identificación de intrones no canónicos. Adicionalmente se llevaron a cabo experimentos de splicing *in vivo* utilizando minigenes de transcritos que podrían sufrir splicing en sitios CT-AC. Nuestros análisis muestran que la maquinaria de splicing celular no es capaz de reconocer las secuencias complementarias a las consenso.

Se buscó la existencia de actividad RdRP en células humanas usando datos de RNA-Seq, sin resultados positivos. A pesar de lo anterior, no es posible descartar del todo que la actividad RdRP exista y sea la responsable de la existencia de algún mirrorRNA.

En diversos trabajos recientes se ha descrito la existencia de pseudogenes procesados polimórficos en la población humana. Múltiples mirrorRNAs pueden ser explicados por la transcripción antisentido de estos pseudogenes procesados polimórficos. Adicionalmente se abre la posibilidad que los mirrorRNAs sean la evidencia transcripcional de eventos de retrotransposición que no estén anotados en el genoma de referencia.

En conclusión los resultados de esta tesis demuestran la existencia de mirrorRNAs en células humanas. Además nuestros resultados respaldan que la maquinaria de splicing celular no es capaz de procesar intrones en los sitios complementarios a los consenso. Finalmente nuestros datos sugieren que los mirrorRNAs son generados desde pseudogenes procesados polimórficos.

ABSTRACT

The advent of new technologies, has uncovered a high degree of complexity in the mammalian transcriptome. The existence of a special type of natural antisense transcripts (NATs), which are perfectly complementary to mature mRNAs over several exons, including splice sites, has been reported. In this thesis these natural antisense transcripts are called mirrorRNAs. Until now, there is still controversy whether mirrorRNAs are real non coding RNAs or experimental artifacts. No systematic analysis has been done to assess the abundance and diversity of mirrorRNAs in the human transcriptome. It has not been studied the mechanism by which the mirrorRNAs are generated. Two hypotheses have been proposed in the literature to explain their biogenesis. The first hypothesis is that these transcripts are generated by the bidirectional transcription of a locus and subsequent antisense transcripts undergo splicing in no consensus CT-AC sites that are complementary to GT-AG consensus sites. The second hypothesis that has been raised is that mirrorRNAs are produced by an RNA-dependent RNA polymerase activity (RdRP) using as a template a mature mRNA, generating an antisense transcript perfectly complementary. In this thesis we propose a third hypothesis: mirrorRNAs come from antisense transcription of processed pseudogenes.

In this thesis is presented a systematic analysis of the existence of mirrorRNAs in the human transcriptome using data from, ESTs, cDNAs and strand specific RNA-Seq. Furthermore, predicted mirrorRNAs were experimentally validated using 3' RACE, 5' RACE and RPA-RT-PCR. Using these approaches we determined that there are hundreds of genes with bioinformatic evidence of possessing mirrorRNAs. Additionally, it was determined that mirrorRNAs have low levels of expression and that some of them would not be polyadenylated. Within the list of genes predicted to possess a mirrorRNA it was identified an enrichment of genes involved in protein translation and genes that have processed pseudogenes. Thereafter, it was analyzed whether mirrorRNAs could generate endo-siRNAs with their sense counterparts. Using RNA-Seq data generated by ENCODE, no endo-siRNAs derivatives from mirrorRNAs were identified. In the search for mirrorRNAs, multiple NATs originated from antisense transcription of processed pseudogenes were identified.

Regarding mirrorRNAs biogenesis, we studied whether the cellular splicing machinery can process introns in sites complementary to the consensus. Using data from ESTs and cDNAs of mouse and human a systematic search of possible CT-AC introns was performed. Subsequently, experimental validations were performed to multiple candidate introns. It was not found introns processed in sites with sequence complementary to the consensus. It was observed that several experimental and alignment artifacts can lead to falsely identifying non canonical introns. Additionally, *in vivo* splicing assays were conducted using minigenes of transcripts that could undergo spliced in CT-AC sites. Our *in vivo* splicing assays show that the cellular splicing machinery is not able to recognize the sequences complementary to the consensus.

The existence of RdRP activity in human cells was explored using RNA-Seq data without positive results. Despite this, it is not possible to exclude that RdRP activity could exist and could be responsible for the existence of some mirrorRNA.

In several recent papers, the existence of polymorphic processed pseudogenes in the human population has been described. Multiple mirrorRNAs could be explained by the antisense transcription of these polymorphic processed pseudogenes. Further, this opens the possibility that mirrorRNAs are transcriptional evidence of retrotransposition events that are not listed in the reference genome.

In conclusion, the results of this thesis demonstrate the existence of mirrorRNAs in human cells. Moreover, our results support that cellular splicing machinery is not able to process the introns in sites complementary to the consensus. Finally, our data suggest that mirrorRNAs are generated from polymorphic processed pseudogenes.

INTRODUCCIÓN

1. Visión proteo-céntrica de la expresión génica

Desde el descubrimiento de la estructura del DNA en 1953, el problema central de la biología molecular ha sido entender como está codificada la información genética y cómo se regula el flujo de la información almacenada en el DNA (Sharp, 2009).

Gracias a los estudios de Jacob y Monod en el operón Lac de *E. coli* (Jacob y Monod, 1961), se propuso a inicio de los años sesenta que el flujo de información genética estaba principalmente regulado a nivel de la transcripción del DNA. Este control era efectuado por proteínas llamadas factores de transcripción los que inhibían o promovían el acceso de la RNA polimerasa a la región promotora del gen, desde donde se iniciaba la transcripción. El transcrito generado se traduciría generando una proteína, que sería el efector final del flujo de la información genética. Este modelo se generalizó también para entender los procesos de expresión génica en eucariontes y se enraizó en la comunidad científica por décadas (Sharp, 2009). Pero gradualmente esta forma de entender el proceso de expresión génica en eucariontes cambió y actualmente se entiende la importancia de los procesos de regulación de la expresión en otros niveles tales como: estructura de la cromatina, splicing, transporte del mRNA entre el núcleo y el citoplasma, regulación post transcripcional mediada por RNAs pequeños, regulación de la eficiencia de la traducción, entre otros.

Durante mucho tiempo la visión proteo-céntrica de la expresión génica relegó al RNA a ser sólo una molécula intermediaria entre el DNA y las proteínas (Mattick, 2009). Actualmente sabemos que existe una gran diversidad de RNAs no codificantes de proteínas (ncRNAs) que constituyen una

compleja capa de regulación de la expresión génica (Amaral et al., 2008; Mattick, 2001; Mattick, 2003; Mattick, 2009).

2. Menos genes codificantes de proteínas que los esperados

La secuenciación del genoma humano y de organismos modelo trajo grandes sorpresas (Consortium, 1998; Lander et al., 2001). En primer lugar se descubrió que menos de un 2% del genoma humano codifica para proteínas (International Human Genome Sequencing, 2004). En segundo lugar se constató que el número de genes codificantes de proteínas era menor del esperado, con un estimado de entre 20.000 y 25.000 (International Human Genome Sequencing, 2004). Los primeros resultados del proyecto ENCODE determinaron que el genoma humano posee 20.687 genes codificantes de proteínas (Consortium et al., 2012).

Es interesante tomar en cuenta que la planta *Arabidopsis thaliana* posee en su genoma aproximadamente 27.400 genes codificantes de proteínas (Lamesch et al., 2012) y que el gusano modelo *C. elegans* posee aproximadamente 20.500 (Yook et al., 2012).

Un *C. elegans* mide aproximadamente 1 mm, posee en total 959 células somáticas y 302 neuronas en el hermafrodita adulto (Kimble y Hirsh, 1979; Sulston y Horvitz, 1977; Sulston et al., 1983; White et al., 1986). En cambio, el cerebro humano posee aproximadamente 86.000 millones de neuronas (Azevedo et al., 2009). Definir la complejidad de un organismo es una tarea sumamente difícil y que va más allá de los objetivos de esta tesis, pero es evidente que un ser humano posee una complejidad estructural y en el desarrollo bastante mayor que la de un *C. elegans*. Sin embargo, el número de genes codificantes de proteínas es prácticamente el mismo, por tanto en esta comparación no es posible explicar el aumento de complejidad por un aumento de genes codificantes de proteínas. A esta aparente falta de correlación entre genes codificantes de proteínas y la

complejidad de los organismos eucariontes se le ha denominado "*G value paradox*" (Hahn y Wray, 2002).

Es interesante además notar que los tamaños de los genomas son muy distintos; 100 Mb para *C elegans* y 3.2 Gb para humano (Flicek et al., 2013). Esta gran diferencia en el tamaño de los genomas se debe a que el genoma humano tiene un alto porcentaje de DNA no codificante (Taft et al., 2007).

3. Amplia transcripción del genoma de los eucariontes

Gracias al desarrollo de iniciativas como el proyecto FANTOM y ENCODE, y al advenimiento de nuevas tecnologías como *Tiling arrays*, CAGE (*cap-analysis of gene expression*) y RNA-seq, se determinó con sorpresa que la mayor parte del genoma de los eucariontes es transcrito (Carninci et al., 2005; Carninci et al., 2006; Consortium et al., 2012; Consortium et al., 2007; Cheng et al., 2005; Kapranov et al., 2010b; Manak et al., 2006; Wilhelm et al., 2008).

En el año 2005 el proyecto FANTOM, utilizando tecnologías convencionales de secuenciación de cDNAs completos, mostró que se transcribe al menos el 63% del genoma de ratón (Carninci et al., 2005). El proyecto FANTOM utilizó muestras de distintos tejidos de ratón y de distintos momentos en el desarrollo. En total se generaron 102.801 cDNAs completos y se identificaron los extremos 5' y 3' de 181.047 transcritos distintos. De los 102.281 cDNAs generados, 34.030 eran lncRNAs (Carninci et al., 2005).

El desarrollo de nuevas tecnologías de secuenciación masiva permitió interrogar el transcriptoma de los eucariontes con una profundidad sin precedentes (Wang et al., 2009). Al analizar el transcriptoma de *Schizosaccharomyces pombe* utilizando RNA-Seq, bajo múltiples condiciones, se

determinó que se transcribe más del 90% de su genoma (Wilhelm et al., 2008). El uso de nuevas tecnologías de secuenciación para el estudio del transcriptoma de *Drosophila melanogaster* (Graveley et al., 2011) y de *C. elegans* (Gerstein et al., 2010; Nam y Bartel, 2012) también reveló la inmensa complejidad de su transcriptoma y la existencia de una gran cantidad de ncRNAs.

La caracterización más completa y profunda que se ha realizado del transcriptoma humano es la que está llevando a cabo el proyecto ENCODE (Consortium et al., 2012). Para caracterizar el transcriptoma humano el proyecto ENCODE utilizó RNA proveniente de 15 líneas celulares (el proyecto no se ha detenido y este número ha ido en aumento). Realizaron fraccionamiento subcelular para obtener RNA nuclear, citoplasmático o de toda la célula. Se separó el RNA de cada muestra en RNAs largos (>200 nt) y RNAs pequeños (<200 nt). Adicionalmente se separaron los RNAs largos en transcritos poliadenilados o no poliadenilados. Estas muestras de RNA fueron analizadas utilizando RNA-seq, CAGE y PET (*paired end tags*) (Djebali et al., 2012).

Una de las principales conclusiones que se ha obtenido a partir del proyecto ENCODE es que al menos el 75% del genoma humano se transcribe (Djebali et al., 2012). Con los datos del proyecto ENCODE se detectaron miles de ncRNAs que no habían sido identificados previamente (Derrien et al., 2012; Djebali et al., 2012; Harrow et al., 2012).

Junto a ENCODE nació un consorcio denominado GENCODE cuyo objetivo es anotar con gran exactitud todas las características, basadas en evidencia, de los genes del genoma humano (Harrow et al., 2012). Según la versión 19 de GENCODE existen 20.345 genes codificantes de proteínas, 13.870 ncRNAs largos y 9.013 ncRNAs pequeños.

4. Existen múltiples tipos de ncRNAs

A principio de los años 90' se reportaron los primeros lincRNAs (*Long intergenic non-coding RNAs*) en mamíferos: H19 (Brannan et al., 1990) y XIST (Brockdorff et al., 1992; Brown et al., 1992). Posteriormente en 1993 el grupo de Victor Ambros describió el primer miRNA (microRNA) en *C. elegans*, lin-4, que es necesario para el correcto desarrollo de las larvas (Lee et al., 1993). En 1999, finalmente se describen los siRNAs (small interfering RNAs) en plantas, que permiten explicar el fenómeno de silenciamiento (Hamilton y Baulcombe, 1999). Pero los lincRNAs, miRNAs y siRNAs no son los únicos tipos de ncRNAs identificados. En los últimos años, principalmente gracias a la aparición de nuevas tecnologías, se ha descubierto una gran cantidad de nuevos tipos de ncRNAs en animales, tales como: endo-siRNAs (Tam et al., 2008; Watanabe et al., 2008), piRNAs (Klattenhoff y Theurkauf, 2008) , crasiRNAs (Carone et al., 2009), spliRNAs (Taft et al., 2010), TASRs (Kapranov et al., 2007), PASRs (Kapranov et al., 2007), PALRs (Kapranov et al., 2007), tiRNAs (Taft et al., 2009a), PROMTs (Preker et al., 2008), TSSa-RNA (Seila et al., 2008), tRFs (Lee et al., 2009), eRNA (Kim et al., 2010), enhancer-ncRNA (Orom et al., 2010), TERRA (Azzalin et al., 2007), sdRNAs (Taft et al., 2009b), circRNAs (Memczak et al., 2013), T-UCRs (Mestdagh et al., 2010), transcritos derivados de pseudogenes (Kalyana-Sundaram et al., 2012) y NATs (Faghihi y Wahlestedt, 2009).

Para muchos de los tipos de ncRNAs identificados existen funciones claramente descritas, pero para otros ncRNAs aún está por verse si tienen funciones regulatorias que aún desconocemos o son sólo productos secundarios de otros procesos celulares sin una función específica.

5. Definición y características de los transcritos antisentido naturales

En esta tesis nos hemos interesado en un tipo particular de RNAs denominados transcritos antisentido naturales (NATs). Los NATs son moléculas de RNAs endógenos que contienen secuencias complementarias a otros transcritos (Lapidot y Pilpel, 2006). Los RNAs sentido y los NATs pueden ser transcritos que codifican proteínas o ser ambos ncRNAs, pero lo más frecuente es que los RNAs sentido sean codificantes de proteínas y los NATs sean ncRNAs (Katayama et al., 2005).

Los NATs poseen diversos orígenes, pueden provenir de promotores independientes, de promotores bidireccionales que originan dos transcritos divergentes (Core et al., 2008; Seila et al., 2008; Trinklein et al., 2004) o desde promotores crípticos que están dentro del gen sentido (Kim et al., 2012; Whitehouse et al., 2007).

Los NATs generalmente se expresan en baja abundancia, siendo su expresión en promedio aproximadamente 10 veces menor que la de los RNAs sentido (He et al., 2008; Ozsolak et al., 2010). Además los NATs en promedio sufren menos eventos de splicing que los transcritos sentido (He et al., 2008). A diferencia de los mRNAs que codifican para proteínas, que se acumulan en el citoplasma, los NATs preferentemente se acumulan en el núcleo (Derrien et al., 2012).

6. Amplia expresión de los NATs en el genoma de mamíferos

El análisis del transcriptoma de mamíferos ha revelado la existencia de un gran número de NATs. Se han utilizado diversas estrategias experimentales para identificar NATs, tales como secuenciación a gran escala de clones de cDNAs completos (Katayama et al., 2005), análisis de bases de datos de ESTs (Chen et al., 2004; Yelin et al., 2003; Zhang et al., 2006), *tiling arrays* (Cheng et al., 2005), *microarrays* hebra específica (Ge et al., 2008), técnicas de hibridación, digestión y clonamiento (Rosok y Sioud, 2004), SAGE (*serial analysis of gene expression*) (Ge et al., 2006), ASSAGE (*asymmetric*

strand-specific analysis of gene expression) (He et al., 2008) y RNA-Seq hebra específica (Djebali et al., 2012).

Los primeros análisis del transcriptoma de ratón y humano que utilizaron ESTs, mostraron que sobre el 20% de los genes de ratón y humano posee un NAT (Chen et al., 2004; Yelin et al., 2003; Zhang et al., 2006). Los resultados del proyecto FANTOM3, que corresponden al mayor esfuerzo realizado para la secuenciación de cDNAs completos, muestran que para el 72% de los transcritos descritos en ratón existe evidencia de transcripción en la hebra complementaria (Katayama et al., 2005). Según la versión 19 de GENCODE (proyecto de anotación del genoma humano derivado de la iniciativa ENCODE) existen 9710 NAT en el transcriptoma humano. Este número está lejos de ser definitivo y crece constantemente (Harrow et al., 2012).

7. Mecanismos de acción y funciones descritas para los NATs

Se ha descrito que los NATs producen diversos efectos sobre la expresión génica lo que indica que sería un grupo heterogéneo de ncRNAs que podrían funcionar mediante diversos mecanismos (Faghihi y Wahlestedt, 2009). Los mecanismos generales mediante los cuales funcionan los NATs son: El enmascaramiento de señales en el RNA sentido, la formación de RNAs doble hebra, la interferencia transcripcional y el reclutamiento de factores remodeladores de la cromatina (Lapidot y Pilpel, 2006; Lavorgna et al., 2004). A continuación se listan procesos y funciones en donde se ha descrito la participación de NATs.

7.1 Interferencia Transcripcional

El proceso de transcripción de un NAT puede generar una disminución de la transcripción del RNA sentido ubicado en *cis*. En este caso no es el NAT el que cumple la función, sino que el fenómeno de la transcripción en sí. Esto puede ocurrir por un "choque" entre los complejos de las RNA polimerasa II (RNA pol II) (Osato et al., 2007; Petruk et al., 2006; Prescott y Proudfoot, 2002). La interferencia transcripcional también se podría producir por un desplazamiento de complejos de pre iniciación o factores de transcripción y por una prolongada oclusión del promotor sentido por una RNA pol II detenida (Palmer et al., 2011). La interferencia transcripcional no pareciera ser el mecanismo predominante mediante el cual los NATs regulan la expresión génica. La transcripción del sentido y del NAT pueden ocurrir a tiempos distintos o al mismo tiempo pero en diferentes cromosomas (Faghihi y Wahlestedt, 2009). La transcripción alelo específica podría explicar porqué el cromosoma X muestra un menor grado de transcripción antisentido que otros cromosomas (Katayama et al., 2005).

7.2 Modificaciones de la cromatina

Se ha descrito que los NATs pueden regular la metilación del DNA y modificaciones post traduccionales de las histonas (Hawkins y Morris, 2010; Modarresi et al., 2012; Morris et al., 2008; Rinn et al., 2007; Tufarelli et al., 2003; Yu et al., 2008). Es interesante que muchas enzimas modificadoras de la cromatina no tienen dominios de unión a DNA, pero sí tienen dominios de unión a RNA (Bernstein y Allis, 2005). Lo que se ha planteado es que los NATs (y otros lncRNAs) pueden unir complejos proteicos reguladores de la cromatina y dirigirlos a un locus específico (Magistri et al., 2012). Experimentos de inmunoprecipitación de RNA y secuenciación (RIP-Seq) dirigidos a la proteína Ezh2, reveló que el complejo PRC2 (*Polycomb repressive complex 2*) se asocia con al menos 10.000 RNAs en células madre embrionarias de ratón y aproximadamente 3.000 de estos RNAs serían NATs

(Zhao et al., 2010). Estos resultados sugieren que los NATs cumplen un importante papel como reguladores epigenéticos del proceso de expresión génica y remodelamiento de la cromatina.

7.3 *Imprinting* genómico

Existen ciertos genes en los cuales de forma específica se expresa sólo un alelo, el paterno o el materno. Esto se produce por marcas epigenéticas tales como modificación de las histonas o metilación del DNA. La mayoría de los genes de mamífero que presentan *imprinting* están agrupados en *clusters* (Verona et al., 2003) y la presencia de NATs es frecuente en esos *clusters* (Katayama et al., 2005; Mohammad et al., 2009; Wan y Bartolomei, 2008). Se han descrito múltiples genes en los cuales el proceso de *imprinting* es regulado por un NAT por ejemplo Igf2r, Slc22a2 y Sls22a3 que son regulados por el NAT Air (Sleutels et al., 2002), Kcnq1 es regulado por Kcnq1ot1 (Thakur et al., 2004), UB3A (Rougeulle et al., 1998) y GNAS (Hayward y Bonthron, 2000). Los NATs regulan el *imprinting* de los genes en *cis* reclutando proteínas modificadoras de la cromatina. Las modificaciones que reprimen la cromatina se extienden a los genes vecinos del sitio en donde se transcribe el NAT (Mohammad et al., 2009).

7.4 Inactivación del cromosoma X

Es un proceso mediante el cual una de las dos copias del cromosoma X presente en las hembras de los mamíferos es inactivado. El lncRNA XIST (X-inactive-specific transcript) sólo es transcrito desde el cromosoma X inactivo (Xi). Xist gatilla la formación de heterocromatina a lo largo del cromosoma X uniéndose directamente al complejo PRC2 y dirigiéndolo al Xi (Lee, 2012). La expresión de Xist es reprimida por su antisentido Tsix, por tanto el cromosoma X que expresa Tsix se mantiene activo (Lee et al., 1999).

7.5 Splicing alternativo

Se han reportado casos de NATs que regulan el proceso de splicing (Beltran et al., 2008; Hastings et al., 1997; Krystal et al., 1990; Munroe y Lazar, 1991). Algunos trabajos han sugerido que esto puede ser por un proceso de enmascaramiento de los sitios de splicing o de las secuencias regulatorias (Krystal et al., 1990; Munroe y Lazar, 1991). También se han reportado NATs que participan en procesos de remodelamiento de la cromatina mediada por RNAs pequeños, y los cambios en la cromatina generan un cambio en el splicing alternativo (Allo et al., 2009; Ameyar-Zazoua et al., 2012).

Mediante análisis de datos de *affymetrix* de 176 líneas celulares linfoblastoides humanas, se estableció que la mayor parte de los genes que poseen expresión sentido y antisentido posee un evento de splicing alternativo que se correlaciona con la expresión del NAT (Morrissy et al., 2011). Este análisis sugiere que puede existir una gran cantidad de eventos de splicing alternativo regulados por NATs en el transcriptoma humano.

7.6 *Editing* de RNA

El *editing* de adenosina a inosina es inducido por la formación de RNAs doble hebra (dsRNA) que reclutan a la enzima ADAR (adenosina deaminasa). Se ha descrito que la interacción del mRNA del gen de *Drosophila melanogaster* Rnp4f y su NAT, Sas10, lleva a que el mRNA de Rnp4f sea hiper editado y se degrade (Peters et al., 2003).

Se ha sugerido que este mecanismo de acción de los NATs no sería frecuente en humanos y ratón (Neeman et al., 2005).

7.7 Cambios en la estabilidad del RNA

Los NATs pueden formar dsRNAs con los transcritos sentido en el citoplasma, modificando su estabilidad. Los elementos ricos en AU (ARE), son regiones de los mRNAs que presentan una alta frecuencia de adeninas y uracilos y que promueven la degradación del mRNA. Se ha mostrado que el NAT de HIF-1α (aHIF) se une al mRNA sentido cambiando su estructura secundaria y exponiendo una región ARE que favorece su degradación (Uchida et al., 2004). El mecanismo opuesto, en donde un NAT cubre el elemento ARE y aumenta la estabilidad de un mRNA, fue sugerido para el gen híbrido Bcl-2/IgH (Capaccioli et al., 1996). Este tipo de regulación también puede ocurrir entre dos genes codificantes de proteínas que son convergentes. En humano, los mRNAs de los genes WDR83 y DHPS se regulan positivamente de forma mutua a través de la formación de un dsRNA en sus extremos 3' UTR (Su et al., 2012).

El NAT del gen iNOS incrementa la estabilidad del mRNA de iNOS. Este efecto es mediado por la interacción del NAT de iNOS con la proteína HuR (*ARE-binding human antigen R*). La proteína HuR puede suprimir la degradación del mRNA de iNOS inhibiendo su desadenilación o a las enzimas exonucleasas (Matsui et al., 2008).

7.8 Enmascaramiento de sitios de miRNAs

Se ha planteado que algunos NATs podrían formar dsRNAs con mRNAs sentido y enmascarar sitios de unión de miRNAs. Se describió que el NAT del gen BACE1 previene la represión del mRNA de BACE1 enmascarando el sitio de unión del miR-485-5p (Faghihi et al., 2010).

7.9 Regulación de la traducción

Se ha demostrado que los NATs pueden regular la traducción uniéndose a los mRNAs sentido. Un ejemplo es el NAT del gen PU.1 que se une al mRNA sentido en el citoplasma y detiene el proceso de traducción entre la iniciación y el proceso de elongación (Ebralidze et al., 2008). Por otra parte, se ha visto la activación de la traducción por parte de un NAT para el gen de ratón Uchl1. El NAT de este gen se une en la región 5´del mRNA sentido y un dominio SINEB2 presente en el NAT induce un incremento en la eficiencia de la traducción (Carrieri et al., 2012).

7.10 Formación de siRNAs endógenos

Los siRNAs endógenos (endo-siRNAs) son RNAs pequeños de aproximadamente 21 nt que derivan de dsRNAs formados por RNAs endógenos y que son procesados por DICER (Okamura y Lai, 2008). Se han descrito endo-siRNAs derivados de NATs en células somáticas de *Drosophila* (Czech et al., 2008; Ghildiyal et al., 2008; Okamura et al., 2008), en ovocitos de ratón (Tam et al., 2008; Watanabe et al., 2008), testículos de ratón (Song et al., 2011) y en líneas celulares humanas (Kawaji et al., 2008; Werner et al., 2014). Estos mismos estudios han mostrado que a pesar de la gran abundancia de NATs en células de mamífero existe una baja abundancia de endo-siRNAs, lo que sugiere que este no sería el modo de acción predominante de los NATs.

8. Clasificaciones de los NATs

Los NATs pueden ser clasificados en dos grupos dependiendo del *locus* de origen. Los cis-NATs se transcriben en el mismo *locus* genómico del gen sentido, pero desde la hebra de DNA complementaria. Por otro lado, los trans-NAT se transcriben de un *locus* genómico distinto al gen con el cual comparten secuencias complementarias (Lapidot y Pilpel, 2006).

Adicionalmente los NATs pueden ser clasificados en función de su orientación y solapamiento con el transcrito sentido; Cabeza a cabeza (Fig. 1A), cola a cola (Fig. 1B) o solapamiento completo (Fig. 1C) (Lapidot y Pilpel, 2006).

9. mirrorRNAs

Mediante diversas técnicas se ha reportado la existencia de un tipo especial de NATs, que son perfectamente complementarios a mRNAs maduros a lo largo de varios exones, incluyendo los sitios de splicing (Fig. 1D) (Cheng et al., 2005; Haeger et al., 2005; Kapranov et al., 2005; Laabi et al., 1994; Merzendorfer et al., 1997; Podlowski et al., 2002; Rosok y Sioud, 2004; Wu et al., 2008). A esta clase de NATs la hemos denominado **mirrorRNAs**.

El primer mirrorRNA reportado fue el del gen BCMA (TNFRSF17) en humanos, el cual fue encontrado en una biblioteca de cDNAs (Laabi et al., 1994). El mirrorRNA de BCMA comparte los mismos 3 exones que el mRNA sentido y está poliadenilado en su extremo 3'. Los autores realizaron ensayos de protección de RNAsas (RPA) con RNAs provenientes de múltiples tejidos humanos y lograron detectar el mirrorRNA de BCMA en bajos niveles.

Otro ejemplo estudiado es el gen de β-globina de ratón (Volloch et al., 1996). Se describió que este gen poseía un NAT perfectamente complementario al mRNA sentido en reticulocitos y células eritroides de bazo. Este mirrorRNA fue detectado por Northern Blot y además describieron sus extremos 3' y 5' mediante un proceso de ligación de adaptadores de RNA y posterior clonamiento (similar estrategia que un 5' y 3' RACE).

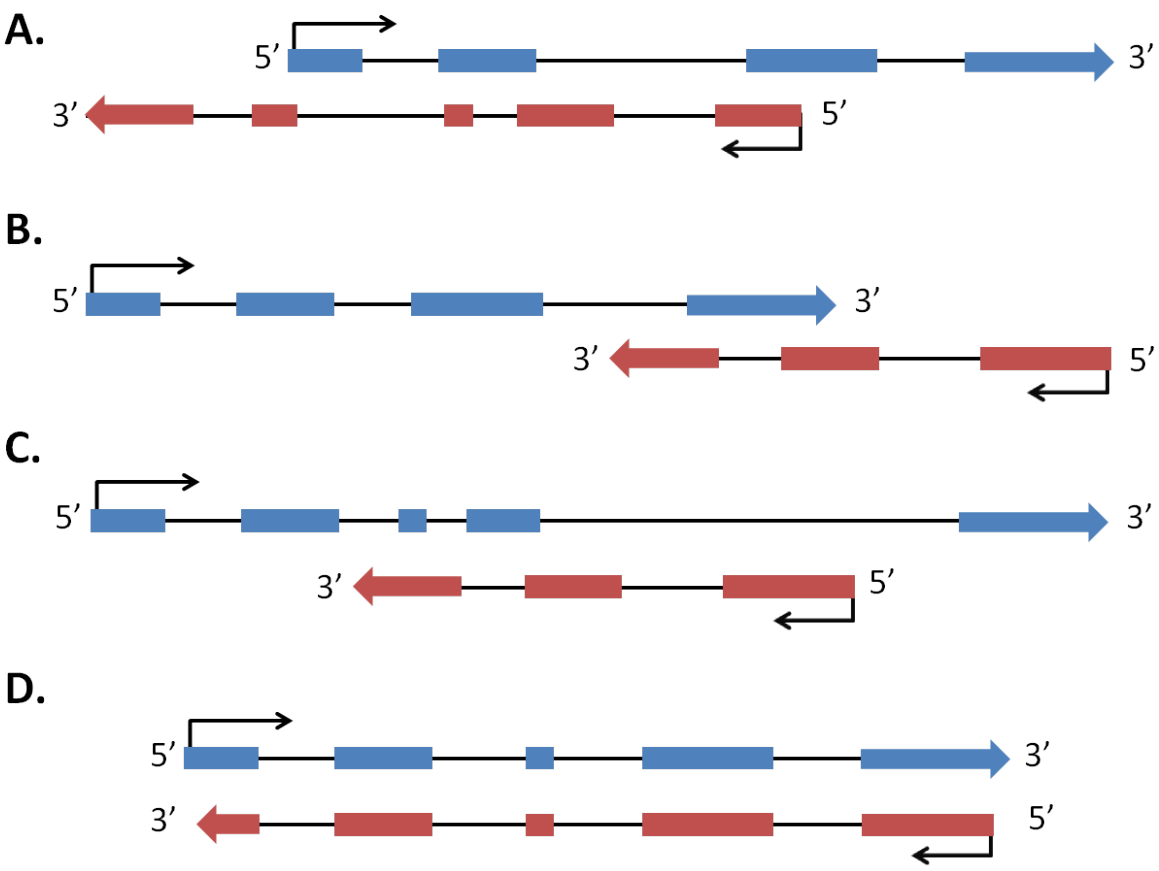


Figura 1. Representación esquemática de los tipos de NATs. **(A)** 5’ con 5’ (cabeza con cabeza). **(B)** 3’ con 3’ (cola con cola). **(C)** Totalmente solapados (un transcrito incluido completamente dentro de la región del otro). **(D)** mirrorRNA, las coordenadas de los exones son las mismas que las del transcrito sentido pero en la hebra complementaria.

Se describió para el gen TNNI3 de rata y humano la existencia de un mirrorRNA en tejido cardíaco. Se detectó el mirrorRNA mediante Northern Blot, RT-PCR hebra específica y por RPA seguido de RT-PCR (Podlowski et al., 2002).

La utilización de nuevas técnicas para caracterizar el transcriptoma de los mamíferos a mayor escala también ha sugerido la presencia de mirrorRNAs. En el año 2005, el grupo de Thomas Gingeras caracterizó el transcriptoma de 10 cromosomas humanos usando *Tiling arrays* con una resolución de 5 nucleótidos (Cheng et al., 2005). Ellos seleccionaron algunos transfrag (fragmentos de transcritos) al azar desde *locis* en donde no había transcritos descritos y los caracterizaron por RACE/array, por RT-PCR hebra específica y secuenciación. Finalmente secuenciaron 170 cDNAs no descritos anteriormente y encontraron que un 14% de ellos era perfectamente complementario a mRNAs sentido.

10. Controversia sobre la existencia de los mirrorRNAs

La existencia de los mirrorRNAs no ha estado libre de controversia. Un grupo reportó que no pudieron detectar el mirrorRNA de β-globina de ratón utilizando Northern Blot (usando como sonda oligonucleótidos) y sugirieron que podría ser un artefacto experimental (Gudima y Taylor, 2001). Llama la atención que en este mismo trabajo los autores realizan un Northern Blot usando una sonda de RNA transcrita *in vitro* y en ese caso sí detectan la presencia del mirrorRNA de β-globina, pero plantean que sería producto de una hibridación no específica o de trazas de transcripción *in vitro* de la hebra complementaria. En este estudio no realizaron ensayos más sensibles como RPA o 3' o 5' RACE utilizando ligación de RNA.

El grupo de Michael Snyder encontró mirrorRNAs utilizando 5' y 3' RACE para caracterizar ciertos transcritos en las regiones definidas por el proyecto piloto de ENCODE (Wu et al., 2008). Al

hibridar cDNA con *microarray* hebra específica se detectaron mirrorRNAs, pero al utilizar RNA marcado en vez de cDNA marcado, no se detectaron los mirrorRNAs. Por esto los autores plantean que los mirrorRNAs serían artefactos de la transcripción inversa producto de que esta enzima puede generar cDNAs doble hebra por su actividad DNA polimerasa dependiente de DNA (DdDP). Previamente se había descrito que parte de las señales de NATs en los *microarrays* hebra específica eran producto de la actividad DdDP de las transcriptasas inversas (RTs) y que este artefacto podía evitarse utilizando Actinomicina D (ActD) en la reacción de RT (Perocchi et al., 2007). El mismo grupo de Snyder plantea que la técnica de RACE permite identificar transcritos con una alta eficiencia y sensibilidad, y concluyen sólo tomando en cuenta la falta de detección de los mirrorRNAs, utilizando *microarrays* con RNA marcado, que son un artefacto experimental. Por tanto hasta el momento no existe claridad si los mirrorRNAs son artefactos experimentales o NATs reales.

11. Posibles mecanismos de biogénesis de los mirrorRNAs

Se han propuesto dos hipótesis que explican la existencia de los mirrorRNAs. Una hipótesis es que estos transcritos son generados por la transcripción bidireccional de un *locus* y que posteriormente los transcritos antisentido sufren splicing en los sitios no consenso CT-AC, que son los sitios complementarios e inversos de los sitios consenso GT-AG utilizados por el mRNA sentido (Haeger et al., 2005; Laabi et al., 1994) (Fig. 2A).

La segunda hipótesis que se ha planteado es que los mirrorRNAs son producidos por una actividad RNA polimerasa dependiente de RNA (RdRP) que utiliza como molde un mRNA maduro,

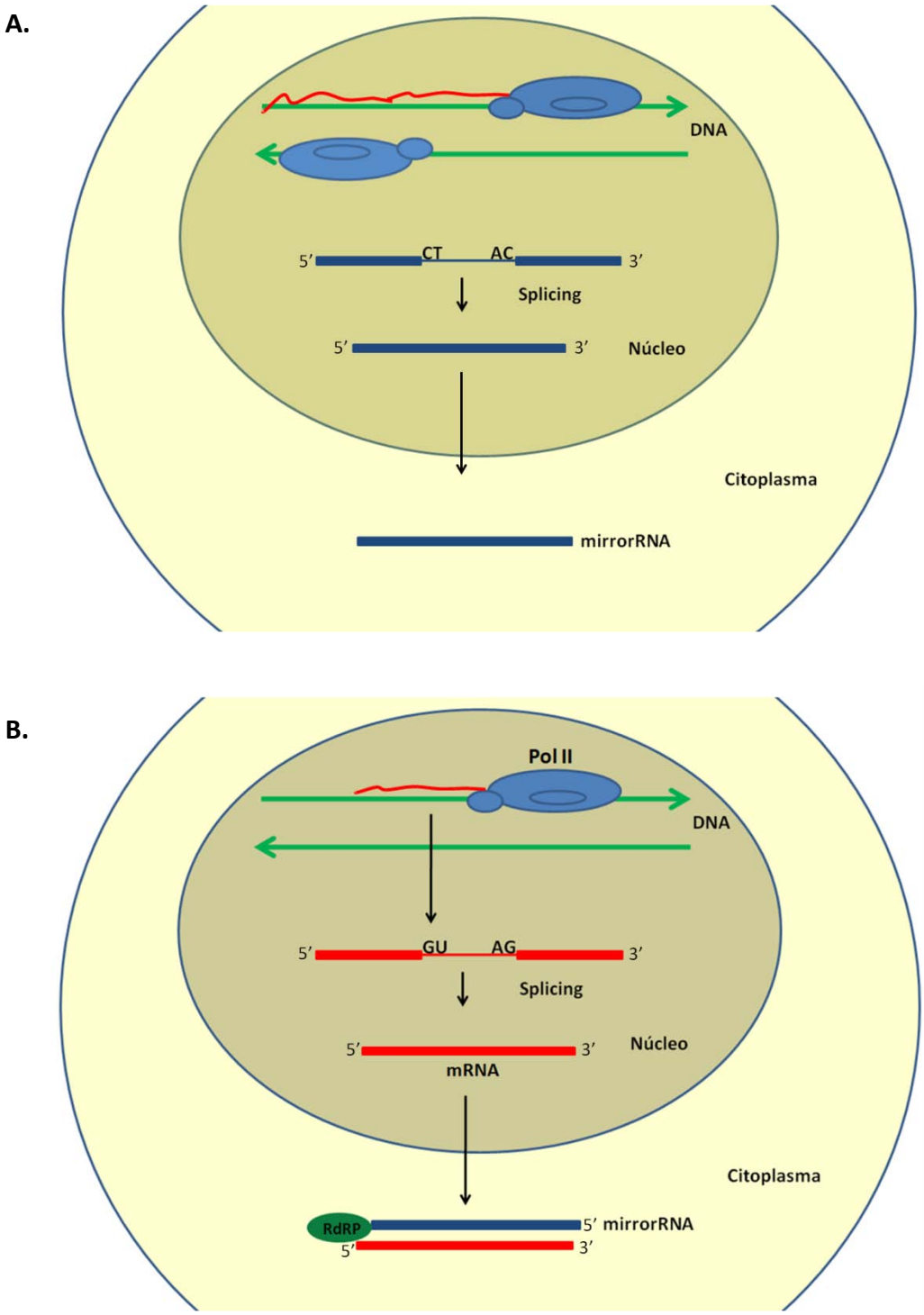


Figura 2. Esquema de los mecanismos propuestos para explicar la biogénesis de los mirrorRNAs. (A) Según esta hipótesis los mirrorRNAs son producidos por la transcripción bidireccional de un *locus* y posterior splicing del NAT en los sitios complementarios a los sitios de splicing consenso. **(B)** Esta hipótesis plantea que los mirrorRNAs son generados por una proteína con actividad RdRP que genera copias de mRNAs maduros.

generando un transcrito antisentido perfectamente complementario al mRNA sentido (Cheng et al., 2005; Haussecker et al., 2008; Merzendorfer et al., 1997; Rosok y Sioud, 2004; Volloch et al., 1996) (Fig. 2B).

12. Splicing de intrones complementarios a los canónicos

La primera hipótesis propuesta, plantea que el *locus* de los mirrorRNAs es transcrito de forma bidireccional, algo que no sería muy extraño a la luz de los resultados del proyecto FANTOM3 que muestra que para el 72% de los transcritos descritos en ratón existe evidencia de transcripción en la hebra complementaria (Katayama et al., 2005). La gran limitación de esta explicación es que la maquinaria de splicing celular tendría que reconocer y utilizar las señales complementarias de las consenso para remover los intrones (Cheng et al., 2005; Frith et al., 2005; Wu et al., 2008).

En los pre-mRNAs nucleares de eucariontes superiores, se han descrito dos tipos de intrones. Estos intrones se diferencian por el spliceosoma que los procesa (Sharp y Burge, 1997). Existen dos spliceosomas descritos, el spliceosoma mayoritario está compuesto por 5 snRNPs (ribonucleoproteínas nucleares pequeñas), U1, U2, U4, U5 y U6. El spliceosoma minoritario está compuesto por 4 snRNPs propios, U11, U12, U4atac y U6atac, y además U5 que es compartido por los dos spliceosomas (Patel y Steitz, 2003). La mayor parte de los intrones son procesados por el spliceosoma mayoritario, y son denominados tipo U2. Tan solo entre un 0,15% y un 0,34% de los intrones de vertebrados son procesados por el spliceosoma minoritario, estos intrones son denominados de tipo U12 (Patel y Steitz, 2003).

Los intrones tipo U2 y los intrones tipo U12 se diferencian por las secuencias conservadas que rodean a los sitios de splicing. Los intrones U2 de vertebrados se caracterizan por una secuencia consenso altamente variable para el sitio 5' dador que sería: **AG/GTRAGT** (R significa purina).

Además poseen un “branch point” pobremente conservado: **CTRACT**. En el extremo 3’ del intrón se encuentra un tramo de polipirimidinas entre el “branch point” y el sitio aceptor que posee el dinucleótido **AG** (Abril et al., 2005). Casi siempre los intrones U2 presentan los dinucleótidos GT y AG en los extremos, aunque existe aproximadamente un 1% de los casos en que se encuentran los dinucleótidos **GC-AG** (Burset et al. 2001).

Los intrones U12 se caracterizan por la secuencia consenso **/[AG]TATCCTT** (el paréntesis cuadrado quiere decir que es A o G) en el sitio dador del splicing. Además poseen un *branch point* **TCCTTAACT**. Estos intrones no poseen tramo de polipirimidinas y el *branch point* está tan solo a 10 o 20 nucleótidos del sitio aceptor **YA[CG]** (La Y simboliza una pirimidina). Aunque inicialmente se describieron por poseer los sitios inusuales AT-AC (Jackson 1991; Hall and Padgett 1994), después se mostró que los intrones U12 también pueden tener los dinucleótidos terminales GT-AG (Dietrich et al. 1997; Sharp and Burge 1997).

Los intrones CT-AC de los mirrorRNAs no se asemejan a ninguno de los tipos de intrones descritos (Laabi et al., 1994), ya que sus secuencias son las complementarias e inversas a las de los intrones de tipo U2.

Por lo expuesto hasta el momento, o realmente no existen los intrones CT-AC o estos serían un nuevo tipo de intrones. Entonces surge la pregunta ¿Existen casos reportados de intrones CT-AC? Se ha descrito la existencia de intrones CT-AC (Delorenzi et al., 2002; French et al., 1999; Haeseleer et al., 2000; He et al., 2009; Kurose et al., 2005; Ladd et al., 2007; Mei et al., 2007; Okada et al., 2007) y la posibilidad de que las células humanas reconozcan y procesen intrones CT-AC (French et al., 1999).

El primer caso de un intrón CT-AC descrito es el del gen U83 del HHV-6 (*Human Herpes Virus-6*) (French et al., 1999). Según los autores cuando el HHV-6 infecta células humanas el gen U83 se expresa y un intrón CT-AC es escindido del transcrito. Se podría argumentar que existen proteínas virales que lleven a cabo este proceso, pero los autores clonaron el gen U83 y lo transfectaron de forma estable en la línea celular HaCaT (derivada de queratinocitos humanos). El resultado es que el transcrito de U83 sigue sufriendo splicing. En genes de mamífero también se han reportado intrones CT-AC. Siah-1 es un gen humano que tiene una variante de splicing en la cual, según los autores, se escinde un intrón CT-AC. Esta variante generaría un codón de término prematuro y por tanto una proteína trunca (Mei et al., 2007).

13. Actividad RdRP en células de mamífero

La segunda hipótesis que se ha planteado para explicar la existencia de los mirrorRNAs es que debe existir una actividad RdRP en células de mamífero (Cheng et al., 2005; Haussecker et al., 2008; Merzendorfer et al., 1997; Rosok y Sioud, 2004; Volloch et al., 1996).

Una RdRP es una enzima que permite generar un transcrito de RNA utilizando como molde otra molécula de RNA. En plantas y en *C. elegans* las RdRPs participan en la biogénesis de siRNAs (Ahlquist, 2002). En mamíferos no existe un homólogo de la RdRP de *C. elegans* o de plantas y por mucho tiempo se sugirió que no existiría una actividad RdRP en células de mamífero (Stein et al., 2003). La actividad RdRP no es necesaria para la generación de siRNAs endógenos en mamíferos ya que los endo-siRNAs pueden provenir de dsRNAs formados por horquillas y NATs derivados de pseudogenes (Tam et al., 2008; Watanabe et al., 2008). Adicionalmente, en células de mamífero no existe un proceso de amplificación de los siRNAs. En cambio, en plantas y *C elegans* la introducción de siRNA (o de dsRNAs) puede inducir un silenciamiento sostenido por RdRPs que amplifican y

mantienen la producción de nuevos siRNAs (Carthew y Sontheimer, 2009; Chapman y Carrington, 2007).

A pesar de lo anterior algunos grupos han postulado la existencia de actividad RdRP en células de mamíferos. Se ha sugerido la presencia de esta actividad en células murinas de eritroleucemia (Volloch, 1986; Volloch et al., 1987). Recientemente Kapranov et al. (2010a) reportaron la existencia de un tipo de RNAs pequeños, antisentidos al extremo 3' de genes conocidos (aTASRs), utilizando secuenciación masiva de moléculas únicas (Helicos). Estos RNAs pequeños antisentido poseen un extremo 5' poli-T, que coincide con el extremo 3' poli(A) de sus transcritos sentido. Los autores sugieren la existencia de una actividad RdRP que copie los mRNAs desde su extremo 3' (incluyendo la cola poliA) generando los aTASRs.

Pero queda preguntarse ¿qué proteína o complejos de proteínas podrían tener esta putativa actividad RdRP en mamíferos? Se reportó que TERT y el RNA RMRP (*RNA component of mitochondrial RNA processing endoribonuclease)* pueden formar un complejo que tiene actividad RdRP (Maida et al., 2009). La proteína TERT es la subunidad catalítica de la telomerasa humana (que tiene actividad transcriptasa inversa). El RNA RMRP forma parte de una endoribonucleasa procesadora del RNA mitocondrial. El complejo ribonúcleoproteico TERT-RMRP produce dsRNAs (lo demostraron para RMRP) que pueden generar endo-siRNAs mediante DICER (Maida et al., 2009). Hasta el momento no se ha demostrado que TERT genere dsRNAs de otros RNAs celulares.

Las RdRPs son esenciales en el proceso de replicación de virus de RNAs que infectan células de mamíferos (Ahlquist, 2002). Las RdRPs que permiten la replicación de genomas virales vienen codificadas en ese mismo genoma. Una notable excepción es el Virus de la Hepatitis Delta (HDV) (Lai, 2005; Taylor, 2009). El HDV es un virus con genoma de RNA y posee sólo un gen. Lo interesante es

que la proteína codificada en el genoma del virus de HDV (en realidad son dos isoformas una larga y otra corta de la proteína HDAg) no tiene actividad RdRP. Estos descubrimientos llevaron a pensar que la replicación de HDV en las células humanas tenía que ser mediado por una polimerasa celular. Actualmente, se sabe que la RNA pol II es la responsable de la replicación y transcripción de HDV mediante una actividad RdRP (Abrahem y Pelchat, 2008; Chang et al., 2008; Chang et al., 2006; Filipovska y Konarska, 2000; Fu y Taylor, 1993; Greco-Stewart et al., 2007; Modahl et al., 2000; Moraleda y Taylor, 2001). Se han realizado estudios cristalográficos para entender los mecanismos moleculares que permiten la actividad RdRP de la RNA pol II (Lehmann et al., 2007). La proteína HDAg del virus de HDV es requerida para la replicación del virus (Chao et al., 1990; Kuo et al., 1989). Se ha planteado que HDAg se une directamente a la RNA pol II y estimula la transcripción mediante el desplazamiento de NELF (*negative elongation factor*) promoviendo la elongación por la RNA pol II (Yamaguchi et al., 2001).

Queda entonces preguntarse si algún RNA celular puede ser utilizado como templado para la actividad RdRP de la RNA pol II, y si esta actividad podría cumplir alguna función. El ncRNA B2 tiene aproximadamente 180 nt y es transcrito por la RNA pol III desde los SINE (*short interspersed elements*) B2, que son elementos móviles que se presentan en un gran número de copias en el genoma de ratón (Kramerov y Vassetzky, 2005). Recientemente se demostró que la RNA pol II extiende el ncRNA B2 por 18 nt en su extremo 3’ utilizando como templado una región interna del mismo ncRNA B2. La extensión del ncRNA B2 mediante la actividad RdRP de la RNA pol II lleva a la desestabilización del ncRNA B2 (Wagner et al., 2013).

14. Transcripción antisentido de pseudogenes procesados

Aunque por el momento no ha sido planteada en la literatura, consideramos posible una tercera alternativa que explique el origen de los mirrorRNAs. Es posible que los mirrorRNAs provengan de la transcripción antisentido de pseudogenes procesados.

Los pseudogenes procesados provienen de un evento de retrotransposición, en el cual el mRNA de un gen es retrotranscrito y el cDNA resultante es insertado en el genoma. Los pseudogenes procesados no poseen los intrones ni el promotor del gen desde el cual provienen (ya que son la copia del mRNA que sufrió splicing) y pueden poseer una cola poli(A) en su extremo 3' (Esnault et al., 2000; Pink et al., 2011). El evento de retrotransposición de pseudogenes procesados es llevado a cabo por la maquinaria enzimática de los LINEs (*Long interspersed elements*) (Esnault et al., 2000).

Han existido múltiples iniciativas para anotar los pseudogenes en el genoma humano (Karro et al., 2007; Khelifi et al., 2005; Ohshima et al., 2003; Pei et al., 2012; Torrents et al., 2003; Zhang et al., 2003). Según la versión 19 de GENCODE existen aproximadamente 10.500 pseudogenes procesados en el genoma humano (Pei et al., 2012). Los pseudogenes procesados poseen, en promedio, una identidad de secuencia del 80,3% con la zona codificante de su gen parental (Pei et al., 2012).

Diversos reportes han demostrado que múltiples pseudogenes procesados son transcritos en los genomas de los mamíferos (Frith et al., 2006; Harrison et al., 2005; Kalyana-Sundaram et al., 2012; Pei et al., 2012; Svensson et al., 2006; Zheng et al., 2007). Según la versión 19 de GENCODE existen 442 pseudogenes procesados que son transcritos en el genoma humano.

Interesantemente también se han detectado NATs provenientes de pseudogenes procesados (Muro y Andrade-Navarro, 2010; Tam et al., 2008; Watanabe et al., 2008; Zhou et al., 1992). Por lo anterior parece plausible que algunos mirrorRNAs puedan provenir de pseudogenes procesados que tengan un alto porcentaje de identidad de secuencia con su gen parental.

Considerando los antecedentes expuestos en esta tesis nos preguntamos si los mirrorRNAs son un artefacto experimental o una nueva sorpresa que nos depara el mundo de los ncRNAs. Si los mirrorRNAs son reales, ¿cuál o cuáles serían los mecanismos mediante los cuales se estarían generando?

La **hipótesis** de esta tesis doctoral es que en células de mamífero existen NATs que son perfectamente complementarios a mRNAs maduros a lo largo de varios exones (mirrorRNAs) y que los mirrorRNAs son producidos por la transcripción antisentido de pseudogenes procesados.

El **objetivo general** fue buscar, caracterizar y validar experimentalmente mirrorRNAs presentes en el transcriptoma humano, y estudiar los mecanismos que podrían generar mirrorRNAs.

Los **objetivos específicos** de esta tesis son:

1. Buscar y caracterizar mirrorRNAs en el transcriptoma humano.
2. Validar experimentalmente mirrorRNAs.
3. Estudiar la biogénesis de los mirrorRNAs.

MATERIALES

1. Material Biológico

1.1 Bacterias

Las células quimiocompetentes *E. coli* DH5α: F^- ϕ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17 (r_k^-, m_k^+) phoA supE44 thi-1 gyrA96 relA1 λ^- fueron obtenidas de Invitrogen Corp. (California, USA).

1.2 Células Eucariontes

HEK293 (CRL-1573): Línea celular derivada de un cultivo primario de riñón de embrión humano transformado con un Adenovirus tipo 5 (Graham et al., 1977). Se obtuvo de ATCC (Manassas, VA, USA).

HEK293T (CRL-11268): Línea celular derivada de células HEK293 y que expresa constitutivamente el antígeno T grande del virus simio 40 (SV40) (DuBridge et al., 1987). Se obtuvo de ATCC (Manassas, VA, USA).

PC12 (CRL-1721): Línea celular derivada de un feocromocitoma de rata (Greene y Tischler, 1976). Se obtuvo de ATCC (Manassas, VA, USA).

HeLa (CCL-2): Línea celular epitelial derivada de un adenocarcinoma humano. Fueron donadas por la Dra. María Estela Andrés.

COS7 (CRL-1651): Línea celular derivada de riñón de monos verdes africanos transformada con SV40 (Gluzman, 1981). Fueron donadas por la Dra. María Estela Andrés.

1.3 Plásmidos

Se obtuvo el plásmido pGEM-T Easy Vector de Promega.

El plásmido pcDNA 3.1 fue donado por la Dra. María Estela Andrés.

Los plásmidos pTEJ8-U83B #52 y #61 fueron donados por el Dr. Hans Luttichau.

Los plásmidos U83A (clon 2-1-43) y U83B (clon HST-3-3) fueron donados por el Dr. Yuji Isegawa.

1.4 RNA

XpressRef™ Human Universal Total RNA y XpressRef™ Mouse Universal Total RNA. Obtenidos de SABiosciences.

FirstChoice® Human Total RNA Survey Panel (RNA total proveniente de 20 tejidos humanos normales). Obtenidos de Ambion.

RNA de células MCF7 (HTB-22). Donado por la Dra. Reini Luco.

2. Reactivos e insumos de biología molecular

Los reactivos e insumos de biología molecular utilizados en la presente tesis fueron obtenidos de los siguientes proveedores:

- Agilent Technologies: PfuUltra II Fusion HS DNA Polymerase.

- Ambion: DNasa *Turbo DNA-free*, FirstChoice® RLM-RACE Kit, RPA III™ Ribonuclease Protection Assay Kit, Poly(A) Polymerase.

- Axygen Biosciences: AxyPrep™ Plasmid Miniprep Kit 250-prep.

- Becton Dickinson: Bacto™ Tryptone.

- Gibco: Select Yeast Extract, Select Agar, Penicilina/estreptomicina/Glutamina 100X, Trypsin-EDTA 10X, Dulbecco`s Modified Eagle Medium, Horse Serum HS, Fetal Bovine Serum FBS, OPTI-MEM®I+GlutaMax™-I Reduced Serum Medium 1X, Dulbecco´s Phosphate-Buffered Saline (PBS).

- Invitrogen: Platinum® Taq DNA Polymerase, PCRx Enhancer System, enzimas de restricción, amortiguadores de reacción de enzimas de restricción, TRIzol, SYBR® Safe DNA gel Stain, LipofectAMINE 2000.

- Merck: etanol, isopropanol, cloroformo.

- New England Biolabs: enzimas de restricción, amortiguador para las reacciones de enzimas de restricción.

- Promega: pGEM®-T Easy Vector System, AMV Reverse Transcriptase.

- Roche Diagnostics: Transcriptor Reverse Transcriptase.

- Thermo Scientific: RevertAid™ Reverse Transcriptase, RevertAid Premium Reverse Transcriptase, dNTP mix, dCTP, Oligo (dT) 18, GeneRuler™ 100 bp Plus DNA Ladder, GeneRuler™ 1 kb DNA Ladder, Random Hexamer, Buffer de carga 6X, IPTG, X-GAL, GeneJET PCR Purification Kit, GeneJET Gel Extraction Kit, Terminal Deoxynucleotidyl Transferase (TdT).

- Winkler: agarosa, solución TAE 50X.

- USBiological: Ampicilina.

-Sigma Aldrich: Actinomicina D.

3. Oligonucleótidos

Los oligonucleótidos utilizados fueron sintetizados por Integrated DNA Technologies, Inc., USA. Se detallan los oligonucleótidos en la Tabla I.

Tabla I. Oligonucleótidos utilizados en la presente tesis.

Nombre	Secuencia	Experimento
QT	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC TTTTTTTTTTTTTTTT	3' RACE
QO	CCAGTGAGCAGAGTGACG	3' RACE y 5' RACE
QI	GAGGACTCGAGCTCAAGC	3' RACE y 5' RACE
AUAP	GGCCACGCGTCGACTAGTAC	3' RACE
AP	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTT	3' RACE
asBCMA-GSP3.1	AAAGTGGCACTGCTCGAGTCGAA	3' RACE
asBCMA-GSP3.2	AGAATGGTTGCGCCTTCCTCCATA	3' RACE
asUBR5-GSP3.1	TTGAGGGCATAGGCTGGAATCCTT	3' RACE y RPA
asUBR5-GSP3.2	CACATTGACTTCACCGCAGCCATT	3' RACE
asKRT8-GSP3.1	ACATTGGCAGAGCTAGCTGAGGTT	3' RACE
asKRT8-GSP3.2	TCCCATCACGTGTCTCGATCTTCT	3' RACE
asKRT8-GSP3.12	TGTTCCCAGTGCTACCCTGCATA	3' RACE
asCWC15-GSP3.1	CTGGGTTAAACCTATTCCCAAGTCC	3' RACE
asCWC15-GSP3.2	AGCCCACACACAATTTAGACAGGG	3' RACE y RPA
asRPS23-GSP3.1	TCAGCTGGACCCTTACACACTT	3' RACE
asRPS23-GSP3.2	TTGGCTGTTTGGCTTCAACTCC	3' RACE
asRPS11-GSP3.1	CAGACATGTTCTTGTGGCGCTT	3' RACE
asRPS11-GSP3.2	TTCTCGAAGCGGTTGTACTTGC	3' RACE y RPA
asCALR-GSP3.1	TGTCCTCATCATCCTCCTTGTCCT	3' RACE y RPA
asCALR-GSP3.2	CCTCCTCCTCTTTGCGTTTCTTGT	3' RACE
asSIDT2-GSP3.1	CCCAGCACAGAGAAGAAGATGACA	3' RACE y RPA
asSIDT2-GSP3.2	ATGCAGAGTCCGGCGATCATGTA	3' RACE
asDDX5-GSP3.1	AGTCCCTGTTGGATTACCAGTCCT	3' RACE y RPA
asDDX5-GSP3.2	AGCAGGCTAGAGTAACCTCTGTCA	3' RACE
asEIF3D-GSP3.1	TCTGTTTGGCACTCTTAGGCAGGA	3' RACE
asEIF3D-GSP3.2	ACGATCTTTGTCTCTGCGGAGGTT	3' RACE
QTg	GAGGACTCGAGCTCAAGCGGGIIGGGIIGGGIIG	5' RACE
5' RACE Outer Primer	GCTGATGGCGATGAATGAACACTG	5' RACE
5' RACE Inner Primer	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG	5' RACE
asBCMA-RT	AGGAACGAATGCGATTCTCTGGAC	5' RACE
asBCMA-GSP5.1	ACGAATGCGATTCTCTGGACCTGT	5' RACE
asBCMA-GSP5.2	TTCTCTGGACCTGTTTGGGACTGA	5' RACE
asUBR5-RT	AATGGCTGCGGTGAAGTCAATGTG	5' RACE
asUBR5-GSP5.1	AAGCTTCTGCAGTTCAAGCGTTGG	5' RACE y RPA
asUBR5-GSP5.2	CAGTTCAAGCGTTGGTTCTGGTCA	5' RACE
asKRT8-RT	CCATTAAGGATGCCAACGCCAAGT	5' RACE
asKRT8-GSP5.1	TGGAGTCTGGGATGCAGAACATGA	5' RACE

Nombre	Secuencia	Experimento
asKRT8-GSP5.2	GTATTCATACGAAGACCACCAGCG	5' RACE
asCWC15-RT	AAGAAAGAGCTGAAGAGCAGGCCA	5' RACE
asCWC15-GSP5.1	CAGGCCAGGAAGGAACAAGAACAA	5' RACE
asCWC15-GSP5.2	ATTCTGAGCGGAAACCCTCTCCTT	5' RACE y RPA
asRPS23-RT	ACGAGACCAGAAGTGGCATGAT	5' RACE
asRPS23-GSP5.1	AGAAAGCTCATTTGGGCACAGC	5' RACE
asRPS23-GSP5.2	AGTGTGTAAGGGTCCAGCTGAT	5' RACE
asRPS11-RT	TCCCGCGGTACTACAAGAACAT	5' RACE
asRPS11-GSP5.1	TACAAGAACATCGGTCTGGGCT	5' RACE y RPA
asRPS11-GSP5.2	TTCAAGACACCCAAGGAGGCTA	5' RACE
asCALR-GSP5.1	ATCACCAACGATGAGGCATACGCT	5' RACE y RPA
asCALR-GSP5.2	ATACGCTGAGGAGTTTGGCAACGA	5' RACE
asSIDT2-GSP5.1	GCTCAGTGCTTGCTATCATGTGTG	5' RACE y RPA
asSIDT2-GSP5.2	CTGCCTGGCCATTGTCATCTTCTT	5' RACE
asDDX5-GSP5.1	TGGAAGAACTGCTCGCAGTACCAA	5' RACE y RPA
asDDX5-GSP5.2	AGCGACCTTATCTCTGTGCTTCGT	5' RACE
asEIF3D-GSP5.1	AGGAACCTCCGCAGAGACAAAGAT	5' RACE
asEIF3D-GSP5.2	TCCTGCCTAAGAGTGCCAAACAGA	5' RACE
qRPS23-R2	CAGCATGACCTTTGCGACC	RPA
qRPS23-F2	GTACCCAATGACGGTTGCTTG	RPA
ACTB-F	CTGGGACGACATGGAGAAAA	RPA
ACTB-R	AAGGAAGGCTGGAAGAGTGC	RPA
ZNF165-F	CCTGCTGCGTGGGCTTCAATTTCA	Validación Intrones
ZNF165-R	ACTGCTCTAGCACCAGCAGTTCCA	Validación Intrones
BCL10-F	CAGGTTGCTTCTTACACAGCGCCA	Validación Intrones
BCL10-R	ACAAGGGTGTCCAGACCTTTTGGGT	Validación Intrones
NEFH-F	GCCGAAAGTGGGTTTGATGACAATG	Validación Intrones
NEFH-R	TTTCAGTGCCTCCAGCTCTGTGGT	Validación Intrones
PRRT-F	CTCCCTAGCTGACTTGCTCCCTCC	Validación Intrones
PRRT-R	GCCCACTGTGTGCCCTTCTCATTC	Validación Intrones
BGN-F	CGTGTCTCTGCTGGCCCTGA	Validación Intrones
BGN-R	TGGAGTAGCGAAGCAGGTCCT	Validación Intrones
LHX9-F	CAGGAGCAGTCCCCAACCCAATCT	Validación Intrones
LHX9-R	CGGTCCAAGTGGTCTGCCTCATTC	Validación Intrones
RAB37-F	GACCAGCGCCAAGACTGGCA	Validación Intrones
RAB37-R	CCCAAACTGGAGCCGCTGACTG	Validación Intrones
SRCRB4D-F	GTGTGCGACGATGACTGGGACTTT	Validación Intrones
SRCRB4D-R	GCGTGGTCTCAGAACCATCCTGCT	Validación Intrones
B15Rik-F	CTGCCAGGAAGTCTGGTATGATGT	Validación Intrones
B15Rik-R	TACCCTCATGATGACGCTGTGCT	Validación Intrones

Nombre	Secuencia	Experimento
M22Rik-F	AGATGCCTTCCGAGCTGTCCTAAA	Validación Intrones
M22Rik-R	GAGCCATCTTTCAGCCCTTGCTTT	Validación Intrones
Alb-F	GTTGTGCAGAGGCTGACAAGGAAA	Validación Intrones
Alb-R	GGCTAAGGCGTCTTTGCATCTAGT	Validación Intrones
Ftl1-F	CTTGGCCATGGAGAAGAACCTGAA	Validación Intrones
Ftl1-R	TGCCTAGTGGCTTGAGAGGTTCAT	Validación Intrones
Kcnk7-F	TTCTGTGTGGTCTATGCAGCCCTT	Validación Intrones
Kcnk7-R	TTCTTCAAGCACTCCTTGGTGCCT	Validación Intrones
Mrps7-F	AGGATCTTCCACGAGGCACTCAAA	Validación Intrones
Mrps7-R	AGCACTCTCCGTCCTTAACTTCCA	Validación Intrones
Nipsnap1-F	AAGCTGGTTCCGTTCCCTCTTTGT	Validación Intrones
Nipsnap1-R	ATTTCGAGTCTCCTCCCGAGATTG	Validación Intrones
Rab2b-F	TCAGTTTACCGACAAGCGGTTCCA	Validación Intrones
Rab2b-R	TTCCCTTACCAGGCTTGCCTTACA	Validación Intrones
Rtkn-F	TCTCTGCATGATCCAGCCCACT	Validación Intrones
Rtkn-R	ATCCTGTTGCTGGTTCCTCTCACA	Validación Intrones
PXR-F	GCTGGTGATTGGCACCGTCATAAA	Validación Intrones
PXR-R	GCATGGTTCCAGCTTTCTTTGGGT	Validación Intrones
U83B-F	CGATGCGGAGCTGTCTTCAG	Minigenes
U83-R	ATTCAAGACTTGTCGGCGATGC	Minigenes
rUcn-F	TGCACTGGATAGACACTCCG	Minigenes
rUcn-R	TCACTTGCCCACCGAATCGAA	Minigenes
U83B-clon-F	ATATATCTCGAGATTGGTATGGCTATCGGATT	Clonamiento
U83B-clon-R	ATATATAAGCTTTTTCATGATTCTTTGTCTAATTTCG	Clonamiento
U83A-clon-F	ATATATGAATTCATTGGTATGGCTATCGGATT	Clonamiento
U83A-clon-R	ATATATGGATCCTTTCATGATTCTTTGTCTAATTTCG	Clonamiento
U83B-AS-clon-F	ATATATGAATTCTCATGATTCTTTGTCTAATTTCGACAATC	Clonamiento
U83B-AS-clon-R	ATATATGGATCCATGTTCATTTGGCTTTTTATTGTTTTTTT	Clonamiento

4. Medios de Cultivo

4.1 Cultivo bacteriano, medio LB

Triptona	10 g/L
Extracto de levadura	5 g/L
NaCl	10 g/L

Se disolvió todo en 1 L de agua destilada, se ajustó a pH 7,0 con NaOH y se esterilizó en autoclave a 21 lb/pul^2 durante 30 min.

4.2 Cultivo de células eucariontes

4.2.1 Dulbecco Modified Eagle's Medium (DMEM base)

Se reconstituyó el DMEM base en polvo en 950 mL de H_2O desionizada estéril, se agregaron 3,7 g de $NaHCO_3$, y se ajustó el pH a 7,2. Una vez completado el volumen a 1 L se esterilizó la mezcla por filtración en filtros de 0,2 µm.

4.2.2 Medio de crecimiento para las células HEK293, HEK293T, HeLa y COS7.

Para 100 mL:

DMEM base	89 mL
Suero Bovino Fetal (FBS)	10 mL
Penicilina Estreptomicina	1 mL

Una vez completado el volumen se esterilizó la mezcla por filtración en filtros de 0,2 µm.

4.2.3 Medio de crecimiento para células PC12

Para 100 mL:

DMEM base	84 mL
Suero de caballo (HS)	10 mL
Suero Bovino Fetal	5 mL
Penicilina Estreptomicina	1 mL

Una vez completado el volumen se esterilizó la mezcla por filtración en filtros de 0,2 µm.

5. Material Fungible

- Axygen: Tubos de PCR 0,2 mL.
- Falcon, Becton Dickinson, NJ, USA: tubos de 15 y 50 mL; pipetas estériles desechables de 10 mL, placas de cultivo celular de 100 mm (353003 y 353803), tubos de poliestireno 14 mL para crecer bacterias (352057).

MÉTODOS

1. Análisis bioinformáticos

1.1 Búsqueda de mirrorRNAs en ESTs y cDNAs humanos

1.1.1 Búsqueda de intrones mirror

Se descargaron los datos de alineamientos de ESTs y cDNAs de humanos generados por BLAT desde el UCSC Table Browser (Karolchik et al., 2004). Utilizando la plataforma web Galaxy (Giardine et al., 2005) se extrajeron las coordenadas, hebra y dinucleótidos de cada uno de los intrones presentes en los alineamientos de los ESTs y cDNAs. Se identificaron pares de intrones que tuvieran las mismas coordenadas pero estuvieran en hebras distintas. Dentro de esos pares el intrón antisentido (identificado por tener los dinucleótidos CT-AC o CT-GC) se le denominó intrón "mirror".

1.1.2 Identificación de ESTs y cDNAs poliadenilados

Para identificar la hebra desde la cual provenían las secuencias de ESTs y cDNAs, se identificaron aquellas secuencias que poseen una cola 3' poli(A). En primer lugar desde el UCSC Table Browser se descargaron las tablas: estOrientinfo y mrnaOrientinfo. A partir de esta tabla se seleccionaron ESTs y cDNAs que tuvieran una cola 3' poli(A) de al menos 6 nt. Como filtros adicionales se exigió: que el término del alineamiento de las secuencias al genoma no estuviera a más de 1 nt del inicio de la cola poli(A), que la cola poli(A) tuviera al menos 4 adeninas que no estuvieran en el genoma y que no hubiera un tramo genómico de más de 6 adeninas antes del inicio de la cola poli(A).

1.1.3 Identificación de genes con ESTs y cDNAs mirror

Para la identificación de putativos mirrorRNAs se seleccionaron los ESTs y cDNAs poliadenilados que sólo poseían intrones mirror. Posteriormente se le asignó el nombre del gen desde el cual podría haberse derivado el mirrorRNA utilizando la tabla RefLink descargada del UCSC Table Browser y desde la herramienta web ID Converter (Alibes et al., 2007).

1.2 Búsqueda de mirrorRNAs en datos de RNA-Seq

1.2.1 Generación de una biblioteca de SJs

A partir de datos de RefSeq de humano se generó una biblioteca con todas los SJs de humano. Se generaron *tags* de 184 nt (92 nt para cada lado del SJ) utilizando la anotación de RefSeq y un *script* escrito en Python. *Tags* con secuencias repetitivas o de baja complejidad fueron detectadas y excluidas con los programas DustMasker (Morgulis et al., 2006) y RepeatMasker (Chen, 2004). Además se excluyeron los *tags* de secuencias de RefSeqs anotadas de forma inversa (que poseían sólo intrones CT-AC). Se generó una biblioteca de SJs como control negativo, para lo cual se invirtió el orden de los exones en los *tags* generados.

1.2.2 Alineamiento de datos de RNA-Seq a la biblioteca de SJs, pseudogenes procesados y al genoma humano

Se utilizaron datos de RNA-Seq hebra específica del proyecto Illumina Body Map 2.0 (IBM 2.0) (Tabla S1) y datos del proyecto ENCODE (Tabla S2). Los datos de RNA-Seq fueron procesados utilizando Galaxy. Se les removieron los adaptadores y los nucleótidos de baja calidad en el extremo 3’ (con un *Phred quality score* < 10). Las lecturas con menos de 50 nt fueron removidas. Se alinearon las lecturas a las bibliotecas de SJs utilizando Bowtie (Langmead et al., 2009) con las opciones -a --best --strata, permitiendo hasta 2 *mismatches*. Las lecturas fueron separadas entre las que alineaban a la hebra sentido y las que alineaban a la hebra antisentido. Las lecturas que alinearon antisentido a los

SJs fueron seleccionadas y posteriormente alineadas al genoma humano de referencia (hg19) y a los pseudogenes procesados presentes en la anotación de GENCODE v12 (la anotación generada por los *pipelines* de GENCODE, Yale y UCSC). Se clasificaron las lecturas en las que alineaban mejor a SJs que a los pseudogenes procesados, las que alineaban con igual *score* y las que alineaban mejor a pseudogenes procesados que a SJs. Aquellas lecturas que alineaban mejor al genoma que a las SJs o a los pseudogenes procesados fueron descartadas. Posteriormente se asignó a cada lectura el nombre del gen desde el que provenía el SJ o el pseudogen procesado.

1.3 Análisis de *Gene Ontology*

Para realizar el análisis de *Gene Ontology* se generó una lista de confianza de genes desde los que se podrían derivar mirrorRNAs. Para esto se seleccionaron genes que tuvieran al menos un cDNA o dos secuencias ya sean ESTs y/o lecturas de RNA-Seq mirror. Para el análisis de *Gene Ontology* se utilizó la plataforma Web GOrilla (Eden et al., 2009).

1.4 Búsqueda de siRNAs endógenos derivados de mirrorRNAs

Se utilizaron datos de RNA-seq de RNAs pequeños producidos por el proyecto ENCODE (Tabla S3). Los datos fueron procesados utilizando FASTX-Toolkit y *scripts* escritos en Python. La librería de RNA-seq de ENCODE fue construida utilizando un protocolo en el cual se ligó un adaptador en el extremo 5' de los RNAs pequeños y se le realizó una poliadenilación *in vitro* en el extremo 3', para posteriormente realizar un RT utilizando un partidor con un adaptador (Djebali et al., 2012). Por ello los datos crudos debieron ser pre-procesados para remover las secuencias propias del proceso de generación de la biblioteca. En primer lugar se removió el adaptador 3' AAAAAAAAAAAAGATCGGAAGAGCGGTTCAGCAGGAATGCCGAGA. Las lecturas en las cuales no se reconoció una secuencia adaptadora fueron analizadas y si existían 6 o más "As" seguidas se les realizó un *trim*. Se eliminaron todas las secuencias de menos de 22 nt (al menos 16 nt de secuencia

del RNA y 6 nt del adaptador 5' NNNNCG). Para garantizar la especificidad de hebra de las secuencias analizadas sólo se utilizaron secuencias que tuvieran el adaptador 5' y que no tuvieran un tramo de poli(A) genómico río abajo del lugar de alineamiento. Se alinearon las secuencias al genoma humano de referencia (hg19) más la biblioteca de SJs. Se analizaron todas las lecturas que alinearan a los SJs de forma única (sólo a un SJ y a ningún otro lugar del genoma). Finalmente para buscar posibles siRNAs endógenos derivados de mirrorRNAs se buscaron lecturas antisentido a SJs que alinearan de forma única y que no tuvieran *mismatches*.

1.5 Búsqueda de *antisense termini-associated short RNAs* (aTASRs)

Para buscar aTASRs se utilizaron datos de RNA-Seq de IBM 2.0 (Tabla S1) y de RNAs pequeños producidos por el proyecto ENCODE (Tabla S3). Los datos de IBM 2.0 recibieron el mismo pre-procesamiento descrito en la sección 1.2.2. Se seleccionaron todas las lecturas que en su extremo 5' tuvieran al menos 6 "Ts" continuas. Las lecturas seleccionadas fueron alineadas al genoma humano de referencia (hg19) de forma cruda y además removiendo las Ts del extremo 5'. Se seleccionaron las lecturas que sólo alineaban al genoma cuando las "Ts" 5' eran removidas. Se cruzaron los alineamientos con las coordenadas de los extremos 3' de transcritos conocidos y con sitios de poliadenilación predichos, según la anotación de GENCODE v17. Los datos de RNAs pequeños de ENCODE que se utilizaron fueron los mismos que se describieron en la sección 1.4 y el análisis posterior para buscar aTASRs es el mismo que el descrito para los datos de IBM 2.0.

2. PCR

2.1 Amplificación de DNA por PCR para clonamientos

Para amplificar el DNA que fue luego insertado en los vectores recombinantes, se utilizó la técnica de PCR. Se utilizó la enzima PfuUltra II Fusion HS DNA Polymerase (Agilent Technologies) en reacciones de 50 µL, siguiendo las instrucciones del fabricante.

2.2 PCR cualitativo

La presencia o ausencia de los transcritos en muestras de cDNAs fue determinada mediante PCRs cualitativos utilizando partidores específicos (Tabla I). Para la reacción se incubó 1µL de cDNA, 0,5 µL de dNTPs (10 mM de cada uno), 1 µL de cada uno de los partidores (10 µM), 0,75 µL de $MgCl_2$ 50 mM, 2,5 µL Buffer 10X, 0,15 µL de Platinum® Taq DNA Polymerase (Invitrogen) y 18,1 µL de H_2O. Se llevaron a cabo las reacciónes de PCR en un termociclador (G-STORM 482) con un protocolo de touchdown descrito en Korbie y Mattick (2008). Las condiciones utilizadas durante el PCR variaban dependiendo de los partidores utilizados y el tamaño del producto esperado. Una alícuota de 10 µL del producto de PCR se analizó por electroforesis en geles de agarosa.

3. Electroforesis de DNA

3.1 Geles de agarosa

Se realizó la separación de fragmentos de DNA por electroforesis en geles horizontales de agarosa (1-2%) preparados en amortiguador TAE (Tris-HCl 40 mM, EDTA 2 mM, CH_3COONa 20 mM ajustado a pH 8,0 con ácido acético glacial) y con SYBR Safe (Invitrogen, life technologies) como marcador de DNA. Los geles analíticos tuvieron un espesor de 5 mm y los preparativos de 5 a 10 mm.

Se sometieron los geles a electroforesis cubiertos por solución amortiguadora TAE a 80-100V. Se cargaron las muestras en el gel preparadas en 0,2 volúmenes de amortiguador de carga 6X. Y se utilizó como estándar de peso molecular *100bp plus* de Thermo Scientific.

3.2 Visualización de DNA

Se visualizaron las bandas de DNA en un transiluminador UV. Se fotografiaron los geles bajo luz UV utilizando el sistema *KODAK Electrophoresis Documentation and Analysis System (EDAS) 290*.

4. Purificación de fragmentos de DNA

Para la purificación de productos de PCR y fragmentos derivados de reacciones enzimáticas se utilizó el GeneJET PCR Purification Kit. Este kit está basado en la unión del DNA a un gel de sílica en condiciones de alta concentración de sales. Para esto se agregó al DNA un volumen de amortiguador de fijación y un volumen de isopropanol y se depositó esta solución en una columna GeneJET que se centrifugó a 12.000 g durante 1 min. Se eliminó el filtrado y se aplicaron 800 µL de amortiguador de lavado a la columna, la cual fue centrifugada a 12.000 g durante 1 min. Una vez descartado el filtrado se centrifugó la columna nuevamente a 12.000 g durante 1 min. Finalmente se eluyó el DNA con 30-50 µL de amortiguador de elución (Tris-HCl 10 mM, pH: 8,5) centrifugando durante 1 min.

5. Determinación de la concentración del DNA/RNA

Se determinó la concentración de DNA por espectrofotometría, realizando mediciones de absorbancia a 260 nm. Además se consideró el valor de la razón DO_{260}/DO_{280} como referencia del grado de pureza de la preparación. Para el DNA doble hebra se utilizó la relación 1 DO_{260}= 50 µg/µL. Para el RNA se consideró la relación 1 DO_{260}= 40 µg/µL.

6. Tratamientos enzimáticos del DNA

6.1 Digestiones con enzimas de restricción

Se incubaron entre 0,2 y 2 µg de DNA con 5 unidades de enzima de restricción en un volumen final de 20 µL, según las condiciones indicadas por el proveedor. El tiempo mínimo de incubación fue de 2 horas a la temperatura correspondiente, según la enzima utilizada (generalmente 37 °C). En el caso de las digestiones preparativas, se incubaron entre 1 y 15 µg de DNA con 5-10 unidades de enzima por un mínimo de 3 horas. Se verificaron las digestiones analizando alícuotas del volumen de digestión por electroforesis en gel de agarosa.

6.2 Ligación de fragmentos de DNA con DNA ligasa del bacteriófago T4

La ligación de fragmentos de DNA para la generación de plásmidos recombinantes fue llevada a cabo utilizando el vector pcDNA3.1. Luego de digerir el vector y el inserto con las enzimas de restricción adecuadas, se mezclaron los fragmentos de DNA en una relación molar de 3:1 (inserto: vector) en cantidades no superiores a 250 ng de DNA total. A esta mezcla se le agregaron 5 µL de T4 DNA Ligation Buffer 2X y H_2O para completar 10 µL. Se agitó la mezcla y se agregó 1 µL de T4 DNA Ligasa.

Para ligar los fragmentos se incubó la mezcla durante 1 hora a temperatura ambiente. Para la transformación de *E. coli* DH5α competentes se utilizaron 2-5 µL de la reacción.

7. Crecimiento y cultivo de *Escherichia coli*

Se crecieron las cepas de *E. coli* utilizadas a 37 °C en placas con medio LB con agar al 1,5% p/v y luego se las mantuvo a 4 °C. Las condiciones normales de crecimiento en medio líquido fueron en medio LB con agitación (250 rpm) durante toda la noche. Las cepas de *E. coli* transformadas con los

distintos plásmidos fueron crecidas en medio LB con ampicilina 50 µg/mL. Para su conservación por tiempos prolongados se las guardó en medio LB con glicerol al 14% v/v en tubos Eppendorf estériles a -80 °C.

8. Obtención de plásmidos recombinantes

8.1 Transformación de células competentes de *E. coli* DH5α

En un tubo de microcentrífuga estéril, se mezclaron 50 µL de una suspensión de células competentes, con 50 a 100 ng de DNA. Se incubó la mezcla por 30 min en hielo, luego se dio un pulso de calor (42 °C por 45 seg) y posteriormente se mantuvo por 2 min en hielo. A continuación se agregaron 950 µL de medio LB y se incubó a 37°C con agitación por 1 hora. De este cultivo se sembró una fracción de 100 a 200 µL en placas con LB agar al 1,5% que contiene como medio de selección ampicilina y se incubó posteriormente por toda la noche a 37 °C.

8.2 Crecimiento de las cepas bacterianas para amplificación de plásmido

Se preparó un inóculo de un clon de *E. coli* en 5 mL de medio LB con ampicilina 50 µg/mL y se dejó crecer toda la noche con agitación a 37 °C.

8.3 Selección de clones positivos mediante PCR en cultivo

Para seleccionar cuáles colonias de bacterias tienen un plásmido recombinante correctamente ligado, se realizó PCR en cultivo. Para ello se hizo una reacción de PCR con Taq polimerasa en condiciones normales, usando los partidores adecuados y 1 µL de cultivo saturado de *E. coli*.

8.4 Purificación de plásmidos a pequeña escala (Miniprep)

Se realizó la preparación de plásmidos a pequeña escala utilizando el kit AxyPrep Plasmid Miniprep Kit de acuerdo a las instrucciones del fabricante. Se eluyó el DNA plasmidial purificado en 80 µL de amortiguador EB.

8.5 Generación de un vector de expresión con el gen U83

Los plásmidos de expresión del gen U83 del virus herpes humano 6 (HHV-6) fueron construidos clonando un fragmento del gen en el vector de expresión pcDNA3.1, utilizando PCR con partidores específicos (Tabla I). Ya que existen dos variantes principales del virus la A y B, se clonó el gen U83 de las dos variantes. El plásmido con el DNA del gen U83A fue donado por el Dr. Yuji Isegawa y el del gen U83B fue donado por el Dr. Hans Luttichau. Además se generó un vector con una versión antisentido del gen U83B en el vector pcDNA3.1. Se confirmaron las secuencias de los plásmidos generados por secuenciación.

9. Extracción de RNA desde células en cultivo

Se retiró el medio de las células cultivadas en placas de 100 mm, se lavó con 2 mL de PBS y se agregaron directamente sobre las células 2-3 mL de reactivo TRIzol, para posteriormente colectar la solución en un tubo de 1,5 mL. Se incubó el homogeneizado de células por 5 min a 30 °C y en seguida se agregaron 0,2 mL de cloroformo. Se agitaron los tubos vigorosamente y luego se incubaron por 3 min a 30 °C. Se centrifugaron las muestras a 12.000 g por 15 min a 4 °C y luego se extrajo la fase acuosa que contiene el RNA a un tubo nuevo. A esta fase se le agregaron 0,5 mL de isopropanol por cada 1 ml de TRIzol, para precipitar el RNA, y se incubó por 1 hora a -20 °C. Se centrifugó a 12.000 g por 10 min a 4 °C y se eliminó el sobrenadante. Se lavó el precipitado de RNA con 1 mL de etanol 75% frío, y se centrifugó a 7.500 g por 5 min a 4 °C. Para conservar el RNA por tiempo prolongado se

guardó a -20 °C precipitado en etanol. Para su utilización inmediata se retiró el etanol y se secó el RNA por aireación. Finalmente se resuspendió el RNA en 40 µL de H_2O libre de nucleasas.

10. Remoción del DNA genómico

Todas las muestras de RNA fueron sometidas a un tratamiento con la DNasa *Turbo DNA-Free* (Ambion), para la remoción del DNA genómico o plasmidial contaminante. Para ello, el volumen total de RNA fue tratado con DNasa siguiendo las instrucciones del fabricante. Luego, se utilizó el RNA para una reacción de RT-PCR o se almacenó a -80 °C.

11. Síntesis de cDNA

Se llevaron a cabo las reacciones de RT utilizando las enzimas RevertAid™ Reverse Transcriptase (Thermo), Transcriptor Reverse Transcriptase (Roche), AMV Reverse Transcriptase (Promega) o RevertAid Premium Reverse Transcriptase (Thermo), siguiendo las instrucciones de cada fabricante. Se realizaron las reacciones utilizando una mezcla de oligo dT (0,5 µg) y random primers (0,2 µg) o un partidor gen específico, dependiendo del experimento. Se utilizó entre 1 y 2 µg de RNA. Los cDNAs fueron guardados a –20 °C o utilizados en una reacción de PCR.

12. Cultivo de células eucariontes

12.1 Mantención y condiciones de crecimiento de células eucariontes

Se cultivaron las líneas celulares HEK293, HEK293T, HeLa y Cos7 en medio DMEM FBS 10%. Las células fueron mantenidas en un incubador NUAIRE modelo Nu-4750 a 37 °C y 5% de CO_2. Se cultivaron las células PC12 en DMEM HS 10%FBS 5% en incubador a 37 °C y 10% de CO_2.

12.2 Descongelamiento de líneas celulares

Para iniciar un cultivo celular los viales que contenían a las distintas células fueron descongelados a 37 °C. Luego se resuspendieron las células rápidamente en 5 mL de medio base y se centrifugó durante 5 min a 1000 rpm. Una vez eliminado el sobrenadante se resuspendió el sedimento en 8 mL de DMEM FBS 20% y se cultivaron las células en placas de 10 cm de diámetro a 37 °C y 5 o 10% de CO_2, dependiendo del tipo celular.

12.3 Congelamiento de líneas celulares

Para almacenar las células durante periodos prolongados, éstas fueron congeladas de la siguiente manera: una vez que las células alcanzaron la confluencia en una placa de 10 cm se retiró el medio de cultivo, se lavó con 3 mL de amortiguador salino fosfato (PBS, NaCl 137 mM, KCl 2,7 mM, Na_2HPO_4 10 mM, KH_2PO_4 1,8 mM, pH 7,4) y después de aspirado éste, se agregó 1 mL de tripsina 0,25% a la placa. Una vez despegadas las células de la placa, se las resuspendió en 2 mL de DMEM FBS 10%. A estas células se les agregó un volumen de una solución FBS 90%, DMSO 10%. Inmediatamente las células fueron guardadas en criotubos y congeladas a -80 °C.

12.4 Transfección de células eucariontes

La transfección de células eucariontes fue realizada utilizando el reactivo LipofectAMINE 2000 (Invitrogen). Para la transfección en placas de 6 pocillos, en el caso de la línea celular HEK293 se sembraron el día anterior entre $5x10^5$ y $7x10^5$ células por pocillo. El complejo DNA-liposoma fue preparado de la siguiente manera: se disponía 1 µg de DNA de vector de expresión en un volumen de 60 µL con agua libre de nucleasas. En ambiente estéril bajo campana se agregaron 90 µL de OptiMEM. Paralelamente se combinaron 150 µL de OptiMEM y 3 µL de LipofectAMINE 2000. Se dejó reposar esta mezcla por 5 min y se mezcló con la solución DNA-OptiMEM, dejando 20 min a

temperatura ambiente. Una vez pasado este tiempo la mezcla DNA-liposoma fue vertida en gotas directamente sobre el medio de cultivo.

13. 3' RACE

El protocolo de 3' RACE fue adaptado del sugerido por (Scotto-Lavino et al., 2006). Las reacciones de RT fueron realizadas con la enzima RevertAid™ Reverse Transcriptase (Thermo Scientific), se utilizaron 2 µg de Human Universal RNA (SAbiosciences) y ActD a una concentración final de 6 µg/ml. Como partidor para la reacción de RT se utilizó QT (Tabla I) excepto para asKRT8 en donde también se usó el partidor AP. Como controles negativos se utilizaron reacciones a las que no se les agregó RT. Los cDNAs obtenidos eran purificados utilizando GeneJET PCR Purification Kit, para remover la ActD. Las reacciones de PCR fueron realizadas utilizando un protocolo de Touchdown PCR (Korbie y Mattick, 2008). Se utilizó una temperatura de desnaturación inicial de 68°C bajando 1°C por cada ciclo hasta llegar a 58°C, para después seguir con 25 ciclos adicionales a 58°C. Los tiempos de elongación utilizados fueron de 2 minutos. Se utilizó 1 µL de una dilución 1/10 del producto del primer PCR, para la segunda ronda de amplificación, que se realizó en un volumen total de 50 µL. Los productos del segundo PCR del 3' RACE fueron analizados en geles de agarosa 1,5%. Algunos productos de PCR fueron clonados directamente en un vector *pGEM-T easy Vector*. En otros casos se purificaron bandas específicas y éstas fueron clonadas en un vector *pGEM-T easy Vector*.

14. Poliadenilación de RNA *in vitro*

Se utilizaron 5 µg de Human Universal RNA (SAbiosciences) el que fue poliadenilado *in vitro* utilizando la enzima Poly(A) Polymerase (AMBION) siguiendo las instrucciones del fabricante. El RNA poliadenilado fue utilizado posteriormente para reacciones de 3' RACE.

15. 5’ RLM RACE

RNA de células humanas MCF7 fue sometido a 5’ RLM RACE. Para este experimento se utilizó First Choice RLM-RACE kit (Ambion) siguiendo las instrucciones del fabricante. Como control negativo de la reacción se utilizó cDNA de células MCF7.

16. RPA-RT-PCR

Se llevó a cabo el protocolo de RPA utilizando 20 μg de RNA total de células HEK293T y RPA III™ Ribonuclease Protection Assay Kit. Se siguieron las instrucciones del fabricante, pero no se adicionó una sonda antisentido, ya que se esperaba detectar RNAs doble hebra formados por mirrorRNAs y sus correspondientes mRNAs sentido. El RNA que no se digirió durante el RPA fue precipitado y utilizado para una reacción de RT utilizando random primers. El cDNA fue utilizado para reacciones de PCR cualitativas para genes específicos.

RESULTADOS

1. Búsqueda y caracterización de mirrorRNAs en el transcriptoma humano

1.1.1 Búsqueda de mirrorRNAs en ESTs y cDNAs de humano

Para identificar la presencia de mirrorRNAs en el transcriptoma humano se utilizaron datos de alineamientos de ESTs y cDNAs generados por BLAT y que fueron descargados del UCSC "Table Browser" (Fujita et al., 2011). La estrategia utilizada para encontrar mirrorRNAs y diferenciarlos de otros NATs consistió en buscar secuencias antisentido que tuvieran intrones en las mismas coordenadas que transcritos sentido pero en la hebra complementaria (Fig. 3A y 3B).

En la Tabla II se resumen los resultados obtenidos en la búsqueda de mirrorRNAs. En la búsqueda inicial se detectó un gran número de ESTs inversos (poseían intrones en las mismas coordenadas que genes sentido pero en la hebra opuesta), de hecho el 7,3 % de los ESTs y el 1,6 % de los cDNAs analizados es inverso. Esto ocurre porque muchas de las bibliotecas de ESTs reportadas fueron preparadas con métodos que no son hebra específica (Shendure y Church, 2002).

Cabe preguntarse: ¿Cómo diferenciar un mirrorRNA real de un EST o cDNA anotado en la hebra equivocada? Para ello se buscaron ESTs y cDNAs antisentido que tuvieran en su extremo 3' una cola poli(A) que no esté codificada en el genoma y que permitiera determinar con mayor seguridad la orientación real de la secuencia (Fig. 3C). Sólo un 1,8 % de todos los ESTs y cDNAs analizados posee una cola poli(A) no genómica en su extremo 3' (Tabla II). De las 144.732 secuencias que poseen una cola poli(A) se encontraron 97 secuencias que podrían ser mirrorRNAs. En la Tabla S4 se detalla información para cada una de las secuencias encontradas. Las 97 secuencias encontradas provienen

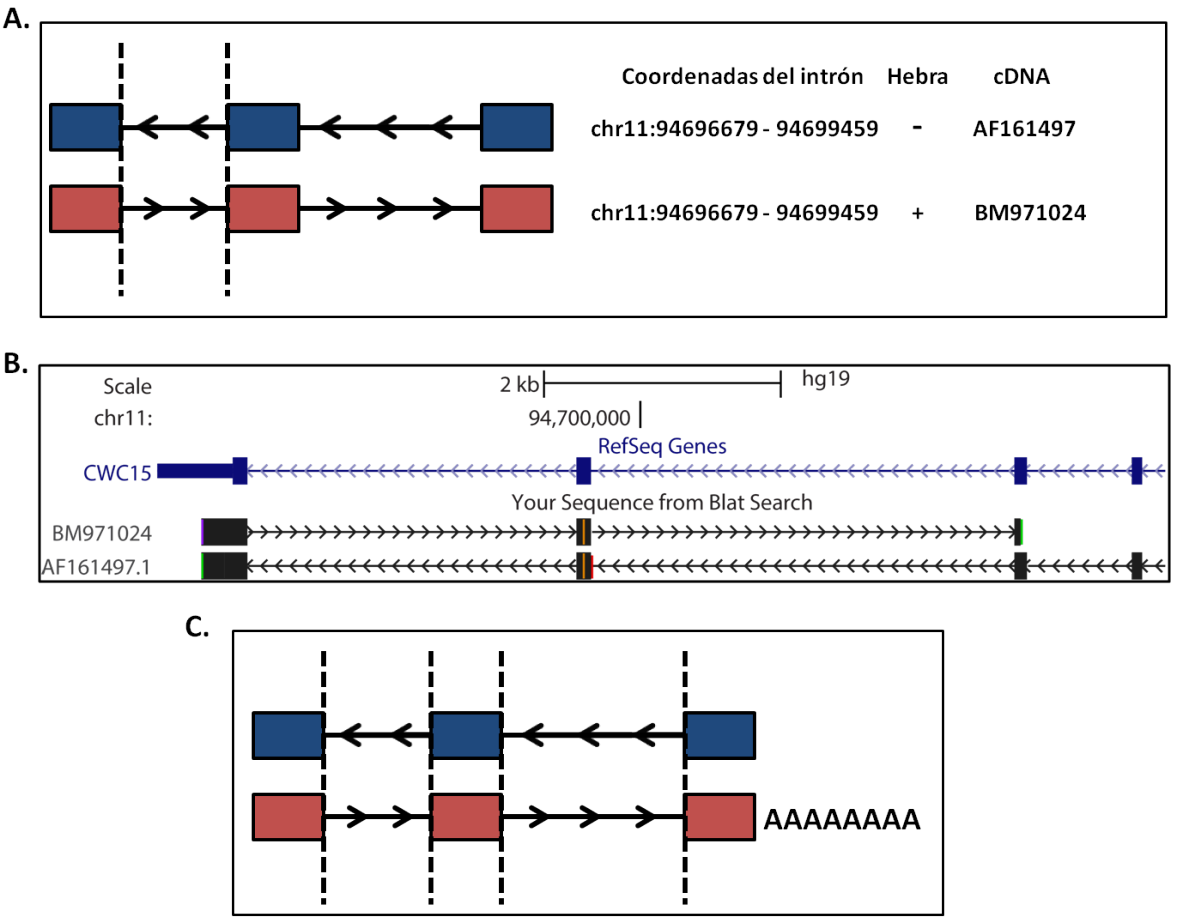


Figura 3. Estrategia utilizada para buscar mirrorRNAs en datos de ESTs y cDNAs de humano. (A) La estrategia utilizada para encontrar mirrorRNAs consistió en buscar secuencias antisentido que tuvieran intrones en las mismas coordenadas que transcritos sentido pero en la hebra complementaria, como se aprecia para el cDNA mirror BM971024. **(B)** Esta imagen obtenida del UCSC Genome Browser muestra el alineamiento del cDNA mirror BM971024 al genoma humano. BM971024 es antisentido al gen CWC15 (y a al cDNA AF161497). Las coordenadas de los dos últimos intrones del gen CWC15 son las mismas que la de los intrones del cDNA BM971024. En negro se muestran bloques de alineamiento entre un cDNA y el genoma, en rojo se indican los *mismatches*, en naranjo las inserciones, las secuencias terminales y las cola poli(A) de un cDNA que no alinean al genoma se muestran en purpura y verde respectivamente. Esta codificación de colores se utilizó en todas las imágenes obtenidas de UCSC Genome Browser. **(C)** Debido a que los ESTs y cDNAs pueden estar anotados en la hebra equivocada, se buscaron secuencias antisentido que tuvieran en su extremo 3' una cola poli(A) que no esté codificada en el genoma, para así determinar con mayor seguridad la orientación de la secuencia.

TABLA II. Búsqueda de mirrorRNAs en datos de cDNAs y de ESTs de humano.

ESTs	**7.576.295**
ESTs inversos	**553.965**
cDNAs	**311.625**
cDNAs inversos	**4.916**
ESTs y cDNAs totales	**7.887.920**
ESTs y cDNAs inversos	**558.881**
ESTs y cDNAs con cola 3' poli(A)	**144.732**
ESTs y cDNAs mirror (inversos y con cola 3' poli(A))	**97**
Genes con ESTs y cDNAs mirror	**68**
ESTs y cDNAs pseudo-mirror	**10**
Genes con ESTs y cDNAs pseudo-mirror	**5**

En la Tabla se resumen los hallazgos de la búsqueda de mirrorRNAs en datos de cDNAs y de ESTs de humano alineados al genoma humano con BLAT.

de 68 genes distintos. En la Tabla III se detalla la lista de estos genes y las secuencias mirror encontradas.

En la Figura 4A-C se muestran ejemplos de alineamientos de ESTs y cDNAs mirror a los genes B2M (Fig. 4A), GAS5 (Fig. 4B) y RPS23 (Fig. 4C).

1.1.2 Señales de poliadenilación en mirrorRNAs

Se analizaron las 97 secuencias de mirrorRNAs encontradas en búsqueda de posibles señales de poliadenilación (se buscó la señal AATAAA en los últimos 40 nt de la secuencia). Se encontró que sólo 3 de las 97 secuencias tienen una señal de poliadenilación consenso (3,1%) (Tabla S5). En cambio al analizar todos los ESTs y cDNAs con cola poli(A) se encontró que 90.607 secuencias tienen la señal de poliadenilación canónica de un total de 144.732 (62,6%).

1.1.3 Los mirrorRNAs pueden provenir de pseudogenes procesados presentes en el genoma humano

En la búsqueda de mirrorRNAs se identificaron 10 ESTs que eran complementarios a mRNAs procesados, pero que además alineaban igual o mejor a pseudogenes procesados presentes en el genoma de referencia. En la Tabla S6 se detalla la información de estas secuencias y de los pseudogenes procesados a los cuales alinean. Para diferenciar a los mirrorRNAs de estos casos en donde las secuencias detectadas son más similares a pseudogenes procesados que a genes sentido procesados, se les denominará pseudo-mirrorRNAs. En la Figura 5A-B se muestra el ejemplo de la secuencia DR978583 que alinea como mirrorRNA al *locus* del gen DGKZ, pero que además alinea a un pseudogen procesado de DGKZ con un mejor puntaje de alineamiento (Fig. 5C-D).

Tabla III. Genes humanos que poseen evidencia de poseer mirrorRNAs a partir de datos de ESTs y cDNAs.

Gen	N° de secuencias	Secuencias
EEF1A1	8	AV735662,AV735984,AV738540,AV740744,AV742230,AV743254,AV743616, AV744581
RPS25	5	AA566093,AA781013,AI312586,AI752208,AI963629
GAS5	4	GD136887,GD137102,GD150301,GD152924
RPL13	4	GD152688,GD152708,GD152714,GD160158
LRRC75A-AS1	3	GD138133,GD152025,GD152782
RPL23A	3	GD143882,GD158185,GD160362
RPS23	3	GD135669,GD140429,GD140430
B2M	2	BM285389,GD160188
DHX29	2	CA432936,GD137882
DTX3	2	DR977956,DR978658
RPL31	2	DR980540,GD157704
TXN	2	GD140231,GD140308
YME1L1	2	DR978824,DR978825
ABCA10	1	AF119885
ACTC1	1	AJ709012
ANUBL1	1	AI624999
BTN3A2	1	GD152943
C2orf89	1	CR936634
C3orf63	1	GD136070
CCDC72	1	BF969899
CD37	1	BM285388
CD46	1	AV682890
CLGN	1	GD158955
COX6C	1	AJ710812
CWC15	1	BM971024
DDX18	1	AV648754
DNAJA2	1	AF116720
DOCK1	1	GD142934
FAM171A1	1	BC089427
FAM177A1	1	GD150565
FAM49B	1	GD157538
FAU	1	GD157815
HERC2	1	DR979211
HNRNPC	1	AV739735
IFI30	1	DR980344

Gen	N° de secuencias	Secuencias
KRT8	1	BC018589
LOC642943	1	AY726558
MAP1S	1	BM021488
MYCT1	1	GD141525
MYL4	1	AI065072
NUCB2	1	GD147401
PHACTR4	1	GD149950
PSMD1	1	GD157744
PTPRC	1	GD135647
RPL13A	1	GD156177
RPL18	1	BM285381
RPL18A	1	CB114618
RPL19	1	GD158226
RPL24	1	BU781776
RPL38	1	GD152785
RPS11	1	AJ710017
RPS16	1	BG033413
RPS20	1	AI132933
RPSA	1	GD142155
RUNX1	1	AY509916
SF3B1	1	GD153942
SLBP	1	DR979183
SLC22A23	1	DR979512
SLC25A26	1	AJ709052
SPPL2A	1	AF090911
STRADB	1	AF116618
SYNRG	1	AF090924
TAF1D	1	GD136763
TSPAN12	1	DA371122
TTLL4	1	GD138416

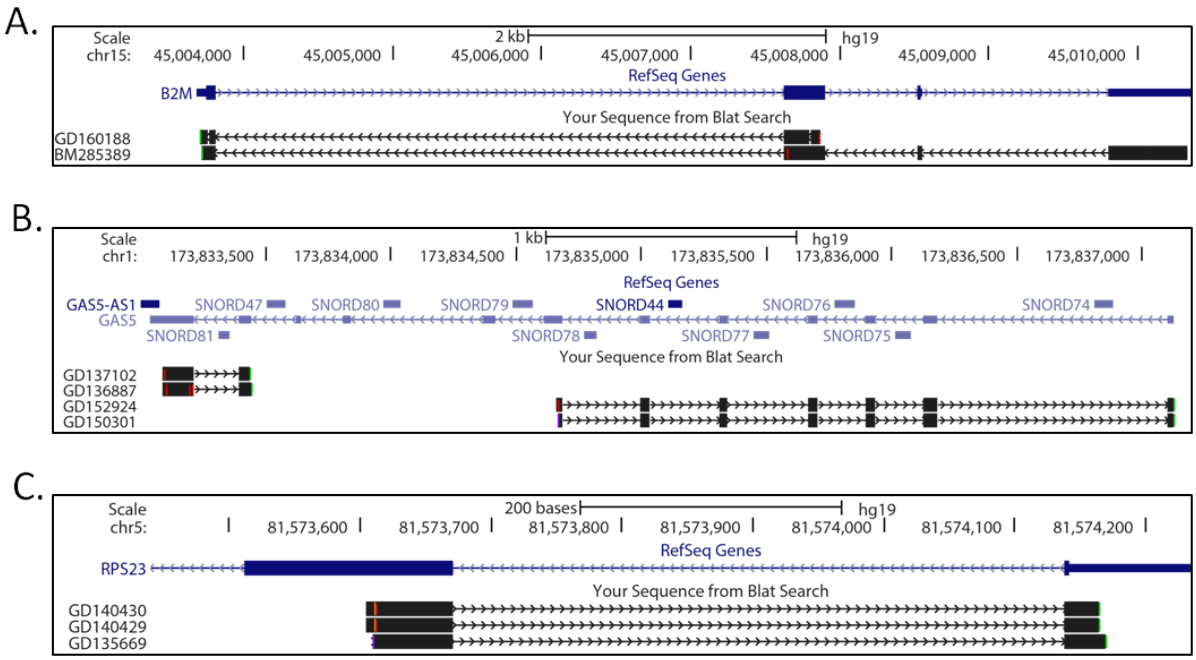
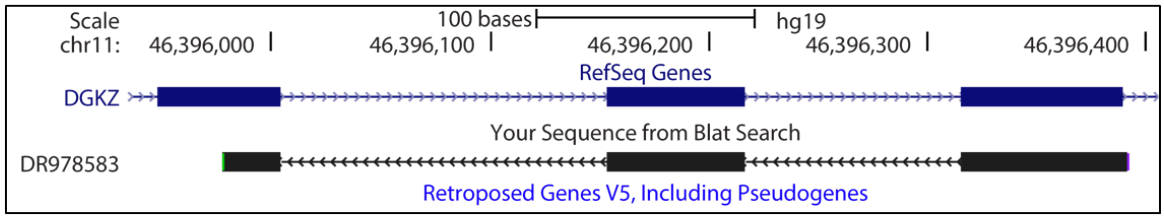


Figura 4. Ejemplos de ESTs y cDNAs mirror alineados al genoma humano. (A-C) Imágenes obtenidas del UCSC Genome Browser de alineamientos de ESTs y cDNAs mirror utilizando BLAT. Al final de las secuencias alineadas se ven líneas verdes que indican la presencia de una cola poli(A).

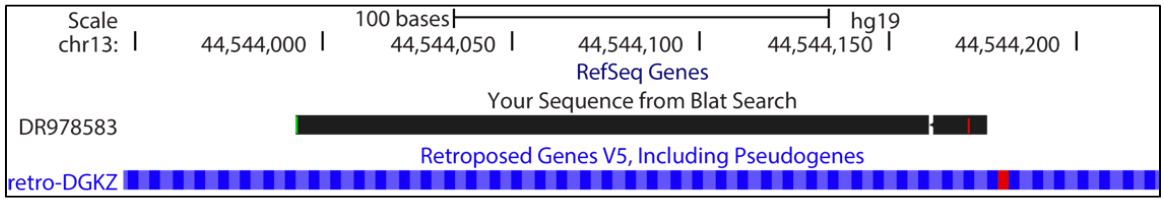
A.



B.

```
catgatgccc gcacgtACCT GGGGATGTTC AGGAAAACAA CACACTGGGG   50
TTTCAGGTCC TGGATCTTGG GAGTCAAGTC CATTCCATCA CACACCACTC  100
GGATGTGCTT GGCCAGGTCC TTGGAGCTGC CCATCAGGAA GTCAGAGAAA  150
GCTGTCCCGG CGTAGAACAT CTTATTCCGA AAaaaaaaa
```

C.



D.

```
CATGaTGCCC GCACGTACCT GGGGATGTTC AGGAAAACAA CACACTGGGG   50
TTTCAGGTCC TGGATCTTGG GAGTCAAGTC CATTCCATCA CACACCACTC  100
GGATGTGCTT GGCCAGGTCC TTGGAGCTGC CCATCAGGAA GTCAGAGAAA  150
GCTGTCCCGG CGTAGAACAT CTTATTCCGA AAaaaaaaa
```

Figura 5. Identificación de un NAT que proviene de un pseudogen procesado del gen DGKZ. (A) La imagen del UCSC Genome Browser muestra el alineamiento de la secuencia DR978583 al genoma humano. DR978583 es antisentido del tipo mirror al gen DGKZ. **(B)** El alineamiento de DR978583 con BLAT al *locus* chr11:46395978-46396391, muestra diferencias en su extremo 5' y una cola 3' poli(A) y entrega un *score* de 164 de 189. Las bases que se muestran en celeste marcan los inicios y términos de los bloques de alineamiento, por ello cuando hay dos bases celestes juntas indican la presencia de una deleción en el cDNA, en este caso las dos deleciones corresponden a dos intrones. **(C)** DR978583 también alinea a un pseudogen procesado de DGKZ. Este alineamiento al genoma humano es el que posee el mejor *score* **(D)** El alineamiento de DR978583 con BLAT al *locus* chr13:44543994-44544176, muestra la presencia de una cola 3' poli(A) y el *score* del alineamiento es de 179 de 189.

1.2 Búsqueda de mirrorRNAs en datos de RNA-seq hebra específica de humano

1.2.1 Búsqueda de mirrorRNAs en datos de RNA-seq de IBM 2.0

Con el advenimiento de nuevas tecnologías de secuenciación se ha podido estudiar el transcriptoma humano a una profundidad sin precedentes. Por ello se utilizaron datos de RNA-Seq hebra específica (bibliotecas en donde los adaptadores para la secuenciación se ligaron a nivel de RNA) para buscar mirrorRNAs. Se utilizaron datos del proyecto IBM 2.0 que están libremente disponibles (detalles en la Tabla S1) y que provienen de una mezcla de RNA extraído de 16 órganos humanos distintos. Los datos ERS025087 provienen de una biblioteca generada con RNA poliadenilado (poli(A)+) y los datos ERS025084 provienen de RNA total al cual se le removió el RNA ribosomal (Tabla S1). Cada conjunto de datos cuenta con aproximadamente 400 millones de lecturas crudas (Tabla IV).

La estrategia utilizada para identificar mirrorRNAs fue identificar lecturas que fueran antisentido a uniones de exones conocidos, para así diferenciar las lecturas antisentido provenientes de mirrorRNAs de lecturas antisentido provenientes de otros tipos de NATs (Fig. 6A). Se generó una biblioteca con todas las uniones de exones presentes en los datos de RefSeq de humano. Las lecturas fueron alineadas a las bibliotecas de uniones de exones utilizando el alineador Bowtie (Langmead et al., 2009). Como control negativo en el proceso de alineamiento se utilizó la misma biblioteca de uniones de exones pero se invirtió el orden de los exones (Fig. 6B).

Las lecturas que alinearon antisentido a las SJs, posteriormente fueron alineadas al genoma (y a pseudogenes procesados que están presentes en el genoma humano). Se compararon los resultados de los alineamientos para ver si existían lecturas que alinearan tanto al genoma como a los SJs. La idea era seleccionar mirrorRNAs que pudieran provenir desde mRNAs o desde pseudogenes procesados y no desde sitios aleatorios del genoma. Finalmente se dividieron los resultados en tres

Tabla IV. Búsqueda de mirrorRNAs en datos de RNA-seq de IBM 2.0.

	Poli(A)+ (ERS025087)	RNA Total (ERS025084)	TOTAL
Número de lecturas	374.983.446	434.395.089	809.378.535
Numero lecturas post-filtros	317.639.170	293.197.295	610.836.465
Lecturas alineadas a SJs	64.463.359	32.952.767	97.416.126
Lecturas alineadas AS a SJs	1.768	2.483	4.251
Lecturas alineadas AS a SJs control	28	24	52
Lecturas mirror a SJs	612	352	964
Lecturas mirror a SJs y pseudogenes	246	624	870
Lecturas mirror a pseudogenes	910	1.507	2.417
Genes con lecturas mirror	361	356	559
Genes con lecturas mirror a SJs	172	69	221
Genes con lecturas mirror a SJs y pseudogenes	71	95	127
Genes con lecturas mirror a pseudogenes	166	270	328

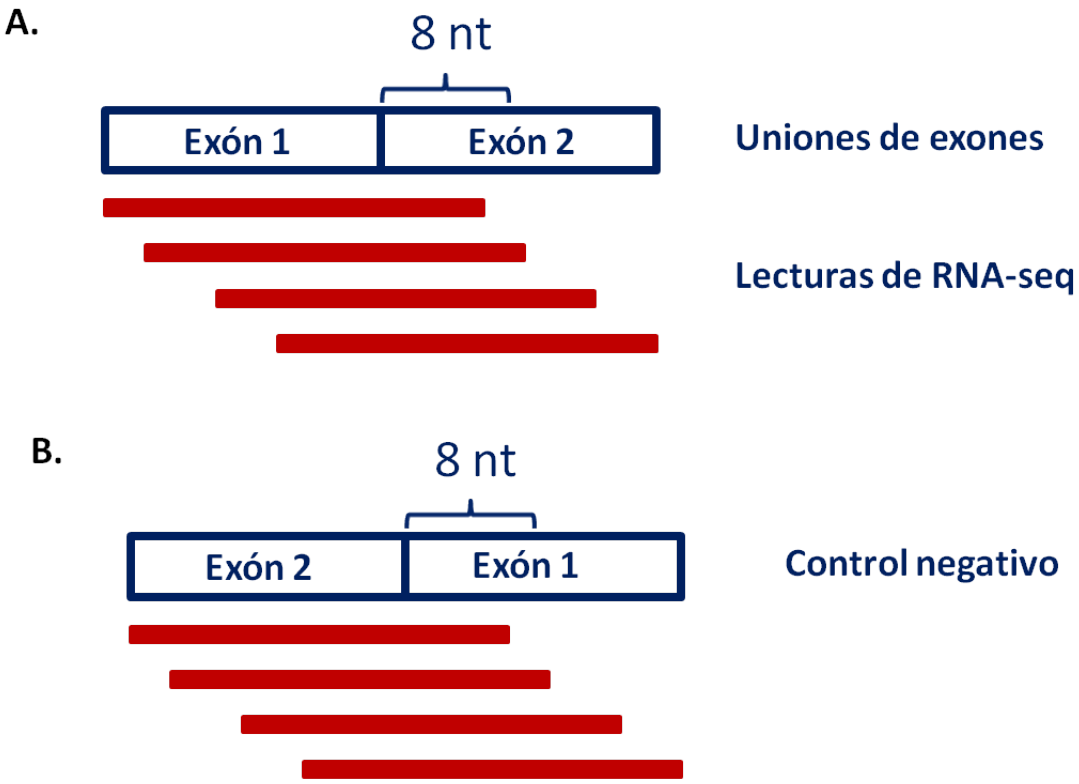


Figura 6. Estrategia utilizada para buscar mirrorRNAs en datos de RNA-Seq hebra específica. (A) Se generaron bibliotecas de SJs anotadas en RefSeq y se alinearon datos de RNA-Seq hebra específica a los SJs. Las lecturas deben alinear al menos 8 nt sobre cada exon para ser consideradas. Las lecturas que alinearon de forma antisentido a las SJs fueron seleccionadas ya que podrían provenir de mirrorRNAs. **(B)** Como control negativo del proceso se utilizaron SJs a las cuales se les invirtió el orden de los exones.

categorías: 1) lecturas que alineaban mejor a los SJs que al genoma, 2) lecturas que alineaban con el mismo puntaje a pseudogenes procesados y a SJs y 3) lecturas que alineaban mejor a pseudogenes procesados que a las SJs. Los resultados se resumen en la Tabla IV. Se encontraron 4251 lecturas antisentido a SJs en los datos de RNA-Seq lo que representa un 0,0044% de las lecturas totales alineadas a los SJs (Tabla S7). Este valor a pesar de ser bajo está muy por sobre el control negativo en donde sólo alinearon 52 lecturas. Las lecturas identificadas podrían provenir mayoritariamente de la transcripción antisentido de pseudogenes procesados, ya que sólo un 22,7% (964 lecturas) de las lecturas alinea mejor a los SJs que a los pseudogenes procesados.

Los datos de la Tabla IV sugieren la existencia de mirrorRNAs y pseudo-mirrorRNAs que están poliadenilados, ya que se encuentran 1768 lecturas antisentido a SJs en la biblioteca poli(A)+, que representa el 0,0027% del total de las lecturas alineadas a SJs en la biblioteca poli(A)+. En la biblioteca de RNA total se encontraron 2483 lecturas lo que representa el 0,0075% del total de las lecturas alineadas a SJs en la biblioteca de RNA total. El mayor porcentaje de lecturas alineadas AS a los SJs en la biblioteca de RNA total se debe a que existe un mayor número de lecturas mirror que alinean a pseudogenes procesados en esa biblioteca (Tabla IV).

Se encontraron 559 genes que poseen lecturas posiblemente derivadas de mirrorRNAs o pseudo-mirrorRNAs (Tabla S8). De estos 559 genes hay 366 genes que poseen evidencia de dos o más lecturas (Tabla S8). Se encontraron 221 genes que poseen lecturas posiblemente derivadas de mirrorRNAs (lecturas antisentido que alinean mejor a SJs que a pseudogenes procesados). De esos 221 genes hay 97 que poseen evidencia de 2 o más lecturas (Tabla S9). Por otra parte, se encontraron 363 genes que poseen lecturas posiblemente derivadas de pseudo-mirrorRNAs (lecturas

antisentido que alinean igual o mejor a pseudogenes procesados que a SJs). De esos 363 genes hay 276 que poseen 2 o más lecturas (Tabla S10).

Para ver si los dos tipos de bibliotecas de RNA-Seq de IBM 2.0 generaban resultados similares. Se compararon que genes poseían lecturas mirror o pseudo-mirror en los dos tipos de bibliotecas. En la Figura 7, y Tablas S11 y S12 se muestran los resultados de las intersecciones entre las dos tipos de bibliotecas. Se observa que las dos bibliotecas comparten 20 genes con evidencia de mirrorRNAs (Fig. 7A) y 132 genes con evidencia de pseudo-mirrorRNAs (Fig. 7B).

Se detectaron genes para los cuales existen lecturas antisentido del tipo mirrorRNA o pseudo-mirrorRNA, para más de uno de sus SJs (Tablas S13 y S14). En la Figura 8A se muestra el ejemplo del gen MYH11, para el cual existen lecturas tipo mirrorRNA para 3 SJs distintos. Un caso muy interesante es el del pseudo-mirrorRNA del gen GK. Como se aprecia en la Figura 8B existen múltiples lecturas tipo pseudo-mirrorRNA en el *locus*, la particularidad es que en los datos de RefSeq está anotado que este pseudogen se transcribe de forma antisentido (GK3P).

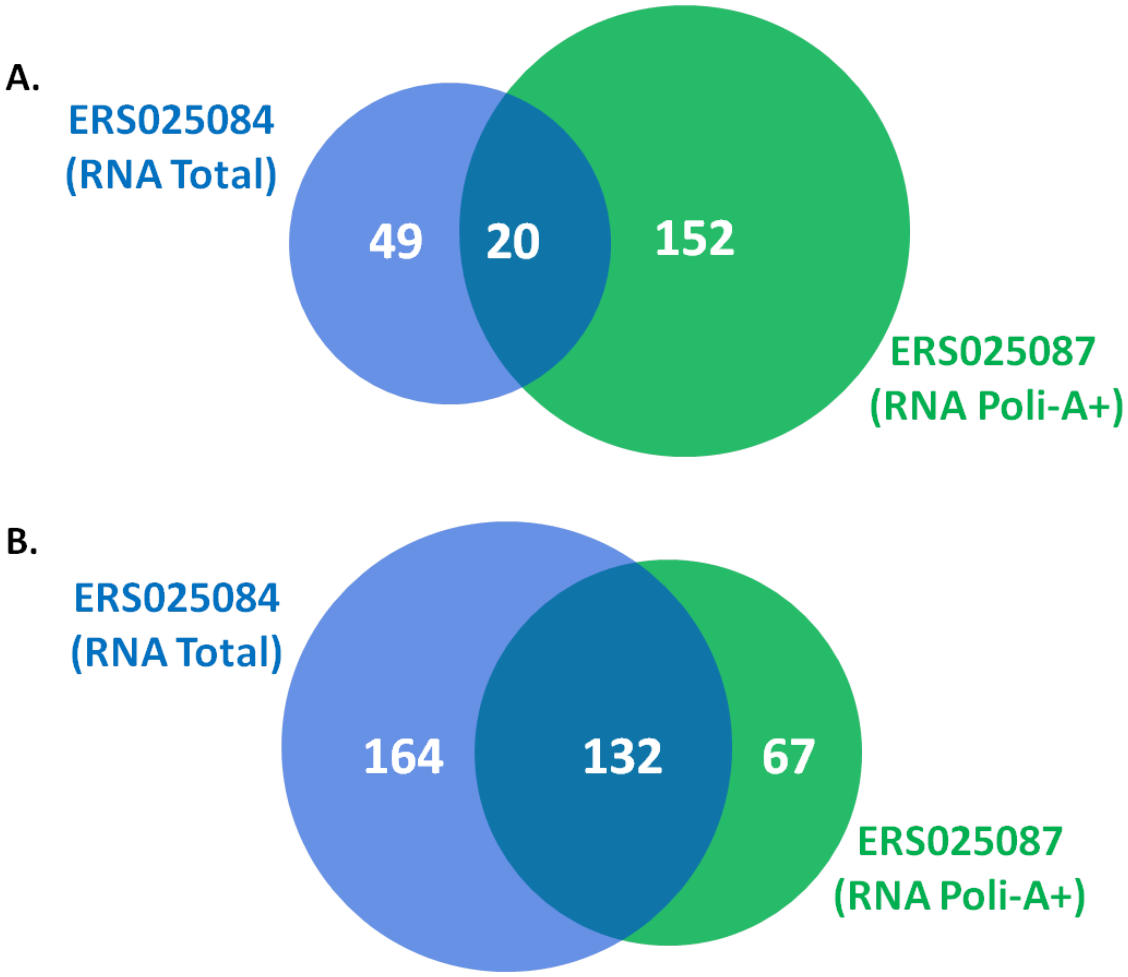


Figura 7. Diagrama de Venn que muestra el número de genes que presentan evidencia de poseer mirrorRNAs o pseudo-mirrorRNAs en datos de RNA-Seq de IBM 2.0 generados con RNA poli(A) + y RNA total. (A) Número de genes que poseen lecturas mirror. **(B)** Número de genes que poseen lecturas pseudo-mirror.

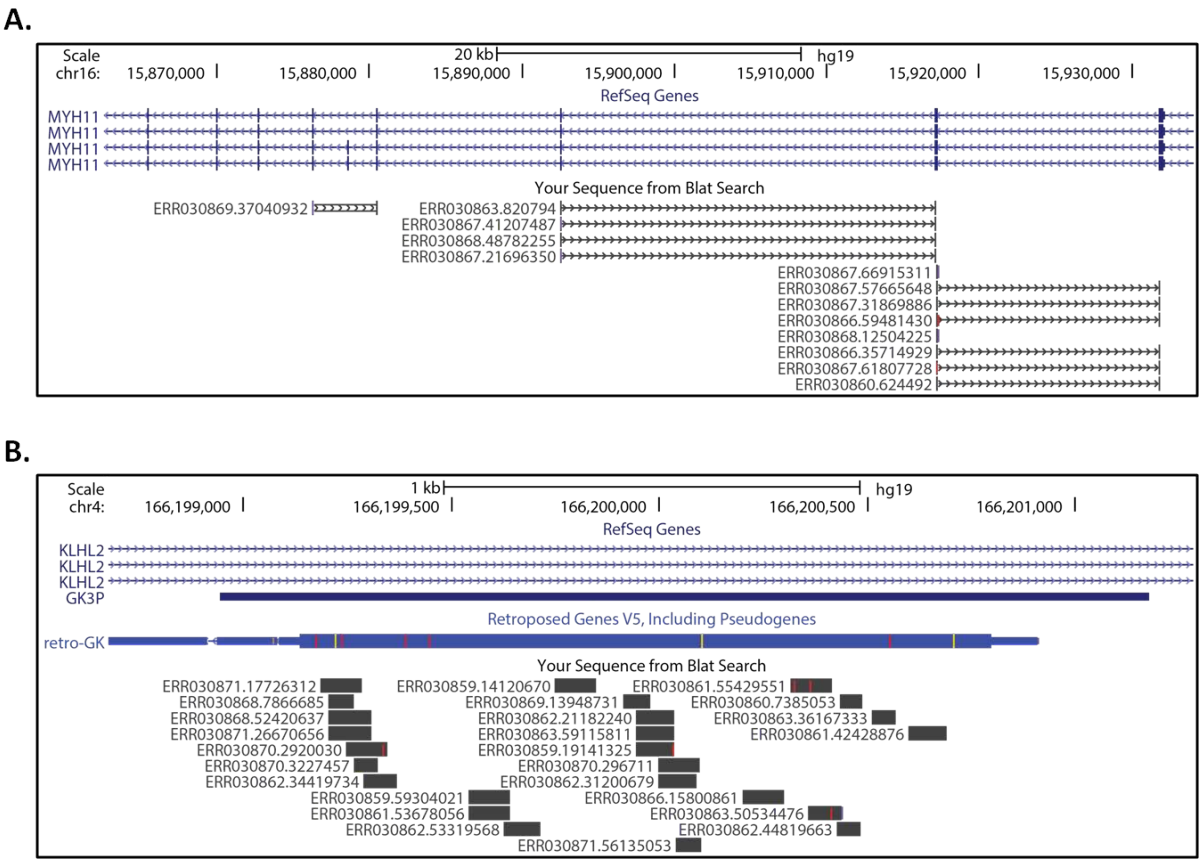


Figura 8. Ejemplos de mirrorRNAs y pseudo-mirrorRNAs que poseen lecturas de RNA-Seq antisentido a varios SJs. (A) Imagen del UCSC Genome Browser con el alineamiento de lecturas de RNA-Seq mirror al gen MYH11. Para el gen MYH11 se detectaron lecturas mirror para tres de sus SJs. **(B)** Alineamiento de lecturas de RNA-Seq antisentido a un pseudogen procesado del gen GK. Según la anotación de RefSeq (GK3P) este pseudogen procesado de GK se transcribe en la dirección antisentido.

1.2.2 Búsqueda de mirrorRNAs en datos de RNA-seq del proyecto ENCODE

El proyecto ENCODE ha caracterizado el transcriptoma de diversas líneas celulares humanas a una gran profundidad usando un protocolo de RNA-Seq hebra específica descrito en Parkhomchuk et al. (2009). Los análisis preliminares mostraron que en las bibliotecas de RNA-Seq de ENCODE se detectaban proporcionalmente 30 veces más mirrorRNAs (datos no mostrados) que en los datos de IBM 2.0. Por tanto no se utilizaron los datos de RNA-Seq de ENCODE preparados con la biblioteca descrita (Parkhomchuk et al., 2009), para evitar la posibilidad de falsos positivos. En cambio existen datos de RNA-Seq hebra específica de ENCODE preparados con el protocolo de ligación de adaptadores a nivel de RNA, que es el mismo utilizado en los datos de IBM 2.0. Los datos de RNA-Seq utilizados están detallados en la Tabla S2 y corresponden a datos provenientes de RNA poli(A)+ de 5 líneas celulares. Se resumen los resultados de la búsqueda de mirrorRNAs en los datos de ENCODE en la Tabla V. Se encontraron 1033 lecturas antisentido a SJs en los datos de RNA-Seq lo que representa un 0,0024% de las lecturas totales alineadas a los SJs (Tabla S15).

Se encontraron 140 genes que poseen lecturas posiblemente derivadas de mirrorRNAs o pseudo-mirrorRNAs (Tabla S16). De estos 140 genes hay 77 genes que poseen evidencia de dos o más lecturas. Se encontraron 16 genes que poseen lecturas posiblemente derivadas de mirrorRNAs. Por otra parte, se encontraron 126 genes que poseen lecturas posiblemente derivadas de pseudo-mirrorRNAs.

Tabla V. Búsqueda de mirrorRNAs en datos de RNA-seq de ENCODE.

Número de lecturas	**366.558.543**
Numero lecturas post-filtros	**192.054.639**
Lecturas alineadas a SJs	**42.654.941**
Lecturas alineadas AS a SJs	**1.033**
Lecturas mirror a SJs	**497**
Lecturas mirror a SJs y pseudogenes	**119**
Lecturas mirror a pseudogenes	**417**
Genes con lecturas mirror	**140**
Genes con lecturas mirror a SJs	**16**
Genes con lecturas mirror a SJs y pseudogenes	**44**
Genes con lecturas mirror a pseudogenes	**112**

1.3 Comparación de datos de cDNAs/ESTs y de RNA-Seq en la búsqueda de mirrorRNAs

Al comparar los mirrorRNAs y pseudo-mirrorRNAs que se encontraron en los datos de ESTs y cDNAs con los de RNA-seq se obtiene la Figura 9A-C. Se observa que existe un gran solapamiento en los genes identificados entre los datos de ENCODE y de IBM 2.0 (9A). Al comparar sólo los datos de mirrorRNAs se observa que los datos de ENCODE se solapan en más de un 50% con los de IBM 2.0, pero los datos de cDNAs/ESTs tienen un bajo grado de solapamiento (9B). De hecho la intersección entre los datos de IBM 2.0 y de cDNAs/ESTs es de dos genes, y no alcanza a ser estadísticamente significativa ($\chi 2$ $p=0.14$). Finalmente los datos de pseudo-mirrorRNAs presentan un altísimo grado de solapamiento entre las tres fuentes de datos.

1.4 Características de los mirrorRNAs

1.4.1 Análisis de *Gene Ontology*

A partir de los datos obtenidos de cDNA/ESTs y de datos de RNA-Seq se generó una lista de confianza, que incluyera genes a partir de los cuales se podrían generar mirrorRNAs. Para estar en esta lista cada gen debe poseer al menos un cDNA mirror o dos secuencias ya sean EST y/o lectura de RNA-seq (Tabla S17). A la lista de genes desde los cuales se pueden generar mirrorRNAs se le realizó un análisis de *Gene Ontology* (GO), para ver si existen genes que participen en cierto tipo de procesos enriquecidos en esta lista. Se detallan Los resultados en la Tabla S18. Se observa que existe un evidente enriquecimiento de genes con mirrorRNAs que participan en el proceso de traducción de proteínas.

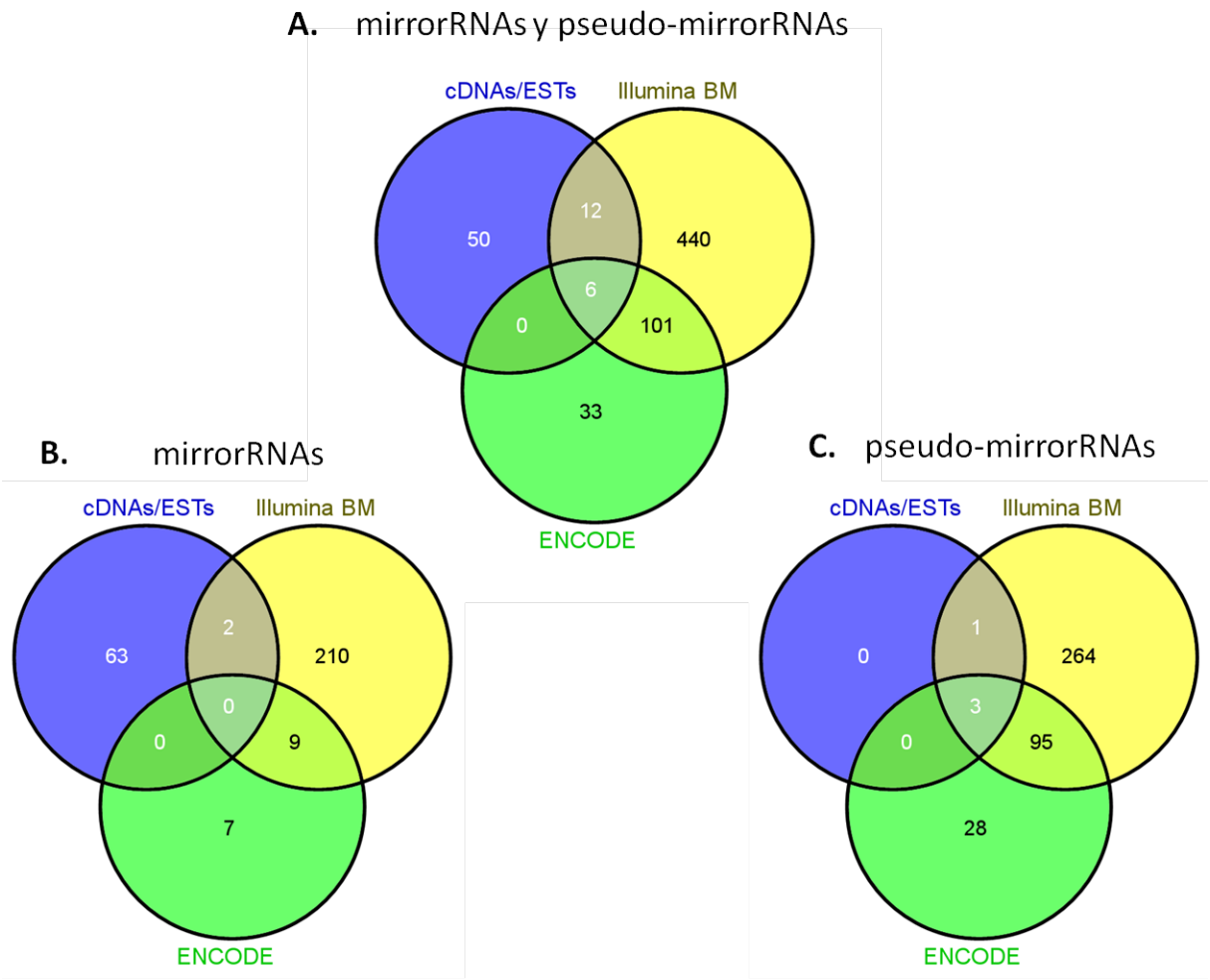


Figura 9. Diagramas de Venn con la comparación de los mirrorRNAs y pseudo-mirrorRNAs detectados en los datos de cDNAS/ESTs, IBM 2.0 y ENCODE. (A) Número de genes con evidencia de mirrorRNAs y pseudo-mirrorRNAs. **(B)** Número de genes con evidencia de mirrorRNAs. **(C)** Número de genes con evidencia de pseudo-mirrorRNAs.

1.4.2 mirrorRNAs y pseudogenes procesados

En el análisis de GO se observó un enriquecimiento de genes que participan en traducción de proteínas, ya que se encontraron mirrorRNAs para los genes: RPS23, EIF3D, RPS14, RPL34, RPL23A, RPS3, RPS25, RPL31, RPL32, EEF1A1 y EIF2S2. Es interesante notar que una característica de esta lista de genes es que 10 de estos 11 genes poseen múltiples pseudogenes procesados en el genoma (entre 5 y 82), sólo EIF3D no tiene pseudogenes procesados en el genoma de referencia. A partir de la anotación de pseudogenes de GENCODE V7 se detectaron pseudogenes procesados para el 13% de los genes codificantes de proteínas. Para la lista de genes a partir de los cuales se podrían generar mirrorRNA se detectó que un 35% de ellos posee pseudogenes procesados anotados en el genoma, lo que implica una clara sobre representación ($\chi 2$ $p<0,0001$).

1.5 Búsqueda de RNAs pequeños derivados de mirrorRNAs

Los niveles de expresión detectados para los mirrorRNAs son muy bajos. Una posibilidad es que estos RNAs sean producidos en muy pequeñas cantidades y la otra es que tengan vidas medias muy cortas. Una vida media muy corta podría deberse a una degradación rápida, o al procesamiento mediado por proteínas como DICER. Si los mirrorRNAs formaran dobles hebras de RNA con sus contrapartes sentido podrían ser reconocidos y procesados por DICER generando siRNAs.

Para evaluar la posibilidad de que existan endo-siRNAs derivados de mirrorRNAs se utilizaron datos de RNA-Seq del proyecto ENCODE de secuenciación de RNAs pequeños. Los datos fueron pre-procesados, para garantizar su especificidad de hebra. Se utilizaron datos de 3 tipos celulares distintos, además se utilizaron datos provenientes de distintos compartimentos celulares (RNA total, nuclear o citoplasmático) y 3 tipos de tratamientos de los RNAs pequeños. El tratamiento con las enzimas CIP y TAP permite un enriquecimiento de RNAs con Cap en la secuenciación realizada por ENCODE. Al utilizar el RNA sin tratamiento, los RNAs que poseen Cap van a estar sub representados

en la secuenciación. Al tratar sólo con TAP, ENCODE obtuvo la representación más amplia de los RNAs pequeños celulares. En total se procesaron más de 1.700 millones de lecturas de RNAs pequeños. Los datos fueron alineados al genoma humano y a una biblioteca de SJs (la misma utilizada para buscar mirrorRNAs en datos de RNA-Seq). Además como control se utilizó una biblioteca de SJs en donde los exones estuvieran en posición inversa. Se resumen los resultados en la Tabla VI. Como se aprecia en la Tabla VI, no se encontraron RNAs pequeños de entre 20-24 nt (que es el tamaño de los siRNAs) antisentido a SJs. Aunque en un bajo número sí se encontraron RNAs pequeños sentido que alinearan a SJs, desde el tamaño de 16 hasta 95 nt (rango de tamaños que se analizó), sin poseer un enriquecimiento en los RNAs pequeños de entre 20-24 nt (dato no mostrado).

Utilizando los datos de RNA-Seq de RNAs pequeños se encontraron posibles endo-siRNAs derivados de pares de transcritos sentido-antisentido. En la Figura 10A se observa el ejemplo del gen NVL que posee un NAT y en esa región del genoma se solapan exones de los dos transcritos. Como se aprecia en los datos de RNA-seq de IBM 2.0 existe transcripción de las dos hebras en esa región del genoma. Se observa que los datos de RNA-seq de RNAs pequeños de H1-neurons y MCF7 muestran evidencia de la existencia de un endo-siRNA derivado de la hebra positiva (la del NAT en este caso). En la Figura 10B se observa el gen ZNF630 que posee un NAT. Para H1-neurons y A549 se identifica la presencia de un endo-siRNA.

A partir de los resultados obtenidos se puede decir que no existe evidencia que los mirrorRNAs sean procesados por DICER y den lugar a siRNAs.

TABLA VI. Búsqueda de siRNAs derivados de mirrorRNAs.

Células	RNA	Biblioteca	N° lecturas post-filtros	N° lecturas alineadas	Alineamientos SJs AS	Alineamientos SJs S	RNAs 20-24 nt SJs AS	RNAs 20-24 nt SJs S
A549	Total	Rep3 sin tratamiento	89.051.064	68.771.740	249 (625)*	5.522 (88)	0 (0)	80 (0)
A549	Total	Rep4 sin tratamiento	82.146.000	62.542.307	380	9.687	0	91
A549	Total	Rep1V2 TAP	54.294.588	32.129.744	220 (2.874)	1.939 (214)	0 (0)	66 (0)
A549	Total	Rep2V2 TAP	60.918.750	38.163.342	224	4.112	0	116
A549	Total	Rep3 CIP-TAP	65.351.054	35.504.814	3.951 (2.096)	3.984 (1.637)	0 (0)	21 (0)
A549	Total	Rep4 CIP-TAP	55.370.700	30.147.294	3.944	3.280	0	13
A549	Citoplasma	Rep3 TAP	52.460.743	34.432.564	2.267	8.238	0	135
A549	Citoplasma	Rep3 CIP-TAP	53.641.974	30.159.730	5.504	6.502	0	26
A549	Nuclear	Rep3 TAP	112.608.743	88.679.404	135	3.053	0	38
A549	Nuclear	Rep3 CIP-TAP	45.191.601	37.728.692	935	2.149	0	19
H1-neurons	Total	Rep1 Sin tratamiento	95.449.257	67.216.974	591	19.784	0	77
H1-neurons	Total	Rep1 TAP	112.417.229	83.469.803	303	26.029	0	87
H1-neurons	Total	Rep1 CIP-TAP	116.886.814	93.416.480	310	21.487	0	36
MCF7	Total	Rep3 sin tratamiento	95.156.050	69.399.966	131	4.658	0	313
MCF7	Total	Rep1V2 TAP	87.440.995	51.377.828	145	6.559	0	237
MCF7	Total	Rep3 CIP-TAP	73.012.247	33.051.936	2.597	6.026	0	50

*** En () y en rojo se entregan los números de los alineamientos realizados a una biblioteca de SJs que se utilizó como control negativo (exones en posiciones inversas).**

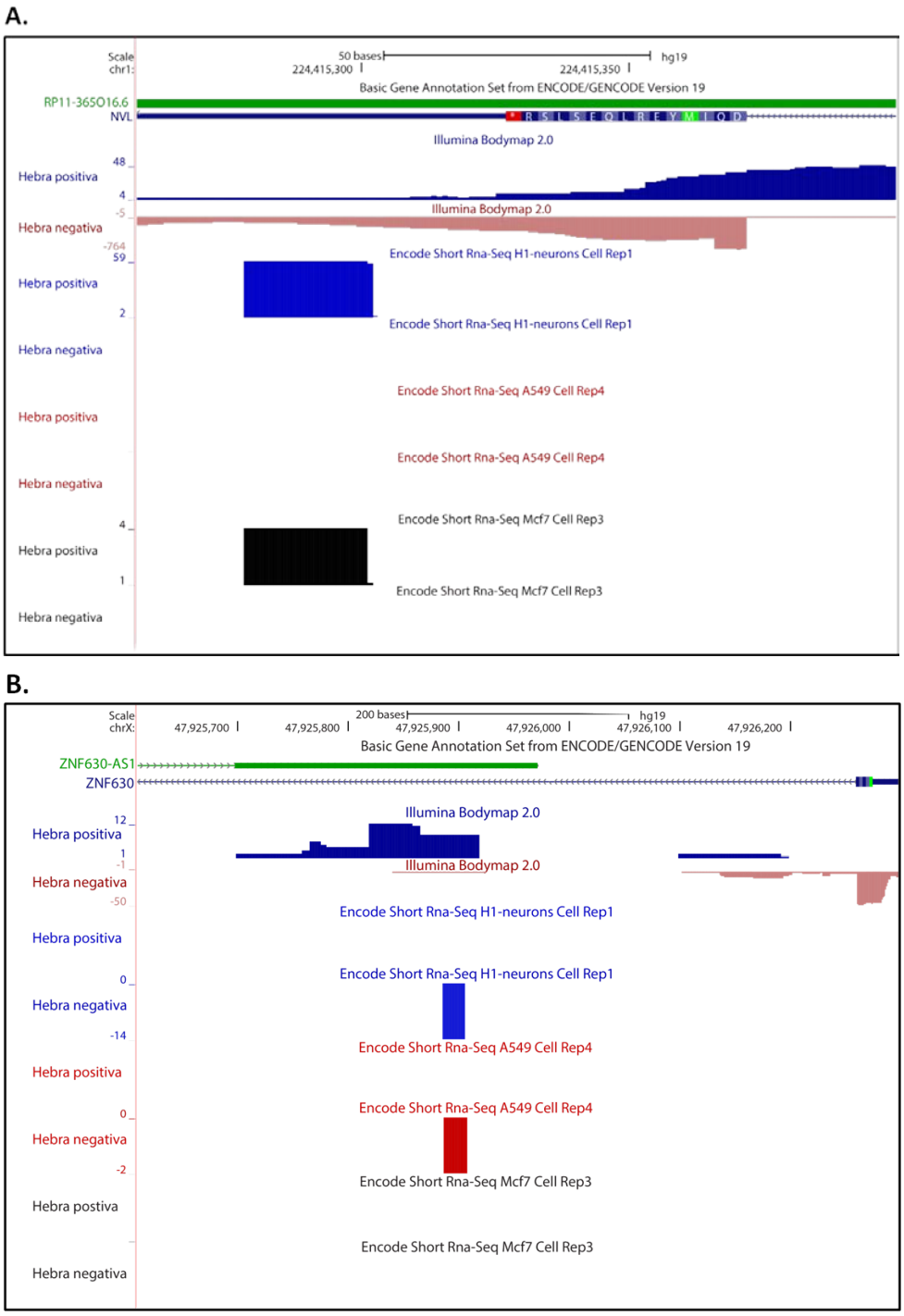


Figura 10. Ejemplos de endo-siRNAs identificados utilizando datos de RNA-Seq de RNAs pequeños del proyecto ENCODE. (A) Imagen del UCSC Genome Browser que muestra la anotación del gen NVL y de su NAT (RP11—365O16.6). La señal del alineamiento de datos de RNA-Seq de IBM 2.0 muestra la existencia de un NAT en esa región (hebra positiva en color azul). Los alineamientos de datos de RNA-seq de RNAs pequeños muestran la existencia de un posible endo-siRNA en células H1-neurons y MCF7. **(B)** Se muestra el gen ZNF630 y su NAT ZNF630-AS1. Los datos de RNA-Seq de IBM 2.0 confirman la existencia del NAT (hebra positiva en color azul). Se detectó un posible endo-siRNA en células H1-neurons y en células A549.

2. Validación experimental de los mirrorRNAs

La búsqueda bioinformática realizada apoya la existencia de mirrorRNAs en el transcriptoma humano. Para estar seguros de la existencia de estos RNAs se hizo necesaria la validación experimental de las predicciones mediante diversas aproximaciones experimentales. Para validar los mirrorRNAs se utilizaron técnicas de alta sensibilidad y a la vez que fueran hebra específica. Por ello se seleccionaron el 3' *Rapid amplification of cDNA ends* (3'RACE), el 5´ *Rapid amplification of cDNA ends* (5' RACE) y una modificación del RPA al cual se le acopló un paso final mediante RT-PCR.

2.1 3' RACE

La técnica de 3' RACE permite identificar el extremo 3' de un RNA de forma sensible y hebra específica. En la Figura 11 se muestra un esquema del procedimiento de 3' RACE. A partir de los análisis bioinformáticos se seleccionaron 9 genes que poseían mirrorRNAs predichos para ser validados: UBR5, CWC15, KRT8, RPS23, RPS11, EIF3D, DDX5, SIDT2 y CALR. Además se seleccionó BCMA que fue el primer mirrorRNA descrito en la literatura (Laabi et al., 1994).

Para realizar el 3' RACE se utilizó una mezcla comercial de RNA humano proveniente de múltiples órganos, denominado RNA Universal de humano. Además a la reacción de transcripción inversa se le agregó Actinomicina D (ActD), para inhibir la actividad DNA polimerasa dependiente de DNA de la enzima RT, que podría generar algún tipo de artefactos experimentales (Perocchi et al., 2007). Todos los mirrorRNAs que se buscaron mediante 3' RACE fueron analizados al menos dos veces, llegando en algunos casos a 6 veces. Después del 3' RACE se clonaron los productos de PCR en un vector pGEM-T y varios clones fueron secuenciados.

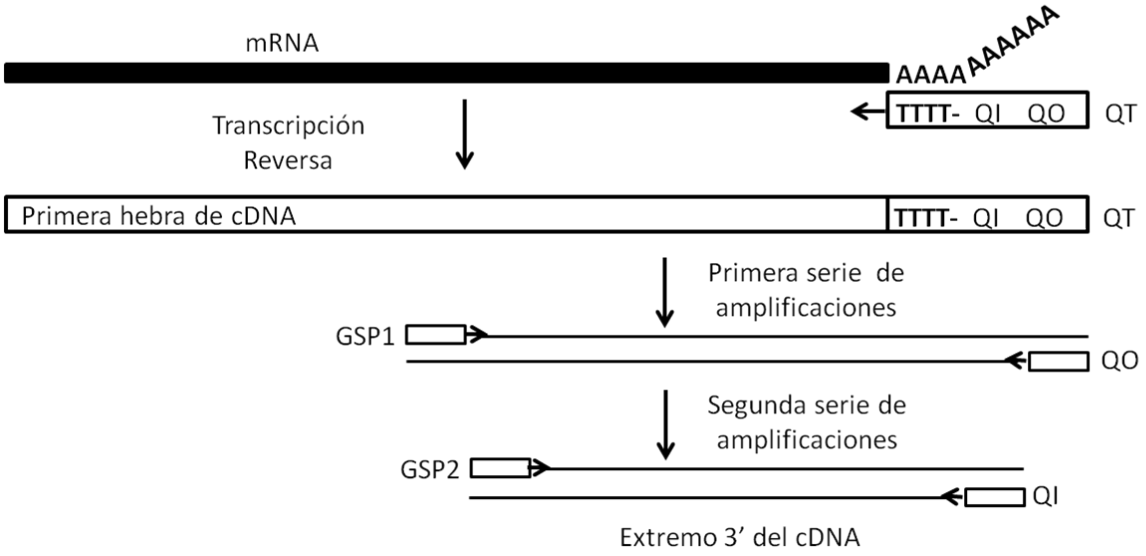


Figura 11. Esquema del protocolo del 3' RACE. Para la validación de mirrorRNAs mediante 3' RACE se utilizó RNA Universal de Humano. QT, QI y QO son los partidores utilizados para la reacción de 3' RACE y sus secuencias se describen en la Tabla I. GSP1 y GSP2 hacen referencia al *Gene Specific Primer* 1 y 2, las secuencias de estos partidores para cada gen se detallan en la Tabla I.

Para los 3’ RACE de los putativos mirrorRNAs de SIDT2 y EIF3D sólo se obtuvieron amplificaciones inespecíficas (datos no mostrados). El 3’ RACE del putativo mirrorRNA de BCMA amplificó NATs provenientes del *locus* del gen, pero ninguno era un mirrorRNA (datos no mostrados).

En el caso del 3’ RACE de CWC15 y RPS23 se lograron amplificar mirrorRNAs como se muestra en la Figura 12A y 12B. Para el caso de RPS23 se detectaron mirrorRNAs con dos extremos 3’.

En el caso de RPS11 se logró amplificar un mirrorRNA como se muestra en la Figura 12C. Sin embargo, la secuencia tiene la característica de que posee un tramo de poli(A) genómico en el extremo 3’ obtenido. Esto sugiere la existencia de este mirrorRNA pero no asegura que ese sea el extremo 3’, ya que se podría haber generado un *mis priming* en el tramo de poli(A) que se muestra en la figura en vez de en la cola poli(A) auténtica.

Los resultados de los 3’ RACE para los mirrorRNAs de DDX5 y de CALR, son especiales, ya que se detectó la presencia de un posible NAT que no posee el ultimo intrón de los genes (como sería el caso de un mirrorRNA en que el SJ del NAT y del mRNA es exactamente el mismo), pero que posteriormente tiene secuencias intrónicas que terminan en un poli(A) que está presente en el genoma (Figura 13A-B). Estos dos casos detectados serían un subtipo especial de mirrorRNAs en donde sólo una parte es perfectamente complementaria a mRNAs sentido, y otra parte tendría secuencias intrónicas.

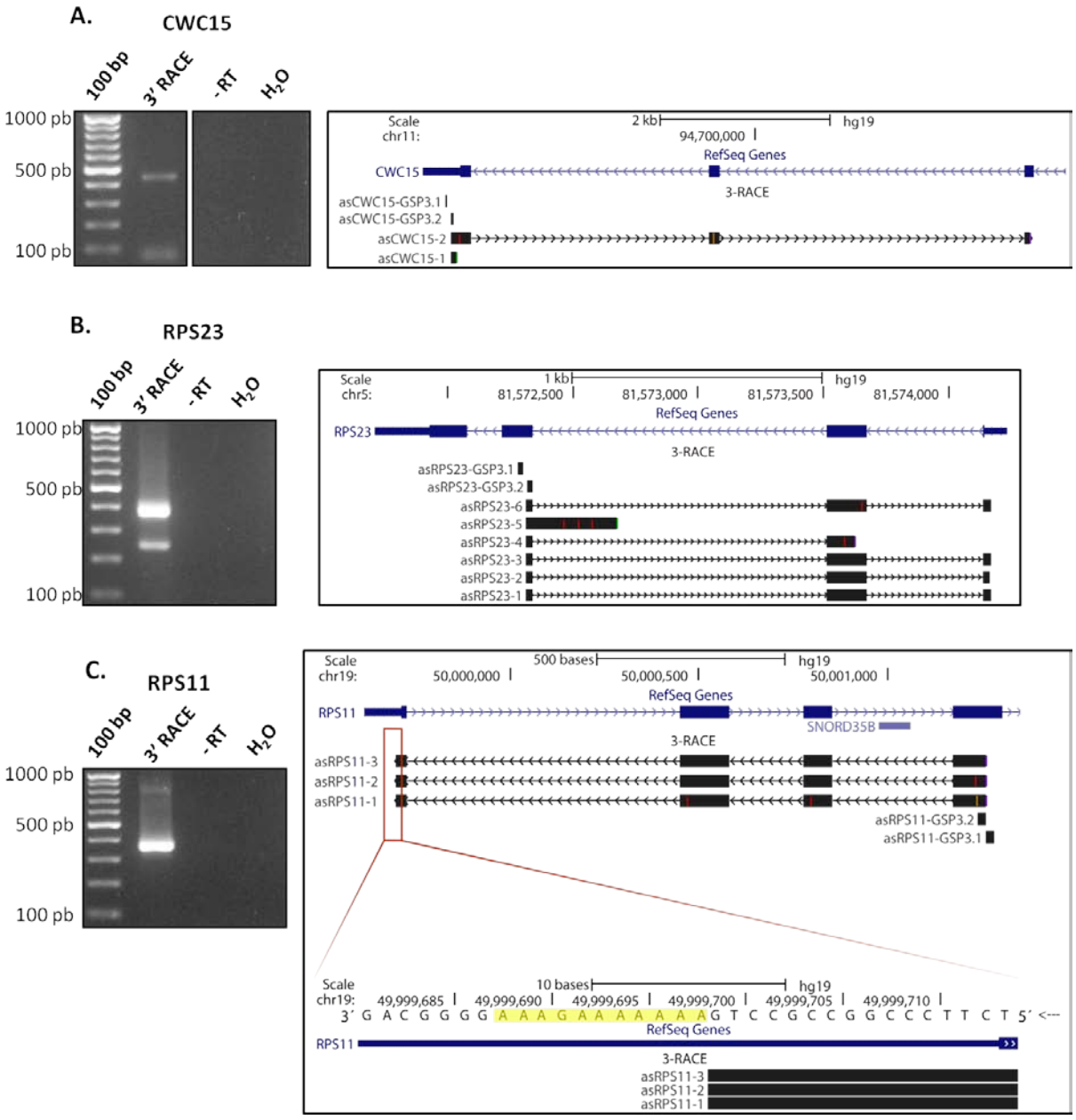


Figura 12. Los mirrorRNAs de CWC15 y RPS23 fueron validados mediante 3' RACE. (A) Resultado del 3' RACE para amplificar un mirrorRNA de CWC15. Los controles utilizados fueron no utilizar enzima durante la reacción de RT (-RT) y utilizar agua en vez de cDNA durante los PCRs anidados (H_2O). El producto del 3' RACE fue clonado en un vector pGEM-T y algunos clones fueron secuenciados. Se muestra una imagen del UCSC Genome Browser con el alineamiento de las secuencias obtenidas. **(B)** Resultado del 3' RACE de RPS23 y alineamiento de las secuencias obtenidas. **(C)** Resultado del 3' RACE de RPS11 y alineamiento de las secuencias obtenidas. En amarillo se destaca un tramo de poli(A) genómico que está justo río abajo del extremo 3' obtenido por 3' RACE. Las líneas rojas en los bloques de alineamiento indican *mismatches* entre las secuencias obtenidas y el genoma.

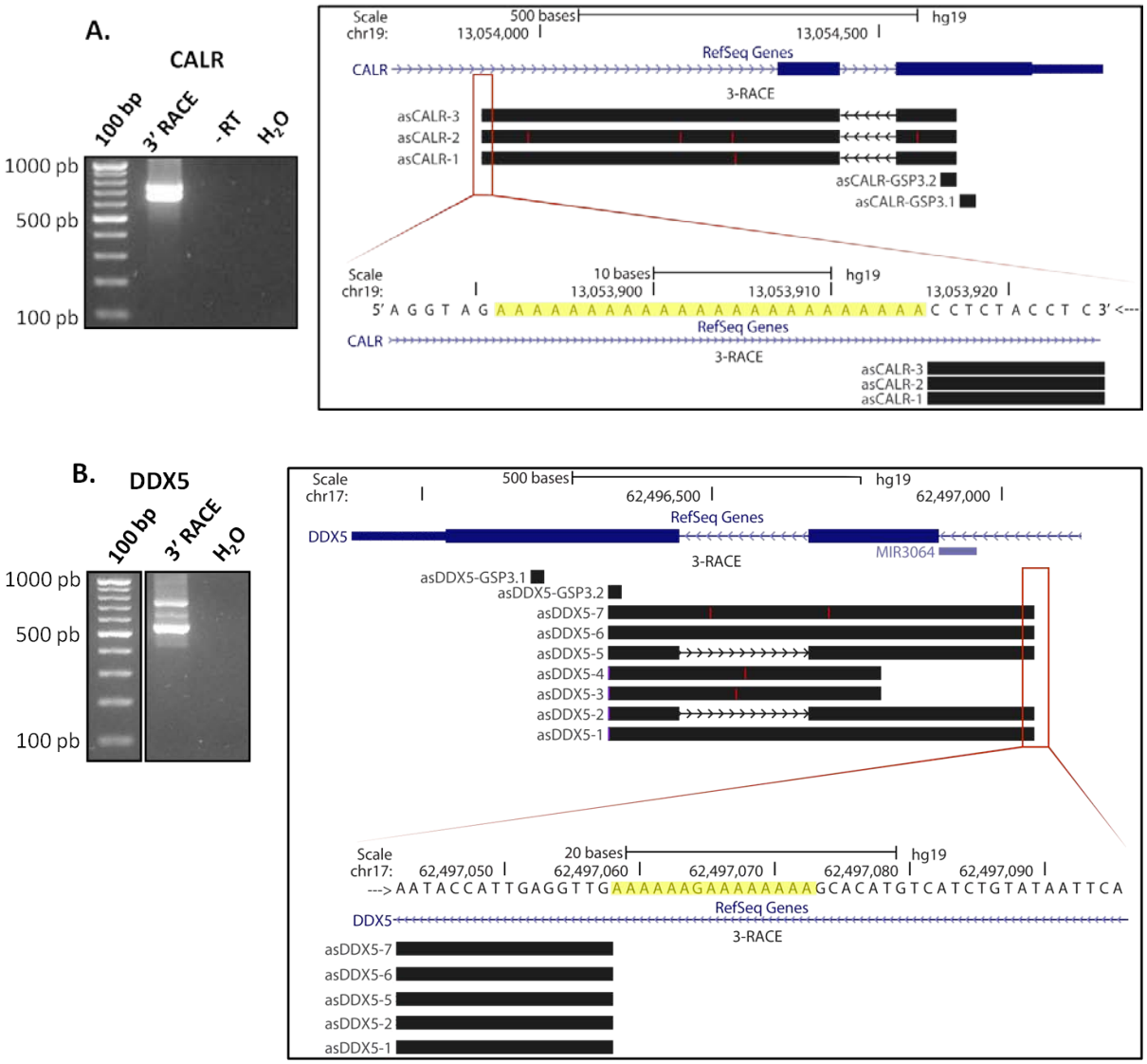


Figura 13. Mediante 3' RACE se identificaron mirrorRNAs parcialmente complementarios a mRNAs sentido para los genes CALR y DDX5. Resultado del 3' RACE para amplificar un mirrorRNA de CALR **(A)** y de DDX5 **(B)** junto al alineamiento de las secuencias obtenidas. En amarillo se destaca un tramo de poli(A) genómico.

2.1.1 Detección de pseudo-mirrorRNAs mediante 3’ RACE

En la búsqueda bioinformática de mirrorRNAs se detectaron casos de genes que tenían mirrorRNAs y además pseudo-mirrorRNAs, como es el caso del gen KRT8. De hecho KRT8 posee múltiples pseudogenes procesados identificados en el genoma humano (34 según la anotación de GENCODE v7). En el caso de KRT8 se usaron dos grupos de partidores distintos para el 3’ RACE, por un lado QT, QO y QI (Fig. 11) pero además AP (el partidor análogo de QT) y AUAP (el partidor análogo de QO y QI). Interesantemente con los distintos partidores se obtuvieron distintos resultados. Al utilizar AP y AUAP se amplificó un mirrorRNA de KRT8 (Fig. 14A). En cambio al utilizar los partidores QT, QO y QI, se amplificaron pseudo-mirrorRNAs de KRT8 (Fig. 14B). Estos resultados experimentales confirman lo encontrado mediante los análisis bioinformáticos y apoyan la posibilidad de que algunos mirrorRNAs provengan de la transcripción antisentido de pseudogenes procesados.

2.1.2 3’ RACE a RNA poliadenilado *in vitro*

Se realizaron repetidamente 3’ RACE con partidores específicos para un posible mirrorRNA de UBR5 y no hubo amplificación positiva. Uno de los problemas del 3’ RACE que utilizamos, es que sólo funciona con RNAs poliadenilados (Fig. 11), y cabe la posibilidad de que los mirrorRNAs puedan en algunos casos ser RNAs no poliadenilados. Para dar cuenta de esta posibilidad el RNA Universal de Humano fue poliadenilado *in vitro*, utilizando una poli(A) polimerasa. Al realizar el 3’ RACE con el RNA poliadenilado *in vitro* se amplificó un mirrorRNA de UBR5 (Fig. 15). Estos resultados sugieren que no siempre los mirrorRNAs son transcritos poliadenilados.

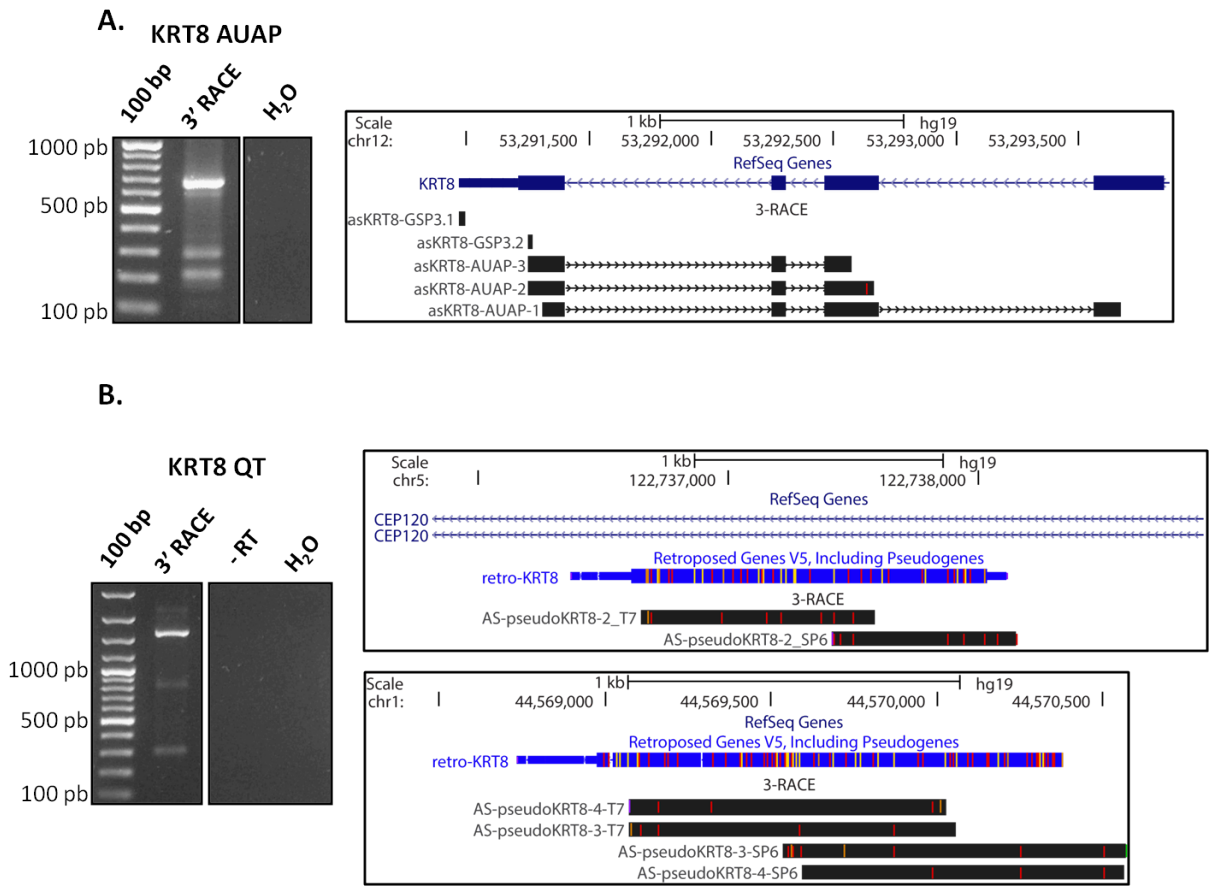


Figura 14. Mediante 3' RACE se validó la existencia de un mirrorRNA y un pseudo-mirrorRNA de KRT8. (A) Resultado del 3' RACE para amplificar un mirrorRNA de KRT8 utilizando el partidor AUAP durante el RT. El alineamiento de las secuencias obtenidas muestra que se amplificaron mirrorRNAs de KRT8. **(B)** Resultado del 3' RACE para amplificar un mirrorRNA de KRT8 utilizando el partidor QT durante el RT. Las secuencias obtenidas alinean a pseudogenes procesados de KRT8, por tanto corresponderían a pseudo-mirrorRNAs.

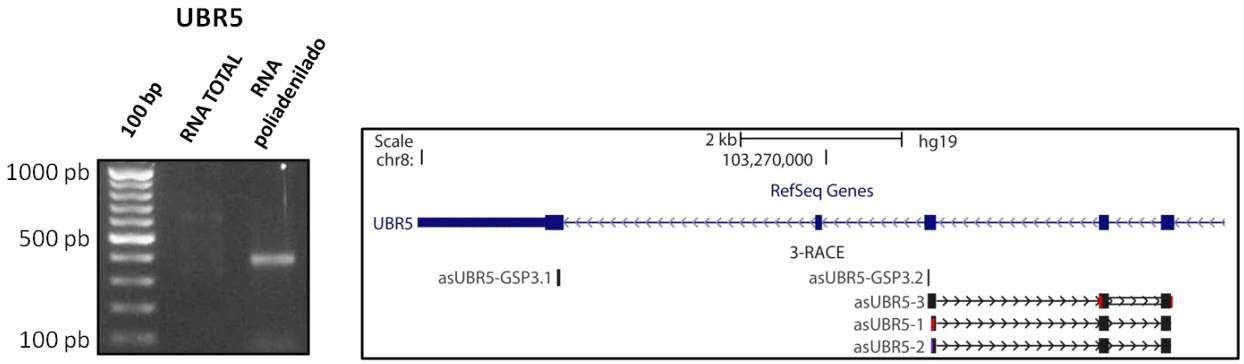


Figura 15. El mirrorRNA de UBR5 en RNA Universal de Humano sólo se pudo amplificar mediante 3' RACE al poliadenilar *in vitro* el RNA. Resultado del 3' RACE para amplificar un mirrorRNA de UBR5 usando RNA Universal de Humano y el mismo RNA poliadenilado *in vitro*. Las secuencias amplificadas utilizando RNA poliadenilado corresponden a mirrorRNAs de UBR5.

2.2 5’ RACE

Además de la utilización de la técnica de 3’ RACE, se utilizó 5’ RACE. Esto permite conocer el extremo 5’ de los transcritos analizados. El 5’ RACE posee la ventaja por sobre el 3’ RACE que funciona tanto con RNAs poli(A)+ como poli(A)-. Se han descrito distintos protocolos de 5’ RACE y en esta tesis se utilizaron dos tipos de protocolos (Fig. 16A-B).

2.2.1 5’ RACE utilizando TdT y dCTP

En primer lugar se realizó un 5’ RACE “clásico”, cuyo protocolo está esquematizado en la Figura 16A. Para este experimento se utilizó RNA universal de humano.

Se seleccionaron los genes KRT8, CWC15, UBR5 y BCMA para validar sus putativos mirrorRNAs. Para BCMA y CWC15 no se detectaron mirrorRNAs (datos no mostrados). Para UBR5 se detectaron varios extremos 5’ que corresponden a mirrorRNAs (Fig. 17A). Nuevamente para KRT8 se encontraron mirrorRNAs y uno de los clones que se amplificó fue de un NAT proveniente de un pseudogen procesado (Fig. 17B).

2.2.2 5’-RLM-RACE

Además de utilizar la técnica de 5’ RACE clásica, se utilizó la técnica de “*RNA Ligase Mediated Rapid Amplification of cDNA Ends*” (RLM-RACE), cuyo protocolo se esquematiza en la Figura 16B. Para este procedimiento se utilizó RNA proveniente de células MCF7.

Se seleccionaron los genes UBR5, CWC15, BCMA, EIF3D, DDX5 y CALR. Para UBR5 y BCMA no se obtuvo una amplificación exitosa. Para CWC15, EIF3D, DDX5 y CALR se obtuvieron secuencias que provenían de mirrorRNAs (Fig. 18A-D).

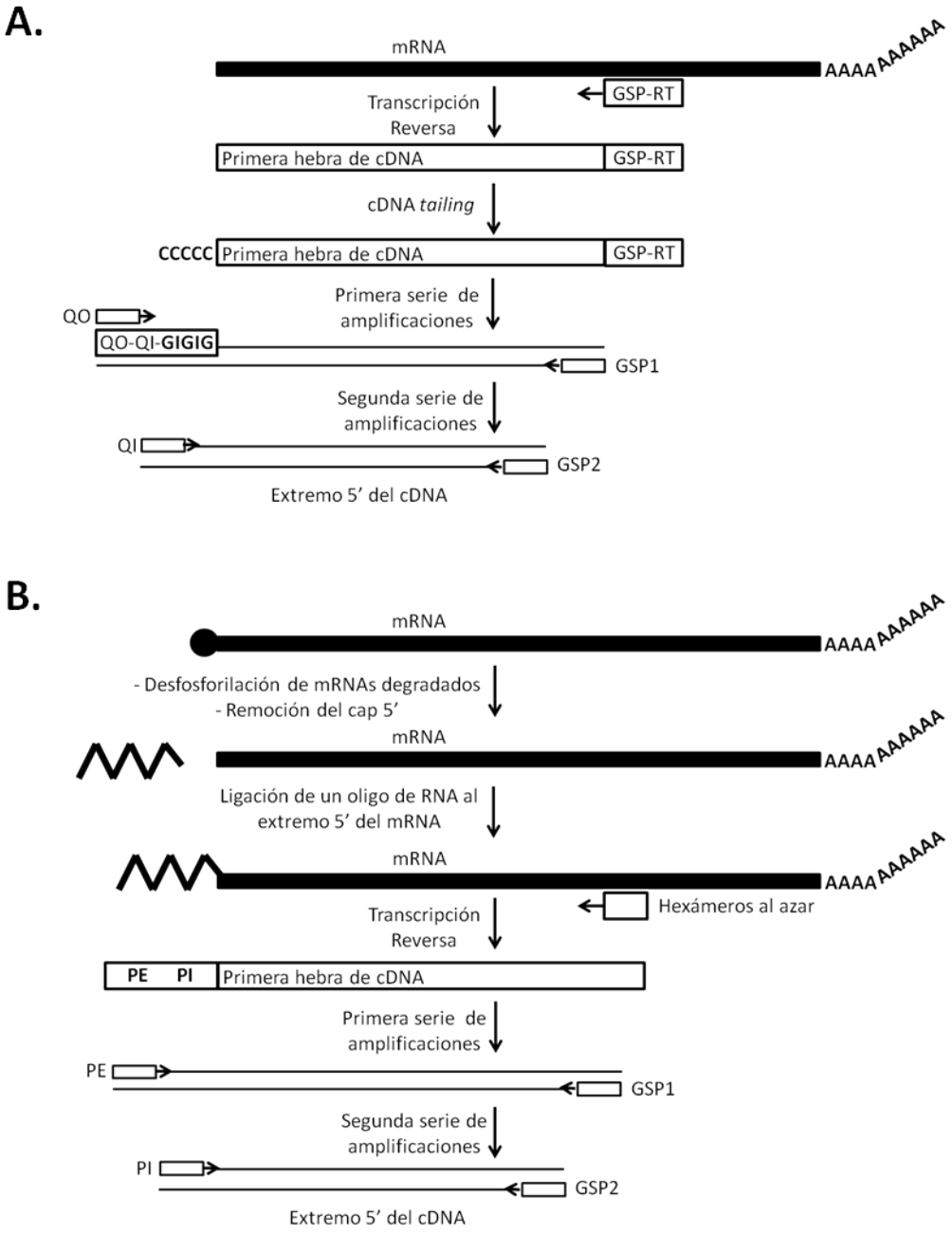


Figura 16. Esquema que muestra los dos protocolos de 5' RACE utilizados. (A) Protocolo de 5' RACE "clásico" en el que se realiza un "cDNA *tailing*" utilizando TdT y dCTP. **(B)** Protocolo de 5' RLM RACE. Las secuencias de los partidores QO, QI, PE (5' RACE *Outer Primer*) y PI (5' RACE *Inner Primer*) están detalladas en la Tabla I.

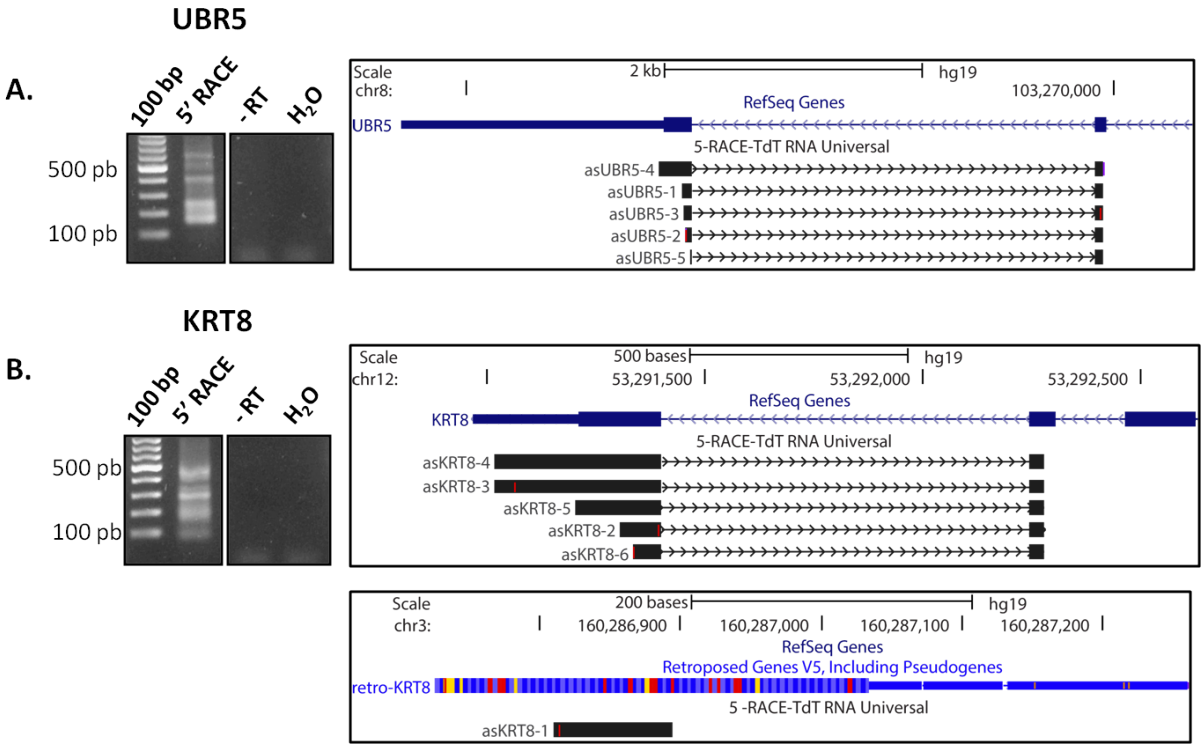


Figura 17. Validación experimental de los mirrorRNAs de los genes UBR5 y KRT8 mediante 5' RACE clásico. (A) Resultado del 5' RACE para amplificar un mirrorRNA de UBR5 utilizando RNA Universal de Humano. Se secuenciaron múltiples clones que corresponden a mirrorRNAs de UBR5. **(B)** En el 5' RACE de KRT8 se detectaron mirrorRNAs y pseudo-mirrorRNAs.

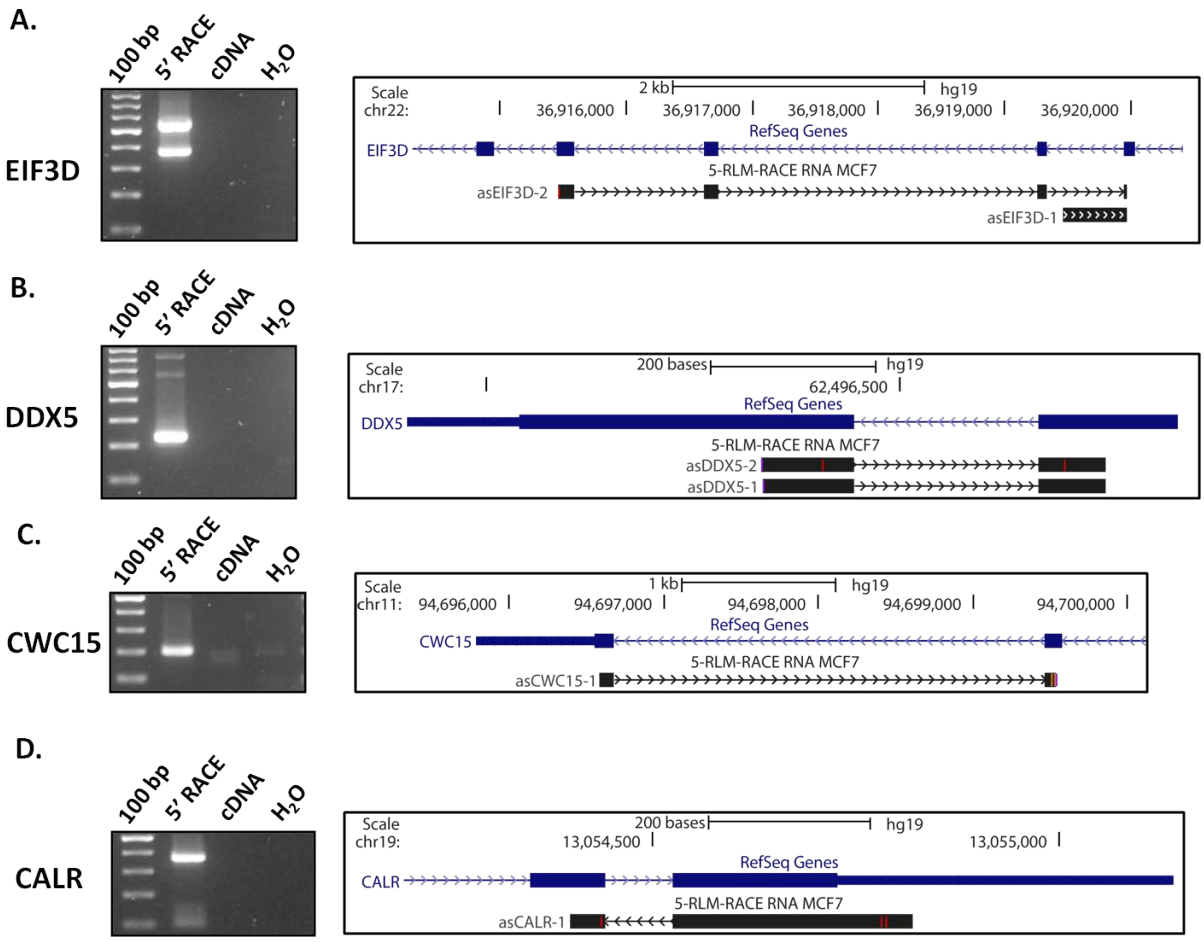


Figura 18. Validación experimental de los mirrorRNAs de los genes EIF3D, DDX5, CWC15 y CALR mediante 5'-RLM RACE. **(A-D)** Resultados del 5' RLM RACE utilizando RNA de células MCF7. Los controles utilizados fueron cDNA al cual no se le ligó el adaptador de RNA y agua en vez de cDNA durante los PCRs anidados (H_2O).

2.3 Validar mirrorRNAs por ensayo de protección de ribonucleasas (RPA)

Otra forma de detectar la presencia de los mirrorRNAs fue la utilización de RPA. Normalmente en esta técnica se utiliza una sonda radioactiva que es complementaria al RNA que se quiere detectar y al formarse dobles hebras de RNA se protegen de la degradación de RNasas específicas para RNAs de simple hebra. Ya que los mirrorRNAs son perfectamente complementarios a los mRNAs sentidos respectivos se realizó un RPA sin una sonda complementaria al RNA que se quiere detectar. El RNA resultante del ensayo de RNasas fue utilizado para realizar RT-PCR y poder detectar el RNA que se protegió de la degradación. Para el procedimiento se utilizó RNA proveniente de células HEK293T.

Se intentaron amplificar los mirrorRNAs predichos: UBR5, RPS11, RPS23, CWC15, CALR y DDX5. Como control negativo se utilizó ACTB. Como se aprecia en la Figura 19, para el caso de UBR5, RPS11 y RPS23 hay amplificación en el RNA tratado con RNAsas. Esto sugiere la presencia de los mirrorRNA que permitieron la protección del fragmento respectivo de la acción de las RNAsas.

En conclusión, a partir de todos los ensayos realizados se obtuvo evidencia que apoya la existencia de mirrorRNAs para 8 de los 10 genes que se sometieron a validación experimental.

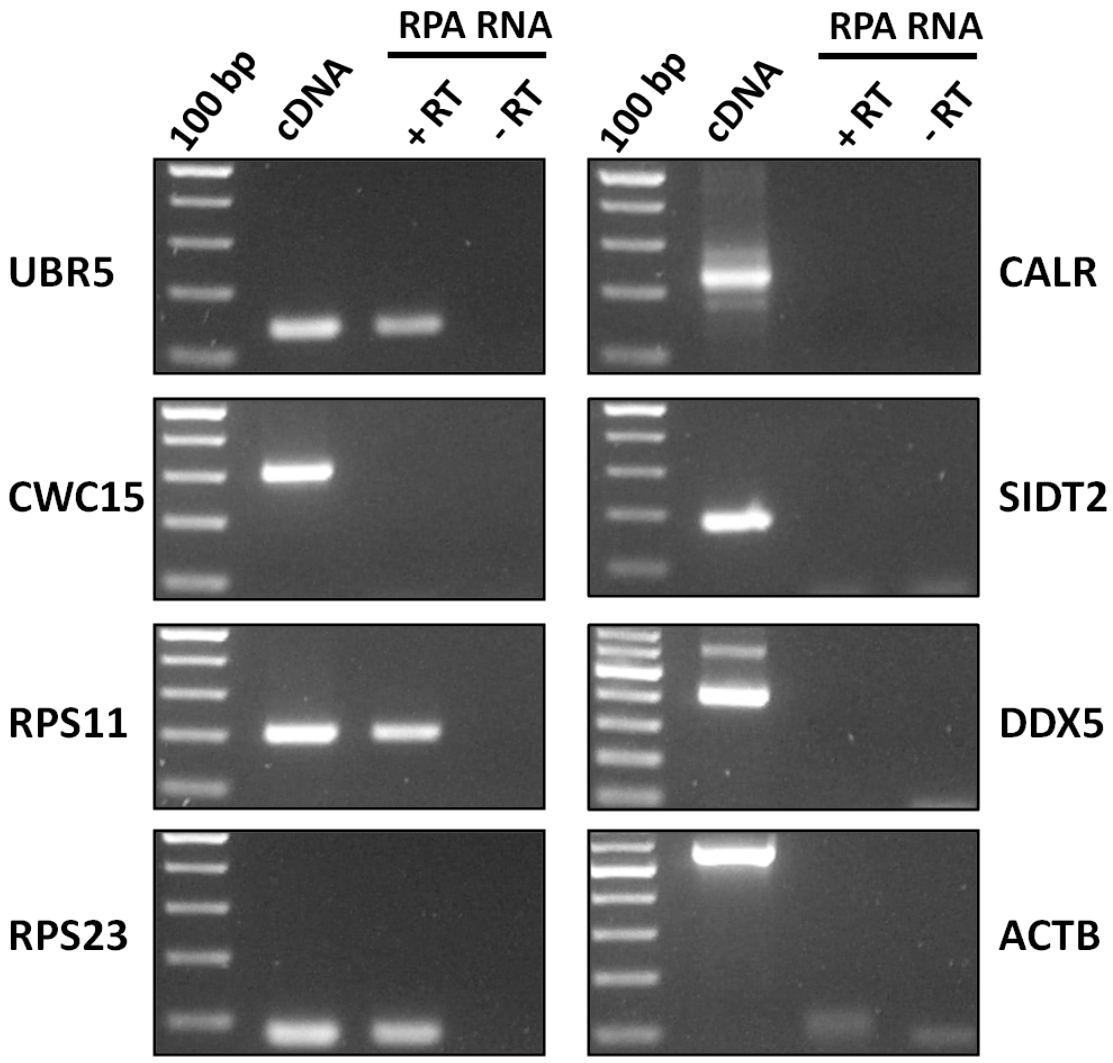


Figura 19. Validación de mirrorRNAs mediante RPA-RT-PCR. PCRs con partidores específicos para los genes indicados, utilizando cDNAs generados desde RNA total de HEK293T (cDNA) y RNA total tratado con RNasas que degradan RNA simple hebra (RPA RNA). El control –RT consiste en RNA tratado con RNasas el que fue tratado con todos los componentes de una reacción de RT pero no se le adicionó la transcriptasa inversa.

3. Biogénesis de los mirrorRNAs

Como ya se describió en la introducción, la existencia de los mirrorRNAs ha sido explicada mediante dos mecanismos en la literatura. La primera explicación plantea que los mirrorRNAs son transcritos desde el mismo *locus* que el mRNA sentido y que sufren splicing en los sitios complementarios e inversos (CT-AC) a los sitios canónicos. La otra explicación que se ha dado es que los mirrorRNAs provengan de la copia de un mRNA sentido por una RdRP. Por último se ha planteado en esta tesis una tercera posibilidad y es que provengan de la trascripción antisentido de pseudogenes procesados.

Por tanto la pregunta es ¿cuál de estos mecanismos planteados podría ser el responsable de la existencia de los mirrorRNAs?

3.1 Búsqueda de intrones CT-AC

Si existiera un RNA que al ser alineado al genoma mostrara evidencia de la remoción de un intron GT-AG y además de un intron CT-AC, esto apoyaría la hipótesis de que los intrones CT-AC pueden ser removidos. Bajo la hipótesis de una RdRP que copie un mRNA sentido, el mirrorRNA sólo debería mostrar evidencia de intrones CT-AC removidos.

3.1.1 Búsqueda de intrones mirror con intrones GT-AG

Definimos intrones mirror, como intrones que poseen los sitios de splicing CT-AC en las mismas coordenadas que las de un intrón canónico GT-AG, pero en la hebra complementaria. Para buscar intrones mirror se utilizaron datos de alineamientos de ESTs y cDNAs humanos. En particular se buscaron secuencias que tuvieran intrones mirror y además al menos un intron GT-AG.

En la búsqueda se encontraron 23 secuencias (Tabla VII). Todas las secuencias encontradas provienen de *loci* que poseen pares de transcritos sentido y antisentido. Las secuencias que poseen intrones mirror e intrones GT-AG parecieran ser la fusión de transcritos sentido y antisentido. En la Figura 20 se muestran dos ejemplos: BCL10 (AF082283 y AJ006288) y ZNF165 (AY083664). En los dos casos el primer intron de las secuencias encontradas es un intron mirror y el resto de los intrones son del tipo GT-AG.

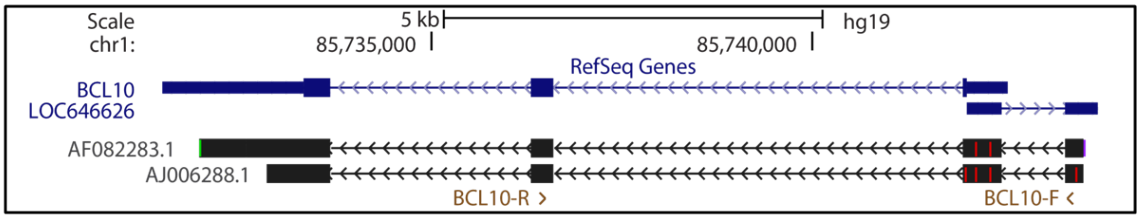
Los ejemplos mostrados en la Figura 20 fueron seleccionados para realizar una validación experimental. Para ello se diseñaron partidores que pudieran amplificar las secuencias descritas (Fig. 20). Se hizo RT-PCR a partir de RNAs de humano proveniente de ocho tejidos distintos: útero, testículo, cerebro, corazón, bazo, pulmón, riñón e hígado. No se observó amplificación positiva en ninguno de los casos estudiados (datos no mostrados).

Los resultados de la validación experimental llevaron a pensar que las secuencias encontradas en las bases de datos podrían ser artefactos experimentales. Un análisis detallado mostró que los dos cDNAs descritos para BCL10 que poseen intrones mirror, provienen de dos publicaciones en donde en ningún momento amplifican por completo los cDNAs que ellos reportaron (Willis et al., 1999; Zhang et al., 1999). Extrapolaron las secuencias reportadas las de ESTs, uniendo equivocadamente ESTs sentido con antisentido. Las secuencias con intrones mirror podrían haberse generado por artefactos producidos en el proceso de creación de las bibliotecas de cDNAs. Esto podría ocurrir si el cDNA sentido y el cDNA antisentido se unen apareándose en la zona complementaria y posteriormente la RT podría generar un cDNA doble hebra que contenga una fusión sentido y antisentido. Se ha descrito que la RT puede generar *in vitro* artefactos en donde se fusionan transcritos sentido y antisentido (Houseley y Tollervey, 2010).

Tabla VII. Secuencias con intrones mirror e intrones GT-AG.

Gen Sentido	Gen Antisentido	Secuencias
ARHGEF39	CCDC107	AY390226
ASF1A	MCM9	AF151856
ASF1A	MCM9	AF161495
BCL10	LOC646626	AF082283
BCL10	LOC646626	AJ006288
CC2D1A	C19orf57	AF536205
CTD-2616J11.16	VSIG10L	CF272609
DNAJB13	RP11-167N4.2	AF419291
EMC10	FAM71E1	AY194293
FLJ35024	VLDLR	AF424541
GINS4	RP11-360L9.7	BX362611
ITFG2	RP4-816N1.6	AF220048
MRPL47	ACTL6A	AF285120
MTIF3	GTF3A	AF265440
NOC2L	SAMD11	AF161376
NOP16	HIGD2A	AF151875
SIDT2	LOC100652768	AF151799
SPCS1	GLT8D1	AX775785
TMOD3	RP11-56B16.4	AF237631
TRMT2A	RANBP1	AW248315
UNC119B	RP11-173P15.5	AK126367
WWTR1	WWTR1-AS1	AJ299431
ZNF165	RP1-313I6.12	AY083664

A.



B.

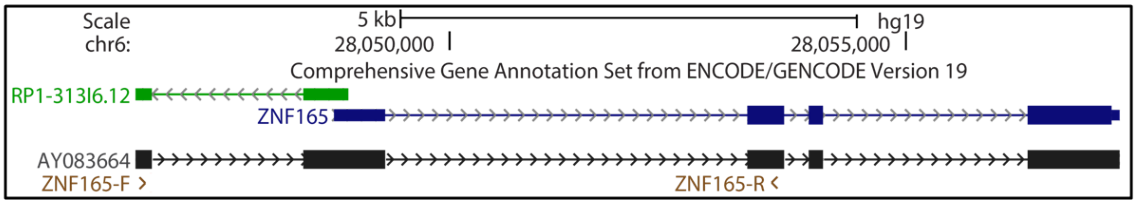


Figura 20. Ejemplos de cDNAs con intrones mirror e intrones GT-AG. (A) Se identificó la secuencia AF082283 como un posible RNA con intrones GT-AG y un intron mirror (CT-AC) que coincide con las coordenadas del intron anotado para la secuencia de RefSeq LOC646626. **(B)** La secuencia AY083664 posee un intrón mirror que coincide con las coordenadas anotadas para la secuencia anotada por GENCODE RP1-313I6.12. En **A y B** se muestran en café la ubicación de los partidores utilizados para realizar una validación experimental de estos RNAs.

3.1.2 Búsqueda de intrones CT-AC en el transcriptoma de ratón y humano

¿Es posible que la maquinaria celular procese los sitios de splicing complementarios e inversos a los canónicos? Para poder responder esta pregunta se realizó una búsqueda de intrones CT-AC. No se restringió esta búsqueda a intrones mirror sino que se efectuó tomando en cuenta todos los posibles intrones CT-AC existentes. Preliminarmente se realizó una búsqueda de intrones CT-AC desde la base de datos ASPICDB (Castrignano et al., 2008). Se buscaron intrones apoyados por al menos dos secuencias (ESTs y/o cDNAs) y que dentro del EST o cDNA además existiera al menos un intrón GT-AG (esto es para evitar tomar cDNAs o ESTs que sólo tengan intrones CT-AC que provengan de secuencias alineadas en la dirección equivocada). Bajo estas condiciones en humano se encontraron 102 intrones y en ratón se encontraron 32 intrones. Producto de estos resultados iniciales se decidió realizar una búsqueda más acabada de intrones CT-AC en los ESTs y cDNAs de humano y ratón.

Desde el UCSC Table Browser se obtuvieron los alineamientos de todos los cDNAs y ESTs de ratón y humano producidos con BLAT (Karolchik et al., 2004). A los datos se les aplicaron los diversos filtros detallados en métodos.

Como ya se comentó anteriormente existe una gran cantidad de ESTs que están anotados en la dirección inversa (todos sus intrones son CT-AC) (Tabla II). Debido al problema de la anotación de los ESTs se ideó una forma de corregir la dirección en la que están anotados. Se utilizó la siguiente fórmula para cada EST y cDNA:

(Intrones GT-AG + GC-AG) – (intrones CT-AC + CT-GC)= Orientación de la secuencia.

Si el número obtenido era mayor o igual a cero no se cambió la dirección de las secuencias. Pero si el resultado era menor a cero, se corrigió la dirección.

Los resultados de las predicciones de los intrones utilizando los datos con la orientación corregida están resumidos en la Tabla VIII. Los resultados fueron llamativos por el alto porcentaje de intrones con dinucleótidos no canónicos. Para el caso de humano los intrones AT-AC (que pertenecen al grupo de intrones del tipo U12) se encuentran recién en el 13° lugar en el ranking de dinucleótidos más frecuentes. Los putativos intrones CT-AC se encuentran en el lugar 53° en el ranking de dinucleótidos más frecuentes para intrones humanos según los datos procesados por BLAT. En el caso de ratón los posibles intrones CT-AC se ubican en el lugar 42°.

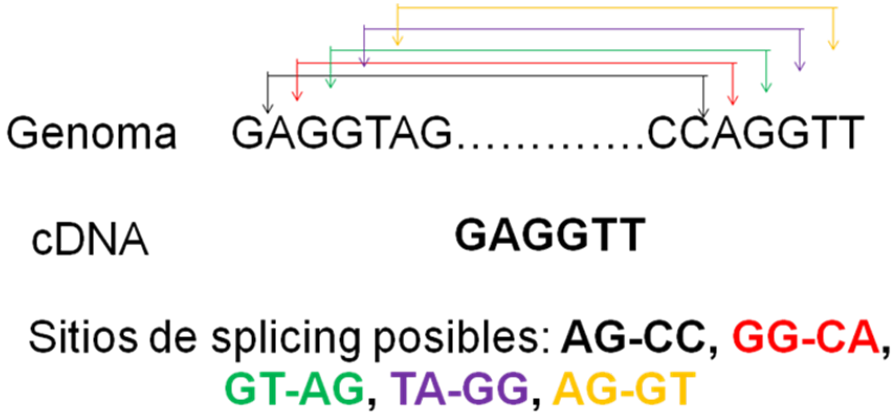
Los intrones CT-AC predichos por BLAT poseen un alto número de secuencias directas repetidas en los sitios de splicing (Fig. 21). Los intrones GT-AG predichos por BLAT poseen un promedio de secuencias directas repetidas en sus sitios de splicing de 2,3 nt (la mediana es 2) En cambio los intrones CT-AC predichos por BLAT poseen secuencias directas repetidas de 6,3 nt (la mediana es 5). La presencia de largas secuencias directas repetidas en los sitios de splicing dificulta el correcto alineamiento de las secuencias al genoma (Fig. 21A). La inspección manual de los alineamientos de algunos intrones CT-AC predichos por BLAT reveló que eran errores (Fig. 21B).

Debido al alto número de intrones no canónicos predichos por BLAT y a la existencia de predicciones de intrones CT-AC producto de errores de alineamiento, se realizó una nueva predicción de intrones utilizando el programa GMAP (Wu y Watanabe, 2005). Este programa está optimizado para encontrar sitios de splicing utilizando ESTs y cDNAs. Después de alinear los datos se utilizaron los

TABLA VIII. Dinucleótidos de los intrones predichos por BLAT a partir de datos de ESTs y cDNAs de humano y ratón.

Humano			Ratón		
Dinucleótido	**Número**	**%**	**Dinucleótido**	**Número**	**%**
GTAG	255.480	92,19	GTAG	212.261	95,6
GCAG	3.658	1,32	GCAG	2.372	1,07
GGAG	956	0,34	GGAG	519	0,23
GTGG	788	0,28	GTGG	411	0,19
GAAG	476	0,17	ATAC	217	0,1
GTTG	412	0,15	GTTG	193	0,09
GTCA	398	0,14	GAAG	182	0,08
CTAG	339	0,12	GTCA	175	0,08
TGAG	325	0,12	TGAG	145	0,07
GTGC	315	0,11	GTGT	142	0,06
NNNN	13.874	5,01	NNNN	5.354	2,41
CTAC	107	0,04	CTAC	49	0,02

A.



B.

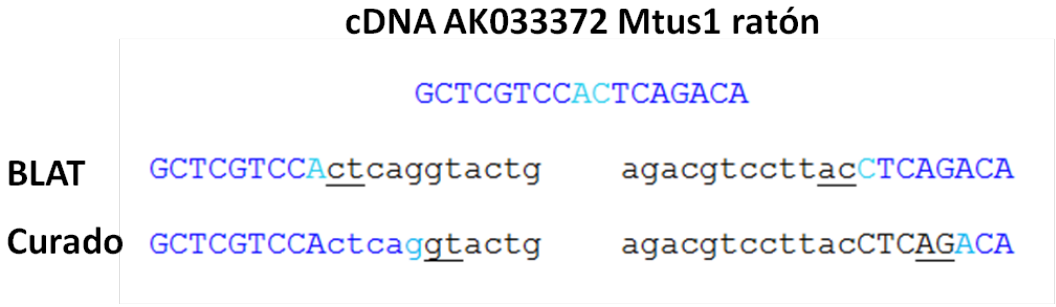


Figura 21. Secuencias directas repetidas en los SJs dificulta el correcto alineamiento de ESTs y cDNAs al genoma. (A) Los sitios de splicing ambiguo dificultan la posibilidad de asignar con certeza la secuencia de los sitios de splicing, en particular la de los intrones no consenso. En el ejemplo existen cinco alineamientos posibles debido a la existencia de una secuencia repetida directa de 4 nt (AGGT). **(B)** Error de alineamiento de BLAT que llevó a la incorrecta identificación de un intrón CT-AC.

TABLA IX. Dinucleótidos de los intrones predichos por GMAP a partir de datos de ESTs y cDNAs de humano y ratón.

Humano			Ratón		
Dinucleótido	**Número**	**%**	**Dinucleótido**	**Número**	**%**
GTAG	267.636	96,321	GTAG	214.253	97,934
GCAG	3.850	1,386	GCAG	2.355	1,076
ATAC	360	0,130	ATAC	256	0,117
GGAG	261	0,094	GGAG	235	0,107
AGAG	140	0,050	TGAG	47	0,021
GTGG	140	0,050	GTGC	36	0,016
TGAG	104	0,037	AGAG	35	0,016
GTGC	100	0,036	CCGG	35	0,016
CCGG	93	0,033	GTAT	34	0,016
GTAT	92	0,033	GTTG	32	0,015
NNNN	5.060	1,821	NNNN	1.450	0,663
CTAC	21	0,008	CTAC	4	0,002

mismos filtros utilizados para los resultados de BLAT. En la Tabla IX se resumen las predicciones de sitios de splicing para humano y ratón. Se observa que el número de intrones no canónicos es menor que el predicho por BLAT. El número de intrones CT-AC también es menor al predicho por BLAT. Los intrones CT-AC están en el lugar 99° en humano y 141° en ratón en la lista de dinucleótidos más frecuentes.

A continuación se tomaron las predicciones de sitios de splicing de BLAT y de GMAP y se realizó una intersección. En la Tabla X se resumen los resultados. Hay 13 intrones CT-AC en humano y 3 intrones CT-AC en ratón que están predichos por BLAT y por GMAP.

Luego se realizó un análisis para determinar si los intrones CT-AC predichos estaban conservados evolutivamente entre ratón y humano. Ninguno de los intrones CT-AC predichos está conservado entre ratón y humano.

El intrón CT-AC del gen Trappc9 parece ser una excepción, ya que posee múltiples ESTs y cDNAs que apoyan su existencia. Además está conservado entre rata y ratón. Y este intrón posee secuencias consenso del tipo U12. Un trabajo previo había reportado su existencia como un intrón U12 no canónico (Lin et al., 2010).

TABLA X. Dinucleótidos de los intrones predichos por BLAT y GMAP a partir de datos de ESTs y cDNAs de humano y ratón.

Humano			Ratón		
Dinucleótido	**Número**	**%**	**Dinucleótido**	**Número**	**%**
GTAG	252.599	97,850	GTAG	210.920	98,576
GCAG	3.392	1,314	GCAG	2.233	1,044
ATAC	252	0,098	ATAC	202	0,094
GGAG	137	0,053	GGAG	139	0,065
GTGC	73	0,028	GTAT	31	0,014
GTAT	69	0,027	GTTG	30	0,014
GTGG	67	0,026	GTGC	24	0,011
AATT	59	0,023	GTGG	22	0,010
GTTG	54	0,021	AGAG	13	0,006
AACT	50	0,019	CGAG	13	0,006
NNNN	1.383	0,536	NNNN	336	0,157
CTAC	13	0,005	CTAC	3	0,001

3.1.3 Validación experimental de los intrones no canónicos CT-AC

A partir de las predicciones realizadas de intrones CT-AC se seleccionaron algunos intrones para realizar una validación experimental mediante RT-PCR. Se seleccionaron seis genes de humano a los cuales se les predijo una variante de splicing con un posible intrón CT-AC: NEFH , RAB37, BGN, PRRT, LHX9 y SRCRB4D (en la Tabla XI se entregan detalles). Para ello, se utilizaron RNAs de humanos y se generó cDNA de 8 tejidos distintos (útero, testículo, cerebro, corazón, bazo, pulmón, riñón e hígado). Se realizaron PCRs mediante un protocolo de *touchdown* PCR a 40 ciclos (Korbie y Mattick, 2008). Se realizaron curvas de Mg^{2+}, curvas de temperatura de *annealing* y se utilizó un *enhancer* de PCR, pero no se encontró evidencia de un intrón CT-AC en los genes analizados (datos no mostrados).

A partir de las predicciones realizadas en los datos de ratón se seleccionaron 9 genes candidatos para ser validados por RT-PCR: 1700012B15Rik, 2900041M22Rik, Alb, Ftl1, Kcnk7, Mrps7, Nipsnap1, Rab2b y Rtkn (en la Tabla XI se entregan detalles). En cada caso se utilizó un par de partidores que permite amplificar las variantes de splicing con el intrón CT-AC procesado y sin procesar. Se utilizó un RNA comercial en el cual están representados todos los tejidos de ratón. Se realizaron las transcripciones inversas utilizando cuatro enzimas distintas: RevertAid (MMLV de Fermentas), AMV (Promega), Transcriptor (Roche), RevertAid Premium (Fermentas).

Se muestran los resultados de los RT-PCRs en la Figura 22. En todos los genes estudiados se amplificó la variante de splicing con el intrón CT-AC no procesado (Fig. 22A). Sólo se observa la banda esperada para el intrón CT-AC procesado (en una muy baja proporción) en el caso de Ftl1 (Fig. 22A y 22B). Las bandas de Ftl1 fueron corroboradas por secuenciación. Al repetir el PCR de Ftl1 con cDNAs generados con cuatro enzimas distintas, sólo se observa la variante procesada al utilizar la RT

Tabla XI. Intrones CT-AC elegidos para ser validados experimentalmente mediante RT-PCR.

Especie	Gen	Coordenadas	Tamaño pb	ESTs	cDNAs
Humano	PRRT	chr16:29824475-29824633	158	BP213249	AK098379
	SRCRB4D	chr7:76022807-76022869	62	BG682948,CV030141	BC015651,JF432453
	BGN	chrX:152770137-152770226	89	DB257240	AK094059
	RAB37	chr17:72741603-72742904	1301	-	BC040547
	LHX9	chr1:197886900-197890535	3635	-	AK097614
	NEFH	chr22:29867016-29868294	1278	-	AB020652
	NEFH	chr22:29868438-29876240	7802	-	AB020652
Ratón	2900041M22Rik	chr11:117473768-117474500	732	BB632596,BB632597	AK038849,AK038850
	1700012B15Rik	chr12:3237205-3237289	84	-	BC060685,BC062897
	KCNK7	chr19:5706904-5706987	83	-	AF012324,AF022820
	FTL1	chr7:52713416-52713575	159	BF322949,CB723511, DV043763, DV046339, DV053030,DV058728	-
	Alb	chr5:90893999-90901647	7648	AI265678,AI303622	-
	Mrps7	chr11:115468438-115468498	60	BY711401	AK012225
	Rab2b	chr14:52893466-52895031	1565	BB173372	AK038576
	Rtkn	chr6:83100022-83102131	2109	AI390876,W14579	-
	Nipsnap	chr11:4784049-4789925	5876	AI037258,BX511822	-

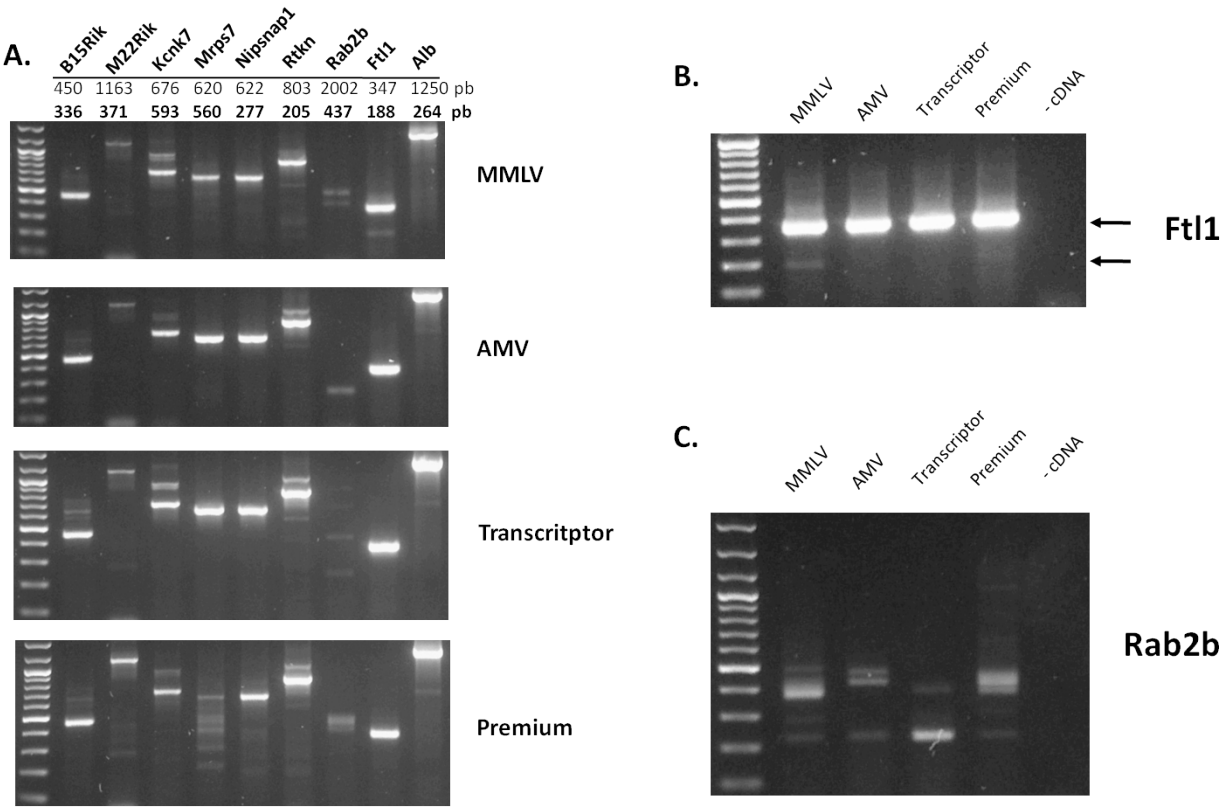


Figura 22. Los intrones CT-AC predichos en genes de ratón y sujetos a verificación experimental no pudieron ser validados. (A) RT-PCRs utilizando partidores que amplifican la variante con intrón CT-AC retenido y procesado. Se muestran los tamaños en pares de bases para los dos productos de PCR esperados. Se utilizó RNA Universal de Ratón y se generó cDNA utilizando cuatro RTs distintas. **(B)** La variante sin el intrón CT-AC del gen Ftl1 sólo se ve en el RT-PCR en el cual se utilizó MMLV. **(C)** Se amplificaron distintos productos de PCR de Rab2b dependiendo de la RT utilizada.

MMLV (Fig. 22B). El caso de Rab2b es bastante complejo, ya que con las 4 RTs se observan distintos productos de PCR (Fig. 22B). Se secuenciaron los productos del RT-PCR de Rab2b y ninguna de las bandas secuenciadas coincide exactamente con lo esperado. Se observa que los productos de PCR de Rab2b presentan múltiples sitios de splicing siendo casi todos no consenso y con largas secuencias directas repetidas (datos no mostrados). En conclusión no se validó ninguno de los intrones CT-AC de ratón elegidos.

3.1.4 Falseamiento de los intrones CT-AC descritos en la literatura

Mediante las validaciones experimentales realizadas no se pudo confirmar la existencia de ningún intrón CT-AC**.** Cabe preguntarse entonces si los intrones CT-AC descritos en la literatura son reales o son errores de anotación y/o artefactos del RT-PCR. Para ello se analizó cada uno de los reportes en donde se sugiere la existencia de un intrón CT-AC.

El grupo de He et al. (2009) reportó que hasta un 6,5% de los sitios de splicing en datos de humano correspondían a sitios CT-AC. Al ser contactado el autor nos cedió sus datos originales. Al ser analizados se verificó que los intrones CT-AC correspondían a intrones de ESTs que estaban anotados de forma inversa a genes conocidos. Como se discutió anteriormente, estos ESTs provienen de bibliotecas que no conservan la hebra original.

El gen CaBP3 es uno de los casos en que se reportó un intrón CT-AC en humano (Haeseleer et al., 2000). Ya no existe el gen CaBP3 en la base de datos *Gene* del NCBI y su acceso fue reemplazado por el de CaBP5, ya que el gen CaBP3 no existía y era un error de alineamiento del cDNA AF169158.

Para el gen SIAH-1 en humano se describió una variante de splicing en la cual se activaría un intrón CT-AC en el exón 2 del gen (Mei et al., 2007). Todo apunta a que el supuesto sitio de splicing CT-AC es en realidad un intrón GT-AG de un NAT de SIAH1. Hay evidencia de ESTs y datos de RNA-Seq hebra específica (IBM 2.0) que muestran la presencia de un NAT con un intrón GT-AG procesado en esa posición (datos no mostrados). Los autores del artículo original realizaron un RT-PCR que no discriminaba mRNA sentido de antisentido y por tanto pudieron amplificar el NAT de SIAH1 con el intrón procesado y confundirlo con el mRNA sentido con un intrón CT-AC procesado.

Se reportó que el gen de β-tubulina en *Karlingiomyces* poseía un intrón CT-AC (Keeling, 2003). Mediante una comunicación personal con el autor, éste manifestó que el intrón CT-AC habría sido generado por un error en el alineamiento de sus datos.

Se describió la existencia de un intrón CT-AC en el gen PIG-O de *Plasmodium falciparum* (Delorenzi et al., 2002). Sin embargo, la anotación del gen PIG-O en PlasmodiumDB muestra que no existe este intrón CT-AC. Además un EST BM274380 de Plasmodium presenta las coordenadas de ese intrón pero en la otra dirección, es decir GT-AG.

PXR es un ejemplo en donde se detectaron dos variantes de splicing con intrones CT-AC (Kurose et al., 2005). Realizamos RT-PCRs con una muestra comercial de RNA Universal de Humano y no pudimos amplificar las variantes descritas en este artículo. Además se buscó evidencia en datos de cDNAs, ESTs y de RNA-seq (IBM 2.0) para estos intrones y no se encontraron las variantes de splicing que ellos reportan (datos no mostrados).

ASFMR1 es un caso al parecer bien caracterizado de un intrón CT-AC, ya que está presente en los datos de RefSeq de humano (Ladd et al., 2007). Pero según la anotación de GENCODE V19 ese

intrón CT-AC en realidad es un intrón GT-AG de un NAT. Este ejemplo es similar al observado para BCL10 y ZNF165 (Fig. 20).

Finalmente, el caso del gen U83 del virus Herpes humano 6 es bastante especial, ya que los autores muestran que puede sufrir splicing en sitios CT-AC al sobre expresarlo en células humanas sin la necesidad de proteínas virales (French et al., 1999). En la Figura 23A se muestran las secuencias de los posibles sitios de splicing de U83. Existen 2 variantes del virus, A y B. Las dos variantes del gen U83 fueron clonadas en un plásmido pcDNA3.1 y expresadas en las líneas celulares PC12, Cos7, HeLa y HEK293. No se observó la variante procesada de U83A o U83B en los ensayos de splicing *in vivo* en esas líneas celulares (datos no mostrados). Al utilizar células HEK293T se observó una banda correspondiente a U83B sin su intrón (Fig. 23B). Además al expresar U83B con mutaciones sitio dirigidas en los sitios de splicing CT-AC se deja de observar la banda de U83 procesado (Fig. 23B). Para controlar que esto no fuera producto de un *template switching* (alrededor de los sitios de splicing de U83B existen secuencias repetidas directas de 6 nt) se realizaron RTs utilizando 3 enzimas distintas y en todos los casos se obtuvo el mismo resultado (datos no mostrados). Además al realizar una transcripción *in vitro* del RNA de U83B y posteriormente mezclarlo con RNA de HEK293T y someterlo a RT-PCR no se observa la variante de U83 procesada (datos no mostrados).

Se repitió el RT-PCR de U83B transfectado en células HEK293T usando RT-PCR hebra específica. Lo que se observó fue que el transcrito de U83B realmente no sufre splicing (Fig 23C). La variante de U83B que sufre splicing corresponde a un NAT. Esto resultados muestran que la banda procesada que se observa es un NAT de U83B en la que se procesaría el intrón en los sitios GT-AG.

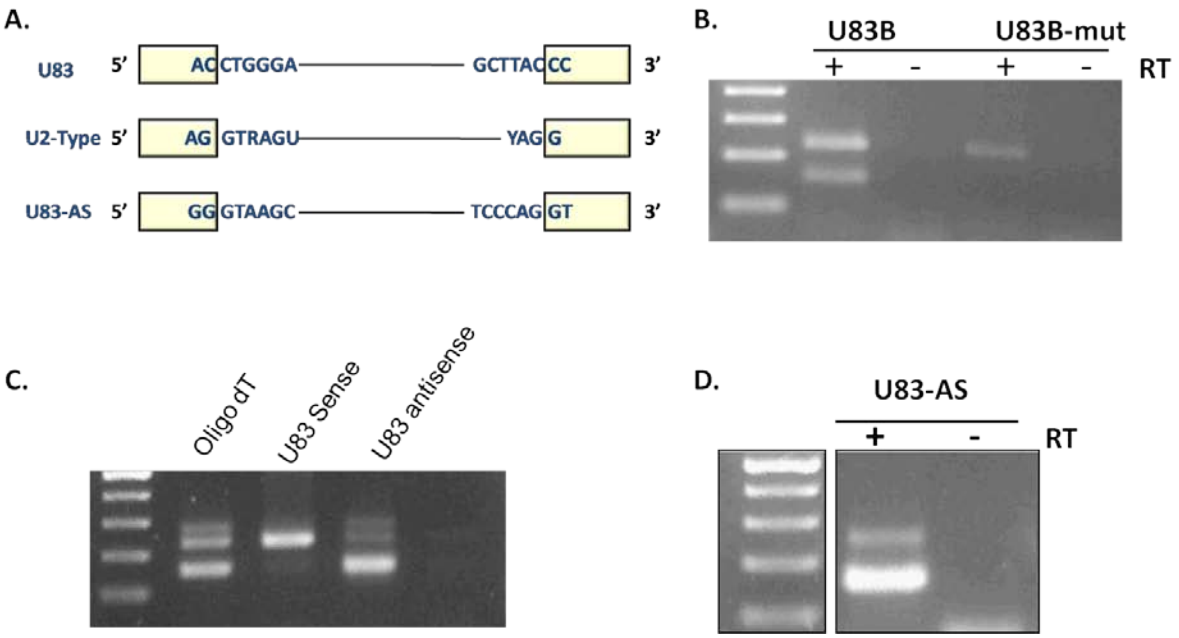


Figura 23. El splicing de un intrón CT-AC del gen U83 de HHV-6 es un artefacto. (A) Secuencias de los sitios de splicing del "intron CT-AC" de U83, reportado en la literatura. **(B)** Ensayo de splicing *in vivo* de U83B en células HEK293T. El vector U83B-mut posee mutaciones sitio dirigidas en los sitios CT-AC. Se realizó el RT-PCR utilizando oligo dT. **(C)** RT-PCR hebra específica de un ensayo de splicing *in vivo* del gen U83B en células HEK293T. **(D)** RT-PCR con oligo dT de un ensayo de splicing *in vivo* en células HEK293T del gen U83B clonado en dirección antisentido.

Esto se verificó clonando U83 en la dirección antisentido y realizando ensayos de splicing *in vivo*. Se observó que el intrón GT-AG de U83-AS es removido (Fig. 23D).

Los resultados sugieren que en el extremo 3' del gen de U83 existiría un promotor críptico que en células HEK293T podría activarse y generar un NAT, el que sufriría el splicing de un intrón GT-AG que tiene señales cercanas al consenso para un intrón U2.

Los resultados obtenidos con los ensayos de splicing *in vivo* para el gen U83 sugieren que la maquinaria celular no puede reconocer y procesar la secuencia complementaria y reversa a la de un intrón consenso del tipo U2. Para poner a prueba nuevamente esta posibilidad se clonó el gen de Ucn de rata en un vector de expresión pcDNA3.1. Se ha descrito la existencia de un mirrorRNA para este gen (Haeger et al., 2005) y se sugirió que este NAT podría generarse por el splicing de un intrón del tipo CT-AC. Por ello además se clonó el gen de Ucn de forma antisentido. Al realizar ensayos de splicing en cuatro líneas celulares distintas: PC12, HeLa, HEK293 y Cos7, se observa que el intrón del gen de Ucn es eficientemente procesado, pero el NAT de Ucn no sufre splicing en el posible intrón CT-AC (Fig. 24).

A partir de nuestros resultados podemos concluir que los intrones CT-AC que se observan en los datos de cDNAs y de ESTs son artefactos producidos por diversos motivos: *template switching* de la RT, generación de fusiones sentido-antisentido, alineamientos anotados en la hebra equivocada y errores de los programas que generan los alineamientos.

Nuestros resultados sugieren que la maquinaria de splicing celular no es capaz de procesar las secuencias complementarias e inversas a las de un intrón consenso por lo que la biogénesis de los

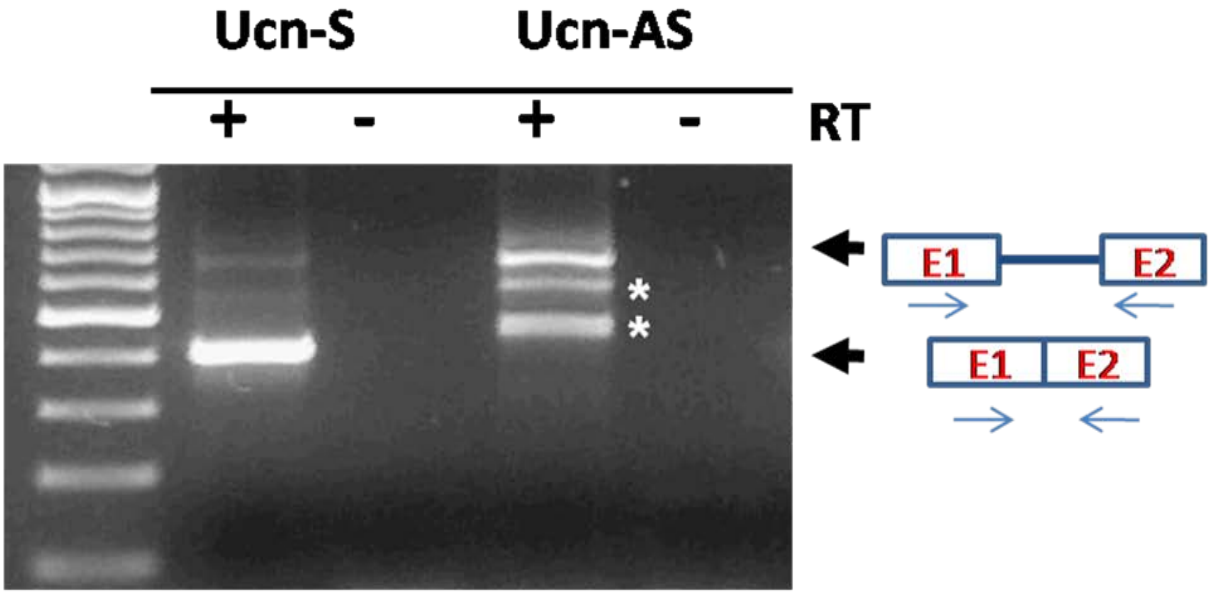


Figura 24. La maquinaria de splicing celular no es capaz de reconocer y procesar las secuencias complementarias e inversas a los sitios de splicing del gen de Ucn de rata. Ensayo de splicing *in vivo* para Ucn sentido y Ucn antisentido en células HEK293T. El transcrito Ucn-AS no sufre splicing en los sitios CT-AC. Con * se marcan mRNAs procesados en sitios GT-AG crípticos. Las flechas indican el tamaño de las dos isoformas esperadas (con el intrón procesado y con el intrón sin procesar).

mirrorRNAs no sería por transcripción bidireccional de un *locus* y splicing de los sitios complementarios e inversos.

3.2 Actividad RdRP

La hipótesis más utilizada para explicar la existencia de los mirrorRNAs es la posible existencia de una actividad RdRP en células de mamíferos (Cheng et al., 2005; Haussecker et al., 2008; Merzendorfer et al., 1997; Rosok y Sioud, 2004; Volloch et al., 1996). En caso de que esto ocurriera, una de las predicciones es que algunos mirrorRNAs deberían tener en su extremo 5' una secuencia poli(T). Este extremo poli(T) no debería estar codificado en el genoma y provendría de la cola poli(A) 3' del mRNA sentido.

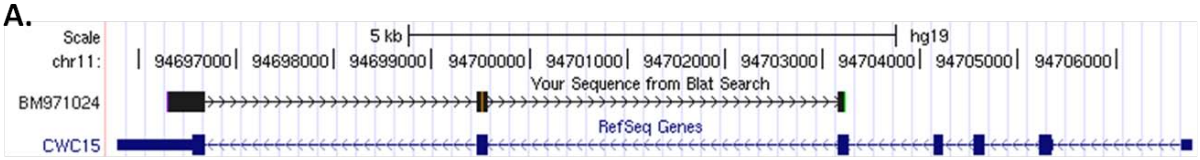
Para poner a prueba esta predicción se realizó una búsqueda de mirrorRNAs con estas características en los datos de ESTs y cDNAs. Se muestran los resultados en la Tabla XII. Se encontraron 5 secuencias tipo mirrorRNA con extremos 3' poli(A) y 5' poli(T). En la Figura 25 se muestra el ejemplo de CWC15.

Para aumentar la profundidad de la búsqueda se utilizaron datos de RNA-Seq. En primer lugar se utilizaron los datos hebra específica de IBM 2.0. Se Analizaron aproximadamente 1.200 millones de lecturas en búsqueda de lecturas que tuvieran al menos 6 Ts no genómicas en su extremo 5' y que alinearan con en el extremo 3' de genes conocidos. Se muestran en detalle los resultados en la Tabla XIII. No se encontró ninguna lectura que apoyara la existencia de RNAs que tuvieran un extremo 5' poli(T) no genómico en donde el extremo 5' poli(T) pudiera provenir de una cola poli(A) de mRNAs de genes conocidos.

Tabla XII. Secuencias de mirrorRNAs con extremos 3' poli(A) y 5' poli(T).

Gen	Nombre	Cromosoma	Inicio	Termino	Hebra	Orientación	Tamaño poli(A) 3'	Tamaño poli(T) 5'
ANUBL1	AI624999	chr10	46111823	46120086	+	-1	7	6
CWC15	BM971024	chr11	94696279	94703214	+	-2	11	15
DHX29	CA432936	chr5	54552072	54557266	+	-2	9	17
LOC129293	CR936634	chr2	85048801	85064628	+	-3	20	17
RPS25	AA781013	chr11	118886421	118889052	+	-4	11	11

A.



B.

cDNA BM971024

```
ttttttttt ttttTTTGG TTATTTAGAT TAAAAATATT TATTACAGGC  50
AGATTTGCTT TTCCCAGGAA AACATATCAC TGGAAACATT TGGAGAATGC  100
AAACTGGGTT AAACCTATTC CCAAGTCCAT TATCAAGTAA GGTATCACTA  150
AAGAAAAAGA TAGGAGTTTA AAACTGGGGA AGCCCACACA CAATTTAGAC  200
AGGGGAAAAG AAAAAAAAAA CTCATAAAAA AAGTCAGAAT GATCCTTTAC  250
GTTTTACAGT CTTTAATTAA GCACATAAAA CTGTACTATT TAATATATTT  300
CTCCATGAAC TTTTTGTGAA ATTCAGATCG CAGTGTGTCA TTTACAAATC  350
TTTTGTCTTT CTTCTGGTCA TCTACACCTT TTGCACAGTT CTTGAAGACA  400
ACGTCATCAT CCCACCTTCT TTTAACTTTG AAGTTGGCCT GAGGCTGGGA  450
TGGGCCAGTG AGATtAAGGA GAGGGTTTCC GCTCAGAATG TTTTCCATAC  500
GAATCCTCTC TTCTTCAGCT TTTTGTTCTT GTTCCTTCCT GGCCTGCTCT  550
TCAGCTCTTT CTTTTTTAAT TTTTTCCAGT TCTGCAAGAA GAGCTGgaaa  600
aaaaaaaa
```

Figura 25. Ejemplo de un mirrorRNA con cola 3' poli(A) y extremo 5' poli(T). (A) Imagen del UCSC Genome Browser que muestra el alineamiento del cDNA BM971024 al genoma humano. El cDNA BM971024 es antisentido al gen CWC15. **(B)** Alineamiento de BM971024 al genoma humano utilizando BLAT, la cola 3' poli(A) y el extremo 5' poli(T) no están codificados en el genoma. Las secuencias que están en negro y en minúscula no alinean al genoma.

Tabla XIII. Búsqueda de RNAs con extremo 5' poli(T) no genómico en datos de RNA-Seq de IBM 2.0.

Datos	N° de Lecturas	N° de lecturas post filtros	Lecturas con Poli(T) 5'	Lecturas con poli(T) 5' alineadas en el extremo 3' de genes conocidos
ERR030868.fastq	72.451.624	48.844.432	49.817	0
ERR030869.fastq	70.743.680	48.088.358	49.705	0
ERR030871.fastq	72.321.018	48.922.330	49.705	0
ERR030863.fastq	73.520.276	49.477.389	45.125	0
ERR030862.fastq	73.420.952	49.444.192	47.298	0
ERR030870.fastq	71.937.539	48.533.977	48.411	0
ERR030859.fastq	76.274.508	64.863.519	66.898	0
ERR030867.fastq	73.773.895	60.137.644	61.995	0
ERR030866.fastq	74.249.497	57.900.535	67.590	0
ERR030860.fastq	75.929.029	66.204.978	67.294	0
ERR030861.fastq	74.756.517	58.993.695	55.587	0
ERR030864.fastq	77.258.890	62.810.257	57.410	0
ERR030865.fastq	75.982.104	61.649.874	45.982	0
ERR030857.fastq	78.243.019	61.886.476	55.746	0
ERR030858.fastq	77.229.855	62.655.875	55.613	0
ERR030856.fastq	76.447.153	62.345.616	57.480	0

En el año 2010, se describió la presencia de RNAs pequeños con extremos 5’ poli(T) (Kapranov et al., 2010a) y los autores atribuyeron su origen a una actividad RdRP. Por ello se intentó replicar sus resultados utilizando datos de RNA-Seq de RNAs pequeños generados por el proyecto ENCODE. Se analizaron datos de 3 tipos celulares distintos y en cada caso se utilizaron tres tipos de tratamientos distintos para generar las bibliotecas. De un total de aproximadamente 1.700 millones de lecturas analizadas, no se encontró ningún alineamiento que tuviera un extremo 5’ poli(T) que pudiera provenir de una cola poli(A) de genes conocidos (Tabla XIV).

A partir de los datos analizados no fue posible encontrar evidencia convincente que apoye la hipótesis de la existencia de una actividad RdRP que genere mirrorRNAs.

3.3 mirrorRNAs derivados desde pseudogenes procesados

En esta tesis se ha propuesto que un posible origen de los mirrorRNAs es la transcripción antisentido de pseudogenes procesados que estén en el genoma humano. Con las búsquedas en datos de cDNAs, ESTs y RNA-seq se detectaron algunos de estos eventos. Además en la validación experimental se detectó el caso de pseudo-mirrorRNAs provenientes de pseudogenes de KRT8. En los casos mencionados los pseudogenes procesados difieren del gen de origen en algunas bases, lo que permite determinar desde que lugar del genoma proviene el transcrito. Por otra parte, existen casos de mirrorRNAs que provienen de genes que no poseen pseudogenes procesados anotados en el genoma de referencia. Por tanto, en principio esos mirrorRNAs no podrían poseer ese origen. Sin embargo, si existieran pseudogenes procesados polimórficos en la población humana que no estén anotados en el genoma humano de referencia, algunos de los mirrorRNAs descritos podrían provenir de ellos.

Tabla XIV. Búsqueda de RNAs con extremo 5' poli(T) no genómico en datos de RNA-Seq de RNAs pequeños de ENCODE.

Células	RNA	Biblioteca	N° de lecturas post-filtros	Lecturas con Poli(T) 5'	Lecturas con poli(T) 5' alineadas en el extremo 3' de genes conocidos
A549	Total	Sin tratamiento	89.051.064	2.189	0
A549	Total	Sin tratamiento	82.146.000	2.793	0
A549	Total	TAP	54.294.588	1.831	0
A549	Total	TAP	60.918.750	1.994	0
A549	Total	CIP-TAP	65.351.054	5.129	0
A549	Total	CIP-TAP	55.370.700	3.592	0
A549	Citoplasmático	TAP	52.460.743	2.581	0
A549	Citoplasmático	CIP-TAP	53.641.974	2.151	0
A549	Nuclear	TAP	112.608.743	2.397	0
A549	Nuclear	CIP-TAP	45.191.601	1.693	0
H1-neurons	Total	Sin tratamiento	95.449.257	724	0
H1-neurons	Total	TAP	112.417.229	390	0
H1-neurons	Total	CIP-TAP	116.886.814	5.857	0
MCF7	Total	Sin tratamiento	95.156.050	1.505	0
MCF7	Total	TAP	87.440.995	3.731	0
MCF7	Total	CIP-TAP	73.012.247	2.648	0

Recientemente se ha descrito que existen pseudogenes procesados polimórficos, que no están reportados en el genoma humano de referencia (Abyzov et al., 2013; Ewing et al., 2013; Schrider et al., 2013). En las publicaciones se detectaron 252 genes desde los cuales se derivaron pseudogenes procesados polimórficos (Tabla S19). Al cruzar esa lista de genes con la de los mirrorRNAs la intersección es de 18 genes de un total de 131, lo que muestra que existe un enriquecimiento de mirrorRNAs presentes en esa lista de genes ($\chi 2$ $p<0,0001$) (Fig. 26).

Al comparar los mirrorRNAs que poseen evidencia para más de un SJ según los datos de IBM 2.0 (Tabla S13) se aprecia que 5 (MYH11, ZNF100, LAPTM4B, CALR y TPT1) de los 13 genes están en la lista de pseudogenes procesados polimórficos ($\chi 2$ $p<0,0001$). Hay casos muy interesantes como el del gen SMAD4. No existe ningún pseudogen procesado anotado en el genoma de referencia para SMAD4, pero 2 de los estudios citados (Abyzov et al., 2013; Schrider et al., 2013) reportan la existencia de un pseudogen procesado de SMAD4 en algunos de los genomas analizados. En los datos de ENCODE se encontraron lecturas mirror para 6 de sus 11 intrones. Esto sugiere fuertemente que estas lecturas mirror provienen de la transcripción antisentido de un pseudogen procesado polimórfico de SMAD4.

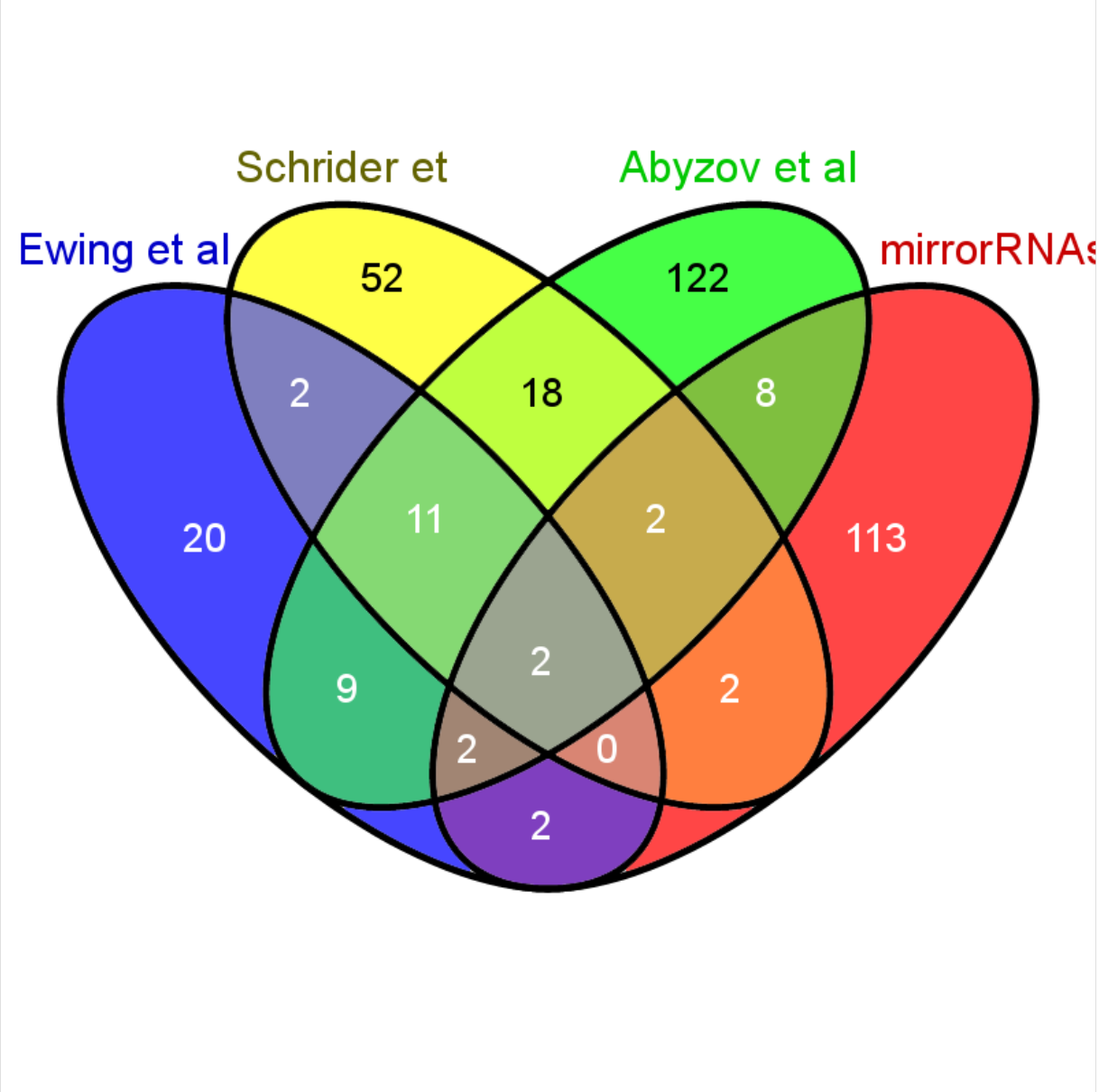


Figura 26. Identificación de mirrorRNAs en genes humanos que poseen pseudogenes procesados polimórficos. El diagrama de Venn muestra la intersección entre las listas de genes que poseen pseudogenes procesados polimórficos identificados en tres estudios independientes con la lista de genes a la que se le identificó mirrorRNAs en esta tesis.

DISCUSIÓN

Durante la presente tesis se puso a prueba la real existencia de los mirrorRNAs para así determinar si eran artefactos experimentales o ncRNAs reales. Nuestros datos muestran mediante distintas aproximaciones tanto bioinformáticas como experimentales que son ncRNAs reales. La existencia de los mirrorRNAs era difícil de ser explicada. Por ello se estudiaron distintas hipótesis que permitían entender cómo se originan los mirrorRNAs. Los resultados presentados en esta tesis indican que los mirrorRNAs son producidos por la transcripción antisentido de pseudogenes procesados.

1. Predicción bioinformática de los mirrorRNAs

Existen dificultades para asignar con total seguridad la hebra desde la cual provienen los datos de ESTs y cDNAs, ya que muchas bibliotecas no fueron preparadas con métodos que mantengan la información de la hebra, o en otros casos en donde si hay anotación de la dirección, pueden estar anotados en la hebra equivocada (Chen et al., 2004; Lavorgna et al., 2004; Shendure y Church, 2002; Yelin et al., 2003). La aproximación utilizada en la presente tesis fue identificar colas 3' poli(A) para así asignar la hebra desde la cual provienen las secuencias. Esta aproximación ha sido anteriormente utilizada para identificar NATs (Chen et al., 2004; Yelin et al., 2003). Es evidente que esta condición restringe la búsqueda sólo a una fracción de los datos ya que sólo el 1,8% de los datos tiene una cola poli(A). Por otra parte una de las limitaciones es que esa búsqueda sólo identificaría mirrorRNAs poliadenilados. A pesar de lo anterior, la búsqueda permitió la identificación de secuencias mirror. No se puede descartar que en la búsqueda realizada se hayan encontrado falsos positivos, producto de errores de secuenciación, polimorfismos en el genoma (polimorfismos podrían llevar a identificar equivocadamente una cola 3' poli(A) que en realidad es una secuencia genómica) u

otros. Es interesante que cinco de los mirrorRNAs que se validaron experimentalmente (UBR5, CWC15, KRT8, RPS23 y RPS11) fueron detectados mediante la búsqueda en ESTs y cDNAs y los cinco dieron resultados positivos.

Los datos de RNA-seq permiten analizar el transcriptoma de los organismos a una gran profundidad, por eso son muy útiles a la hora de detectar RNAs que estén presentes en bajos niveles. En este caso la utilización de datos de RNA-seq hebra específica era esencial, para poder distinguir mirrorRNAs de mRNAs sentido. La forma en la que se construye la biblioteca para el RNA-seq hebra específica puede influir en el grado de especificidad de hebra de las lecturas (Levin et al., 2010). Sólo se utilizaron datos de RNA-Seq que estuvieran generados con bibliotecas en la cuales se ligaron los adaptadores a nivel de RNA, para así evitar artefactos (Levin et al., 2010).

En los datos de RNA-Seq de IBM 2.0 provenientes de RNA total (ERS025084) se encontró una mayor proporción de lecturas provenientes de datos de pseudo-mirrorRNAs que en los datos poli(A)+ (ER025087). Esto puede deberse a que muchos de los pseudogenes procesados desde los que se derivan los pseudo-mirrorRNAs se encuentran dentro de intrones de genes (antisentidos a la dirección de la transcripción del gen), por tanto es posible que lecturas mirror provengan de RNAs poli(A)-, ya sea pre-mRNAs, intrones procesados o transcritos intrónicos no poliadenilados. Por otra parte, el hecho que se encuentren mirrorRNAs en los datos de RNA-Seq poli(A)+ en cantidades similares a la de datos de RNA total, sugiere que gran parte de los mirrorRNAs podrían estar poliadenilados.

Mediante el análisis bioinformático desarrollado se encontraron múltiples posibles mirrorRNAs y pseudo-mirrorRNAs en los datos de RNA-Seq. Debido a que las lecturas de RNA-seq utilizadas son pequeñas (entre 50 nt y 100 nt) existieron muchos casos en donde no se podía

diferenciar si las lecturas eran mirror o pseudo-mirror, ya que alineaban con igual *score* a un SJ que a un pseudogen procesado. Ante esta dificultad se prefirió clasificar las lecturas con estas características como pseudo-mirror. Esto puede llevar a detectar menos mirrorRNAs de los que realmente existen en las muestras analizadas, producto de la imposibilidad de diferenciar con certeza el origen de las lecturas.

Al comparar los resultados de las dos estrategias de búsqueda de mirrorRNAs se observó que los datos de pseudo-mirrorRNAs encontrados en los datos de ESTs y cDNAs están todos contenidos en los datos de pseudo-mirrorRNAs encontrados en datos de RNA-seq (Fig. 9C). Para el caso de los mirrorRNAs la intersección es menor (Fig. 9B), pero si se incluyen los datos de mirrorRNAs y pseudo-mirrorRNAs encontrados en datos de RNA-seq, la intersección con los datos de ESTs y cDNAs si es significativa (Fig. 9A). Esto puede deberse a que varios mirrorRNAs fueron clasificados como pseudo-mirrorRNAs en los datos de RNA-seq producto de que el pequeño tamaño de las lecturas no permitía diferenciar entre mirror o pseudo-mirror. Otra razón por la que la identificación de pseudo-mirrorRNAs puede ser más reproducible (en distintas muestras) que la identificación de mirrorRNAs, es que los pseudo-mirrorRNAs son generados desde pseudogenes anotados en el genoma de referencia y que están presentes en toda (o la gran mayoría) de la población. En cambio, si los mirrorRNAs provienen de pseudogenes procesados de origen reciente y que son polimórficos en la población encontrándose algunos de ellos quizás en un bajo porcentaje de individuos, la posibilidad de reproducir la identificación de mirrorRNAs con fuentes de datos distintas se hace menos probable.

La intersección de los datos de mirrorRNAs y pseudo-mirrorRNAs derivados de datos de IBM 2.0 y de ENCODE es bastante grande, lo que muestra que el método de búsqueda es robusto y reproducible con datos de RNA-seq independientes (Fig. 9A). Es interesante notar que en los datos de IBM 2.0 se encontró un mayor número de genes que posiblemente tengan mirrorRNAs en

comparación a los datos de ENCODE (Fig. 9B). Esto no se puede explicar por la profundidad de los datos analizados. Los datos de ENCODE provienen de cinco líneas celulares y pudiese ser que en esas líneas celulares exista una menor diversidad de pseudogenes procesados polimórficos en los genomas de las células. En cambio los datos de IBM 2.0 provienen de 16 tejidos (de personas distintas). Esto representa una mayor diversidad de "individuos" y por tanto mayor posibilidad de que los genomas contengan pseudogenes procesados no anotados en el genoma de referencia que den origen a mirrorRNAs. Además, cabe la posibilidad que dentro de un tejido de un humano adulto existan poblaciones de células que posean eventos de retrotransposición que no sean compartidas por todas las células del mismo tejido. Esto ya se ha visto en datos de cerebro humano en donde en tejido adulto existe una heterogeneidad de eventos de retrotransposición (Baillie et al., 2011; Coufal et al., 2009). Adicionalmente puede ser que algunos de los 16 tejidos analizados posean tasas de retrotransposición más elevadas que el resto de los tejidos (cerebro por ejemplo), incluidos las de las líneas celulares, esto podría producir una mayor cantidad de pseudogenes procesados y posiblemente una mayor cantidad de mirrorRNAs.

2. Validación experimental de los mirrorRNAs

La validación experimental de NATs presenta ciertos desafíos, ya que existe la posibilidad de identificar de forma incorrecta la hebra desde la cual proviene el RNA. Para el caso de los mirrorRNAs también existe la dificultad de que se expresan en muy bajos niveles, por tanto para validarlos experimentalmente eran necesarios métodos altamente específicos y sensibles. En un inicio se intentó realizar RT-PCR hebra específica, pero distintos experimentos realizados en el laboratorio junto a datos de la literatura muestran que las técnicas convencionales de RT-PCR hebra específica no son del todo específicas (Craggs et al., 2001; Haddad et al., 2007; Lanford et al., 1994). En el laboratorio se desarrolló una técnica de RT-PCR hebra específica que mejora la especificidad de la detección (Cerda, 2014) pero finalmente no fue utilizada para la identificación de mirrorRNAs.

Uno de los posibles artefactos que se debe controlar es la actividad DNA polimerasa dependiente de DNA (DdDP) de la RT. Se ha observado que experimentos de affymetrix hebra específica detectan NATs que son artefactos de la RT, debido a la generación de cDNAs doble hebra (Perocchi et al., 2007). Esto hace que sea necesario interpretar con cuidado la identificación de mirrorRNAs por metodologías que no tomen en cuenta este fenómeno. Nuestros ensayos de 3' RACE y 5' RACE con TdT y dCTP fueron realizados utilizando ActD durante la reacción de RT. Se describió que la utilización de ActD durante la RT inhibe la actividad DdDP de la RT (Perocchi et al., 2007).

Los ensayos de 5' y 3' RACE son altamente sensibles y hebra específica. Para el caso del 3' RACE se lograron validar 5 (UBR5, CWC15, KRT8, RPS23 y RPS11) de los 10 mirrorRNAs analizados. Para el caso de RPS11 (Fig. 12C) se identificó un mirrorRNA que termina en un tramo poli(A) genómico. En este caso es difícil saber si este es el extremo 3' real del mirrorRNA. Es posible que este no sea el extremo 3' real del mirrorRNA y haya sido generado por un *priming* en el tramo poli(A) genómico, generando un cDNA más corto que el RNA original. Se ha observado que el *priming* de secuencias poli(A) internas puede afectar la generación de cDNAs completos o llevar a la transcripción inversa de RNAs que no tienen una cola poli(A) (Nam et al., 2002; Ravasi et al., 2006). Una última opción es que este clon sea un artefacto experimental, que se haya producido por la formación de cDNA desde el mRNA sentido y posterior *priming* del partidor QT en la secuencia poli(A) y generación de un cDNA doble hebra. Para evitar este artefacto se utilizó ActD, compuesto que inhibe la formación de cDNAs doble hebra, por tanto esta posibilidad es menos probable.

En el caso del 3' RACE de DDX5 y CALR, se obtuvieron mirrorRNAs parciales (Fig. 13A y B), en los que sólo un SJ es mirror, pero el NAT detectado además tiene secuencias intrónicas del gen sentido. Una posibilidad es que las secuencias que son parcialmente mirrorRNAs de CALR y DDX5

sean artefactos del 3' RACE, en donde se amplificó un cDNA doble hebra generado a partir de un pre-mRNA en donde se removió un intrón pero no todos los intrones del pre-mRNA y el primer QT se unió a la secuencia poli(A) interna. Esta posibilidad existe y no se puede descartar del todo, pero como ya se planteó se intentó evitarla utilizando ActD durante la reacción de RT, para evitar la generación de cDNAs doble hebra.

Los datos de 5' RACE muestran que se logró validar la existencia de 6 de los 7 mirrorRNAs analizados. El protocolo de 5' RACE utilizando TdT y dCTP, podría llegar a presentar artefactos en la especificidad de hebra si es que se forma cDNA doble hebra y se agrega dCTP en el extremo de la segunda hebra de cDNA. Para evitar esto se utilizó ActD en la reacción. Por otra parte, en el 5' RLM RACE se evitan posibles artefactos de la RT, ya que se liga un adaptador de RNA al extremo 5' de los RNAs que poseen cap. Está estrategia de ligación de un adaptador de RNA en principio debería garantizar la total especificidad de hebra de la reacción.

La utilización de un protocolo que combina el RPA con el RT-PCR permitió la detección de mirrorRNAs (Fig. 19). Un protocolo similar fue utilizado por Podlowski et al. (2002) para detectar un mirrorRNA de la Troponina I cardíaca en rata. Algo importante en este protocolo es controlar que la acción de las RNasas (A y T1) sea completa y que no quede RNA simple hebra sin degradar. En nuestro caso varios de los posibles mirrorRNAs no amplifican y ACTB tampoco amplificó lo que muestra que la acción de las RNAsas fue completa. Este ensayo podría realizarse acoplado a qPCR para identificar de forma cuantitativa la presencia de mirrorRNAs.

Una aproximación experimental que podría ser interesante para detectar NATs y en particular mirrorRNAs, es realizar un RPA acoplado a RNA-Seq. Una muestra de RNA total podría ser tratada para remover el RNA ribosomal y posteriormente hacer un tratamiento con RNAsas que degraden

RNA simple hebra. Se debería precipitar el RNA y posteriormente ligarle adaptadores para generar una biblioteca hebra específica y finalmente secuenciar. Con esta aproximación se podrían detectar las regiones de RNA que se protegen de la acción de las RNAsas por la generación de RNAs doble hebra. Ya existe en la literatura dos reportes de experimentos en los que se utilizó esta aproximación en ratón (Shen et al., 2011; Zheng et al., 2010). Los datos del RPA-RNA-seq con RNA de ratón fueron pedidos a los autores del trabajo para buscar mirrorRNAs, pero la calidad de los datos no permitió hacer un análisis, ya que prácticamente no había lecturas que alinearan al genoma de ratón. Esta aproximación propuesta se podría complementar con la secuenciación del genoma desde donde proviene la muestra de RNA, para asi poder detectar mirrorRNAs que provengan de pseudogenes procesados que no estén anotados en el genoma de referencia.

Tomando en cuenta todos los experimentos de validación experimental que se realizaron en la presente tesis se validó la existencia de 8 de los 10 posibles mirrorRNAs estudiados. Las distintas aproximaciones experimentales no son 100% consistentes para encontrar o no la presencia de un mirrorRNA (por ejemplo el mirrorRNA de CWC15 se encontró en el 3’ RACE y en el 5’ RLM RACE, pero no en el RPA-RTPCR y en el 5’ RACE con TdT y dCTP). Esto puede deberse a que se utilizaron muestras de RNA distintas, para el 3’ RACE y 5’ RACE con TdT se utilizó RNA Universal de humano, para el 5’ RLM RACE se utilizó RNA de MCF7 y para el RPA-RT-PCR se utilizó RNA de HEK293T. Por tanto la discrepancia a la hora de encontrar mirrorRNAs puede ser explicada por el hecho que se utilizaron muestras distintas. Además podrían existir diferencias en la sensibilidad de la detección con los distintos métodos y partidores utilizados. Una buena aproximación sería realizar 3’ RACE, 5’ RLM RACE y RPA-RT-PCR con la misma muestra de RNA (por ejemplo HEK293T). Posteriormente se realizaría PCR con DNA genómico y partidores específicos, para ver si es posible encontrar pseudogenes procesados en el genoma que permitan explicar la existencia de los mirrorRNAs encontrados mediante esa validación.

3. Características de los mirrorRNAs

Los mirrorRNAs se expresan a muy bajos niveles. Los datos de RNA-seq son consistentes y sugieren que los niveles de expresión de los mirrorRNAs están entre el 0,0024 y 0,0075% en relación a los niveles de los mRNAs. Esto puede ser porque los mirrorRNAs Son generados en muy pequeñas cantidades o porque son degradados muy rápido (o las dos). La detección de mirrorRNAs en datos poli(A)+ de RNA-Seq y de forma experimental utilizando 3' RACE, muestran que los mirrorRNAs pueden estar poliadenilados. Generalmente se entiende la poliadenilación como un proceso que aumenta la vida media de los mRNAs. Pero se ha descrito que la poliadenilación puede participar en procesos de degradación de RNAs mediados por el exosoma (Beaulieu et al., 2012; Slomovic et al., 2010; West et al., 2006; Wyers et al., 2005). La mayoría de los mirrorRNAs encontrados en los datos de ESTs y cDNAs no presentaban secuencias de poliadenilación canónicas. Es posible que los mirrorRNAs sean poliadenilados como mecanismo que los lleve a degradación mediada por el exosoma. Para probar esta hipótesis se podrían realizar ensayos de *knockdown* de proteínas importantes para este proceso y posteriormente cuantificar los niveles de los mirrorRNAs, para ver si estos aumentan.

Otra posibilidad que se analizó es que los mirrorRNAs tuvieran bajos niveles de expresión debido a que se unieran a sus mRNAs sentido y al formar RNAs doble hebra fueran procesados por DICER y generaran endo-siRNAs. Los análisis no pudieron identificar ningún siRNA que proviniera de mirrorRNAs (Tabla VI). Por tanto la evidencia sugiere que los mirrorRNAs no se procesarían por este mecanismo y no estarían formando RNAs doble hebra *in vivo*.

Los análisis de *GO* revelaron que en la lista de genes que poseen mirrorRNAs predichos existe un enriquecimiento de genes que participan en el proceso de traducción de proteínas. Este

enriquecimiento podría sugerir que los mirrorRNAs participan en procesos regulatorios de genes importantes para el proceso de traducción. Pero por otra parte esta asociación podría deberse a que se encontraron mirrorRNAs para 11 genes que participan en traducción de proteínas (8 de ellos proteínas ribosomales) de los cuales 10 poseen pseudogenes procesados en el genoma de referencia. Como se mostró en resultados existe un enriquecimiento de mirrorRNAs en genes que poseen pseudogenes procesados. Es interesante notar que se ha detectado una amplia transcripción de pseudogenes procesados de proteínas ribosomales, algunos de ellos con expresión tejido específica (Tonner et al., 2012). Es posible que no solo se transcriban de forma sentido esos pseudogenes procesados de proteínas ribosomales sino que también se transcriban en la dirección antisentido dando origen a pseudo-mirrorRNAs y mirrorRNAs (en el caso de pseudogenes procesados que no estén anotados en el genoma de referencia).

4. Mecanismo de biogénesis de los mirrorRNAs

Después de mostrar que los mirrorRNAs existen, queda preguntarse ¿Cuál es el origen de estos ncRNAs? Como se describió previamente existen dos mecanismos propuestos en la literatura: 1) Splicing de los mirrorRNAs en los sitios complementarios a los canónicos (Haeger et al., 2005; Laabi et al., 1994). 2) La existencia de una actividad RdRP en células humanas que copie un mRNA sentido, generando un mirrorRNA (Cheng et al., 2005; Haussecker et al., 2008; Merzendorfer et al., 1997; Rosok y Sioud, 2004; Volloch et al., 1996). En la presente tesis hemos planteado una tercera opción que consiste en que los mirrorRNAs pueden derivarse de la transcripción antisentido de pseudogenes procesados, algunos de ellos polimórficos en la población y que no están anotados en el genoma de referencia.

4.1 Splicing de los mirrorRNAs en sitios CT-AC

En la presente tesis se describe que nuestras distintas aproximaciones experimentales y bioinformáticas, no encontraron evidencias de que la maquinaria de splicing celular pudiera procesar los intrones complementarios e inversos a los canónicos.

La búsqueda de intrones mirror identificó algunos cDNAs (Tabla VII), pero es bastante probable que estos sean ejemplos de artefactos de la RT como se ha mostrado antes (Houseley y Tollervey, 2010). Es posible que la unión de un cDNA sentido y un cDNA antisentido que solapen en su extremo 3' pueda generar un cDNA quimérico como los mostrados en la Figura 20, producto de la generación de cDNAs de doble hebra. Los intentos de validar experimentalmente dos casos fueron infructuosos.

La búsqueda de intrones CT-AC en el transcriptoma de ratón y de humano no llevó a la identificación de ningún intrón CT-AC de confianza que además tuviera secuencias consensos complementarias a las canónicas. Las diversas validaciones experimentales muestran que estos intrones son artefactos de diversos tipos. Existen tres problemas principales en la identificación de intrones no canónicos. En primer lugar, se ha descrito que las RTs pueden producir un artefacto durante la síntesis de cDNA, denominado *template switching* (Cocquet et al., 2006; Houseley y Tollervey, 2010; Mader et al., 2001). Bajo ciertas condiciones en las cuales existe una estructura secundaria y secuencias directas repetidas, la RT puede "saltar" de un sitio a otro y dejar sin copiar una zona del RNA que puede ser interpretada como un intrón (los supuestos sitios de splicing de Ftl1 tienen una secuencia directa repetida de 6 nt). Ya se han reportado algunos casos en los cuales supuestas variantes de splicing eran en realidad artefactos del RT-PCR (Geiszt et al., 2004). Distintas RTs pueden generar distintas variantes de *template switching* (Houseley y Tollervey, 2010). Además existe un reporte de que enzimas termoestables como Transcriptor tienen inhibida esta propiedad

(Cocquet et al., 2006). Tomando en cuenta lo anterior consideramos que el supuesto splicing del intrón CT-AC de Ftl1 (Fig. 22B) no es otra cosa que un artefacto de la MMLV. El segundo problema para identificar intrones no canónicos es que existen diferencias entre el genoma de referencia y el genoma desde donde provienen los datos de ESTs y cDNAs, inserciones o deleciones polimórficas que pueden llevar a la identificación errónea de intrones. El tercer problema es generado por los errores de los alineadores, ya que ciertos errores en el alineamiento pueden llevar a la falsa identificación de un intron no canónico (Fig. 21).

Hasta el momento existe sólo un trabajo publicado en el que se describió que la maquinaria de splicing de células humanas puede procesar un intron CT-AC. En el trabajo de French et al. (1999) se describe que el gen U83 del HHV-6 es procesado en los sitios de splicing CT-AC. Además al clonar el gen de U83 y expresarlo en células HaCaT observaron el supuesto splicing del transcrito de U83 en los sitios CT-AC. Se intentó replicar los resultados sin éxito. Sólo se observó una variante procesada de U83 en las células HEK293T, pero esta variante procesada era un NAT de U83 (Fig. 23). Este fenómeno es observado al expresar U83B en el vector pcDNA 3.1 y en el vector pTEJ-8 (datos no mostrados) en células HEK293T. Nuestros resultados sugieren que en las células HEK293T se activa un promotor críptico que está en el extremo 3' del gen U83 y que lleva a la transcripción antisentido del gen, el transcrito antisentido puede ser procesado eficientemente en los sitios GT-AG presentes, ya que las secuencias de los sitios de splicing son similares a las canónicas (Fig. 23A). Estos datos sugieren que el resultado de French et al. (1999) es fruto de un artefacto experimental, ya que al realizar un RT-PCR que no discrimina entre transcritos sentido y antisentido, amplificaron la hebra equivocada.

En la literatura existen ejemplos de otros supuestos intrones CT-AC, pero como se presenta en resultados, éstos son fruto de errores experimentales, artefactos de las técnicas o una mala

manipulación de los datos de ESTs y cDNAs. Por tanto es necesario tomar con cautela los reportes en donde se presente la existencia de intrones CT-AC.

4.2 Actividad RdRP en células humanas

Hasta el momento existe controversia sobre la posibilidad de la existencia de actividad RdRP en células de mamífero (Cheng et al., 2005; Kapranov et al., 2010a; Lai, 2005; Maida et al., 2009; Volloch, 1986; Wagner et al., 2013). Distintos experimentos han mostrado que la RNA pol II puede tener actividad RdRP (Abrahem y Pelchat, 2008; Chang et al., 2008; Filipovska y Konarska, 2000; Fu y Taylor, 1993; Lehmann et al., 2007; Modahl et al., 2000; Moraleda y Taylor, 2001). De hecho recientemente se demostró que la RNA pol II puede elongar 18 nt el ncRNA B2 de ratón en ciertas condiciones (Wagner et al., 2013). Análisis *in vitro* han mostrado que la RNA pol II puede actuar como una RdRP si utiliza como sustrato un fragmento del genoma de HDV (Filipovska y Konarska, 2000). La actividad RdRP es altamente específica para ese fragmento identificado, ya que al utilizar otras partes del genoma de HDV u otros RNAs como snRNAs, 7SL RNA, tRNAs no se observa transcripción (Filipovska y Konarska, 2000). Adicionalmente la interacción de la RNA pol II con ese fragmento del genoma de HDV es dependiente más de la estructura secundaria que de la secuencia del RNA (Filipovska y Konarska, 2000). Estos datos sugieren que la actividad RdRP de la RNA pol II es un fenómeno altamente específico y que estaría limitado a ciertas condiciones celulares y sólo a ciertos sustratos. Esto lleva a pensar que difícilmente la actividad RdRP de la RNA pol II pueda explicar la diversidad de mirrorRNAs encontrados. Adicionalmente se ha determinado que la actividad RdRP de la RNA pol II es lenta y poco procesiva comparado con su actividad DdRP (Lehmann et al., 2007). HDV lograría superar estas limitaciones para replicar su genoma, mediante la unión del antígeno delta, la única proteína que codifica este virus, a la RNA pol II estimulando la elongación (Yamaguchi et al., 2001; Yamaguchi et al., 2007). Esto sugiere que es poco probable que en condiciones basales la actividad RdRP de la RNA pol II sea capaz de elongar un transcrito cientos de nucleótidos como se ha

observado en el caso de algunos mirrorRNAs detectados. Los estudios de la actividad RdRP de la RNA pol II han mostrado que es capaz de elongar un extremo 3' OH libre, ya que el templado se pliega sobre si mismo generando un *self-priming*, pero no han mostrado que pueda iniciar la transcripción *de novo* (Filipovska y Konarska, 2000; Lehmann et al., 2007). Para generar un mirrorRNA desde un mRNA mediante actividad RdRP se esperaría que la actividad RdRP empezara *de novo*, a menos que utilice como partidor a otro RNA, como por ejemplo un miRNA.

Se describió que la proteína TERT junto al RNA RMRP podrían unirse y poseer actividad RdRP que llevara a copias del mismo RNA RMRP (Maida et al., 2009). Hasta el momento no se ha mostrado que TERT se una a otro RNA y genere copias de este RNA. El análisis inicial reveló que TERT se podía unir a otros RNAs, principalmente tRNAs mitocondriales, secuencias Alu y RNA ribosomal 5.8S. Los resultados experimentales revelan que TERT/RMRP puede elongar un RNA pero no iniciar un RNA *de novo*. Los datos apuntan a que es poco probable que TERT se asocie a mRNAs y genere copias de esos mRNAs dando lugar a la diversidad de mirrorRNAs observados. Un experimento para poner a prueba el posible papel de TERT en la generación de mirrorRNAs es realizar un *knockdown* de TERT en células en las cuales se han detectado mirrorRNAs, para ver si esto afecta la expresión de mirrorRNAs. Se realizó este experimento pero los resultados de los RPA-RT-PCR no fueron concluyentes.

En el trabajo de Kapranov et al (2010a), se planteó la existencia de una actividad RdRP en células humanas ya que encontraron aTASRs. Estos RNAs pequeños que ellos describen poseen extremos 5' poli(T), que no están codificados en el genoma, y son antisentidos a extremos 3' de genes conocidos. Por tanto plantean que se originarían por una actividad RdRP que se iniciaría desde la cola poli(A) de mRNAs. En la presente tesis se utilizaron datos de RNA-seq de gran profundidad para poder identificar estos RNAs y no fue posible reproducir sus resultados. En el trabajo citado secuenciaron RNAs utilizando Helicos, que permite secuenciación de cDNAs molécula única. Los datos utilizados en

esta tesis fueron obtenidos de ENCODE y fueron generados utilizando un secuenciador de Illumina. El uso de distintas tecnologías, o el hecho de que las bibliotecas de cDNA fueron preparadas con protocolos distintos (los datos de ENCODE utilizados fueron generados usando 3 protocolos distintos en 3 líneas celulares diferentes) podrían explicar las diferencias. Por tanto es difícil discernir con total seguridad si estos RNAs son reales o un artefacto de la secuenciación (o biblioteca) de Helicos.

Tomando en consideración los datos disponibles consideramos que es poco probable que todos los mirrorRNAs sean generados por una actividad RdRP, pero no es del todo descartable que algunos de ellos pudieran generarse mediante un mecanismo RdRP dependiente.

4.3 Transcripción antisentido de pseudogenes procesados

Ya se ha descrito que los pseudogenes procesados pueden ser fuente de NATs (Muro y Andrade-Navarro, 2010; Tam et al., 2008; Watanabe et al., 2008; Zhou et al., 1992). En el presente trabajo identificamos NATs de pseudogenes procesados de KRT8, mediante 3' RACE. Además los datos de RNA-Seq, ESTs y cDNAs también muestran la existencia de NATs derivados de pseudogenes procesados anotados en el genoma humano de referencia.

Normalmente los pseudogenes procesados poseen secuencias similares a las de sus genes parentales, pero mayoritariamente poseen mutaciones suficientes que permiten distinguir de donde proviene un mRNA, si del gen parental o de la transcripción de un pseudogen procesado (Kalyana-Sundaram et al., 2012). En esta tesis se denominó pseudo-mirrorRNAs a NATs que poseen una alta identidad de secuencia con el mRNA de un gen, pero que son más similares en secuencia a un pseudogen procesado presente en el genoma de referencia. La gran mayoría de los pseudo-mirrorRNAs alinea dentro de pseudogenes procesados presentes en la hebra antisentido de intrones de genes codificantes de proteínas.

Pero entonces ¿de dónde provienen los mirrorRNAs que poseen secuencias idénticas a las del mRNA de un gen anotado y que posee un pseudogen procesado con suficientes cambios nucleotídicos como para ser diferenciados? Como se describió en resultados, existe un enriquecimiento de genes que poseen pseudogenes procesados, en la lista de genes con mirrorRNAs. Se ha mostrado que genes que poseen pseudogenes procesados en el genoma de referencia, poseen mayor probabilidad de unir la proteína ORF1p, que es parte del sistema de retrotransposición de los LINEs, y de ser sujetos a transcripción inversa por la proteína de los LINEs ORF2p (Mandal et al., 2013). Esto sugiere que genes que ya poseen pseudogenes procesados anotados en el genoma de referencia poseen mayor probabilidad de generar nuevos eventos de retrotransposición. Por tanto existe la posibilidad de que los mirrorRNAs provengan de la transcripción antisentido de pseudogenes procesados de origen reciente en la población humana y que por tanto la secuencia del pseudogen procesado y la del gen parental sean la misma o muy similares. Recientemente diversos trabajos han reportado la existencia de pseudogenes procesados que son polimórficos en la población y que no están anotados en el genoma de referencia (Abyzov et al., 2013; Ewing et al., 2013; Schrider et al., 2013). Existe un gran enriquecimiento de mirrorRNAs predichos en la lista de pseudogenes procesados polimórficos encontrados en los estudios citados. Hay casos muy interesantes como el del gen SMAD4. No existe ningún pseudogen procesado anotado en el genoma de referencia para SMAD4, pero 2 de los estudios citados (Abyzov et al., 2013; Schrider et al., 2013) reportan la existencia de un pseudogen procesado de SMAD4 en algunos de los genomas analizados. En los datos de ENCODE se encontraron lecturas mirror para 6 de sus 11 intrones. Esto sugiere fuertemente que estas lecturas mirror provienen de la transcripción antisentido de un pseudogen procesado polimórfico de SMAD4.

También es posible que los pseudogenes procesados polimórficos no estén presentes en todas las células somáticas de un individuo, y que el evento haya ocurrido durante el proceso de desarrollo y diferenciación celular. Se ha descrito que en cerebro humano existen diferencias entre los genomas de las células, debido a un mosaicismo en eventos de retrotransposición (Baillie et al., 2011; Coufal et al., 2009). Por lo anterior, dentro de una población de células de un mismo tejido podrían existir diferencias en la presencia de mirrorRNAs.

¿Cómo es posible que los pseudogenes procesados se transcriban? Se ha propuesto mecanismos que pueden llevar a que esto ocurra (Harrison et al., 2005). En primer lugar los pseudogenes procesados pueden insertarse cerca de promotores de genes codificantes de proteínas. En segundo lugar, es posible que los pseudogenes procesados estén dentro de intrones o en los UTRs de genes que se transcriben. Un tercer mecanismo propuesto es que los pseudogenes procesados se transcriban por el uso de promotores crípticos presentes en el DNA intergénico. Dichos promotores podrían provenir de retrotransposones, de duplicaciones genéticas de promotores existentes o en algunos casos de promotores que se generen *de novo*. La transcripción de pseudogenes no es un proceso poco frecuente. Utilizando datos de RNA-Seq provenientes desde distintos tejidos humanos y muestras de tejido tumoral se determinó que al menos 2082 pseudogenes se transcriben en el genoma humano (Kalyana-Sundaram et al., 2012).

El completo entendimiento de la biogénesis de los mirrorRNAs puede abrir las puertas a comprender si es que realmente existe una actividad RdRP en células de mamífero que cumpla funciones regulatorias y por otra parte, entender como los procesos de retrotransposición pueden modificar el genoma humano y tener impactos a nivel del transcriptoma.

Queda preguntarse, ¿Tienen funciones los mirrorRNAs? Se ha descrito que trans-NATs derivados de pseudogenes pueden llevar a la formación de RNAs doble hebra que se procesen por DICER generando endo-siRNAs en ovocitos de ratón (Tam et al., 2008; Watanabe et al., 2008). En principio los mirrorRNAs también podrían generar endo-siRNAs, pero nuestro análisis no identificó ningún endo-siRNA (Tabla VI). En principio los mirrorRNAs y pseudo-mirrorRNAs también podrían regular la expresión de sus mRNAs sentido generando una inhibición de la traducción y/o dirigir procesos de *editing* (Hatzoglou et al., 2002).

Los mirrorRNAs son transcritos que están en muy bajos niveles, probablemente debido a una baja tasa transcripcional y quizás a una alta tasa de degradación mediada por el exosoma. Esto lleva a pensar *a priori* que la mayoría de los mirrorRNA no serían RNAs funcionales, pero es posible que bajo ciertas condiciones sí pudieran adquirir una función asociada a su posibilidad de regular la expresión génica mediante mecanismos que se han observado en otras clases de NATs.

CONCLUSIONES

En esta tesis mediante distintas aproximaciones bioinformáticas y experimentales se ha demostrado la existencia en células humanas de un tipo especial de NATs denominados por nosotros mirrorRNAs.

El estudio del transcriptoma de tejidos humanos llevó a la identificación sistemática de NATs provenientes de pseudogenes procesados presentes en el genoma de referencia, denominados por nosotros pseudo-mirrorRNAs.

Los mirrorRNAs tienen bajos niveles de expresión y algunos están poliadenilados. Aunque en principio podrían formar RNAs doble hebra con sus mRNAs sentido y dar origen a endo-siRNAs, estos no fueron detectados.

Nuestros datos bioinformáticos y experimentales respaldan que la maquinaria de splicing celular no es capaz de procesar intrones en los sitios complementarios e inversos a los canónicos. Por tanto no es posible que los mirrorRNAs sean generados por transcripción del mismo *locus* de un gen y posterior remoción de los intrones en los mismos sitios que el transcrito sentido.

Nuestros datos sugieren que los mirrorRNAs pueden ser generados desde pseudogenes procesados polimórficos.

BIBLIOGRAFÍA

Abrahem, A. y Pelchat, M. (2008). Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis delta virus RNA genome. Nucleic Acids Res 36, 5201-11.

Abyzov, A., Iskow, R., Gokcumen, O., Radke, D.W., Balasubramanian, S., Pei, B., Habegger, L., Genomes Project, C., Lee, C. y Gerstein, M. (2013). Analysis of variable retroduplications in human populations suggests coupling of retrotransposition to cell division. Genome Res 23, 2042-52.

Ahlquist, P. (2002). RNA-dependent RNA polymerases, viruses, and RNA silencing. Science 296, 1270-3.

Alibes, A., Yankilevich, P., Canada, A. y Diaz-Uriarte, R. (2007). IDconverter and IDClight: conversion and annotation of gene and protein IDs. BMC Bioinformatics 8, 9.

Allo, M., Buggiano, V., Fededa, J.P., Petrillo, E., Schor, I., de la Mata, M., Agirre, E., Plass, M., Eyras, E., Elela, S.A., Klinck, R., Chabot, B. y Kornblihtt, A.R. (2009). Control of alternative splicing through siRNA-mediated transcriptional gene silencing. Nat Struct Mol Biol 16, 717-24.

Amaral, P.P., Dinger, M.E., Mercer, T.R. y Mattick, J.S. (2008). The eukaryotic genome as an RNA machine. Science 319, 1787-9.

Ameyar-Zazoua, M., Rachez, C., Souidi, M., Robin, P., Fritsch, L., Young, R., Morozova, N., Fenouil, R., Descostes, N., Andrau, J.C., Mathieu, J., Hamiche, A., Ait-Si-Ali, S., Muchardt, C., Batsche, E. y Harel-Bellan, A. (2012). Argonaute proteins couple chromatin silencing to alternative splicing. Nat Struct Mol Biol 19, 998-1004.

Azevedo, F.A., Carvalho, L.R., Grinberg, L.T., Farfel, J.M., Ferretti, R.E., Leite, R.E., Jacob Filho, W., Lent, R. y Herculano-Houzel, S. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. J Comp Neurol 513, 532-41.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. y Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. Science 318, 798-801.

Baillie, J.K., Barnett, M.W., Upton, K.R., Gerhardt, D.J., Richmond, T.A., De Sapio, F., Brennan, P.M., Rizzu, P., Smith, S., Fell, M., Talbot, R.T., Gustincich, S., Freeman, T.C., Mattick, J.S., Hume, D.A., Heutink, P., Carninci, P., Jeddeloh, J.A. y Faulkner, G.J. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. Nature 479, 534-7.

Beaulieu, Y.B., Kleinman, C.L., Landry-Voyer, A.M., Majewski, J. y Bachand, F. (2012). Polyadenylation-dependent control of long noncoding RNA expression by the poly(A)-binding protein nuclear 1. PLoS Genet 8, e1003078.

Beltran, M., Puig, I., Pena, C., Garcia, J.M., Alvarez, A.B., Pena, R., Bonilla, F. y de Herreros, A.G. (2008). A natural antisense transcript regulates Zeb2/Sip1 gene expression during Snail1-induced epithelial-mesenchymal transition. Genes Dev 22, 756-69.

Bernstein, E. y Allis, C.D. (2005). RNA meets chromatin. Genes Dev 19, 1635-55.

Brannan, C.I., Dees, E.C., Ingram, R.S. y Tilghman, S.M. (1990). The product of the H19 gene may function as an RNA. Mol Cell Biol 10, 28-36.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S. y Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515-26.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J. y Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71, 527-42.

Capaccioli, S., Quattrone, A., Schiavone, N., Calastretti, A., Copreni, E., Bevilacqua, A., Canti, G., Gong, L., Morelli, S. y Nicolin, A. (1996). A bcl-2/IgH antisense transcript deregulates bcl-2 gene expression in human follicular lymphoma t(14;18) cell lines. Oncogene 13, 105-15.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V.B., Brenner, S.E., Batalov, S., Forrest, A.R., Zavolan, M., Davis, M.J., Wilming, L.G., Aidinis, V., Allen, J.E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R.N., Bailey, T.L., Bansal, M., Baxter, L., Beisel, K.W., Bersano, T., Bono, H., Chalk, A.M., Chiu, K.P., Choudhary, V., Christoffels, A., Clutterbuck, D.R., Crowe, M.L., Dalla, E., Dalrymple, B.P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C.F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T.R., Gojobori, T., Green, R.E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T.K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S.P., Kruger, A., Kummerfeld, S.K., Kurochkin, I.V., Lareau, L.F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K.C., Pavan, W.J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J.F., Ring, B.Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S.L., Sandelin, A., Schneider, C., Schonbach, C., Sekiguchi, K., Semple, C.A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S.L., Tang, S., Taylor, M.S., Tegner, J., Teichmann, S.A., Ueda, H.R., van Nimwegen, E., Verardo, R., Wei, C.L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S.M., Teasdale, R.D., Liu, E.T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J.S., Hume, D.A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y., Consortium, F., Group, R.G.E.R. y Genome Science, G. (2005). The transcriptional landscape of the mammalian genome. Science 309, 1559-63.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., Forrest, A.R., Alkema, W.B., Tan, S.L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S.M., Wells, C.A., Orlando, V., Wahlestedt, C., Liu, E.T., Harbers, M., Kawai, J., Bajic, V.B., Hume, D.A. y Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38, 626-35.

Carone, D.M., Longo, M.S., Ferreri, G.C., Hall, L., Harris, M., Shook, N., Bulazel, K.V., Carone, B.R., Obergfell, C., O'Neill, M.J. y O'Neill, R.J. (2009). A new class of retroviral and satellite encoded small RNAs emanates from mammalian centromeres. Chromosoma 118, 113-25.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., Forrest, A.R., Carninci, P., Biffo, S., Stupka, E. y Gustincich, S. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491, 454-7.

Carthew, R.W. y Sontheimer, E.J. (2009). Origins and Mechanisms of miRNAs and siRNAs. Cell 136, 642-55.

Castrignano, T., D'Antonio, M., Anselmo, A., Carrabino, D., D'Onorio De Meo, A., D'Erchia, A.M., Licciulli, F., Mangiulli, M., Mignone, F., Pavesi, G., Picardi, E., Riva, A., Rizzi, R., Bonizzoni, P. y

Pesole, G. (2008). ASPicDB: a database resource for alternative splicing analysis. Bioinformatics 24, 1300-4.

Cocquet, J., Chong, A., Zhang, G. y Veitia, R.A. (2006). Reverse transcriptase template switching and false alternative transcripts. Genomics 88, 127-31.

Consortium, C.e.S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282, 2012-8.

Consortium, E.P., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C. y Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.

Consortium, E.P., Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S., Taylor, C.M., Neph, S., Koch, C.M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., Andrews, R.M., Flicek, P., Boyle, P.J., Cao, H., Carter, N.P., Clelland, G.K., Davis, S., Day, N., Dhami, P., Dillon, S.C., Dorschner, M.O., Fiegler, H., Giresi, P.G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K.D., Johnson, B.E., Johnson, E.M., Frum, T.T., Rosenzweig, E.R., Karnani, N., Lee, K., Lefebvre, G.C., Navas, P.A., Neri, F., Parker, S.C., Sabo, P.J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F.S., Dekker, J., Lieb, J.D., Tullius, T.D., Crawford, G.E., Sunyaev, S., Noble, W.S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I.L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H.A., Sekinger, E.A., Lagarde, J., Abril, J.F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J.S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M.C., Thomas, D.J., Weirauch, M.T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K.G., Sung, W.K., Ooi, H.S., Chiu, K.P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M.L., Valencia, A., Choo, S.W., Choo, C.Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T.G., Brown, J.B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C.N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J.S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R.M., Rogers, J., Stadler, P.F., Lowe, T.M., Wei, C.L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S.E., Fu, Y., Green, E.D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L.A., Wetterstrand, K.A., Good, P.J., Feingold, E.A., Guyer, M.S., Cooper, G.M., Asimenos, G., Dewey, C.N., Hou, M., Nikolaev, S., Montoya-Burgos, J.I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N.R., Holmes, I., Mullikin, J.C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W.J., Stone, E.A., Program, N.C.S., Baylor College of Medicine Human Genome Sequencing, C., Washington University Genome Sequencing, C., Broad, I., Children's Hospital Oakland Research, I., Batzoglou, S., Goldman, N., Hardison, R.C., Haussler, D., Miller, W., Sidow, A., Trinklein, N.D., Zhang, Z.D., Barrera, L., Stuart, R., King, D.C., Ameur, A., Enroth, S., Bieda, M.C., Kim, J., Bhinge, A.A., Jiang, N., Liu, J., Yao, F., Vega, V.B., Lee, C.W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M.J., Inman, D., Singer, M.A., Richmond, T.A., Munn, K.J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J.C., Couttet, P., Bruce, A.W., Dovey, O.M., Ellis, P.D., Langford, C.F., Nix, D.A., Euskirchen, G., Hartman, S., Urban, A.E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T.H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C.K., Rosenfeld, M.G., Aldred, S.F., Cooper, S.J., Halees, A., Lin, J.M., Shulha, H.P., Zhang, X., Xu, M., Haidar, J.N., Yu, Y., Ruan, Y., Iyer, V.R., Green, R.D., Wadelius, C., Farnham, P.J., Ren, B., Harte, R.A., Hinrichs, A.S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A.S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R.M., Karolchik, D., Armengol, L., Bird, C.P., de Bakker, P.I., Kern, A.D., Lopez-Bigas, N., Martin, J.D., Stranger, B.E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I.B., Huppert, J., Zody, M.C., Abecasis, G.R., Estivill, X., Bouffard, G.G., Guan, X., Hansen, N.F., Idol, J.R., Maduro, V.V., Maskeri, B., McDowell, J.C., Park, M., Thomas, P.J.,

Young, A.C., Blakesley, R.W., Muzny, D.M., Sodergren, E., Wheeler, D.A., Worley, K.C., Jiang, H., Weinstock, G.M., Gibbs, R.A., Graves, T., Fulton, R., Mardis, E.R., Wilson, R.K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D.B., Chang, J.L., Lindblad-Toh, K., Lander, E.S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B. y de Jong, P.J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816.

Core, L.J., Waterfall, J.J. y Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322, 1845-8.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V. y Gage, F.H. (2009). L1 retrotransposition in human neural progenitor cells. Nature 460, 1127-31.

Craggs, J.K., Ball, J.K., Thomson, B.J., Irving, W.L. y Grabowska, A.M. (2001). Development of a strand-specific RT-PCR based assay to detect the replicative form of hepatitis C virus RNA. J Virol Methods 94, 111-20.

Czech, B., Malone, C.D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., Perrimon, N., Kellis, M., Wohlschlegel, J.A., Sachidanandam, R., Hannon, G.J. y Brennecke, J. (2008). An endogenous small interfering RNA pathway in Drosophila. Nature 453, 798-802.

Chang, J., Nie, X., Chang, H.E., Han, Z. y Taylor, J. (2008). Transcription of hepatitis delta virus RNA by RNA polymerase II. J Virol 82, 1118-27.

Chang, J., Nie, X., Gudima, S. y Taylor, J. (2006). Action of inhibitors on accumulation of processed hepatitis delta virus RNAs. J Virol 80, 3205-14.

Chao, M., Hsieh, S.Y. y Taylor, J. (1990). Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. J Virol 64, 5066-9.

Chapman, E.J. y Carrington, J.C. (2007). Specialization and evolution of endogenous small RNA pathways. Nat Rev Genet 8, 884-96.

Chen, J., Sun, M., Kent, W.J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R.Z. y Rowley, J.D. (2004). Over 20% of human transcripts might form sense-antisense pairs. Nucleic Acids Res 32, 4812-20.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics Chapter 4, Unit 4 10.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D.K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D.S. y Gingeras, T.R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 308, 1149-54.

Delorenzi, M., Sexton, A., Shams-Eldin, H., Schwarz, R.T., Speed, T. y Schofield, L. (2002). Genes for glycosylphosphatidylinositol toxin biosynthesis in Plasmodium falciparum. Infect Immun 70, 4510-22.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J. y Guigo, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res 22, 1775-89.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H.,

Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigo, R. y Gingeras, T.R. (2012). Landscape of transcription in human cells. Nature 489, 101-8.

DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H. y Calos, M.P. (1987). Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 7, 379-87.

Ebralidze, A.K., Guibal, F.C., Steidl, U., Zhang, P., Lee, S., Bartholdy, B., Jorda, M.A., Petkova, V., Rosenbauer, F., Huang, G., Dayaram, T., Klupp, J., O'Brien, K.B., Will, B., Hoogenkamp, M., Borden, K.L., Bonifer, C. y Tenen, D.G. (2008). PU.1 expression is modulated by the balance of functional sense and antisense RNAs regulated by a shared cis-regulatory element. Genes Dev 22, 2085-92.

Eden, E., Navon, R., Steinfeld, I., Lipson, D. y Yakhini, Z. (2009). GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10, 48.

Esnault, C., Maestre, J. y Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24, 363-7.

Ewing, A.D., Ballinger, T.J., Earl, D., Broad Institute Genome, S., Analysis, P., Platform, Harris, C.C., Ding, L., Wilson, R.K. y Haussler, D. (2013). Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. Genome Biol 14, R22.

Faghihi, M.A. y Wahlestedt, C. (2009). Regulatory roles of natural antisense transcripts. Nat Rev Mol Cell Biol 10, 637-43.

Faghihi, M.A., Zhang, M., Huang, J., Modarresi, F., Van der Brug, M.P., Nalls, M.A., Cookson, M.R., St-Laurent, G., 3rd y Wahlestedt, C. (2010). Evidence for natural antisense transcript-mediated inhibition of microRNA function. Genome Biol 11, R56.

Filipovska, J. y Konarska, M.M. (2000). Specific HDV RNA-templated transcription by pol II in vitro. RNA 6, 41-54.

Flicek, P., Ahmed, I., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Garcia-Giron, C., Gordon, L., Hourlier, T., Hunt, S., Juettemann, T., Kahari, A.K., Keenan, S., Komorowska, M., Kulesha, E., Longden, I., Maurel, T., McLaren, W.M., Muffato, M., Nag, R., Overduin, B., Pignatelli, M., Pritchard, B., Pritchard, E., Riat, H.S., Ritchie, G.R., Ruffier, M., Schuster, M., Sheppard, D., Sobral, D., Taylor, K., Thormann, A., Trevanion, S., White, S., Wilder, S.P., Aken, B.L., Birney, E., Cunningham, F., Dunham, I., Harrow, J., Herrero, J., Hubbard, T.J., Johnson, N., Kinsella, R., Parker, A., Spudich, G., Yates, A., Zadissa, A. y Searle, S.M. (2013). Ensembl 2013. Nucleic Acids Res 41, D48-55.

French, C., Menegazzi, P., Nicholson, L., Macaulay, H., DiLuca, D. y Gompels, U.A. (1999). Novel, nonconsensus cellular splicing regulates expression of a gene encoding a chemokine-like protein that shows high variation and is specific for human herpesvirus 6. Virology 262, 139-51.

Frith, M.C., Wilming, L.G., Forrest, A., Kawaji, H., Tan, S.L., Wahlestedt, C., Bajic, V.B., Kai, C., Kawai, J., Carninci, P., Hayashizaki, Y., Bailey, T.L. y Huminiecki, L. (2006). Pseudo-messenger RNA: phantoms of the transcriptome. PLoS Genet 2, e23.

Fu, T.B. y Taylor, J. (1993). The RNAs of hepatitis delta virus are copied by RNA polymerase II in nuclear homogenates. J Virol 67, 6965-72.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., Diekhans, M., Dreszer, T.R., Giardine, B.M., Harte, R.A., Hillman-Jackson, J., Hsu, F., Kirkup, V., Kuhn, R.M., Learned, K., Li, C.H., Meyer, L.R., Pohl, A., Raney,

B.J., Rosenbloom, K.R., Smith, K.E., Haussler, D. y Kent, W.J. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res 39, D876-82.

Ge, X., Rubinstein, W.S., Jung, Y.C. y Wu, Q. (2008). Genome-wide analysis of antisense transcription with Affymetrix exon array. BMC Genomics 9, 27.

Ge, X., Wu, Q., Jung, Y.C., Chen, J. y Wang, S.M. (2006). A large quantity of novel human antisense transcripts detected by LongSAGE. Bioinformatics 22, 2475-9.

Geiszt, M., Lekstrom, K. y Leto, T.L. (2004). Analysis of mRNA transcripts from the NAD(P)H oxidase 1 (Nox1) gene. Evidence against production of the NADPH oxidase homolog-1 short (NOH-1S) transcript variant. J Biol Chem 279, 51661-8.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., Alves, P., Chateigner, A., Perry, M., Morris, M., Auerbach, R.K., Feng, X., Leng, J., Vielle, A., Niu, W., Rhrissorrakrai, K., Agarwal, A., Alexander, R.P., Barber, G., Brdlik, C.M., Brennan, J., Brouillet, J.J., Carr, A., Cheung, M.S., Clawson, H., Contrino, S., Dannenberg, L.O., Dernburg, A.F., Desai, A., Dick, L., Dose, A.C., Du, J., Egelhofer, T., Ercan, S., Euskirchen, G., Ewing, B., Feingold, E.A., Gassmann, R., Good, P.J., Green, P., Gullier, F., Gutwein, M., Guyer, M.S., Habegger, L., Han, T., Henikoff, J.G., Henz, S.R., Hinrichs, A., Holster, H., Hyman, T., Iniguez, A.L., Janette, J., Jensen, M., Kato, M., Kent, W.J., Kephart, E., Khivansara, V., Khurana, E., Kim, J.K., Kolasinska-Zwierz, P., Lai, E.C., Latorre, I., Leahey, A., Lewis, S., Lloyd, P., Lochovsky, L., Lowdon, R.F., Lubling, Y., Lyne, R., MacCoss, M., Mackowiak, S.D., Mangone, M., McKay, S., Mecenas, D., Merrihew, G., Miller, D.M., 3rd, Muroyama, A., Murray, J.I., Ooi, S.L., Pham, H., Phippen, T., Preston, E.A., Rajewsky, N., Ratsch, G., Rosenbaum, H., Rozowsky, J., Rutherford, K., Ruzanov, P., Sarov, M., Sasidharan, R., Sboner, A., Scheid, P., Segal, E., Shin, H., Shou, C., Slack, F.J., Slightam, C., Smith, R., Spencer, W.C., Stinson, E.O., Taing, S., Takasaki, T., Vafeados, D., Voronina, K., Wang, G., Washington, N.L., Whittle, C.M., Wu, B., Yan, K.K., Zeller, G., Zha, Z., Zhong, M., Zhou, X., mod, E.C., Ahringer, J., Strome, S., Gunsalus, K.C., Micklem, G., Liu, X.S., Reinke, V., Kim, S.K., Hillier, L.W., Henikoff, S., Piano, F., Snyder, M., Stein, L., Lieb, J.D. y Waterston, R.H. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330, 1775-87.

Ghildiyal, M., Seitz, H., Horwich, M.D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E.L., Zapp, M.L., Weng, Z. y Zamore, P.D. (2008). Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science 320, 1077-81.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., Miller, W., Kent, W.J. y Nekrutenko, A. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15, 1451-5.

Gluzman, Y. (1981). SV40-transformed simian cells support the replication of early SV40 mutants. Cell 23, 175-82.

Graham, F.L., Smiley, J., Russell, W.C. y Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol 36, 59-74.

Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., Brown, J.B., Cherbas, L., Davis, C.A., Dobin, A., Li, R., Lin, W., Malone, J.H., Mattiuzzo, N.R., Miller, D., Sturgill, D., Tuch, B.B., Zaleski, C., Zhang, D., Blanchette, M., Dudoit, S., Eads, B., Green, R.E., Hammonds, A., Jiang, L., Kapranov, P., Langton, L., Perrimon, N., Sandler, J.E., Wan, K.H., Willingham, A., Zhang, Y., Zou, Y., Andrews, J., Bickel, P.J., Brenner, S.E., Brent, M.R., Cherbas, P., Gingeras, T.R., Hoskins, R.A., Kaufman, T.C., Oliver, B. y Celniker, S.E. (2011). The developmental transcriptome of Drosophila melanogaster. Nature 471, 473-9.

Greco-Stewart, V.S., Miron, P., Abrahem, A. y Pelchat, M. (2007). The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. Virology 357, 68-78.

Greene, L.A. y Tischler, A.S. (1976). Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. Proc Natl Acad Sci U S A 73, 2424-8.

Gudima, S.O. y Taylor, J.M. (2001). Search for antisense copies of beta-globin mRNA in anemic mouse spleen. BMC Biochem 2, 3.

Haddad, F., Qin, A.X., Giger, J.M., Guo, H. y Baldwin, K.M. (2007). Potential pitfalls in the accuracy of analysis of natural sense-antisense RNA pairs by reverse transcription-PCR. BMC Biotechnol 7, 21.

Haeger, P., Cuevas, R., Forray, M.I., Rojas, R., Daza, C., Rivadeneira, J. y Gysling, K. (2005). Natural expression of immature Ucn antisense RNA in the rat brain. Evidence favoring bidirectional transcription of the Ucn gene locus. Brain Res Mol Brain Res 139, 115-28.

Haeseleer, F., Sokal, I., Verlinde, C.L., Erdjument-Bromage, H., Tempst, P., Pronin, A.N., Benovic, J.L., Fariss, R.N. y Palczewski, K. (2000). Five members of a novel Ca(2+)-binding protein (CABP) subfamily with similarity to calmodulin. J Biol Chem 275, 1247-60.

Hahn, M.W. y Wray, G.A. (2002). The g-value paradox. Evol Dev 4, 73-5.

Hamilton, A.J. y Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950-2.

Harrison, P.M., Zheng, D., Zhang, Z., Carriero, N. y Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. Nucleic Acids Res 33, 2374-83.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J.M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigo, R. y Hubbard, T.J. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22, 1760-74.

Hastings, M.L., Milcarek, C., Martincic, K., Peterson, M.L. y Munroe, S.H. (1997). Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res 25, 4296-300.

Hatzoglou, A., Deshayes, F., Madry, C., Lapree, G., Castanas, E. y Tsapis, A. (2002). Natural antisense RNA inhibits the expression of BCMA, a tumour necrosis factor receptor homologue. BMC Mol Biol 3, 4.

Haussecker, D., Cao, D., Huang, Y., Parameswaran, P., Fire, A.Z. y Kay, M.A. (2008). Capped small RNAs and MOV10 in human hepatitis delta virus replication. Nat Struct Mol Biol 15, 714-21.

Hawkins, P.G. y Morris, K.V. (2010). Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Transcription 1, 165-175.

Hayward, B.E. y Bonthron, D.T. (2000). An imprinted antisense transcript at the human GNAS1 locus. Hum Mol Genet 9, 835-41.

He, C., Zhou, F., Zuo, Z., Cheng, H. y Zhou, R. (2009). A global view of cancer-specific transcript variants by subtractive transcriptome-wide analysis. PLoS One 4, e4732.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N. y Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science 322, 1855-7.

Houseley, J. y Tollervey, D. (2010). Apparent non-canonical trans-splicing is generated by reverse transcriptase in vitro. PLoS One 5, e12271.

International Human Genome Sequencing, C. (2004). Finishing the euchromatic sequence of the human genome. Nature 431, 931-45.

Jacob, F. y Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3, 318-56.

Kalyana-Sundaram, S., Kumar-Sinha, C., Shankar, S., Robinson, D.R., Wu, Y.M., Cao, X., Asangani, I.A., Kothari, V., Prensner, J.R., Lonigro, R.J., Iyer, M.K., Barrette, T., Shanmugam, A., Dhanasekaran, S.M., Palanisamy, N. y Chinnaiyan, A.M. (2012). Expressed pseudogenes in the transcriptional landscape of human cancers. Cell 149, 1622-34.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermuller, J., Hofacker, I.L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H. y Gingeras, T.R. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science 316, 1484-8.

Kapranov, P., Drenkow, J., Cheng, J., Long, J., Helt, G., Dike, S. y Gingeras, T.R. (2005). Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res 15, 987-97.

Kapranov, P., Ozsolak, F., Kim, S.W., Foissac, S., Lipson, D., Hart, C., Roels, S., Borel, C., Antonarakis, S.E., Monaghan, A.P., John, B. y Milos, P.M. (2010a). New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. Nature 466, 642-6.

Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C.P., Sorensen, P.H., Reaman, G., Milos, P., Arceci, R.J., Thompson, J.F. y Triche, T.J. (2010b). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. BMC Biol 8, 149.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D. y Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32, D493-6.

Karro, J.E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrrison, P. y Gerstein, M. (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. Nucleic Acids Res 35, D55-60.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K.C., Hallinan, J., Mattick, J., Hume, D.A., Lipovich, L., Batalov, S., Engstrom, P.G., Mizuno, Y., Faghihi, M.A., Sandelin, A., Chalk, A.M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C., Group, R.G.E.R., Genome Science, G. y Consortium, F. (2005). Antisense transcription in the mammalian transcriptome. Science 309, 1564-6.

Kawaji, H., Nakamura, M., Takahashi, Y., Sandelin, A., Katayama, S., Fukuda, S., Daub, C.O., Kai, C., Kawai, J., Yasuda, J., Carninci, P. y Hayashizaki, Y. (2008). Hidden layers of human small RNAs. BMC Genomics 9, 157.

Keeling, P.J. (2003). Congruent evidence from alpha-tubulin and beta-tubulin gene phylogenies for a zygomycete origin of microsporidia. Fungal Genet Biol 38, 298-309.

Khelifi, A., Duret, L. y Mouchiroud, D. (2005). HOPPSIGEN: a database of human and mouse processed pseudogenes. Nucleic Acids Res 33, D59-66.

Kim, T., Xu, Z., Clauder-Munster, S., Steinmetz, L.M. y Buratowski, S. (2012). Set3 HDAC mediates effects of overlapping noncoding transcription on gene induction kinetics. Cell 150, 1158-69.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G. y Greenberg, M.E. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182-7.

Kimble, J. y Hirsh, D. (1979). The postembryonic cell lineages of the hermaphrodite and male gonads in Caenorhabditis elegans. Dev Biol 70, 396-417.

Klattenhoff, C. y Theurkauf, W. (2008). Biogenesis and germline functions of piRNAs. Development 135, 3-9.

Korbie, D.J. y Mattick, J.S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc 3, 1452-6.

Kramerov, D.A. y Vassetzky, N.S. (2005). Short retroposons in eukaryotic genomes. Int Rev Cytol 247, 165-221.

Krystal, G.W., Armstrong, B.C. y Battey, J.F. (1990). N-myc mRNA forms an RNA-RNA duplex with endogenous antisense transcripts. Mol Cell Biol 10, 4180-91.

Kuo, M.Y., Chao, M. y Taylor, J. (1989). Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. J Virol 63, 1945-50.

Kurose, K., Koyano, S., Ikeda, S., Tohkin, M., Hasegawa, R. y Sawada, J. (2005). 5' diversity of human hepatic PXR (NR1I2) transcripts and identification of the major transcription initiation site. Mol Cell Biochem 273, 79-85.

Laabi, Y., Gras, M.P., Brouet, J.C., Berger, R., Larsen, C.J. y Tsapis, A. (1994). The BCMA gene, preferentially expressed during B lymphoid maturation, is bidirectionally transcribed. Nucleic Acids Res 22, 1147-54.

Ladd, P.D., Smith, L.E., Rabaia, N.A., Moore, J.M., Georges, S.A., Hansen, R.S., Hagerman, R.J., Tassone, F., Tapscott, S.J. y Filippova, G.N. (2007). An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet 16, 3174-87.

Lai, M.M. (2005). RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. J Virol 79, 7951-8.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A. y Huala, E. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res 40, D1202-10.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R.,

Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J. y International Human Genome Sequencing, C. (2001). Initial sequencing and analysis of the human genome. Nature 409, 860-921.

Lanford, R.E., Sureau, C., Jacob, J.R., White, R. y Fuerst, T.R. (1994). Demonstration of in vitro infection of chimpanzee hepatocytes with hepatitis C virus using strand-specific RT/PCR. Virology 202, 606-14.

Langmead, B., Trapnell, C., Pop, M. y Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10, R25.

Lapidot, M. y Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. EMBO Rep 7, 1216-22.

Lavorgna, G., Dahary, D., Lehner, B., Sorek, R., Sanderson, C.M. y Casari, G. (2004). In search of antisense. Trends Biochem Sci 29, 88-94.

Lee, J.T. (2012). Epigenetic regulation by long noncoding RNAs. Science 338, 1435-9.

Lee, J.T., Davidow, L.S. y Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. Nat Genet 21, 400-4.

Lee, R.C., Feinbaum, R.L. y Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-54.

Lee, Y.S., Shibata, Y., Malhotra, A. y Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes Dev 23, 2639-49.

Lehmann, E., Brueckner, F. y Cramer, P. (2007). Molecular basis of RNA-dependent RNA polymerase II activity. Nature 450, 445-9.

Levin, J.Z., Yassour, M., Adiconis, X., Nusbaum, C., Thompson, D.A., Friedman, N., Gnirke, A. y Regev, A. (2010). Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat Methods 7, 709-15.

Lin, C.F., Mount, S.M., Jarmolowski, A. y Makalowski, W. (2010). Evolutionary dynamics of U12-type spliceosomal introns. BMC Evol Biol 10, 47.

Mader, R.M., Schmidt, W.M., Sedivy, R., Rizovski, B., Braun, J., Kalipciyan, M., Exner, M., Steger, G.G. y Mueller, M.W. (2001). Reverse transcriptase template switching during reverse transcriptase-polymerase chain reaction: artificial generation of deletions in ribonucleotide reductase mRNA. J Lab Clin Med 137, 422-8.

Magistri, M., Faghihi, M.A., St Laurent, G., 3rd y Wahlestedt, C. (2012). Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. Trends Genet 28, 389-96.

Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W.C. y Masutomi, K. (2009). An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. Nature 461, 230-5.

Manak, J.R., Dike, S., Sementchenko, V., Kapranov, P., Biemar, F., Long, J., Cheng, J., Bell, I., Ghosh, S., Piccolboni, A. y Gingeras, T.R. (2006). Biological function of unannotated transcription during the early development of Drosophila melanogaster. Nat Genet 38, 1151-8.

Mandal, P.K., Ewing, A.D., Hancks, D.C. y Kazazian, H.H., Jr. (2013). Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles. Hum Mol Genet 22, 3730-48.

Matsui, K., Nishizawa, M., Ozaki, T., Kimura, T., Hashimoto, I., Yamada, M., Kaibori, M., Kamiyama, Y., Ito, S. y Okumura, T. (2008). Natural antisense transcript stabilizes inducible nitric oxide synthase messenger RNA in rat hepatocytes. Hepatology 47, 686-97.

Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep 2, 986-91.

Mattick, J.S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays 25, 930-9.

Mattick, J.S. (2009). Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. Ann N Y Acad Sci 1178, 29-46.

Mei, Y., Xie, C., Xie, W., Wu, Z. y Wu, M. (2007). Siah-1S, a novel splice variant of Siah-1 (seven in absentia homolog), counteracts Siah-1-mediated downregulation of beta-catenin. Oncogene 26, 6319-31.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S.D., Gregersen, L.H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F. y Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. Nature 495, 333-8.

Merzendorfer, H., Harvey, W.R. y Wieczorek, H. (1997). Sense and antisense RNA for the membrane associated 40 kDa subunit M40 of the insect V-ATPase. FEBS Lett 411, 239-44.

Mestdagh, P., Fredlund, E., Pattyn, F., Rihani, A., Van Maerken, T., Vermeulen, J., Kumps, C., Menten, B., De Preter, K., Schramm, A., Schulte, J., Noguera, R., Schleiermacher, G., Janoueix-Lerosey, I., Laureys, G., Powel, R., Nittner, D., Marine, J.C., Ringner, M., Speleman, F. y Vandesompele, J. (2010). An integrative genomics screen uncovers ncRNA T-UCR functions in neuroblastoma tumours. Oncogene 29, 3583-92.

Modahl, L.E., Macnaughton, T.B., Zhu, N., Johnson, D.L. y Lai, M.M. (2000). RNA-Dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. Mol Cell Biol 20, 6030-9.

Modarresi, F., Faghihi, M.A., Lopez-Toledano, M.A., Fatemi, R.P., Magistri, M., Brothers, S.P., van der Brug, M.P. y Wahlestedt, C. (2012). Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. Nat Biotechnol 30, 453-9.

Mohammad, F., Mondal, T. y Kanduri, C. (2009). Epigenetics of imprinted long noncoding RNAs. Epigenetics 4, 277-86.

Moraleda, G. y Taylor, J. (2001). Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. J Virol 75, 10161-9.

Morgulis, A., Gertz, E.M., Schaffer, A.A. y Agarwala, R. (2006). A fast and symmetric DUST implementation to mask low-complexity DNA sequences. J Comput Biol 13, 1028-40.

Morris, K.V., Santoso, S., Turner, A.M., Pastori, C. y Hawkins, P.G. (2008). Bidirectional transcription directs both transcriptional gene activation and suppression in human cells. PLoS Genet 4, e1000258.

Morrissy, A.S., Griffith, M. y Marra, M.A. (2011). Extensive relationship between antisense transcription and alternative splicing in the human genome. Genome Res 21, 1203-12.

Munroe, S.H. y Lazar, M.A. (1991). Inhibition of c-erbA mRNA splicing by a naturally occurring antisense RNA. J Biol Chem 266, 22083-6.

Muro, E.M. y Andrade-Navarro, M.A. (2010). Pseudogenes as an alternative source of natural antisense transcripts. BMC Evol Biol 10, 338.

Nam, D.K., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., Chen, J., Rowley, J.D. y Wang, S.M. (2002). Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. Proc Natl Acad Sci U S A 99, 6152-6.

Nam, J.W. y Bartel, D.P. (2012). Long noncoding RNAs in C. elegans. Genome Res 22, 2529-40.

Neeman, Y., Dahary, D., Levanon, E.Y., Sorek, R. y Eisenberg, E. (2005). Is there any sense in antisense editing? Trends Genet 21, 544-7.

Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y. y Okada, N. (2003). Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol 4, R74.

Okada, T., Takagi, M., Murata, S., Onuma, M. y Ohashi, K. (2007). Identification and characterization of a novel spliced form of the meq transcript in lymphoblastoid cell lines derived from Marek's disease tumours. J Gen Virol 88, 2111-20.

Okamura, K., Balla, S., Martin, R., Liu, N. y Lai, E.C. (2008). Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol 15, 581-90.

Okamura, K. y Lai, E.C. (2008). Endogenous small interfering RNAs in animals. Nat Rev Mol Cell Biol 9, 673-8.

Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., Guigo, R. y Shiekhattar, R. (2010). Long noncoding RNAs with enhancer-like function in human cells. Cell 143, 46-58.

Osato, N., Suzuki, Y., Ikeo, K. y Gojobori, T. (2007). Transcriptional interferences in cis natural antisense transcripts of humans and mice. Genetics 176, 1299-306.

Ozsolak, F., Kapranov, P., Foissac, S., Kim, S.W., Fishilevich, E., Monaghan, A.P., John, B. y Milos, P.M. (2010). Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. Cell 143, 1018-29.

Palmer, A.C., Egan, J.B. y Shearwin, K.E. (2011). Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. Transcription 2, 9-14.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H. y Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res 37, e123.

Pei, B., Sisu, C., Frankish, A., Howald, C., Habegger, L., Mu, X.J., Harte, R., Balasubramanian, S., Tanzer, A., Diekhans, M., Reymond, A., Hubbard, T.J., Harrow, J. y Gerstein, M.B. (2012). The GENCODE pseudogene resource. Genome Biol 13, R51.

Perocchi, F., Xu, Z., Clauder-Munster, S. y Steinmetz, L.M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res 35, e128.

Peters, N.T., Rohrbach, J.A., Zalewski, B.A., Byrkett, C.M. y Vaughn, J.C. (2003). RNA editing and regulation of Drosophila 4f-rnp expression by sas-10 antisense readthrough mRNA transcripts. RNA 9, 698-710.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W. y Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell 127, 1209-21.

Pink, R.C., Wicks, K., Caley, D.P., Punch, E.K., Jacobs, L. y Carter, D.R. (2011). Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17, 792-8.

Podlowski, S., Bramlage, P., Baumann, G., Morano, I. y Luther, H.P. (2002). Cardiac troponin I sense-antisense RNA duplexes in the myocardium. J Cell Biochem 85, 198-207.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H. y Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science 322, 1851-4.

Prescott, E.M. y Proudfoot, N.J. (2002). Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci U S A 99, 8796-801.

Ravasi, T., Suzuki, H., Pang, K.C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M.C., Gongora, M.M., Grimmond, S.M., Hume, D.A., Hayashizaki, Y. y Mattick, J.S. (2006).

Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. Genome Res 16, 11-9.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. y Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311-23.

Rosok, O. y Sioud, M. (2004). Systematic identification of sense-antisense transcripts in mammalian cells. Nat Biotechnol 22, 104-8.

Rougeulle, C., Cardoso, C., Fontes, M., Colleaux, L. y Lalande, M. (1998). An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. Nat Genet 19, 15-6.

Scotto-Lavino, E., Du, G. y Frohman, M.A. (2006). 3' end cDNA amplification using classic RACE. Nat Protoc 1, 2742-5.

Schrider, D.R., Navarro, F.C., Galante, P.A., Parmigiani, R.B., Camargo, A.A., Hahn, M.W. y de Souza, S.J. (2013). Gene copy-number polymorphism caused by retrotransposition in humans. PLoS Genet 9, e1003242.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A. y Sharp, P.A. (2008). Divergent transcription from active promoters. Science 322, 1849-51.

Sharp, P.A. (2009). The centrality of RNA. Cell 136, 577-80.

Shen, M., Eyras, E., Wu, J., Khanna, A., Josiah, S., Rederstorff, M., Zhang, M.Q. y Stamm, S. (2011). Direct cloning of double-stranded RNAs from RNase protection analysis reveals processing patterns of C/D box snoRNAs and provides evidence for widespread antisense transcript expression. Nucleic Acids Res 39, 9720-30.

Shendure, J. y Church, G.M. (2002). Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol 3, RESEARCH0044.

Sleutels, F., Zwart, R. y Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature 415, 810-3.

Slomovic, S., Fremder, E., Staals, R.H., Pruijn, G.J. y Schuster, G. (2010). Addition of poly(A) and poly(A)-rich tails during RNA degradation in the cytoplasm of human cells. Proc Natl Acad Sci U S A 107, 7407-12.

Song, R., Hennig, G.W., Wu, Q., Jose, C., Zheng, H. y Yan, W. (2011). Male germ cells express abundant endogenous siRNAs. Proc Natl Acad Sci U S A 108, 13159-64.

Su, W.Y., Li, J.T., Cui, Y., Hong, J., Du, W., Wang, Y.C., Lin, Y.W., Xiong, H., Wang, J.L., Kong, X., Gao, Q.Y., Wei, L.P. y Fang, J.Y. (2012). Bidirectional regulation between WDR83 and its natural antisense transcript DHPS in gastric cancer. Cell Res 22, 1374-89.

Sulston, J.E. y Horvitz, H.R. (1977). Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. Dev Biol 56, 110-56.

Sulston, J.E., Schierenberg, E., White, J.G. y Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol 100, 64-119.

Svensson, O., Arvestad, L. y Lagergren, J. (2006). Genome-wide survey for biologically functional pseudogenes. PLoS Comput Biol 2, e46.

Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., Irvine, K., Arakawa, T., Nakamura, M., Kubosaki, A., Hayashida, K., Kawazu, C., Murata, M., Nishiyori, H., Fukuda, S., Kawai, J., Daub, C.O., Hume, D.A., Suzuki, H., Orlando, V., Carninci, P., Hayashizaki, Y. y Mattick, J.S. (2009a). Tiny RNAs associated with transcription start sites in animals. Nat Genet 41, 572-8.

Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P. y Mattick, J.S. (2009b). Small RNAs derived from snoRNAs. RNA 15, 1233-40.

Taft, R.J., Pheasant, M. y Mattick, J.S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays 29, 288-99.

Taft, R.J., Simons, C., Nahkuri, S., Oey, H., Korbie, D.J., Mercer, T.R., Holst, J., Ritchie, W., Wong, J.J., Rasko, J.E., Rokhsar, D.S., Degnan, B.M. y Mattick, J.S. (2010). Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans. Nat Struct Mol Biol 17, 1030-4.

Tam, O.H., Aravin, A.A., Stein, P., Girard, A., Murchison, E.P., Cheloufi, S., Hodges, E., Anger, M., Sachidanandam, R., Schultz, R.M. y Hannon, G.J. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature 453, 534-8.

Taylor, J.M. (2009). Chapter 3. Replication of the hepatitis delta virus RNA genome. Adv Virus Res 74, 103-21.

Thakur, N., Tiwari, V.K., Thomassin, H., Pandey, R.R., Kanduri, M., Gondor, A., Grange, T., Ohlsson, R. y Kanduri, C. (2004). An antisense RNA regulates the bidirectional silencing property of the Kcnq1 imprinting control region. Mol Cell Biol 24, 7855-62.

Tonner, P., Srinivasasainagendra, V., Zhang, S. y Zhi, D. (2012). Detecting transcription of ribosomal protein pseudogenes in diverse human tissues from RNA-seq data. BMC Genomics 13, 412.

Torrents, D., Suyama, M., Zdobnov, E. y Bork, P. (2003). A genome-wide survey of human pseudogenes. Genome Res 13, 2559-67.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. y Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome Res 14, 62-6.

Tufarelli, C., Stanley, J.A., Garrick, D., Sharpe, J.A., Ayyub, H., Wood, W.G. y Higgs, D.R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34, 157-65.

Uchida, T., Rossignol, F., Matthay, M.A., Mounier, R., Couette, S., Clottes, E. y Clerici, C. (2004). Prolonged hypoxia differentially regulates hypoxia-inducible factor (HIF)-1alpha and HIF-2alpha expression in lung epithelial cells: implication of natural antisense HIF-1alpha. J Biol Chem 279, 14871-8.

Verona, R.I., Mann, M.R. y Bartolomei, M.S. (2003). Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol 19, 237-59.

Volloch, V. (1986). Cytoplasmic synthesis of globin RNA in differentiated murine erythroleukemia cells: possible involvement of RNA-dependent RNA polymerase. Proc Natl Acad Sci U S A 83, 1208-12.

Volloch, V., Schweitzer, B. y Rits, S. (1987). Synthesis of globin RNA in enucleated differentiating murine erythroleukemia cells. J Cell Biol 105, 137-43.

Volloch, V., Schweitzer, B. y Rits, S. (1996). Antisense globin RNA in mouse erythroid tissues: structure, origin, and possible function. Proc Natl Acad Sci U S A 93, 2476-81.

Wagner, S.D., Yakovchuk, P., Gilman, B., Ponicsan, S.L., Drullinger, L.F., Kugel, J.F. y Goodrich, J.A. (2013). RNA polymerase II acts as an RNA-dependent RNA polymerase to extend and destabilize a non-coding RNA. EMBO J 32, 781-90.

Wan, L.B. y Bartolomei, M.S. (2008). Regulation of imprinting in clusters: noncoding RNAs versus insulators. Adv Genet 61, 207-23.

Wang, Z., Gerstein, M. y Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57-63.

Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., Chiba, H., Kohara, Y., Kono, T., Nakano, T., Surani, M.A., Sakaki, Y. y Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature 453, 539-43.

Werner, A., Cockell, S., Falconer, J., Carlile, M., Alnumeir, S. y Robinson, J. (2014). Contribution of natural antisense transcription to an endogenous siRNA signature in human cells. BMC Genomics 15, 19.

West, S., Gromak, N., Norbury, C.J. y Proudfoot, N.J. (2006). Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. Mol Cell 21, 437-43.

White, J.G., Southgate, E., Thomson, J.N. y Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci 314, 1-340.

Whitehouse, I., Rando, O.J., Delrow, J. y Tsukiyama, T. (2007). Chromatin remodelling at promoters suppresses antisense transcription. Nature 450, 1031-5.

Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J. y Bahler, J. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453, 1239-43.

Willis, T.G., Jadayel, D.M., Du, M.Q., Peng, H., Perry, A.R., Abdul-Rauf, M., Price, H., Karran, L., Majekodunmi, O., Wlodarska, I., Pan, L., Crook, T., Hamoudi, R., Isaacson, P.G. y Dyer, M.J. (1999). Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. Cell 96, 35-45.

Wu, J.Q., Du, J., Rozowsky, J., Zhang, Z., Urban, A.E., Euskirchen, G., Weissman, S., Gerstein, M. y Snyder, M. (2008). Systematic analysis of transcribed loci in ENCODE regions using RACE sequencing reveals extensive transcription in the human genome. Genome Biol 9, R3.

Wu, T.D. y Watanabe, C.K. (2005). GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21, 1859-75.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.C., Dufour, M.E., Boulay, J., Regnault, B., Devaux, F., Namane, A., Seraphin, B., Libri, D. y Jacquier, A. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell 121, 725-37.

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M.M. y Handa, H. (2001). Stimulation of RNA polymerase II elongation by hepatitis delta antigen. Science 293, 124-7.

Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S. y Handa, H. (2007). Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. Genes Cells 12, 863-75.

Yelin, R., Dahary, D., Sorek, R., Levanon, E.Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K. y Rotman, G. (2003). Widespread occurrence of antisense transcription in the human genome. Nat Biotechnol 21, 379-86.

Yook, K., Harris, T.W., Bieri, T., Cabunoc, A., Chan, J., Chen, W.J., Davis, P., de la Cruz, N., Duong, A., Fang, R., Ganesan, U., Grove, C., Howe, K., Kadam, S., Kishore, R., Lee, R., Li, Y., Muller, H.M., Nakamura, C., Nash, B., Ozersky, P., Paulini, M., Raciti, D., Rangarajan, A., Schindelman, G., Shi, X., Schwarz, E.M., Ann Tuli, M., Van Auken, K., Wang, D., Wang, X., Williams, G., Hodgkin, J., Berriman, M., Durbin, R., Kersey, P., Spieth, J., Stein, L. y Sternberg, P.W. (2012). WormBase 2012: more genomes, more data, new website. Nucleic Acids Res 40, D735-41.

Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A.P. y Cui, H. (2008). Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. Nature 451, 202-6.

Zhang, Q., Siebert, R., Yan, M., Hinzmann, B., Cui, X., Xue, L., Rakestraw, K.M., Naeve, C.W., Beckmann, G., Weisenburger, D.D., Sanger, W.G., Nowotny, H., Vesely, M., Callet-Bauchu, E., Salles, G., Dixit, V.M., Rosenthal, A., Schlegelberger, B. y Morris, S.W. (1999). Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). Nat Genet 22, 63-8.

Zhang, Y., Liu, X.S., Liu, Q.R. y Wei, L. (2006). Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. Nucleic Acids Res 34, 3465-75.

Zhang, Z., Harrison, P.M., Liu, Y. y Gerstein, M. (2003). Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res 13, 2541-58.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M. y Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol Cell 40, 939-53.

Zheng, D., Frankish, A., Baertsch, R., Kapranov, P., Reymond, A., Choo, S.W., Lu, Y., Denoeud, F., Antonarakis, S.E., Snyder, M., Ruan, Y., Wei, C.L., Gingeras, T.R., Guigo, R., Harrow, J. y Gerstein, M.B. (2007). Pseudogenes in the ENCODE regions: consensus annotation, analysis of transcription, and evolution. Genome Res 17, 839-51.

Zheng, Q., Ryvkin, P., Li, F., Dragomir, I., Valladares, O., Yang, J., Cao, K., Wang, L.S. y Gregory, B.D. (2010). Genome-wide double-stranded RNA sequencing reveals the functional significance of base-paired RNAs in Arabidopsis. PLoS Genet 6, e1001141.

Zhou, B.S., Beidler, D.R. y Cheng, Y.C. (1992). Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene. Cancer Res 52, 4280-5.