



**Caracterización de la región cromosómica  
15q11-q13 del genoma humano**  
Variabilidad genómica en el autismo e  
identificación de ncRNAs

Celia Cerrato Rivera

Memoria presentada para optar al grado de doctora por la Universidad Pompeu Fabra

Esta tesis ha sido dirigida por el Dr. Xavier Estivill Pallejà del programa de Genes y Enfermedad,  
del Centro de Regulación Genómica

Tesis doctoral inscrita en el Departamento de Ciencias Experimentales y de la Salud,  
Universidad Pompeu Fabra

Programa de doctorado de Ciencias de la Salud y de la Vida  
Bienio 2003-2005

Xavier Estivill Pallejà

Celia Cerrato Rivera

Barcelona, Abril 2007

Esta tesis doctoral ha sido soportada por las becas FPU del Ministerio de Educación y Ciencia, el Fondo de investigaciones sanitarias (Red de psiquiatría genética, G03 y CIBER-CB06/02/0058), el Departament d'Universitats i Societat de la Informació, Generalitat de Catalunya (2005SGR00008) y Genoma España.

D.L.: B.57455-2007

ISBN: 978-84-691-1204-5

## Índice

<b>Agradecimientos</b>	<b>V</b>
<b>Prefacio</b>	<b>IX</b>
<b>Introducción</b>	<b>1</b>
1. Organización génica de la región cromosómica 15q11-q13	3
1.1 snoRNAs	4
2. <i>Imprinting</i>	7
2.1 Modificaciones de los genes improntados	7
2.2 <i>Imprinting</i> y enfermedades genómicas	8
2.3 Factores ambientales que afectan al <i>imprinting</i>	9
2.4 Dominio improntado de la región cromosómica 15q11-q13	10
3. Variabilidad genómica	12
3.1 Duplicaciones Segmentarias y reordenamientos genómicos	12
3.2 <i>Copy Number Variants</i> (CNVs)	14
3.3 Duplicaciones segmentarias de la región 15q11-q13	15
3.4 Variabilidad genómica de 15q11-q13	16
3.5 <i>Gaps</i> de secuencia	17
4. Autismo	19
4.1 Causas genéticas del autismo	19
4.1.1 Análisis de ligamiento y de asociación	20
4.1.2 Alteraciones genómicas	22
4.1.3 Otras enfermedades y síndromes asociados al autismo	23
5. Región cromosómica 15q11-q13 y enfermedades genómicas	24
5.1 Cromosomas marcadores inv dup(15)	24
5.2 Síndromes de Angelman y Prader-Willi	24
5.3 Autismo	26

6. Non-coding RNAs	28
6.1 Identificación y detección de ncRNAs	29
6.2 Biogénesis de miRNAs	31
6.3 Regulación de la expresión génica por miRNAs	33
6.4 Funciones de ncRNAs	34
6.5 ncRNAs y enfermedades complejas	36
6.6 Región 15q11.2 del genoma humano: candidata para codificar ncRNAs	39
<b>Objetivos</b>	<b>44</b>
<b>Materiales y métodos</b>	<b>46</b>
1. Reordenamientos de 15q11-q13	48
1.1 Muestras	48
1.2 Selección y amplificación de marcadores microsatélites	48
1.3 Análisis estadístico	50
1.4 Identificación de la inserción en el gen <i>GABRG3</i>	50
2 Identificación de ncRNAs	51
2.1 Predicción bioinformática de nuevos ncRNAs	51
2.2 Aislamiento de RNA	51
2.3 <i>Northern Blot</i>	51
2.4 <i>Primer Extension</i>	52
2.5 Herramientas bioinformáticas para el análisis de ncRNAs	54
2.6 Predicción de <i>targets</i> de miRNAs	54

<b>Resultados</b>	<b>57</b>
1. Definición de reordenamientos y <i>Copy Number Variants</i> (CNVs) de la región cromosómica 15q11.2 en muestras controles y de pacientes autistas	59
1.1 Identificación de reordenamientos polimórficos en 15q11.2	59
1.2 Asociación de las deleciones identificadas en 15q11.2 con el trastorno autista	65
1.3 Ausencia de reordenamientos en la parte más telomérica de 15q11-q13	67
1.4 Inserción en el gen <i>GABRG3</i>	67
2. Identificación de ncRNAs específicos de primates en la región cromosómica 15q11.2 del genoma humano	69
2.1 Selección de ncRNAs candidatos	69
2.2 Expresión y caracterización de nuevos ncRNAs	70
2.3 Definición de las secuencias de los ncRNAs identificados en 15q11.2	74
2.4 Análisis bioinformático de los nuevos ncRNAs	76
2.5 Predicción de posibles <i>targets</i> de los miRNAs identificados en 15q11.2	78
2.6 Material suplementario	80
<b>Discusión</b>	<b>87</b>
<b>Conclusiones</b>	<b>99</b>
<b>Referencias</b>	<b>101</b>



## ***Agradecimientos***

Sin duda estas páginas son el resultado del trabajo de muchas personas y estoy segura de que no podría haberme rodeado de mejores personas ni mejores profesionales. Gracias a todos los momentos duros no han sido tan duros, las satisfacciones han sido mayores y volvería a vivir cada uno de ellos sólo por compartirlo con vosotros. Espero no olvidarme de nadie.

A Xavier Estivill, por darme la oportunidad de conocer el mundo de la ciencia, por haber confiado en mí y en este proyecto, porque siempre encuentra las mejores alternativas y las más novedosas.

A Mira, por el tiempo que le ha dedicado a este proyecto, por preocuparse de que saliera adelante, por sus consejos, por decirme mil veces "Celia, don't waste your time" y sobre todo por ser tan especial.

A Aurea, porque con ella la burocracia es más fácil y porque siempre está pendiente de todo y de todos.

A todo el laboratorio P13, por el día a día, por lo que nos hemos reído, por estar siempre dispuestos a escuchar y opinar y por supuesto, por enseñarme lo mejor que sabéis. A Mònica Gratacòs, porque es la mejor resolviendo dudas, sabe de todo y lo comparte encantada, por tu buen humor, por las aventuras en radiactividad con Marta Morell y por un mail que me encantó. A Lluís, por su ayuda con la informática, el tiempo que me ha ahorrado, su interés en este proyecto y los desayunos de los viernes a las ocho. A Pili y Mili: a Marta Ribasés por ser el primer ejemplo y un buen ejemplo, por su sinceridad, su espontaneidad, las demostraciones de flamenco y no ocultar sus emociones, y a Marta Morell porque nadie me entiende mejor que ella, en lo nerviosa que te puede poner el desorden. A pesar de todo, no se rinde y gracias a ella el laboratorio funciona mucho mejor. A Imma porque no hace falta que hable mucho para saber que siempre se puede contar con ella. Nos parecemos demasiado para que haga falta decir nada más. A Rafa por ser un buen científico y mejor persona. A Yolanda por las tensiones que hemos liberado juntas y porque siempre que la he necesitado ha estado disponible, aunque no haya sido fácil. A Mònica Bayés por todas sus ideas y críticas constructivas. A Txema y a mis niñas Ester, Nina y Maya, por ser los mejores amigos y los mejores compañeros. A Ester por ser la referencia más próxima, porque saber que está ella me hace estar más tranquila, porque es capaz de dejarlo todo cuando la necesitas, por mil momentos, mil conversaciones y porque es una gran persona. A Nina porque sabe hacerte reír cuando estás deseando llorar, por enseñarme los mejores sitios de Barcelona,

por ser la mejor organizadora de fiestas, porque es la única que no se da cuenta de lo patosa que soy, si no nunca hubiera ido conmigo a esquiar ni a una clase de danza africana. A Maya porque es la alegría en persona, por ser tan cariñosa, porque tiene el móvil siempre disponible y por preocuparse de mi alimentación (se acabó el whiskas). A Txema por su ayuda con el FBAT y su increíble sentido del humor, no cambies nunca!!!. A Marina por contagiarnos con sus rarezas. A Mónica Guidi por compartir ideas y puntos de vista. A Ester Saus porque el poco tiempo que hemos coincidido ha sido suficiente para estar contigo como con los demás, se nota que eres del P13. A Sergi, por compartir el interés por el cromosoma 15. A todos los técnicos: Anna Puig, Anna Carreras, Cecilia, Kristin y Carles, por su dedicación. A Anna Carreras por tener siempre lista una placa de DNAs, por integrarme en Barcelona y porque casi sin conocerme, se atrevía a compartir piso conmigo. Gracias a todos por organizarme una fiesta de “despedida” que nunca olvidaré.

A todo el programa de Genes y Enfermedad, por ser el programa más divertido del CRG y por hacer posible una convivencia ejemplar. Al P15 por ser los mejores vecinos. A Meri por ser la persona más dulce que conozco. A Anna Cascante por sufrir juntas los inconvenientes de la FPU. A Dani por estar siempre listo para las actividades extraexcolares y por recordarme mis primeros meses en Barcelona. A Nuria porque disfruta de cualquier conversación. A Jon porque es un buen amigo, por enseñarme la fiesta de Barcelona, por todas las tardes de confesiones y risas en mil sitios diferentes y porque sigue acordándose de mi cuando va al “Chatelet”.

Gracias a Manel Esteller y a todo el laboratio de Epigenética del CNIO por hacerme un hueco y tratarme tan bien en esta última etapa. Gracias a Amaia por enseñarme a trabajar en cultivos, por compartir la inquietud por los miRNAs y por recordarme a mí hace cuatro años. No pierdas nunca la ilusión!!. Gracias a Isabel por sus buenos consejos, por hacerme ver más allá y por todo lo que desinteresadamente ha aportado a este proyecto. Gracias a Cecilia, Rocio, Paz y Filipe por hacer de nuestro rincón el mejor ambiente de trabajo. A Paz por su confianza, los cafés y esos minutos de descanso que nos hacen ver todo de otra manera. A Rocio por enviarme todos los papers durante las últimas semanas, acompañados de mucho ánimo y mucho cariño, por intercambiar experiencias y por ser como eres. A Cecilia por ser la primera en acompañarme por las mañanas, enseñarme a clonar y aportar un poco de diversidad al grupo. A Filipe por preocuparse de cómo esto iba avanzando, ser un gran científico, acoger tan bien a los nuevos, y porque es un tio genial. A Irene porque siempre tiene lista una sonrisa, su cariño, sus consejos y porque sabe escuchar. A Agus por intentar poner orden y porque no hace falta hablar mucho con él para saber que es otra buena

persona. A Fer por hacer todos los pedidos, preparar las células y por su indispensable labor en el laboratorio.

A José M<sup>a</sup> Fernández por su apoyo en los análisis bioinformáticos, instalar y ejecutar todos los programas que he necesitado, analizar todas las secuencias que han hecho falta y por hacerlo siempre de buena gana y con buen humor.

A Ana y Elena, por convertir una hora de viaje, en una hora de risas, de buena e interesante conversación y de terapia, cuando ha hecho falta. A Ana por animarme a ir en coche hasta Madrid y por no pasar miedo (o por lo menos disimularlo muy bien) el primer día que yo conducía. A Elena, por contagiar alegría y no parar de reír a las siete de la mañana, y por alegrarnos la vista con las mejores fotos.

A Olaia, Lorena, Mary, Sonia y Marta, porque con vosotras no importa el tiempo que pase sin vernos, ni que cada una esté en una punta del país. Cada vez que hablamos es como cuando estábamos juntas y siempre sereis mis amigas.

A David, María y Nuria, los compañeros de piso más originales que podría haber encontrado. Gracias por enseñarme un poco de arte, a ver el trabajo y la vida de otra manera, por intentar entender qué es esto de la ciencia y no hacerme sentir un bicho raro entre tantos artistas. Vivir con vosotros ha sido una aventura desde el principio y tengo un millón de buenos momentos en la memoria.

A Alberto porque sólo tube que llamarle una vez para que me diera el apoyo de la familia, y la complicidad, confianza y consejos del mejor amigo para siempre. De ti he aprendido que hay que luchar por conseguir los sueños y no ponerse límites. Espero que seas feliz como tú te lo mereces y darte un abrazo muy pronto.

A María, Antonio, Fátima y Ana por ser lo más próximo a la familia y porque en su casa me siento como en la mía. A Ana por algunas noches de aventuras y a Fátima porque no recuerdo ninguna etapa de mi vida sin que estuviera cerca.

A Santiago y a Carmen, por hacer que me sienta una más de la familia.

A mi abuela Antonia, por mirarme como sólo ella me mira, decir a voces que soy su nieta preferida y porque se le llena la boca cada vez que dice “mi Celia”. Por acompañarme en los mejores y en los peores momentos. A mi abuela Edu, por toda la ilusión que tiene en mí. A mi tía Micaela, por no cansarse nunca de enseñarme sus creencias. Yo sé que ella cree que es lo mejor que puede hacer por mí.

A todos mis tios y mis primos, porque llegar hasta aquí no tendría sentido si no pudiera compartirlo con vosotros. A mi tío Manolo por ser la persona más divertida y desinteresada que conozco. A mi tío Toni porque no hace falta que hablemos para decírnoslo todo, es mi mejor amigo. A mi tía Rubia porque defiende a los suyos por encima de todo y disfruta y anima como una niña cada reunión familiar. A mi tía Celia por

ser la mejor madrina, su confianza y pensar en mí cuando cocina cada navidad. A mi tía Mari por ser tan buena persona. A mi tía Maribel por su naturalidad, espontaneidad y por muchos detalles. A mi tío Riquero por sus charlas llenas de consejos y reconocimiento. A mi tío Miguel por sus comentarios y recomendaciones sobre libros y películas. A Alicia por algunas visitas que me hicieron mucha ilusión, por entenderme y apoyarme en muchas decisiones. A Manuel porque te dice guapa cuando menos te lo esperas, por hacernos reír con sus despistes y sobre todo porque es muy sincero y auténtico. A Jose por saber disfrutar de las cosas de siempre y porque derrocha cariño por donde va. A Carlos por sorprenderme siempre para bien y por ser tan familiar. A Miguel Ángel por sus abrazos, por contar conmigo y por saber integrarse donde va sin dejar de ser él mismo. A Jesús por ser el primero en hacerme sentir responsable, por todos los domingos que pasábamos juntos y porque espero que pasemos muchos más. Y a mis niñas, Mamen, Isabel y Belén, porque son la alegría de la casa, porque cuando llego me están esperando y cuando me voy van a despedirme, porque siempre salen corriendo para darme un beso y por dejarme jugar con ellas.

A mi madre porque es la persona que mejor me conoce, porque es la que más sufre mi mal humor y nunca se queja y por apoyarme siempre. A mi padre por confiar en mí y en mis decisiones, muchas veces sin compartirlas; por ser capaz de callarse, en vez de decir lo que piensa porque sabe que eso cambiaría mis planes y por no acostumbrarse a tenerme lejos.

A mi hermana, por todo, por ser exactamente como es, por intentar entenderme y no enfadarse nunca, por cuidar de mis padres cuando yo no estoy y por enseñarme a ser feliz.

A José Julián, por su generosidad y por estar con nosotros.

A Santi, por mil cosas: por ponerle color a esta tesis, por aguantar los nervios y la presión del final, por sus consejos y sus críticas siempre constructivas, por tener siempre una respuesta y sobre todo por su apoyo y soportarme cada día.

Muchas gracias a todos!!!!

## ***Prefacio***

La secuenciación del genoma humano ha supuesto uno de los mayores avances de la biología durante los últimos años. Hoy es más fácil saber en que nos parecemos unos a otros y porqué somos tan diferentes, pero hasta a quienes trabajamos diariamente con ello, nos resulta increíble cómo una simple secuencia de bases puede llegar a acumular tanta información.

Durante estos años mucha gente se sorprendía al saber que iba a dedicar como mínimo cuatro años a estudiar “una región del cromosoma 15” y hoy la más sorprendida soy yo, porque el proyecto inicial era sobre una región de 4 Mb, nos hemos centrado en 1 Mb y sólo conocemos una pequeña parte de ella. Cuando esta historia comenzó no pensábamos en CNVs, ni en ncRNAs ni miRNAs y hoy parecen ser casi imprescindibles para entender el funcionamiento del cuerpo humano. Me parece muy positivo habérmelos encontrado por el camino porque sobre todo significa que la ciencia avanza y sigue un buen camino. Por otra parte, me parece emocionante no poder hacerme una idea de lo que se sabrá de esta pequeña región dentro de otros cuatro años.

Me quedo con que el ser humano es muy complejo, de que hay muchos factores ambientales y genéticos que hacen esto posible pero que todos juntos forman la naturaleza humana y hacen que seamos tan parecidos y tan diferentes a la vez.

Barcelona, 2007







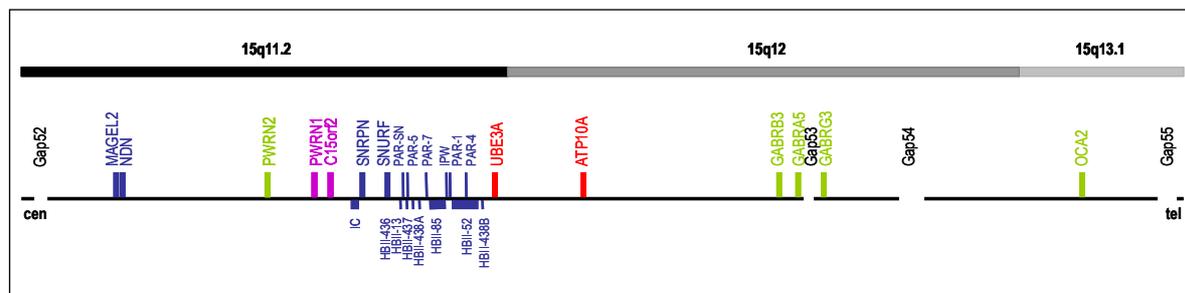
# ***Introducción***



## 1. Organización génica de la región cromosómica 15q11-q13

La región cromosómica 15q11-q13 es una de las regiones más complejas del genoma humano, en cuanto a inestabilidad genómica, *imprinting* y expresión génica; siendo así de gran interés para entender enfermedades genómicas.

La correcta expresión de los genes codificados en 15q11-q13 es fundamental para el desarrollo y funcionamiento del Sistema Nervioso Central (SNC). La transcripción de una parte de ellos en el cerebro es uniparental y está regulada por *imprinting*. Los genes *MAGEL2*, *NDN* y *SNURF-SNRPN* se transcriben a partir del alelo paterno [1-3]; *UBE3A* y *ATP10A* en el alelo materno [4-7] y tres subunidades del receptor de GABA (*GABRA5*, *GABRB3* y *GABRG3*) y *OCA2* son de expresión bialélica [8] (Figura 1.1).



**Figura 1.1 Organización génica de la región 15q11-q13 del genoma humano.**

Los genes de expresión paterna, materna y bialélica están representados en azul, rojo y verde respectivamente. En morado están marcados los genes que presentan expresión monoalélica en cerebro fetal pero no se conoce si son de origen materno o paterno.

Los genes *MAGEL2* y *NDN* se expresan principalmente en cerebro aunque también se ha detectado *MAGEL2* en otros tejidos fetales y *NDN* en la placenta. Los dos codifican proteínas que participan en la diferenciación neuronal [2, 9].

Más cerca del telómero que *NDN*, se han identificado los genes *PWRN1*, *PWRN2* y *C15orf2*. Los tres se expresan de manera bialélica principalmente en testículo, pero *PWRN1* y *C15orf2* se han observado también en cerebro fetal donde su expresión es monoalélica, aunque se desconoce si es materna o paterna [10, 11].

El locus *SNURF-SNRPN* del cromosoma 15 es uno de los más complejos del genoma humano. Es un gen bicistrónico que codifica dos proteínas [12]: *SNRPN upstream reading frame* (*SNURF*), de función desconocida y *small nuclear ribonucleoprotein complex* (*SNRPN*), que pertenece a la familia snRNP SMB/SMN, se expresa fundamentalmente en cerebro y participa en el procesamiento de pre-mRNAs específicos de dicho tejido [13]. En el extremo 5'UTR de este gen se encuentra el *Imprinting Center* (IC), necesario para mantener la expresión monoalélica del locus [14].

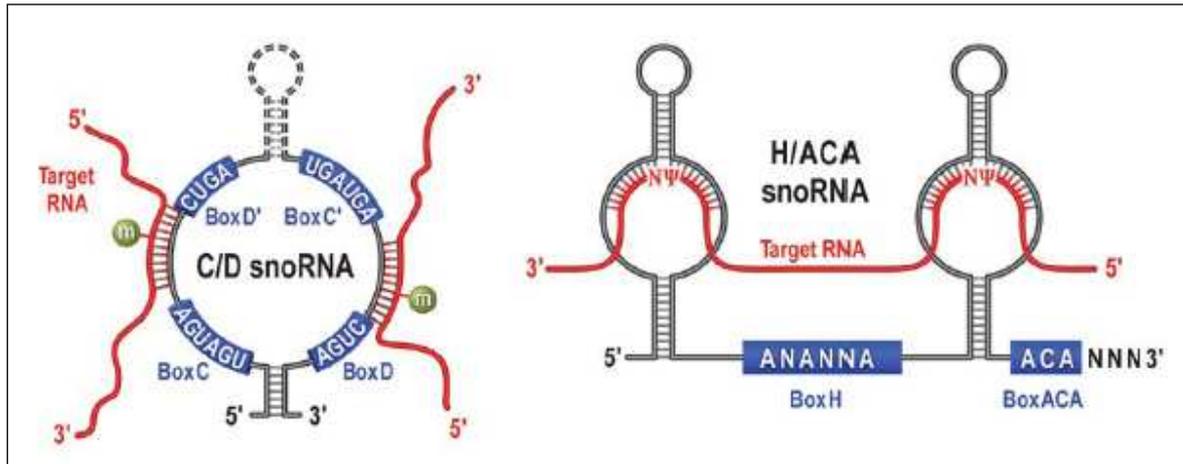
Entre *SNRPN* y *UBE3A* se han identificado una serie de transcritos (*PAR-SN*, *PAR-5*, *PAR-7*, *IPW*, *PAR-1* y *PAR-5*) que no codifican ninguna proteína y a los que se atribuyen funciones reguladoras aún no bien definidas [15, 16]. Sin embargo, en los intrones de estos genes está codificado un *cluster* de *small nucleolar RNAs* (snoRNAs): HBII-436, HBII-13, HBII-437, HBII-438A, HBII-85 y HBII-52, que modifican a otros RNAs y podrían ser la parte funcional de estos transcritos [17].

El gen *UBE3A* codifica una *E3 ubiquitin-protein ligase* [18-20] y *ATP10A* a una translocasa encargada del transporte de fosfolípidos, que se ha especulado podría participar en la señalización del SNC [6, 7]. Ambos genes se expresan en otros tejidos además del cerebro, pero sólo en este su expresión está regulada por *imprinting* [4-7].

Fuera del dominio improntado se encuentran tres subunidades del receptor de GABA (*GABRA5*, *GABRB3* y *GABRG3*), el mayor inhibidor de la neurotransmisión en el cerebro humano [21]; y el gen *OCA2*, también llamado gen P, que codifica una proteína integral de membrana de los melanocitos, actúa como transportador de tirosinas y es importante para la pigmentación [22, 23].

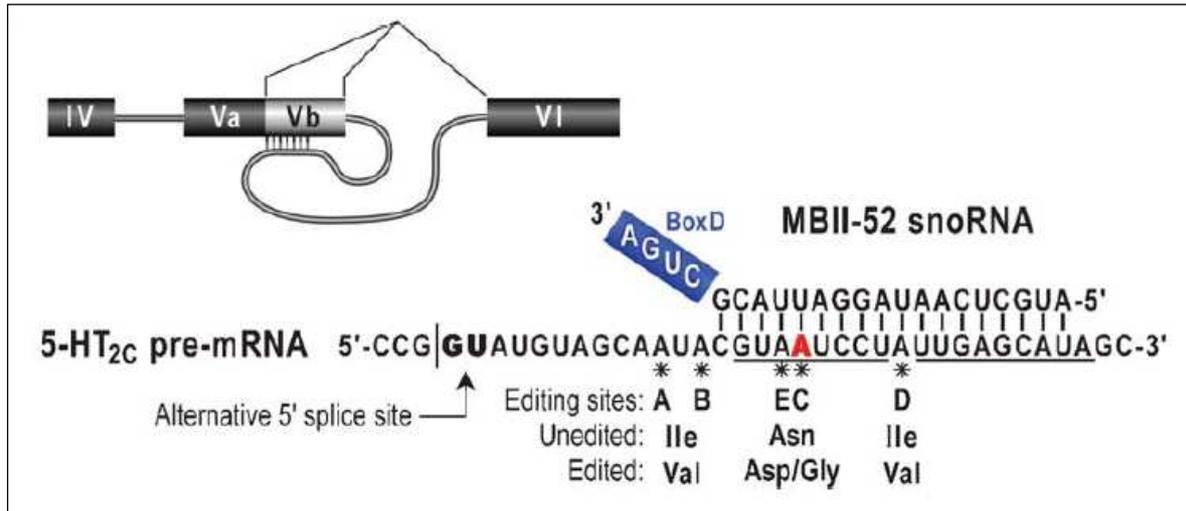
## 1.1 snoRNAs

Los snoRNAs como su nombre indica se localizan en el nucleolo y allí modifican a otros RNAs. Inicialmente se pensó que sólo actuaban sobre RNA ribosómicos (rRNA) pero hoy es evidente que pueden interactuar con otros RNAs: *small nuclear RNAs* (snRNAs) y RNA mensajeros (mRNAs) [24, 25]. Se distinguen dos tipos de snoRNAs caracterizados por la presencia de secuencias específicas, denominadas *boxes* (C/D y H/ACA) y por estar asociados a proteínas concretas. Los C/D snoRNAs se asocian a fibrilarina y modifican al RNA con el que interactúan, metilando ribosas en sitios específicos. Los H/ACA se asocian a la proteína GAR1 y convierten a una uridina en pseudouridina. Las moléculas de RNA a las que modifican presentan una región complementaria al snoRNAs. En el caso de los C/D snoRNAs es de 10 a 21 nt y comprende la región 5' de los dominios D o D' y en los H/ACA es una región más pequeña (de 3 a 10 nt) localizada en los *loops* de los dos *hairpins* formados por el snoRNAs (Figura 1.2). En los sitios de complementariedad es donde se producen las modificaciones correspondientes [26, 27].



**Figura 1.2** snoRNAs con dominios C/D y H/ACA

Los snoRNAs codificados en 15q11.2 (HBII-436, HBII-13, HBII-437, HBII-438A, HBII-85, HBII-52 y HBII-438B) tienen el dominio C/D y son de especial interés debido a: a) se expresan principalmente en cerebro; b) existen 3 copias en tandem de HBII-13, 24 de HBII-85, y 47 copias igualmente organizadas de HBII-52 en la región 15q11.2, mientras que sus homólogos en ratón aparecen en 1, 15 y 17 copias respectivamente; c) los tres snoRNAs específicos de cerebro no tienen ninguna región complementaria a ningún snRNA, rRNA u otros RNAs no codificantes conocidos, pero sí presentan al menos 14-15 nt complementarios a algún mRNA. Concretamente 18 nt de HBII-52 son complementarios al mRNA de receptor de serotonina (5-HT<sub>2C</sub>) [28] (Figura 1.3). La región de interacción de este mRNA con HBII-52 es de particular interés ya que posee dos sitios susceptibles de modificaciones: *splicing* alternativo y desaminación de una adenina, transformando esta base en inosina [29, 30]. Por *splicing* alternativo se pueden generar dos isoformas de 5-HT<sub>2C</sub>, una inactiva y trunca que sólo contiene al exón Va y otra activa que incluye a Va y Vb. Se ha comprobado MBII-52 (el homólogo a HBII-52 en ratón) participa en el *splicing* del mRNA [31]. La segunda modificación es más frecuente en cerebro que en otros tejidos [32] y es facilitada por la metilación de una ribosa [33]. En el caso de 5-HT<sub>2C</sub> la modificación implica una disminución de la actividad del receptor y sólo queda por demostrar si MBII-52 interviene en dicha modificación. Las peculiaridades de los snoRNAs de la región 15q11.2 sugieren que han podido tener un papel importante en la evolución y función del cerebro humano, que aún queda por descubrir y puede ser relevante para conocer la etiología de enfermedades genómicas.



**Figura 1.3** Región del mRNA 5HT2C susceptible de interactuar con el snoRNA HBII-52

## 2. *Imprinting*

Todas las células del organismo tienen la misma información genética, la misma secuencia de DNA, sin embargo, la expresión diferencial de genes en cada una de ellas hace que sean muy diferentes. Esto es posible gracias a los mecanismos que regulan la expresión génica. Algunos de ellos son específicos de cada momento, como la presencia de activadores o represores, su concentración celular y las interacciones entre ellos; pero otros son lo suficientemente estables para ser transmitidos a la descendencia, afectan a la estructura de la cromatina y se conocen como factores epigenéticos.

El *imprinting* es una parte de la epigenética por el cual los genes de un locus presentan una expresión diferencial, dependiendo de su origen parental [34]. Se estima que en el genoma humano existen unos 200 genes improntados y suelen estar implicados en el desarrollo durante las primeras etapas de la vida (fetal y embrionaria). Así, muchos se expresan en tejidos fetales y en la placenta. El *imprinting* tiene lugar durante la gametogénesis y/o se modifica en fases embrionarias precoces. Puede ser específico de un tejido, tipo celular o momento concreto del desarrollo [35].

### 2.1 Modificaciones de los genes improntados

La expresión diferencial de un gen dependiendo de su origen parental requiere modificaciones que afectan a la cromatina [34, 36]. La modificación más frecuente es la metilación de citosinas que forman parte de dinucleótidos CpG y suelen estar en las regiones promotoras formando las llamadas islas CpG. La metilación puede interferir con la unión de factores de transcripción u otras proteínas al DNA [34, 37, 38]. Por este mecanismo se encuentran silenciados los genes *Insulin-like growth factor 2 (IGF2)* y *SNRPN* en el alelo materno y tiene lugar la inactivación de un cromosoma X en las mujeres [34].

Las demás modificaciones afectan al extremo amino-terminal de las histonas, asociadas al DNA: metilación de argininas o lisinas, acetilación o ubiquitinación de lisinas y fosforilación de serinas. En general, la acetilación de las histonas H3 y H4 están asociadas a la formación de eucromatina y a la activación de la transcripción. La reacción está catalizada por las histonas acetiltransferasas (HTAs). Las histonas deacetilasas (HDACs) promueve la condensación de la cromatina, reprimiendo la transcripción. Del mismo modo, la metilación de la Lys4 de la histona H3 está relacionada con la activación de la transcripción, mientras que la metilación de la Lys9 de la misma histona, se asocia al silenciamiento génico.

La metilación del DNA puede modular la deacetilación de las histonas a través del reclutamiento de la proteína *methyl-CpG-binding protein 2* (MeCP2). Esta proteína nuclear codificada en el cromosoma X forma parte del complejo HDAC [34].

## 2.2 *Imprinting* y enfermedades genómicas

En principio se pensaba que las mutaciones eran la principal causa de las enfermedades pero el descubrimiento del papel de la epigenética en la regulación génica ha sido relevante para entender la etiología de muchas de ellas.

Dos tipos de variaciones relacionadas con el epigenoma se han asociado a enfermedades: mutaciones en genes que participan en el establecimiento y mantenimiento del *imprinting* (alteraciones *trans*) o cambios en la estructura de la cromatina [34] (alteraciones *cis*). Entre las primeras destacan por ejemplo, las mutaciones del gen *DNMT3B*, de *MECP2*, de *MTHFR* y de *ATRX*. *DNMT3B* codifica una *de novo* DNA metiltransferasa y mutaciones de este gen producen el síndrome de inmunodeficiencia, inestabilidad de la región centromérica y anomalías faciales (ICF) [39]. Mutaciones de *MECP2* se han detectado en pacientes con síndrome de Rett (RTT), un trastorno del desarrollo caracterizado por retraso mental y ataxias [40]. Deficiencias de *MTHFR* alteran la metilación de las histonas y producen retraso mental [34]. Mutaciones de *ATRX* afectan a la regulación del locus de las  $\alpha$ -globinas alterando la metilación de secuencias muy repetidas en el genoma, como repeticiones subteloméricas, y satélites específicos del cromosoma Y. Los hombres con mutaciones en el gen *ATRX* desarrollan  $\alpha$ -talasemias, presentan deficiencias mentales y muchos no pueden caminar, mientras que mujeres heterocigotas para la mutación son normales [34, 39, 41].

Entre las alteraciones de la cromatina destacan el síndrome del cromosoma X frágil, debido a mutaciones en el gen *FMR1*. La expansión de los trinucleótidos (CCG)<sub>n</sub> en el extremo 5' del gen están asociadas a hipermetilación del DNA [42]. *FMR1* codifica a la proteína FMRP, una proteína de unión a RNA que participa en la regulación de la expresión génica formando parte del complejo de actuación de los miRNAs [43]. También se han observado deleciones de elementos que regulan la transcripción en dominios improntados. Dos ejemplo son las deleciones del *Imprinting Center* (IC) de la región 15q11-q13 que causan los síndromes de Angelman (AS) y Prader-Willi (PWS) [44] y las deleciones del *Locus Control Region* (LCR), situado 10 Kb *upstream* de las  $\beta$ -globinas, relacionadas con el desarrollo de talasemias [34]. Cambios que afectan a la cromatina son también responsables de la distrofia muscular facioescapulohumeral (FSHD). En este caso se debe a un reordenamiento de la región 4q35 que en individuos sanos incluye

entre 10-150 repeticiones de los nucleótidos GCs. Estas repeticiones suponen ~3,3 Kb y se las conoce como *D4Z4*. En pacientes con FSHD la región *D4Z4* tiene menos de 10 repeticiones. Parece que la reducción de GCs hace que la región adopte una conformación menos represiva y así la transcripción resulta activada [34, 39].

Otro mecanismo que afecta a la expresión de los genes improntados y se ha asociado a enfermedades es la disomía uniparental, que ocurre cuando un par de cromosomas homólogos son heredado del mismo progenitor. Este fenómeno se ha observado en pacientes con AS, PWS, síndrome de Russell-Silver (SRS) y diabetes neonatal transitoria.

Estos son sólo algunos ejemplos de enfermedades en las que se han observado alteraciones epigenéticas, pero conocida la relevancia del *imprinting* en el desarrollo embrionario, debería ser considerado en el estudio de otras enfermedades, sobre todo aquellas relacionadas con el crecimiento, la conducta y el desarrollo de células anormales.

### **2.3 Factores ambientales que afectan al *imprinting***

Las modificaciones epigenéticas pueden explicar cómo los factores ambientales afectan a la expresión génica sin modificar el código genético. Entre estos factores han despertado especial interés los efectos de la fecundación *in vitro* y el ácido fólico.

Las técnicas de fecundación *in vitro* se desarrollan hoy con éxito para solucionar los problemas de fertilidad, pero existen evidencias de que suponen un riesgo mayor que un embarazo natural para el desarrollo normal del feto. Entre sus causas destacan la manipulación de los gametos y del embrión antes de la implantación, que pueden modificar la impronta genética. Normalmente personas que se someten a este tratamiento tienen problemas de fertilidad, sus gametos pueden tener alteraciones genéticas, y entre ellas algunas que afecten al *imprinting*, transmisibles a la descendencia. También se estimula hormonalmente a la mujer y aunque no se sabe cómo las gonadotropinas afectan al *imprinting*, parece que podrían estimular la liberación de óvulos inmaduros en los que no se han completado las modificaciones epigenéticas [35, 45]. Diferentes estudios se refieren casos de Síndrome de Angelman (AS) y de Beckwith-Wiedemann (BWS) concebidos mediante fecundación *in vitro* [46-49]. Estos síndromes están relacionados con alteraciones epigenéticas de las regiones 15q11-q12 y 11p15 respectivamente [35].

El ácido fólico es de particular interés en epigenética. Contribuye a la biosíntesis de purinas y pirimidinas y es esencial para la producción de la cantidad adecuada de SAM, el principal donante de grupos metilos en todas las reacciones de metilación. Existen evidencias de que la ingesta de ácido fólico afecta a la expresión de genes improntados en humanos y de que su efecto puede transmitirse a la descendencia. Por

otro lado, se sabe que suplementos de ácido fólico en la dieta de las mujeres embarazadas, reduce el riesgo de que el feto desarrolle alteraciones del tubo neural, posiblemente a través de la metilación del DNA y/o de las histonas, relacionando la epigenética con el desarrollo de malformaciones en el cerebro y en la médula espinal [34].

## 2.4 Dominio improntado de la región cromosómica 15q11-q13

Los genes de expresión paterna de este locus (*MAGEL2*, *NDN*, *SNURF-SNRPN*) están improntados en todos los tejidos en los que se expresan [15, 50]. Sin embargo, los de expresión materna (*UBE3A* y *ATP10A*), sólo lo están en el cerebro, e incluso *UBE3A* sólo en ciertas regiones del cerebro [4-6]. En ratón se ha observado que *UBE3A* está improntado en neuronas y no en células gliales [51].

El establecimiento y mantenimiento del *imprinting* en la región cromosómica 15q11-q13 depende del *Imprinting Center* (IC). Este lo forman dos fragmentos: el PWS-IC y AS-IC. Es PWS-IC se encarga de mantener el estado paterno, mide ~4,3 Kb, y se solapa con el promotor del gen *SNRPN*. El AS-IC mantiene el *imprinting* materno, mide ~0,9 Kb y está situado ~35 Kb más centromérico que PWS-IC. En PWS-IC está completamente metilado en el alelo materno y en él se han observado modificaciones en las histonas propias de regiones silenciadas. Sin embargo, en el alelo paterno no está metilado y las modificaciones de las histonas son propias de regiones transcripcionalmente activas (hiperacetilación de las histonas H3 y H4). Estas características facilitan que el PWS-IC se transcriba sólo a partir del alelo paterno. También se ha encontrado metilación diferencial en los promotores de los genes *MAGEL2*, *NDN* y *SNRPN*. Lo que no se conoce aún es cómo el PWS-IC regula la transcripción de los demás genes de la región. Runte et al. [17] han identificado un transcrito de >500 Kb entre PWS-IC y *UBE3A*, de expresión exclusivamente paterna y en dirección opuesta al gen *UBE3A*, que podría reprimir la expresión de *UBE3A* en dicho cromosoma [50]. No se han observado diferencias en la metilación del AS-IC entre ambos alelos pero si en la conformación de la cromatina, que facilitaría la transcripción de AS-IC sólo en el alelo materno [36]. El AS-IC podría codificar un sitio de unión a un factor regulador de la transcripción de este locus. Este factor, hoy desconocido, se uniría al AS-IC, guiando la metilación del PWS-IC en el alelo materno e impidiendo su transcripción y la de los demás genes de expresión paterna, incluido el *UBE3A antisense*. Al no transcribirse AS-IC en el alelo paterno, este factor regulador no puede unirse y el PWS-IC se mantiene no metilado y activo (Figura 1.4) [52].

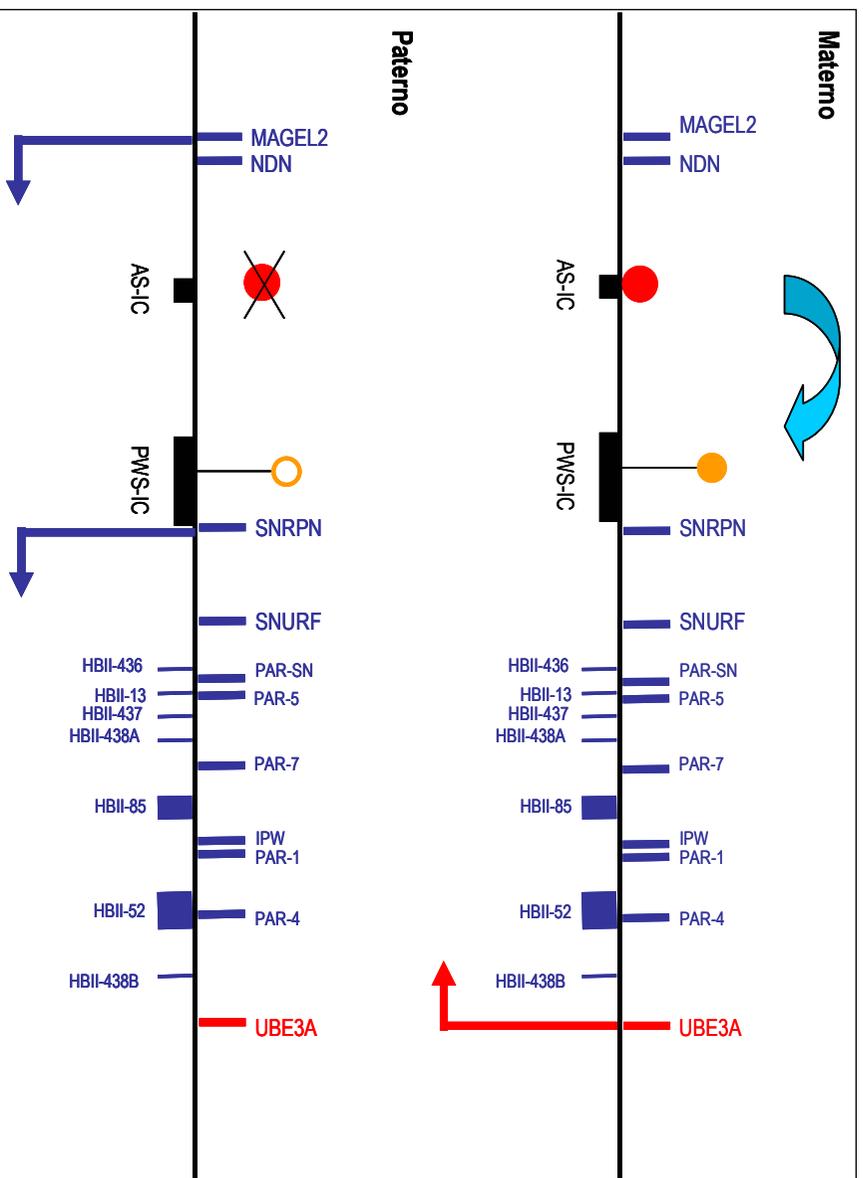


Figura 1.4 Regulación de la expresión génica por imprinting en la región 15q11-q13

### 3. Variabilidad genómica

La variabilidad genómica se manifiesta de diversas formas, incluyendo *Single Nucleotide Polymorphisms* (SNPs), microsatélites, minisatélites, *Variable Number of Tandem Repeats* (VNTR), inserción de secuencias tipo *Alu* y reordenamientos genómicos (duplicaciones, deleciones, inversiones y translocaciones) [53]. Hasta hace poco se pensaba que los SNPs eran el principal recurso de variabilidad genética, pero los avances en las técnicas de laboratorio (*Genome-scanning array*) y en las herramientas bioinformáticas para estudiar el genoma, han descubierto un tipo de variación que afecta a fragmentos de DNA pequeños para poder ser observados al microscopio, pero de mayor escala a los detectados por las técnicas convencionales de secuenciación. Se conocen como variaciones submicroscópicas e incluyen inversiones y *Copy Number Variants* (CNVs) desde ~1 Kb hasta 3 Mb [54-56].

Estas variaciones pueden afectar a millones de bases de DNA, incluyendo genes enteros y sus secuencias reguladoras [55, 57, 58]. Aunque este tipo de variación en algunas regiones no suponen un cambio a nivel de fenotipo [57, 59], otras conllevan alteraciones en el número de copias de un gen, pudiendo ser la causa de algunas enfermedades por si solas, o combinándose con otros factores genéticos o ambientales.

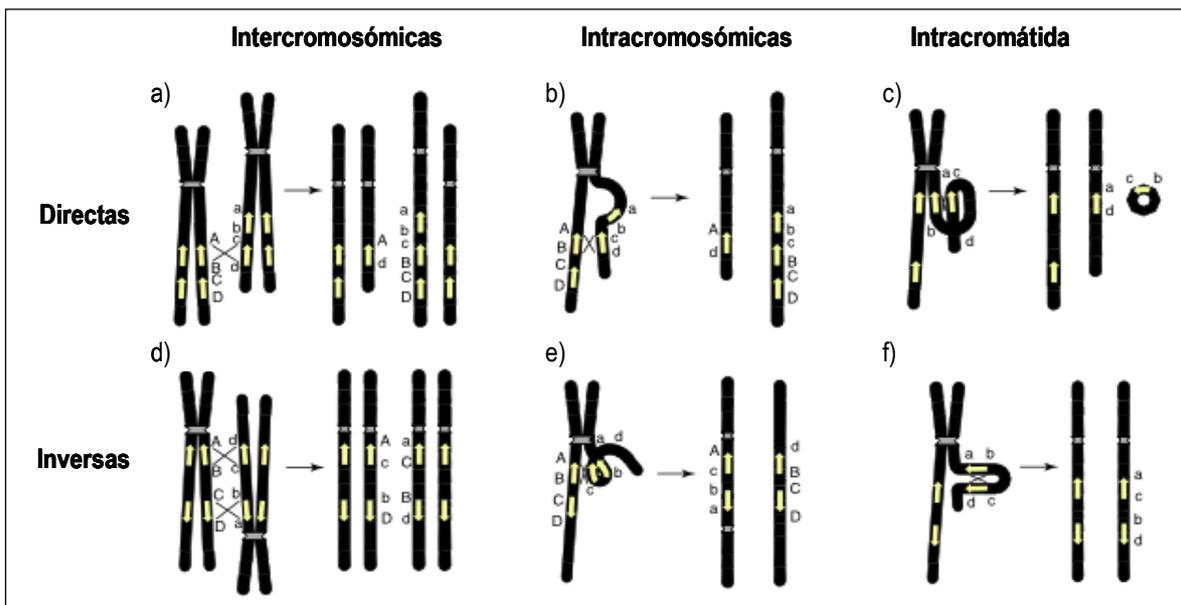
#### 3.1 Duplicaciones Segmentarias y reordenamientos genómicos

Las duplicaciones segmentarias o duplicones se han definido como fragmentos de DNA de >1 kb, representadas más de una vez en el genoma y con una identidad entre ellas superior al 90% [60]. No están distribuidas de manera uniforme en el genoma, por ejemplo, el cromosoma Y presenta un enriquecimiento significativo de duplicones, y en cada cromosoma, suelen concentrarse en las regiones centroméricas y subteloméricas [61].

Las duplicaciones segmentarias se dividen en dos clases: intercromosómicas e intracromosómicas. Las intercromosómicas están duplicadas en cromosomas no homólogos y normalmente se disponen en las regiones centroméricas o subteloméricas [60].

Un *cluster* de duplicaciones segmentarias se forma a partir de varios segmentos de material genómico que mediante transposición se integran en una región determinada. Duplicaciones totales o parciales de cada uno de estos fragmentos explican la estructura y complejidad de algunas regiones del genoma [61].

Los duplicones facilitan la recombinación homóloga no alélica (NAHR), durante la meiosis, dando lugar a reordenamientos genómicos: duplicaciones, deleciones, inversiones, cromosomas marcadores y translocaciones [62, 63] (Figura 1.5). Muchos de ellos están asociados a enfermedades genómicas. Entre ellas que destacan la enfermedad de Charcot–Marie–Tooth tipo 1A (CMT1A) (17p11.2), la neurofibromatosis tipo 1 (NF1) (17q11.2), los síndromes de Angelman (AS) y Prader-Willi (PWS) (15q11–q13), de Williams–Beuren (WBS) (7q11.23) y de DiGeorge (22q11) [60].



**Figura 1.5 Mecanismos de recombinación homóloga no alélica guiados por duplicaciones segmentarias**

Se estima que las duplicaciones segmentarias suponen ~5% de nuestro genoma y han aparecido durante los últimos 35 millones de años de evolución [61]. Comparaciones del genoma humano con el de otras especies demuestran que han sido relevantes para la evolución genómica. Un estudio de los genomas de humano y de ratón revela que los duplicones coinciden con regiones de pérdida de sintenia entre las dos especies y que fragmentos de DNA flanqueados por duplicaciones segmentarias y asociados a enfermedades genómicas, aparecen en diferente orden u orientación en las dos especies [64]. Por otro lado, se ha observado que regiones y genes duplicados en el genoma humano, aparecen representados sólo una vez en el de chimpancé, y viceversa, lo que indica que en cada especie se han generado duplicaciones de novo que suponen diferencias en la expresión génica. Entre las duplicaciones de novo del genoma humano se incluyen las asociadas a enfermedades genómicas [65].

### 3.2 Copy Number Variants

Además de estos reordenamientos de mayor tamaño, recientemente se han identificado variaciones en el número de copia (CNVs) e inversiones que afectan a fragmentos de DNA de tamaño submicroscópicos. Los CNVs se definen como segmentos de DNA de 1 Kb o más, que aparecen en el genoma en diferente número de copias en comparación con un genoma de referencia [55].

Las técnicas utilizadas para su detección a gran escala son principalmente *Comparative Genomic Hybridization* (CGH) *arrays* y *SNP arrays*. Este último además tiene la ventaja de añadir información sobre el genotipo, se pueden detectar pérdidas de la heterocigosidad que soportan la presencia de deleciones o de disomías uniparentales. Para el análisis de CNVs que afectan a regiones concretas se pueden utilizar PCR cuantitativas o métodos como *Quantitative Multiplex PCR of Short Fluorescent Fragments* (QMPSF), *Multiplex Amplifiable Probe Hybridization* (MAPH) y *Multiplex ligation-dependent Probe Amplification* (MLPA) en los que se pueden analizar CNVs de hasta 40 regiones diferentes en cada ensayo. Otra posibilidad es analizar SNPs variables entre dos copias de un duplicón, llamados *Paralogous Sequence Variants*. Para ello se existen técnicas de genotipación que permitan distinguir y comparar la intensidad que corresponde a cada uno de los alelos [53].

Recientemente Redon et al. [55] han identificado 1.447 CNVs cubriendo más de 360 Mb (12%) del genoma humano. Utilizaron 270 muestras pertenecientes a cuatro poblaciones diferentes del proyecto HapMap. Unos meses más tarde Wong et al. [66] han definido 3.654 CNVs en 95 controles, de 16 etnias y estimaron una media de 155 variaciones por cada muestra. Estos resultados ponen en evidencia la importancia de los CNVs en la diversidad humana.

Es frecuente que CNVs coincidan con duplicones posiblemente porque los reordenamientos se deban a recombinación homóloga no alélica [54, 55]. Además Redon et al. observaron un enriquecimiento de duplicaciones segmentarias en los intervalos que podrían contener los puntos de rotura o *breakpoints* (BP) de los CNVs.

Más del 58% de los CNVs identificados en el genoma humano se solapan con algún gen y el 99% con regiones no codificantes conservadas del genoma, señalando la contribución de los CNVs en la diversidad de fenotipos y en la predisposición a enfermedades. Entre los genes que coinciden con CNVs predominan los que participan en la adhesión celular y los relacionados con la percepción sensorial y los procesos neurofisiológicos. También 21 de los miRNAs conocidos hasta ahora se solapan con CNVs, implicando a los miRNAs en la variabilidad humana. Por ejemplo, las regiones

asociadas a los síndromes de DiGeorge, Smith-Magenis, WBS, PWS, AS y la región 3p21.3, delecionada en pacientes con cáncer de pulmón, resultaron ser muy polimórficas. Se solapa con un CNV el gen *CCL3L1*, un ligando para el co-receptor del HIV. Menor número de copias de este gen se ha asociado a la susceptibilidad para desarrollar el SIDA. También los genes *DISC1* y *COMT* relacionados con esquizofrenia, *LPA* y *BTNL2*, asociados a arterosclerosis y soriasis respectivamente y oncogenes o genes supresores de tumores como *VAV2*, *TNFRSF25*, *CDKN1C*, coinciden con regiones polimórficas del genoma [55, 66].

Otro aspecto interesante de los CNVs es su posible relación con las diferencias entre poblaciones y la capacidad de adaptación al medio. Por ejemplo se han observado diferencias entre poblaciones en el número de copias de los CNVs que incluyen a los genes *CCL3L1* y *UGT2B17* [55, 60]. Es de esperar que los genes implicados en estos aspectos sean más variables y puedan evolucionar más deprisa.

Las inversiones suponen otra clase de variación estructural pero aunque su frecuencia en el genoma humano no es tan conocida, se estima que son menos frecuentes que los CNVs. Algunas están bien caracterizadas y parecen estar asociadas a la predisposición de ciertas enfermedades. Inversiones de las regiones delecionadas en pacientes con WBS y AS no son la causa de la enfermedad pero si parecen suponer un riesgo para el desarrollo de la misma en la descendencia [67].

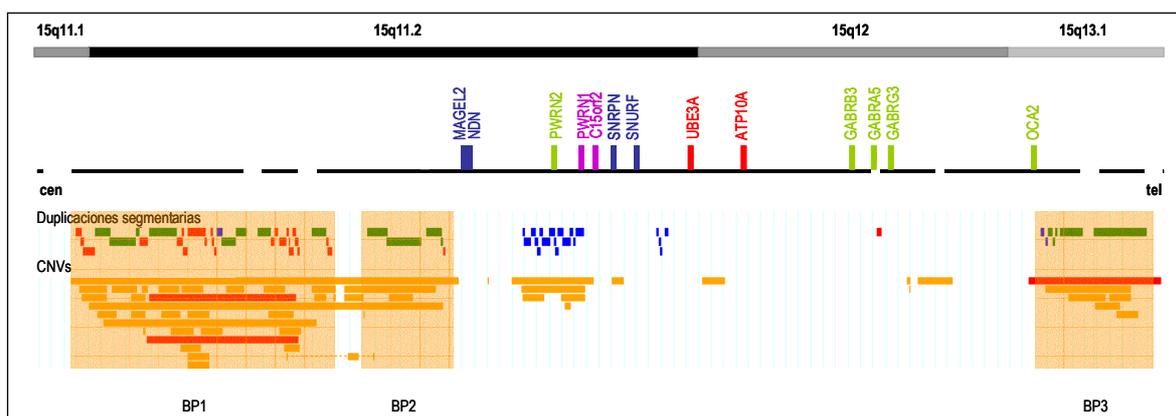
### 3.3 Duplicaciones segmentarias de la región cromosómica 15q11-q13

El cromosoma 15 es uno de los siete cromosomas humanos más enriquecido en duplicaciones segmentarias y en él se distinguen dos regiones especialmente duplicadas, situadas en los extremos del brazo largo. Los dos clusters parecen haber evolucionado independientemente a partir de un ancestro común. La más centromérica coincide con 15q11-q13 [68].

Los distintos *cluster* de duplicaciones segmentarias de 15q11-q13 lo forman principalmente dos tipos de duplicaciones: *15 low-copy repeat* (LCR15) y duplicaciones que contienen al gen *Hect Domain and RCC1-LIKE domain2* (*HERC2*). LCR15 incluye varios genes no procesados o pseudogenes: *Dynamamin 1* (*DNM1*), *Cholinergic Receptor*, *Neuronal Nicotinic*, *Alpha Polypeptide 7* (*CHRNA7*), *SH3 domain containing protein* (*SH3P18*) y *Golgin-like protein* (*GLP*). Los dos tipos están parcialmente relacionados por compartir secuencias del gen *GLP*. Once copias diferentes de la duplicación *HERC2* se han identificado en los cromosomas 15 y 16 pero la mayoría de ellas están en la región 15q11-q13, repartidas entre BP2 y BP3, mientras que no se ha observado en BP1. Existen copias de LCR15 en los cromosomas 2, 5, 7, 9, 10, 12, 16, 19 e Y, pero igual que

HERC2 se concentran en el cromosoma 15, entre las regiones 15q11-q14, 15q24-q26 y 15q21-q22. LCR15 son de origen relativamente reciente (20 millones de años) y la primera copia en el genoma humano parece ser una de la región 9q33-q34 [69, 70].

Un análisis más detallado de los duplicones de la región 15q11 muestra que esta la componen hasta 24 duplicaciones segmentarias diferentes, incluidas LCR15 y HERC2, a pesar de que el 8,5% de su secuencia son *gaps*. La mayoría de ellas con representaciones en dos o tres regiones más del genoma, normalmente pericentroméricas. Entre las duplicaciones de 15q11 las más próximas al centrómero parecen ser las de aparición más reciente (8-15 millones de años), mientras que las más teloméricas podrían haber surgido hace más de 15 millones de años [71]. Esto soporta los datos anteriores que suponían el origen de LCR15, localizadas en la región en BP2 y no en BP1, hace más de 20 millones de años (Figura 1.6).



**Figura 1.6 Duplicaciones segmentarias y CNVs de la región 15q11-q13 (chr15:18.000.000-27.000.000)**

La figura muestra la compleja arquitectura de las duplicaciones segmentarias de 15q11-q13, los BPs asociados a los reordenamientos identificados de la región y solapados con bloques de duplicaciones y los diferentes CNVs. Los *gaps* de secuencia están dibujados como interrupciones en la línea que representa al cromosoma.

### 3.4 Variabilidad genómica de 15q11-q13

La región 15q11-q13 del genoma humano es susceptible de frecuentes alteraciones cromosómicas, microscópicas y submicroscópicas, posiblemente debidas a la compleja arquitectura de las duplicaciones segmentarias que la flanquean y facilitan la recombinación homóloga no alélica.

Entre los reordenamientos de 15q11-q13 se han identificado deleciones causantes de los AS y PWS, duplicaciones en algunos casos de autismo [72], cromosomas marcadores inv dup(15) [73], duplicaciones intersticiales [74], triplicaciones [75], inversiones y translocaciones [76]. Algunos reordenamientos son relativamente frecuentes, por ejemplo, se producen deleciones en 1/10.000 nacimientos e invdup15 en 2/10000 [70]. Inversiones

de la región se han observado en madres sanas de pacientes con AS, que podrían suponer un factor de riesgo para que se de la delección causante del síndrome en la descendencia [77].

También CNVs (duplicaciones y delecciones) se han identificado en individuos sanos contribuyendo a la complejidad estructural de 15q11-q13 [54, 55, 57, 59, 66, 78-81]. Los más frecuentes coinciden con la posición de las duplicaciones segmentarias que flanquean la región y con los BPs definidos en la misma (Figura 1.6). Además de estos, se han caracterizado CNVs, coincidiendo con otro bloque de duplicones, situado entre los genes *NDN* y *SNURF-SNRPN* y otros que incluyen total o parcialmente a los genes *PWRN1*, *PWRN2*, *C15orf2*, *SNRPN*, *ATP10A* y *GABRG3* (Figura 1.6). Estos son menos frecuentes y podrían suponer un factor de riesgo para el desarrollo de las alteraciones fenotípicas asociadas a la región y/o otras enfermedades. Un trabajo publicado recientemente soporta esta hipótesis definiendo CNVs de esta región en pacientes con autismo [82]. También 4 de los 6 gaps de 15q11-13 coinciden con CNVs. Así para entender cómo afecta 15q11-q13 a la diversidad humana y elucidar su papel en la etiología de alteraciones genómicas sería conveniente identificar las secuencias funcionales de los CNVs que no se solapan con genes y cubrir los *gaps* de la región.

### 3.5 Gaps de secuencia

Actualmente se conoce más del 99% de la porción de eucromatina del genoma humano y ~94% de la secuencia total está ordenada y orientada. Esto ha supuesto uno de los mayores avances de la biología en los últimos años. El 6% restante lo forman 379 *gaps* divididos en dos tipos: 58 heterocromáticos que suponen ~200Mb y 321 eucromáticos (24,4 Mb) [83]. Diferentes evidencias soportan que la mayor parte de estos *gaps* se correspondan con duplicaciones segmentarias: a) las secuencias más largas e idénticas son las más difíciles de secuenciar y ensamblar; b) el 54% de los *gaps* eucromáticos están flanqueados por duplicones y si estos suponen un ~5% del genoma, parece que los *gaps* están enriquecidos en duplicaciones c) el 48% de los *gaps* están flanqueados o se solapan con CNVs, regiones también enriquecidas en duplicaciones [55, 83].

En la región cromosómica 15q11-q13 existen tres *gaps* (Figura1.6) que se solapan con CNVs [55, 59, 78] y dos ellos además afectan a intrones de los genes *GABRA5* y *GABRG3*, por lo que es posible que estos genes estén parcialmente duplicados. La caracterización de los *gaps* de 15q11-q13 podría explicar el papel de estos CNVs en la variabilidad genómica y en el desarrollo de las enfermedades asociadas (AS, PWS y autismo) o no a esta región.

## 4. Autismo

El término autista fue utilizado por primera vez en 1943 por Kanner y Asperger, para describir un trastorno infantil caracterizado por la ausencia de interacción social. Hoy en día se entiende por autismo un trastorno del desarrollo, que da lugar a diferentes grados de alteración del lenguaje, la comunicación, las competencias sociales y la imaginación. Con frecuencia, estos síntomas se acompañan de comportamientos anormales, tales como actividades e intereses de carácter repetitivo y estereotipado, movimientos de balanceo, y obsesiones insólitas hacia ciertos objetos o acontecimientos [84-87].

El autismo se incluye en el grupo de Trastornos Generalizados del Desarrollo (*Pervasive Developmental Disorder*, PDD) junto con los síndromes de Asperger, de Rett, el trastorno desintegrativo de la niñez y el Trastorno Generalizado del Desarrollo No Especificado, conocido por las siglas PDDNOS [87, 88].

Una de las características del trastorno autista es su difícil diagnóstico, debido a que existen gran variedad de fenotipos, niveles de afectación, diferentes síntomas durante el desarrollo, incluso en el mismo individuo, y a la ausencia de un marcador biológico específico. Hoy se diagnostica mediante la entrevista diagnóstico de autismo revisada (ADI-R) y la escala de observación del autismo (ADOS), antes de los tres años [86, 89].

Los estudios epidemiológicos estiman que actualmente el trastorno autista afecta a 15-20/10.000 nacimientos y los PDD a 60/10.000, siendo más frecuente en el sexo masculino que en el femenino, en una proporción de 4 a 1 [90].

Diferentes estudios relacionan factores ambientales con el autismo. Entre ellos se mencionan la exposición a drogas, alcohol o ciertas medicinas durante el periodo intrauterino, infecciones pre- y postnatales con organismos como *Haemophilus influenzae*, virus de la rubéola y la exposición al mercurio. En cualquier caso, estos estudios se refieren a un número pequeño de pacientes y no se han hallado suficientes evidencias para llegar a una conclusión clara sobre el papel de estos agentes [88, 89].

### 4.1. Causas genéticas del autismo

Se han encontrado evidencias convincentes sobre las bases genéticas del autismo: a) la probabilidad de que una familia tenga algún miembro autista, es de un 2-8% superior en familias que ya tienen un individuo afectado; b) los estudios de gemelos para el autismo son más concordantes entre gemelos monozigóticos (60%-91% dependiendo del grado de afectación considerado) que entre los dizigóticos (0%-10%); y c) algunos síntomas del

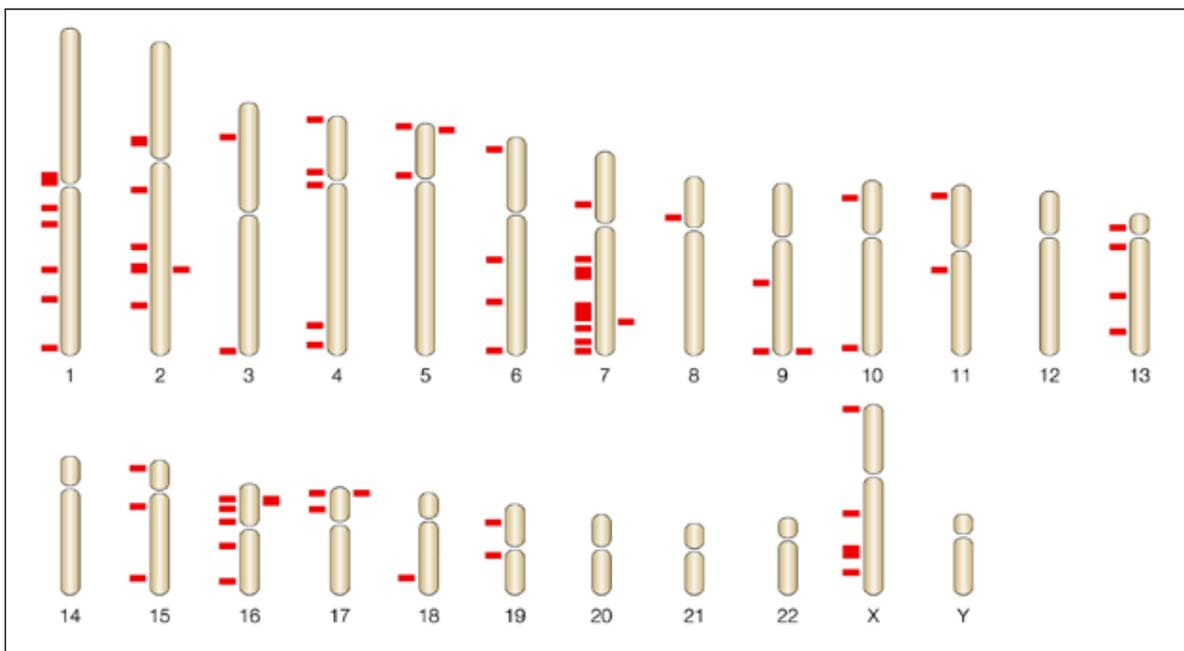
autismo como fobias sociales, preferencias por la rutina y dificultades en la comunicación y adaptación, son más frecuentes en familiares de autistas que en la población general [85-88, 91-94].

Las bases genéticas del autismo son complejas, su transmisión no sigue la herencia simple mendeliana, ni está asociado a un gen, sino que son varios genes y la interacción entre ellos lo que conducen a su desarrollo.

Para identificar las bases genéticas del trastorno autista se han realizado estudios citogenéticos [95-97], de asociación [98-124], de desequilibrio de ligamiento [91, 94, 125-138], de expresión [8, 139, 140] y con modelos animales [141].

#### 4.1.1 Análisis de ligamiento y asociación

Análisis de ligamiento han revelado regiones implicadas en la etiología del autismo en casi todos los cromosomas (Figura 1.7). Sin embargo, la asociación de algunas ha sido soportada por más de un estudio; por incluir genes asociados al trastorno autista; genes que participan en sistemas alterados en el autismo (GABAérgico, glutamatérgico o de la serotonina); y/o genes cuyas alteraciones causan características del fenotipo autista. Por ejemplo varios estudios coinciden al mostrar asociación entre el autismo y las regiones 2q, 3q25-q27, 3p25, 6q14-q21, 7q31-q36, 13q, 16p y 17q11-q21 [86, 88]. Las regiones 3p25-p26, 6p21, 17q11-q12, 15q11-q13 y varias zonas de los cromosomas 7 y X son además interesantes para entender la etiología del autismo por su contenido génico.



**Figura 1.7** Regiones cromosómicas asociadas al autismo

La región 3p25-p26 resulta relevante porque en ella está codificado el receptor de la oxitocina (OT), asociado con el trastorno autista. En pacientes se han detectado niveles de OT más bajos que en controles [142] y estudios en ratas y ratones demuestran que los niveles de OT afectan a la capacidad de relacionarse, explicando cómo este gen participa en el fenotipo autista [143, 144].

Mutaciones del receptor de glutamato 6 (*GluR6*), codificado en 6q21, se han observado en autistas [100]. Sólo algunas podrían afectar a la función de la proteína pero debido a la importancia del glutamato en el desarrollo del cerebro, el aprendizaje y la memoria y a las alteraciones del sistema glutamatérgico observadas en pacientes, este gen resulta relevante en la patogénesis de dicho trastorno [85, 88].

El cromosoma 7 es uno de los que más se repite en los análisis de ligamiento y la mayor parte de los genes asociados al autismo están codificados en este cromosoma [88]. Entre ellos destacan *FOXP2*, *LAMB1*, *EN2*, *HOXA1*, *GRM8*, *NPTX2*, *WNT2*, *RELN* y *CAPS2* [85, 101, 103-105, 145-149]. Mutaciones de *FOXP2* están relacionadas con alteraciones del lenguaje, *LAMB1* promueve la migración neuronal y el crecimiento de las neuritas durante el desarrollo del SNC, *EN2* es un factor de transcripción homeobox que actúa durante la formación del cerebro y el cerebelo, *HOXA1* participa en el desarrollo del cerebro posterior, *GMR8* es otro receptor de glutamato y *NPTX2* es importante para la sinaptogénesis [85, 88]. Experimentos con ratones muestran que mutaciones de *WNT2* afectan a la capacidad de relacionarse [150]. *RELN* participa en la migración neural, en la formación de la corteza cerebral y en la sinaptogénesis. Niveles más bajos de esta proteína se han identificado en pacientes [151] y se ha observado que en familias autistas existe una tendencia a la transmisión de un alelo de *RELN* que contiene una expansión de trinucleótidos en su extremo 5' [102]. Alteraciones de este gen afectan al desarrollo del cerebelo y la corteza, unas de las anomalías encontradas en autistas [85]. Ratones *knockout* para *CAPS2* muestran que la adecuada liberación de neurotrofinas (NT-3 y BDNF) mediada por *CAPS2* es indispensable para el desarrollo y funcionamiento normal del cerebelo, incluyendo la diferenciación neuronal, morfogénesis, sinapsis y el aprendizaje y control motor [152].

La región 17q11-q12 es relevante porque en ella se localiza el transportador de serotonina (*SLC6A4*), y además de haberse identificado mutaciones de este gen en autistas [116-118], se ha comprobado que los niveles de serotonina en plasma son más altos en pacientes que en la población control [153].

Mutaciones en *MECP2*, *NLGN3* y *NLGN4* codificados en el cromosoma X también se han identificado en individuos afectados [123, 124, 139, 154, 155]. *NLGN3* y *NLGN4* son esenciales para la sinaptogénesis y *MECP2* participa en el *imprinting*.

---

Un estudio reciente soporta la asociación del gen *adenosine deaminase (ADA)* (20q13.12) con el autismo. Alteraciones en el metabolismo de las purinas y actividad reducida de ADA se han observado en autistas [156, 157].

Todos estos genes por las funciones que desempeñan y los sistemas en los que actúan son buenos candidatos para participar en el desarrollo del autismo pero algunos estudios de asociación aún son contradictorios, posiblemente por la heterogeneidad de fenotipos de la enfermedad [88]. En este sentido se ha comprobado que determinadas regiones podrían estar más relacionadas con el desarrollo del autismo dependiendo del sexo del paciente o de su origen ancestral. Las regiones 5p15.33, 9p24.1 y 11p13-p12 parecen estar asociadas al desarrollo del autismo en mujeres, 5q12.3 y 9q33.3 en hombres y 2q y 7q en familias de origen europeo, reflejando la complejidad de los factores genéticos que afectan al autismo [82].

#### 4.1.2 Alteraciones genómicas

Diferentes estudios muestran que 3-5% de los autistas presentan alteraciones citogenéticas [95, 96]. La más frecuente es una duplicación de la región 15q11-q13 [158-161], pero también se han observado deleciones de 2q37 [162, 163], 7q31 [164, 165], 22q11 [166] y 22q13.3 [167]. Las deleciones de 2q37 están a menudo asociadas a rasgos dismórficos, hipotonía, disfunción renal y braquidactilia. En la región 7q31 están codificados los genes *FOXP2*, *WNT2* y *GMR8*, asociados también al autismo. Deleciones de 22q11 también se han observado en casos de síndromes velocardiocfacial, de DiGeorge y anomalías conotruncales, síndromes con características clínicas comunes con el autismo [88].

Szatmari et al. [82] han identificado CNVs asociados al autismo. El reordenamiento más frecuente relacionado con dicho trastorno son duplicaciones de la región 15q11-q13, como demostraban los estudios anteriores. Observaron además 4 deleciones de 2p16 que afectan al gen *NRXN1*, interesante para el origen del trastorno autista, porque participa en la sinaptogénesis glutamatérgicas y posiblemente GABérgicas, e interaccionan con las neuroliginas, también asociadas al autismo. Así estas alteraciones refuerzan la asociación del sistema glutamatérgico con el trastorno autista. Identificaron 1 duplicación y 3 deleciones de la región 17p12 y una duplicación de 22q11.2. Duplicaciones de 17p12 se han detectado en pacientes con CMT1A, deleciones de la misma región en neuropatías, microduplicaciones en pacientes con retraso mental y déficit del lenguaje y microdeleciones en el síndrome de Smith-Magenis.

### 4.1.3 Otras enfermedades y síndromes asociados al autismo

Enfermedades génicas se relacionan con un mayor riesgo de desarrollar el trastorno autista. Las más comunes son la epilepsia, la esclerosis tuberosa (TSC) y el síndrome del cromosoma X frágil. La epilepsia es más frecuente en pacientes con retraso mental y que no han desarrollado la capacidad de hablar. TSC se debe a mutaciones en los genes *TSC1* (9q34) y *TSC2* (16p13). Estudios epidemiológicos reflejan que la prevalencia de TSC en niños con autismo es 100 veces mayor a la esperada. El cromosoma X frágil es una de las causas más frecuentes de retraso mental en niños y sobre 2-5% de los autistas presentan mutaciones en el gen *FMR1*. No son tan frecuentes pero también se conocen casos de fenilcetonuria, síndromes de Smith-Lemli-Opitz (SLO), de Sotos, WBS, AS, distrofia muscular de Duchenne y neurofibromatosis con características autistas [85, 88].

Es importante conocer la asociación del autismo con otras enfermedades porque algunas de ellas como SLO y la fenilcetonuria tienen tratamiento y de otras se conoce la causa genética (TSC y cromosoma X frágil) y puede ser interesante desde el punto de vista del consejo genético.

## 5. Chr15q11-q13 y enfermedades genómicas

La compleja organización de la región cromosómica 15q11-q13 hace que sea de especial interés para entender enfermedades genómicas. Las mejores caracterizadas son los Síndromes de Angelman (AS) y Prader-Willi (PWS) pero también anomalías citogenéticas de la región están asociadas a un fenotipo particular o a enfermedades complejas como el autismo.

### 5.1 Cromosomas marcadores inv dup(15)

Los cromosomas marcadores inv dup(15) son duplicaciones invertidas de la región proximal del cromosoma 15. Aparecen en 2/10.000 nacimientos y suponen el 57% de los cromosomas marcadores conocidos, siendo la anomalía autonómica más frecuente en humanos después de la trisomía del cromosoma 21 [73, 158, 168].

Se distinguen dos tipos: cromosomas marcadores cortos cuyos puntos de rotura coinciden con BP1 o BP2, no se les ha asociado ningún efecto clínico y son normalmente heredados; y cromosomas marcadores largos cuyos puntos de rotura se solapan con BP3 o YAC 810f11, se producen siempre *de novo* y son de origen materno (la duplicación paterna es letal). Estos llevan asociado un fenotipo particular caracterizado por retraso mental, hipotonía, alteraciones del comportamiento, ataques, estrabismo y dermatoglífico alterado [73, 158, 168]. Este reordenamiento se ha observado también en pacientes con epilepsia y autismo [169].

### 5.2 Síndromes de Angelman y Prader-Willi

Los síndromes de Angelman y Prader-Willi son trastornos neurológicos debidos alteraciones de la expresión génica en la región 15q11-q13 del genoma humano. Se dan con una frecuencia de 1/15.000 nacimientos [15, 36]. A pesar de estar ligados a la misma región, sus fenotipos son diferentes como consecuencia de si son los genes de expresión materna o paterna los que resultan alterados (Figuras 1.8 y 1.9). El AS se caracteriza por retraso mental moderado, dificultad para aprender, ausencia del habla, ataques epilépticos, ataxias y apariencia feliz [170]. Los pacientes con PWS presentan también retraso mental y en el crecimiento, hipotonía infantil, hipoplasia gonadal, obesidad y manos y pies pequeños [171].

La mayor parte de los casos de AS y PWS (70-75%) se deben a deleciones *de novo* de una región de ~4 Mb de 15q11-q13. Si la deleción afecta al alelo paterno, se produce el PWS y si afecta al materno el AS. Se conocen dos tipos de deleciones, las que

se producen entre BP1 y BP3 y las que tienen lugar entre BP2 y BP3. Son más frecuentes las deleciones entre BP2 y BP3 (Figura 1.6).

La siguiente causa de PWS de acuerdo a su frecuencias (20-30%) es una disomía uniparental (UPD) de la misma región con pérdida de la copia paterna (matUPD). patUPD se da en 2-3% de los casos de AS.

Otra causa común de ambos síndromes son mutaciones en el IC. Afectan a menos del 5% de los pacientes. Se trata de pequeñas deleciones en las regiones AS-IC o PWS-IC dependiendo del síndrome, y resultan en el epigenotipo contrario, es decir, pacientes con AS presentan una deleción en el AS-IC en su cromosoma materno y este desarrolla un epigenotipo paterno [170] y deleciones del PWS-IC provocan el epigenotipo materno en el cromosoma paterno portador de la deleción, dando lugar al PWS.

De las alteraciones de 15q11-q13 conocidas en AS y PWS la menos común (0,1%) son tranlocaciones que dependiendo del cromosoma (materno o paterno) al que afecten dan lugar a un síndrome u otro.

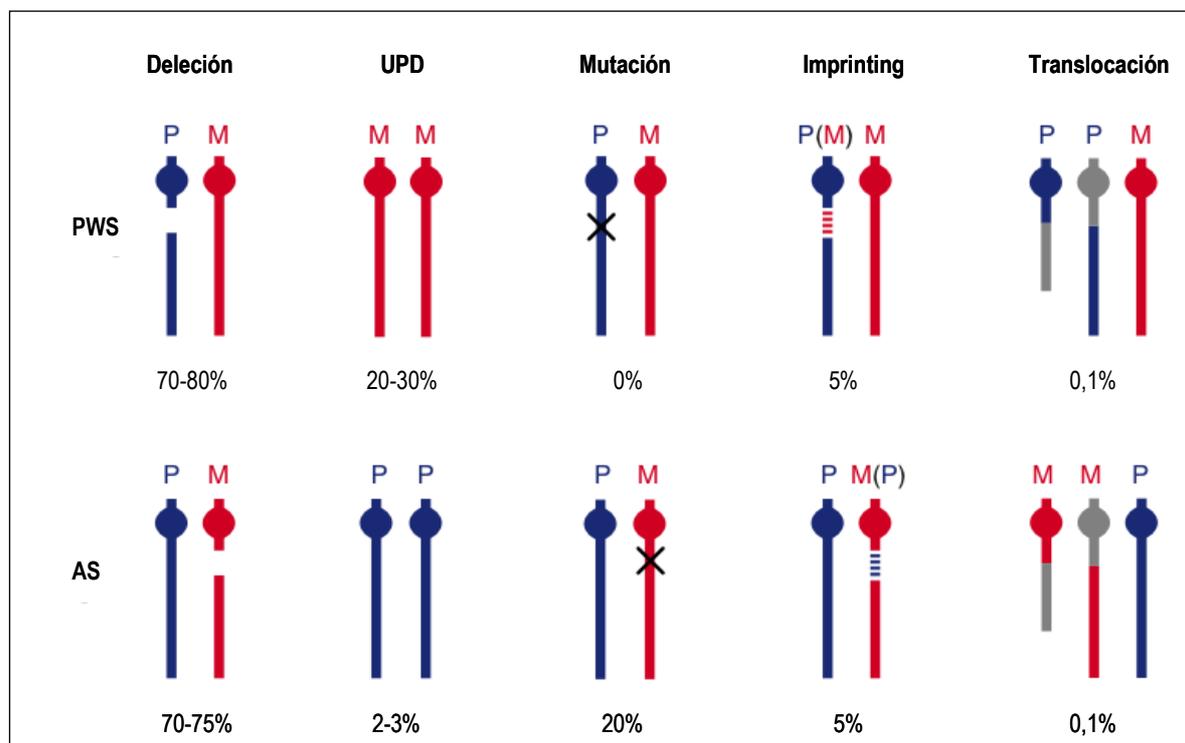
El 20% de los casos de AS se deben a mutaciones del gen *UBE3A* [172]. Se han identificado 52 mutaciones en este gen y todas son de novo o heredadas del alelo materno. El ~33% de las mutaciones identificadas son deleciones o duplicaciones y ~20% son mutaciones sin sentido. Los mecanismos más comunes por los que se producen son errores de replicación del DNA o en el caso de mutaciones puntuales C/T por la desaminación de la 5'-metilcitosina de una isla CpG. El efecto de varias mutaciones de *UBE3A* ha sido probado *in vitro* y se ha comprobado que impiden la ubiquitinación de los sustratos de la proteína, indicando que el AS se debe a un defecto en la degradación de proteínas, a través del proteasoma. Tanto estas mutaciones como las del IC, en los dos síndromes, se pueden transmitir a la descendencia sin que afecten al fenotipo durante varias generaciones, pero suponen un riesgo del 50% para desarrollar la enfermedad, cuando se transmiten a la descendencia del sexo contrario, por lo que se debe tener en cuenta para el consejo genético [15, 36, 170, 171].

La ausencia de mutaciones asociadas al PWS ha llevado a estudiar la contribución de los genes de expresión paterna de 15q11-q13 en su fenotipo, pero aún no se conoce si es la pérdida de expresión de uno o varios genes la causa del síndrome. Ratones *knockout* para *NDN* o con mutaciones para este gen no desarrollaron el fenotipo de PWS o resultaron letales, no siendo informativos para conocer las causas del síndrome [36]. Se generaron ratones con una deleción desde *SNURF-SNRPN* hasta *UBE3A* que desarrollaron hipotonía y retraso en el crecimiento por lo que los transcritos generados en este locus, incluyendo a los snoRNAs podrían explicar estas características [36]. Sin embargo, se han identificado tranlocaciones que afectan desde *IPW* hasta *UBE3A* que

---

no causan el fenotipo de PWS, por lo que la región crítica debe estar comprendida entre *SNRPN* y *IPW* y debe ser la pérdida de expresión de HBII-85 la causa del PWS [173]. Por otro lado, conocida la relación ente HBII-52 y las modificaciones de 5-HT2C, se ha observado que en pacientes con PWS los niveles de 5-HT2C con la modificación A/I son más bajos de lo normal, sugiriendo que HBII-52 y el sistema de la serotonina pueden estar relacionados con el síndrome. Igualmente debe ser considerada la posibilidad de que los genes *C15orf2*, *PWRN1* y *PWRN2* también participen en la etiología del síndrome [10, 11].

Aún quedan por conocer las bases moleculares de ~10% de los casos de AS [36].



**Figura 1.8 Causas de los síndromes de Angelman y Prader-Willi**

### 5.3 Autismo

Alteraciones citogenéticas de la región 15q11-q13 son el reordenamiento más frecuente identificado en autistas, se han observado en el 4% de los pacientes y se trata de duplicaciones de origen materno que pueden ser intersticiales (1% de loa autistas) o inv dup(15) (2% de los pacientes) [87, 158, 160, 161, 174]. Esto sugiere que la sobre-expresión de los genes improntados de expresión materna sea la causa de los síntomas autistas. Las duplicaciones de 15q11-q13 se asocian además a un fenotipo particular de autismo, caracterizado por mayor incidencia de epilepsia, hipotonía muscular, problemas en la coordinación motora, hiperactividad y retraso mental y del lenguaje [88].

La relación de esta región con el autismo está reforzada por diferentes estudios de asociación y/o ligamiento [174, 175], que han encontrado asociación de marcadores alrededor de los genes *GABRB3* y *GABRG3*, del extremo 5' del gen *UBE3A* (D15S122) y del gen *ATP10A* [107, 109, 112-115, 176].

El fenotipo autista comparte características clínicas con el síndrome de Rett (RTT) como retraso mental, deficiencias en la comunicación y comportamientos repetitivos y estereotipados. Aproximadamente el 80% de los casos de RTT se deben a una mutación del gen *MECP2* y se ha demostrado que la deficiencia de *MECP2* reduce la expresión de *UBE3A* y *GABRB3* en el cerebro de pacientes con síndrome de RTT [177]. Por otro lado, expresión reducida de *UBE3A* y *GABRB3* se ha observado también en autistas [140] y mientras que se ha demostrado que la expresión de las tres subunidades del receptor de GABA codificadas en 15q11-q13, son de expresión bialélica en el cerebro, estos genes en autistas y pacientes con RTT presentan expresión monoalélica, posiblemente paterna [8]. Estos resultados sugieren que *MECP2* podría estar implicado en las alteraciones de la expresión de *UBE3A* y *GABRB3* en autistas y RTT.

Microrreordenamientos de 15q11.2 (una delección materna del marcador *D15S646* y dos duplicaciones paternas de *D15S817*) se han observado también en RTT [178], sugiriendo que las características clínicas y las alteraciones en la regulación génica que comparten, en algunos casos, pueden ser debidas microrreordenamientos comunes de esta región.

Uno de los snoRNAs codificados en 15q12, HBII-52, hemos mostrado que regula en splicing del mRNA del receptor de serotonina [28, 31] e independientemente, se ha comprobado que los niveles de serotonina son más altos en autistas [153], por lo los snoRNAs codificados en 15q12 podrían así participar en la etiología del autismo. Además HBII-52 coincide con un bloque de duplicaciones segmentarias que podrían facilitar reordenamientos y producir la pérdida o ganancia de algunas copias.

## 6. Non-coding RNAs

Datos recientes sobre la actividad transcripcional del genoma humano sugieren que aproximadamente la mitad del DNA genómico se transcribe [179-181]. Sólo el 1,2% de los transcritos humanos codifican proteínas y ~98% son RNAs no codificantes o *non-coding* RNAs (ncRNAs). La importancia de los ncRNAs se hace más evidente al comprobar que un incremento de la complejidad de los organismos va acompañado de la acumulación de secuencias de DNA que no se traducen [181, 182].

En el término ncRNA se incluyen todos los RNAs que no codifican proteínas y funcionan directamente como RNAs. Los primeros ncRNAs descritos hace varios años fueron los RNAs ribosómicos (rRNAs) y los RNA de transferencia (tRNAs) que participan en la traducción del RNA mensajero (mRNA). Los siguieron los *small nuclear RNAs* (snRNAs) y *small nucleolar RNAs* (snoRNAs) que forman parte del complejo de *splicing* y la modificación de rRNAs respectivamente. Recientemente, se han descrito dos nuevas clases de ncRNAs: los microRNAs (miRNAs) y *small interfering RNAs* (siRNAs), reguladores de la expresión génica [183]. Finalmente, diferentes trabajos publicados durante el año pasado hacen referencia a una nueva clase ncRNAs llamados *Piwi Interacting RNAs* (piRNAs) que parecen estar implicados en la gametogénesis [184-187]. Al mismo tiempo, otros ncRNAs se han identificado y agrupado en clases dependiendo de su función o patrones de expresión, como por ejemplo, ncRNAs improntados, RNAs inducidos por estrés, RNAs del SNC, RNAs específicos de un tejido o momento del desarrollo, etc... [183, 188]. Así, los ncRNAs se han convertido en elementos fundamentales para entender cualquier proceso importante en células procariontas y eucariotas.

En general el tamaño de los ncRNAs conocidos varía desde ~21 nt que miden los miRNAs y siRNAs hasta 500nt, pero se han descrito ncRNAs muy largos como XIST que mide ~17 Kb [189].

Los patrones de expresión de los ncRNAs son muy heterogéneos: pueden ser específicos de un tejido o momento del desarrollo [28, 190-193]; pueden estar codificados en intrones o exones de genes codificantes o no codificantes, en regiones intergénicas, en los extremos 3' y 5' de otros genes, e incluso se han identificado miRNAs en regiones repetitivas del genoma [194-197]. Se pueden transcribir a partir de genes improntados [198, 199] o ser *antisense* a un gen codificante [200-202]. También tienden a agruparse en clusters, transcribiéndose como genes policistrónicos o cada uno puede tener su promotor independiente [203, 204].

## 6.1 Identificación y detección de ncRNAs

En un principio, la clonación de RNAs pequeños de librerías de cDNAs supuso un éxito para la identificación de nuevos ncRNAs [203-205]. Sin embargo, esta técnica está limitada por la secuenciación y sólo resulta adecuada para los ncRNAs ubicuos y con niveles de expresión altos, ya que serían necesarias librerías de cada tejido y momento del desarrollo, así como secuenciar muchos clones para detectar todos los ncRNAs con este método. Las limitaciones de la secuenciación se han resuelto con el desarrollo de técnicas de *Massively Parallel Signature Sequencing* (MPSS), pero hasta hoy sólo se ha realizado un trabajo con este método para identificar miRNAs en humanos [206].

La siguiente aproximación utilizada son las predicciones bioinformáticas. El primer problema para diseñar estas herramientas es la ausencia de señales específicas de ncRNAs en la secuencia primaria, por lo que la mayor parte de ellas se basan en la tendencia de los ncRNAs a formar estructuras secundarias estables [207]. Algunos programas como Mfold [208, 209] o RNashapes [210] se basan simplemente en la predicción de estas estructuras y se han utilizado con éxito para identificar ncRNAs, apoyándose en la conservación a lo largo de la evolución, para disminuir el número de falsos positivos [205, 211, 212]. Se han desarrollado programas que además de predecir la estructura secundaria hacen una estimación de su estabilidad, es decir, miden si la probabilidad de que se forme una determinada estructura secundaria es mayor a la formación de cualquier otra combinando las bases de su secuencia aleatoriamente. De este tipo es el programa Randfold que se ha utilizado para filtrar miRNAs candidatos, porque se ha comprobado que al 95% de los miRNAs les corresponde una  $P < 0,01$  [213, 214]. Basado sólo en la estructura secundaria se ha creado el software miRabela para predecir miRNAs [215]. El siguiente paso, ha sido el desarrollo de programas que utilizan la conservación de la estructura secundaria de un determinado ncRNA en otras especies para predecir nuevos ncRNAs. Son por ejemplo QRNA, RNAz, Dynalign, MiRscan, MiRseeker y miRalign [211, 216-220]. Los tres últimos son también específicos para miRNAs. Entre ellos RNAz ha sido considerado el más efectivo para predecir ncRNAs, por combinar la estabilidad de la estructura secundaria con su conservación en otros genomas. También existen software como CMsearch o FastR para buscar secuencias homólogas a otros ncRNAs previamente identificados [221, 222]. Las limitaciones de las herramientas bioinformáticas son sus propios parámetros de selección, así los programas basados en la conservación no son útiles para identificar ncRNAs específicos de una especie, los basados en secuencias homólogas sólo sirven para identificar nuevos miembros de familias de ncRNAs ya conocidas y los que se basan en las características

de las estructuras secundarias no resultan adecuados para identificar ncRNAs con otro tipo de estructura secundaria. También la expresión de los candidatos debe ser comprobada experimentalmente.

Los métodos de hibridación son los más utilizados para comprobar la expresión de ncRNAs predichos. Entre ellos destacan el *Northern Blot* y *Primer extension* que además dan información sobre el tamaño y el extremo 5' del transcrito [223]. Estas técnicas no son muy sensibles y requieren mucho tiempo para detectar un solo ncRNA, no resultando adecuadas para estudios de gran escala. Ambos métodos han evolucionado para la identificación de miRNAs. En el caso de *Primer Extension*, para detectar miRNAs, se ha desarrollado *RNA-prime array based-Klenow extension* (RAKE), en el cual los miRNAs se hibridan con sondas específicas que son utilizadas como *primers* por la enzima Klenow para una reacción de extensión. Así se puede comprobar la expresión de muchos miRNAs a la vez y además conocer su extremo 3'. Con este método se pueden detectar miRNAs de tejidos conservados en parafina, abriendo nuevas oportunidades para el análisis de RNAs de tejidos humanos archivados, ya que es un recurso limitado [194, 205, 223]. En el caso del *Northern Blot* se han utilizado con éxito sondas marcadas con digoxigenina y *Locked-nucleic acid modified oligonucleotides* (LNA) que no resuelven las limitaciones del tiempo pero si son más sensibles [194, 224].

Los métodos de hibridación descritos no permiten distinguir la expresión de ncRNAs entre células próximas. Con este fin se han desarrollado técnicas que permiten visualizar la expresión de un miRNA *in vivo*. Se utilizan transgenes "sensores" que expresan un gen reportero, con la secuencia complementaria a un miRNA en su extremo 3'UTR. En las células que no expresan el miRNA, tendrá lugar la expresión del reportero, pero en la que se transcriba el miRNA, este se une al extremo 3' del reportero impidiendo su expresión. Este método permite diferenciar la expresión espacio-temporal de miRNAs pero aún no se sabe si es útil para aquellos con niveles de expresión bajos o si la expresión del transgen puede interferir con la de los miRNAs endógenos de la célula [194, 223, 225].

Se pueden detectar ncRNAs mediante *real time* RT-PCR. Esta técnica se utiliza más para estudios cuantitativos que para comprobar la expresión de nuevos ncRNAs [194, 223, 226].

El método más utilizado para la identificación de nuevos ncRNAs a gran escala son los *arrays* por ser más sensibles y porque permiten diferenciar miRNAs relacionados [212, 227-230].

Todos estos métodos han servido para la identificación de ncRNAs pero aún es difícil estimar su número en el genoma humano. Sólo de miRNAs se estima que pueden

existir más de un millar, de entre los que 300-400 serían específicos de humanos y de los que sólo 474 están actualmente validados experimentalmente y registrados en la base de datos miRBase [212, 231]. Mediante *Genome Tiling Array* y MPSS se ha estimado que existen más de 50.000 transcritos no codificantes en el genoma humano, de los que se conocen alrededor de 800 [188]. Sólo piRNAs, específicos de la línea germinal masculina y con un tamaño de ~30 nt ya se conocen más de 50.000 en cada una de las especies de ratón, rata y humano [184-187].

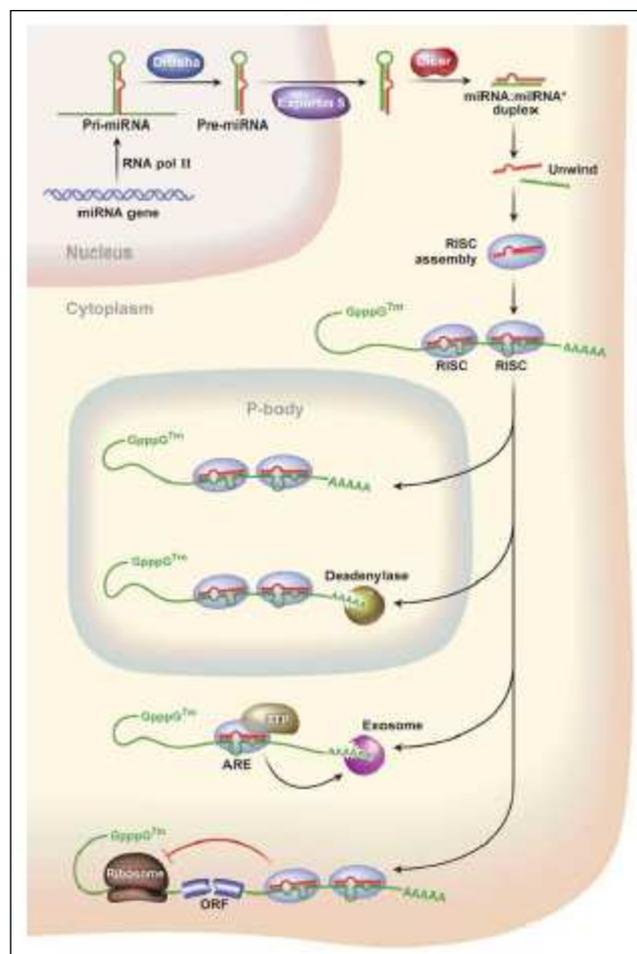
## 6.2 Biogénesis de miRNAs

A pesar de los esfuerzos dedicados a la búsqueda de nuevos ncRNAs, se sabe muy poco sobre cómo se transcriben muchos de ellos. En algunos casos no se conoce ni la polimerasa que lo sintetiza. Sin embargo, en otros casos se han identificado promotores con sitios de unión a factores de transcripción, una cola poliA e incluso parece que son modificados post-transcripcionalmente, indicando que podrían ser sintetizados por la RNA polimerasa II [189].

El único tipo del que se conoce todo el proceso de biosíntesis son los miRNAs. El transcrito primario o pri-miRNA es sintetizado por la RNA polimerasa II y tiene el resto de 7-metilguanósina y la cola de poliA, característicos de la transcripción por esta enzima. En algunos miRNAs se han identificado elementos reguladores como la caja TATA pero no se puede descartar la posibilidad de que otros sean sintetizados por otras polimerasas. El pri-miRNA es reconocido por Drosha, una RNAsa nuclear que da lugar al pre-miRNA, de ~70 nt y con forma de *hairpin*. El resto del pri-miRNA es degradado en el núcleo. Drosha es una proteína de 160 KDa, conservada en animales pero no en plantas. En humanos requiere la presencia del cofactor *DiGeorge syndrome critical region gene 8* (DGCR8). DGCR8 tiene dos dominios de unión a RNA y se cree que participa en el reconocimiento del pri-miRNA. El pre-miRNA es transportado hacia el citoplasma por otra proteína llamada exportina-5 (exp-5). Una vez en el citoplasma, Dicer, que es otra RNAsa, y una helicasa dan lugar a la forma madura. Dicer, de ~200 KDa, también está muy conservada y en humanos se une a la proteína *HIV-1 TAR RNA Binding* (TRBP). Los cofactores de Dicer no parecen necesarios para la reacción en sí, si no que estabilizan al miRNA. El producto de Dicer son dos cadenas de RNA de ~22 nt: la más estable es el miRNA madura y se une al complejo proteico RISC (*RNA-induced silencing complex*) y la otra es degradada. Si las dos cadenas resultan igual de estables, las dos pueden unirse a RISC y funcionar como miRNAs maduros. Los principales componentes de RISC son diferentes miembros de la familia de

proteínas *Argonaute* (AGO) con un dominio PIWI/PAZ. RISC tiene múltiples funciones: estabiliza al mRNA diana, guía al miRNA hacia el *target* y si la complementariedad entre ellos es perfecta, guía la degradación del mRNA y si es imperfecta facilita la inhibición de la traducción. Junto a la proteína TTP puede desviar mRNAs ricos AU a los exosomas para su degradación (Figura 1.9) [194, 232-235].

La regulación de la expresión génica por miRNAs tiene lugar en los P-bodies, orgánulos citoplasmáticos libres de ribosomas, donde se agrupan mRNAs no traducidos, proteínas de degradación de mRNAs y represores de la traducción [236, 237].



**Figura 1.9 Biosíntesis de miRNAs**

### 6.3 Regulación de la expresión génica por miRNAs

Muchos de los ncRNAs reconocen al RNA al que modifican o regulan a través de una secuencia complementaria. La identificación de *targets* de miRNAs no resulta sencilla, sobre todo para los miRNAs de animales, ya que la complementariedad entre ellos no es perfecta. Tampoco se han desarrollado hasta ahora técnicas que permitan analizar experimentalmente muchos *targets* a la vez [233].

Dependiendo de cómo sea la interacción entre el target y el miRNA, los sitios de unión se clasifican en dos categorías: 5' dominantes y 3' compensatorios. En la primera categoría la región de complementariedad entre ambos incluye 2-8 nt, del extremo 5' del miRNA, que suelen estar conservados en familias de miRNAs. La unión resulta suficiente para que este sea funcional. Se dividen a la vez en otros dos grupos: los que tienen otra región complementaria en el extremo 3' y los que no requieren la unión en este extremo para realizar su función. La segunda categoría la componen los miRNA:mRNA que dependen de una complementariedad fuerte en el extremo 3' del miRNA. En humanos algunos miRNAs presentan una complementariedad perfecta con el mRNA de más de 10 nt [194, 233].

Se han desarrollado diferentes programas para predecir targets de miRNAs (Diana Micro Test [238], MiRanda [239], PicTar [240], RNAhybrid [241], TargetBoost [242] y TargetScan [243]) y aunque difieren entre ellos en detalles técnicos, todos se basan en la complementariedad entre el miRNA y el extremo 3'UTR de los mRNAs, la estabilidad de la unión miRNA:mRNA, y algunos además en la conservación del sitio de unión del mRNA en los genomas de especies relacionadas. En humanos se estima que cada miRNA puede regular hasta 200 mRNAs diferentes y se ha demostrado que su capacidad reguladora depende de la unión a múltiples sitios del mRNA. Independientemente, cada mRNA puede ser *target* de distintos miRNAs y se predice que la expresión del 30% de los genes humanos está regulada por miRNAs. Hay que tener en cuenta que sólo una parte de los *targets* han sido comprobados experimentalmente [27, 194, 233, 244].

El método más utilizado para comprobar si un mRNA es *target* de un miRNA es un transfección celular con un vector que contiene la región de unión del mRNA al miRNA, *downstream* la región codificante de un reportero (luciferasa). A la vez, otras células son transfectadas con la misma construcción pero con mutaciones en la región de unión al miRNA. El *target* se comprueba por una disminución de la actividad del reportero *wild-type* respecto al mutado. Una aproximación complementaria es la inhibición de los miRNAs con un oligorribonucleótido modificado *antisense* al miRNA. En este caso un incremento de la actividad luciferasa es interpretado como la inhibición de la actividad del miRNA.

---

También se puede incrementar la concentración de un miRNA mediante transfección de RNAs sintéticos, homólogos a un determinado miRNA, y medir la expresión génica mediante *microarrays* [225].

Además de comprobar los *targets* de un miRNA, se pueden estudiar sus funciones en determinados procesos celulares sobre-expresándolos mediante la transfección de RNAs de doble cadena, que mimetizan el producto de *Dicer*, o silenciándolos por la inactivación o mutación de *Dicer*, por la generación de ratones *knockout* de miRNAs, induciendo mutaciones de los sitios de unión del miRNA en el mRNA o con oligonucleótidos *antisense* al miRNA [225].

## 6.4 Funciones de ncRNAs

Durante los últimos años los ncRNAs han sido implicados en una gran variedad de procesos moleculares y celulares. Para definir su función se han utilizado experimentos de RNAi, ratones *knockout* y vectores virales [189], además de los específicos para miRNAs descritos previamente.

En este punto se describirán algunas de las funciones de los ncRNAs en mamíferos.

### Epigenética

Los ncRNAs han sido considerados importantes factores para las modificaciones epigenéticas por su actuación en la inactivación del cromosoma X, la estructura de la cromatina, el *imprinting* y la metilación del DNA.

Los ncRNAs XIST y TSIX, codificados en el cromosoma X, son fundamentales para la inactivación de dicho cromosoma. Ambos se expresan en la misma región pero en sentido contrario. XIST inactiva un cromosoma X en las hembras y TSIX en el otro cromosoma X inhibe a XIST [245].

AIR es un ncRNA que se origina en el segundo intrón del gen *IGF2r*, está improntado y se expresa sólo en el alelo paterno. Su expresión inhibe la transcripción de tres genes que se transcriben en el sentido contrario: *Igf2r*, *Slc22a2* y *Slc22a3* [246].

La sobre-expresión de *KHPS1* causa la demetilación de islas CpG en el promotor de *SPHK1*, un gen con metilación específica de tejido [247].

El ncRNA SRA actúa como co-activador en la transcripción de receptores de esteroides [248].

HP1 se une a la histona H3 metilada y otros RNAs en regiones de heterocromatina [245].

Mir-1 y mir-133 se transcriben juntos, son específicos de tejidos específicos y se expresan durante el desarrollo. Mir-1 promueve la miogénesis regulando la expresión de

histona deacetilasa 4 (*HDAC4*), que a la vez reprime a genes musculares. Mir-133 estimula la proliferación de mioblastos reprimiendo a *SRF* [249].

### Desarrollo

Diferentes observaciones demuestran que los miRNAs son esenciales para el desarrollo normal de los mamíferos: las células madre embrionarias expresan miRNAs específicos, células madre deficientes en *Dicer* son viables pero no forman miRNAs maduros y no se diferencian y embriones de ratones con mutaciones en el gen que codifica a *Dicer* mueren en el estado de gástrula [250].

Mir-196 regula la expresión de *HOXB8* durante el desarrollo embrionario [251].

Mir-1, un miRNA específico del músculo, regula el balance entre la diferenciación y la proliferación de los cardiomiocitos durante el desarrollo del corazón en ratones [252].

Mir-181 modula la diferenciación del linaje hematopoyético también en ratones [253].

Mir-143 regula la diferenciación de los adipocitos [254].

### Procesos fisiológicos

Algunos miRNAs actúan en diversos procesos fisiológicos y celulares:

Mir-375 se expresa en las islas pancreáticas, donde regula al gen *MTPN* y estimula la secreción de insulina [255].

Mir-122 inhibe la traducción de transportadores de aminoácidos catiónicos en el hígado [256].

mir-16 participa en la degradación de mRNAs ricos en AU [257].

### ncRNAs del SNC

Muchos ncRNAs se han relacionado con el desarrollo del cerebro, la diferenciación neuronal y funciones más complejas como el aprendizaje y la memoria.

BC1 y BC200 son ncRNAs citoplasmáticos específicos del cerebro. Ambos se expresan en neuronas y se transportan hasta las dendritas, pero BC1 es específico de roedores y BC200 de primates. Se unen a FMRP, una proteína que inhibe la traducción de mRNAs específicos en las sinapsis. BC1 y BC200 probablemente guían a FMRP hacia sus *targets* (*Arc*, *MAP1B*, y  $\alpha$ CaMKII), proteínas importantes para el aprendizaje y la memoria. BC1 y BC200 también interactúan con proteínas de unión a poliA, reguladoras del inicio de la traducción [235, 258-260].

NRSE (*Neuron Restictive Silencer Element*) es un ncRNA de doble cadena de 20 bp. La unión del NRSE al factor de transcripción NRSF/REST impide su unión a proteínas represoras como HDACs o *methyl-DNA binding proteins*, activando la transcripción de este directamente y la de otros genes neuronales indirectamente.

Evf-2 es otro ncRNA regulado durante el desarrollo y actúa como co-activador de la transcripción en células neuronales [261].

La mayor parte de los snoRNAs son ubicuos, sin embargo HBII-52, HBII-13 y HBII-85 son específicos de cerebro y parecen regular mRNAs en lugar de rRNAs, por lo que pueden ser relevantes para el desarrollo y funcionamiento correcto del SNC [28].

Entre 20-40% de los miRNAs expresados en el cerebro están regulados durante el desarrollo y algunos se expresan sólo o principalmente en un tipo celular: mir-124 y mir-128 se expresan principalmente en neuronas, mir-23, mir-26 y mir-29 en astrositos y mir-9 y mir-25 están distribuidos entre los dos tipos celulares. mir-124 es el uno de los más abundantes en cerebro y es capaz de inhibir transcritos no neuronales reprimiendo a un factor transcripcional llamado REST. Así regula la diferenciación y dirige a las neuronas. Entre los *targets* que se han predicho para los miRNAs expresados en cerebro predominan los que actúan en la división celular asimétrica de neuroblastos y en la diferenciación neuronal, pero también están representados genes que regulan el crecimiento de las dendritas y los axones, participan en las sinapsis, receptores de la neurotransmisión y canales iónicos [233].

## 6.5 ncRNAs y enfermedades

Después de conocer la diversidad de procesos regulados por ncRNAs y su relevancia, no resulta sorprendente que muchos estén asociados a enfermedades.

### Cáncer

El cáncer es una enfermedad compleja en la que intervienen múltiples factores genéticos y epigenéticos. Cambios en la expresión de algunos ncRNAs se han asociado a distintos tipos de cáncer.

Alteraciones de la metilación de las islas CpG *upstream* del gen *H19* resulta en la pérdida del imprinting de los genes *H19* y *IGF2* y su expresión bialélica induce el crecimiento de células malignas [262].

Carcinomas de colon presentan niveles elevados de OCC-1 que está ausente o sus niveles de expresión son muy bajos en células de la mucosa normales [263].

Dos ncRNAs (DD3 y PCGEM1) están sobre-expresados en tumores de próstata comparado con tejidos sanos. DD3 se utiliza como marcador tumoral y PCGEM1 parece estar implicado en el crecimiento celular [264, 265].

El RNA HOST-2 es específico de cáncer de ovarios en humanos aunque su función biológica en el desarrollo de la enfermedad es aún desconocida [266].

La expresión de SRA es mayor en cáncer de mama, útero y ovario que en tejido normal [248].

En células cancerígenas de pulmón la metástasis está asociada a la expresión de *MALAT-1* (*Metastasis Associated in Lung Adenocarcinoma Transcript 1*). La medida de su nivel de expresión contribuye a la indentificación de la enfermedad en las primeras etapas [267].

Más de la mitad de los miRNAs está localizados en regiones cromosómicas relacionadas con el cáncer, sugiriendo que estos tienen un papel en la tumorigénesis.

El cluster de mir-15 y mir-16 están delecionados en más de la mitad de las leucemias [268].

La expresión de mir-143 y mir-145 está alterada en células cancerígenas de colon [269].

Mir-155 y BIC se concentran en diferentes linfomas. Los linfomas de células B tienen peor pronóstico que los linfomas de centros germinales y se refleja en los niveles de mir-155 [270, 271]. A la vez, mir-17 mir-92 están en una región que resulta amplificada en linfomas de células B. Su sobre-expresión previene la apoptosis, y en ratones se ha comprobado que acelera la formación de este tipo de linfomas mediante la inducción de *c-Myc* [272].

El miRNA let-7 impide el crecimiento de las células cancerígenas de pulmón donde su nivel de expresión está disminuido [273].

Células de glioblastoma sobre-expresan mir-21 que se cree puede funcionar como factor antiapoptótico [274].

Mir-124 se encuentra hipermetilado en células tumorales y se ha relacionado la pérdida de su estado epigenético con la activación de un oncogen [275].

### **Enfermedades neurológicas**

Diferentes ncRNAs se han relacionado con alteraciones del SNC durante los últimos años:

El síndrome de Beckwith-Wiedemann (BWS) resulta como consecuencia de alteraciones del imprinting de dos genes no codificantes (*H19* y *LIT1*) de la región 11p15.5, bien por alteraciones de la metilación o por microdeleciones de la región metilada de *H19* [276].

La ataxia espinocerebral (SCA) es una de las enfermedad neurodegenerativa causadas por extensiones de poliglutaminas. En este caso el gen afectado (*SCA8*) no se traduce. [277]

Los niveles de transcripción de BC200 son el 70% inferior en pacientes con alzheimer que en individuos controles [278].

Algunos pacientes con esquizofrenia son portadores de la translocación t(1:11)(q43, q14). En la región cromosómica 1q43 están codificados *DISC1* y *DISC2*.

---

DISC2 es un ncRNA que se transcribe en dirección opuesta a *DISC1* y podría participar en su regulación. En 11q14 se localiza otro ncRNA (PSZA11q14) cuya expresión está reducida en pacientes [279].

Otra translocación t(7:13)(q31.2, q21) que afecta a los ncRNAs ST7OT4 y ST7OT3 se ha encontrado en un niño autista. La función de estos transcritos es aún desconocida pero parece que podrían inhibir la traducción de los genes de la región (locus *RAY1/ST7*) que se transcriben en sentido contrario [165].

La sobre-expresión de *PRINS* está asociada con la susceptibilidad a desarrollar soriasis. Su expresión se induce cuando las células se someten a condiciones de estrés, como la irradiación con UV-B o la infección por HSV. Se ha sugerido que PRINS puede ser un protector celular frente al estrés [280].

El síndrome de Tourette se ha asociado a mutaciones en la región que reconocería el miRNA mir-189 del gen *SLITRK1*. Los dos genes presentan el mismo patrón de expresión en regiones cerebrales implicadas en el síndrome [281].

El RNA DGCR8, cofactor de Drosha, está codificado en la región 22q11.2, delecionada en pacientes con síndrome de DiGeorge. Esto sugiere que alteraciones en la biogénesis de miRNAs pueden afectar a los pacientes con este síndrome [233].

Otro ejemplo es el síndrome del cromosoma X frágil, causado por la expansión de los trinucleótidos CGG en el gen *FMR1*. La proteína codificada por *FMR1* (FMRP) regula la traducción de algunos genes neuronales, por la vía de los miRNAs [233].

Las encefalopatías espongiiformes están causadas por la infección de priones (PrP<sup>Sc</sup>) que para propagarse necesitan un cambio conformacional (PrP<sup>C</sup>) y algunos trabajos muestran que en este cambio participa un ncRNA [282].

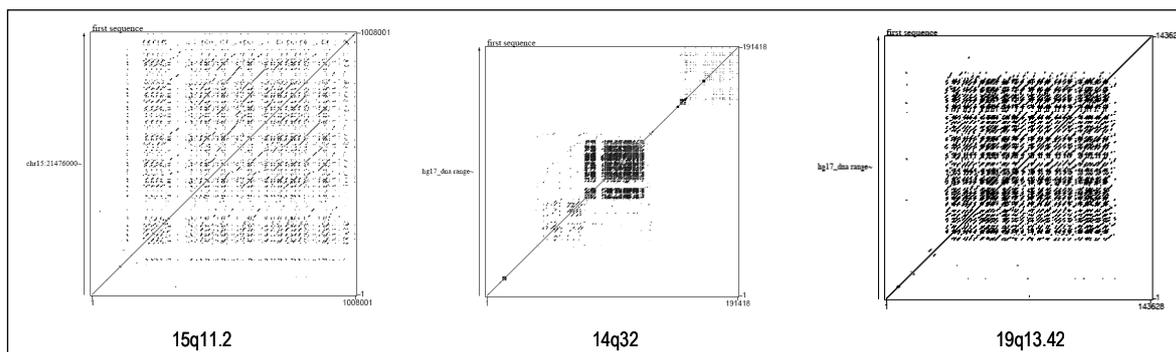
### **Infecciones virales**

MiRNAs se han identificado también en virus de mamíferos (herpesvirus, Epstein-Barr virus, virus asociados al sarcoma de Kaposi, HIV...). De muchos no se conoce su función pero se espera que regulen la expresión de genes virales y del organismo infectado, para sobrevivir y propagarse. Por ejemplo, dos miRNAs del SV40 reducen la expresión de antígenos T virales, convirtiendo a las células infectadas en menos susceptibles a las células T citotóxicas y facilitando así su propagación. El mir-32 previene la acumulación del virus PFV-1 en células humanas. Sin embargo, una proteína codificada por el virus (TAS) suprime la inhibición inducida por mir-32, indicando que los miRNAs regulan diversos aspectos de las infecciones virales en humanos [250].

## 6.6 Región 15q11.2 del genoma humano: candidata para codificar ncRNAs

La región 15q11-q13 incluye ~1 Mb en las que hasta hace unos meses no se conocían genes pero si se habían predicho algunos y se habían identificado ESTs, indicando que existía actividad transcripcional en 15q11.2. Observado su organización se encuentran evidencias para pensar que un *cluster* de ncRNAs puede ser la parte funcional de dicha región.

Un alineamiento con Pipmaker [283] de la secuencia chr15:21.483.000-22.477.000 (hg18) con ella misma, eliminando previamente las repeticiones [284], reveló que la secuencia contenía muchos elementos repetitivos en *tandem* (Figura 1.10). Hertel et al. [285] han publicado que las duplicaciones en *tandem* son un mecanismo frecuente para la expansión de miRNAs y *blots* similares resultaron del mismo tipo de análisis con secuencias de otras regiones donde están codificados clusters de miRNAs (14q32 y 19q13.42) (Figura 1.10), añadiendo evidencias de que ncRNAs pueden estar codificados en esta megabase.



**Figura 1.10 Resultados de analizar con pipmaker la secuencia de 15q11.2 y de otras dos regiones más (14q32 y 19q13.42) que codifican miRNAs**

También se ha establecido una relación entre regiones de inestabilidad y la presencia de ncRNAs [286, 287]. Entendiendo como regiones inestables aquellas que presentan pérdida o ganancia de fragmentos de DNA o alteraciones de la expresión génica, 15q11.2 puede ser considerada inestable y candidata para codificar ncRNAs, ya que se han identificado CNVs y BPs en su secuencia o flanqueándola (Figura 1.6) [55, 59, 66, 70].

La región 15q11-q13 comparte características con 14q32 (un cluster de snoRNAs y regulación génica por imprinting) y en esta se han observado además miRNAs.

Recientemente Buiting et al. [11] identificando dos ncRNAs (PWRN1 y PWRN2) en 15q11.2, pero se han observado repeticiones de la región más pequeñas y otros

aspectos de su organización parecen indicar que también podrían estar codificados miRNAs, sugiriendo una mayor complejidad de la actividad transcripcional de la región. Además, las predicciones de ncRNAs estiman que aún quedan por conocer cientos de ncRNAs en el genoma humano [205, 212, 214] y 15q11.2 reúne condiciones para codificar una parte de ellos.



# ***Objetivos***



## **Objetivos**

El objetivo principal de esta tesis doctoral ha sido contribuir a la caracterización de la región cromosómica 15q11-q13 del genoma humano, centrándonos en la variabilidad genómica y su significado a nivel de fenotipo. Elegimos el trastorno autista para analizar reordenamientos de dicha región cromosómica asociados a un fenotipo particular; e intentamos identificar la parte funcional de la región 15q11.2, caracterizada por su variabilidad e inestabilidad, para entender cómo las alteraciones en este locus, afectan a la diversidad humana y al desarrollo de enfermedades. En este punto, nos encontramos con la presencia de los ncRNAs, los cuales son claramente indispensables para entender casi cualquier proceso fisiológico y/o del desarrollo.

Así definimos los objetivos puntuales de este proyecto:

- Analizar la presencia de reordenamientos genómicos (duplicaciones y/o deleciones) de la región 15q11-q13 en pacientes con autismo.
- Definir los límites de una deleción polimórfica observada en la región 15q11.2.
- Identificar posibles ncRNAs en la región cromosómica 15q11.2.



# ***Materiales y métodos***



## 1. Reordenamientos de 15q11-q13

### 1.1 Muestras

Las muestras clínicas sobre las que se ha realizado este estudio las forman 42 familias caucásicas, con dos o más hijos autistas, de diferente sexo, diagnosticados según la Entrevista Diagnóstica del Autismo Revisada (ADI-R) [288]. En todas las familias los padres eran sanos.

La población control está formada por un grupo de 64 donantes de sangre (32 hombres y 32 mujeres) sanos, de origen español y no relacionados entre ellos y por 29 tríos CEPH [289-292]. Las familias CEPH son familias también caucásicas, residentes en Utah, y con antecesores del norte y oeste de Europa. También se las conoce como familias CEU.

En este estudio no se ha llevado a cabo ninguna intervención clínica ni de diagnóstico sobre animales o humanos. Los análisis genéticos fueron realizados sobre DNA genómico extraído de linfocitos por el sistema *Chemagic Magnetic Separation Module I System* (Chemagen). Las líneas celulares de los pacientes autistas y controles CEPH fueron compradas a las asociaciones *Autism Genetic Resource Exchange* (AGRE) [161] y *Coriell Institute for Medical Research* [159] respectivamente. Ambas organizaciones se encargan de la recogida y distribución de muestras humanas siguiendo los principios éticos de la declaración de Helsinki. La identidad de los donantes de sangre españoles utilizados como control, está protegida por la ley española 15/1999 de protección de datos de carácter personal. (Ley Orgánica 15/1999).

Para aumentar el poder estadístico del test FBAT, genotipamos 10 familias AGRE adicionales con las mismas características que las anteriores, sólo para los marcadores D15S1548 y D15S1549

### 1.2 Selección y amplificación de marcadores microsatélites

Los microsatélites de la región cromosómica 15q11-q13 fueron seleccionados desde la base de datos *University of California Santa Cruz (UCSC) Genome Browser* [168] ensamblaje hg17. Para su selección consideramos que tuvieran una única representación en el genoma y una heterocigosidad superior al 50%. La heterocigosidad de cada marcador fue calculada previamente a su selección, genotipando 64 muestras controles no relacionadas entre ellas.

Diseñamos los *primers* para la amplificación de los marcadores microsatélites con los programas Primer 3 [74] y Fast PCR [93] (Tabla 2.1). Uno de cada pareja necesaria para una amplificación estaba marcado con un fluorocromo (Fam, Ned o Vic) y todos fueron comprados a las compañías comerciales Sigma Aldrich o Applied Biosystems. Las amplificaciones fueron llevadas a cabo en un volumen total de 20  $\mu$ l incluyendo 50 ng de DNA, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl<sub>2</sub>, 200 mM de dNTP (Pharmacia Biotech), 10 pmol de cada oligonucleótido y 0.25 unidades de *Taq DNA polymerase* (Roche). Las condiciones de las amplificaciones fueron: una desnaturalización inicial de 4 minutos a 94°C; 25 ciclos de 30 segundos a 94°C, 30 segundos a la temperatura de anillamiento específica para cada marcador (Tabla 2.1) y otros 30 segundos a 94°C; seguidos de una extensión final de 10 minutos a 74°C. Los productos de PCR se cargaron en un secuenciador de Applied Biosystems (ABI3730XL) y los genotipos fueron analizados con el programa GeneScan, de la misma compañía.

**Tabla 2.1. Secuencias de los primers y las temperaturas de anillamiento necesarios para amplificar los marcadores microsatélites de la región 15q11-q13.**

Marcador	Primer forward	Primer reverse	Temperatura anillamiento
D15S646	5'_GCTAGATGACGGGTTAGTGG_3'	5'_GCAATAATGCAAAGGACAGG_3'	63°C
D15S817	5'_GGTTATCGAGAGCAATGGAAC_3'	5'_ATGGTACCCAACAAGCATCCT_3'	59°C
D15S1548	5'_AGTGTACCTGTACATC_3'	5'_TATCACAGCACAATTCACAAC_3'	55°C
D15S1549	5'_GGTTTTCTAGAGCAACAGAACGA_3'	5'_AACATGCATCCCAGGATCTC_3'	60°C
D15S1550	5'_TTTGGGTTGTGAGGAGGTTTC_3'	5'_TGGAGACCTAGGAAGCGTGT_3'	57°C
D15S1021	5'_CCTGGCAGGTGGAAGT_3'	5'_ACAAAGATTAACCTCTATGTTTTCG_3'	45°C
D15S128	5'_GCTGTGTGTAAGTGTGTTTTATATC_3'	5'_GCAAGCCAGTGGAGAG_3'	63°C
D15S1551	5'_GCTCCAAAGGTTACGGTCAA_3'	5'_CAAGGACCAACTGGAACA_3'	57°C
D15S1552	5'_TGTTGGCAATAATGTTGTGA_3'	5'_GGGACCCGTGTTGTAAGGTT_3'	60°C
D15S511	5'_AGTCGATTTATCAGCCTCC_3'	5'_CCATGATAGTGGTGAATAG_3'	56°C
D15S1553	5'_TTTTCCAAAGCCATTTTTGC_3'	5'_CATCGCTGATGGATCTTCAA_3'	55°C

### 1.3 Análisis estadístico

La heterocigosidad de los marcadores y si estaban o no en equilibrio de Hardy-Weinberg (HWE) fueron calculados con los programas HET (versión 1.8) y HWE (versión 1.05) respectivamente. Para calcular las frecuencias alélicas en la población autista seleccionamos aleatoriamente 40 pacientes (20 hombres y 20 mujeres sin relación entre ellos) entre las 42 familias objeto de estudio. Utilizamos el programa SPSS 11.0 para analizar las frecuencias alélicas en pacientes y controles mediante el test de la Chi-cuadrado ( $\chi^2$ ) y el programa para genética de poblaciones de la escuela de salud pública de Harvard (versión 1.7.2) para realizar el test de asociación familiar *Family-Based Association Test* (FBAT).

### 1.4 Identificación de la inserción en el gen *GABRG3*

Una inserción polimórfica de 627 bp en el gen *GABRG3* fue amplificada por PCR con los siguientes *primers*: F: 5'-GCTTCTCACATGCAGACCAA-3' (chr15: 25303923-2530943, hg17) y R: 5'-GAGACCTGTGGGACGTCAGT-3' (chr15: 25304259-25304279, hg17). La amplificación se realizó con la enzima *Expand High Fidelity Taq polymerase* (Roche) en un volumen final de 50  $\mu$ l (50ng de DNA, 10 pmol de cada oligonucleótido (Sigma Aldrich), 1x *Expand High Fidelity buffer* con  $MgCl_2$  y 2,6 unidades de la enzima) y en las mismas condiciones que los microsatélites, a excepción del número de ciclos, que en este caso fueron 30 y la temperatura de anillamiento 63°C. Las coordenadas de la inserción son 5788500 y 5789488 de AC\_000058.1 (ensamblaje de Celera).

## 2. Identificación de ncRNAs

### 2.1 Predicción bioinformática de nuevos ncRNAs

Comenzamos haciendo una predicción de todas las estructuras secundarias de 100 nt que se podrían formar con la secuencia de la región chr15:21.483.000-22.477.000, utilizando el programa Mfold versión 3.2 [110, 209] del paquete Viena [293]. Las secuencias corresponden al ensamblaje hg18 de la base de datos UCSC [168]. El análisis con Mfold se realizó de las dos cadenas de DNA independientemente. Las secuencias se eliminaron de la lista de candidatas si: a) la estructura secundaria formada resultaba inestable ( $dG > -30$  Kcal/mol); b) si coincidían con un *Simple Repeat* o cualquier elemento repetitivo del genoma (Alu, SINE, LINE, LTR...), a menos que el *hairpin* predicho coincidiera con dos elementos repetitivos adyacentes y cada mitad del *hairpin* coincidiera con uno de ellos, como han descrito Smalheiser et al [197] que se disponen algunos miRNAs.

### 2.2 Aislamiento de RNA

RNA total de 16 tejidos humanos diferentes (adiposo, vejiga, cerebro, útero, colon, esófago, corazón, pulmón, placenta, próstata, testículo, riñón, hígado, ovario y músculo esquelético) fueron comprados a la compañía Ambion. La fracción de RNAs pequeños (<200 nt) fue aislada del RNA total con *Mirvana™ Isolation Kit* (Ambion) siguiendo las instrucciones del fabricante.

### 2.3 Northern Blot

La expresión de los ncRNAs candidatos previamente seleccionados, fue primero comprobada mediante *Northern Blot*. Se utilizaron sondas de <45 bases (Sigma Aldrich) seleccionadas en una o las dos partes del *hairpin*. Marcamos 200 ng de oligo con 25 $\mu$ Ci de  $\gamma$ -<sup>32</sup>P ATP (Amersham Pharmacia) y 10 unidades de T4 *Polynucleotide Kinase* (Roche) durante 1 hora, a 37°C, en un volumen final de 10  $\mu$ l. Las sondas marcadas fueron purificadas con columnas de Sephadex C-50 (Amersham Pharmacia).

Por otro lado, la fracción de RNAs pequeños, obtenida a partir de 50  $\mu$ g de RNA total, fue mezclada con el mismo volumen de Gel Loading Buffer II (Ambion), desnaturalizada durante 3 minutos a 95°C y separada en un gel de 15% poliacrilamida/7M urea. Realizamos una electroforesis a 25-30 mA hasta que el azul de bromofenol alcanza

la base del gel. El RNA ya separado por tamaño, se transfiere a una membrana de Nylon (Hybond N+, Amersham), mediante electrotransferencia a 300 mA, durante una hora. Después, el RNA se fija a la membrana con UV y tras una incubación de una hora con 15 ml de buffer de pre-hibridación (5X SSPE, 1X Denhardt's, 0.5% SDS y 150 µg/ml de tRNA de levaduras (Roche)), se sustituye este por el de hibridación, de la misma composición que el anterior más la sonda marcada. Se mantiene la hibridación toda la noche a 37°C en agitación. El día siguiente se realizan dos lavados de 10 minutos, a temperatura ambiente, con el buffer de lavado (2x SSPE y 0,1% SDS), se expone de 4-24 horas, dependiendo de la intensidad de la señal, y se revela.

El marcador de tamaño molecular utilizado es *Decade<sup>TM</sup> Marker System* de Ambion.

Las secuencias de las sondas utilizadas para detectar ncRNAs en la región 15q11.2 están especificadas en la Tabla 2.2.

## 2.4 Primer Extension

Para los ensayos de *Primer Extension* seleccionamos *primers* específicos de 17 nt en la secuencia del ncRNA candidato, solapándose con la secuencia de la sonda con la que ya habíamos detectado un ncRNA (si era posible) y/o lo más cerca posible al extremo 5' del transcrito predicho, ya que los transcritos más largos son más difíciles de sintetizar (Tabla 2.2).

Los *primers* fueron marcados con  $\gamma$ -<sup>32</sup>P ATP del mismo modo que las sondas empleadas para *Northern Blot*. Estos se purifican en un gel para separar los oligos de 17 bases, de los posibles productos de degradación, y de los nucleótidos marcados no incorporados. Para esta purificación, se mezclan los *primers* con el mismo volumen de *Gel Loading Buffer II* (Ambion), se desnaturalizan durante 3 minutos a 95°C, se cargan en un gel de 15% poliacrilamida/7M urea y se realiza una electroforesis a 25 mA hasta el azul de bromofenol alcanza la base del gel. A continuación, al gel cubierto con plástico se le hacen unas marcas con un *autoradiography pen* (Sigma) y se expone durante 30 segundos. La película se revela y se coloca sobre el gel haciendo coincidir las marcas hechas previamente con el *autoradiography pen*. Una vez orientados el gel y la película, se corta el área del gel que contiene al *primer* completo con un bisturí y este se deja incubando durante toda la noche a 37°C en 300 µl de tampón de elución (0.5 M NH<sub>4</sub>OAc, 1 mM EDTA, 0.1% SDS). El día siguiente, los *primers* se extraen del buffer de elución mediante extracción con fenol-cloroformo y una precipitación con etanol. Una vez el *primer* está purificado, se mezclan 10 µg de RNA total con ~5000 cpm del *primer* marcado

y se desnaturalizan a 95°C 3 minutos. En un volumen final de 20 µl se mezcla esto con 1mM de cada dNTP, 1x *buffer* de AMV RTase y 25 unidades AMV-RTase (Roche) y se incubaba a 42°C durante 2 horas. Por cada transcrito a testar en cada tejido, se realiza un control- al que se le trata exactamente igual que se ha descrito hasta ahora pero no se le añade enzima. El cDNA resultante se resuelve en un gel de 15% poliacrilamida/7M urea.

**Tabla 2.2. Sondas y primers utilizados para identificar ncRNAs en la región 15q11.2 del genoma humano.**

ncRNAs	Sondas	Primers
RNA5a	5'_CGCCGCCCGCCCTGCCCGTCG_3'	5'_CGCCGCCCGCCCTGCC_3'
RNA5b	5'_GCGCCTTCCGCAGCGACGGGCAGGG_3'	
RNA7	5'_GCGCACTTCCACTAGAGGCTCTGGGGTGA_3'	5'_AGGCTCTGGGGTGACCC_3'
RNA11	5'_GCTACAAGAGCAAACTCCATCCCA_3'	5'_GCTACAAGAGCAAACTC_3' 5'_CTGAGATGGAGTCACTG_3'
RNA21	5'_AGATAGCAAGAAGCAAACC_3'	
RNA23a	5'_AACCTCCTAAACTATGGGAGTTTC_3'	5'_TCCTAAACTATGGGAG_3'
RNA28	5'_TGGTTCTAGGAATGAGCACAG_3' 5'_CTGTGCCCATCTGAAGCCA_3'	
RNA34	5'_CAGCATCTCCCCAGCCCCAACCACTCCCCTCCCC_3'	5'_GCTGCAAGGGTTAGAAC_3'
RNA49a	5'_TGAGGGCAGAGGGTCCCCCTTTGTCCCTCT_3'	5'_TGAGGGCAGAGGGTCCC_3'
RNA49b	5'_CTAGGAGCACTGGCCACCCCTTTGTCTCTCAGCTCT_3'	5'_CTAGGAGCACTGGCCAC_3'
RNA51	5'_AAGAGAGTGCAAGGCCGGAAGCCCCCTCC_3'	5'_GGAGGGCTCACGGCTTG_3'
RNA61	5'_CCATACCCTTCGGCAGCTACC_3' 5'_CCGTGACTACCAAAAGGGTAATAG_3'	
RNA67	5'_AGCATGGCTGTCTTGCAACAAGCCCAAGGATGAGTGGTCATTCA_3' 5'_TGAATCCTTCATCTTATTCTGTCTGCAAAAAGACAGATGCT_3'	
RNA70	5'_GGCATGTGCTCTTCAGTGTGCTGCCACTGCTGC_3'	5'_AGCAGTCCAGCAACCA_3'
RNA78	5'_CCCACGGCAGAGGCCAGCAGGTAC_3' 5'_CCCTCGCACAGACACCTTCCGCAGGGA_3'	
RNA106	5'_ACCTGGCCTCTGCTCCACCCCAACAGCTCC_3'	5'_TCCACCCCAACAGCTCC_3'
RNA195a	5'_GGCCACGCGGTGCAACACCTGCCACCGGC_3'	5'_GGCCACGCGGTGCAACA_3'
RNA195b	5'_CGTGGCCTCTATGAGGGAAGTCTG_3'	
RNA274	5'_CCCTGTCCCCGGGGAGGCTCCTGGGC_3'	5'_CCCTGTCCCCGGGGAG_3'
RNA478	5'_CATATACAACATATGCTCAACCCTCCATGCCATGATCC_3'	5'_CATATGCTCAACCCTCC_3'
RNA514a	5'_GGATCCTATGGCCCATCACTGT_3'	5'_GGATCCTATGGCCCATC_3'

## 2.5 Herramientas bioinformáticas para el análisis de ncRNAs

En primer lugar comprobamos la conservación de todos los ncRNAs candidatos mediante BLAT [294] contra los genomas de chimpancé, macaco, ratón, rata, pollo y vaca (ensamblajes panTro2, rheMac2, mm8, rn4, CanFam2 y bosTau2 respectivamente) disponibles en la base de datos UCSC [168].

Las secuencias fueron analizadas con el software MiRscan (*score* >10) [295] para predecir miRNAs entre los ncRNAs identificados en 15q11.2. En este caso, el análisis también resultó adecuado para predecir la posición del miRNA maduro dentro de la secuencia del pre-miRNA candidato. Para cada predicción utilizamos las correspondientes secuencias ortólogas de los genomas de chimpancé y macaco.

Utilizamos los programas FastR [221] y CMsearch [296] para buscar secuencias homólogas a los ncRNAs identificados entre las familias Rfam [297].

Las predicciones de las estructuras secundarias correspondientes a cada ncRNA fueron calculadas por Mfold [110, 209].

Los programas Randfold [213] y RNAz versión 1.0 [217] fueron utilizados para analizar la estabilidad de las estructuras secundarias de cada ncRNA. Además RNAz ( $P > 0,5$ ) es el software más fiable de los desarrollados hasta ahora para predecir ncRNAs, por lo que sirvió también para soportar los resultados experimentales.

## 2.6 Predicción de *targets* de miRNAs

El software miRanda [289] fue utilizado para escanear las secuencias 3'UTR de todos los transcritos del genoma humano, anotados en la base de datos UCSC (hg17), como posibles *targets* de RNA 61 y RNA274. Debido al elevado número de predicciones generadas por este programa, establecimos los parámetros  $S > 130$  y  $dG < -35 \text{Kcal/mol}$  para filtrar posibles *targets* y minimizar el rango de falsos positivos.



# ***Resultados***





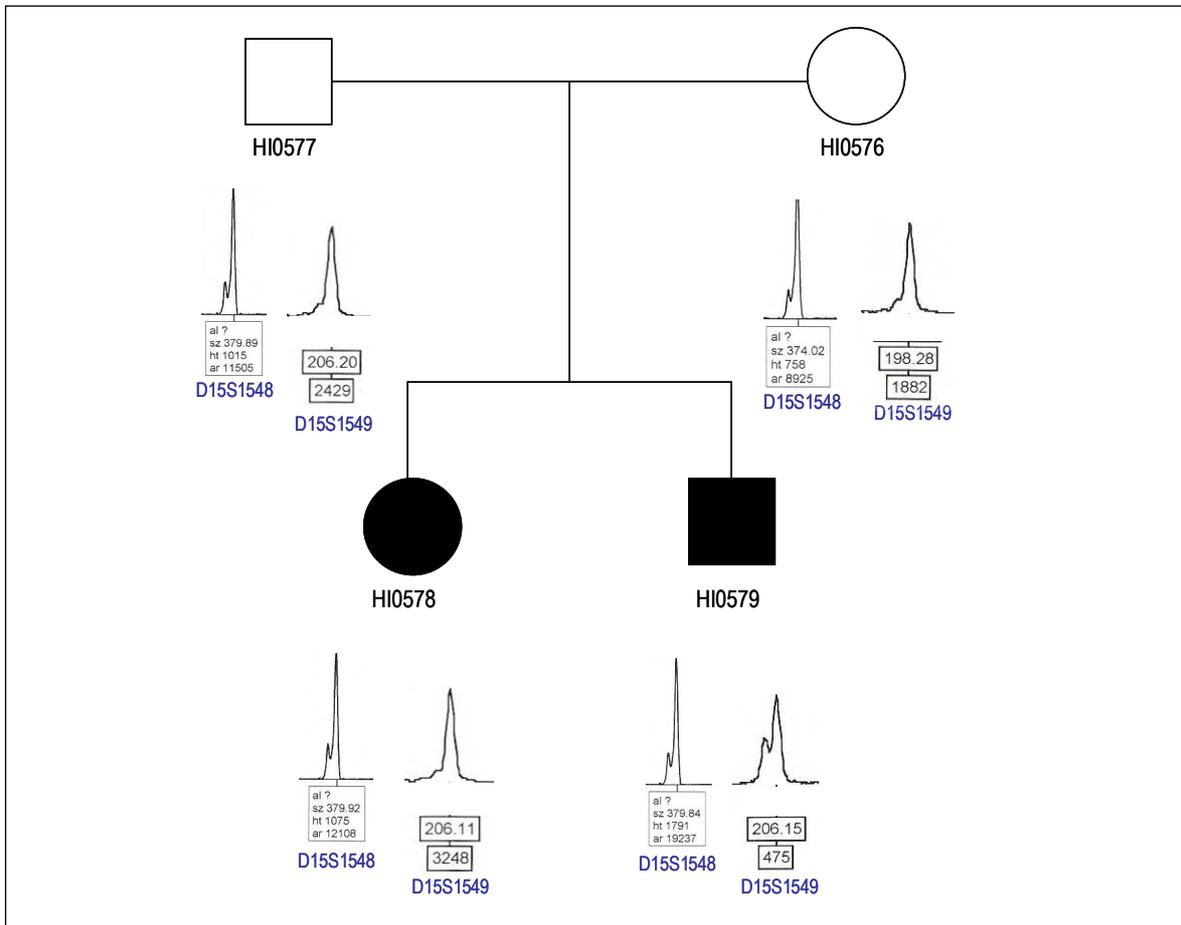
segregaban los alelos, observamos que era más probable que estuviéramos amplificando dos regiones a la vez, a que estos errores correspondieran a reordenamientos, por lo que no consideramos los resultados de este marcador para posteriores análisis.

**Tabla 3.1. HWE y errores mendelianos correspondientes a los microsatélites de 15q11-q13 y la inserción en el gen GABRG3 analizados en muestras de autistas y controles.**

Marcadores	Posición	Nº alelos	Heterocigosidad	Autismo (42 familias)			Controles (62 muestras)			Errores Mendelianos
				Het. Obs.	Het. Esp.	HWE P valor	Het. Obs.	Het. Esp.	HWE P valor	
D15S646	chr15:21900146-21900176	4	65,56%	69,70%	63,22%	0,90	66,67%	65,63%	0,67	0
D15S817	chr15:22155388-22155423	8	83,11%	66,67%	79,67%	0,16	85,71%	83,54%	0,95	10
D15S1548	chr15:22229594-22229622	8	71,43%	48,72%	76,29%	0,00*	57,14%	71,61%	0,00*	23
D15S1549	chr15:22263209-22263267	5	72,88%	47,50%	65,59%	0,00*	54,84%	72,35%	0,00*	16
D15S1550	chr15:22496447-22496478	6	55,09%	54,55%	58,95%	0,08	59,68%	55,02%	0,95	0
D15S1021	chr15:22566926-22566956	8	63,08%	63,64%	67,18%	0,41	68,33%	63,20%	0,99	0
D15S128	chr15:22681957-22682008	7	83,86%	78,13%	78,44%	0,35	86,84%	83,95%	0,62	0
D15S1551	chr15:22866050-22866113	4	55,53%	61,76%	61,47%	0,39	66,10%	54,67%	0,60	0
D15S1552	chr15:23391018-23391079	7	64,27%	42,42%	65,05%	0,10	62,07%	64,37%	0,88	0
D15S511	chr15:24598888-24598934	10	79,14%	78,13%	82,86%	0,25	76,92%	79,60%	0,06	0
D15S1553	chr15:25329587-25329630	8	77,82%	81,82%	79,88%	0,90	73,68%	79,26%	0,15	0
Inserción	chr15: 25303918-25304279	2	50%	33,33%	33,59%	0,96	46,15%	49,62%	0,61	0

El número de errores mendelianos se refiere a la suma de los detectados en familias AGRE y controles. HWE, Equilibrio de Hardy Weinberg; Het. Obs., heterocigosidad observada; Het. Esp., heterocigosidad esperada, \*, P valor estadísticamente significativo.

Seleccionamos dos microsátelites más (D15S1548 y D15S1549) en la misma región de duplicaciones segmentarias (~0,5 Mb) en la que están localizados D15S646 y D15S817. Observamos que ninguno de estos marcadores estaba en HWE en las dos poblaciones analizadas, sugiriendo la presencia de reordenamientos polimórficos que los incluyan. Comprobamos cómo se heredaban los alelos de D15S1548 y detectamos errores mendelianos que indicaban deleciones en 14 familias autistas (20 pacientes) y 3 individuos controles (Tabla 3.2, Figura 3.2). A pesar de que la frecuencia de deleción es mayor en la población autista (16,1%) que en la control (10,7%), las diferencias no resultaron estadísticamente significativas (Tabla 3.2) y tampoco observamos diferencias respecto al origen, materno o paterno, de las deleciones observadas.

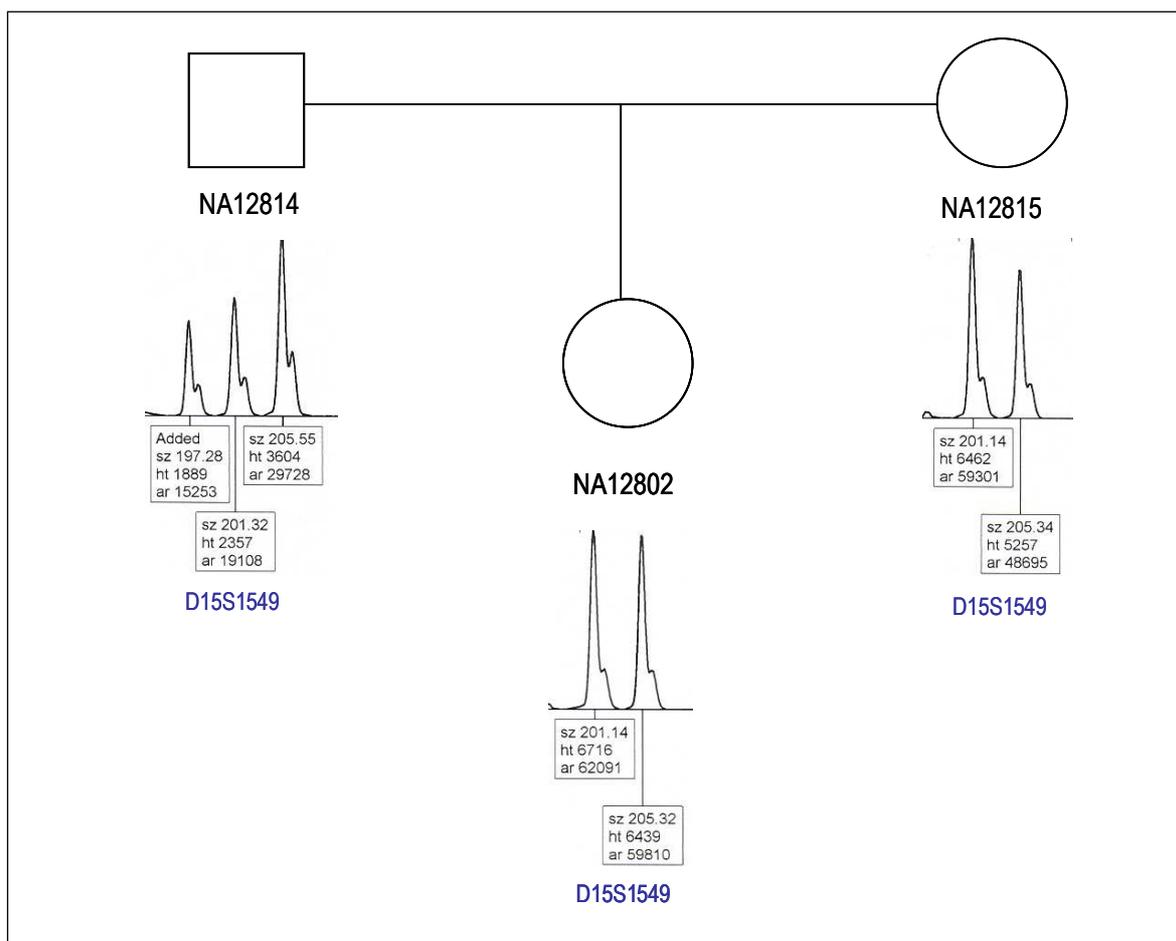


**Figura 3.2. Deleción de los marcadores D15S1548 y D15S1549.**

Los dos hijos autistas (HI0589 y HI0579) presentan una deleción de los dos marcadores heredada de la madre.

Del mismo modo, identificamos deleciones de D15S1549 en 12 pacientes autistas de 9 familias diferentes y en 4 muestras controles. En este caso las diferencias en las frecuencias de deleción tampoco resultaron significativas (Figura 3.2, Tablas 3.1 y 3.2). En una muestra control detectamos tres alelos para D15S1549, lo que indica una duplicación

de dicho marcador (Figura 3.3), que bien podría ser un polimorfismo raro o poco frecuente porque sólo lo observamos en una muestra. De ser así, podría estar interfiriendo en la heterocigosis de D15S1549 y enmascarando otras deleciones. Este caso es una limitación de la técnica empleada, debido a que no podemos cuantificar el número de alelos representados por cada pico.



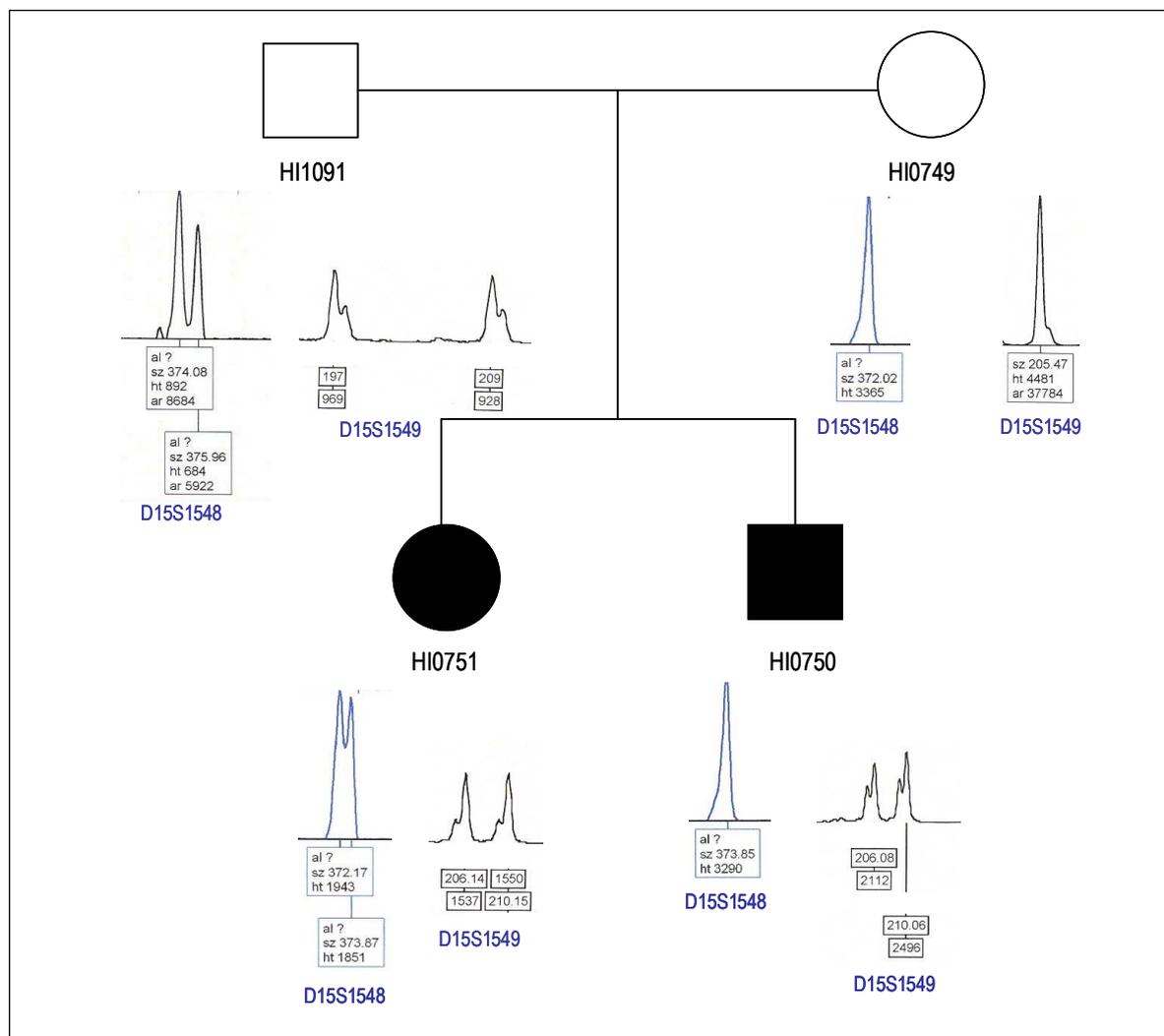
**Figura 3.3. Duplicación del marcador D15S1549.**

El individuo NA12814 (familia CEPH 1454-14) presenta una duplicación de D15S1549 representada por tres alelos.

Once muestras (9 autistas y 2 controles) eran hemocigotos para D15S1548 y D15S1549 y las deleciones de los dos marcadores tenían el mismo origen, indicando la presencia de deleciones más largas, de al menos 33,5 Kb en estos individuos (Figura 3.2). Las diferencias en las frecuencias resultaron de nuevo no significativas. A pesar de la elevada heterocigosis calculada para los dos marcadores, aproximadamente el 30% de las familias no eran informativas para al menos uno de ellos, por lo que la frecuencia de deleciones más largas, que incluyen D15S1548 y D15S1549, podría ser más alta de lo

que hemos estimado. Entendemos por familia no informativa para un marcador, aquella en la que los dos progenitores son homocigotos para el mismo alelo.

En varios casos observamos hemicigosis de D15S1548 y heterocigosis de D15S1549 (Figura 3.4) demostrando que las deleciones identificadas en esta región son variables en tamaño entre diferentes individuos. Sin embargo, no encontramos ningún caso que tuviera delecionado el marcador D15S1549 y dos alelos diferentes para D15S1548, lo que sugiere que deleciones de D15S1549 podrían ser parte de deleciones más largas que incluyen ambos marcadores.



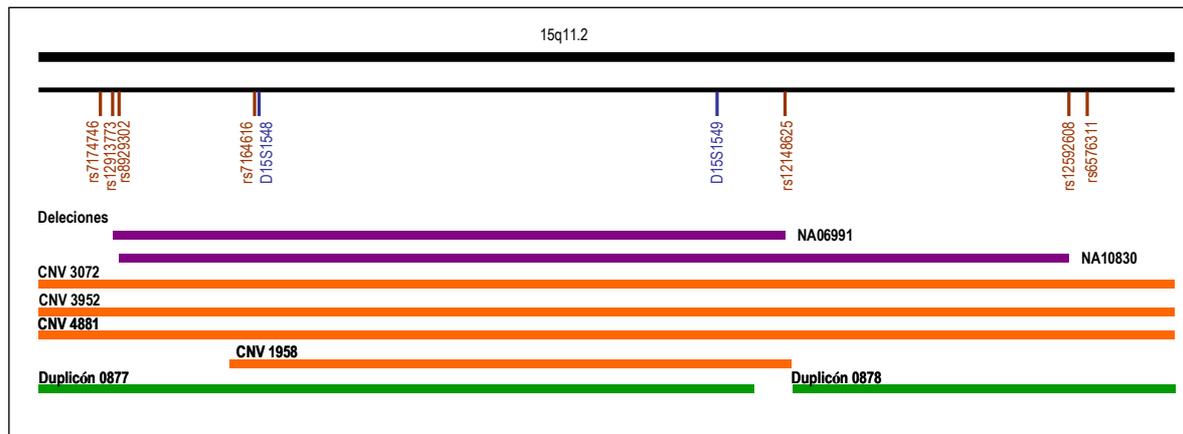
**Figura 3.4. Delección del marcador D15S1548.**

El autista HI0750 es heterocigoto para D15S1549 y hemicigoto para el alelo paterno de D15S1548, teniendo una delección heredada de la madre de dicho marcador. La delección no ha sido heredada por su hermana (HI0751).

**Tabla 3.2. Frecuencias de las deleciones de los marcadores D15S1548 y D15S1549 en pacientes autistas y controles CEU seleccionados aleatoriamente.**

Marcador	Origen	Autismo	CEPH	X <sup>2</sup>	P valor
<b>D15S1548</b>		16,12%	10,7%	0,46	0,50
	Materno	6,45%	3,57%	0,31	0,58
	Paterno	8,06%	7,14%	0,15	0,70
<b>D15S1549</b>		12,5%	10%	0,12	0,73
	Materno	3,57%	3,33%	0,00	0,95
	Paterno	8,92%	6,66%	0,13	0,71
<b>D15S1548- D15S1549</b>		6,94%	5%	0,36	0,55
	Materno	2,78%	0%	1,31	0,25
	Paterno	4,17%	5%	0,00	0,97

Con el fin de caracterizar mejor estos reordenamientos, comparamos nuestros resultados con los procedentes de genotipar SNPs de la misma región, en las muestras controles CEU que utilizamos en nuestro estudio. La existencia de deleciones sólo del marcador D15S1548 fue confirmada al identificar una muestra (NA06991) hemicigota para D15S1548, homocigota no informativa para D15S1549 y heterocigota para rs12148625, un SNP separado de D15S1549 sólo por 865 bp (Figura 3.5). El tamaño de la deleción de este individuo es inferior a 44 Kb ya que es heterocigoto para los SNPs rs12913773 y rs12148625. Curiosamente, rs7164616 localizado 407 bp hacia el centrómero desde la posición de D15S1548, tampoco está en HWE ( $P < 0,005$ ) en las familias controles CEU, soportando la presencia de reordenamientos polimórficos en esta región. Deleciones más largas que incluyen D15S1548 y D15S1549 también fueron comprobadas con los resultados de SNPs en el individuo NA10830. Este tiene una deleción paterna de D15S1548 y D15S1549 y es hemicigoto para rs12148625. La deleción esta limitada por rs8029302 y rs12592608 y es inferior a 62 Kb (Figura 3.5). El otro individuo control con una deleción larga no es informativo para ningún SNP en el intervalo de 137 Kb analizado alrededor de D15S1548 y D15S1549.



**Figura 3.5. Amplificación de la región delecionada (chr15:22.197.281-22.297.280 (hg18)).**

En azul están marcados los microsatélites delecionados (D15S1548 y D15S1549) y en rojo los SNPs del proyecto HapMap, que resultaron informativos para caracterizar las deleciones de dos individuos controles CEU (NA06991 y NA10830). Las regiones delecionadas en cada uno de ellos están representadas en malva, los CNVs identificados por Redon et al. (CNV 3072 y CNV 3952), McCarroll et al. (CNV 1958) y Wong et al. (CNV 4881) en naranja y las duplicaciones segmentarias en verde.

## 1.2 Asociación de las deleciones identificadas en 15q11.2 con el trastorno autista

Comprobamos si las deleciones polimórficas identificadas en 15q11.2 estaban asociadas con el trastorno autista mediante el test *Family-Based Association Test* (FBAT), más adecuado para estudios de asociación que el test de la  $\chi^2$ . Para este test consideramos sólo las familias informativas para la ausencia o presencia de deleciones, es decir, aquellas en las que los progenitores tenían alelos de diferentes tamaños. Como consecuencia de la reducción en el número de familias para considerar, genotipamos diez adicionales sólo para los marcadores en los que observamos reordenamientos, y así conseguimos incrementar el poder estadístico del test. Analizamos las deleciones de D15S1548 y D15S1549 por separado, así como las que incluyen a los dos marcadores (Tabla 3.3). Según los resultados de FBAT, existe una transferencia del alelo D15S1548 delecionado superior a la esperada en las familias autistas, que resultó estadísticamente significativa usando cualquiera de los modelos del test: aditivo, dominante, recesivo o de genotipo. También observamos que las deleciones largas y las de D15S1549 parecen transmitirse con mayor frecuencia a la esperada en las familias autistas pero los valores no alcanzan significación estadística.

**Tabla 3.3. FBAT de las deleciones de D15S1548 y D15S1549 en las familias autistas.**

Marcador	Modelo	Alelo Genotipo	Frecuencia	Familias	S	E (S)	Var (S)	Z	P
<b>D15S1548</b>	Aditivo	1	0,911	16	35,0	43,5	7,25	-3,157	0,00**
		2	0,089	16	23,0	14,5	7,25	3,157	0,00**
	Dominante	1	0,911	0	*****	*****			
		2	0,089	16	23,0	14,5	7,25	3,157	0,00**
	Recesivo	1	0,911	16	6,0	14,5	7,25	-3,157	0,00**
		2	0,089	0	*****	*****			
	Genotipo	11	0,821	16	6,0	14,5	7,25	-3,157	0,00**
		12	0,179	16	23,0	14,5	7,25	-3,157	0,00**
		22	0	0	*****				
<b>D15S1549</b>	Aditivo	1	0,931	11	28,0	32,0	6,0	-1,63	0,1
		2	0,069	11	16,0	12,0	6,0	-1,63	0,1
	Dominante	1	0,931	1	*****				
		2	0,069	11	16,0	11,5	5,38	1,94	0,052
	Recesivo	1	0,931	11	6,0	10,5	5,38	-1,94	0,052
		2	0,069	1	*****				
	Genotipo	11	0,862	11	6,00	10,5	5,38	-1,94	0,052
		12	0,138	11	16,00	11,0	5,5	2,13	0,033
		22	0	0	*****				
<b>D15S1548- D15S1549</b>	Aditivo	1	0,956	8	20,0	24,0	4,0	-2,0	0,045
		2	0,044	8	12,0	8,0	4,0	2,0	0,045
	Dominante	1	0,956	0	*****				
		2	0,044	8	12,0	8,0	4,0	2,0	0,045
	Recesivo	1	0,956	8	20,0	24,0	4,0	-2,0	0,045
		2	0,044	0	*****				
	Genotipo	11	0,911	8	4,0	8,0	4,0	-2,0	0,045
		12	0,089	8	12,0	8,0	4,0	2,0	0,045
		22	0	0					

Análisis de las deleciones observadas de 15q11.2 en 18 familias de autistas informativas. Alelo "1" es el alelo no delecionado y el alelo "2" se corresponde con el delecionado. S, Transmisión del genotipo a la descendencia observada; E, Transmisión del genotipo esperada según el modelo de herencia mendeliano; Var, varianza; P, P valor; \*\* P valor significativo después de la corrección de Bonferroni ( $P < 0.002$ ).

### 1.3 Ausencia de reordenamientos en la parte más telomérica de 15q11-q13

Continuamos con el análisis de reordenamientos de 15q11-q13 en autistas genotipando siete marcadores más, situados en sitios estratégicos, a lo largo de la región, coincidiendo o lo más cerca posible a CNVs, duplicaciones segmentarias y genes candidatos para explicar la etiología del autismo (Figura 3.1).

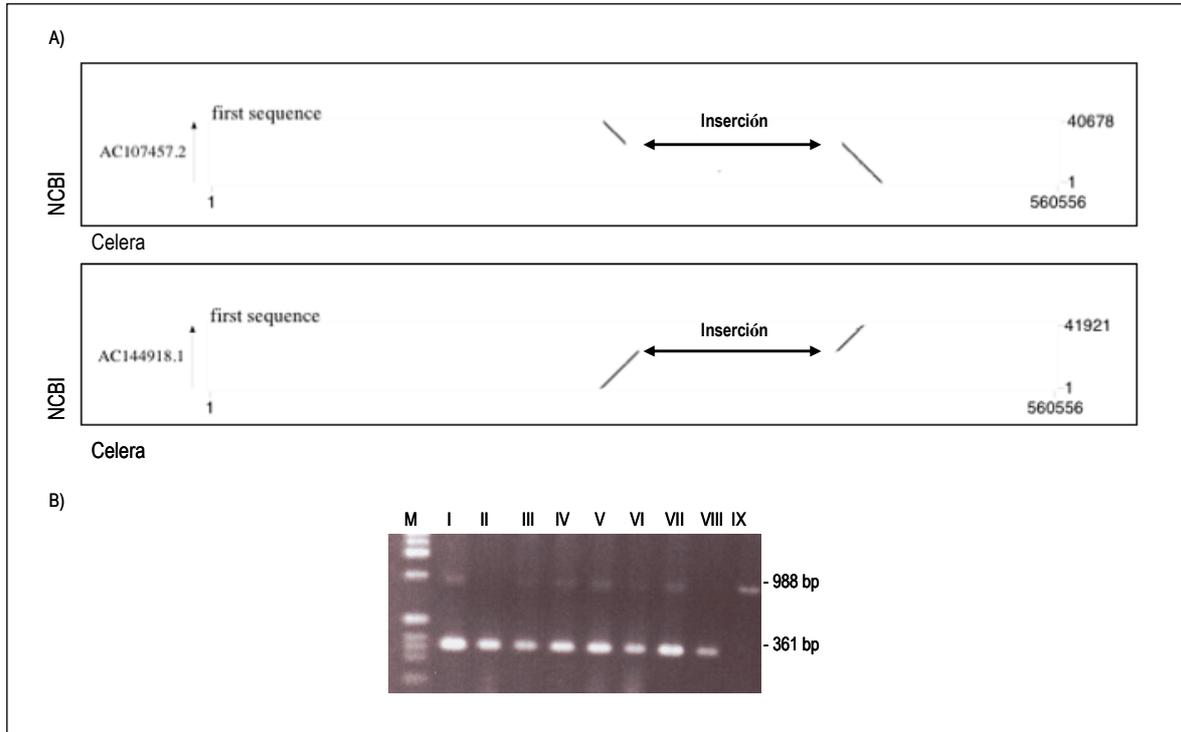
Seleccionamos cuatro microsatélites más en 15q11.2 (D15S1550, D15S1021, D15S128 y D15S1551) (Figura 3.1). D15S128 y D15S1551 coinciden con la posición de CNV [55, 82] y además con los genes SNURF/SNRPN y el *cluster* de snoRNAs HBII-85 respectivamente. Sin embargo, no encontramos errores mendelianos ni evidencias de ellos en ninguno de los cuatro marcadores y en ninguna de las dos poblaciones analizadas.

Genotipamos también el marcador D15S1552 que coincide con CNVs [58, 82] y situado entre los genes *UBE3A* y *ATP10A*, genes expresados en el alelo materno y asociados a autismo [114], pero tampoco observamos ningún reordenamiento.

Finalmente, buscamos variaciones en el número de copias, en la región donde está codificado un *cluster* de subunidades del receptor de GABA (*GABRB3*, *GABRA5* y *GABRG3*) que a la vez son genes asociados a autismo [107, 112, 113, 176, 294]. Analizamos D15S511 situado cerca del *gap* de secuencia que existe en la subunidad *GABRA5* y D15S1553 próximo al gen *GABRG3* y en un CNV, pero no observamos reordenamientos en este locus.

### 1.4 Inserción en el gen *GABRG3*

Comparando la secuencia del gen *GABRG3* según los ensamblajes de NCBI (NC\_000015.8) y Celera (AC\_000058.1), detectamos *in silico* una inserción de 627 bp en el sexto intrón de dicho gen (Figura 3.6), coincidiendo además con un *gap* de 100 Kb. Confirmamos experimentalmente la presencia de esta inserción mediante PCR y secuenciación en nuestras muestras controles y de familias de autistas. Observamos que la frecuencia de la inserción estaba en HWE y era similar en ambas poblaciones (Tabla 3.1). Así, nuestros datos demuestran que se trata de un polimorfismo que no parece estar asociado con el desarrollo del autismo.



**Figura 3.6. Inserción en el sexto intrón del gen GABRG3.**

A) Pipmaker de las secuencias de dos BACs (AC107457.2 y AC144918.1) contra la secuencia de la misma región según Celera. Ambos dibujos reflejan la inserción considerando como referencia la secuencia de Celera. B) Comprobación experimental de la inserción mediante PCR. Se observan individuos heterocigotos (I, III, IV, V, VI, VII), homocigotos para la inserción (IX) y homocigotos para el alelo pequeño (II y VIII).

## **2. Identificación de ncRNAs específicos de primates en la región cromosómica 15q11.2 del genoma humano.**

### **2.1 Selección de secuencias y estructuras secundarias en la región cromosómica 15q11.2 que podrían corresponder a nuevos ncRNAs.**

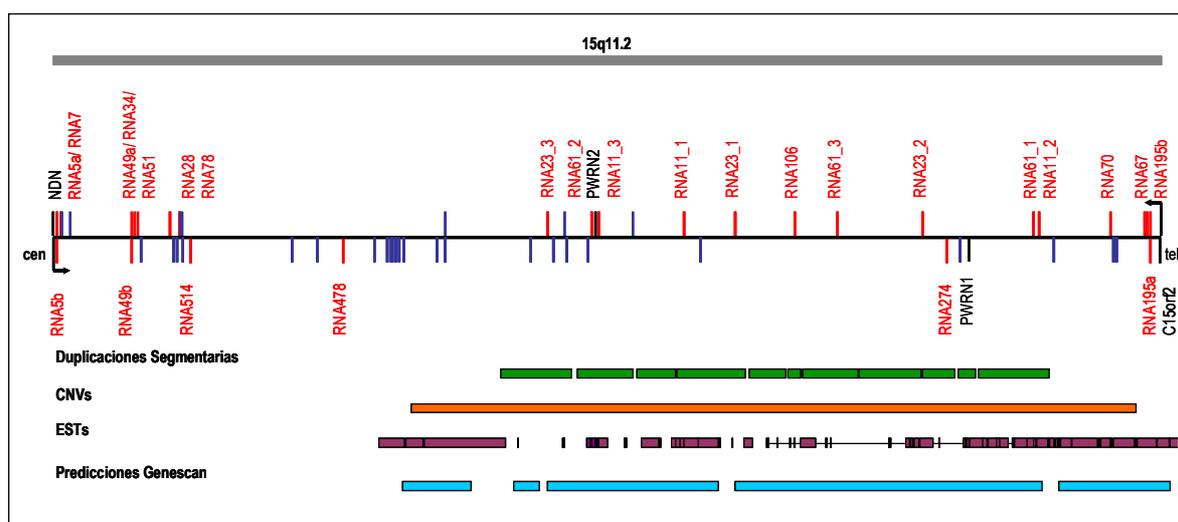
Diferentes clases de ncRNAs, como snoRNAs, snRNAs, miRNAs y piRNAs se han identificado durante los últimos años, pero se estima que sólo una parte de ellos se conocen y de los identificados hasta ahora, pocos han sido parcialmente caracterizados en cuanto a función y patrones de expresión [27, 188, 189, 216, 298].

Consideramos que la región cromosómica 15q11.2 reunía condiciones que la convertían en una región candidata para codificar ncRNAs. Entre ellas destacan la presencia de pequeñas repeticiones en tandem (Figura 1.10), un bloque de duplicaciones segmentarias y CNVs [55, 59, 66] que contribuyen a la inestabilidad de la región y genes improntados alrededor (Figura 3.7) [15, 72, 299]. Comenzamos su identificación buscando estructuras secundarias estables ( $\Delta G < -30$  Kcal/mol) de 100 nt a lo largo de aproximadamente 1 Mb (chr15:21.483.000-22.477.000 (hg18)), mediante el programa Mfold [110, 209]. Así, entre  $\sim 2 \cdot 10^6$  secuencias iniciales obtuvimos 246 candidatas. Como una característica común entre los ncRNAs es su conservación a lo largo de la evolución, comprobamos si estas 246 secuencias estaban conservadas en los genomas de chimpancé, macaco, ratón, rata, pollo y vaca y observamos que la mayor parte de ellas eran específicas de primates, por lo que el tan usado parámetro de la conservación para identificar ncRNAs, no resultó útil para nuestro estudio. Entre estas 246 secuencias eliminamos las que coincidían con elementos repetitivos (LINEs, SINEs, Alu, *Simple repeats*...) y nos quedaron 86 candidatos. Cuatro coincidían en la misma posición pero se transcribirían en sentido contrario y las llamamos con el mismo nombre terminado en a o b.

Entre los 86 candidatos seleccionamos 27 para comprobar su expresión experimentalmente: los 20 que formaban estructuras secundarias más estables; los dos únicos candidatos que coincidían con dos elementos repetitivos adyacentes, una disposición que había observado Smalheiser et al. [197] en algunos miRNA y 5 candidatos más que estarían codificados en los extremos 5'UTR de los genes *Necdin* y *C15orf2*. Incluimos estos 5 candidatos porque 1041 ncRNAs registrados en la base de datos fRNAdb están codificados en las regiones 5'UTR de otros genes [178].

## 2.2 Expresión y caracterización de nuevos ncRNAs

En primer lugar comprobamos la expresión de los 27 ncRNAs candidatos previamente seleccionado, mediante *Northern Blot*, detectando 21 de ellos (77,8% de los seleccionados) en tejidos de cerebro, hígado, riñón y testículo. Los niveles de expresión de estos 21 ncRNAs son muy bajos, ya que no fue posible detectarlos utilizando menos de 50  $\mu\text{g}$  de RNA total, mientras que normalmente es suficiente con 20  $\mu\text{g}$  [300, 301]. La mayor parte de ellos se transcribían a partir de la cadena (-) del DNA (Figura 3.7). Los ncRNAs RNA61, RNA11\_2 y RNA70 coinciden con la posición de intrones del gen *PWRN1* pero se transcriben en sentido contrario. RNA 61\_2 se solapa con un exón de *PWRN2* y se podría producir como consecuencia de diferentes procesos de *splicing* del gen [194]. De todas formas RNA61\_2 ha sido identificado por homología con RNA61\_1 y no podemos asegurar que se transcriba, correspondiendo en este caso la transcripción a *PWRN2*.



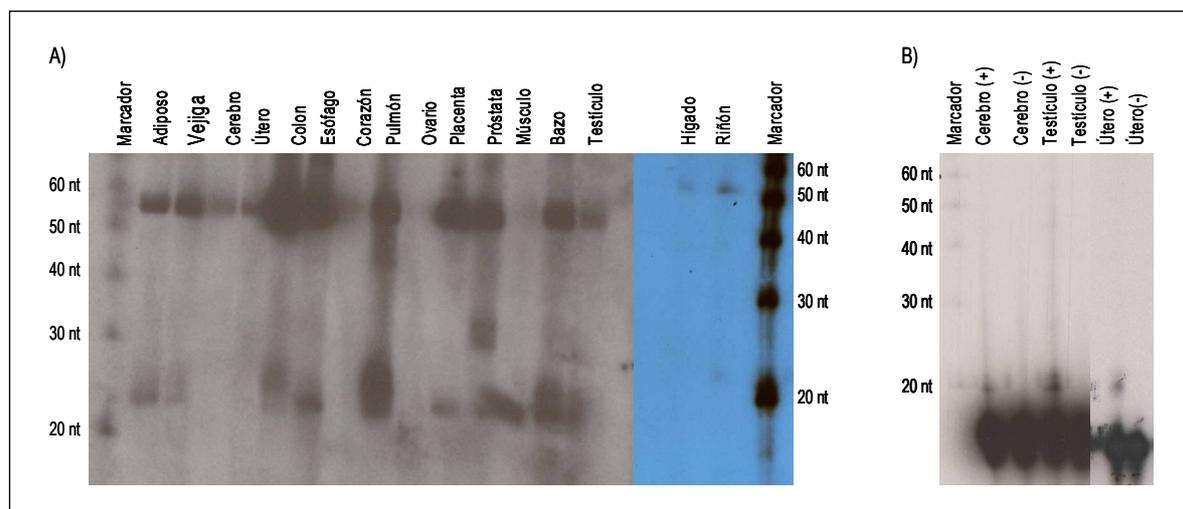
**Figura 3.7. Disposición de los ncRNAs identificados en la región cromosómica 15q11.2.**

Sobre la línea que representa al cromosoma están representados los genes que se transcriben desde la cadena (-) del DNA (de telómero a centrómero) y debajo de ella los codificados en la cadena (+) del DNA (de centrómero a telómero). Las barras azules corresponden a piRNAs, las negras a genes codificantes (NDN, *PWRN1* y *PWRN2* y *C15orf2*) y las rojas a los ncRNAs identificados en este estudio.

Entre los 6 candidatos que no detectamos mediante *Northern blot* se encuentran dos de los que estaban localizados en la misma posición que otro candidato, pero se expresaban en sentido contrario. Según los resultados de Randfold [213], eran más estables las estructuras secundarias formadas por los candidatos que se expresan.

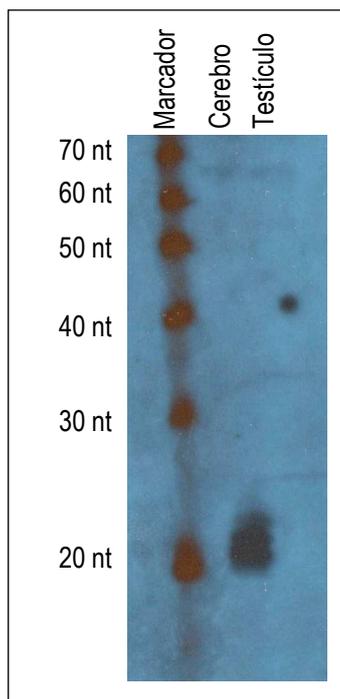
Observamos bandas del tamaño de un miRNA maduro en los *blots* correspondientes a los candidatos RNA61 y RNA274 (Tabla 3.4, Figuras 7R y 8R) y

comprobamos sus patrones de expresión con la misma técnica utilizando 16 tejidos diferentes: adiposo, vejiga, cerebro, útero, colon, esófago, corazón, pulmón, placenta, próstata, bazo, testículo, hígado, riñón, ovario y músculo esquelético. Detectamos los precursores de ambos miRNAs en todos los tejidos excepto en ovario y músculo esquelético, con un tamaño de 63nt el pre-miRNA RNA61 y de 53 nt el pre-miRNA RNA274. Sin embargo, la forma madura de RNA274 no se apreciaba ni en útero ni en corazón y el RNA61 maduro sólo se observaba en cerebro (Figura 3.8 y 9R). Para comprobar si la ausencia de banda correspondiente a las formas maduras se debía a los límites de detección de la técnica, realizamos experimentos de *Primer Extension*. Utilizamos RNAs de tejidos de cerebro, testículo y útero para comprobar la expresión RNA274 y así vimos la forma madura también en útero, por lo que probablemente la ausencia de banda correspondiente a la forma madura de RNA274 en algunos tejidos, se deba a que sus niveles de expresión son más bajos a los límites de detección del *Northern*. Además, observando con detalle el *blot*, vemos que los tejidos donde la banda del precursor es más débil son los mismos que en los que no se detectan las formas maduras, corroborando el resultado anterior. En el caso de RNA61 no fue posible desarrollar con éxito los experimentos de *Primer Extension* debido a la ausencia de *primers* específicos en la secuencia del precursor. Es de esperar que al igual que RNA274, RNA61 es procesado en los tejidos donde se expresa su precursor y la ausencia de banda se debe a su bajo nivel de expresión.



**Figura 3.8. Expresión del miRNA RNA274.**

A) *Blot* mostrando el patrón de expresión de RNA274 en 16 tejidos diferentes. B) Comprobación de la expresión del miRNA maduro en cerebro, testículo y útero por *Primer extension*.



**Figura 3.9. Expresión del miRNA RNA61.**

El *blot* muestra el pre-miRNA (63 nt) en los tejidos de cerebro y testículo mientras que la forma madura (21 nt) sólo se aprecia en cerebro

Según el patrón de banda observado para los demás ncRNAs identificados, se pueden diferenciar dos tipos: el primero lo forman los ncRNAs para los que sólo hemos observado una banda con tamaños comprendidos entre 61 y 93 nt (RNA21, RNA28, RNA67, RNA106, RNA195a, RNA195b and RNA478) (Tabla 3.4 y Figura S1) y el segundo lo forman aquellos ncRNAs para los que hemos detectado más de una banda (RNA5a, RNA5b, RNA7, RNA11, RNA23, RNA34, RNA49a, RNA49b, RNA51, RNA78 y RNA514) (Tabla 3.4 y Figura S2).

Debido a que este patrón de bandas observado no corresponde a ninguna clase de ncRNAs conocidos hasta ahora, también comprobamos la expresión de estos ncRNAs mediante *Primer Extension* que además aporta información sobre la posición del extremo 5' de los transcritos.

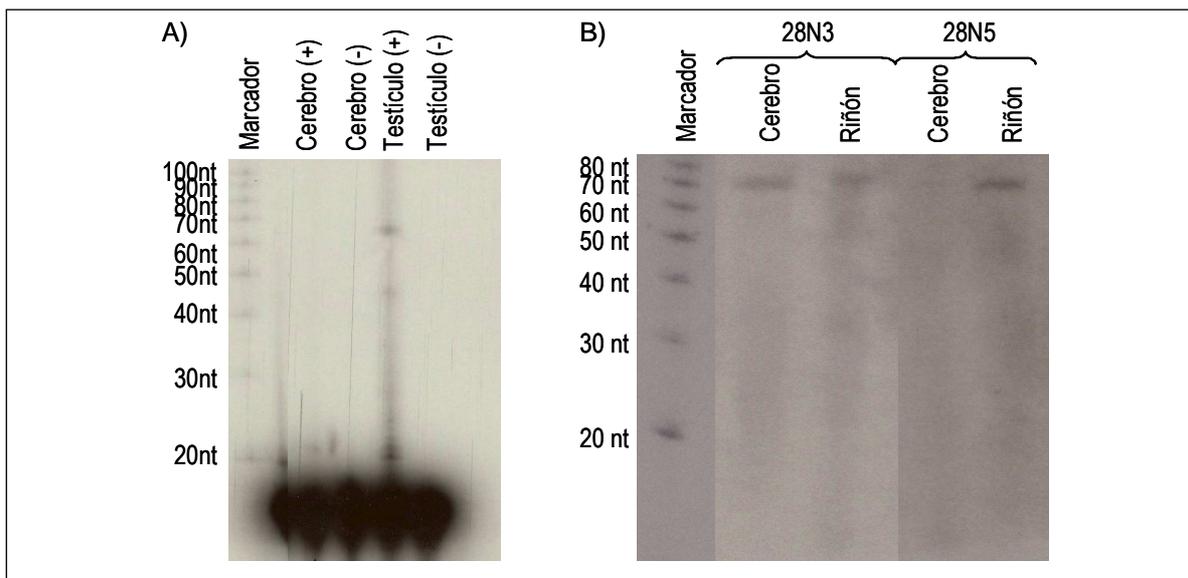
Para RNA49a detectamos dos bandas por *Primer Extension* (Figura 3.10a), de 22 y 26 nt, que podrían corresponder a cualquiera de las observadas en el *blot* (28 y 45 nt).

En el caso de RNA49b, los dos transcritos observados en el *blot* de 39 y 89 nt, fueron detectados por *Primer Extension* como bandas de 22 y 42 nt respectivamente. Además con esta última técnica observamos una banda adicional de 63 nt que podría ser el extremo 5' del transcrito primario.

El tamaño de banda observada para RNA70 fue de 67 nt, el mismo que tenía su correspondiente banda en el *blot*. Así identificamos los extremos 3' y 5' del ncRNA.

El transcrito sintetizado para RNA106 mediante *Primer Extension* era de 55 nt y corresponde con el extremo 5' del ncRNA de 70 nt detectado por *Northern Blot*.

Comprobamos sin éxito con esta técnica la expresión de RNA23, posiblemente porque el único primer específico que encontramos en su secuencia estaba en su extremo 3' y la transcripción reversa no funcionó o el nivel de transcrito sintetizado estaba por debajo de los límites de detección. Tampoco fue posible detectar RNA5a, RNA7, RNA11, RNA34, RNA51, RNA61, RNA78, RNA195a, RNA478 y RNA514 mediante transcripción reversa, en estos casos posiblemente por la falta de especificidad de los *primers*. Por este motivo no intentamos esta técnica con los demás ncRNAs para los que no encontramos *primers* específicos. La expresión de RNA28, RNA61, RNA67 y RNA78 fue comprobada por *Northern Blot* utilizando dos sondas diferentes (Figura 3.10b).



**Figura 3.10. Comprobación de la expresión de ncRNAs.**

A) *Primer extension* para detectar RNA49b. B) *Blot* de RNA28 con dos sondas diferentes: 28N5 y 28N3 diseñada en los lados 5' y 3' del *hairpin*.

**Tabla 3.4. Validación experimental y caracterización bioinformática de los ncRNAs identificados en la región 15q11.2.**

ncRNA	Northern Blot Tamaño de la banda (nt)	Primer extension Tamaño de la banda [transcrito] (nt)	MiRscan score	RandFold P valor	RNAz		
					z-score	SCI	P valor
RNA5a	40, 60, 70, 90	0		0.58	2,23	0,95	0,13
RNA5b	83,100	****		0.72	0,28	0,98	0,03
RNA7	50, 62, 72, 90, 99	0		0.13	-1,44	0,91	0,34
RNA11	28, 56,100,120	0		0.00	-4,63	1	0,99
RNA21	63	****	5,79	0.00	-7,36	0,99	0,92
RNA23	50, 106	0		0.00	-11,84	0,96	0,51
RNA28	75*	****	12,25	0.00	-7,01	1	0,79
RNA34	28, 45, 60, 82	0		0.00	-3,05	0,95	0,98
RNA49a	28, 45	22, 26[28 o 45], 76[?]		0.00	-2,41	0,86	0,83
RNA49b	39, 89	22[39], 42[89], 63[?]		0.00	-3,42	0,9	0,98
RNA51	41, 60	0		0.00	-3,07	0,79	0,94
RNA61	22, 63*	0		0.00	-3,78	0,98	0,99
RNA67	93*	****	0	0.00	-2,82	0,96	0,98
RNA70	67	67[67]	14,34	0.36	-1,41	1	0,72
RNA78	65, 90, 100*	0		0.34	-1,15	0,57	0,06
RNA106	70	55[70]	-19	0.25	-1,32	0,93	0,57
RNA195a	89	0	6,39	0.26	-0,45	0,78	0,08
RNA195b	61	****	7,43	0.74	-0,65	0,97	0,09
RNA274	21, 53	21[21]		0.00	-1,79	1	0,91
RNA478	74	0	10	0.00	-5,94	0,98	0,91
RNA514	50, 72, 100	0		0.00	-3,06	0,85	0,91

\*, ncRNA detectados mediante *Northern Blot* utilizando dos sondas diferentes; los números entre corchetes se refieren al tamaño del transcrito al que corresponde la banda observada con *Primer Extensión*; \*\*\*\*, ncRNAs para los que no hemos realizado transcripción reversa.

### 2.3 Definición de las secuencias de los ncRNAs identificados en 15q11.2

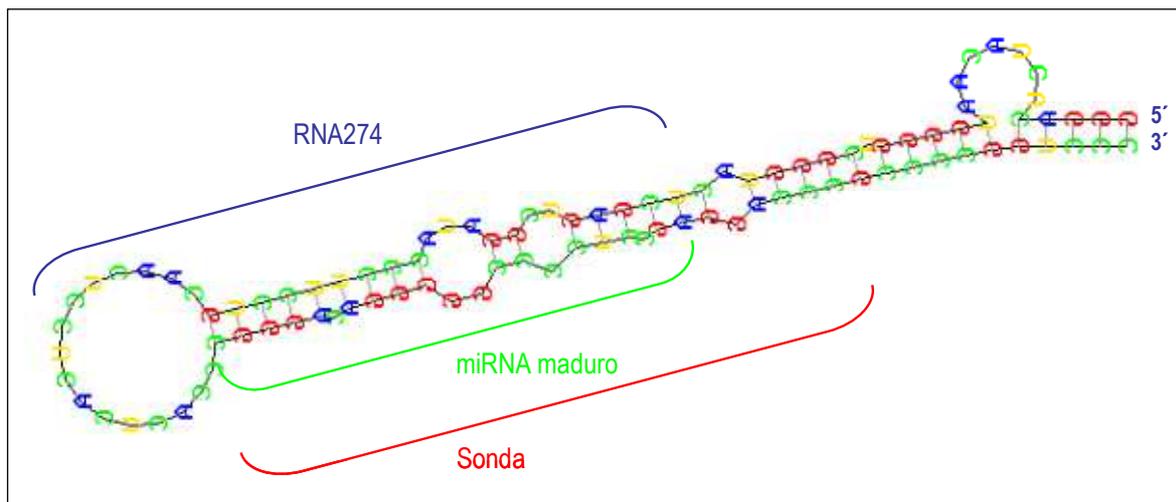
En este punto de la caracterización de los ncRNAs identificados en 15q11.2, compilamos la información que teníamos sobre ellos (tamaño real, estructura secundaria y extremo 5' en algunos casos) para limitar qué bases del primer ncRNA predicho de 100 nt, corresponden al ncRNA que se expresa y con estructura secundaria estable. Eliminamos base por base desde los extremos 5' y 3' del *hairpin* de 100 nt, hasta conseguir la estructura más estable, del tamaño adecuado y que incluía las sondas y *primers* utilizados para la validación de su expresión (Figura 3.11). En los casos en los

que se detectaron más de una banda, analizamos la estructura secundaria que le correspondería a cada uno de los transcritos observados, mayores de 50 bases, y consideramos la más estable para el resto del estudio. En todos los casos esta resultó ser la de mayor tamaño, a excepción de RNA11 en el que la secuencia que al plegarse puede formar un *hairpin* es la de 56 nt. La energía libre de todas estas formas también fue calculada por Mfold [110, 209]. Las secuencias están descritas en la Tabla S1 y sus estructuras secundarias finales en la Figura S3.

En el caso de RNA49a como los transcritos detectados son pequeños para formar una estructura secundaria estable, nos basamos en los resultados del estudio bioinformático explicado en el siguiente apartado para predecir la secuencia de su posible precursor, aunque este no haya sido detectado experimentalmente.

Suponemos que los transcritos más pequeños son producto del procesado de los *hairpins*, pero con la información disponible no es posible predecir sus secuencias, ni identificar cuál es la forma funcional. Las bandas mayores de 56 nt detectadas con la sonda para RNA11 deben ser sus precursores pero tampoco se pueden predecir las secuencias de estos, hasta que no se conozcan señales características de ncRNAs en la secuencia primaria o se tengan más detalles sobre su transcripción.

Los dos miRNAs (RNA61 y RNA274) fueron además analizados con MiRscan [177], un software que se ha utilizado para filtrar secuencias que podrían corresponder a miRNAs [300] y que además hace una predicción de la posición de la forma madura en la secuencia del *hairpin*. La limitación de este programa en nuestro estudio es que para filtrar miRNAs candidatos ( $score > 10$ ) [300] se basa en la conservación a lo largo de la evolución y los miRNAs identificados en 15q11.2 son específicos de primates, por lo que en este caso sólo consideraremos la predicción del miRNA maduro. Para el análisis con MiRscan de RNA61 utilizamos sus correspondientes secuencias ortólogas en los genomas de chimpancé y macaco. El miRNA maduro predicho por MiRscan se solapa con la sonda con la que lo detectamos y está localizado en el extremo 5' del pre-miRNA. Sin embargo, las predicciones de MiRscan para RNA274 no fueron reproducibles, posiblemente debido a que sólo está conservado en el genoma de chimpancé. Observando su estructura secundaria, vemos que eliminando las bases que forman el *loop* final, el lado 5' del *hairpin* mide 21 nt, el tamaño del miRNA maduro, y coincide con la región a la que se unirían la sonda y el *primer* utilizado para su detección. Así deducimos que estos 21 nt forman la secuencia del miRNA maduro (Figura 3.11).



**Figura 3.11. Definición de las secuencia del pre-miRNA y miRNA maduro RNA274.**

El dibujo se refiere al *hairpin* predicho para el primer candidato RNA274 de 100nt. En rojo está señalada la región donde se uniría la sonda utilizada para su detección mediante *Northern Blot*. Como el tamaño observado del pre-miRNA es 53 nt, la región azul es la parte del primer *hairpin* que corresponde al pre-miRNA. Finalmente, como la forma madura está en el extremo 5' y mide 21 nt, su secuencia la componen las 21 bases que van desde el *loop* hasta el extremo 5' del dicho pre-miRNA (verde).

## 2.4 Análisis bioinformático de los nuevos ncRNAs

Una vez comprobada la expresión de los nuevos ncRNAs y definidas sus secuencias, realizamos un análisis bioinformático de las mismas con el fin de encontrar evidencias adicionales que soporten los resultados experimentales y caracterizar mejor los ncRNAs de 15q11.2.

Consideramos la posibilidad de que los ncRNAs para los que sólo habíamos detectado una banda de tamaños entre 61 y 93 nt fueran pre-miRNAs, para los que no habíamos detectado la forma madura, y analizamos éstas secuencias con MiRscan. La sensibilidad de este programa considerando un *score* >10 es del 75% [302], y teniendo en cuenta que para asignar este *score* se basa en la conservación, resulta fácil entender que miRNAs como RNA61 y RNA274 no pasen este filtro. Por el contrario, puede ser una herramienta útil para seleccionar candidatos. De hecho tres de los ncRNAs identificados en 15q11.2 (RNA28, RNA70 y RNA478) podrían ser miRNAs según las predicciones del programa MiRscan, a pesar de que la forma madura correspondiente no han sido detectadas experimentalmente (Tabla 3.4).

También analizamos la estabilidad de las estructuras secundarias formadas por los ncRNAs recién identificados con el programa Randfold [213]. Dicha estabilidad se cuantifica con un P valor <0,01 y se ha estimado que el 90% de los miRNAs tienen un P valor <0.01, característica de 13 de los ncRNAs identificados en 15q11.2, incluidos los dos miRNAs (Tabla 3.4) detectados y dos de los predichos. Sin embargo, en el caso de otros

ncRNAs con estructuras secundarias conocidas, como los tRNAs y rRNAs, este P valor no es significativo para su caracterización, ya que sólo menos del 50% de los rRNAs y del 8% de los tRNAs tienen P valores  $<0,01$  [213]. Esto significa que la estabilidad de las estructuras secundarias medidas con este programa es un parámetro adecuado para filtrar predicciones de miRNAs, pero que para otros tipos de ncRNAs, este P valor sólo supone información adicional sobre la estabilidad de sus estructuras secundarias y no se debe utilizar como filtro de selección hasta que se hagan estudios más detallados que incluyan otros tipos de ncRNAs.

Con los programas CMsearch [296] y FastR [55, 221] concluimos que los ncRNAs de la región 15q11.2 no forman parte de las familias de ncRNAs identificadas hasta ahora.

Finalmente, las 21 secuencias fueron analizadas con RNAz [85, 217], un software desarrollado para predecir ncRNAs basándose en la estabilidad de la estructura secundaria y su conservación en diferentes especies. De este programa consideramos tres parámetros: *z-score*, Structural Conservation Index (SCI) y su P valor. Valores negativos del *z-score* indican que la estructura secundaria resultante es la más estable de las formadas, combinando aleatoriamente las bases de una secuencia dada, por lo que tiene el mismo significado que el P valor de Randfold. *z-scores*  $<-3$  corresponden a estructuras secundarias muy estables. Sin embargo, este valor no es muy fiable cuando el ncRNA está conservado en pocas especies y el porcentaje de identidad entre las diferentes secuencias es elevado como ocurre en nuestro estudio. Así en este caso, el P valor de Randfold es más adecuado para medir la estabilidad de las estructuras secundarias que el *z-score* de RNAz. Comparando los resultados obtenidos de ambos programas en este estudio, los ncRNAs con P valor  $<0,01$  son los mismos que los que tiene *z-score*  $<-3$  (Tabla 3.4). El valor del SCI debe ser próximo o superior al porcentaje de identidad entre las diferentes secuencias ortólogas para considerar que la estructura secundaria de un determinado ncRNA está conservada en las diferentes especies. Esta característica la cumplen todos los ncRNAs identificados en 15q11.2. Finalmente, el *z-score* y el SCI se combinan para dar un P valor. Se ha establecido que se pueden considerar ncRNAs los alineamientos con P valor  $>0,5$  [217]. Según esto, tenemos datos adicionales que soportan que 15 (71%) de los ncRNAs identificados en 15q11.2 sean ncRNAs (Tabla 3.4). Al observar la Tabla 3.4 vemos que los ncRNAs con P valor  $<0,5$  son los que tienen *z-scores* más altos y la contribución de este parámetro al P valor final explicaría este resultado. Es decir, que 6 ncRNAs no son considerados como tal por el programa RNAz porque la estabilidad de sus estructuras secundarias no es estadísticamente significativa, una característica que como hemos explicado antes no es necesario ni suficiente para seleccionar ncRNAs, por lo que serían necesarios otros

programas basados en otras características de los ncRNAs para considerar que lo que hemos detectado experimentalmente con estas seis sondas no corresponde a un ncRNA.

## 2.5 Predicción de posibles *targets* de los miRNAs identificados en 15q11.2

Predecir la función de un determinado ncRNA a partir sólo de su secuencia, es hoy imposible, pero en el caso de los miRNAs, que se conocen sus mecanismos de acción, se han desarrollado programas que predicen qué mRNA pueden ser *targets* o dianas de un determinado miRNA.

Escaneamos la secuencia de los extremos 3'UTR, de todos los transcritos del genoma humano registrados en el ensamblaje hg17, como posibles *targets* de los miRNAs RNA61 y RNA274 con el programa miRanda [289]. Identificamos 289 y 528 posibles *targets* de RNA61 y RNA274 respectivamente. Sorprendentemente 269 (93%) de las predicciones para RNA61 son comunes para RNA274, lo que indica que ambos miRNAs podrían participar en los mismos procesos celulares. 37 de estos mRNAs podrían ser reconocidos por cualquiera de los dos miRNAs identificados en 15q11.2 en más de una posición y de los 259 *targets* específicos de RNA274, 122 también podrían ser reconocidos en varias posiciones.

Analizamos las funciones de los *targets* predichos con miRanda para hacer una estimación de la función de estos miRNAs. Dado que la mayor parte de los *targets* de RNA61 (93%) lo son también de RNA274, los porcentajes dados a continuación se refieren a los 548 *targets* juntos (269 comunes, 20 específicos de RNA61 y 259 específicos de RNA274). Las variaciones considerando los tres grupos por separado son mínimas y en la Tabla S2 están especificados cuales son comunes y cuales específicos de cada miRNA. Observamos que el grupo con más representantes (11%) es el de *targets* que participan en el desarrollo y mantenimiento del SNC, seguidos con diferencia, por 5 grupos con porcentajes de representación similares: el grupo de genes que participan en la regulación de la expresión génica (factores de transcripción, proteínas que actúan en splicing, ribonucleoproteínas, etc), y los que codifican proteínas asociadas al desarrollo del cáncer (oncogenes, supresores tumorales, genes sobre-expresado en cáncer, etc) suponen un 6,6% de los *targets* totales cada uno; los que participan en las rutas de señalización celular y los que codifican Open Reading Frames (ORF), son un 5,1% cada grupo; y el menos representativo pero por poca diferencia, es el grupo de genes que actúan en la defensa frente a otros organismos (4,7%). Curiosamente, dos *targets* comunes de RNA 61 y RNA274 (ATXN2L y NUDT18) y uno específico de RNA274 (APTX) están asociados con distintos tipos de ataxias, una enfermedad

neurodegenerativa; dos de los *targets* predicho para RNA274 forman parte del sistema GABAérgico (ABAT y SLC3A1), un sistema alterado en pacientes con autismo [303-306] y dos predicciones más (DGCR2 y CLTCL1) están asociados con el desarrollo del síndrome de Di George (Tabla S2), sugiriendo un papel de estos miRNAs en la etiología de estas alteraciones o enfermedades.

## Material Suplementario

**Tabla S1. ncRNAs identificados en la región 15q11.2: posición, energía libre, conservación y secuencias.**

		Secuencias
<b>RNA5a</b>		
Posición	chr15:21483570- 21483629	5'GCAGCGACGGGCAGGGCGGGGGCGGCGCGGGACGCGGCACUGCGC CUGCGCGACCGCGGC_3'
dG (Kcal/Mol)	-35.2	
Conservación	Pan, Rhe	
<b>RNA5b</b>		
Posición	chr15: 21483532-21483631	5'AGGCCGCGGUCGCGCAGGGCGCAGUGCCGCGUCCCGCCGCCGCC GCCUGCCCGUCGUCGCGGAAGGGCGCCGCGCGCAGCAACGCGCACUU CCUCUCCA_3'
dG (Kcal/Mol)	-46.1	
Conservación	Pan, Rhe	
<b>RNA7</b>		
Posición	chr15:21486775-21486874	5'GGAAGGGGGGUCACCCCAGAGCCUCUAGUGGGAAGUGCGCAUGCUC GCAGUGCAGGGGCCAGAGGCCUGGAUGAAUGUGAAUCCUCUGGC UCUUC_3'
dG (Kcal/Mol)	-41.9	
Conservación	Pan, Rhe	
<b>RNA11_1</b>		
Posición	chr15:22043534-22043589	5'UUGGGAUGGAGUUUUGCUCUUGUAGCCCAGGCUGGAGUGCAGUGAC UCCAUCUCAG_3'
dG (Kcal/Mol)	-29.9	
Conservación	Pan	
<b>RNA11_2</b>		
Posición	chr15:22356680-22356735	5'UUGGGAUGGAGUUUUGCUCUUGUAGCCCAGGCUGGAGUGCAGUGAC UCCAUCUCAG_3'
dG (Kcal/Mol)	-29.9	
Conservación	Pan	
<b>RNA11_3</b>		
Posición	chr15:21968500-21968555	5'_UGGGAUGGAGUUUGCUCUCGUAGCCCAGGCUGGAGUGCAGUGA CUCCAUCUCAG_3'
dG (Kcal/Mol)	-29.9	
Conservación	Pan	
<b>RNA21</b>		
Posición	chr15:22464516-22464578	5'_CUCUUGAGGUUUGCUUCUUGCUAUCUGAAGUUUAUACCAGAUGGCA AGGAGCAAACCAAGGAG_3'
dG (Kcal/Mol)	-37.4	
Conservación	Pan, Rhe	
<b>RNA23_1</b>		
Posición	chr15:22085280-2085385	5'_CAAAUUCAGUCAACCUCCUAAAUUAUGUGAGUUUUUGUUUUUGUA UGUGAGUUUCUGCAUACACAUAAAGAAACUCCCAUAGUUUJAGGAGGUU GACUGAAUUUUG_3'
dG (Kcal/Mol)	-59.5	
Conservación	Pan, Rhe	
<b>RNA23_2</b>		
Posición	chr15:22251959-2252064	5'_CAAAUUCAGUCAACCUCCUAAAUUAUGUGAGUUUUUGUUUUUGUA UGUGAGUUUCUGCAUACACAUAAAGAAACUCCCAUAGUUUJAGGAGGUU GACUGCAUUUUG_3'
dG (Kcal/Mol)	-59.5	
Conservación	Pan, Rhe	
<b>RNA23_3</b>		
Posición	chr15:21921339-1921444	5'_CAAAUUCAGUCAACCUCCUAAAUUAUGUGAGUUUUUGUUUUUGUA UGUGAGUUUCUGCAUACACAUAAAGAAACUCCUUAUAGUUUJAGGAGGUU GACUGAAUUUUG_3'
dG (Kcal/Mol)	-59.5	
Conservación	Pan, Rhe	

<b>RNA28</b>		
Posición	chr15:21584052-21584126	5'_UCUAU <b>CCUCAUCAUCUGUGCUCAUUC</b> CUAGAACCAUUUUCUGGCCUU CAGAAUGGGCACAGCUGCUGAGGACAUGC_3'
dG (Kcal/Mol)	-33.6	
Conservación	Pan	
<b>RNA34</b>		
Posición	chr15:21557858-21557939	5'_GGGGAGGGGAGUGGUUGGGGGCUGGGGAGAUGCUGGCCUGUGCUG CUUCUCUGUUCUAACCCUUGCAGCCUAUUCUUCUCUUU_3'
dG (Kcal/Mol)	-34.8	
Conservación	Pan, Rhe	
<b>RNA49a</b>		
Posición	chr15:21557063-21557155	5'_AGGGAGGAGAGAGGGACAAAAGGGGGACCCUCUGCCUCAGGAGA UUCUGCUAGGAGCACUGGCCACCCUUUGUCUCUCAGCUCUCUUCUCU _3'
dG (Kcal/Mol)	-47.9	
Conservación	Pan, Rhe	
<b>RNA49b</b>		
Posición	chr15:21557066-21557154	5'_GAGAAGAGAGCUGAGAGACAAAGGGGUGGCCAGUGCUCUAGCAG AAUCUCCUGAGGGCAGAGGGUCCCCUUUGUCCUCUCUCCUC_3'
dG (Kcal/Mol)	-40.5	
Conservación	Pan, Rhe	
<b>RNA51</b>		
Posición	chr15:21558473-21558532	5'_GGGAGGGCAAGCCGUGAGCCUCCGUUGAGAGAGGGGGCUUC CGCCUUGCACUCUC_3'
dG (Kcal/Mol)	-37.2	
Conservación	Pan, Rhe, Mm, Rn, Can, Bos	
<b>RNA61_1</b>		
Posición	chr15:22349946-22350008	5'_ <b>GUGCAGAGGUAGCUGCCGGA</b> AGGGUAUGGACUUCUAUUACCCUUU UGGUAGUCACGGUGGCAC_3'
dG (Kcal/Mol)	-31.8	
Conservación	Pan, Rhe	
<b>RNA61_2</b>		
Posición	chr15:21961819-21961881	5'_ <b>GUGCAGAGGUAGCUGCAGGA</b> AGGGUAUGGACUACUAUUACCCUUU UGGUAGUCACAGUGGCAC_3'
dG (Kcal/Mol)	-31.8	
Conservación	Pan, Rhe	
<b>RNA61_3</b>		
Posición	chr15:22177395-22177456	5'_ <b>GUGCAGAGGUAGCUGCAGGA</b> AGGGUAUGGACUACUAUUACCCUUU UGGCAGUCACGGUGGCA_3'
dG (Kcal/Mol)	-31.8	
Conservación	Pan, Rhe	
<b>RNA67</b>		
Posición	chr15:22451398-22451490	5'_UGAAUGACCACUCAUCCUUGGGCUUGUUGCAAGACAGCCAUGCUAU GUAAAAGCAUCUGUCUUUUUGCAGACAGGAAUAAGAUGAAGGAUUCA_3
dG (Kcal/Mol)	-30.2	
Conservación	Pan, Rhe	
<b>RNA70</b>		
Posición	chr15:22417931-22417997	5'_GCAG <b>UGGCAGCACACUGAAGAGCAC</b> AUGCCUCAGCUGGGAUAGGG CACUGGUGGUUGCUGGACUGCU_3'
dG (Kcal/Mol)	-25.2	
Conservación	Pan	
<b>RNA78</b>		
Posición	chr15:21594294-21594358	5'_CACGUACCUGCUGGGCCUCUGCCGUGGGACGAGGUCCUCGCGAA GGUGUCUGUGCGAGGGCGUG_3'
dG (Kcal/Mol)	-30.8	
Conservación	Pan, Rhe, Mm, Rn	
<b>RNA106</b>		
Posición	chr15:22138304-22138373	5'_AGGCCGGCGCUAAGGGCCCCGUAUGGCGGUACCCGAGUGGGAGCU GUUGGGUGGAGCAGAGGCCAGGUG_3'
dG (Kcal/Mol)	-29.3	
Conservación	Pan, Rhe	

<b>RNA195a</b>		
Posición	chr15:22471784-22471872	5'_UGCCGGUGGCAGGUGUUGCACCGCGUGGCCUCUAUGAGGGAACUG CUGGGGCCUGCGGGCCAUCGCCAGAGCUCUUGUCACCUGGGCG_3'
dG (Kcal/Mol)	-43.5	
Conservación	Pan, Rhe	
<b>RNA195b</b>		
Posición	chr15:22471785-22471845	5'_GCCCAGGUGACAAGAGCUCUGGCGAUGGGCCCGCAGGCCCCAGCA GUUCCUCAUAGAGGC_3'
dG (Kcal/Mol)	-23.1	
Conservación	Pan, Rhe	
<b>RNA274</b>		
Posición	chr15:22273478-22273530	5'_AGCCUCCCCGGGGGACAGGGCCACUCACUCCUCAACGUCCUCCCC AUAGGCU_3'
dG (Kcal/Mol)	-16	
Conservación	Pan	
<b>RNA478</b>		
Posición	chr15:21739950-21740023	5'_GCACAUUGCACAAUCCUCAUUCAGCCUGUUCCAUCUCUUGGAUC AUGGCAUGGAGGGUUGAGCAUAUGUUGU_3'
dG (Kcal/Mol)	-33.1	
Conservación	Pan, Rhe	
<b>RNA514</b>		
Posición	chr15:21605773-21605872	5'_GGCAGCCUGCGCCCAAUAGGUGGGAUUCUAUGUCAAUCCACUGUG UUUUCACAGUGAUGGGCCAUAGGAUCCUGAGCUCAUUUGGAGCUAUU CAGGCUA_3'
dG (Kcal/Mol)	-45.3	
Conservación	Pan, Rhe	

Pan, PanTro2 (chimpancé); Rhe, RheMac2 (macaco); Mm, mm8 (Ratón); Rn, rn4 (Rata); Can, CanFam2 (Perro); Bos, BosTau2 (Vaca). Las bases marcadas en rojo son las secuencias de los miRNAs maduros comprobados experimentalmente y las marcadas en azul son los miRNAs predichos según miRscan.

Tabla S2. Predicción de targets para los miRNAs RNA61 y RNA274a.

Función del <i>target</i>	RNA61	RNA274	RNA61 y RNA274
Desarrollo y funcionamiento del SNC		<b>NFASC</b> , SYT7, EML1, CES2, FXR2, CDK5R1, C1QL1, SGSH, <b>BRE</b> , ACCN4, DGCR2, <b>CLTCL1</b> , HIYPB, BSN, PRR7, <b>LAT2</b> , CNTNAP2, ADAM9, SLC32A1	LHX3, KIAA0562, <b>CHD5</b> , <b>ACOT7</b> , EPHA8, BAI2, DAB1, PTGER3, NBPF4, EFNA3, ABAT, ATXN2L, DOC2A, SMPD3, FA2H, <b>GAS7</b> , <b>PIPEK1C</b> , CTXN1, KIF1A, DRD3, DCAMKL2, TPPP, SYNGAP1, GFRA2, NUDT18, NPDC1, CABP1, OPRS1
Regulación de la expresión génica	SOX11, SOX18	ILF3, TRMU, PADI3 KIF17, CUGBP2, RPP30, VENTX, MTG1, <b>SFRS8</b> , GLIS2, BRUNOL5, RFX1, SNRP70, HDAC4	ANKRD30A, MRPL43, DDX11, SOLH, PARN, <b>MNT</b> , CAMTA2, MED9, MRPL12, LMNB2, JUND, <b>SF4</b> , <b>GRLF1</b> , HNF4A, RREB1, HOXA3, PPARD, LHX3
Crecimiento, proliferación, división y diferenciación celular	FGFR2, PPP2R5B	RASGRP4, CRX, RBBP9, CDC25A, MPP2	
<i>Open reading frames</i>	C1orf27	C1orf38, C1orf113, C1orf183, C11orf64, C11orf65, C12orf47, C14orf166, C14orf43, C21orf59, C7orf20, C8orf13, C9orf16, C9orf7	C1orf102, <b>C11orf9</b> , C16orf7, C17orf82, C17orf56, C18orf45, C2orf24, C20orf71, C21orf34, C21orf125, C22orf26, C5orf20
Desarrollo de cáncer	ABTB1, CDCP1	<b>PRDM2</b> , ITIH5, IGF2AS, <b>TRAF7</b> , <b>AMFR</b> , RAD51C, TMEM16G, BAGE4, BAGE5, TIAM1, LIF, C1QTNF6, THBS2, ABL1, LOXL1	PPARD, <b>MUC1</b> , <b>PCNXL2</b> , THRSP, BCL7A, WNT3, RAB4B, HIC2, DLEC1, TNFAIP3, KIAA0828, <b>MTUS1</b> , WNT5B, PTGES, STC2
Asociados a síndromes y enfermedades neurológicas		APTX, SLC32A1, DGCR2, <b>CLTCL1</b> , FXR2, HYPB, EML1, <b>LAT2</b>	ATXN2L, NUDT18,
Defensa	CD5	<b>CSF1</b> , CXCL12, PLA2G4D, ATRN, RALY, MUC4, RAET1G, ILF3	PTGES, IRF4, IL28RA, CXCL12, MMP25, CD300LB, CLEC4M, <b>IGSF4C</b> , HS1BP3, OPRS1
Señalización celular		GPR153, ETNK2, ADRBK1, MRGPRF, P2RX7, MAP3K3, ADCYAP1, DIRAS1, DUSP2, <b>PREX1</b> , MAPKAPK3, PTPDC1, MPP2	ITPKB, LPHN1, VASP, FARP2, GNAZ, PTK7, PCTK1, CABP1
Transportadores		SLC32A1, TRPC4AP, SLC9A1, YKT6, ABCA2, <b>CACNA1H</b> , RANBP3	APOL4, SLC22A7, SLC9A1, STK40, CACNA2D4, ABCB9, ABCD4, ABCC12, ABCC3

Genes marcados con letra negra son aquellos en los que se han predicho más de un sitio de interacción mRNA-miRNA.

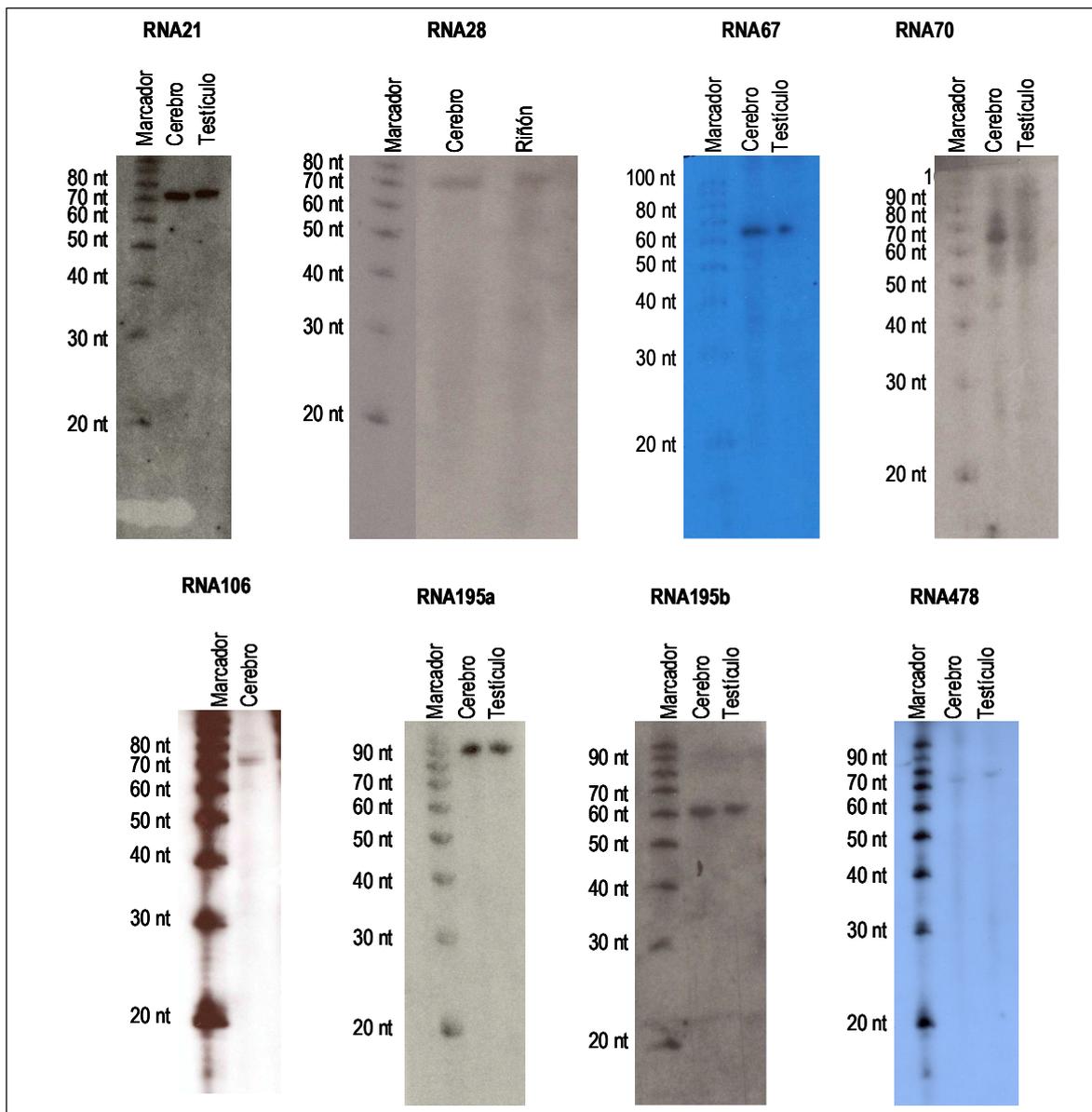


Figura S1. Blot de ncRNAs para los que sólo hemos observado una banda.

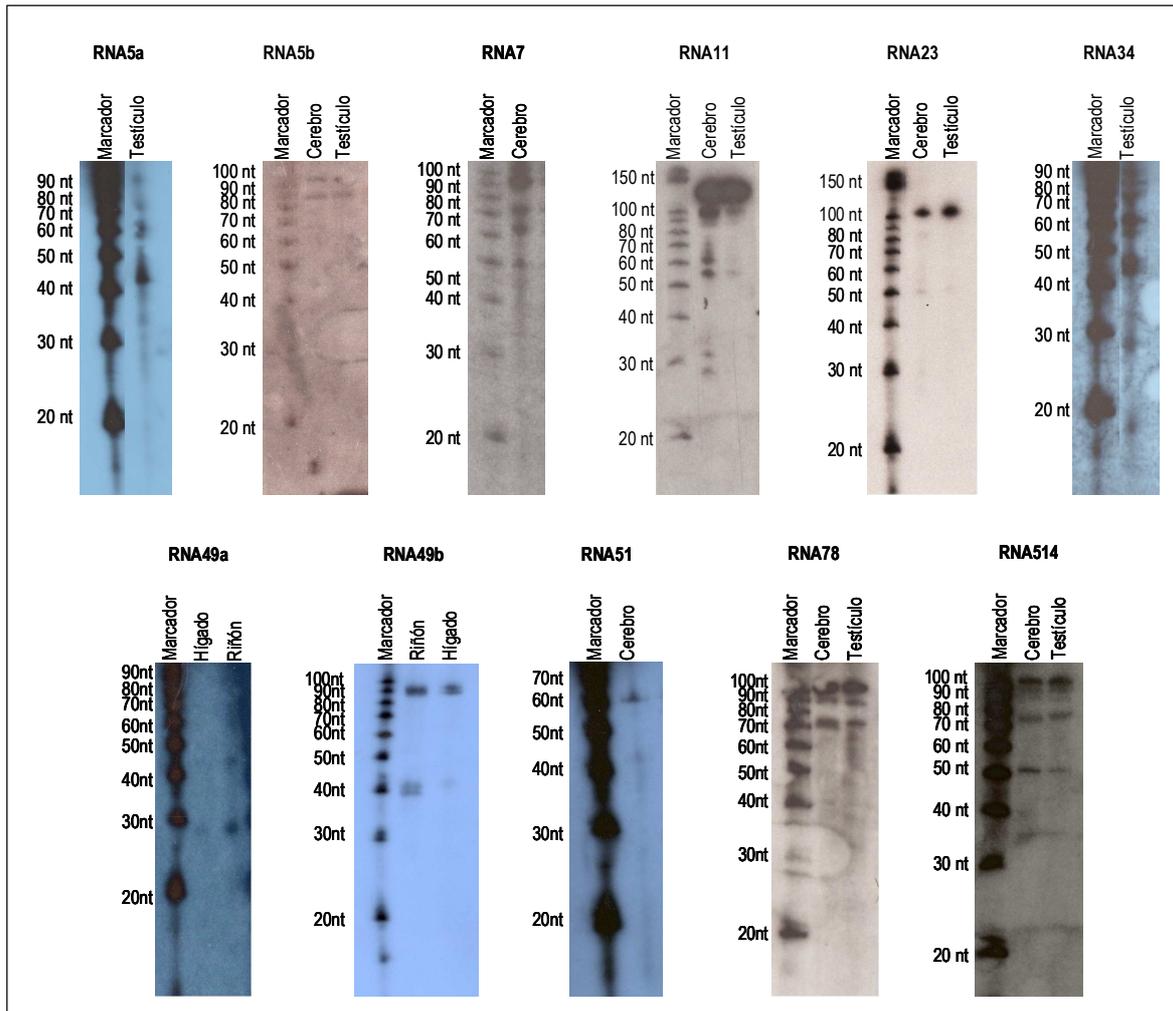
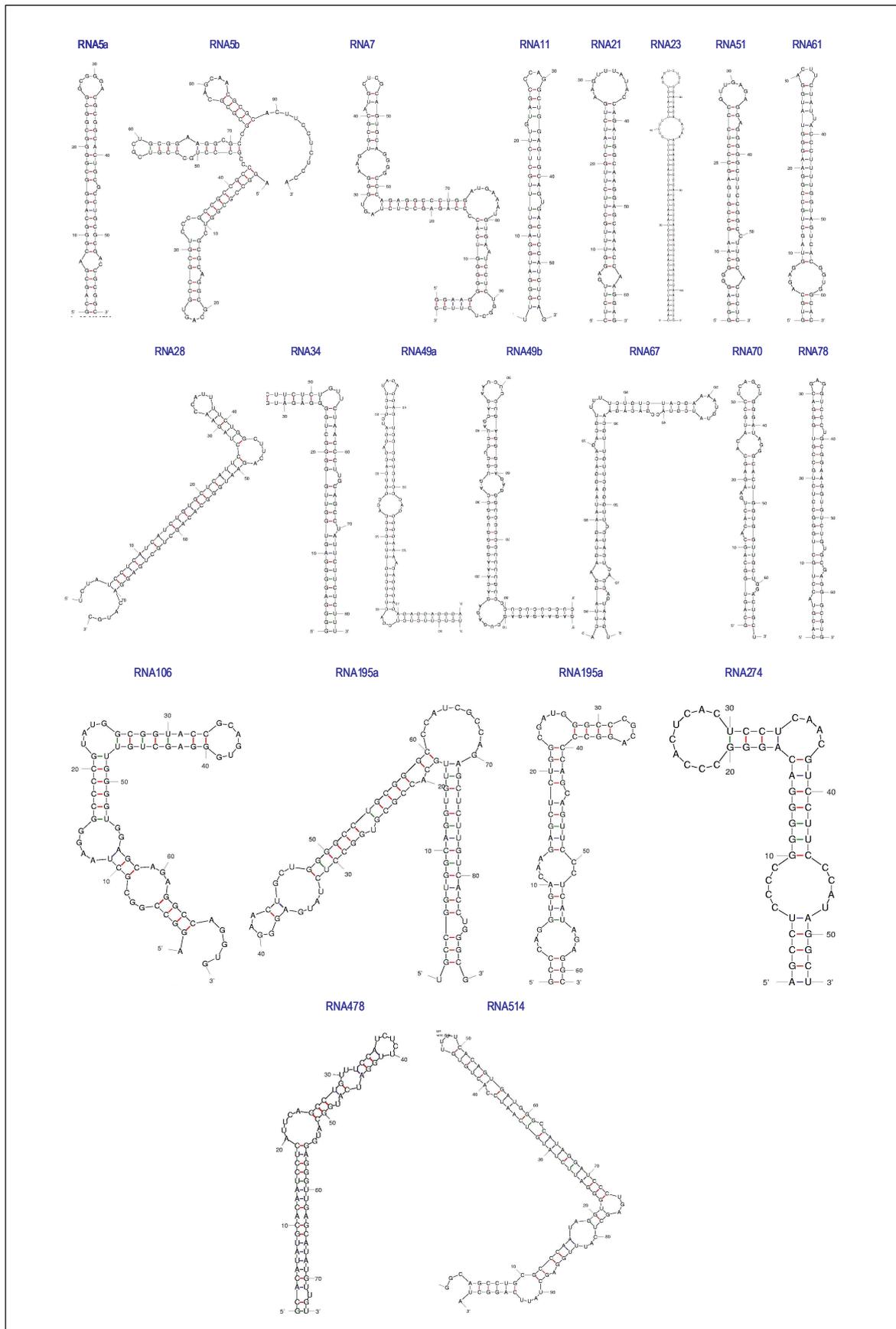


Figura S2. Blot de ncRNAs para los que se ha detectado más de una banda.



**Figura S3. Estructuras secundarias predichas para los nuevos 21 ncRNAs.**

# ***Discusión***



## **Discusión**

En este apartado se describirá la relevancia de los resultados previamente expuestos para entender la organización de la región cromosómica 15q11-q13 y su papel en el desarrollo del autismo y otros trastornos del SNC, integrándolos con los datos y resultados de estudios anteriores.

La compleja estructura de las duplicaciones segmentarias de 15q11-q13, los CNVs y la presencia de *gaps*, convierten a 15q11-q13 en una región inestable y muy variable. Algunas de estas variaciones han sido asociadas a enfermedades. Hoy sabemos que deleciones de 15q11-q13 son la principal causa de AS y PWS [307, 308] y duplicaciones de la misma es la alteración citogenética más frecuente en pacientes con autismo [74, 158, 168, 169, 309, 310]; pero también se ha encontrado asociación de genes codificados en 15q11-q13 con el trastorno autista [107, 110, 112-114, 136, 174-176, 294, 311] y microrreordenamientos de 15q11.2 en pacientes con RTT [178], por lo que es probable que existan variaciones submicroscópicas de esta región asociadas al autismo u otras enfermedades. Por otro lado, mientras diferentes estudios sugieren la asociación de las variaciones genómicas con un determinado fenotipo [67, 79], aún se conocen variaciones sin fenotipo asociado. Para esta tarea es importante conocer las secuencias funcionales que contiene cada variación. Dada la importancia que están adquiriendo en nuestros días las regiones no codificantes del genoma y las cifras de predicciones de ncRNAs frente a la de genes codificantes [230, 312], la identificación de nuevos ncRNAs puede ser clave para explicar diferencias entre individuos de la misma especie.

### **Variabilidad genómica de 15q11.2**

En la región 15q11.2 se habían observado microrreordenamientos en tres pacientes con RTT mediante la genotipación de los marcadores D16S646 y D15S817. En las muestras analizadas en nuestro estudio no encontramos evidencias de reordenamientos ni en las frecuencias, ni en la transmisión alélica de D15S646. Longo et al. [178] identificaron deleciones maternas de este marcador, en 2 de 63 pacientes con RTT no relacionados entre ellos y en dos hermanos sanos de uno de los pacientes. La presencia de la deleción en individuos sanos sugiere que no está asociada al desarrollo de la enfermedad, sin embargo, no identificaron esta deleción en ninguno de los 200 muestras controles analizadas. Es probable que se trate de una deleción rara y/o más frecuente en la población italiana o que esté asociada a RTT pero no a autismo. En el mismo estudio identificaron una duplicación paterna de D15S817 en un solo paciente, y

como la delección de D15S646, este reordenamiento tampoco se observó en ningún control. A pesar de intentar amplificar D15S817 de manera específica, con diferentes parejas de *primers*, incluidos los utilizados por Longo et al., la forma en que se heredaban los alelos nos indicaba que estábamos amplificando dos regiones a la vez, por lo que no consideramos D15S817 un marcador informativo para definir reordenamientos de 15q11.2.

En la misma región de duplicaciones segmentarias nosotros identificamos delecciones de tamaños variables a ~330 Kb de D15S646. Las delecciones incluyen sólo el marcador D15S1548 o pueden incluir también a D15S1549, pero nuestros datos sugieren que las delecciones de este último son sólo parte de delecciones más largas. De todas formas debido a los límites de la genotipación de microsatélites y de los SNPs del proyecto HapMap, no podemos descartar la posibilidad de que existan delecciones de D15S1549 independientes de D15S1548.

Para definir los límites de la delección combinamos nuestros datos con los del proyecto HapMap. Sólo dos casos resultaron informativos para este tipo de análisis. Así, limitamos la delección de D15S1548 a ~44 Kb y las delecciones D15S1548-D15S1549 a ~62 Kb (Figura 3.5). Como nuestros datos están basados sólo en casos informativos es posible que las delecciones tengan tamaños más variables. También observamos una duplicación de D15S1549 en una muestra control lo que sugiere que la región es muy inestable y que ambos tipos de reordenamientos pueden ser frecuentes y de tamaños polimórficos.

Las delecciones identificadas coinciden con la posición de cuatro CNVs [55, 59, 66] (Figura 3.5). Redon et al [55]. han identificado variaciones de ~503 Kb y ~647 Kb mediante los *arrays* 500K SNP Affymetrix and CGH WGTP (*Whole Genome TilePath Array*) respectivamente. (Figura 3.5, variaciones 3072 y 3959). Por WGTP CGH identificaron 8 individuos (3%) con pérdida y 4 (1,5%) con ganancia de material genético de este locus, mientras que por el *array* de Affimatrix detectaron 16 delecciones y 5 ganancias del mismo locus. Para los dos *arrays* utilizaron las mismas 270 muestras del proyecto HapMap. Siete de las ocho delecciones observadas por Redon et al. con los dos *arrays* incluyen los marcadores D15S1548 y D15S1549. Sin embargo, los límites de cada CNV no están bien definidos. Según el *array* CGH WGTP la delección del individuo NA12056 incluye a D15S1548 y D15S1549, según el *array* de Affimatrix sólo a D15S1549 y según nuestros datos NA12056 es homocigoto para D15S1547 y heterocigoto para D15S1549. Así parece que la delección de este individuo es de diferente tamaño, afecta a una región entre los dos marcadores y comparando todos los datos hemos limitado sus bordes entre el SNP A\_2135235 y el marcador D15S1549 (~30,7 Kb). Esto supone mayor

variabilidad del locus. Entre las deleciones detectadas sólo por el array de Affimetrix, dos incluyen los dos marcadores y una solo a D15S1548, soportando nuestra hipótesis de que las deleciones de D15S1549 son sólo parte de deleciones más largas. Redon et al. también identificaron ganancia de material genético en el mismo individuo en el que nosotros identificamos tres alelos para el marcador D15S1549. Según sus resultados, la duplicación es más larga, incluye también a D15S1548 y es heredada por el hijo. En nuestro estudio ambos eran homocigotos no informativos para D15S1548 y el hijo era heterocigoto para D15S1549, en este caso la heterocigosidad no era informativa para la duplicación ya que ambos progenitores tenían una copia de cada uno de los alelos.

Aunque hemos visto que nuestros datos son compatibles con los resultados de Redon et al., nosotros definimos deleciones más pequeñas en dos individuos CEU. Es posible que en esta región se solapen CNVs de tamaño diferentes, tanto pérdidas como ganancias de material genético. Soportando esta hipótesis, McCarroll et al.[59] identificó deleciones que incluyen D15S1548 y D15S1549, pero mucho más pequeñas (~ 37Kb) de las observadas por Redon et al., en 3 de 269 muestras (1%) del proyecto HapMap (CNV 1959, Figura 3.5). Nuestros datos demuestran que el CNV descrito por McCarroll et al. corresponde a deleciones de tamaños variables de mayor frecuencia en la población europea.

Recientemente, Wong et al. [66] y Szatmari et al. [82] han publicado dos trabajos independientes dónde describen ganancias de material genético de esta misma región mediante los arrays CGH WGTP y Affimetrix 10K respectivamente. Las variaciones observadas de este locus incluyen a D15S1548 y D15S1549 pero las identificadas por Wong et al. son más pequeñas (~183,2 Kb) incluso que las observadas por Redon et al. y las detectadas por Szatmari et al. miden ~4,2Kb y ~3,8Kb. Los primeros utilizaron 95 muestras representando 16 etnias diferentes y observaron CNVs en 12 de ellas (12,6%), mientras que Redon et al. con el mismo método observaron ganancias de este locus en 4 de 270 muestras (1,5%). Sería necesario conocer la proporción de muestras con origen europeo en el estudio de Wong et al. portadoras de la duplicación para corroborar que los CNVs de este locus del genoma humano son más frecuentes en la población de origen europeo. Szatmari et al. identificaron el mismo reordenamiento en siete pacientes con autismo reforzando la asociación de 15q11-q13 con dicho trastorno.

### **Asociación de la delección polimórfica identificada en la región 15q11.2 con el autismo**

Comparamos las frecuencias de la delección de 15q11.2 en las poblaciones de autistas y controles y no encontramos diferencias estadísticamente significativas, pero si observamos una tendencia a la transmisión del alelo delecionado D15S1548 en las familias con hijos autistas. Debido al número de familias no informativas de las dos poblaciones, es necesario utilizar una técnica más cuantitativa para obtener resultados más fiables. Mediante FBAT, un test limitado a las familias informativas y más potente que los basados en las frecuencias, observamos una fuerte asociación de la delección de D15S1548 con el trastorno autista utilizando cualquiera de los modelos de herencia ( $P < 0,00$ ) y una tendencia a la preferencia de transmisión de las delecciones de D15S1549 ( $P = 0,052$ ) y de D15S1548-D15S1549 ( $P = 0,045$ ). La falta de significación estadística en los últimos casos es posiblemente debida a que el número de familias informativas para estas delecciones era más bajo. Nuestros resultados soportan que las delecciones polimórficas que incluyen D15S1548 y D15S1549 estén asociadas al trastorno autista. En la región 15q11.2 no se ha descrito ningún gen pero si se ha predicho uno (NT\_026446.20) y se han identificado varios ESTs, indicando transcripción, pero es necesario conocer qué genes están codificados en esta región para entender el significado funcional de tanta variabilidad genómica en este locus y cómo esta delección participa en el desarrollo del autismo y/o otras enfermedades.

### **Posible asociación de otros CNVs de 15q11-q13 con el trastorno autista**

Estudios anteriores de asociación y ligamiento con autistas mostraban a los genes de expresión materna (*UBE3A* y *ATP10A*) y el locus de los receptores de GABA como los genes candidatos para explicar el papel de 15q11-q13 en la etiología del autismo [107, 110, 113-115, 160, 175, 177, 294]. Sin embargo, nosotros no encontramos evidencias de reordenamientos en esta región. En el caso de los marcadores D15S128, D15S1552 y D15S1553 se puede explicar por la baja frecuencia en que aparecen los CNVs con los que se solapan en la población caucásica (3/270 individuos, 1/55 individuos y 1/270 individuos para las variaciones 3073, 0183 y 3953 respectivamente) [55, 58, 59, 78]. Otra posibilidad es que como estos CNVs están identificados por CGH array, sus bordes no estén bien definidos y los marcadores que nosotros hemos genotipado no coincidan con la región variable. Así, la asociación de estos CNVs con el autismo no debe dejar de ser considerada.

## Identificación de ncRNAs en la región 15q11.2 de cromosoma humano

Con el fin de identificar la parte funcional de las variaciones genómicas observadas en la región 15q11.2 del genoma humano, y entender su contribución a la variabilidad humana, reforzar su asociación con el desarrollo del autismo y establecer nuevas relaciones con otras enfermedades o síndromes, analizamos con detalle la arquitectura de la región y consideramos la posibilidad de que estas secuencias funcionales podrían corresponder a ncRNAs.

Desde que se conoce que más de la mitad de la actividad transcripcional del genoma humano corresponde a ncRNAs, su papel en diversos procesos fisiológicos y del desarrollo, así como su asociación con diferentes tipos de cáncer y enfermedades, la identificación y caracterización de nuevos ncRNAs han sido el objetivo de numerosos trabajos científicos [203, 204, 212, 214, 215, 301, 313-315]. Al principio, la secuenciación de librerías de cDNAs supuso un éxito para la identificación de ncRNAs [203-205], pero los codificados en 15q11.2 tienen niveles de expresión tan bajos que sería necesario secuenciar muchos clones para identificarlos, no siendo detectables por este método. Los límites de la secuenciación actualmente se resuelven con técnicas como *Masively Parallel Sequencing*. Hasta hoy sólo se ha publicado un estudio basado en esta técnica referido al genoma humano [206]. Berezikov et al. han identificado así 447 nuevos miRNAs, muchos de ellos específicos de primates, pero ninguno de la región 15q11.2. Nuestros resultados indican que la mayor parte de los ncRNAs codificados en 15q11.2 no son miRNAs y Berezikov et al. sólo consideraron las secuencias que podrían corresponder a este tipo de ncRNAs. También se sabe que algunos miRNAs resultan difíciles de clonar, bien por sus propiedades físicas, la composición de su secuencia o porque sufren modificaciones post-transcripcionales [55, 316, 317], y entre ellos podrían estar los de la región 15q11.2.

Desde 2004 se han realizado experimentos de *Genome Tiling Array* para la identificación de ncRNAs [227, 230, 318] y ninguno de ellos hace referencia a regiones transcripcionalmente activas (TAR) en 15q11.2. En algunos casos posiblemente por la falta de cobertura de la región y en otros porque el elevado número de TARs detectado en cada experimento, obliga a los autores a seleccionar las secuencias a analizar con más detalle. Hasta ahora se han centrado en la identificación de miRNAs, y sólo el recién publicado trabajo de Zhang et al. [230] considera otros RNA funcionales. En este estudio se identificaron 17.198 TARs de los que han intentado caracterizar 353. Los ncRNAs de 15q11.2 pueden estar representados entre los 16.845 restantes. *Genome Tiling Arrays* suponen una potente herramienta para la identificación de nuevos RNAs funcionales.

Una alternativa para identificar ncRNAs son las predicciones bioinformáticas. Estas como consecuencia de la ausencia de señales características de ncRNAs en su

secuencia primaria [207], se basan principalmente en la tendencia de los ncRNAs a formas estructuras secundarias [212, 319]. Nosotros seleccionamos secuencias de 100 nt en 15q11.2 que formaban *hairpins* simples y estables, que es la forma que adoptan principalmente los ncRNAs pequeños, incluidos los pre-miRNAs [320]. En otros trabajos se han utilizado programas que introducen filtros basados en la conservación de la estructura secundaria a lo largo de la evolución, como miRAlign [220], MiRscan [321], RNAz [217], Dynalign [216, 218] o QRNA [216] y así consiguen disminuir el número de falsos positivos. Dado que los ncRNAs candidatos seleccionados en 15q11.2 eran específicos de primates, decidimos no utilizar estas herramientas para filtrar secuencias candidatas.

Los ncRNAs predichos con herramientas bioinformáticas deben ser validados experimentalmente y los métodos de hibridación, *Northern Blot* y *Primer extension*, son las más adecuadas para comprobar la expresión de ncRNAs. *Northern Blot* además proporciona información sobre el tamaño del transcrito. Sin embargo, como todas las técnicas tienen sus limitaciones: la primera es la sensibilidad, que impide detectar ncRNAs expresados a muy bajo nivel, como las formas maduras de los miRNAs RNA61 y RNA274, en algunos tejidos; y la otra es que no se pueden detectar varios candidatos a la vez y si además el ncRNA tiene más de una forma, como los miRNAs (pre-miRNA y miRNA maduro), pueden ser necesarias varias sondas para detectarlas. *Primer extensión* es más sensible y además es útil para identificar el extremo 5' de los transcritos, pero en este caso el éxito del resultado depende de la especificidad del *primer* y de la distancia entre este y el extremo 5' del transcrito, siendo necesarios también varios *primers* para detectar un ncRNA. Esto hace que estas técnicas no sean recomendadas para comprobar muchos candidatos, siendo sustituidas por los *arrays*.

### ¿ncRNAs o sólo productos de transcripción?

Debido a que el patrón de bandas observado en 19 de los ncRNAs identificados en este estudio no corresponde a ninguna clase de ncRNAs conocida hasta ahora, su función es desconocida, pero esto no justifica que sean sólo productos de transcripción. De hecho, patrones similares habían sido observado antes por Berezikov et al. [214]. Ellos se centraban en la búsqueda de miRNAs y consideraron que estas bandas correspondían a pre-miRNAs, pero realmente sólo se ha comprobado que lo sean 5 de los 19 candidatos para los que observaban una o más bandas del tamaño de un pre-miRNA [212, 215, 255, 322]. La única diferencia entre ambos estudios es que los candidatos de Berezikov et al. están conservados en vertebrados y los nuestros son específicos de primates. Según las

predicciones de MiRscan, tres de los ncRNAs identificados en 15q11.2 podrían ser miRNAs, pero a pesar de la coincidencia entre los dos trabajos y de las predicciones bioinformáticas, en los casos en los que no se han detectado las formas maduras no podemos dejar de considerar que se trate de otro tipo de ncRNA, aunque la observación repetida de *blots* con patrones de bandas similares soporta la hipótesis de que los RNAs observados son funcionales.

Por otro lado, el trabajo de Zhang et al. [230] muestra que los miRNAs que no se conocían antes tenían niveles de expresión muy bajos, que más de la mitad eran específicos de primates y el 8% específicos de humanos, mientras que el 75% de los miRNAs previamente identificados estaban conservados en primates y mamíferos, es decir, que la fracción desconocida de miRNAs y posiblemente de ncRNAs, son aquellos con los niveles más bajos de expresión y menos conservados, como son los ncRNAs de 15q11.2. Esto se debe a los límites de las técnicas empleadas hasta ahora: los métodos experimentales no eran muy sensibles y las predicciones bioinformáticas dependen mucho de la conservación

Las secuencias de los ncRNAs identificados en 15q11.2 fueron también analizadas con el programa RNAz, considerado el más efectivo de los desarrollados hasta ahora para predecir ncRNAs [323]. Su efectividad se debe a que es el único que combina la estabilidad de la estructura secundaria con su conservación en otras especies [217]. Considerando una  $P > 0,5$  la sensibilidad del programa es del 84% y la especificidad 96,5%. A 15 de los 21 ncRNAs identificados en 15q11.2 les corresponde una  $P > 0,5$ , e incluso 10 tienen  $P > 0,9$ , valor para el que han calculado una especificidad del 98,9%, demostrando que las secuencias analizadas corresponden a RNAs funcionales. Los 6 restantes posiblemente no alcanzan este  $P$  valor porque la estabilidad de su estructura secundaria no es estadísticamente significativa, como hemos comprobado también con el programa RandFold. El valor de este parámetro para predecir ncRNAs, sin tener en cuenta la conservación, sólo ha sido calculado para miRNAs, tRNAs y rRNAs y sólo es informativo en el caso de los miRNAs. Así, son necesarios más estudios para poder utilizar la significación estadística de la estabilidad de las estructuras secundarias como filtro de ncRNAs candidatos.

Otros estudios han mostrado que la expresión de la mayoría de los ncRNAs está regulada por los mismos mecanismos que controlan la expresión de genes codificantes [324, 325], soportando la hipótesis de que la gran mayoría de los ncRNAs son funcionales.

Todos los puntos expuestos previamente soportan que los transcritos detectados experimentalmente se correspondan con ncRNAs.

### ncRNAs identificados en 15q11.2 y enfermedades genómicas

Predecir la función de un ncRNA del que sólo se conoce su secuencia es hoy muy difícil, por eso, en este apartado más que discutir las posibles funciones de los ncRNAs identificados en la región 15q11.2, vamos a intentar relacionarlos con la variabilidad genómica entre personas y con el desarrollo del autismo y otras enfermedades del neurodesarrollo, que era además el objetivo de esta parte del proyecto.

La causa más frecuente de los AS y PWS son deleciones de 15q11-q13 de origen materno y paterno respectivamente [72], y duplicaciones maternas de la misma región son la alteración citogenética más común en pacientes con autismo [160, 161]. Así alteraciones en el número de copia de estos ncRNAs podrían ser un factor implicado en la etiología de dichas enfermedades. Igualmente 15q11.2 es una región muy polimórfica en cuanto al número de copias y al tamaño de estas variaciones [55, 59, 66], y diferencias en el número de copias de todos o algunos ncRNAs podrían explicar la variabilidad entre humanos y/o la predisposición al desarrollo de otras enfermedades.

La expresión de los genes que flanquean a los ncRNAs de 15q11.2 está regulada por *imprinting* y se expresan en el alelo paterno en el cerebro [15], por lo que estos ncRNAs también pueden estar improntados. Una posibilidad es que ellos mismos, participen en el mantenimiento de dicho proceso, porque aunque el papel de los ncRNAs improntados, ya identificados, es aún desconocido, su integridad y su expresión en el alelo correcto, parece crucial para el mantenimiento de la actividad transcripcional del resto de genes, localizados en la misma región improntada [198, 200, 326]. Alteraciones del *imprinting* de los genes codificados en 15q11-q13 están asociadas al desarrollo de los AS y PWS, sugiriendo otra posible forma de asociación de estos ncRNAs con dichos síndromes.

La relación entre los ncRNAs de 15q11.2 con trastornos del neurodesarrollo está soportada por el alto porcentaje de *targets* predichos para RNA61 y RNA274 que participan en el desarrollo y mantenimiento del SNC. Otros reordenamientos que incluyen ncRNAs se ha encontrado en un niño autista [165, 179] y este trabajo junto con las duplicaciones maternas de 15q11-q13 detectadas en autistas, la asociación de la deleción de 15q11.2 con el mismo trastorno, los ncRNAs identificados en 15q11.2 y los *targets* predichos para RNA274 que forman parte del sistema GABAérgico, implican a los ncRNAs en el origen del autismo.

Confirmar los *targets* de RNA61 y RNA274 puede ser de gran utilidad para conocer en que procesos participan y cómo influyen en la variabilidad humana, el desarrollo del autismo, AS, PWS u otras enfermedades. Sin embargo, atribuir la función de una nueva

clase suele requerir ensayos *in vitro* e *in vivo*, resultando lento y complicado y más aún en el caso de ncRNAs específicos de primates.

Nuestros resultados demuestran que el mundo de los RNAs es todavía el gran desconocido, quedando aún por investigar muchas clases de ncRNAs, su localización celular, sus patrones de expresión, mecanismos moleculares de actuación, interacciones con DNA, proteínas y otros RNAs y cómo afectan a la variabilidad inter- e intra específica y al desarrollo de enfermedades.



# ***Conclusiones***



## Conclusiones

Las aportaciones de este trabajo pueden resumirse en los siguientes puntos:

- Hemos identificado dos polimorfismos en la región cromosómica 15q11-q13: una deleción de tamaño y origen variable en 15q11.2 y una inserción en el sexto intrón del gen *GABRG3*.
- Mediante FBAT hemos comprobado que la deleción del marcador D15S1548 está asociada con el trastorno autista y que existe una tendencia a la asociación de las deleciones de D15S1549 y D15S1548-D15S1549 con el mismo trastorno, aunque estas no alcanzaron significación estadística.
- Hemos caracterizado la parte funcional de la región variable 15q11.2 identificando un *cluster* de ncRNAs específicos de primates. De acuerdo a los resultados experimentales hemos comprobado que dos de ellos son miRNAs y no hemos podido clasificar a los 19 restantes entre las clases de ncRNAs hasta ahora conocidas.
- Todos los ncRNAs identificados en 15q11.2 se expresan al menos, en cerebro, testículo, hígado y riñón. Los miRNAs RNA61 y RNA274 se expresan además en tejido adiposo, vejiga, útero, colon, esófago, corazón, pulmón, próstata y bazo. Los niveles de expresión de todos ellos son muy bajos.
- El análisis bioinformático de los ncRNAs codificados en 15q11.2, con diferentes programas, demuestra que no pertenecen a ninguna familia de ncRNAs hasta ahora conocida, apoyando que 15 de ellos son RNAs funcionales y prediciendo que RNA28, RNA70 y RNA478 también podrían ser miRNAs.
- Con la información disponible no es posible asignar una función a estos ncRNAs, pero su localización en el genoma y la predicción de *targets* de RNA61 y RNA274 apoyan el papel de los ncRNAs de 15q11.2, en el desarrollo y mantenimiento del SNC, así como sugieren una posible participación en las bases moleculares del autismo, AS, PWS, las ataxias, síndrome de DiGeorge y el cáncer. Serán necesarios estudios funcionales para comprobar esta hipótesis.



# ***Referencias***



## Referencias

1. Boccaccio I, Glatt-Deeley H, Watrin F, Roeckel N, Lalande M, Muscatelli F: **The human MAGEL2 gene and its mouse homologue are paternally expressed and mapped to the Prader-Willi region.** *Hum Mol Genet* 1999, **8**(13):2497-2505.
2. MacDonald HR, Wevrick R: **The neccin gene is deleted in Prader-Willi syndrome and is imprinted in human and mouse.** *Hum Mol Genet* 1997, **6**(11):1873-1878.
3. Glenn CC, Porter KA, Jong MT, Nicholls RD, Driscoll DJ: **Functional imprinting and epigenetic modification of the human SNRPN gene.** *Hum Mol Genet* 1993, **2**(12):2001-2005.
4. Vu TH, Hoffman AR: **Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain.** *Nat Genet* 1997, **17**(1):12-13.
5. Rougeulle C, Glatt H, Lalande M: **The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain.** *Nat Genet* 1997, **17**(1):14-15.
6. Meguro M, Kashiwagi A, Mitsuya K, Nakao M, Kondo I, Saitoh S, Oshimura M: **A novel maternally expressed gene, ATP10C, encodes a putative aminophospholipid translocase associated with Angelman syndrome.** *Nat Genet* 2001, **28**(1):19-20.
7. Herzing LB, Kim SJ, Cook EH, Jr., Ledbetter DH: **The human aminophospholipid-transporting ATPase gene ATP10C maps adjacent to UBE3A and exhibits similar imprinted expression.** *Am J Hum Genet* 2001, **68**(6):1501-1505.
8. Hogart A, Nagarajan RP, Patzel KA, Yasui DH, Lasalle JM: **15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders.** *Hum Mol Genet* 2007.
9. Lee S, Kozlov S, Hernandez L, Chamberlain SJ, Brannan CI, Stewart CL, Wevrick R: **Expression and imprinting of MAGEL2 suggest a role in Prader-willi syndrome and the homologous murine imprinting phenotype.** *Hum Mol Genet* 2000, **9**(12):1813-1819.
10. Farber C, Gross S, Neesen J, Buiting K, Horsthemke B: **Identification of a testis-specific gene (C15orf2) in the Prader-Willi syndrome region on chromosome 15.** *Genomics* 2000, **65**(2):174-183.
11. Buiting K, Nazlican H, Galetzka D, Wawrzik M, Grobeta S, Horsthemke B: **C15orf2 and a novel noncoding transcript from the Prader-Willi/Angelman syndrome region show monoallelic expression in fetal brain.** *Genomics* 2007.
12. Gray TA, Saitoh S, Nicholls RD: **An imprinted, mammalian bicistronic transcript encodes two independent proteins.** *Proc Natl Acad Sci U S A* 1999, **96**(10):5616-5621.
13. Ozcelik T, Leff S, Robinson W, Donlon T, Lalande M, Sanjines E, Schinzel A, Francke U: **Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader-Willi syndrome critical region.** *Nat Genet* 1992, **2**(4):265-269.
14. Farber C, Ditttrich B, Buiting K, Horsthemke B: **The chromosome 15 imprinting centre (IC) region has undergone multiple duplication events and contains an upstream exon of SNRPN that is deleted in all Angelman syndrome patients with an IC microdeletion.** *Hum Mol Genet* 1999, **8**(2):337-343.
15. Nicholls RD, Saitoh S, Horsthemke B: **Imprinting in Prader-Willi and Angelman syndromes.** *Trends Genet* 1998, **14**(5):194-200.
16. Wevrick R, Kerns JA, Francke U: **Identification of a novel paternally expressed gene in the Prader-Willi syndrome region.** *Hum Mol Genet* 1994, **3**(10):1877-1882.
17. Runte M, Huttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K: **The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A.** *Hum Mol Genet* 2001, **10**(23):2687-2700.
18. Huibregtse JM, Scheffner M, Beaudenon S, Howley PM: **A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase.** *Proc Natl Acad Sci U S A* 1995, **92**(7):2563-2567.
19. Huibregtse JM, Scheffner M, Howley PM: **Cloning and expression of the cDNA for E6-AP, a protein that mediates the interaction of the human papillomavirus E6 oncoprotein with p53.** *Mol Cell Biol* 1993, **13**(2):775-784.
20. Scheffner M, Nuber U, Huibregtse JM: **Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade.** *Nature* 1995, **373**(6509):81-83.
21. Greger V, Knoll JH, Woolf E, Glatt K, Tyndale RF, DeLorey TM, Olsen RW, Tobin AJ, Sikela JM, Nakatsu Y *et al*: **The gamma-aminobutyric acid receptor gamma 3 subunit gene (GABRG3) is**

- tightly linked to the alpha 5 subunit gene (GABRA5) on human chromosome 15q11-q13 and is transcribed in the same orientation.** *Genomics* 1995, **26**(2):258-264.
22. Ancans J, Tobin DJ, Hoogduijn MJ, Smit NP, Wakamatsu K, Thody AJ: **Melanosomal pH controls rate of melanogenesis, eumelanin/phaeomelanin ratio and melanosome maturation in melanocytes and melanoma cells.** *Exp Cell Res* 2001, **268**(1):26-35.
  23. Lee ST, Nicholls RD, Jong MT, Fukai K, Spritz RA: **Organization and sequence of the human P gene and identification of a new family of transport proteins.** *Genomics* 1995, **26**(2):354-363.
  24. Bachellerie JP, Cavaille J, Huttenhofer A: **The expanding snoRNA world.** *Biochimie* 2002, **84**(8):775-790.
  25. Kiss T: **Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions.** *Cell* 2002, **109**(2):145-148.
  26. Filipowicz W: **Imprinted expression of small nucleolar RNAs in brain: time for RNomics.** *Proc Natl Acad Sci U S A* 2000, **97**(26):14035-14037.
  27. Mattick JS, Makunin IV: **Small regulatory RNAs in mammals.** *Hum Mol Genet* 2005, **14 Spec No 1**:R121-132.
  28. Cavaille J, Buiting K, Kieffmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A: **Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization.** *Proc Natl Acad Sci U S A* 2000, **97**(26):14311-14316.
  29. Burns CM, Chu H, Rueter SM, Hutchinson LK, Canton H, Sanders-Bush E, Emeson RB: **Regulation of serotonin-2C receptor G-protein coupling by RNA editing.** *Nature* 1997, **387**(6630):303-308.
  30. Canton H, Emeson RB, Barker EL, Backstrom JR, Lu JT, Chang MS, Sanders-Bush E: **Identification, molecular cloning, and distribution of a short variant of the 5-hydroxytryptamine-2C receptor produced by alternative splicing.** *Mol Pharmacol* 1996, **50**(4):799-807.
  31. Kishore S, Stamm S: **The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C.** *Science* 2006, **311**(5758):230-232.
  32. Paul MS, Bass BL: **Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA.** *Embo J* 1998, **17**(4):1120-1127.
  33. Yi-Brunozzi HY, Easterwood LM, Kamilar GM, Beal PA: **Synthetic substrate analogs for the RNA-editing adenosine deaminase ADAR-2.** *Nucleic Acids Res* 1999, **27**(14):2912-2917.
  34. Jiang YH, Bressler J, Beaudet AL: **Epigenetics and human disease.** *Annu Rev Genomics Hum Genet* 2004, **5**:479-510.
  35. Paoloni-Giacobino A, Chaillet JR: **Genomic imprinting and assisted reproduction.** *Reprod Health* 2004, **1**(1):6.
  36. Nicholls RD, Knepper JL: **Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes.** *Annu Rev Genomics Hum Genet* 2001, **2**:153-175.
  37. Preece MA, Moore GE: **Genomic imprinting, uniparental disomy and foetal growth.** *Trends Endocrinol Metab* 2000, **11**(7):270-275.
  38. Surani MA: **Imprinting and the initiation of gene silencing in the germ line.** *Cell* 1998, **93**(3):309-312.
  39. Bickmore WA, van der Maarel SM: **Perturbations of chromatin structure in human genetic disease: recent advances.** *Hum Mol Genet* 2003, **12 Spec No 2**:R207-213.
  40. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY: **Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2.** *Nat Genet* 1999, **23**(2):185-188.
  41. Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H, Higgs DR: **Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation.** *Nat Genet* 2000, **24**(4):368-371.
  42. Primerano B, Tassone F, Hagerman RJ, Hagerman P, Amaldi F, Bagni C: **Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations.** *Rna* 2002, **8**(12):1482-1488.
  43. Jin P, Zarnescu DC, Ceman S, Nakamoto M, Mowrey J, Jongens TA, Nelson DL, Moses K, Warren ST: **Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway.** *Nat Neurosci* 2004, **7**(2):113-117.
  44. Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, Muralidhar B, Bilienska B, Krajewska-Walasek M, Driscoll DJ *et al*: **Imprinting-mutation mechanisms in Prader-Willi syndrome.** *Am J Hum Genet* 1999, **64**(2):397-413.

45. Cox GF, Burger J, Lip V, Mau UA, Sperling K, Wu BL, Horsthemke B: **Intracytoplasmic sperm injection may increase the risk of imprinting defects.** *Am J Hum Genet* 2002, **71**(1):162-164.
46. Halliday J, Oke K, Breheny S, Algar E, D JA: **Beckwith-Wiedemann syndrome and IVF: a case-control study.** *Am J Hum Genet* 2004, **75**(3):526-528.
47. DeBaun MR, Niemitz EL, Feinberg AP: **Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19.** *Am J Hum Genet* 2003, **72**(1):156-160.
48. Maher ER, Afnan M, Barratt CL: **Epigenetic risks related to assisted reproductive technologies: epigenetics, imprinting, ART and icebergs?** *Hum Reprod* 2003, **18**(12):2508-2511.
49. Orstavik KH, Eiklid K, van der Hagen CB, Spetalen S, Kierulf K, Skjeldal O, Buiting K: **Another case of imprinting defect in a girl with Angelman syndrome who was conceived by intracytoplasmic semen injection.** *Am J Hum Genet* 2003, **72**(1):218-219.
50. Soejima H, Wagstaff J: **Imprinting centers, chromatin structure, and disease.** *J Cell Biochem* 2005, **95**(2):226-233.
51. Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, Mukai T, Niikawa N, Ogawa M, Wagstaff J, Kishino T: **Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a.** *Hum Mol Genet* 2003, **12**(8):837-847.
52. Shemer R, Hershko AY, Perk J, Mostoslavsky R, Tsuberi B, Cedar H, Buiting K, Razin A: **The imprinting box of the Prader-Willi/Angelman syndrome domain.** *Nat Genet* 2000, **26**(4):440-443.
53. Feuk L, Carson AR, Scherer SW: **Structural variation in the human genome.** *Nat Rev Genet* 2006, **7**(2):85-97.
54. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Se Graves R *et al*: **Segmental duplications and copy-number variation in the human genome.** *Am J Hum Genet* 2005, **77**(1):78-88.
55. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W *et al*: **Global variation in copy number in the human genome.** *Nature* 2006, **444**(7118):444-454.
56. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME *et al*: **Copy number variation: new insights in genome diversity.** *Genome Res* 2006, **16**(8):949-961.
57. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M *et al*: **Large-scale copy number polymorphism in the human genome.** *Science* 2004, **305**(5683):525-528.
58. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**(9):949-951.
59. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ *et al*: **Common deletion polymorphisms in the human genome.** *Nat Genet* 2006, **38**(1):86-92.
60. Emanuel BS, Shaikh TH: **Segmental duplications: an 'expanding' role in genomic instability and disease.** *Nat Rev Genet* 2001, **2**(10):791-800.
61. Samonte RV, Eichler EE: **Segmental duplications and the evolution of the primate genome.** *Nat Rev Genet* 2002, **3**(1):65-72.
62. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: **Recent segmental duplications in the human genome.** *Science* 2002, **297**(5583):1003-1007.
63. Cheung J, Estivill X, Khaja R, MacDonald JR, Lau K, Tsui LC, Scherer SW: **Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence.** *Genome Biol* 2003, **4**(4):R25.
64. Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X: **Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements.** *Hum Mol Genet* 2003, **12**(17):2201-2208.
65. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S *et al*: **A genome-wide comparison of recent chimpanzee and human segmental duplications.** *Nature* 2005, **437**(7055):88-93.
66. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE *et al*: **A comprehensive analysis of common copy-number variations in the human genome.** *Am J Hum Genet* 2007, **80**(1):91-104.

67. Feuk L, Marshall CR, Wintle RF, Scherer SW: **Structural variants: changing the landscape of chromosomes and design of disease studies.** *Hum Mol Genet* 2006, **15 Spec No 1**:R57-66.
68. Zody MC, Garber M, Sharpe T, Young SK, Rowen L, O'Neill K, Whittaker CA, Kamal M, Chang JL, Cuomo CA *et al*: **Analysis of the DNA sequence and duplication history of human chromosome 15.** *Nature* 2006, **440**(7084):671-675.
69. Pujana MA, Nadal M, Gratacos M, Peral B, Csiszar K, Gonzalez-Sarmiento R, Sumoy L, Estivill X: **Additional complexity on human chromosome 15q: identification of a set of newly recognized duplicons (LCR15) on 15q11-q13, 15q24, and 15q26.** *Genome Res* 2001, **11**(1):98-111.
70. Pujana MA, Nadal M, Guitart M, Armengol L, Gratacos M, Estivill X: **Human chromosome 15q11-q14 regions of rearrangements contain clusters of LCR15 duplicons.** *Eur J Hum Genet* 2002, **10**(1):26-35.
71. Locke DP, Jiang Z, Pertz LM, Misceo D, Archidiacono N, Eichler EE: **Molecular evolution of the human chromosome 15 pericentromeric region.** *Cytogenet Genome Res* 2005, **108**(1-3):73-82.
72. Amos-Landgraf JM, Ji Y, Gottlieb W, Depinet T, Wandstrat AE, Cassidy SB, Driscoll DJ, Rogan PK, Schwartz S, Nicholls RD: **Chromosome breakage in the Prader-Willi and Angelman syndromes involves recombination between large, transcribed repeats at proximal and distal breakpoints.** *Am J Hum Genet* 1999, **65**(2):370-386.
73. Wandstrat AE, Leana-Cox J, Jenkins L, Schwartz S: **Molecular cytogenetic evidence for a common breakpoint in the largest inverted duplications of chromosome 15.** *Am J Hum Genet* 1998, **62**(4):925-936.
74. Browne CE, Dennis NR, Maher E, Long FL, Nicholson JC, Sillibourne J, Barber JC: **Inherited interstitial duplications of proximal 15q: genotype-phenotype correlations.** *Am J Hum Genet* 1997, **61**(6):1342-1352.
75. Schinzel AA, Brecevic L, Bernasconi F, Binkert F, Berthet F, Wuilloud A, Robinson WP: **Intrachromosomal triplication of 15q11-q13.** *J Med Genet* 1994, **31**(10):798-803.
76. Horsthemke B, Maat-Kievit A, Slegers E, van den Ouweland A, Buiting K, Lich C, Mollevanger P, Beverstock G, Gillessen-Kaesbach G, Schwanz G: **Familial translocations involving 15q11-q13 can give rise to interstitial deletions causing Prader-Willi or Angelman syndrome.** *J Med Genet* 1996, **33**(10):848-851.
77. Gimelli G, Pujana MA, Patricelli MG, Russo S, Giardino D, Larizza L, Cheung J, Armengol L, Schinzel A, Estivill X *et al*: **Genomic inversions of human chromosome 15q11-q13 in mothers of Angelman syndrome patients with class II (BP2/3) deletions.** *Hum Mol Genet* 2003, **12**(8):849-858.
78. Conrad DF, Andrews TD, Carter NP, Hurler ME, Pritchard JK: **A high-resolution survey of deletion polymorphism in the human genome.** *Nat Genet* 2006, **38**(1):75-81.
79. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM *et al*: **Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome.** *Am J Hum Genet* 2006, **79**(2):275-290.
80. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D *et al*: **Fine-scale structural variation of the human genome.** *Nat Genet* 2005, **37**(7):727-732.
81. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE: **An initial map of insertion and deletion (INDEL) variation in the human genome.** *Genome Res* 2006, **16**(9):1182-1190.
82. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L *et al*: **Mapping autism risk loci using genetic linkage and chromosomal rearrangements.** *Nat Genet* 2007, **39**(3):319-328.
83. Eichler EE, Clark RA, She X: **An assessment of the sequence gaps: unfinished business in a finished human genome.** *Nat Rev Genet* 2004, **5**(5):345-354.
84. Coon H: **Current perspectives on the genetic analysis of autism.** *Am J Med Genet C Semin Med Genet* 2006, **142**(1):24-32.
85. Muhle R, Trentacoste SV, Rapin I: **The genetics of autism.** *Pediatrics* 2004, **113**(5):e472-486.
86. Veenstra-Vanderweele J, Christian SL, Cook EH, Jr.: **Autism as a paradigmatic complex genetic disorder.** *Annu Rev Genomics Hum Genet* 2004, **5**:379-405.
87. Folstein SE, Rosen-Sheidley B: **Genetics of autism: complex aetiology for a heterogeneous disorder.** *Nat Rev Genet* 2001, **2**(12):943-955.
88. Freitag CM: **The genetics of autistic disorders and its clinical relevance: a review of the literature.** *Mol Psychiatry* 2007, **12**(1):2-22.

89. Medical Council Research: **Review of autism research**. 2001.
90. Chakrabarti S, Fombonne E: **Pervasive developmental disorders in preschool children: confirmation of high prevalence**. *Am J Psychiatry* 2005, **162**(6):1133-1141.
91. Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, Palmer AA, Grunn A, Juo SH, Terwilliger JD, Liu J *et al*: **A genomewide screen of 345 families for autism-susceptibility loci**. *Am J Hum Genet* 2003, **73**(4):886-897.
92. Turner M, Barnby G, Bailey A: **Genetic clues to the biological basis of autism**. *Mol Med Today* 2000, **6**(6):238-244.
93. Vorstman JA, Staal WG, van Daalen E, van Engeland H, Hochstenbach PF, Franke L: **Identification of novel autism candidate regions through analysis of reported cytogenetic abnormalities associated with autism**. *Mol Psychiatry* 2006, **11**(1):1, 18-28.
94. Gutknecht L: **Full-genome scans with autistic disorder: a review**. *Behav Genet* 2001, **31**(1):113-123.
95. Lauritsen M, Mors O, Mortensen PB, Ewald H: **Infantile autism and associated autosomal chromosome abnormalities: a register-based study and a literature survey**. *J Child Psychol Psychiatry* 1999, **40**(3):335-345.
96. Gillberg C: **Chromosomal disorders and autism**. *J Autism Dev Disord* 1998, **28**(5):415-425.
97. Jacquemont ML, Sanlaville D, Redon R, Raoul O, Cormier-Daire V, Lyonnet S, Amiel J, Le Merrer M, Heron D, de Blois MC *et al*: **Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders**. *J Med Genet* 2006, **43**(11):843-849.
98. Rabionet R, Jaworski JM, Ashley-Koch AE, Martin ER, Sutcliffe JS, Haines JL, DeLong GR, Abramson RK, Wright HH, Cuccaro ML *et al*: **Analysis of the autism chromosome 2 linkage region: GAD1 and other candidate genes**. *Neurosci Lett* 2004, **372**(3):209-214.
99. Hamilton SP, Woo JM, Carlson EJ, Ghanem N, Ekker M, Rubenstein JL: **Analysis of four DLX homeobox genes in autistic probands**. *BMC Genet* 2005, **6**:52.
100. Jamain S, Betancur C, Quach H, Philippe A, Fellous M, Giros B, Gillberg C, Leboyer M, Bourgeron T: **Linkage and association of the glutamate receptor 6 gene with autism**. *Mol Psychiatry* 2002, **7**(3):302-310.
101. Conciatori M, Stodgell CJ, Hyman SL, O'Bara M, Militeri R, Bravaccio C, Trillo S, Montecchi F, Schneider C, Melmed R *et al*: **Association between the HOXA1 A218G polymorphism and increased head circumference in patients with autism**. *Biol Psychiatry* 2004, **55**(4):413-419.
102. Persico AM, D'Agruma L, Maiorano N, Totaro A, Militeri R, Bravaccio C, Wassink TH, Schneider C, Melmed R, Trillo S *et al*: **Reelin gene alleles and haplotypes as a factor predisposing to autistic disorder**. *Mol Psychiatry* 2001, **6**(2):150-159.
103. Serajee FJ, Zhong H, Nabi R, Huq AH: **The metabotropic glutamate receptor 8 gene at 7q31: partial duplication and possible association with autism**. *J Med Genet* 2003, **40**(4):e42.
104. Skaar DA, Shao Y, Haines JL, Stenger JE, Jaworski J, Martin ER, DeLong GR, Moore JH, McCauley JL, Sutcliffe JS *et al*: **Analysis of the RELN gene as a genetic risk factor for autism**. *Mol Psychiatry* 2005, **10**(6):563-571.
105. Zhang H, Liu X, Zhang C, Mundo E, Macciardi F, Grayson DR, Guidotti AR, Holden JJ: **Reelin gene alleles and susceptibility to autism spectrum disorders**. *Mol Psychiatry* 2002, **7**(9):1012-1017.
106. Li J, Nguyen L, Gleason C, Lotspeich L, Spiker D, Risch N, Myers RM: **Lack of evidence for an association between WNT2 and RELN polymorphisms and autism**. *Am J Med Genet B Neuropsychiatr Genet* 2004, **126**(1):51-57.
107. Buxbaum JD, Silverman JM, Smith CJ, Greenberg DA, Kilifarski M, Reichert J, Cook EH, Jr., Fang Y, Song CY, Vitale R: **Association between a GABRB3 polymorphism and autism**. *Mol Psychiatry* 2002, **7**(3):311-316.
108. Bass MP, Menold MM, Wolpert CM, Donnelly SL, Ravan SA, Hauser ER, Maddox LO, Vance JM, Abramson RK, Wright HH *et al*: **Genetic studies in autistic disorder and chromosome 15**. *Neurogenetics* 2000, **2**(4):219-226.
109. Cook EH, Jr., Courchesne RY, Cox NJ, Lord C, Gonen D, Guter SJ, Lincoln A, Nix K, Haas R, Leventhal BL *et al*: **Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers**. *Am J Hum Genet* 1998, **62**(5):1077-1083.
110. Curran S, Powell J, Neale BM, Dworzynski K, Li T, Murphy D, Bolton PF: **An association analysis of candidate genes on chromosome 15 q11-13 and autism spectrum disorder**. *Mol Psychiatry* 2006, **11**(8):709-713.

111. Curran S, Roberts S, Thomas S, Veltman M, Browne J, Medda E, Pickles A, Sham P, Bolton PF: **An association analysis of microsatellite markers across the Prader-Willi/Angelman critical region on chromosome 15 (q11-13) and autism spectrum disorder.** *Am J Med Genet B Neuropsychiatr Genet* 2005, **137**(1):25-28.
112. Ma DQ, Whitehead PL, Menold MM, Martin ER, Ashley-Koch AE, Mei H, Ritchie MD, DeLong GR, Abramson RK, Wright HH *et al*: **Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism.** *Am J Hum Genet* 2005, **77**(3):377-388.
113. McCauley JL, Olson LM, Delahanty R, Amin T, Nurmi EL, Organ EL, Jacobs MM, Folstein SE, Haines JL, Sutcliffe JS: **A linkage disequilibrium map of the 1-Mb 15q12 GABA(A) receptor subunit cluster and association to autism.** *Am J Med Genet B Neuropsychiatr Genet* 2004, **131**(1):51-59.
114. Nurmi EL, Amin T, Olson LM, Jacobs MM, McCauley JL, Lam AY, Organ EL, Folstein SE, Haines JL, Sutcliffe JS: **Dense linkage disequilibrium mapping in the 15q11-q13 maternal expression domain yields evidence for association in autism.** *Mol Psychiatry* 2003, **8**(6):624-634, 570.
115. Kim SJ, Herzing LB, Veenstra-VanderWeele J, Lord C, Courchesne R, Leventhal BL, Ledbetter DH, Courchesne E, Cook EH, Jr.: **Mutation screening and transmission disequilibrium study of ATP10C in autism.** *Am J Med Genet* 2002, **114**(2):137-143.
116. Devlin B, Cook EH, Jr., Coon H, Dawson G, Grigorenko EL, McMahon W, Minshew N, Pauls D, Smith M, Spence MA *et al*: **Autism and the serotonin transporter: the long and short of it.** *Mol Psychiatry* 2005, **10**(12):1110-1116.
117. Mulder EJ, Anderson GM, Kema IP, Brugman AM, Ketelaars CE, de Bildt A, van Lang ND, den Boer JA, Minderaa RB: **Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder.** *Am J Med Genet B Neuropsychiatr Genet* 2005, **133**(1):93-96.
118. Sutcliffe JS, Delahanty RJ, Prasad HC, McCauley JL, Han Q, Jiang L, Li C, Folstein SE, Blakely RD: **Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors.** *Am J Hum Genet* 2005, **77**(2):265-279.
119. Wu S, Guo Y, Jia M, Ruan Y, Shuang M, Liu J, Gong X, Zhang Y, Yang J, Yang X *et al*: **Lack of evidence for association between the serotonin transporter gene (SLC6A4) polymorphisms and autism in the Chinese trios.** *Neurosci Lett* 2005, **381**(1-2):1-5.
120. Blasi F, Bacchelli E, Pesaresi G, Carone S, Bailey AJ, Maestrini E: **Absence of coding mutations in the X-linked genes neuroligin 3 and neuroligin 4 in individuals with autism from the IMGSAC collection.** *Am J Med Genet B Neuropsychiatr Genet* 2006, **141**(3):220-221.
121. Gauthier J, Bonnel A, St-Onge J, Karemera L, Laurent S, Mottron L, Fombonne E, Joober R, Rouleau GA: **NLGN3/NLGN4 gene mutations are not responsible for autism in the Quebec population.** *Am J Med Genet B Neuropsychiatr Genet* 2005, **132**(1):74-75.
122. Yan J, Oliveira G, Coutinho A, Yang C, Feng J, Katz C, Sram J, Bockholt A, Jones IR, Craddock N *et al*: **Analysis of the neuroligin 3 and 4 genes in autism and other neuropsychiatric patients.** *Mol Psychiatry* 2005, **10**(4):329-332.
123. Li H, Yamagata T, Mori M, Yasuhara A, Momoi MY: **Mutation analysis of methyl-CpG binding protein family genes in autistic patients.** *Brain Dev* 2005, **27**(5):321-325.
124. Shibayama A, Cook EH, Jr., Feng J, Glanzmann C, Yan J, Craddock N, Jones IR, Goldman D, Heston LL, Sommer SS: **MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism.** *Am J Med Genet B Neuropsychiatr Genet* 2004, **128**(1):50-53.
125. Bartlett CW, Goedken R, Vieland VJ: **Effects of updating linkage evidence across subsets of data: reanalysis of the autism genetic resource exchange data set.** *Am J Hum Genet* 2005, **76**(4):688-695.
126. McCauley JL, Li C, Jiang L, Olson LM, Crockett G, Gainer K, Folstein SE, Haines JL, Sutcliffe JS: **Genome-wide and Ordered-Subset linkage analyses provide support for autism loci on 17q and 19p with evidence of phenotypic and interlocus genetic correlates.** *BMC Med Genet* 2005, **6**:1.
127. **A genomewide screen for autism: strong evidence for linkage to chromosomes 2q, 7q, and 16p.** *Am J Hum Genet* 2001, **69**(3):570-581.
128. **Further characterization of the autism susceptibility locus AUTS1 on chromosome 7q.** *Hum Mol Genet* 2001, **10**(9):973-982.
129. Philippe A, Martinez M, Guilloud-Bataille M, Gillberg C, Rastam M, Sponheim E, Coleman M, Zappella M, Aschauer H, Van Maldergem L *et al*: **Genome-wide scan for autism susceptibility genes. Paris Autism Research International Sibpair Study.** *Hum Mol Genet* 1999, **8**(5):805-812.

130. Lamb JA, Barnby G, Bonora E, Sykes N, Bacchelli E, Blasi F, Maestrini E, Broxholme J, Tzenova J, Weeks D *et al*: **Analysis of IMGSAC autism susceptibility loci: evidence for sex limited and parent of origin specific effects.** *J Med Genet* 2005, **42**(2):132-137.
131. Alarcon M, Cantor RM, Liu J, Gilliam TC, Geschwind DH: **Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families.** *Am J Hum Genet* 2002, **70**(1):60-71.
132. Shao Y, Raiford KL, Wolpert CM, Cope HA, Ravan SA, Ashley-Koch AA, Abramson RK, Wright HH, DeLong RG, Gilbert JR *et al*: **Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder.** *Am J Hum Genet* 2002, **70**(4):1058-1061.
133. Molloy CA, Keddache M, Martin LJ: **Evidence for linkage on 21q and 7q in a subset of autism characterized by developmental regression.** *Mol Psychiatry* 2005, **10**(8):741-746.
134. Buxbaum JD, Silverman J, Keddache M, Smith CJ, Hollander E, Ramoz N, Reichert JG: **Linkage analysis for autism in a subset families with obsessive-compulsive behaviors: evidence for an autism susceptibility gene on chromosome 1 and further support for susceptibility genes on chromosome 6 and 19.** *Mol Psychiatry* 2004, **9**(2):144-150.
135. Nurmi EL, Dowd M, Tadevosyan-Leyfer O, Haines JL, Folstein SE, Sutcliffe JS: **Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13.** *J Am Acad Child Adolesc Psychiatry* 2003, **42**(7):856-863.
136. Shao Y, Cuccaro ML, Hauser ER, Raiford KL, Menold MM, Wolpert CM, Ravan SA, Elston L, Decena K, Donnelly SL *et al*: **Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes.** *Am J Hum Genet* 2003, **72**(3):539-548.
137. Stone JL, Merriman B, Cantor RM, Yonan AL, Gilliam TC, Geschwind DH, Nelson SF: **Evidence for sex-specific risk alleles in autism spectrum disorder.** *Am J Hum Genet* 2004, **75**(6):1117-1123.
138. Lauritsen MB, Als TD, Dahl HA, Flint TJ, Wang AG, Vang M, Kruse TA, Ewald H, Mors O: **A genome-wide search for alleles and haplotypes associated with autism and related pervasive developmental disorders on the Faroe Islands.** *Mol Psychiatry* 2006, **11**(1):37-46.
139. Talebizadeh Z, Lam DY, Theodoro MF, Bittel DC, Lushington GH, Butler MG: **Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism.** *J Med Genet* 2006, **43**(5):e21.
140. Samaco RC, Nagarajan RP, Braunschweig D, LaSalle JM: **Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders.** *Hum Mol Genet* 2004, **13**(6):629-639.
141. Tordjman S, Drapier D, Bonnot O, Graignic R, Fortes S, Cohen D, Millet B, Laurent C, Roubertoux PL: **Animal Models Relevant to Schizophrenia and Autism: Validity and Limitations.** *Behav Genet* 2007, **37**(1):61-78.
142. Modahl C, Green L, Fein D, Morris M, Waterhouse L, Feinstein C, Levin H: **Plasma oxytocin levels in autistic children.** *Biol Psychiatry* 1998, **43**(4):270-277.
143. Insel TR, O'Brien DJ, Leckman JF: **Oxytocin, vasopressin, and autism: is there a connection?** *Biol Psychiatry* 1999, **45**(2):145-157.
144. Winslow JT, Insel TR: **The social deficits of the oxytocin knockout mouse.** *Neuropeptides* 2002, **36**(2-3):221-229.
145. Li H, Yamagata T, Mori M, Momoi MY: **Absence of causative mutations and presence of autism-related allele in FOXP2 in Japanese autistic patients.** *Brain Dev* 2005, **27**(3):207-210.
146. Gong X, Jia M, Ruan Y, Shuang M, Liu J, Wu S, Guo Y, Yang J, Ling Y, Yang X *et al*: **Association between the FOXP2 gene and autistic disorder in Chinese population.** *Am J Med Genet B Neuropsychiatr Genet* 2004, **127**(1):113-116.
147. Bonora E, Bacchelli E, Levy ER, Blasi F, Marlow A, Monaco AP, Maestrini E: **Mutation screening and imprinting analysis of four candidate genes for autism in the 7q32 region.** *Mol Psychiatry* 2002, **7**(3):289-301.
148. Gharani N, Benayed R, Mancuso V, Brzustowicz LM, Millonig JH: **Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder.** *Mol Psychiatry* 2004, **9**(5):474-484.
149. Wassink TH, Piven J, Vieland VJ, Huang J, Swiderski RE, Pietila J, Braun T, Beck G, Folstein SE, Haines JL *et al*: **Evidence supporting WNT2 as an autism susceptibility gene.** *Am J Med Genet* 2001, **105**(5):406-413.
150. Lijam N, Paylor R, McDonald MP, Crawley JN, Deng CX, Herrup K, Stevens KE, Maccaferri G, McBain CJ, Sussman DJ *et al*: **Social interaction and sensorimotor gating abnormalities in mice lacking Dvl1.** *Cell* 1997, **90**(5):895-905.
151. Fatemi SH, Stary JM, Halt AR, Realmuto GR: **Dysregulation of Reelin and Bcl-2 proteins in autistic cerebellum.** *J Autism Dev Disord* 2001, **31**(6):529-535.

152. Sadakata T, Kakegawa W, Mizoguchi A, Washida M, Katoh-Semba R, Shutoh F, Okamoto T, Nakashima H, Kimura K, Tanaka M *et al*: **Impaired cerebellar development and function in mice lacking CAPS2, a protein involved in neurotrophin release.** *J Neurosci* 2007, **27**(10):2472-2482.
153. Anderson GM, Freedman DX, Cohen DJ, Volkmar FR, Hoder EL, McPhedran P, Minderara RB, Hansen CR, Young JG: **Whole blood serotonin in autistic and normal subjects.** *J Child Psychol Psychiatry* 1987, **28**(6):885-900.
154. Beyer KS, Blasi F, Bacchelli E, Klauck SM, Maestrini E, Poustka A: **Mutation analysis of the coding sequence of the MECP2 gene in infantile autism.** *Hum Genet* 2002, **111**(4-5):305-309.
155. Carney RM, Wolpert CM, Ravan SA, Shahbazian M, Ashley-Koch A, Cuccaro ML, Vance JM, Pericak-Vance MA: **Identification of MeCP2 mutations in a series of females with autistic disorder.** *Pediatr Neurol* 2003, **28**(3):205-211.
156. Hettinger JA, Liu X, Holden JJ: **The G22A Polymorphism of the ADA Gene and Susceptibility to Autism Spectrum Disorders.** *J Autism Dev Disord* 2007.
157. Bottini N, De Luca D, Saccucci P, Fiumara A, Elia M, Porfirio MC, Lucarelli P, Curatolo P: **Autism: evidence of association with adenosine deaminase genetic polymorphism.** *Neurogenetics* 2001, **3**(2):111-113.
158. Bolton PF, Dennis NR, Browne CE, Thomas NS, Veltman MW, Thompson RJ, Jacobs P: **The phenotypic manifestations of interstitial duplications of proximal 15q with special reference to the autistic spectrum disorders.** *Am J Med Genet* 2001, **105**(8):675-685.
159. Bolton PF, Veltman MW, Weisblatt E, Holmes JR, Thomas NS, Youngs SA, Thompson RJ, Roberts SE, Dennis NR, Browne CE *et al*: **Chromosome 15q11-13 abnormalities and other medical conditions in individuals with autism spectrum disorders.** *Psychiatr Genet* 2004, **14**(3):131-137.
160. Cook EH, Jr., Lindgren V, Leventhal BL, Courchesne R, Lincoln A, Shulman C, Lord C, Courchesne E: **Autism or atypical autism in maternally but not paternally derived proximal 15q duplication.** *Am J Hum Genet* 1997, **60**(4):928-934.
161. Schroer RJ, Phelan MC, Michaelis RC, Crawford EC, Skinner SA, Cuccaro M, Simensen RJ, Bishop J, Skinner C, Fender D *et al*: **Autism and maternally derived aberrations of chromosome 15q.** *Am J Med Genet* 1998, **76**(4):327-336.
162. Lukusa T, Vermeesch JR, Holvoet M, Fryns JP, Devriendt K: **Deletion 2q37.3 and autism: molecular cytogenetic mapping of the candidate region for autistic disorder.** *Genet Couns* 2004, **15**(3):293-301.
163. Ghaziuddin M, Burmeister M: **Deletion of chromosome 2q37 and autism: a distinct subtype?** *J Autism Dev Disord* 1999, **29**(3):259-263.
164. Ashley-Koch A, Wolpert CM, Menold MM, Zaeem L, Basu S, Donnelly SL, Ravan SA, Powell CM, Qumsiyeh MB, Aylsworth AS *et al*: **Genetic studies of autistic disorder and chromosome 7.** *Genomics* 1999, **61**(3):227-236.
165. Vincent JB, Herbrick JA, Gurling HM, Bolton PF, Roberts W, Scherer SW: **Identification of a novel gene on chromosome 7q31 that is interrupted by a translocation breakpoint in an autistic individual.** *Am J Hum Genet* 2000, **67**(2):510-514.
166. Fine SE, Weissman A, Gerdes M, Pinto-Martin J, Zackai EH, McDonald-McGinn DM, Emanuel BS: **Autism spectrum disorders and symptoms in children with molecularly confirmed 22q11.2 deletion syndrome.** *J Autism Dev Disord* 2005, **35**(4):461-470.
167. Manning MA, Cassidy SB, Clericuzio C, Cherry AM, Schwartz S, Hudgins L, Enns GM, Hoyme HE: **Terminal 22q deletion syndrome: a newly recognized cause of speech and language disability in the autism spectrum.** *Pediatrics* 2004, **114**(2):451-457.
168. Repetto GM, White LM, Bader PJ, Johnson D, Knoll JH: **Interstitial duplications of chromosome region 15q11q13: clinical and molecular characterization.** *Am J Med Genet* 1998, **79**(2):82-89.
169. Borgatti R, Piccinelli P, Passoni D, Dalpra L, Miozzo M, Micheli R, Gagliardi C, Balottin U: **Relationship between clinical and genetic features in "inverted duplicated chromosome 15" patients.** *Pediatr Neurol* 2001, **24**(2):111-116.
170. Clayton-Smith J, Laan L: **Angelman syndrome: a review of the clinical and genetic aspects.** *J Med Genet* 2003, **40**(2):87-95.
171. State MW, Dykens EM: **Genetics of childhood disorders: XV. Prader-Willi syndrome: genes, brain, and behavior.** *J Am Acad Child Adolesc Psychiatry* 2000, **39**(6):797-800.
172. **Human Mutation Database** [<http://www.hgmd.cf.ac.uk/ac/index.php>]
173. Wirth J, Back E, Huttenhofer A, Nothwang HG, Lich C, Gross S, Menzel C, Schinzel A, Kioschis P, Tommerup N *et al*: **A translocation breakpoint cluster disrupts the newly defined 3' end of the SNURF-SNRPN transcription unit on chromosome 15.** *Hum Mol Genet* 2001, **10**(3):201-210.

174. Shao Y, Wolpert CM, Raiford KL, Menold MM, Donnelly SL, Ravan SA, Bass MP, McClain C, von Wendt L, Vance JM *et al*: **Genomic screen and follow-up analysis for autistic disorder**. *Am J Med Genet* 2002, **114**(1):99-105.
175. Barrett S, Beck JC, Bernier R, Bisson E, Braun TA, Casavant TL, Childress D, Folstein SE, Garcia M, Gardiner MB *et al*: **An autosomal genomic screen for autism. Collaborative linkage study of autism**. *Am J Med Genet* 1999, **88**(6):609-615.
176. Menold MM, Shao Y, Wolpert CM, Donnelly SL, Raiford KL, Martin ER, Ravan SA, Abramson RK, Wright HH, DeLong GR *et al*: **Association analysis of chromosome 15 gabaa receptor subunit genes in autistic disorder**. *J Neurogenet* 2001, **15**(3-4):245-259.
177. Samaco RC, Hogart A, LaSalle JM: **Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3**. *Hum Mol Genet* 2005, **14**(4):483-492.
178. Longo I, Russo L, Meloni I, Ricci I, Ariani F, Pescucci C, Giordano CT, Canitano R, Hayek G, Zappella M *et al*: **Three Rett patients with both MECP2 mutation and 15q11-13 rearrangements**. *Eur J Hum Genet* 2004, **12**(8):682-685.
179. Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J: **A new frontier for molecular medicine: noncoding RNAs**. *Biochim Biophys Acta* 2005, **1756**(1):65-75.
180. Mattick JS: **Non-coding RNAs: the architects of eukaryotic complexity**. *EMBO Rep* 2001, **2**(11):986-991.
181. Mattick JS: **Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms**. *Bioessays* 2003, **25**(10):930-939.
182. Shabalina SA, Spiridonov NA: **The mammalian transcriptome and the function of non-coding DNA sequences**. *Genome Biol* 2004, **5**(4):105.
183. Huttenhofer A, Brosius J, Bachellerie JP: **RNomics: identification and function of small, non-messenger RNAs**. *Curr Opin Chem Biol* 2002, **6**(6):835-843.
184. Grivna ST, Beyret E, Wang Z, Lin H: **A novel class of small RNAs in mouse spermatogenic cells**. *Genes Dev* 2006, **20**(13):1709-1714.
185. Grivna ST, Pyhtila B, Lin H: **MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis**. *Proc Natl Acad Sci U S A* 2006, **103**(36):13415-13420.
186. Girard A, Sachidanandam R, Hannon GJ, Carmell MA: **A germline-specific class of small RNAs binds mammalian Piwi proteins**. *Nature* 2006, **442**(7099):199-202.
187. Kim VN: **Small RNAs just got bigger: Piwi-interacting RNAs (piRNAs) in mammalian testes**. *Genes Dev* 2006, **20**(15):1993-1997.
188. Mattick JS, Makunin IV: **Non-coding RNA**. *Hum Mol Genet* 2006, **15 Spec No 1**:R17-29.
189. Costa FF: **Non-coding RNAs: Lost in translation?** *Gene* 2006.
190. Sutcliffe JG, Milner RJ, Bloom FE, Lerner RA: **Common 82-nucleotide sequence unique to brain RNA**. *Proc Natl Acad Sci U S A* 1982, **79**(16):4942-4946.
191. Watson JB, Sutcliffe JG: **Primate brain-specific cytoplasmic transcript of the Alu repeat family**. *Mol Cell Biol* 1987, **7**(9):3324-3327.
192. Mineno J, Okamoto S, Ando T, Sato M, Chono H, Izu H, Takayama M, Asada K, Mirochnitchenko O, Inouye M *et al*: **The expression profile of microRNAs in mouse embryos**. *Nucleic Acids Res* 2006, **34**(6):1765-1771.
193. Strauss WM, Chen C, Lee CT, Ridzon D: **Nonrestrictive developmental regulation of microRNA gene expression**. *Mamm Genome* 2006, **17**(8):833-840.
194. Kong Y, Han JH: **MicroRNA: biological and computational perspective**. *Genomics Proteomics Bioinformatics* 2005, **3**(2):62-72.
195. Lin SL, Miller JD, Ying SY: **Intronic MicroRNA (miRNA)**. *J Biomed Biotechnol* 2006, **2006**(4):26818.
196. Gu J, He T, Pei Y, Li F, Wang X, Zhang J, Zhang X, Li Y: **Primary transcripts and expressions of mammal intergenic microRNAs detected by mapping ESTs to their flanking sequences**. *Mamm Genome* 2006, **17**(10):1033-1041.
197. Smalheiser NR, Torvik VI: **Mammalian microRNAs derived from genomic repeats**. *Trends Genet* 2005, **21**(6):322-326.
198. Seitz H, Royo H, Lin SP, Youngson N, Ferguson-Smith AC, Cavaille J: **Imprinted small RNA genes**. *Biol Chem* 2004, **385**(10):905-911.
199. Gabory A, Ripoché MA, Yoshimizu T, Dandolo L: **The H19 gene: regulation and function of a non-coding RNA**. *Cytogenet Genome Res* 2006, **113**(1-4):188-193.

200. Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G: **In search of antisense.** *Trends Biochem Sci* 2004, **29**(2):88-94.
201. Numata K, Okada Y, Saito R, Kiyosawa H, Kanai A, Tomita M: **Comparative analysis of cis-encoded antisense RNAs in eukaryotes.** *Gene* 2006.
202. Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, Shoshan A, Diber A, Biton S, Tamir Y, Khosravi R *et al*: **Widespread occurrence of antisense transcription in the human genome.** *Nat Biotechnol* 2003, **21**(4):379-386.
203. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: **Identification of novel genes coding for small expressed RNAs.** *Science* 2001, **294**(5543):853-858.
204. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T: **New microRNAs from mouse and human.** *Rna* 2003, **9**(2):175-179.
205. Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, Vos J, Verloop R, van de Wetering M, Guryev V, Takada S *et al*: **Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis.** *Genome Res* 2006.
206. Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RH: **Diversity of microRNAs in human and chimpanzee brain.** *Nat Genet* 2006.
207. Rivas E, Eddy SR: **Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs.** *Bioinformatics* 2000, **16**(7):583-605.
208. **Mfold** [<http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi>]
209. Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**(13):3406-3415.
210. **RNAShapes** [<http://bibiserv.techfak.uni-bielefeld.de/rnashapes/submission.html>]
211. Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP: **Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region.** *Hum Mol Genet* 2002, **11**(13):1527-1538.
212. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U, Meiri E *et al*: **Identification of hundreds of conserved and nonconserved human microRNAs.** *Nat Genet* 2005, **37**(7):766-770.
213. Bonnet E, Wuyts J, Rouze P, Van de Peer Y: **Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences.** *Bioinformatics* 2004, **20**(17):2911-2917.
214. Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E: **Phylogenetic shadowing and computational identification of human microRNA genes.** *Cell* 2005, **120**(1):21-24.
215. Sewer A, Paul N, Landgraf P, Aravin A, Pfeffer S, Brownstein MJ, Tuschl T, van Nimwegen E, Zavolan M: **Identification of clustered microRNAs using an ab initio prediction method.** *BMC Bioinformatics* 2005, **6**:267.
216. Rivas E, Eddy SR: **Noncoding RNA gene detection using comparative sequence analysis.** *BMC Bioinformatics* 2001, **2**:8.
217. Washietl S, Hofacker IL, Stadler PF: **Fast and reliable prediction of noncoding RNAs.** *Proc Natl Acad Sci U S A* 2005, **102**(7):2454-2459.
218. Uzilov AV, Keegan JM, Mathews DH: **Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change.** *BMC Bioinformatics* 2006, **7**:173.
219. Lai EC, Tomancak P, Williams RW, Rubin GM: **Computational identification of Drosophila microRNA genes.** *Genome Biol* 2003, **4**(7):R42.
220. Wang X, Zhang J, Li F, Gu J, He T, Zhang X, Li Y: **MicroRNA identification based on sequence and structure alignment.** *Bioinformatics* 2005, **21**(18):3610-3614.
221. Bafna V, Zhang S: **FastR: fast database search tool for non-coding RNA.** *Proc IEEE Comput Syst Bioinform Conf* 2004:52-61.
222. Zhang S, Haas B, Eskin E, Bafna V: **Searching genomes for noncoding RNA using FastR.** *IEEE/ACM Trans Comput Biol Bioinform* 2005, **2**(4):366-379.
223. Berezikov E, Cuppen E, Plasterk RH: **Approaches to microRNA discovery.** *Nat Genet* 2006, **38 Suppl**:S2-7.
224. Ramkissoon SH, Mainwaring LA, Sloand EM, Young NS, Kajigaya S: **Nonisotopic detection of microRNA using digoxigenin labeled RNA probes.** *Mol Cell Probes* 2006, **20**(1):1-4.
225. Krutzfeldt J, Poy MN, Stoffel M: **Strategies to determine the biological function of microRNAs.** *Nat Genet* 2006, **38 Suppl**:S14-19.
226. Lao K, Xu NL, Yeung V, Chen C, Livak KJ, Straus NA: **Multiplexing RT-PCR for the detection of multiple miRNA species in small samples.** *Biochem Biophys Res Commun* 2006, **343**(1):85-89.

227. Kapranov P, Drenkow J, Cheng J, Long J, Helt G, Dike S, Gingeras TR: **Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays.** *Genome Res* 2005, **15**(7):987-997.
228. Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS: **A microRNA array reveals extensive regulation of microRNAs during brain development.** *Rna* 2003, **9**(10):1274-1281.
229. Lim LP, Lau NC, Garrett-Engle P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM: **Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs.** *Nature* 2005, **433**(7027):769-773.
230. Zhang Z, Pang AW, Gerstein M: **Comparative analysis of genome tiling array data reveals many novel primate-specific functional RNAs in human.** *BMC Evol Biol* 2007, **7** Suppl 1:S14.
231. miRBase [<http://microrna.sanger.ac.uk/sequences/index.shtml>]
232. Kim VN, Nam JW: **Genomics of microRNA.** *Trends Genet* 2006, **22**(3):165-173.
233. Presutti C, Rosati J, Vincenti S, Nasi S: **Non coding RNA and brain.** *BMC Neurosci* 2006, **7** Suppl 1:S5.
234. He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation.** *Nat Rev Genet* 2004, **5**(7):522-531.
235. Cao X, Yeo G, Muotri AR, Kuwabara T, Gage FH: **Noncoding RNAs in the Mammalian Central Nervous System.** *Annu Rev Neurosci* 2006.
236. Eulalio A, Behm-Ansmant I, Izaurralde E: **P bodies: at the crossroads of post-transcriptional pathways.** *Nat Rev Mol Cell Biol* 2007, **8**(1):9-22.
237. Carthew RW: **Gene regulation by microRNAs.** *Curr Opin Genet Dev* 2006, **16**(2):203-208.
238. Diana micro test [[http://diana.pcbi.upenn.edu/cgi-bin/micro\\_t.cgi](http://diana.pcbi.upenn.edu/cgi-bin/micro_t.cgi)]
239. miRanda [<http://www.microrna.org>]
240. Krek A, Grun D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M *et al*: **Combinatorial microRNA target predictions.** *Nat Genet* 2005, **37**(5):495-500.
241. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R: **Fast and effective prediction of microRNA/target duplexes.** *Rna* 2004, **10**(10):1507-1517.
242. Saetrom O, Snove O, Jr., Saetrom P: **Weighted sequence motifs as an improved seeding step in microRNA target prediction algorithms.** *Rna* 2005, **11**(7):995-1003.
243. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS: **Human MicroRNA targets.** *PLoS Biol* 2004, **2**(11):e363.
244. Costa FF: **Non-coding RNAs: new players in eukaryotic biology.** *Gene* 2005, **357**(2):83-94.
245. Andersen AA, Panning B: **Epigenetic gene regulation by noncoding RNAs.** *Curr Opin Cell Biol* 2003, **15**(3):281-289.
246. Sleutels F, Zwart R, Barlow DP: **The non-coding Air RNA is required for silencing autosomal imprinted genes.** *Nature* 2002, **415**(6873):810-813.
247. Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K: **Non-coding RNA directed DNA demethylation of Sphk1 CpG island.** *Biochem Biophys Res Commun* 2004, **322**(2):593-600.
248. Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW: **A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex.** *Cell* 1999, **97**(1):17-27.
249. Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang DZ: **The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation.** *Nat Genet* 2006, **38**(2):228-233.
250. Wienholds E, Plasterk RH: **MicroRNA function in animal development.** *FEBS Lett* 2005, **579**(26):5911-5922.
251. Yekta S, Shih IH, Bartel DP: **MicroRNA-directed cleavage of HOXB8 mRNA.** *Science* 2004, **304**(5670):594-596.
252. Zhao Y, Samal E, Srivastava D: **Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis.** *Nature* 2005, **436**(7048):214-220.
253. Chen CZ, Li L, Lodish HF, Bartel DP: **MicroRNAs modulate hematopoietic lineage differentiation.** *Science* 2004, **303**(5654):83-86.
254. Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV, Sun Y, Koo S, Perera RJ, Jain R *et al*: **MicroRNA-143 regulates adipocyte differentiation.** *J Biol Chem* 2004, **279**(50):52361-52365.
255. Poy MN, Eliasson L, Krutzfeldt J, Kuwajima S, Ma X, Macdonald PE, Pfeffer S, Tuschl T, Rajewsky N, Rorsman P *et al*: **A pancreatic islet-specific microRNA regulates insulin secretion.** *Nature* 2004, **432**(7014):226-230.

256. Chang J, Nicolas E, Marks D, Sander C, Lerro A, Buendia MA, Xu C, Mason WS, Moloshok T, Bort R *et al*: **miR-122, a mammalian liver-specific microRNA, is processed from hcr mRNA and may downregulate the high affinity cationic amino acid transporter CAT-1.** *RNA Biol* 2004, **1**(2):106-113.
257. Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, Chen J, Di Padova F, Lin SC, Gram H, Han J: **Involvement of microRNA in AU-rich element-mediated mRNA instability.** *Cell* 2005, **120**(5):623-634.
258. Cristofanilli M, Iacoangeli A, Muslimov IA, Tiedge H: **Neuronal BC1 RNA: microtubule-dependent dendritic delivery.** *J Mol Biol* 2006, **356**(5):1118-1123.
259. Muddashetty R, Khanam T, Kondrashov A, Bundman M, Iacoangeli A, Kremerskothen J, Duning K, Barnekow A, Huttenhofer A, Tiedge H *et al*: **Poly(A)-binding protein is associated with neuronal BC1 and BC200 ribonucleoprotein particles.** *J Mol Biol* 2002, **321**(3):433-445.
260. Wang H, Iacoangeli A, Popp S, Muslimov IA, Imataka H, Sonenberg N, Lomakin IB, Tiedge H: **Dendritic BC1 RNA: functional role in regulation of translation initiation.** *J Neurosci* 2002, **22**(23):10232-10241.
261. Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD: **The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator.** *Genes Dev* 2006, **20**(11):1470-1484.
262. Ulaner GA, Vu TH, Li T, Hu JF, Yao XM, Yang Y, Gorlick R, Meyers P, Healey J, Ladanyi M *et al*: **Loss of imprinting of IGF2 and H19 in osteosarcoma is accompanied by reciprocal methylation changes of a CTCF-binding site.** *Hum Mol Genet* 2003, **12**(5):535-549.
263. Pibouin L, Villaudy J, Ferbus D, Muleris M, Prosperi MT, Renvikos Y, Goubin G: **Cloning of the mRNA of overexpression in colon carcinoma-1: a sequence overexpressed in a subset of colon carcinomas.** *Cancer Genet Cytogenet* 2002, **133**(1):55-60.
264. Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB: **DD3: a new prostate-specific gene, highly overexpressed in prostate cancer.** *Cancer Res* 1999, **59**(23):5975-5979.
265. Petrovics G, Zhang W, Makarem M, Street JP, Connelly R, Sun L, Sesterhenn IA, Srikantan V, Moul JW, Srivastava S: **Elevated expression of PCGEM1, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients.** *Oncogene* 2004, **23**(2):605-611.
266. Rangel LB, Sherman-Baust CA, Wernyj RP, Schwartz DR, Cho KR, Morin PJ: **Characterization of novel human ovarian cancer-specific transcripts (HOSTs) identified by serial analysis of gene expression.** *Oncogene* 2003, **22**(46):7225-7232.
267. Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E *et al*: **MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer.** *Oncogene* 2003, **22**(39):8031-8041.
268. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K *et al*: **Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.** *Proc Natl Acad Sci U S A* 2002, **99**(24):15524-15529.
269. Michael MZ, SM OC, van Holst Pellekaan NG, Young GP, James RJ: **Reduced accumulation of specific microRNAs in colorectal neoplasia.** *Mol Cancer Res* 2003, **1**(12):882-891.
270. Eis PS, Tam W, Sun L, Chadburn A, Li Z, Gomez MF, Lund E, Dahlberg JE: **Accumulation of miR-155 and BIC RNA in human B cell lymphomas.** *Proc Natl Acad Sci U S A* 2005, **102**(10):3627-3632.
271. Metzler M, Wilda M, Busch K, Viehmann S, Borkhardt A: **High expression of precursor microRNA-155/BIC RNA in children with Burkitt lymphoma.** *Genes Chromosomes Cancer* 2004, **39**(2):167-169.
272. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ *et al*: **A microRNA polycistron as a potential human oncogene.** *Nature* 2005, **435**(7043):828-833.
273. Takamizawa J, Konishi H, Yanagisawa K, Tomida S, Osada H, Endoh H, Harano T, Yatabe Y, Nagino M, Nimura Y *et al*: **Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival.** *Cancer Res* 2004, **64**(11):3753-3756.
274. Chan JA, Krichevsky AM, Kosik KS: **MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells.** *Cancer Res* 2005, **65**(14):6029-6033.
275. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F, Casado S, Suarez-Gauthier A, Sanchez-Cespedes M, Gitt A *et al*: **Genetic unmasking of an epigenetically silenced microRNA in human cancer cells.** *Cancer Res* 2007, **67**(4):1424-1429.

276. Sparago A, Cerrato F, Vernucci M, Ferrero GB, Silengo MC, Riccio A: **Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and Beckwith-Wiedemann syndrome.** *Nat Genet* 2004, **36**(9):958-960.
277. Moseley ML, Zu T, Ikeda Y, Gao W, Mosemiller AK, Daughters RS, Chen G, Weatherspoon MR, Clark HB, Ebner TJ *et al*: **Bidirectional expression of CUG and CAG expansion transcripts and intranuclear polyglutamine inclusions in spinocerebellar ataxia type 8.** *Nat Genet* 2006, **38**(7):758-769.
278. Lukiw WJ, Handley P, Wong L, Crapper McLachlan DR: **BC200 RNA in normal human neocortex, non-Alzheimer dementia (NAD), and senile dementia of the Alzheimer type (AD).** *Neurochem Res* 1992, **17**(6):591-597.
279. Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH *et al*: **Disruption of two novel genes by a translocation co-segregating with schizophrenia.** *Hum Mol Genet* 2000, **9**(9):1415-1423.
280. Sonkoly E, Bata-Csorgo Z, Pivarsci A, Polyanka H, Kenderessy-Szabo A, Molnar G, Szentpali K, Bari L, Megyeri K, Mandi Y *et al*: **Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene, PRINS.** *J Biol Chem* 2005, **280**(25):24159-24167.
281. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M *et al*: **Sequence variants in SLITRK1 are associated with Tourette's syndrome.** *Science* 2005, **310**(5746):317-320.
282. Deleault NR, Lucassen RW, Supattapone S: **RNA molecules stimulate prion protein conversion.** *Nature* 2003, **425**(6959):717-720.
283. PipMaker [<http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?advanced>]
284. RepeatMasker [<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>]
285. Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF: **The expansion of the metazoan microRNA repertoire.** *BMC Genomics* 2006, **7**:25.
286. Huppi K, Volfovsky N, Mackiewicz M, Runfola T, Jones TL, Martin SE, Stephens R, Caplen NJ: **MicroRNAs and genomic instability.** *Semin Cancer Biol* 2006.
287. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, Shimizu M, Rattan S, Bullrich F, Negrini M *et al*: **Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers.** *Proc Natl Acad Sci U S A* 2004, **101**(9):2999-3004.
288. Lord C, Rutter M, Le Couteur A: **Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders.** *J Autism Dev Disord* 1994, **24**(5):659-685.
289. **A haplotype map of the human genome.** *Nature* 2005, **437**(7063):1299-1320.
290. **Integrating ethics and science in the International HapMap Project.** *Nat Rev Genet* 2004, **5**(6):467-475.
291. **The International HapMap Project.** *Nature* 2003, **426**(6968):789-796.
292. Thorisson GA, Smith AV, Krishnan L, Stein LD: **The International HapMap Project Web site.** *Genome Res* 2005, **15**(11):1592-1593.
293. Hofacker IL: **Vienna RNA secondary structure server.** *Nucleic Acids Res* 2003, **31**(13):3429-3431.
294. Collins AL, Ma D, Whitehead PL, Martin ER, Wright HH, Abramson RK, Hussman JP, Haines JL, Cuccaro ML, Gilbert JR *et al*: **Investigation of autism and GABA receptor subunit genes in multiple ethnic groups.** *Neurogenetics* 2006, **7**(3):167-174.
295. MiRscan [<http://genes.mit.edu/mirscan/>]
296. Eddy SR: **A memory-efficient dynamic programming algorithm for optimal alignment of a sequence to an RNA secondary structure.** *BMC Bioinformatics* 2002, **3**:18.
297. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR: **Rfam: an RNA family database.** *Nucleic Acids Res* 2003, **31**(1):439-441.
298. Eddy SR: **Noncoding RNA genes.** *Curr Opin Genet Dev* 1999, **9**(6):695-699.
299. Kantor B, Makedonski K, Green-Finberg Y, Shemer R, Razin A: **Control elements within the PWS/AS imprinting box and their function in the imprinting process.** *Hum Mol Genet* 2004, **13**(7):751-762.
300. Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J: **A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain.** *Genome Res* 2004, **14**(9):1741-1748.
301. Fu H, Tie Y, Xu C, Zhang Z, Zhu J, Shi Y, Jiang H, Sun Z, Zheng X: **Identification of human fetal liver miRNAs by a novel method.** *FEBS Lett* 2005, **579**(17):3849-3854.
302. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP: **Vertebrate microRNA genes.** *Science* 2003, **299**(5612):1540.

303. Barnby G, Abbott A, Sykes N, Morris A, Weeks DE, Mott R, Lamb J, Bailey AJ, Monaco AP: **Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT.** *Am J Hum Genet* 2005, **76**(6):950-966.
304. Dhossche D, Applegate H, Abraham A, Maertens P, Bland L, Bencsath A, Martinez J: **Elevated plasma gamma-aminobutyric acid (GABA) levels in autistic youngsters: stimulus for a GABA hypothesis of autism.** *Med Sci Monit* 2002, **8**(8):PR1-6.
305. Hussman JP: **Suppressed GABAergic inhibition as a common factor in suspected etiologies of autism.** *J Autism Dev Disord* 2001, **31**(2):247-248.
306. Schmitz C, van Kooten IA, Hof PR, van Engeland H, Patterson PH, Steinbusch HW: **Autism: neuropathology, alterations of the GABAergic system, and animal models.** *Int Rev Neurobiol* 2005, **71**:1-26.
307. Carrozzo R, Rossi E, Christian SL, Kittikamron K, Livieri C, Corrias A, Pucci L, Fois A, Simi P, Bosio L *et al*: **Inter- and intrachromosomal rearrangements are both involved in the origin of 15q11-q13 deletions in Prader-Willi syndrome.** *Am J Hum Genet* 1997, **61**(1):228-231.
308. Vogels A, Fryns JP: **The Prader-Willi syndrome and the Angelman syndrome.** *Genet Couns* 2002, **13**(4):385-396.
309. Boyar FZ, Whitney MM, Lossie AC, Gray BA, Keller KL, Stalker HJ, Zori RT, Geffken G, Mutch J, Edge PJ *et al*: **A family with a grand-maternally derived interstitial duplication of proximal 15q.** *Clin Genet* 2001, **60**(6):421-430.
310. Wolpert CM, Menold MM, Bass MP, Qumsiyeh MB, Donnelly SL, Ravan SA, Vance JM, Gilbert JR, Abramson RK, Wright HH *et al*: **Three probands with autistic disorder and isodicentric chromosome 15.** *Am J Med Genet* 2000, **96**(3):365-372.
311. Sinkkonen ST, Homanics GE, Korpi ER: **Mouse models of Angelman syndrome, a neurodevelopmental disorder, display different brain regional GABA(A) receptor alterations.** *Neurosci Lett* 2003, **340**(3):205-208.
312. Claverie JM: **Fewer genes, more noncoding RNA.** *Science* 2005, **309**(5740):1529-1530.
313. Kim J, Krichevsky A, Grad Y, Hayes GD, Kosik KS, Church GM, Ruvkun G: **Identification of many microRNAs that copurify with polyribosomes in mammalian neurons.** *Proc Natl Acad Sci U S A* 2004, **101**(1):360-365.
314. Zhang S, Borovok I, Aharonowitz Y, Sharan R, Bafna V: **A sequence-based filtering method for ncRNA identification and its application to searching for riboswitch elements.** *Bioinformatics* 2006, **22**(14):e557-565.
315. Zhou H, Zhao J, Yu CH, Luo QJ, Chen YQ, Xiao Y, Qu LH: **Identification of a novel box C/D snoRNA from mouse nucleolar cDNA library.** *Gene* 2004, **327**(1):99-105.
316. Yang Z, Ebright YW, Yu B, Chen X: **HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide.** *Nucleic Acids Res* 2006, **34**(2):667-675.
317. Luciano DJ, Mirsky H, Vendetti NJ, Maas S: **RNA editing of a miRNA precursor.** *Rna* 2004, **10**(8):1174-1177.
318. Schadt EE, Edwards SW, GuhaThakurta D, Holder D, Ying L, Svetnik V, Leonardson A, Hart KW, Russell A, Li G *et al*: **A comprehensive transcript index of the human genome generated using microarrays and computational approaches.** *Genome Biol* 2004, **5**(10):R73.
319. Li SC, Pan CY, Lin WC: **Bioinformatic discovery of microRNA precursors from human ESTs and introns.** *BMC Genomics* 2006, **7**:164.
320. Pedersen JS, Bejerano G, Siepel A, Rosenbloom K, Lindblad-Toh K, Lander ES, Kent J, Miller W, Haussler D: **Identification and classification of conserved RNA secondary structures in the human genome.** *PLoS Comput Biol* 2006, **2**(4):e33.
321. Ohler U, Yekta S, Lim LP, Bartel DP, Burge CB: **Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification.** *Rna* 2004, **10**(9):1309-1322.
322. Cummins JM, He Y, Leary RJ, Pagliarini R, Diaz LA, Jr., Sjoblom T, Barad O, Bentwich Z, Szafarska AE, Labourier E *et al*: **The colorectal microRNAome.** *Proc Natl Acad Sci U S A* 2006, **103**(10):3687-3692.
323. Fontaine A. TH: **How to detect non-coding RNAs?** In.; 2005.
324. Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, Okunishi R, Fukuda S, Ru K, Frith MC, Gongora MM *et al*: **Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome.** *Genome Res* 2006, **16**(1):11-19.
325. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ *et al*: **Unbiased mapping of transcription factor binding sites along**

- human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs.** *Cell* 2004, **116**(4):499-509.
326. Royo H, Bortolin ML, Seitz H, Cavaille J: **Small non-coding RNAs and genomic imprinting.** *Cytogenet Genome Res* 2006, **113**(1-4):99-108.