



UNIVERSITAT DE  
BARCELONA

# Una nueva domesticación molecular en el origen de los euterios

## La familia génica *Bex/Tceal*

Enrique Navas Pérez

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

**Una nueva domesticación molecular en el origen de los euterios:**

La familia génica *Bex/Tceal*



Enrique Navas Pérez







*Thus strangely are our souls constructed, and by slight ligaments are we  
bound to prosperity and ruin.*

— Mary Wollstonecraft Shelley, *Frankenstein*



## Agradecimientos

En primer lugar, quiero agradecer a Jordi la confianza, la cercanía y la libertad que siempre me ha brindado. Gracias también por intentar mantener al grupo lo más unido posible. Difícilmente encontraré otro jefe con quien sea tan cómodo trabajar y crecer como científico.

Luego, gracias a Demi. Por ser un «tron», un gil y un futuro Dr. Grijander. Por haberme enseñado buena parte de lo que sé sobre el mundo de la ciencia, por atender con paciencia todas mis dudas y estar siempre ahí para responderlas de forma estimulante. Por tantas discusiones, y por tu preocupación sincera. Y *last but not the least*, por ser un gran amigo prácticamente desde el día en que nos conocimos. Este trabajo no se puede entender sin tu ayuda.

Por supuesto, a todos los que han pasado por el grupo de Jordi. A Carlos, por intentar (y conseguirlo siempre) hacer más alegre y agradable cada día que hemos compartido, y por ser demasiado *freak* para morir. A Bea y Aina, por aguantar a un desastre de persona... ¡Pero bien está lo que bien acaba! Y a Ari, Roser, Anna y demás compañer@s y alumñ@s. También a la *old school*: sobre todo a Manu, Nacho y Chelo, por ser unos referentes.

A toda la gente de otros laboratorios con la que he colaborado. Especialmente a Serena y Cristina, dos grandes personas con una paciencia infinita.

A las planarias (Eudald, Miquel, Nidia, Jose, Sisco, María, Marta, Susanna, Sara, Noelia, Nieves, Àlex y esclavitos), celos y moscas. Nunca me habéis negado un puñado de hielo.

A mis amig@s, tanto los de Barcelona (sobre todo a mis compañeros de zulo, Andreu, Martí y Tito) como los del pueblo: no os escribo nada, porque sé que no vais a leerlo, cabrones.

Y por último, a mi familia, especialmente a mis padres y hermanos. Os quiero.

¡Gracias!





# Índice

<b>Introducción</b>	<b>11</b>
Mecanismos de formación de nuevos genes	14
Introducción a los transposones	16
<i>Retrotransposones</i>	17
<i>Transposones de ADN</i>	20
Los transposones como facilitadores de la adaptación evolutiva	20
<i>Los transposones domesticados</i>	22
<i>Impacto evolutivo de los TDs</i>	25
<i>TDs en mamíferos</i>	27
<b>Objetivos</b>	<b>31</b>
<b>Métodos</b>	<b>35</b>
<b>Resultados</b>	<b>41</b>
Búsqueda de genes domesticados	43
Análisis de la exaptación (I) – Nacimiento de un nuevo gen	45
Análisis de la exaptación (II) – Nacimiento de una nueva familia	50
Análisis de la exaptación (III) – Coopción de regiones reguladoras	55
Análisis de la exaptación (IV) – Impacto del motivo BGW	59
Expresión de la familia <i>Bex/Tceal</i>	63

Generación y caracterización de ratones <i>Bex3</i> mutantes	65
Resumen de resultados	73
<b>Discusión</b>	<b>75</b>
Métodos para la detección de TDs	77
Una nueva domesticación molecular en el origen de los euterios	78
El motivo BGW y la regulación de la expresión de los genes BGW	80
Generación y caracterización de ratones <i>Bex3</i> mutantes	84
Integración funcional de los genes <i>Bex/Tceal</i>	92
<b>Conclusiones</b>	<b>99</b>
<b>Material suplementario</b>	<b>103</b>
<b>Referencias</b>	<b>109</b>
<b>Anexos</b>	<b>129</b>

## **Introducción**



Una de las cuestiones básicas de la biología evolutiva atañe a la comprensión de las bases genéticas subyacentes a los cambios fenotípicos. Junto a los cambios en la regulación génica, la renovación del *pool* genético es uno de los principales mecanismos genéticos capaces de producir innovaciones evolutivas. Por eso, su estudio ha despertado gran interés entre los pioneros de la genética y la evolución. Ya en la década de 1930, se hipotetizó que la duplicación de material genético podría desempeñar un importante papel evolutivo<sup>1</sup>. Posteriormente, estas ideas fueron desarrolladas por Susumu Ohno, quien propuso que la duplicación era el principal mecanismo detrás de la generación de nuevos genes<sup>2</sup>. Además, intentó responder a una pregunta con implicaciones evolutivas fundamentales: ¿Qué repercusiones funcionales desencadenaría la aparición de estos nuevos actores? Propuso Ohno que, en ciertos casos, después de la duplicación, una de las copias retendría la función ancestral mientras que la otra quedaría libre para evolucionar nuevas funciones: ese proceso se denominó neofuncionalización. Numerosos estudios han contribuido a establecer que la aparición de nuevos genes, bien sea vía duplicación o por otros mecanismos, ha permitido la evolución de nuevas funciones génicas que han mediado importantes cambios fenotípicos<sup>3</sup>. Más adelante, esas teorías se refinaron y, al mismo tiempo, se plantearon escenarios alternativos. Por ejemplo, la duplicación podría conllevar la subfuncionalización de la función ancestral entre las copias resultantes<sup>4</sup>. Por otro lado, mediante la modificación de los dominios de la proteína para la que codifica, un gen puede adquirir una nueva función sin necesidad de duplicarse<sup>5</sup>.

En el marco de la biología evolutiva del desarrollo (la llamada «evo-devo»), se ha propuesto que la innovación fenotípica ocurre mayormente vía la modificación de elementos reguladores en *cis*, afinando de ese modo los patrones de expresión espaciotemporales durante el desarrollo de una serie de factores evolutivamente conservados, a los que se ha dado en llamar *genetic*

*toolkit*<sup>6</sup>. Sin embargo, en ese sentido también se ha reconocido la importancia tanto de la pérdida como de la ganancia de genes<sup>7,8</sup>. De hecho, la creciente disponibilidad de genomas secuenciados nos muestra una significativa variabilidad en el número de genes codificantes, y se hace evidente que en la práctica totalidad de los genomas se produce una renovación más o menos dinámica de una parte considerable del *pool* genético mediante un proceso continuado de muerte y nacimiento de genes<sup>7-11</sup>. Comprender la lógica de estas dinámicas y los mecanismos que las posibilitan es, por lo tanto, comprender una parte importante de los cambios evolutivos.

## **Mecanismos de formación de nuevos genes**

La era genómica ha facilitado la comprensión de los procesos que determinan la aparición de nuevos genes. De entre ellos, el más común y mejor estudiado es la duplicación. El proceso de duplicación puede darse de distintas formas, y estar mediado por ADN o ARN. Por ejemplo, debido a defectos durante la recombinación meiótica pueden duplicarse segmentos cromosómicos que comprendan genes completos o fragmentos de ellos. A mayor escala, también se ha observado que, de modo recurrente en la historia evolutiva de los eucariotas, el genoma completo puede duplicarse<sup>12,13</sup>, resultando en la generación de miles de nuevos duplicados. Aunque la mayoría de ellos acumularán mutaciones y devendrán pseudogenes, una proporción considerable puede conservarse<sup>14-16</sup>.

Por otro lado, el proceso de retroposición, retrocopia o retroduplicación consiste en la retrotranscripción del producto de la transcripción génica —es decir, de una molécula de ARN mensajero (ARNm)— y su posterior integración en el genoma. Al haber pasado el ARNm por el proceso de *splicing* y tener, por lo tanto, los exones empalmados, estas retrocopias (o retrogenes) carecen en un principio de intrones<sup>3</sup>.

La alteración y la recombinación de estructuras génicas es otro modo de producir nuevos genes. Fue Walter Gilbert quien, en los años 1970, propuso que los exones codificantes y dominios proteicos de distintos genes podían recombinarse (en inglés, *shuffle*), creando así nuevas estructuras quiméricas<sup>17</sup>. Con el tiempo, se ha confirmado que la recombinación génica es un fenómeno común a todos los eucariotas. Un caso ilustrativo es el del gen *jingwei*, uno de los primeros cuya formación pudo ser estudiada en profundidad. En el linaje de dos especies africanas de *Drosophila*, hace unos dos millones de años, una retrocopia del gen *Adh* se insertó en un intrón del gen *yande*. *Jingwei*, el nuevo gen resultante, se originó a partir de la recombinación de los primeros exones de *yande* y el material genético aportado por la retrocopia de *Adh*, y, a diferencia de los genes parentales, evolucionó una función en el metabolismo hormonal<sup>18</sup>.

Finalmente, existen otras fuentes más minoritarias pero no por ello menos interesantes desde un punto de vista evolutivo. Una de ellas es la transferencia genética horizontal (TGH), que consiste en el intercambio de genes entre los genomas de especies filogenéticamente distantes. Muy común en procariontes, también ocurre en eucariotas<sup>19</sup>. De hecho, en los últimos años han ido emergiendo multitud de casos que demuestran cómo genes de organismos eucariotas originados vía TGH son capaces de conferir ventajas adaptativas<sup>20-22</sup>. Además, es fascinante comprobar que, probablemente debido al íntimo contacto físico, muchos de ellos se dan, de forma bidireccional, entre huésped y parásito<sup>23</sup>. Por otro lado, nuevas estructuras génicas pueden surgir totalmente *de novo*, es decir, a partir de material genómico previamente no génico. A pesar de que, en un principio, este mecanismo se consideró altamente improbable<sup>2</sup>, la importancia evolutiva de estos genes se va poniendo de relieve<sup>24</sup>. Por último, la domesticación molecular de genes transposónicos es otra fuente de nuevos genes. Al ser el foco de este trabajo, se hace necesario introducir algunos conceptos relativos a la naturaleza de los transposones.



## Introducción a los transposones

Descubiertos por Barbara McClintock en la década de 1940, los transposones —también llamados elementos móviles o transponibles— se definen como segmentos de ADN capaces de movilizarse de un locus a otro del genoma. Estos elementos están presentes, de forma activa o «fósil», en prácticamente todos los genomas analizados: por ejemplo, constituyen casi el 50% del genoma humano (algunas estimaciones lo elevan hasta el 70%<sup>25</sup>). Pese a su omnipresencia, se ha identificado una gran variabilidad en la abundancia y diversidad de los transposones, tanto en procariotas como en eucariotas, e incluso entre distintos grupos de vertebrados. La clasificación de estos elementos móviles y la comprensión de los mecanismos de transposición son de gran importancia debido a sus implicaciones evolutivas.

Al abordar la clasificación de los transposones, en este trabajo seguiremos los sistemas de Wicker<sup>26</sup>, debido a su amplia difusión (para otras propuestas, ver el trabajo de Piégu et al.<sup>27</sup>). Tradicionalmente, los transposones han sido divididos en dos tipos, de acuerdo con sus mecanismos de transposición: los elementos de clase I, también llamados retrotransposones, y los de clase II, o transposones de ADN (fig. 1).

Además, es importante recalcar que podemos diferenciar los transposones aún a otro nivel: entre autónomos y no autónomos. Los primeros codifican para todos los componentes necesarios durante su ciclo vital, mientras que los segundos carecen de alguno o de todos ellos. Para realizar su transposición, los elementos no autónomos se apropian de factores provenientes de otros transposones autónomos simultáneamente activos. Por ejemplo, ese es el caso de los transposones de ADN *Ac* (*Activator*) y *Ds* (*Dissociator*; elemento no autónomo que necesita de *Ac* para su movilización), descubiertos por McClintock<sup>28</sup>.

## *Retrotransposones*

Para movilizarse, los retrotransposones utilizan un mecanismo de «copia y pega», que consiste en transcribir un intermediario de ARN, que será posteriormente retrotranscrito a ADN e insertado en otro locus genómico. Como resultado de este proceso, se genera una copia nueva del transposón original. Los retrotransposones más presentes en mamíferos son los elementos nucleares dispersos largos y cortos (*long y short interspersed nuclear elements*, LINEs y SINEs), y los retrotransposones de repeticiones terminales largas (*long terminal repeat*, LTR).

Los retrotransposones LTR están evolutivamente relacionados con los retrovirus<sup>29</sup> y se caracterizan por tener unas repeticiones terminales largas directas, de centenares de bases, flanqueando los genes transposónicos. Dentro de los LTRs, los retrovirus endógenos (*endogenous retrovirus*, ERV) son copias de retrovirus que han infectado e integrado su genoma viral en la línea germinal. Estos provirus heredables pueden quedar grabados durante millones de años en el genoma del huésped, a modo de fósiles de infecciones víricas ancestrales. Los ERVs constan de tres genes: *pol*, necesario para la retrotranscripción e integración del ADN viral; *gag*, que codifica para una poliproteína que dará lugar a la matriz y las cápsides, y *env*, cuya proteína codificada es la encargada de reconocer los receptores de membrana y mediar la fusión virus-célula.

Tabla 1. Clasificación de los transposones

<b>Clase I (retrotransposones)</b>		<b>Clase II (transposones ADN) - Subclase I</b>	
Orden	Superfamilia	Orden	Superfamilia
LTR	<i>Copia</i>	TIR	<i>Tc1-Mariner</i>
	<i>Gypsy</i>		<i>hAT</i>
	<i>Bel-Pao</i>		<i>Mutator</i>
	<i>Retrovirus</i>		<i>Merlin</i>
	<i>ERV</i>		<i>Transib</i>
DIRS	<i>DIRS</i>		<i>P</i>
	<i>Ngaro</i>		<i>PiggyBac</i>
	<i>VIPER</i>		<i>PIF-Harbinger</i>
PLE	<i>Penelope</i>		<i>CACTA</i>
LINE	<i>R2</i>	Crypton	<i>Crypton</i>
	<i>RTE</i>		
	<i>Jockey</i>		
	<i>L1</i>		
SINE	<i>I</i>		
	<i>tRNA</i>		
	<i>7SL</i>		
	<i>5S</i>		

<b>Clase II (transposones ADN) - Subclase II</b>	
Orden	Superfamilia
Helitron	<i>Helitron</i>
Maverick	<i>Maverick</i>

<b>Clase II (transposones ADN) - Subclase II</b>	
Orden	Superfamilia
Helitron	<i>Helitron</i>
Maverick	<i>Maverick</i>

→ LTR (repeticiones terminales largas)    → TIR (repeticiones terminales invertidas)    ■ Región codificante  
 — Región no codificante    —/— Región que puede contener uno o más ORFs    — Elemento diagnóstico de región no codificante

AP: Proteínasa aspártica    APE: Endonucleasa apurínica    ATP: ATPasa    C-INT: C-integrasa    CYP: Cisteín-proteasa    EN: Endonucleasa  
 ENV: Proteína de la envoltura    GAG: Proteína de la cápside    INT: Integrasa    HEL: Helicasa    ORF: Pauta de lectura abierta    POL B: Polimerasa de ADN B  
 RH: ARNasa H    RPA: Proteína de replicación A    RT: Transcriptasa reversa    Tase: Transposasa    YR: Tirosina recombinasa    Y2: YR con motivo YY

Los elementos L1 (o LINE-1) son una de las superfamilias de LINEs, y están ampliamente distribuidos en los genomas de vertebrados. El caso más notable es el de los mamíferos: más del 20% de su genoma puede estar compuesto por copias de este tipo de transposones. La evolución de los L1 en estos animales sigue una pauta muy particular: habitualmente, una sola familia es responsable de la actividad replicativa, hasta que es silenciada y su testigo lo recoge otra familia más «moderna»<sup>30,31</sup>. Así, en una especie de carrera armamentística (entre el huésped y el transposón), de relevos (entre las distintas familias), se forma una característica filogenia en escalones, similar a la del virus de la gripe. Se ha de remarcar que, en humanos, son la única clase de transposones autónomos activos en la actualidad<sup>32</sup>. Por otro lado, en vertebrados no mamíferos —como los reptiles escamosos y los teleosteos— los L1 están representados por múltiples linajes altamente divergentes y paralelamente activos, pero constituyen un porcentaje mucho menor del genoma<sup>33</sup>.

Típicamente, un elemento L1 completo consta de unas 6 kilobases (kb), que codifican para dos marcos abiertos de lectura (*open reading frame*, ORFs) separados por una región intergénica. Como en la mayoría de los retrotransposones, los insertos están flanqueados por duplicaciones del sitio diana (*target site duplication*, TSDs), que se forman durante la transposición. En los elementos L1 presentes en vertebrados, el primer ORF (ORF1) codifica para una proteína (ORF1p) de unión a RNA con capacidad chaperona de ácidos nucleicos<sup>34</sup>. Estructuralmente, presenta tres dominios: una hélice superenrollada (*coiled-coil*, CC) mediante la cual forma trímeros (Martin et al., 2003); un dominio de reconocimiento de ARN (*RNA-recognition-motif*, RRM), y un dominio carboxi-terminal (*carboxy-terminal domain*, CTD). ORF1p necesita de estos últimos dos dominios para poder unir ARN<sup>35,36</sup>. El ORF2 codifica una proteína algo mayor (ORF2p) con actividades endonucleasa y retrotranscriptasa<sup>37,38</sup>. Es de interés señalar que, en mamíferos y plantas, la capacidad que tiene la proteína ORF2p de los L1 de retrotranscribir el ARN en

ADN es la responsable de generar las retrocopias de las que se ha hablado anteriormente<sup>39,40</sup>.

Estudios recientes han demostrado que este esquema no engloba la complejidad real de la familia L1. Algunos elementos son capaces de codificar para otras proteínas en antisentido codificación<sup>41,42</sup>, así como de deshacerse del ORF2, convirtiéndose en elementos no autónomos<sup>43</sup>. Además de los ORFs, los L1 contienen 5'UTRs y 3'UTRs. En la 5'UTR, inmediatamente después del inicio de transcripción, un sitio de unión para el represor transcripcional YY1 (Yin Yang 1) asegura que esta se lleve a cabo a partir del nucleótido adecuado<sup>44</sup>. Tanto ORF1p como ORF2p son necesarias para una correcta transposición (Moran et al., 1996), que se lleva a cabo en un proceso llamado retrotranscripción dirigida por el sitio diana (*target-primed reverse transcription*, TPRT)<sup>45</sup>.

### *Transposones de ADN*

A diferencia de los retrotransposones, los transposones de ADN se movilizan en su mayoría mediante un mecanismo de «corta y pega». La transposasa para la que codifican reconoce unas repeticiones terminales invertidas (*terminal inverted repeats*, TIRs) situadas en cada extremo del transposón, generando un corte de doble cadena y permitiendo su movilización. Los transposones *Helitron* y *Maverick* forman parte de otra subclase de transposones de ADN. Su transposición requiere la replicación de la copia, y no producen cortes de doble cadena.

## **Los transposones como facilitadores de la adaptación evolutiva**

A causa de su aparente falta de función y su capacidad para proliferar a expensas del huésped y producirle mutaciones perjudiciales, los transposones han sido catalogados por una parte de la comunidad científica como ADN

basura. En las últimas dos décadas, han empezado a trascender algunos aspectos sobre la relevancia evolutiva de estos elementos móviles, aunque el debate sobre la consideración que merecen —ADN basura o reservorio evolutivo— sigue candente hoy en día<sup>46,47</sup>.

Al ser capaces de movilizarse y recombinar entre sí, generar inserciones, duplicaciones, deleciones y translocaciones genómicas, uno de los efectos de la actividad transposónica es la generación de mutaciones. Pese a que esa capacidad dinamizadora tiene frecuentemente efectos perniciosos en los organismos<sup>48</sup>, a nivel de especie puede favorecer la especiación y la adaptación, por ejemplo como fuente de innovaciones evolutivas<sup>49</sup>.

Artículos muy bellos publicados recientemente muestran las consecuencias evolutivas de la actividad transposónica, a veces mediante una sola inserción<sup>50,51</sup>. Uno de ellos es el de Van't Hof y colaboradores<sup>52</sup>, donde se describe la mutación causante del famoso melanismo industrial de la mariposa de los abedules, *Biston betularia*. Durante la Revolución Industrial, la inserción de un transposón en el intrón del gen *cortex* alteró la expresión de este y se produjo el cambio fenotípico (de coloración moteada a negra), que acabó extendiéndose entre la población. Además de la capacidad mutagénica para insertarse en nuevos loci, la similitud entre las distintas copias de un transposón puede facilitar recombinaciones no alélicas que conlleven duplicaciones o deleciones. Por ejemplo, la recombinación entre elementos LINE parece haber potenciado la expansión de familias génicas en ratón y hombre<sup>53</sup>.

Asimismo, la secuencia genética del transposón alberga motivos susceptibles de ser cooptados por el genoma huésped, tanto a pequeña como a gran escala. Por un lado, al tener en cuenta la extensa distribución genómica y el elevado número de copias a que pueden llegar estos elementos móviles, se pone de manifiesto el potencial evolutivo que guardan. Siguiendo esta línea, numerosos artículos han demostrado que las copias de ciertas familias de transposones contienen sitios de unión para factores de transcripción que han

sido utilizados para establecer redes de regulación génica<sup>54</sup>. Otro fenómeno al que han sido asociados los transposones es la generación de nuevos intrones. Este proceso suele caracterizarse por largos periodos con escaso o nulo incremento en el número de intrones de grandes linajes evolutivos, seguidos de episodios muy puntuales de ganancia explosiva en alguno de sus clados<sup>55</sup>. Huff et al. han identificado dos transposones de ADN no autónomos como los responsables de la reciente generación de miles de nuevos intrones en sendas especies de algas unicelulares<sup>56</sup>. Dada la antigüedad y ubicuidad de los transposones de ADN, los autores sugieren que un mecanismo similar podría estar detrás de algunos de los episodios de ganancia de intrones en eucariotas.

En cuanto a su contribución a la estructura génica, fragmentos de transposones han aportado material genético a ARNs largos no codificantes<sup>57</sup>, así como a exones codificantes<sup>58</sup>. Aquí encontramos uno de los casos más impactantes de exaptación de transposones. Los genes o dominios proteicos codificados por los transposones pueden dar lugar a nuevos genes «controlados» por el huésped y evolucionar funciones beneficiosas. Este proceso se denomina domesticación molecular<sup>59</sup>, coopción<sup>60</sup> o exaptación<sup>61</sup> de elementos transponibles. En este trabajo se utilizará el término domesticación, pues, además de ser la acepción original para describir esta clase de eventos, en la literatura concerniente a los transposones exaptación y coopción han sido utilizados principalmente para referirse a su aprovechamiento como *enhancers*.

### *Los transposones domesticados*

Los transposones domesticados (TDs) son copias de un transposón ancestral que han originado un gen funcional para el huésped. A diferencia de una copia no domesticada, los TDs se comportan y evolucionan de forma similar a los genes convencionales, y esto nos ayuda a diferenciarlos. En primer lugar, un TD ha dejado de ser móvil. Por lo tanto, su posición se estabiliza, las secuencias flanqueantes necesarias para su movilización suelen degenerar, y, al perder la capacidad de transposición, su potencial para multiplicarse es menor.

Normalmente, las copias de un transposón evolucionan de forma neutral, y con el tiempo el ORF para el que codifican termina perdiéndose. Así, si se añade la reducida movilidad propia de los genes convencionales, la conservación del ORF de un transposón entre especies en una posición ortóloga podría señalar una domesticación. Por otro lado, para ejercer su función un TD debe expresarse, mientras que en la mayoría de tejidos los transposones se mantienen silenciados gracias a diversos mecanismos (ref).

Otra característica de los TDs (al igual que las retrocopias y los genes originados vía TGH o *de novo*) es la necesidad de adquirir señales que permitan su transcripción y controlen específicamente cuándo y en qué tejido o tipo celular —así como con qué intensidad— se expresan, de modo que puedan manifestar y desarrollar su función. La posibilidad de estudiar de cerca estos procesos hace que esos genes sean especialmente interesantes desde un punto de vista evolutivo.

Primeramente, para que se dé la unión de la polimerasa que transcribirá el gen, es necesaria la existencia de un promotor. Se han caracterizado diversas estrategias mediante las cuales estos genes adquieren o cooptan uno. Por ejemplo, pueden aprovechar el promotor previamente funcional de un gen vecino. Así, en un locus en el que el gen vecino se transcriba en sentido opuesto al nuevo gen, el promotor del primero puede ser o convertirse en bidireccional, permitiendo la transcripción del segundo. También, si el nuevo gen ha aparecido (o se ha insertado) en el intrón o *downstream* de otro gen, puede aprovechar el promotor así como algunos de los exones, codificantes o no, del gen vecino. El resultado, por lo tanto, sería una suerte de fusión de elementos génicos.

Por otro lado, otras secuencias tienen la capacidad de desempeñar *de novo* el papel de promotor: en el caso específico de los TDs y las retrocopias, este parece ser el escenario más común<sup>62,63</sup>. En mamíferos, las secuencias con un alto porcentaje de G-Cs (también llamadas islas CpG) son una importante



fuelle de nuevos promotores, debido a su predisposición a abrir la cromatina y mantenerse accesibles a factores de transcripción y regiones reguladoras<sup>64,65</sup>. Existen, por lo tanto, ciertas regiones características con una capacidad latente de devenir promotores. En este aspecto, son notables los resultados de Carelli y colaboradores<sup>62</sup>. Comparando regiones ortólogas de ratón y rata, encuentran que una parte importante de los promotores de retrocopias específicas de rata derivan de secuencias que ancestralmente eran potenciadores (*enhancers*). Esto remarca la naturaleza ambivalente de estas regiones reguladoras, así como su potencial para ser cooptadas por genes nacientes.

Finalmente, aunque en la mayoría de casos se captan regiones reguladoras del «ecosistema» genómico circundante, también existen algunas excepciones. Como se ha comentado anteriormente, los retrogenes derivan de copias carentes de intrones pero que incluyen las regiones no traducidas (*untranslated regions*, UTRs) en 5' y 3' del gen parental. En mamíferos, un pequeño porcentaje (~3%) de estas retrocopias es capaz, una vez integrado, de heredar el promotor parental incluido en la 5'UTR<sup>62</sup>. De forma similar, los genes adquiridos por TGH pueden arrastrar consigo parte de las regiones reguladoras que permitirán su transcripción<sup>66</sup>. En cuanto a los TDs, no se conocen casos análogos.

Otra consecuencia de la peculiar naturaleza de su origen (de transposones, derivados de ARNm retrotranscrito, *de novo*, etc.) es que estos genes empiezan el proceso de exaptación sin intrones ¿Cómo y por qué, pues, la mayoría de ellos los adquieren durante el proceso? Los estudios de Fablet et al., y de Kordis y Kokosar sobre la ganancia de intrones (normalmente mediante la intronización de secuencias no codificantes) por parte de retrogenes y TDs, respectivamente, ayudan a responder estas preguntas<sup>67,68</sup>. Una vez captado el promotor por parte del nuevo gen, la evolución de sitios de *splicing* (o utilización de sitios crípticos) permitiría descartar del transcrito la secuencia *upstream* al codón ATG. Así, se eliminarían posibles inicios de traducción alternativos y la aparición de estructuras secundarias deletéreas en el ARNm.

Además, el mantenimiento de estos nuevos intrones podría haber tenido otras funcionalidades, pues con el tiempo contribuirían a regular los niveles de ARN y proteína<sup>69</sup>.

A pesar de los rasgos definitorios arriba mencionados, la identificación sistemática de TDs siguiendo estos y otros criterios solo ha sido abordada recientemente por Hoen y Bureau<sup>61</sup>. Su estudio, centrado en la planta modelo *Arabidopsis thaliana*, ha doblado el número de candidatos a TDs en esta especie, que han pasado de 31 a 67. Es tentador concluir que, con la excepción de clados intensamente estudiados, como los mamíferos, la mayoría de casos aún permanecen por descubrir.

### *Impacto evolutivo de los TDs*

Las proteínas transposónicas son capaces de desempeñar funciones muy diversas: unir, cortar y degradar ácidos nucleicos, interactuar con proteínas del huésped o modificar la cromatina. Por lo tanto, es sugerente pensar que los transposones son una fuente recurrente de nuevos genes, debido a las propiedades de los motivos estructurales y dominios de las proteínas que codifican.

Las transposasas propias de los transposones de ADN constituyen la fuente más común y extendida de TDs. Su estudio ilustra bien de qué forma han sido aprovechados los dominios proteicos transposónicos. Típicamente, una transposasa consta de un dominio de unión al ADN (*DNA-binding domain*, DBD) y un dominio catalítico que lleva a cabo las reacciones de corte y unión del ADN durante la transposición. Para ejercer su función, debe translocarse al núcleo y acceder al ADN diana, modificando o aprovechando relajaciones transitorias de la cromatina.

Un ejemplo de transposasa domesticada es la involucrada en la aparición, hace unos 500 millones de años, del sistema inmune adaptativo de los gnatóstomos. Este subgrupo de vertebrados es capaz de generar, mediante un

proceso de recombinación genética llamado V(D)J, un amplio repertorio de inmunoglobulinas y receptores de linfocitos T. El proceso de recombinación es muy parecido a la transposición de «corta y pega», y es mediado por la acción de las proteínas RAG1 y 2. El complejo RAG1/2 reconoce mediante su DBD unas secuencias (parecidas a los TIRs de los transposones de ADN) situadas en los extremos de cada segmento V, D y J, y crea una horquilla de ADN que será posteriormente escindida y descartada, y sus extremos unidos. La eliminación del ADN que separa los múltiples segmentos resulta en el ensamblaje de una sola unidad V(D)J. Hacia el final de un largo camino de investigaciones que han ido completando el rompecabezas, se ha podido confirmar que ambas proteínas derivan de genes pertenecientes a un mismo transposón ADN<sup>70</sup>.

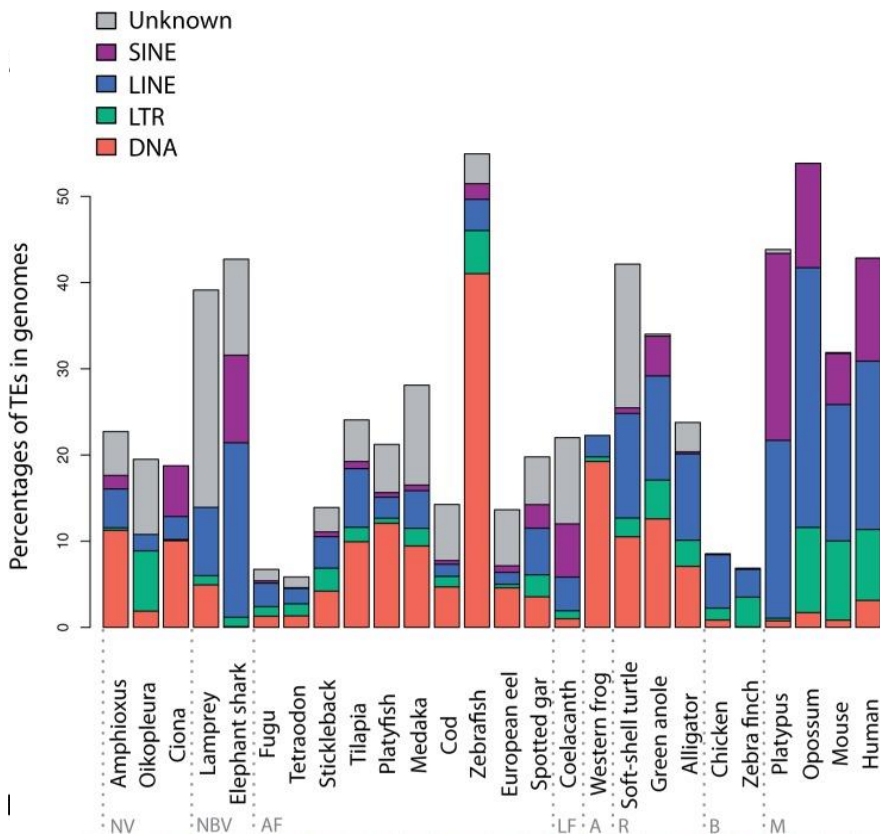
Aunque en muchos casos las transposasas son domesticadas en su totalidad, su capacidad para unirse en *trans* a elementos de ADN dispersos por el genoma parece el principal motivo de la ubicuidad de estos TDs. De hecho, la mayoría de transposasas domesticadas carecen de capacidad catalítica y han sido cooptadas para desempeñar un papel de reguladores transcripcionales. Por ejemplo, en plantas, la familia de factores de transcripción FHY3 derivada de transposones *Mutator* está involucrada en la integración de estímulos lumínicos. O los genes *abp1*, *cbh1* y *cbh2*, presentes en levaduras, un caso de domesticación especialmente interesante por dos razones. Primero, porque constituye un caso de convergencia evolutiva: los tres genes de levadura y el gen *CENP-B* de mamíferos fueron independientemente domesticados a partir de transposasas de la familia *Pogo*, y todos ellos se unen a los centrómeros permitiendo una correcta segregación cromosómica. Y segundo porque, curiosamente, la proteína Abp1 también ha sido reclutada para el silenciamiento de retrotransposones LTR<sup>71,72</sup>.

Por lo tanto, la domesticación de transposones es un fenómeno filogenéticamente extendido, que ha desempeñado un importante papel evolutivo tanto en animales y plantas como en hongos. Sin embargo, el taxón

donde mejor se han caracterizado y donde más diversidad de casos se ha encontrado es el de los mamíferos.

### *TDs en mamíferos*

De entre los animales amniotas con el genoma secuenciado, los mamíferos son los que contienen un mayor porcentaje de transposones<sup>73</sup> (fig. 1). Un ~40% está compuesto por elementos móviles activos o fósiles, con una clara predominancia de los retrotransposones; hay algunas excepciones, como el género de murciélagos *Myotis*<sup>74</sup>.



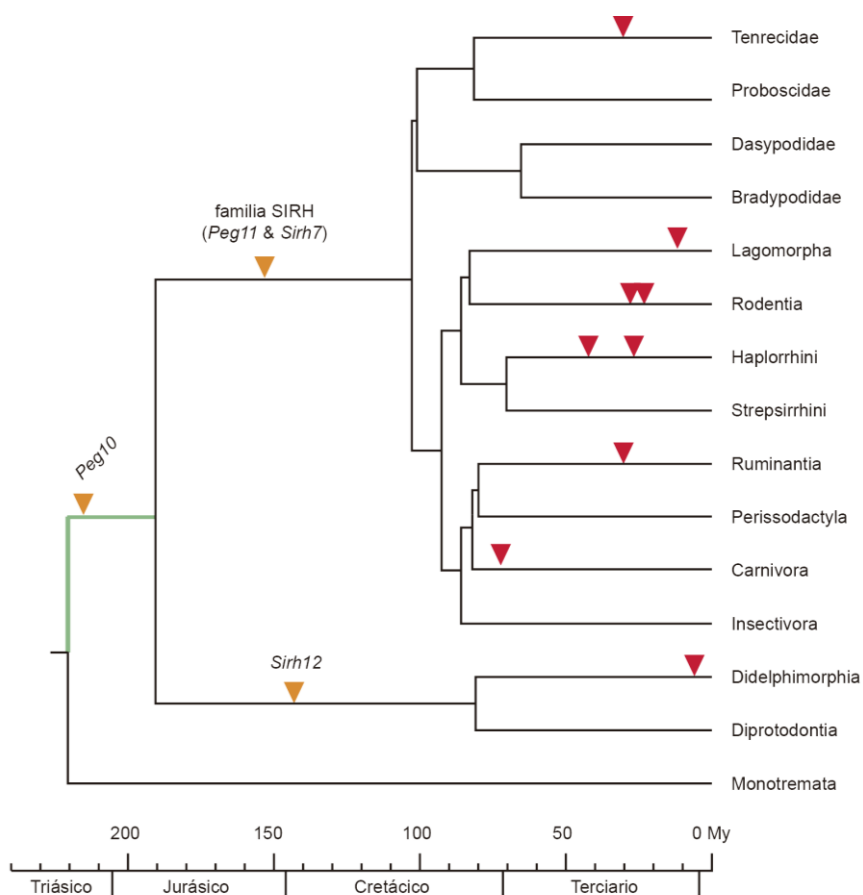
**Figura 1.** Porcentaje de transposones de ADN, retrotransposones LTR, LINE y SINE, y elementos desconocidos (*Unknown*) en diversos genomas de deuteróstomos. Estimación basada en las anotaciones RepeatMasker. NV, no vertebrados; NBV, vertebrados no óseos; AF, peces actinopterygios; LF, peces de aletas lobuladas; A, anfibios; R, reptiles (excluyendo pájaros); B, aves; M, mamíferos. Figura adaptada de Chalopin et al.<sup>73</sup>

Como en el caso del sistema inmune adaptativo de los gnatóstomos, la placenta de los mamíferos terios es otra innovación evolutiva en la que los TDs han participado de un modo fundamental. Los terios son una subclase de mamíferos que agrupa a los mamíferos metaterios y euterios (incluyendo los clados que agrupan el último ancestro común de sus descendientes vivos hoy en día, es decir, los marsupiales y los placentarios, respectivamente). Las hembras de estos animales vivíparos gestan internamente a sus embriones gracias a la formación de una placenta, permitiendo así el intercambio materno-fetal de gases y nutrientes. Aunque ambos grupos desarrollan esta estructura, la de los placentarios es mucho más duradera y compleja.

Los genes de las sincitinas se identificaron a principios del siglo XXI como los primeros TDs derivados de ERVs en mamíferos, y constituyen un ejemplo paradigmático de cómo las distintas propiedades de los transposones (su variedad, ubicuidad y capacidad de persistencia, así como su función intrínseca) pueden ser aprovechadas para generar nuevas y cruciales estructuras. Estos genes, cooptados a partir de un gen *env*, se expresan en la placenta y mantienen la actividad inmunosupresora y/o fusogénica de la proteína transposónica ancestral. Poco tiempo después de su identificación, otros dos genes con características y funcionalidades similares fueron detectados en roedores. Mediante la generación de ratones knockout, se comprobó que las sincitinas son esenciales para la fusión de las células que formarán el sincitiotrofoblasto, una estructura multinucleada clave durante la placentación. Además, ayudarían a la tolerancia materno-fetal mediante su capacidad inmunosupresora.

Sorprendentemente, las sincitinas murinas y humanas, así como las identificadas en otros clados de mamíferos y marsupiales, han surgido a partir de eventos de domesticación independientes específicos de cada linaje (fig. 2). Dado que todos ellos generan una placenta con sincitiotrofoblasto, surge la pregunta de cómo explicar el desfase en la coopción de genes aparentemente esenciales durante su formación. Para resolver esta paradoja, se propuso el

siguiente modelo<sup>75</sup>: antes de la diversificación de los mamíferos placentarios (incluidos los marsupiales), el viviparismo y la formación de la placenta habrían sido facilitados por la domesticación de una sincitina ancestral, que fue sustituida paulatinamente en cada linaje por recurrentes domesticaciones de genes *env*. Esta atractiva hipótesis, que se ve respaldada por la convivencia en ciertos linajes de nuevas y «viejas» sincitinas en proceso de pseudogenización, podría explicar también parte de la variabilidad estructural observable en las placentas de los mamíferos.



**Figura 2.** Filogenia de los mamíferos mostrando algunos eventos de domesticación molecular. Los triángulos naranjas marcan las dos domesticaciones independientes de genes *SIRH*, y la expansión de la familia en el ancestro de los mamíferos placentarios. Los genes *env* de ERVs han sido recurrentemente domesticados en los linajes de terios para dar lugar a nuevas sincitinas (triángulos rojos). La rama delineada en verde marca la aparición de la placenta en el origen de los terios. Figura adaptada de Cornelis et al.<sup>236</sup> y Kaneko-Ishino e Ishino<sup>76</sup>.

Al mismo tiempo que se caracterizaron las sincitinas, se identificaron los genes de la familia *SIRH* (*Sushi-Ichi-related Retrotransposon Homologs*), también conocida como *Mart* (*Mammalian RetroTransposon*), derivados de los genes *gag* y *pol* de un retrotransposón LTR Ty3/Gypsy. La familia surgió antes de la separación de los terios a partir del gen *Peg10/Mart2*, y fue amplificada en euterios vía duplicación hasta los once miembros (fig. 2). Ocho de los genes *SIRH* de euterios se expresan en la placenta, algunos de ellos con funciones esenciales. Por ejemplo, los genes *Peg10/Mart2* y *Peg11/Rtl1/Mart1*, cuya ablación en ratones produce malformaciones letales en la placenta, son necesarios para la formación y posterior mantenimiento de esta estructura, respectivamente<sup>76</sup>.

Además, otros genes *SIRH* estarían implicados en la evolución del cerebro de los euterios: ratones knockout para *Zcchc16/Mart4/Sirh11* muestran diversas deficiencias cognitivas asociadas al sistema noradrenérgico<sup>77</sup>. Podemos concluir que las sincitinas y la familia *SIRH* son un ejemplo de cómo los TDs participan en el desarrollo de innovaciones macroevolutivas del grupo de los mamíferos, como la aparición de la placenta en terios (domesticación de las sincitinas y *Peg10/Mart2*), y su mantenimiento prolongado en euterios (gracias a *Peg11/Rtl1/Mart1*).

Por lo tanto, el estudio de los TDs pone de relieve otro de los mecanismos genéticos que permiten la innovación evolutiva, además de brindarnos la oportunidad de investigar los procesos asociados al nacimiento de nuevos genes. Es muy probable que solo conozcamos una pequeña parte de los eventos de domesticación molecular; recientemente, el desarrollo de estrategias *ad hoc* ha empezado a llenar este vacío de conocimiento<sup>61</sup>. Por otro lado, la comprensión de la función de los escasos TDs investigados es todavía limitada.

## **Objetivos**





Este trabajo comienza con una búsqueda en el genoma humano de nuevos casos de TDs. Al identificar como candidato el gen *Tceal7* (miembro de la familia génica *Bex/Tceal*), se procede a estudiar el origen, la regulación y la función de esta familia y, por extensión, del clúster génico específico de mamíferos euterios al que pertenece. Después de encontrar evidencias concluyentes de que los *Bex/Tceal* derivan de un evento de domesticación molecular de fragmentos de retrotransposones L1 en el origen de los euterios, se pasa a estudiar la adquisición de regiones reguladoras, ya desde la óptica de los TDs. En este punto, se propone una reconstrucción de la historia evolutiva del clúster, intentando resolver a su vez algunas de las preguntas respecto del origen y la función de las regiones reguladoras características de estos genes. Una vez propuesto un panorama plausible del establecimiento del clúster, se pasa a ahondar en la caracterización de la función de los genes *Bex/Tceal*, tratando de arrojar luz sobre su integración y las implicaciones evolutivas de su aparición.

1. Identificación de candidatos a TD en el genoma humano mediante el solapamiento de las anotaciones de regiones codificantes y transposones.
2. Análisis del proceso de domesticación molecular del gen *Tceal7*, así como del origen y la regulación de la familia y del clúster génico al que pertenece.
3. Estudio de la expresión tisular durante el desarrollo de los genes de la familia *Bex/Tceal*.
4. Generación y análisis del fenotipo de líneas de ratones mutantes para *Bex3*.



## **Métodos**



## Búsqueda de transposones domesticados

Las anotaciones de los genes codificantes y de los elementos repetitivos del ensamblaje del genoma humano hg38 se descargaron del navegador de genomas UCSC, utilizando el navegador de tablas y seleccionando «UCSC genes» y «RepeatMasker», respectivamente. Para descartar elementos no transposónicos, se filtró el archivo del RepeatMasker como en Kapusta *et al.*<sup>78</sup>. Solo se tuvieron en cuenta los candidatos con un txCdsPredict superior a 800 (predictivo en un ~90% de genes codificantes).

## Obtención de secuencias

Para identificar los genes de la familia *Bex/Tceal* en especies de euterios, se llevaron a cabo búsquedas con BLASTn o BLAT en la base de datos NCBI y UCSC, utilizando las secuencias nucleotídicas de humano. Para establecer las relaciones de ortología entre los genes que sufren conversión génica, se incluyeron también, al realizar la búsqueda, regiones circundantes al gen humano correspondiente. Los ensamblajes de los genomas analizados son los siguientes: GRCh38/hg38 para humano (*Homo sapiens*); GRCm38/mm10 para ratón (*Mus musculus*); Broad CanFam3.1/canFam3 para perro (*Canis familiaris*); Bos\_taurus\_UMD\_3.1.1/bosTau8 para vaca (*Bos taurus*); Broad/equCab2 para caballo (*Equus caballus*); Baylor/dasNov3 para el armadillo de nueve bandas (*Dasypus novemcinctus*); Broad/choHof1 para el perezoso didáctilo de Hoffmann (*Choloepus hoffmanni*); Broad/monDom5 para el colicorto gris (*Monodelphis domestica*); WTSI Devil\_ref v7.0/sarHar1 para el demonio de Tasmania (*Sarcophilus harrisii*), y TWGS Meug\_1.1/macEug2 para el ualabí (*Macropus eugenii*). A su vez, las secuencias consenso de las familias transposónicas fueron descargadas de la librería RepeatMasker derivada de RepBase (actualización del 27/01/2017; <http://www.girinst.org/server/RepBase/>).

## **Análisis bioinformático**

Los análisis estructurales de las proteínas BEX/TCEAL se llevaron a cabo con las versiones web de los programas PONDR-FIT y PCOILS, seleccionando los parámetros por defecto. Las secuencias de nucleótidos y las de aminoácidos fueron alineadas con el algoritmo L-INS-i del programa MAFFT<sup>79</sup>, y los alineamientos resultantes fueron editados con JalView<sup>80</sup>. Los dominios proteicos fueron identificados con el buscador de dominios del NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) o con Pfam 31.0 (<http://pfam.xfam.org/>). Cuando fue necesario traducir una secuencia nucleotídica en aminoácidos, se utilizó el servidor ExPASy<sup>81</sup>. Se utilizó el programa MEGA6<sup>82</sup> para calcular los ratios Nei-Gojobori de mutaciones no sinónimas vs sinónimas (dN/dS), y para obtener ficheros de entrada para los análisis filogenéticos con inferencia bayesiana, que fueron llevados a cabo utilizando MrBayes 3.2.6<sup>83</sup>. Con los programas MrModelTest2 y ProtTest 2.4<sup>84</sup> se escogieron los modelos según el criterio de información de Akaike (AIC) (GTR+G en la fig. 4a; GTR+G+I en las fig. 4b y 4c; y JTT+G+I en la fig. 9). Las muestras con una desviación estándar de las frecuencias divididas >0.01 fueron descartadas. Finalmente, los árboles filogenéticos fueron visualizados y editados con el programa FigTree v1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

## **Análisis de regiones reguladoras**

Las secuencias seleccionadas fueron amplificadas mediante reacción en cadena de la polimerasa (PCR) a partir de ADN genómico humano, murino o de koala (las secuencias amplificadas se pueden consultar en la tabla X), y clonadas direccionalmente en un vector reportero *hsp68-lacZ* con el sistema Gateway. Las secuencias se clonaron, en relación al promotor mínimo *hsp68*, en la orientación original respecto al gen humano o murino (en el caso del proto-motivo BGW de koala, la secuencia se clonó en la orientación contraria a la transcripción de *Gla*).

Los plásmidos se purificaron, se linearizaron y posteriormente fueron purificados de nuevo a partir de gel. El ADN fue diluido hasta una concentración de 2 ng/μl e inyectado en pronúcleos de embriones FVB, de acuerdo a protocolos aprobados por el Comité de Experimentación Animal de la Generalitat de Catalunya. Los embriones fueron recolectados a 13,5 días post-coito (dpc) y diseccionados, aislando el saco vitelino. Se extrajo ADN del saco vitelino digiriéndolo a 55°C durante >12 horas en tampón de cola (10 mM Tris, pH 7.5; 400mM NaCl; 100 mM EDTA; 0,6% SDS), el cual se utilizó para genotipar los embriones mediante PCR, utilizando los cebadores UNIVERSAL\_F (5'-AGCAGGCTCCGAATTCGCCC-3') y β\_GAL\_R2 (5'-CGACGGGATCATCGCGAGCC-3'). La expresión de la β-galactosidasa en los embriones transgénicos se detectó tiñendo con X-Gal, siguiendo los pasos descritos en Dodou et al.<sup>85</sup>. Las imágenes se obtuvieron con una lupa Leica MZ16 F.

### **Hibridaciones *in situ***

Parejas de cebadores fueron diseñadas para amplificar mediante PCR regiones de los genes *Bex/Tceal* de *Mus musculus* a partir de ADNc. Los fragmentos amplificados fueron clonados en vectores pGEMTeasy (Promega), a partir de los cuales se sintetizaron las ribosondas con DIG RNA labeling mix (Roche). Los embriones de ratón fueron recolectados a 13.5 dpc, crioprotegidos, cortados al criostato en secciones de 20 μm, sometidos a hibridación *in situ* y montados, tal y como se ha descrito anteriormente<sup>86</sup>. Se utilizó nitroazul de tetrazolio/5-bromo-4-cloro-3-indolil fosfato (NBT/BCIP) o BMP *purple* (Roche) como sustrato cromogénico para la fosfatasa alcalina. Las imágenes se tomaron con una lupa Leica MZ16 F.

### **Generación de ratones KO**

Con el objetivo de generar deleciones en el ORF del gen *Bex3* mediante el sistema CRISPR/Cas9, se diseñaron ARNg con CRISPRscan<sup>87</sup> y la herramienta



para el diseño de CRISPR del grupo de Feng Zhang (<http://crispr.mit.edu/>). Los cebadores utilizados para genotipar las líneas de *Bex3* fueron Bex3-F2 (5'-AGCCCCACTCCACACTACT-3') y Bex3-R (5'-TGGTGATCGTGGTGGTTAGA-3').

### ***Western Blot***

Las muestras de ratones mutantes o WT fueron diseccionadas y se procedió inmediatamente a la extracción (alternativamente, se congelaron en nitrógeno líquido para su posterior uso). El tejido se homogeneizó en *buffer* de radioinmunoprecipitación (RIPA) con inhibidores de la fosfatasa (fluoruro sódico, ortovanadato sódico y pirofosfato ácido de sodio) y de la proteasa (cOmplete™, Roche), con la ayuda de un POLYTRON PT 1200 E. Las concentraciones totales de proteína se cuantificaron con un Micro BCA Protein Assay Kit (Thermo Fisher Scientific), y las muestras de cada carril se diluyeron hasta la misma concentración. Los anticuerpos primarios utilizados en este estudio fueron Bex3 (1:1000, ThermoFisher, PA5-20076; 1:1000, Novus Biologicals, NBP1-77149), fosfo-S6 (1:1000, Cell Signaling, #2211), S6 (1:1000, Cell Signaling, #2217), GAPDH (1:10000, Sigma, G9545). Como secundario, se utilizó un anticuerpo conjugado con HRP (Polyclonal Swine Anti-Rabbit Immunoglobulins/HRP, DAKO, P0217), diluido 1:1000 en TBS-T con un 5% de leche en polvo.

Las proteínas se visualizaron mediante quimioluminiscencia con ECL plus (Amersham Pharmacia Biotech), y la cuantificación se llevó a cabo con ImageJ. En los WB para medir la activación de la vía mTOR en el encéfalo de ratones Bex3-L2, se utilizaron ratones de entre 12 y 14 semanas. Los gráficos representan al menos 4 animales por cada grupo.

## **Resultados**



## Búsqueda de genes domesticados

Con la idea de explorar la contribución de los TDs al repertorio de proteínas en humano, nos propusimos identificar genes codificantes que solapasen con transposones (en colaboración con Pol Cuscó, del Centro de Regulación Genómica). Reuniendo las anotaciones en el genoma humano para estos genes y para los elementos transponibles, filtramos los casos de acuerdo a los siguientes criterios: (i) genes con un solapamiento entre el ORF y secuencias de origen transposónico superior al 50%, (ii) potencial codificante significativo y (iii) conservación del ORF en más de una especie. Aunque la no conservación respecto a especies cercanas no excluye que el candidato sea un TD, este último criterio se incluyó para disminuir el número de falsos positivos. Así, obtuvimos una lista de 67 candidatos (tabla 1), incluyendo casos de TDs ya documentados, como los genes de las *sincitinas-1* y  $2^{88,89}$ , o las familias génicas *ZBED*<sup>90</sup> y *PNMA*<sup>91</sup>. Además, también se identificó un nuevo gen potencialmente interesante, llamado *Tceal7*, que se decidió investigar a fondo (para los candidatos *AK127846* y *MGC12965*, ver anexos I y II, respectivamente).

**Tabla 1. Genes codificantes con >50% del ORF solapante con un transposón (txCdsPredict >800)**

	ID	% Solap.	txCdsPredict		ID	% Solap.	txCdsPredict
1	<i>L1TD1</i>	53,81	5296,67	35	<i>PNMA2</i>	92,7	2302
2	<i>L1TD1</i>	53,81	5296,67	36	<i>PNMA1</i>	96,14	2279
3	<i>GTF2IRD2B</i>	100	4988,33	37	<i>ERVFRD-1</i>	57,05	2188
4	<i>GTF2IRD2</i>	55,99	4911	38	<i>AK127846</i>	100	2124
5	<i>GTF2IRD2</i>	99,93	4564,33	39	<i>MOAP1</i>	96,5	2045,33
6	<i>GTF2IRD2B</i>	75,81	3849,33	40	<i>MOAP1</i>	96,5	2045,33
7	<i>ZBED5</i>	97,94	3590,5	41	<i>PNMA6A</i>	73,19	1972
8	<i>ZBED5</i>	97,94	3587,5	42	<i>ERVMER34-1</i>	53,81	1965,67
9	<i>SETMAR</i>	71,84	3181,83	43	<i>ERVMER34-1</i>	53,81	1873,67
10	<i>ZMYM6</i>	69,92	3128,17	44	<i>SETMAR</i>	80,31	1852,83
11	<i>C5orf54</i>	100	3097,83	45	<i>ERV3-1</i>	100	1829,5
12	<i>GTF2IRD2B</i>	68,3	3075,33	46	<i>SETMAR</i>	84,64	1717,5
13	<i>ZBED5</i>	97,94	3056	47	<i>SETMAR</i>	50,34	1691,17
14	<i>C5orf54</i>	100	3036,83	48	<i>ERVW-1</i>	100	1630,5
15	<i>JRK</i>	90,57	3028,5	49	<i>pol</i>	99,87	1607,5
16	<i>FAM200B</i>	100	2993	50	<i>ERVK6</i>	100	1583,5
17	<i>JRK</i>	92,47	2938,5	51	<i>ERVK6</i>	99,93	1583,5
18	<i>JRK</i>	92,47	2938,5	52	<i>ERVW-1</i>	100	1257
19	<i>FAM200A</i>	94,02	2929,5	53	<i>pol</i>	100	1192
20	<i>PNMA3</i>	72,72	2872,33	54	<i>MGC12965</i>	54,66	1148
21	<i>PNMA3</i>	73,94	2854,33	55	<i>DQ587413</i>	100	986
22	<i>FAM200A</i>	94,02	2818	56	<i>AK124340</i>	100	973
23	<i>PNMA3</i>	72,72	2805,33	57	<i>AX747630</i>	52,94	897
24	<i>GTF2IRD2</i>	63,1	2576	58	<i>env</i>	100	882
25	<i>GTF2IRD2B</i>	56,02	2576	59	<i>env</i>	99,86	882
26	<i>PNMA5</i>	60,76	2491,67	60	<i>AX747977</i>	58,85	851,5
27	<i>PNMA5</i>	60,76	2491,67	61	<i>HKR1</i>	50,19	844,667
28	<i>ZBED5</i>	97,94	2489	62	<i>TMEM220-AS1</i>	54,48	834,667
29	<i>SETMAR</i>	63,09	2486,83	63	<i>AX746793</i>	99,73	824,5
30	<i>PNMA5</i>	60,76	2444,33	64	<i>JAKMIP2-AS1</i>	63,3	811,667
31	<i>PNMA5</i>	60,76	2436,67	65	<i>TCEAL7</i>	76,32	804
32	<i>TIGD1</i>	100	2366	66	<i>SDCCAG8</i>	58,67	801
33	<i>ZMYM5</i>	51,11	2329	67	<i>SDCCAG8</i>	58,67	801
34	<i>PNMA6A</i>	73,19	2306,67				

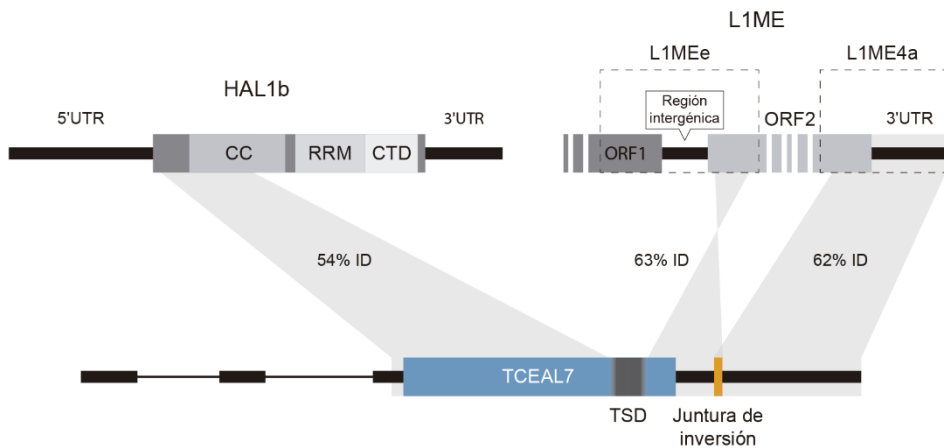
Negro: TDs previamente documentados

Rojo: candidatos no anotados o anotados como no codificantes en GENCODE v24

Verde: candidatos a TD

## Análisis de la exaptación (I) – Nacimiento de un nuevo gen

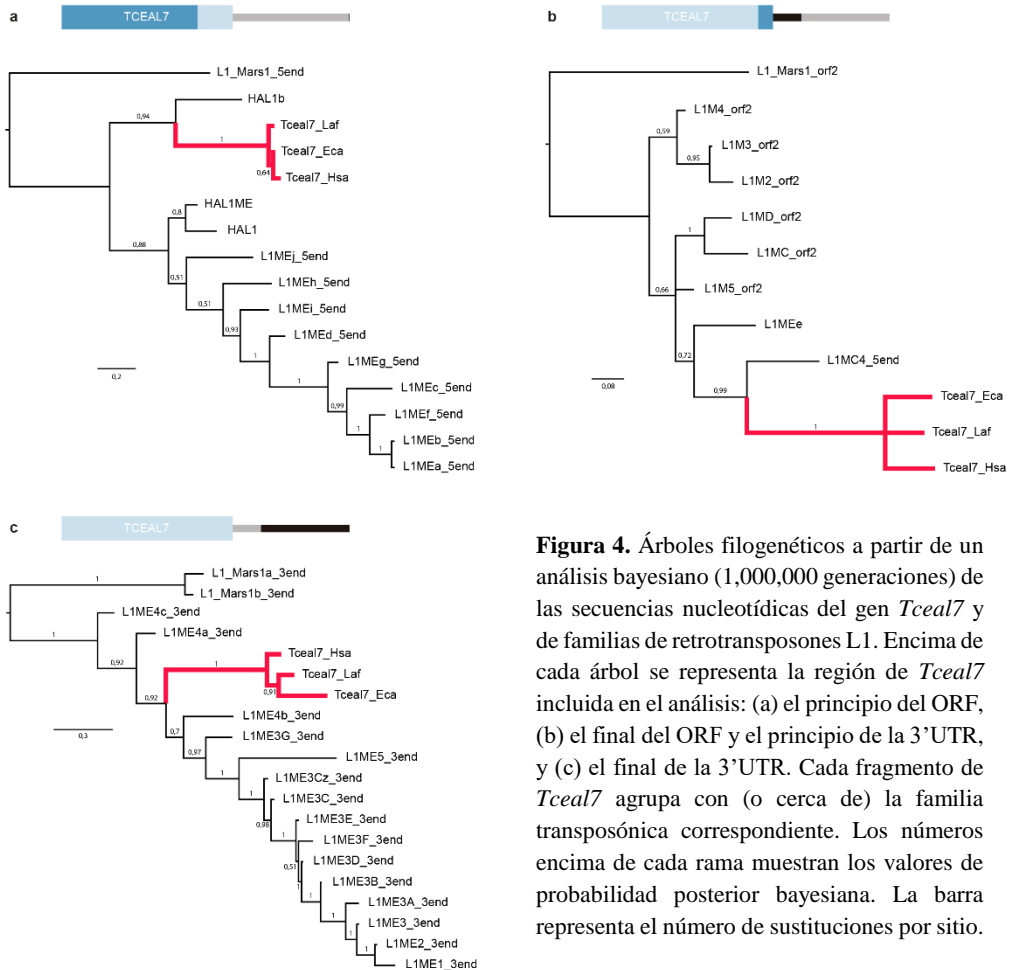
*Tceal7* es un gen de pequeño tamaño que consta de dos exones no codificantes en 5' y un tercer exón que incluye la totalidad del ORF y la 3'UTR. Según nuestro análisis, el 76% de su ORF solapa con la copia de un retrotransposón L1 llamado HAL1b. Sorprendentemente, mientras se realizaba una consulta en la versión hg19 del genoma humano en el buscador UCSC, se vio que la contribución de los transposones a la estructura de *Tceal7* parecía ser aún mayor: según la anotación de elementos repetitivos implementada por el software RepeatMasker, los últimos nucleótidos del ORF, el codón stop y la mayor parte de la 3'UTR solapaban con copias de otras dos familias L1 (un inserto L1Mee —invertido respecto a la orientación de *Tceal7*—, y un L1ME4a) (fig. 3).



**Figura 3.** Esquema de la domesticación molecular del ORF y la 3'UTR de *Tceal7* a partir de retrotransposones L1. El porcentaje de identidad (ID) se obtuvo al alinear las secuencias transposónicas y el fragmento correspondiente de *Tceal7* humano. Las subfamilias L1MEe y L1ME4a se consideran como pertenecientes a un mismo transposón L1ME. En el ORF1 de HAL1b están representados los dominios proteicos de ORF1p: CC, *coiled-coil*; RRM, *RNA-recognition motif*; y CTD, *carboxy-terminal domain*.

Como se comentó en la introducción, un elemento L1 típico contiene dos ORFs separados por una región intergénica, además de 5' y 3'UTRs. Por otro lado, es importante recalcar que RepBase Update<sup>92</sup>, la base de datos de la que deriva RepeatMasker, elabora la secuencia consenso de cada familia a partir de aquellas copias que compartan al menos un 75% de identidad, siguiendo, por lo tanto, un criterio filogenético basado en ADN. De este modo, las secuencias consenso intentan reconstruir la familia originalmente activa. Debido al estado fragmentario y pobremente conservado de las copias de transposones antiguas, algunas familias (o, más apropiadamente, subfamilias) representan solo partes del inserto original. Específicamente, la subfamilia L1MEe comprende el final del ORF1 y el principio del ORF2 de un L1 ancestral, mientras que L1ME4a incluye el final del ORF2 y la 3'UTR. A su vez, HAL1b (HALf L1 b) forma parte de una categoría especial de L1 no autónomos que carecen de ORF2<sup>43</sup>.

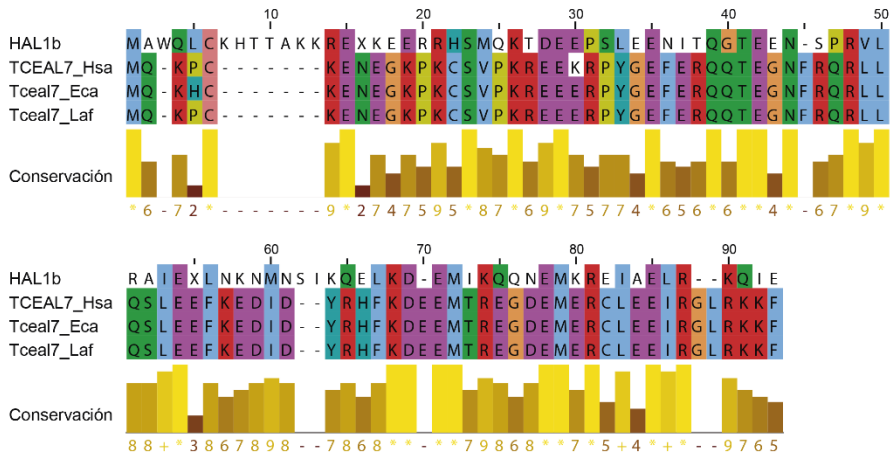
De acuerdo con lo anterior, nos propusimos comprender la enigmática naturaleza de *Tceal7*. El alineamiento de los segmentos de las tres familias de L1 y las secuencias correspondientes del gen *Tceal7* arrojó unos valores de similitud e identidad que oscilan entre el 51 y el 63% (fig. 3). Para confirmar la homología, se llevó a cabo un análisis filogenético para cada uno de los tres fragmentos de *Tceal7* —solapantes a HAL1b, L1Mee y L1ME4a— junto con otras subfamilias de L1 ancestrales<sup>93</sup>. Encontramos que cada segmento de *Tceal7* se agrupa junto a, o cerca de, la familia correspondiente (fig. 4). Además, la secuencia aminoacídica de *Tceal7* derivada de HAL1b dio unos valores de identidad del 28% y una similitud del 34% cuando se la alineó con el ORF1p de HAL1b (fig. 5). Con estos resultados, podemos determinar que gran parte del gen *Tceal7* surgió a partir de un mosaico de familias L1, reteniendo el ORF de la secuencia transposónica parental.



**Figura 4.** Árboles filogenéticos a partir de un análisis bayesiano (1,000,000 generaciones) de las secuencias nucleotídicas del gen *Tceal7* y de familias de retrotransposones L1. Encima de cada árbol se representa la región de *Tceal7* incluida en el análisis: (a) el principio del ORF, (b) el final del ORF y el principio de la 3' UTR, y (c) el final de la 3' UTR. Cada fragmento de *Tceal7* agrupa con (o cerca de) la familia transposónica correspondiente. Los números encima de cada rama muestran los valores de probabilidad posterior bayesiana. La barra representa el número de sustituciones por sitio.

Los elementos L1 se insertan en el genoma mediante un proceso llamado retrotranscripción dirigida por el sitio diana (*target-primed reverse transcription*, TPRT)<sup>45</sup> (fig. 6a, b). Pese a contar con centenares de miles de copias en el genoma humano, prácticamente todos los elementos L1 presentan alteraciones en la arquitectura genética que imposibilitan su actividad. De hecho, gran parte de las copias están truncadas (o truncadas e invertidas) debido a defectos o variantes de la TPRT<sup>94</sup>. Uno de los mecanismos no canónicos propuestos para explicar las inserciones invertidas es el de *twin priming*<sup>95</sup> (fig. 6c, d). Estos insertos comparten una serie de características, entre ellas la presencia de 1-4 pares de bases (pb) en la juntura de la inversión (*inversion*

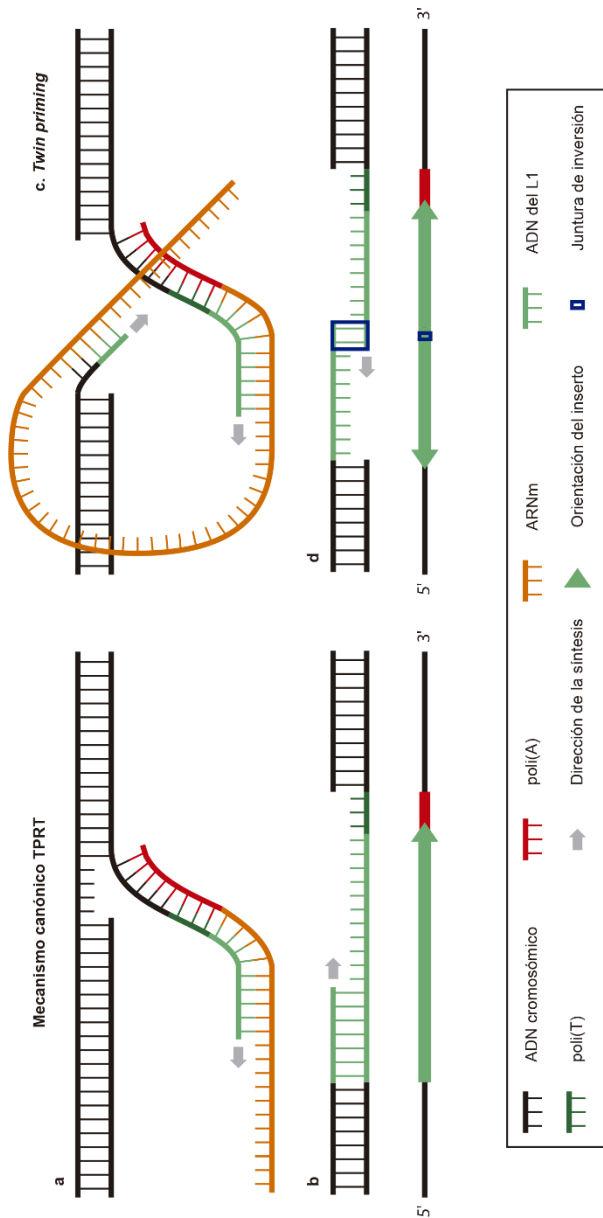




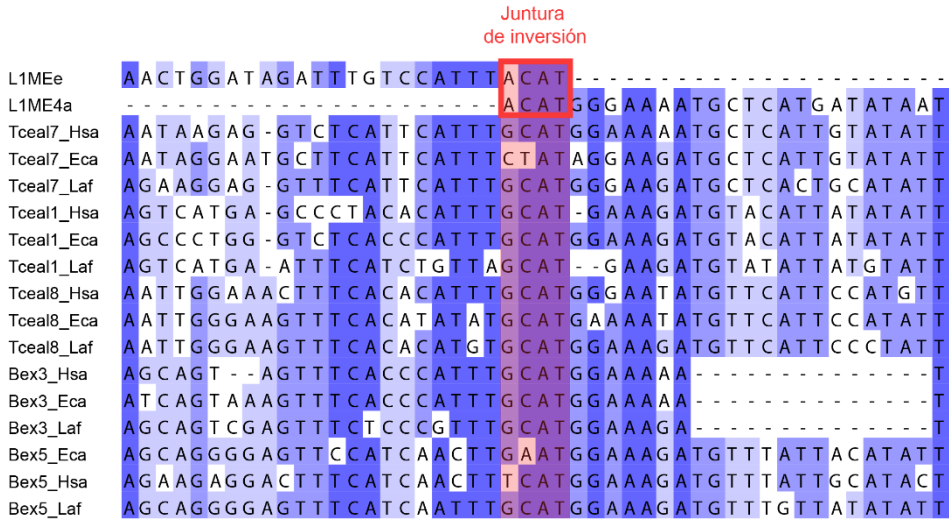
**Figura 5.** Alineamiento visualizado con JalView (colores Clustalx) de las proteínas TCEAL7 y ORF1p de HAL1b. Eca, caballo; Hsa, humano; Laf, elefante africano.

*junction*), que pueden proceder tanto de la secuencia invertida del L1 como de la no invertida.

Relevantemente, encontramos 4 pb en la 3'UTR de *Tceal7* que se podían asignar a la secuencia consenso del L1MEe invertido o a la del L1ME4a (fig. 7). Esta microsimilitud induce a pensar que las secuencias L1MEe y L1ME4a que componen parte del gen *Tceal7* son en realidad fragmentos de un mismo inserto L1 invertido y truncado en 5'. Esta copia L1ME se integró en un elemento HAL1b preexistente, generándose una TSD en su extremo 5' truncado que se corresponde con los nucleótidos comprendidos entre las partes derivadas de HAL1b y L1MEe de *Tceal7* (fig. 3). Por lo tanto, podemos concluir que la inserción de un elemento L1 ancestral en una copia HAL1b, seguida de un proceso de coopción, originó el ORF y la 3'UTR del gen *Tceal7*.



**Figura 6.** Mecanismo de transposición canónico de los elementos L1 y su variante *twin priming*. (a) El mecanismo canónico de transposición comienza con la proteína ORF2p creando un corte en la cadena complementaria del sitio diana (con secuencia consenso 5'-TTTT/A-3') que genera un grupo hidroxilo libre<sup>38,237</sup>. La poli-A del extremo 3' del mRNA transposónico se hibrida al corte, y la actividad transcriptasa reversa de ORF2p inicia la síntesis del ADN complementario (ADNc) del L1 utilizando los nucleótidos del grupo hidroxilo como cebador<sup>238</sup>. (b) A continuación, la cadena superior del ADN genómico sufre otro corte unos nucleótidos más abajo del primero, y se sintetiza la secuencia complementaria restante del ADNc del L1. El resultado de la TPRT es una nueva inserción con unos rasgos característicos: TSDs intactos a ambos lados del inserto, motivo endonucleasa en el sitio de escisión, y una poli-A de longitud variable<sup>94,239</sup>. (c) El intrincado proceso de *twin priming* transcurre del mismo modo que la TPRT hasta el punto en que se inicia la síntesis de la primera cadena de ADNc. Antes de que ésta termine, ORF2p crea el segundo corte en la cadena de ADN superior, dando lugar a un grupo hidroxilo libre adicional y una secuencia de ADN de cadena sencilla. Esta secuencia, llamada cebador interno, se hibrida al mRNA del L1 y permite el inicio de la retrotranscripción desde un punto interno y alternativo al original. Una vez eliminado el mRNA del transposón, ambas secuencias de ADNc se unen en una región de microsimitud de 1 a 4 nucleótidos llamada junta de inversión, y la síntesis restante de ADN es completada. (d) El resultado es una inserción parcialmente invertida, que puede estar internamente deletionada (en función de cuán extensa haya sido la transcripción del primer ADNc) y truncada (si la transcripción del segundo ADNc no es completa), flanqueada por TSDs. Figura adaptada de Meyer et al.<sup>147</sup>.



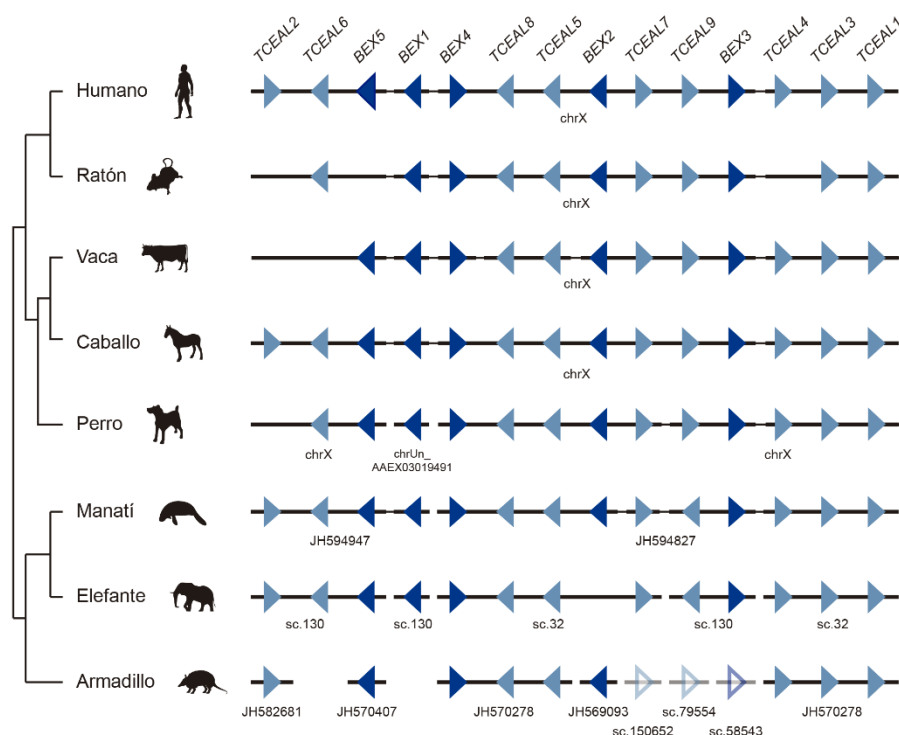
**Figura 7.** Alineamiento nucleotídico visualizado con JalView de la región de las secuencias consenso de L1MEe y L1ME4a, y de los genes *Bex/Tceal* correspondiente a la juntura de inversión (sombreada en rojo). Encuadrada también en rojo está la región de microsimilitud entre L1MEe (cuya secuencia está invertida en el alineamiento) y L1ME4a involucrada en el proceso de *twin priming*.

## Análisis de la exaptación (II) – Nacimiento de una nueva familia

*Tceal7* pertenece a la familia génica *Bex/Tceal*, formada por las subfamilias *Bex* (*Brain-expressed X-linked*) y *Tceal* (*Transcription elongation factor A-like*). Esta familia forma un clúster en el cromosoma X de los mamíferos placentarios, y carece de ortólogos fuera de ese clado<sup>96</sup>. En humanos, el clúster consta de 14 pequeños genes (5 *BEX* y 9 *TCEAL*) y abarca ~1,5 megabases. En ratón, debido a pérdidas y ganancias específicas de linaje, la familia consta de 12 miembros. Aunque Winter y Ponting sugirieron que la familia era específica de euterios, solo ha sido caracterizada en su totalidad en estas dos especies<sup>96</sup>. Utilizando secuencias de humano, analizamos la presencia de miembros de la familia *Bex/Tceal* en los genomas de varias especies representativas. Así, se pudieron detectar ortólogos de la mayoría de los genes humanos en las especies analizadas, con la excepción del linaje de los xenartros, que parece haber sufrido numerosas pérdidas —el armadillo de nueve bandas

(*Dasybus novemcinctus*), por ejemplo, carece de 5 de los 14 genes *Bex/Tceal*— (fig. 8). Si consideramos que los placentarios se dividen en atlantogenados y boreoeuterios, o en afroterios y exafroplacentarios<sup>97</sup>, podemos determinar que, esencialmente, el origen y la expansión de la familia *Bex/Tceal* sucedió entre la divergencia de los metaterios y los euterios, y la radiación de los mamíferos placentarios. Apoyando esta teoría, las familias HAL1b, L1MEe y L1ME4a solo se encuentran en euterios, y estuvieron activas durante el origen de estos, hace unos 150 millones de años<sup>98</sup>.

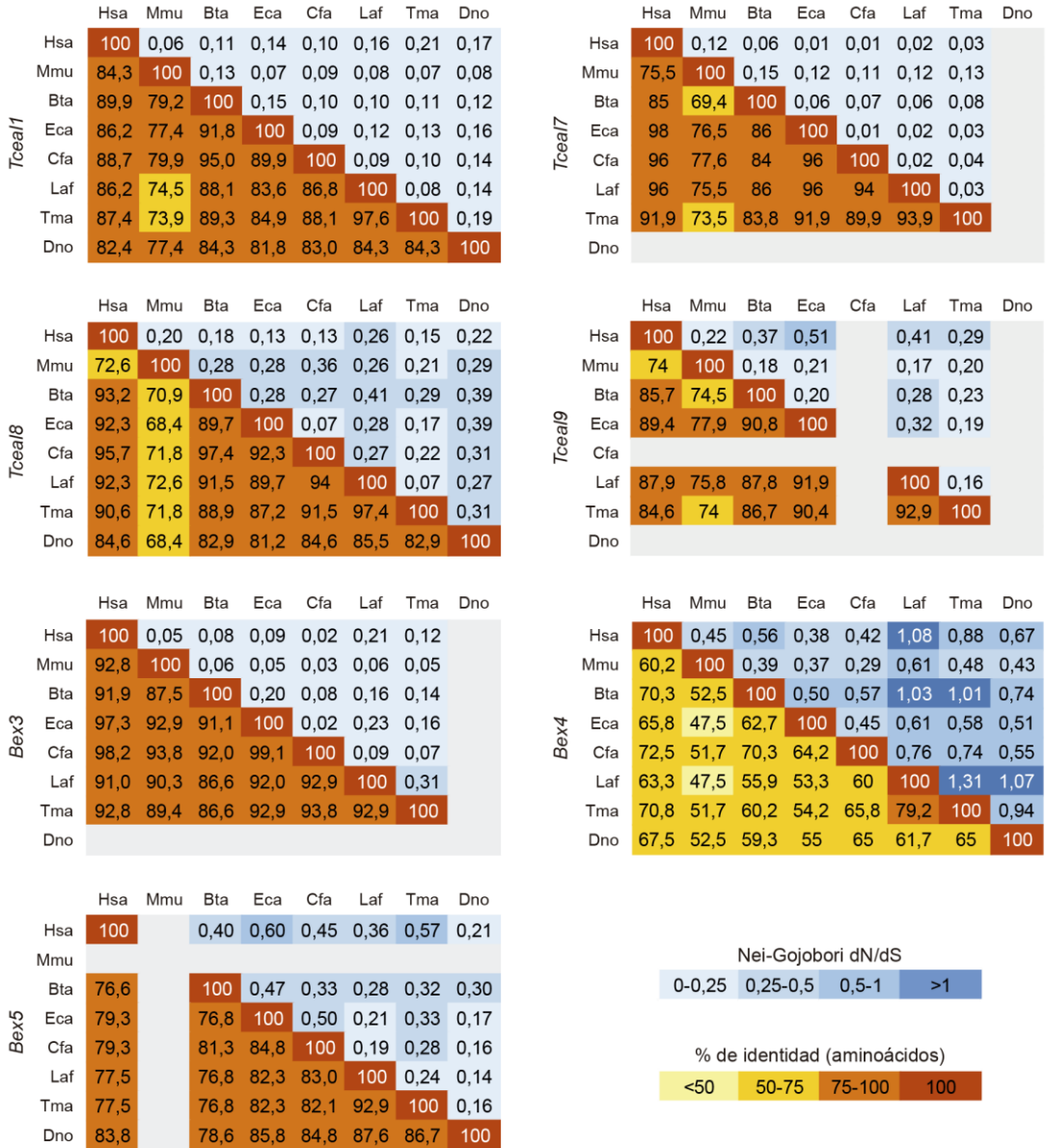
Según estudios anteriores<sup>96,99</sup>, las regiones codificantes de los genes *Bex/Tceal* sufren episodios recurrentes de conversión génica. Con el doble objetivo de estudiar las relaciones entre los miembros de la familia y ver si los



**Figura 8.** Comparativa de la presencia de los genes *Bex/Tceal* en especies de mamíferos euterios representativas de los grandes linajes de este grupo. Cada triángulo representa un gen, y un triángulo vacío, un pseudogén. Los pseudogenes transparentados de la fila correspondiente al armadillo se detectaron en el genoma de otro xenartro: el perezoso didáctilo de Hoffmann (*Choloepus hoffmanni*). El árbol filogenético de las especies de euterios fue construido de acuerdo a trabajos anteriores<sup>240</sup>.

eventos de conversión génica son generalizados, construimos un árbol filogenético con las proteínas *BEX/TCEAL* de especies representativas de los cuatro grandes clados de placentarios: euarcontoglires, laurasiaterios, afroterios y xenartros (fig. 9). Estos resultados nos permiten confirmar que los eventos de conversión génica entre las regiones codificantes de ciertos genes *Bex/Tceal* son comunes y recurrentes en todas las especies de placentarios analizadas. Por otro lado, se ha comprobado que los genes *Bex* evolucionaron bajo una fuerte selección positiva antes de la diversificación de los placentarios<sup>99</sup>. Para investigar la conservación de la secuencia de los distintos ortólogos y comprobar si, una vez diversificados los placentarios, los genes *Bex/Tceal* evolucionaron bajo selección positiva o purificadora, se calculó el porcentaje de identidad y el ratio dN/dS entre pares de ortólogos (fig. 10). Los genes que sufren conversión génica entre sus regiones codificantes (*Bex1-2* y *Tceal2-6*) fueron excluidos del análisis. En cuanto al resto, se detectaron altos índices de similitud y un ratio dN/dS < 1, salvo para *Bex4*, que mostró una marcada relajación de las constricciones evolutivas (dN/dS > o cercano a 1). Resulta interesante comprobar que *BEX4* es el único gen *Bex/Tceal* al que se ha relacionado con la acetilación y estabilidad microtubular<sup>100</sup>. En conclusión, con la excepción de *Bex4*, los distintos ortólogos de los genes *Bex/Tceal* muestran un alto grado de conservación en sus secuencias codificantes y evolucionan bajo selección purificadora desde la diversificación de los mamíferos placentarios.





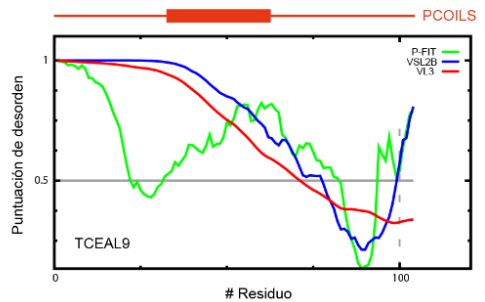
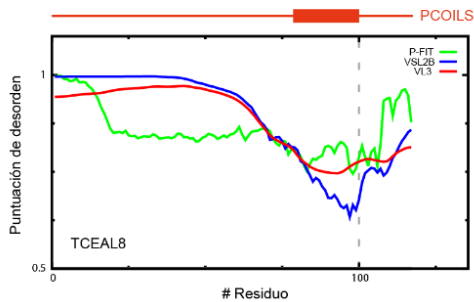
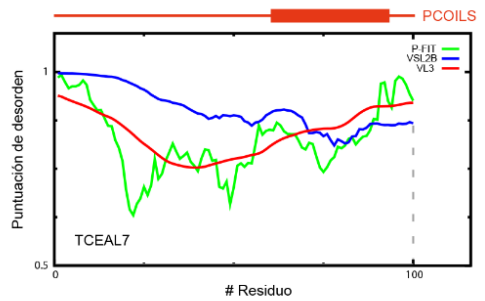
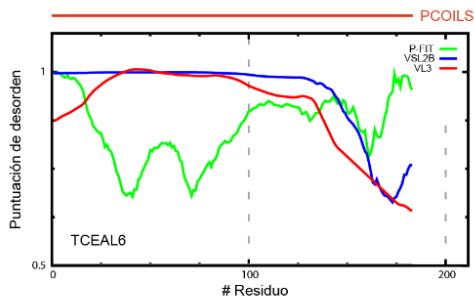
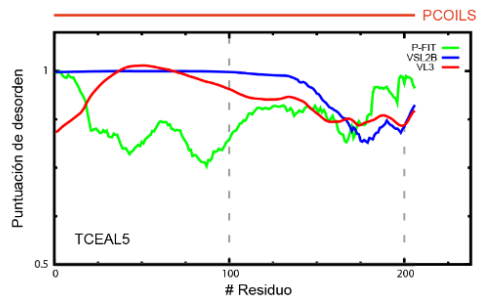
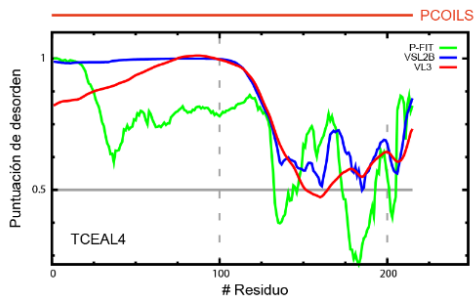
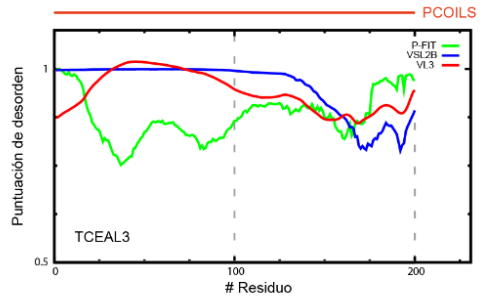
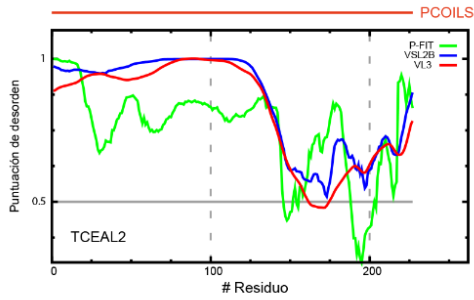
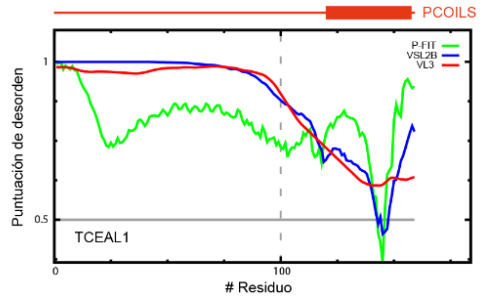
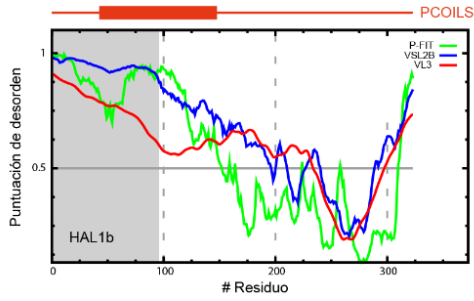
**Figura 10.** Conservación de secuencia y selección natural en los genes *Bex/Tceal*. Los genes *Tceal2-6* y *Bex1-2* se excluyeron del análisis. Se presentan en tonos rojizos los valores de % identidad al comparar las secuencias aminoacídicas entre pares de ortólogos, y en tonos azules el ratio Nei-Gojobori de mutaciones no sinónimas vs sinónimas (dN/dS). Bta, vaca; Cfa, perro; Dno, armadillo de nueve bandas; Eca, caballo; Hsa, humano; Laf, elefante africano; Mmu, ratón; Tma, manatí del Caribe.

Las proteínas BEX/TCEAL derivan del extremo amino-terminal de la proteína ORF1p de HAL1b (fig. 3). Es en esta región de ORF1p donde se encuentra el *coiled-coil*, así como la mayoría de las regiones estructuralmente desordenadas<sup>101,102</sup>. Curiosamente, las proteínas BEX han sido catalogadas como intrínsecamente desordenadas (*intrinsically disordered proteins*, IDPs), con regiones *coiled-coil* en su extremo carboxi-terminal<sup>103,104</sup>. Para analizar si las proteínas TCEAL también tienen una estructura desordenada y *coiled-coils*, analizamos bioinformáticamente sus secuencias con los programas PONDR-FIT<sup>105</sup>, VSL2B<sup>106</sup>, VL3<sup>107</sup> y PCOILS<sup>108</sup>. La mayor parte de la secuencia de todas las proteínas TCEAL mostró una estructura desordenada, y se detectaron dominios *coiled-coil* en los extremos carboxi-terminales de TCEAL1 y TCEAL7-9 (fig. 11). Dadas las similitudes estructurales entre la región amino-terminal del ORF1p y las proteínas BEX/TCEAL, parece probable que estas últimas hayan heredado y conservado sus características estructurales de la proteína transposónica ancestral.

### **Análisis de la exaptación (III) – Coopción de regiones reguladoras**

En la introducción se ha remarcado que una parte indispensable del proceso de domesticación molecular de los transposones es la adquisición de regiones reguladoras. La familia *Bex/Tceal*, junto con los genes *ArmcX* (*armadillo repeat containing, X-linked*), forman el clúster BGW<sup>96</sup>. Aunque no se detectó similitud entre ellos, los genes de estas familias se aunaron porque tienen en común la estructura génica (un ORF concentrado en el último de tres exones) y un motivo en su región promotora llamado motivo BGW. Este motivo consta de unos 60 pb, y contiene una *E-box* (*E* de *enhancer*, potenciador) en su centro que parece clave para promover la transcripción<sup>109,110</sup>.





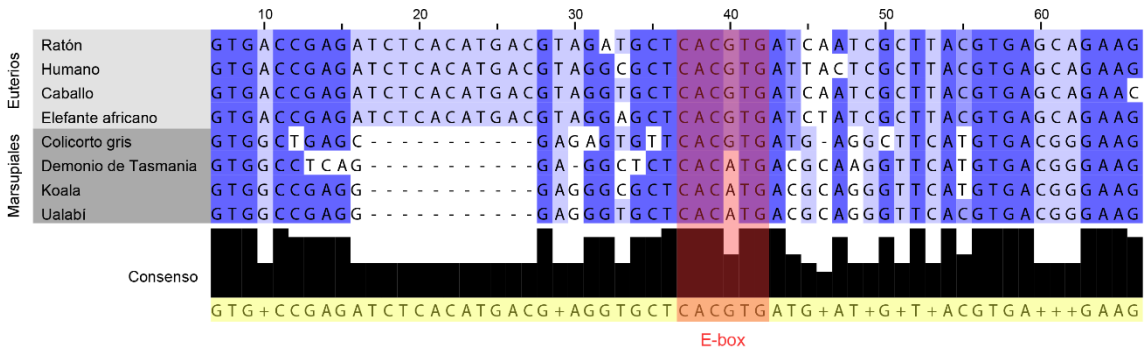
[Viene de la página anterior]

**Figura 11.** Predicción bioinformática de regiones desordenadas (PONDR-FIT, VSL2B y VL3) y *coiled-coils* (PCOILS). En el cuadro correspondiente al ORF1p de HAL1b, se sombrea en gris la región que originó el gen *Bex/Tceal* ancestral.

Esta discrepancia entre la homología de los *ArmcX* y los *Bex/Tceal* a nivel de promotor y la aparente ausencia de relación a nivel de región codificante, se intentó resolver en el artículo que describió por primera vez el clúster. Winter y Ponting propusieron que, antes de la diversificación de los mamíferos euterios, los *Bex/Tceal* se habrían originado a partir de un gen *ArmcX* ancestral. La secuencia codificante del proto-*Bex/Tceal* divergió rápidamente y el gen se duplicó hasta dar lugar a ambas subfamilias, aunque se conservó en los parálogos resultantes la región del promotor correspondiente al motivo BGW. Ahora que conocemos que los *ArmcX* derivan de la retrocopia de un gen autosómico llamado *Armc10*<sup>111</sup>, y que los *Bex/Tceal* derivan de la domesticación de elementos L1, se hace necesaria una nueva teoría que explique la formación del clúster. Complicando aún más el panorama evolutivo, el promotor de otro retrogén específico de euterios, situado a 5' del clúster y llamado *Hnrnp2*, alberga también este motivo<sup>96</sup>. Surge, por lo tanto, una pregunta: ¿Cómo genes con tres historias evolutivas aparentemente independientes llegaron a tener la misma secuencia en su promotor?

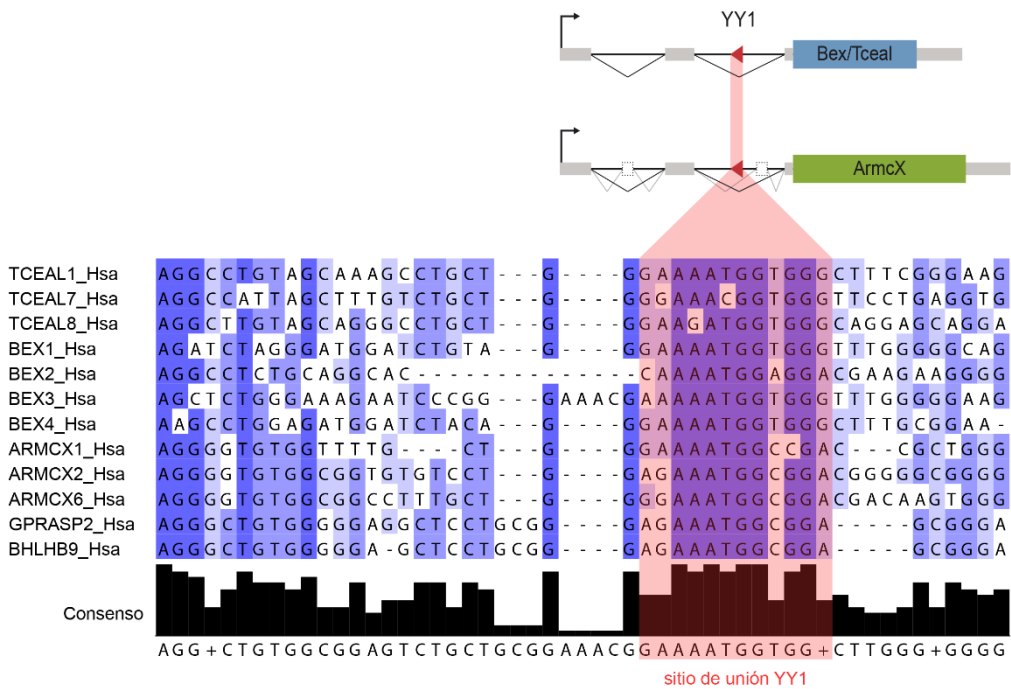
El promotor de *HNRNPH2* es bidireccional, compartido con el gen de la galactosidasa alfa (*GLA*)<sup>112</sup>. *GLA* ya estaba presente en el cromosoma X del terio ancestral, puesto que se encuentra en posiciones ortólogas tanto en marsupiales como en placentarios. Teniendo en cuenta esto, decidimos buscar secuencias cercanas al promotor de *Gla* que pudieran haber sido cooptadas en el ancestro de los placentarios por el retrogén de *Hnrnp2*. En los genomas marsupiales disponibles, encontramos una secuencia parecida al motivo BGW, con una *E-box* conservada, justo *upstream* del promotor de *Gla* (fig. 12). De este modo, pudimos trazar el origen del motivo BGW a una secuencia presente

por lo menos en el ancestro terio. Posteriormente, vía múltiples duplicaciones y coopciones en el ancestro de los placentarios, este motivo pudo haber desempeñado un papel clave en la regulación y la transcripción del retrogén *Hnrnp2* y las familias *ArmcX* y *Bex/Tceal*.



**Figura 12.** Alinamiento nucleotídico del motivo BGW de *Hnrnp2* y su región ortóloga en especies marsupiales. Sombreado en rojo, la *E-box* conservada, y abajo, en amarillo, el motivo BGW.

Además de los ORFs, los elementos L1 contienen 5'UTRs y 3'UTRs. En la 5'UTR, inmediatamente después del inicio de transcripción, un sitio de unión en la cadena complementaria para el represor transcripcional Yin Yang 1 (YY1) asegura que esta se lleve a cabo a partir del nucleótido adecuado (Athankar et al., 2004). Sorprendentemente, en la cadena complementaria de la región intrónica comprendida entre el segundo y el tercer exón de diversos genes *Bex/Tceal* y *ArmcX* se detectó un sitio de unión conservado para YY1 (fig. 13). Es más, una consulta de los datos de CHIP-Seq del consorcio ENCODE mostró que estas secuencias posiblemente sean funcionales, puesto que unen YY1 en al menos una de las líneas celulares humanas testadas.



**Figura 13.** Alineamiento de la región intrónica de genes *Bex/Tceal* y *ArmcX* de humano circundante al sitio de unión para el factor de transcripción YY1. Todos los sitios de unión mostrados unen YY1 en al menos una línea celular, según los experimentos de ChIP-Seq del consorcio ENCODE. Arriba, representación esquemática de la posición y orientación del sitio de unión respecto al ORF (en azul y verde) y los exones no codificantes (en gris los constitutivos, en blanco los alternativos).

## Análisis de la exaptación (IV) – Impacto del motivo BGW

Además de determinar el origen evolutivo del motivo BGW, quedaba por comprender por qué este fue cooptado originalmente y se ha mantenido en un rango tan amplio de genes. Con esta pregunta en mente, procedimos a testar su potencial intrínseco para controlar la expresión génica. Para investigar este punto, amplificamos mediante PCR algunas regiones de especies marsupiales y placentarias, clonándolas en el vector de expresión hsp68-lacZ, y creamos embriones de ratón transgénicos mediante inyección pronuclear (véase *Métodos*). A continuación, se examinó la actividad del gen reportero (*lacZ*) en embriones E13.5 tiñéndolos con X-gal, y los patrones fueron anotados de

acuerdo al siguiente criterio: partiendo de un mínimo de cinco muestras, se consideró que el elemento testado promueve la expresión en un determinado tejido cuando observamos una tinción consistente en la mitad o más de los embriones<sup>113</sup>. Es necesario hacer notar que el hecho que una secuencia no produzca un patrón en un determinado estadio embrionario no significa que no pueda producirlo en otros, o en distintas condiciones fisiológicas.

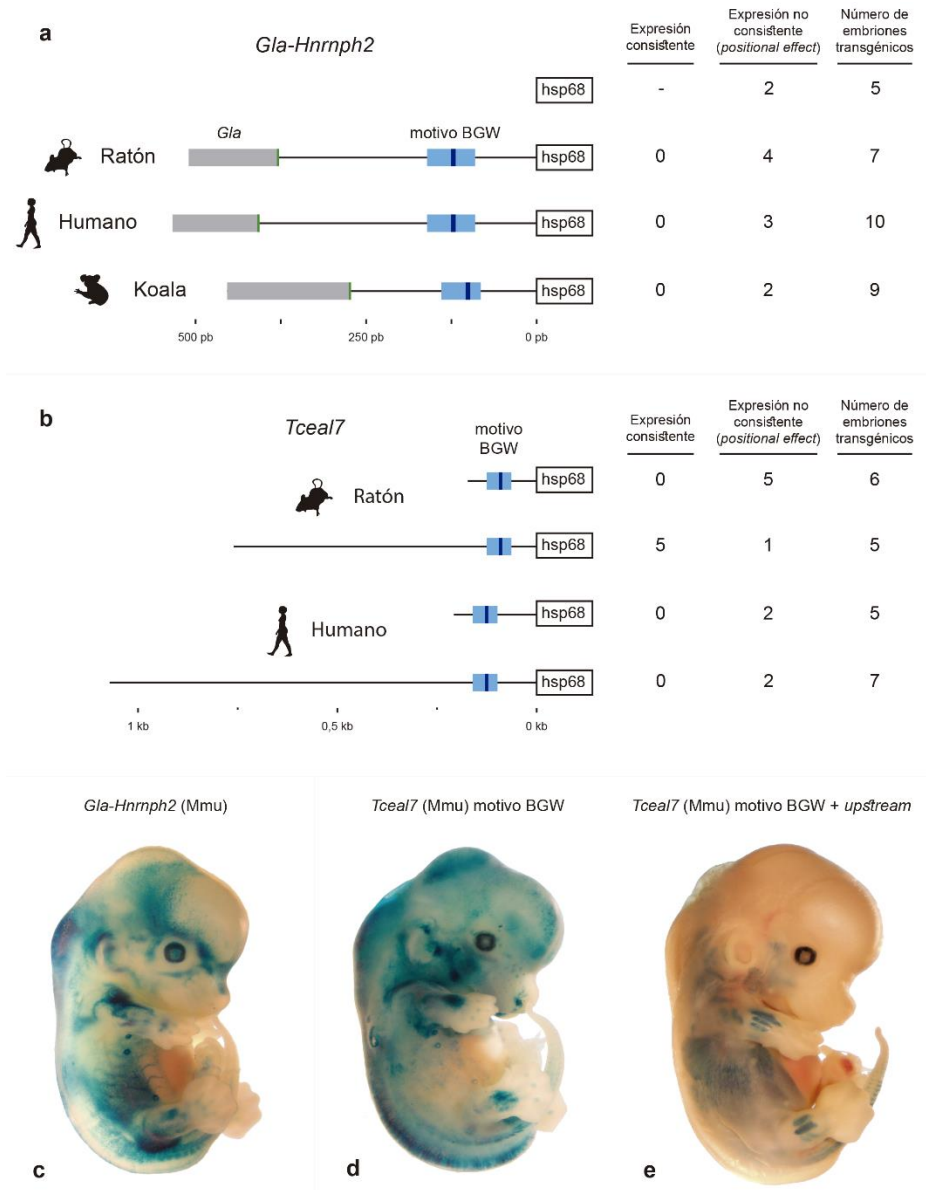
Primero, para inferir el potencial regulador del motivo BGW ancestral, decidimos estudiar la región *upstream* de *GLA* en especies marsupiales (pues contiene una secuencia parecida al motivo BGW), y compararla con la de euterios modernos. Para ello, clonamos la región *upstream* del gen *GLA* de koala correspondientes al promotor de *GLA* y el motivo proto-BGW, y las secuencias homólogas murinas y humanas, es decir, el promotor bidireccional *GLA-Hnrnp2*. Ninguno de los tres elementos produjo un patrón consistente (fig. 14). Además, al conocer el patrón de expresión de *Hnrnp2* en E13.5 (fig. S1), podemos afirmar que el promotor bidireccional murino no lo recapitula. Por lo tanto, estos resultados sugieren que, en el estadio estudiado, el motivo proto-BGW de koala no es capaz por sí solo de promover la transcripción génica, y que otras regiones reguladoras condicionan la expresión tisular del gen *Hnrnp2*.

Segundo, examinamos la regulación del gen *Tceal7* con el objetivo de diseccionar la influencia del motivo BGW. Shi y Garry han mostrado que, durante el desarrollo, la expresión del gen *Tceal7* de ratón se restringe al linaje miogénico, y que una región promotora de ~700 pb es responsable de producir este patrón<sup>109</sup>. Esta región incluye el motivo BGW, cuya *E-box* es necesaria para la actividad del promotor. Además, mientras que la expresión de *Tceal7* en el ratón adulto también es muscular, en humano es predominantemente neural<sup>114</sup>. Para entender estas variaciones interespecíficas en la expresión tisular de *Tceal7*, tratamos de ver si la explicación se encontraba en las pequeñas alteraciones en el motivo BGW entre linajes, o si estarían implicadas otras regiones reguladoras *upstream*. Así, testamos los motivos BGW del *Tceal7* de

ratón y humano (añadiendo regiones circundantes con una fuerte conservación), la mencionada región promotora de 700 pb de ratón, y su contraparte homóloga humana. Tan solo el elemento murino de 700 pb produjo una expresión consistente, reproduciendo el patrón específico del linaje muscular de *Tceal7* durante el desarrollo (fig. 14 y 15).

Es necesario aclarar que, una vez insertado el vector en el genoma, el promotor mínimo que dirige la expresión (aquel que consta de las secuencias estrictamente necesarias para iniciar la transcripción, pero carece de la potencia suficiente para promoverla por sí solo) puede ser influenciado por regiones reguladoras presentes en el entorno de inserción. Los patrones inespecíficos resultantes de este fenómeno son considerados *positional effect*. Por lo tanto, aunque la región bajo escrutinio carezca de potencial, es esperable que pueda llegar a observarse una expresión no consistente del gen reportero en alguno de los ejemplares. Sin embargo, al testar el motivo BGW de *Tceal7* de ratón encontramos una expresión dispar en 5/6 embriones, principalmente en tejidos neurales y esqueléticos (fig. 14). Estos resultados son interesantes, puesto que la proporción obtenida de embriones con transcripción espuria —debido a *positional effect* del promotor mínimo— es menor tanto en nuestro control experimental como en resultados previamente publicados<sup>115</sup>. Es tentador proponer que, más que conferir especificidad de tejido, el motivo BGW podría ser una secuencia altamente sensible a determinadas señales reguladoras circundantes. Las secuencias BGW humanas de *TCEAL7* no sufrirían tanto *positional effect* por ser menos sensibles a las señales reguladoras de ratón, debido a un proceso de refinamiento específico de cada linaje. Aun así, se hace necesario aumentar el número de embriones y realizar cortes de los mismos,

tanto para incrementar la significación estadística como para facilitar la interpretación de los resultados.



**Figura 14.** Potencial de expresión de secuencias con el motivo BGW durante el desarrollo de ratón. En (a) y (b) se representan esquemáticamente las secuencias testadas y los resultados del análisis. En azul claro, el motivo BGW, y en azul oscuro, la *E-box*. (c), (d) y (e) muestran embriones E13.5 transgénicos teñidos con X-Gal. Las imágenes (c) y (d) revelan los patrones de expresión no consistentes (*positional effect*) de los constructos indicados. (e) muestra la expresión consistente en el linaje muscular del motivo BGW de *Tceal7* de ratón, incluyendo las regiones *upstream*.

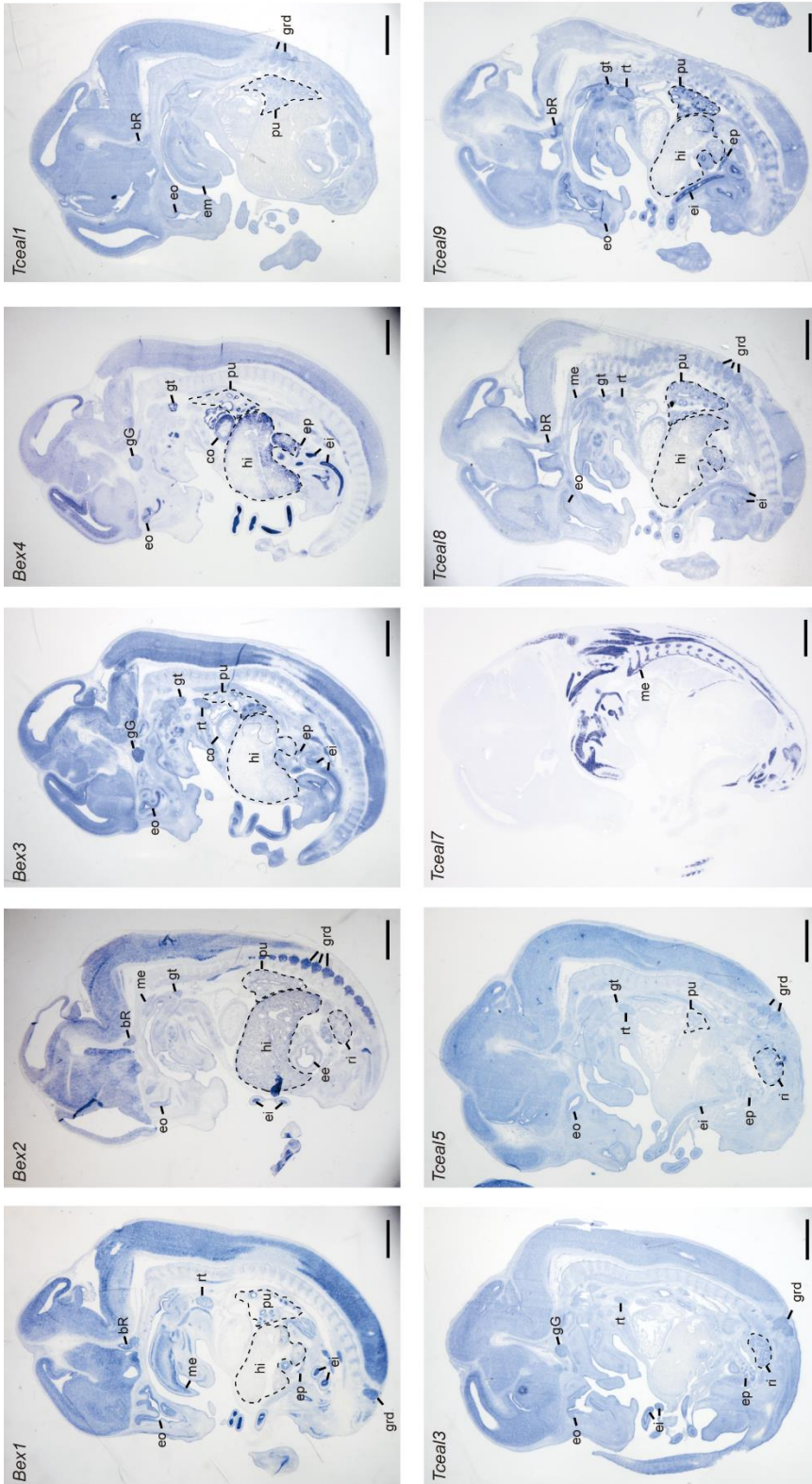
## **Expresión de la familia *Bex/Tceal***

Aunque algunos de los genes *Bex/Tceal* parecen desempeñar un papel en procesos como la miogénesis o la neurogénesis<sup>109,116-119</sup>, y se expresan abundantemente durante el desarrollo embrionario<sup>109,120-123</sup>, para la mayoría de ellos hay escasa o ninguna información sobre la expresión tisular, tanto en adulto como durante el desarrollo<sup>109,120,121,124-128</sup>.

Primero, decidimos caracterizar la expresión en embriones E13.5 de ratón mediante hibridación *in situ* (ISH) de los genes *Bex/Tceal* (*Tceal6* no pudo ser analizado, debido a la extensiva conversión génica con *Tceal3* y *5*). A excepción de *Tceal7*, que reprodujo la expresión específica de músculo esquelético previamente reportada por Shi y Garry<sup>109</sup> (fig. 15), los demás genes *Bex/Tceal* mostraron unos dominios de expresión muy extensos. Para todos ellos, se detectó marcaje en el epitelio olfatorio, renal y respiratorio, así como en otros tejidos y órganos glandulares, como la bolsa de Rathke, el esbozo pancreático y la glándula tiroides. Sin embargo, también se encontraron diferencias en las intensidades de expresión, así como ciertas particularidades (tabla S1 y fig. 15 y S2). Ejemplos de ello serían la señal de *Bex4*, que es especialmente intensa en el corazón, y los genes *Bex1-3*, que muestran un destacado marcaje en tejidos neuronales.

Con estos resultados, podemos concluir que, con remarcables especificidades, los genes *Bex/Tceal* muestran unos patrones de expresión extensos y parcialmente solapantes.





**Figura 15.** Patrones de expresión de los genes *Bex/Tceal* en ratones E13.5, mediante hibridación *in situ* sobre cortes sagitales. Con la excepción de *Tceal7*, que se expresa específicamente en el músculo esquelético, los demás genes *Bex/Tceal* muestran extensos patrones de expresión, aunque con diferentes intensidades (véase tabla S1). Se aprecia especialmente un marcaje en el SNC y en el SNP (en ganglios de la raíz dorsal y de Gasser), en el epitelio olfatorio, renal y pulmonar, además de en otros tejidos glandulares como la tiroides, el esbozo pancreático y la bolsa de Rathke (véanse imágenes complementarias en fig. S2). Algunos genes muestran patrones específicos prominentes, como la intensa expresión de los genes *Bex1-3* en tejidos neurales y de *Bex4* en el corazón. *br*, bolsa de Rathke; *co*, corazón; *ee*, epitelio estomacal; *ei*, epitelio intestinal; *eo*, epitelio olfatorio; *ep*, epitelio olfatorio; *ep*, epitelio olfatorio; *gG*, ganglio de Gasser; *grd*, ganglios de la raíz dorsal; *gt*, glándula tiroides; *hi*, hígado; *me*, músculo esquelético; *pu*, pulmón; *ri*, riñón; *rt*, rudimento tímico. Barra=1 mm.

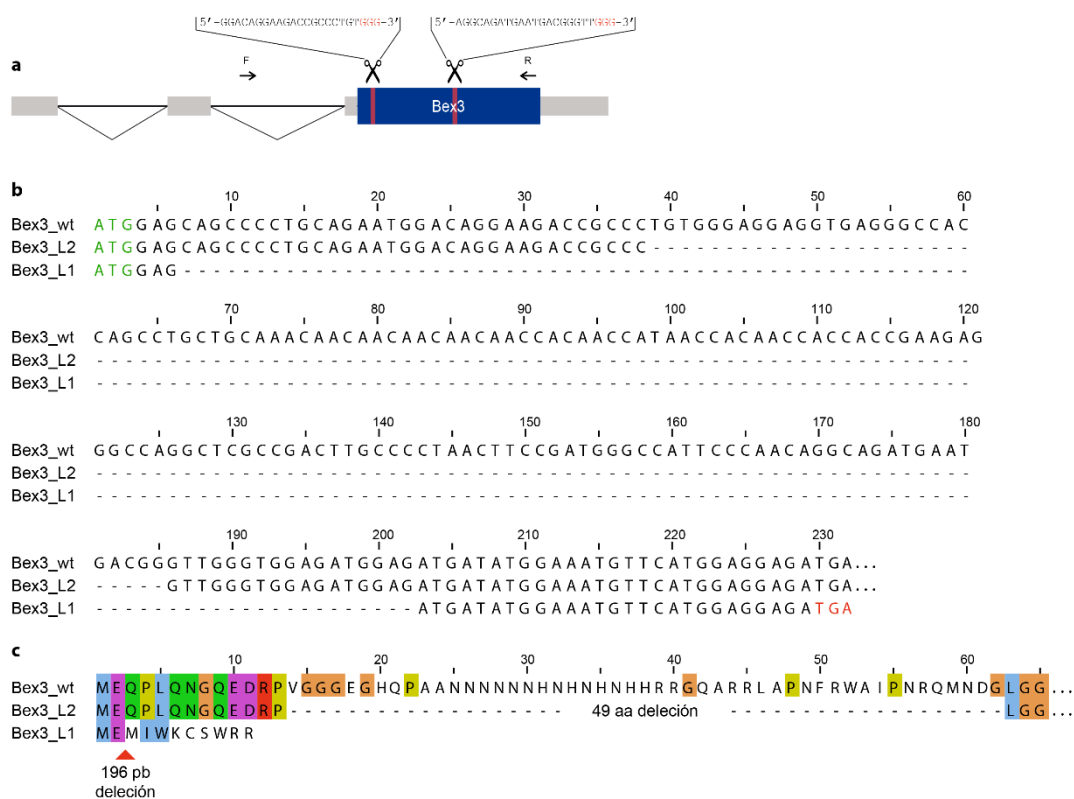
## Generación y caracterización de ratones *Bex3* mutantes

Los genes *Bex/Tceal* codifican para pequeñas proteínas con una expresión predominantemente neural. En cuanto a los genes *Tceal*, con la excepción de *Tceal7*, la única información disponible sobre los miembros de esta subfamilia es su asociación con algunos tipos de cáncer<sup>128-130</sup>. La expresión muscular de *Tceal7* durante el desarrollo de ratón está regulada por factores miogénicos como MyoD<sup>109</sup> y, mediante la inhibición de la vía NF- $\kappa$ B<sup>131</sup>, su función parece ser antiproliferativa y prodiferenciación.

Por el contrario, los genes *Bex* están mejor estudiados. Las proteínas de esta subfamilia muestran una gran promiscuidad en sus interacciones proteicas y participan en multitud de vías de señalización, regulando el ciclo celular, la apoptosis y la diferenciación. Han sido implicadas en regeneración<sup>116,117</sup> y en cáncer, principalmente como genes supresores de tumores, aunque también como oncogenes<sup>132</sup>. Pueden asociarse a mitocondrias<sup>120</sup>, a proteínas del citoesqueleto<sup>100</sup> o a receptores de membrana<sup>121</sup>, y localizarse tanto en el citoplasma como en el núcleo<sup>127</sup>. Por ejemplo, BEX3 interactúa con los receptores de neurotrofinas p75<sup>NTR</sup> y TrkA, regulando así la supervivencia neuronal<sup>118,133,134</sup>. Esta promiscuidad de las proteínas BEX, sumada al hecho que adoptan una estructura intrínsecamente desordenada, ha llevado a que se las proponga como candidatas a proteínas «foco» o *hub*<sup>104</sup>. Las proteínas *hub* son los puntos de convergencia en un interactoma proteico, y resultan esenciales para su funcionalidad y estabilidad. Sin embargo, tanto de los genes *Bex* como de los *Tceal* se han realizado pocos estudios que investiguen su función durante el desarrollo e interpreten su papel desde un punto de vista evolutivo.

Para avanzar en la caracterización funcional de esta familia génica, escogimos el gen *Bex3* como candidato más interesante, basándonos en la literatura disponible. A continuación, nos propusimos generar líneas de ratones mutantes para este gen mediante el sistema CRISPR/Cas9. Primero, se

diseñaron ARN guía (ARNg) dirigidos al tercer exón, concretamente al principio del ORF (véase *Métodos* y fig. 16). Así, se generaron las líneas L1 y L2, con deleciones de 196 pb y 147 pb, respectivamente. Tanto la generación como la caracterización del fenotipo de estos ratones se llevaron a cabo en colaboración con el laboratorio de Jaime Carvajal (Centro Andaluz de Biología del Desarrollo, CABD). La deleción asociada a la línea L1 rompía la pauta de lectura y generaba una proteína truncada. En cambio, la de L2 mantenía la pauta de lectura, aunque eliminando 49 aminoácidos (de un total de 114 en la isoforma corta y 124 en la larga) del centro de la proteína (fig. 16).

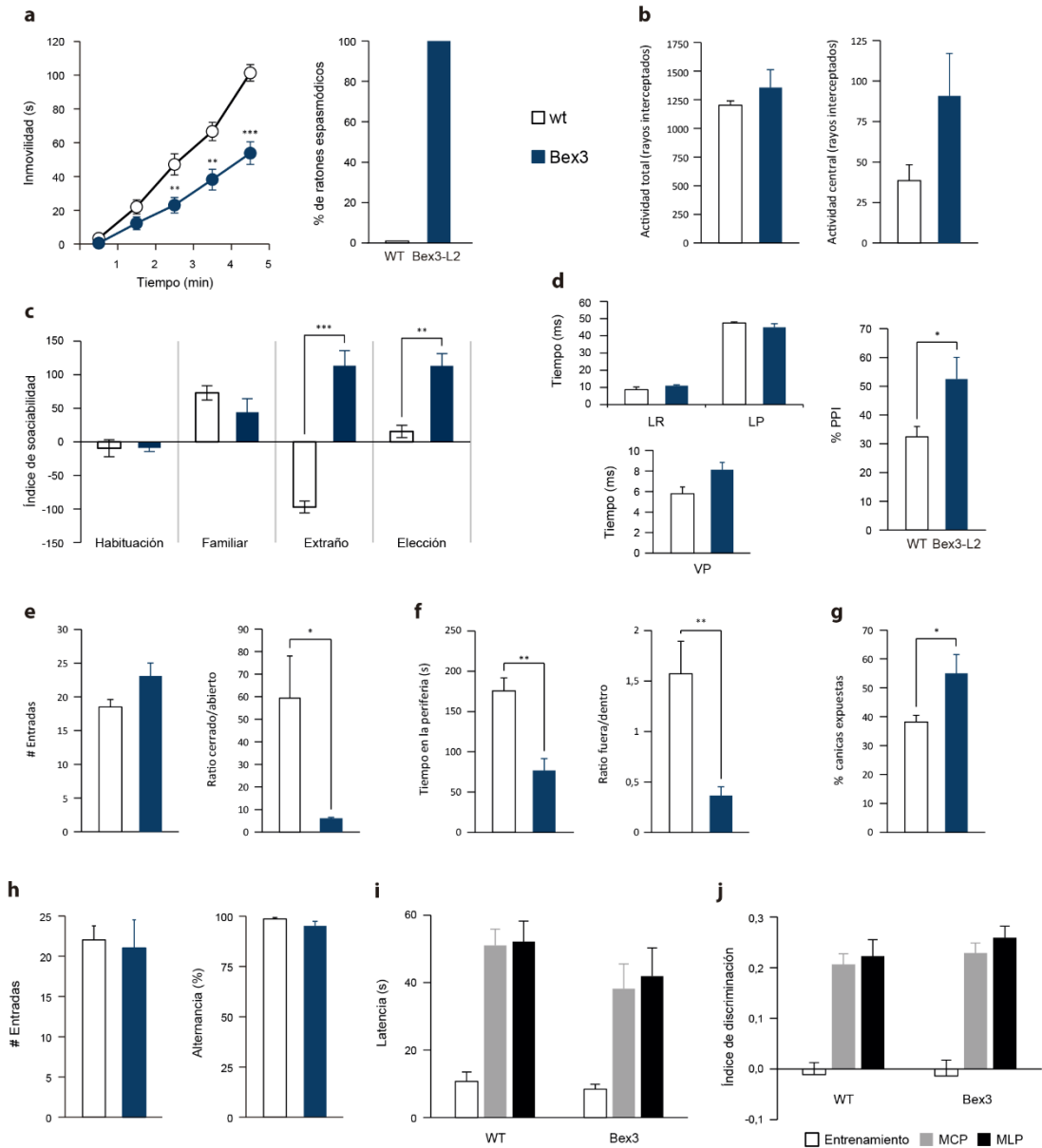


**Figura 16.** (a) Representación esquemática del locus de *Bex3*. Se indican las secuencias diana de los ARNg, con el motivo adyacente de protospaciador (*protospacer-adjacent motif*, PAM) en rojo, y los *primers* utilizados para el genotipado. (b) Secuencias del ORF de *Bex3* wild type y de las líneas mutantes L1 y L2, mostrando las deleciones de 196 y 147 pares de bases, respectivamente. (c) Secuencias aminoacídicas de los ORFs anteriores. La deleción de la línea Bex3-L1 produce un cambio de pauta de lectura que genera un codón stop prematuro. La de la línea Bex3-L2 está *in-frame*, produciendo una deleción de 49 aminoácidos.

Para determinar el impacto de estas mutaciones, decidimos comprobar la presencia de la proteína en extracciones de cerebro de ratones adultos L2 y *wild type* (WT) mediante *western blot* (WB), escogiendo anticuerpos cuyo epítipo se encontrara *downstream* de la zona delecionada en la línea L2. Los dos anticuerpos comerciales testados mostraron múltiples bandas inespecíficas tanto en muestras mutantes como WT, y ninguna del peso esperado para Bex3 (~23 kDa). Por lo tanto, la presencia de proteína Bex3 en los ratones de la línea L2 permanece por descartar (véase *Discusión*).

Como se ha visto, en el estadio E13.5 *Bex3* se expresa intensamente en el SNC (fig. 15). Además, estudios anteriores apuntan a un posible rol de *Bex3* durante el desarrollo del SNC: interacciona con los receptores de neurotrofinas p75<sup>NTR</sup> y TrkA, promoviendo la supervivencia y la diferenciación neuronal, así como el crecimiento neurítico<sup>118,133</sup>. Para caracterizar el fenotipo etológico de esta línea, sometimos a los ratones Bex3-L2 a una serie de test de comportamiento (en colaboración con Ángel Carrión, de la Universidad Pablo de Olavide, UPO). Al suspenderlos por la cola, el 100% de los mutantes mostraron movimientos espasmódicos en las patas traseras (fig. 17a). Esta disminución de la parálisis puede ser atribuida a un escaso control motor. En el test de campo abierto (fig. 17b), que mide la actividad motora, no se encontraron diferencias significativas. Sin embargo, tanto en la caja de preferencia social (c) como en el reflejo de sobresalto (fig. 17d) se detectaron alteraciones significativas en los ratones Bex3-L2. El tiempo que dedicaron los ratones mutantes a investigar a un ratón desconocido fue menor ( $p < 0.001$ ), y se detectó una potenciación ( $p < 0,05$ ) de la inhibición prepulso (abreviada PPI, de *prepulse inhibition*: se basa en la atenuación de la respuesta a un pulso una vez se ha sometido al animal a un prepulso). Ambos resultados indicarían un comportamiento autístico en los ratones Bex3-L2. Otros test capaces de medir la ansiedad arrojaron resultados aparentemente contradictorios. En el laberinto elevado en cruz (fig. 17e), los ratones mutantes tendieron a explorar más los brazos abiertos ( $p < 0,05$ ), señal de una menor ansiedad. Sin embargo, en el test

de emergencia o *emergency test* (fig. 17f), que consiste en situar una caja oscura en el centro de un campo abierto, los Bex3-L2 se aventuraron significativamente menos al exterior y pasaron menos tiempo en la periferia que sus compañeros de camada WT (en ambos casos,  $p < 0.01$ ). Por otro lado, al depositar en la caja una serie de canicas, se midió el porcentaje de objetos enterrados por el animal (fig. 17g). Los ratones mutantes tendieron a dejar más canicas expuestas ( $p < 0,05$ ), lo que sería indicativo de ansiedad. No se apreciaron diferencias significativas en el laberinto en Y (fig. 17h), que mide la memoria espacial. Por último, en los test de evitación pasiva (fig. 17i) y reconocimiento de objetos (fig. 17j) no se detectaron alteraciones en la memoria a corto o largo plazo de los Bex3-L2. Por lo tanto, los ratones Bex3-L2 muestran diversas alteraciones comportamentales, entre ellas algunas asociadas al autismo.



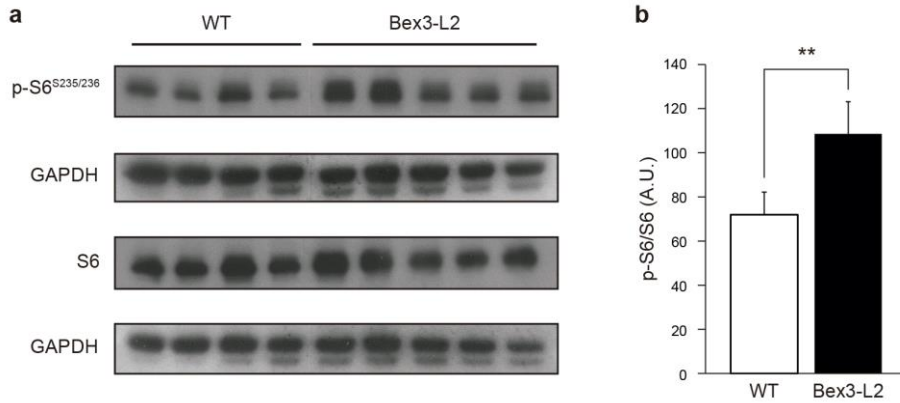
**Figura 17.** Alteraciones en el comportamiento de ratones Bex3-L2 homocigotos. (a) Tiempo en inmovilidad al ser suspendidos por la cola. (b) Actividad en campo abierto, medida por el número de intercepciones de los rayos infrarrojos. (c) Caja de preferencia social, donde se calcula el índice de sociabilidad sustrayendo al tiempo de interacción con la caja del ratón familiar el tiempo de interacción con la del ratón extraño. (d) Test de reflejo de sobresalto. La inhibición prepulso (*prepulse inhibition*, PPI) se obtiene a partir de los VP del prepulso y del pulso. LP, latencia de aparición del pico de la respuesta; LR, latencia de inicio de la respuesta; VP, valor del pico máximo de la respuesta. (e) Laberinto elevado en cruz. Se calculan el número de cambios entre brazos cerrados y abiertos, y el ratio del tiempo transcurrido en cada tipo de brazo.

[Viene de la página anterior]

(f) Test de emergencia. Se consideran el tiempo en la periferia y fuera o dentro de la caja central. (g) Porcentaje de canicas expuestas respecto a las totales. (h) Laberinto en Y. El porcentaje de alternancia tiene en cuenta el número de entradas y el brazo anteriormente visitado. (i) Test de evitación pasiva. Tiempo en pasar a la caja oscura. (j) Test de reconocimiento de objetos. El índice de discriminación se calcula en función del tiempo transcurrido investigando el objeto nuevo o el conocido. MCP, memoria a corto plazo; MLP, memoria a largo plazo. Todos los valores representan medias  $\pm$  desviación estándar. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  señalan diferencias significativas entre grupos.

Desregulaciones en la vía mTOR han sido frecuentemente asociadas a la aparición de fenotipos relacionados con el autismo<sup>135</sup>. A su vez, se conoce que en humano BEX3 interacciona con la hamartina, *Tsc1*<sup>136</sup>, un inhibidor de esta vía, y que la proteína Bex2 murina es una efectora de la misma, mediando sus efectos tumorigénicos<sup>137</sup>.

Por lo tanto, en colaboración con el laboratorio de Eduardo Soriano (Parc Científic de Barcelona, PCB), decidimos analizar si el fenotipo comportamental observado en los ratones Bex3-L2 podría estar relacionado con alteraciones en esta vía. Sorprendentemente, al medir un indicador de la activación de la vía mTOR, como es el ratio p-S6/S6, detectamos una hiperactivación significativa ( $p < 0.01$ ) en el encéfalo de ratones homocigotos Bex3-L2 adultos (fig. 18). Con estos resultados, podemos aventurar una implicación de *Bex3* en la regulación de la vía mTOR, cuya hiperactivación sería responsable del comportamiento autístico observado en los ratones Bex3-L2.

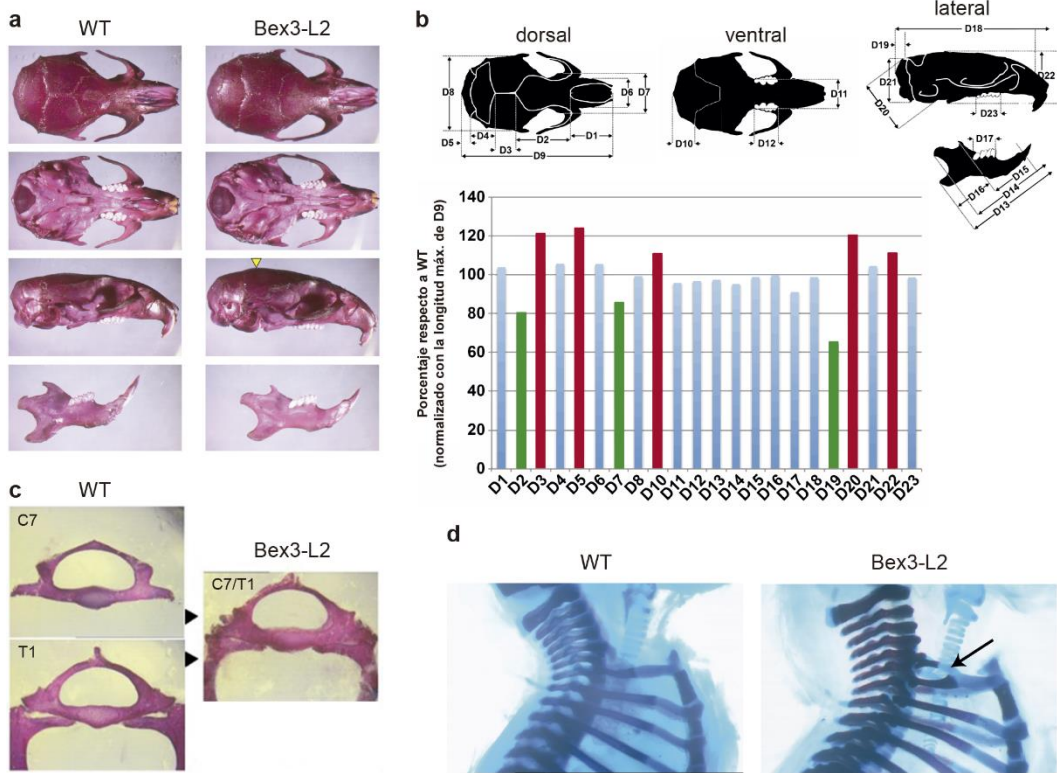


**Figura 18.** Hiperactivación de la vía mTOR en el encéfalo de ratones Bex3-L2. (a) Análisis por *western blot* de las proteínas p-S6 y S6 en el encéfalo de ratones WT y mutantes homocigotos Bex3-L2. GAPDH se utilizó como control interno. (b) Ratio p-S6/S6 relativo a GAPDH de los westerns anteriores. Los ratones Bex3-L2 muestran un aumento significativo del ratio, indicativo de una hiperactivación de la vía mTOR. Todos los valores representan medias  $\pm$  desviación estándar. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  señalan diferencias significativas entre grupos.

Finalmente, un examen anatómico de los ratones homocigotos Bex3-L2 arrojó alteraciones morfológicas evidentes. Al ser comparados con sus compañeros de camada WT, los ratones mutantes mostraban malformaciones craneales evidentes. Con el objetivo de analizar el conjunto de las estructuras óseas y cartilaginosas, se tiñeron ratones de esta línea con azul alcian y rojo de alizarina. Las mediciones a las que fueron sometidos los cráneos mostraron claras alteraciones morfológicas, particularmente llamativas en los huesos occipital y parietal (fig. 19a, b).

Además, se detectó —tanto en adultos como en embriones 16 días post coito (dpc)— una transformación homeótica que resulta en la aparición de una costilla asociada a la última vértebra cervical (C7), que se fusiona con la costilla de la primera vértebra torácica (T1) (fig. 19c, d). Por lo tanto, los ratones Bex3-L2 sufren una serie de malformaciones esqueléticas.





**Figura 19.** Malformaciones esqueléticas de ratones Bex3-L2. (a) Vistas del cráneo (dorsal, ventral y lateral) y la mandíbula de ratones *wild type* (WT) y Bex3-L2, teñidos con azul de alcian y rojo de alizarina. El hueso parietal de los Bex3-L2 se halla abombado (triángulo amarillo). (b) Comparación de las dimensiones del cráneo de ratones Bex3-L2 versus ratones WT. Las barras en rojo y verde corresponden respectivamente a mediciones significativamente mayores y menores en los ratones Bex3-L2. (c) y (d) Tinción de vértebras y costillas con azul alcian y rojo de alizarina de ratones adultos y embriones 16 dpc. Se aprecia la transformación homeótica resultante en una fusión de la costilla de la última vértebra cervical (C7) con la de la primera torácica (T1) en ratones Bex3-L2.

## Resumen de resultados

- Elaboración de una lista con 67 candidatos a transposones domesticados en el genoma humano, entre ellos, el gen *Tceal7*.
- El gen *Tceal7* se originó a partir de la domesticación molecular de los retrotransposones HAL1b y L1ME.
- La familia *Bex/Tceal*, de la que *Tceal7* forma parte, se originó y expandió antes de la radiación de los placentarios. Posteriormente, algunos linajes sufrieron pérdidas y ganancias génicas.
- Los eventos de conversión génica entre las regiones codificantes de los genes *Bex/Tceal* son comunes en todos los linajes analizados.
- La secuencia aminoacídica de los genes *Bex/Tceal*, a excepción de *Bex4*, muestran un alto grado de conservación, y han evolucionado bajo selección purificadora desde la diversificación de los placentarios.
- Las proteínas TCEAL presentan una estructura mayormente desordenada y, en algunos casos, con regiones *coiled-coil*.
- El motivo BGW, presente en los promotores de las familias *Bex/Tceal* y *ArmcX*, y en el gen *Hnrnp2*, se originó a partir de una secuencia *upstream* del promotor del gen *Gla* presente en el ancestro terio.
- Identificación de sitios de unión para YY1 conservados y funcionales en la región intrónica de los genes *Bex/Tceal* y *ArmcX*, posiblemente derivados del transposón ancestral HAL1b.
- El ensayo del transgén reportero mostró que los motivos BGW, en el estadio analizado, no son capaces de recapitular la expresión tisular de los genes cuya transcripción promueven.

- A excepción de *Tceal7*, que se expresa intensamente en el linaje muscular, los genes *Bex/Tceal* muestran patrones de expresión extensos y parcialmente solapantes en el estadio E13.5 del desarrollo murino.
- Generación de dos líneas de ratones mutantes para *Bex3* mediante el sistema CRISPR/Cas9.
- Los ratones homocigotos mutantes de la línea Bex3-L2 muestran alteraciones del comportamiento, algunas asociadas a trastornos del espectro autista.
- Un primer análisis apunta a una hiperactivación de la vía mTOR en el encéfalo de los ratones homocigotos mutantes de la línea Bex3-L2.
- Los ratones homocigotos mutantes de la línea Bex3-L2 muestran malformaciones craneales y desarrollan una costilla asociada a la última vértebra cervical, que se fusiona con la de la primera torácica.

## **Discusión**



## Métodos para la detección de TDs

Trabajos anteriores en los que se estudiaba la contribución de los transposones al proteoma humano han utilizado estrategias similares a la nuestra<sup>58,138,139</sup>. Sin embargo, probablemente debido a una (por entonces) deficiente anotación tanto de los genes codificantes como de los transposones, se reportaron falsos eventos de TDs y no se detectaron otros, como *Tceal7*. Malas anotaciones aparte, uno de los principales problemas de esta metodología es que, al depender de las secuencias consenso transposónicas construidas por RepBase (de la cual derivan a su vez las anotaciones del RepeatMasker), tiene limitada la edad de los TDs que es capaz de detectar. RepBase Update<sup>92</sup>, en un intento de reconstruir el elemento originalmente activo, elabora la secuencia consenso de cada familia a partir de aquellas copias que compartan un mínimo de similitud. Por lo tanto, las copias de transposones que hayan divergido más allá de un determinado umbral no serán detectables. Por ilustrarlo metafóricamente, las copias de un transposón serían como las ruinas de una antigua ciudad: con el examen de cada edificación podemos hacernos una idea de la arquitectura original, cada vez más precaria a medida que el tiempo transcurre. Llegados a un determinado punto, las ruinas serán totalmente irreconocibles, y ningún patrón será extrapolable. De este modo, en RepeatMasker podemos encontrar algunas pocas familias que se expandieron en el ancestro de los mamíferos (~180 millones de años), pero ninguna cuya actividad pueda ser trazada al origen de los vertebrados (~500 millones de años). Igualmente, tampoco se detectarán eventos de domesticación más recientes pero que hayan sufrido un aumento de la tasa de mutaciones, debido por ejemplo a selección positiva.

Otros métodos, en cambio, como el desarrollado por Zdobnov y colegas para *Homo sapiens*<sup>140</sup>, o el más exhaustivo de Hoen y Bureau para *Arabidopsis thaliana*<sup>61</sup>, se basan en la conservación de dominios proteicos característicos de elementos transponibles. Siguiendo esta estrategia, se pueden detectar TDs como *Tigd5*, una transposasa de ADN domesticada en el ancestro de los

sarcopterigios, hace más de 400 millones de años. Sin embargo, este enfoque no deja de estar limitado por la presencia y conservación del dominio proteico transposónico.

Volviendo a *Tceal7*, este gen no conserva ningún dominio con esas características, pues su ORF corresponde a parte de la región *coiled-coil* de la proteína ORF1p (fig. 3). Esta región, encargada de la oligomerización de ORF1p, evoluciona muy rápidamente<sup>141,142</sup>, lo que hace inviable su caracterización como dominio. El ejemplo de los *Bex/Tceal* nos muestra que no solo son susceptibles de domesticación los dominios con una funcionalidad relativamente definida, sino también motivos estructurales más sencillos. Por lo tanto, es necesaria una combinación de ambos métodos (solapamiento de transposones y regiones codificantes, y búsqueda basada en dominios transposónicos) para optimizar la identificación de TDs. Finalmente, es necesario recordar que, para reducir el número de falsos positivos, ambas metodologías requieren la conservación del ORF al considerar un candidato a TD. Así, puede suceder que excluyan TDs para los que carecemos de un genoma suficientemente cercano que demuestre conservación, o los que sean una innovación específica del linaje estudiado.

### **Una nueva domesticación molecular en el origen de los euterios**

La domesticación molecular de cierto tipo de proteínas transposónicas, como las transposasas de los transposones ADN o las Gag y Env de los transposones LTR, es relativamente común<sup>72</sup>. En cambio, pese a constituir una parte sustancial del genoma de muchas especies (fig. 1), la domesticación de elementos sin LTRs parece ser mucho más infrecuente. En el hongo *Blumeria graminis* se han comunicado los casos de *Avrkl* y *Avra10*<sup>143</sup>, y el único conocido en metazoos es el del gen específico de euterios *LITD1*<sup>144</sup>. Por lo tanto, este trabajo caracteriza el segundo evento de domesticación molecular de un retrotransposón sin LTRs en metazoos, y el primero en originar una familia génica.

Por otro lado, los resultados presentados arrojan luz sobre la evolución de los mecanismos de transposición, especialmente relevantes debido a sus repercusiones sobre la integridad y plasticidad genómica. En este trabajo se ha prestado especial atención al ensamblaje del material genético que dio lugar al ancestro de los genes *Bex/Tceal*. Como se ha detallado, el proceso empezó con la transposición de un elemento HAL1b en el cromosoma X del ancestro de los placentarios. Sobre esta copia, se integró otra de la familia L1ME que, al hacerlo mediante la variante *twin priming* de la TPRT (véase fig. 6), quedó invertida e internamente deletada (fig. 7). El mecanismo de *twin priming* solo se ha observado en ratón, perro y primates catarrinos<sup>145-147</sup>. El análisis del origen de los *Bex/Tceal* nos permite sugerir que este mecanismo ya actuaba durante la transposición de los elementos L1 activos en el ancestro de los placentarios.

Como ha quedado demostrado en estudios anteriores, además de los genes *Bex/Tceal* otras familias derivadas de TDs se diversificaron en el ancestro de los placentarios<sup>63</sup>: es el caso de los antígenos paraneoplásicos Ma (*PNMA*) y los *SIRH*, sobre los que se habló en la introducción. Algunos de los miembros de estas familias (*Pnma* y *Sirh12*) ya estaban presentes en marsupiales, pero fue después de la separación del linaje de los metaterios y de los euterios, y antes de la diversificación de los placentarios, cuando ambas se expandieron. En el caso de la familia *Bex/Tceal*, se podría dar una situación llamativa: que la expansión de una familia derivada de elementos L1 haya sido facilitada por esos mismos elementos. Hay tres factores que inducen a pensar que esto es posible: el cromosoma X está enriquecido en transposones L1<sup>148</sup>; los mamíferos son un clado donde los L1 están y han estado especialmente activos<sup>42</sup>, y se ha demostrado que estos elementos pueden facilitar la duplicación génica<sup>149</sup>. Siguiendo estrategias recientes<sup>149</sup>, se podría analizar si existe un enriquecimiento en L1 específicos de euterios cerca de los genes *Bex/Tceal* que hubieran podido fomentar la expansión de la familia. Por otro lado, curiosamente, del mismo modo que en los *Bex/Tceal* (fig. 8), la disposición sinténica de las familias *PNMA* y *SIRH* también quedó establecida antes de la



diversificación de los placentarios<sup>63</sup>, con lo que parece que este periodo fue especialmente propicio para el origen y la expansión de familias génicas derivadas de transposones. Es tentador especular con que estas familias hayan podido estar implicadas en la aparición y desarrollo de estructuras cruciales para la evolución de los placentarios.

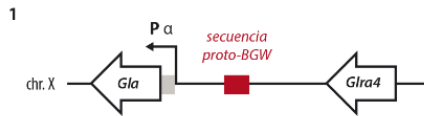
### **El motivo BGW y la regulación de la expresión de los genes BGW**

Uno de los puntos atractivos de la historia evolutiva de la familia *Bex/Tceal* es su interrelación con la de la familia *ArmcX* y el gen *Hnrnp2*, así como las consecuencias que se derivan sobre la regulación transcripcional de estos genes. Trazar el origen del motivo BGW nos permite reconstruir con relativa seguridad la formación del clúster, y supone un ejemplo de cómo los genes derivados de transposones y retrocopias cooptan regiones reguladoras para su transcripción (fig. 20).

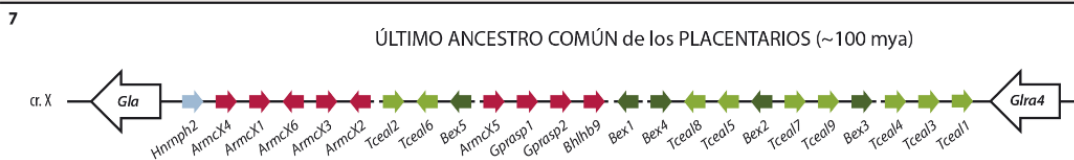
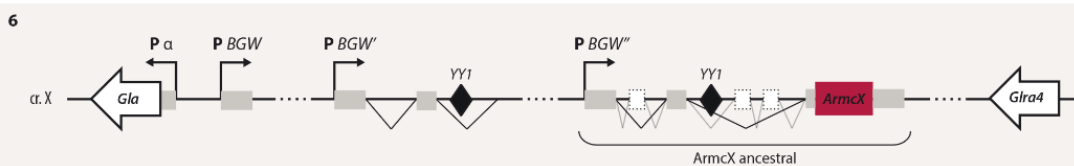
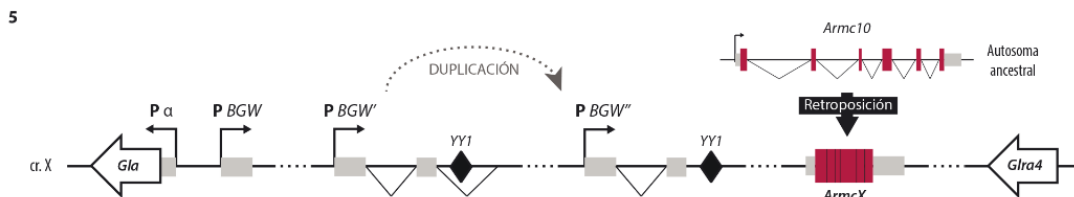
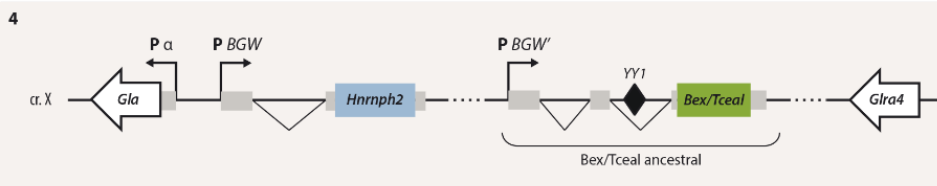
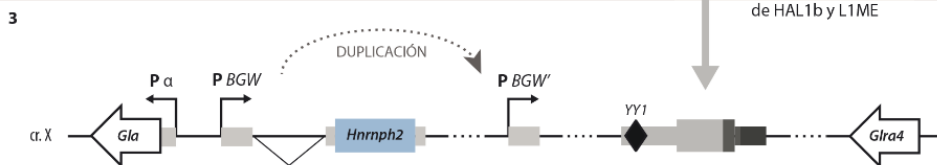
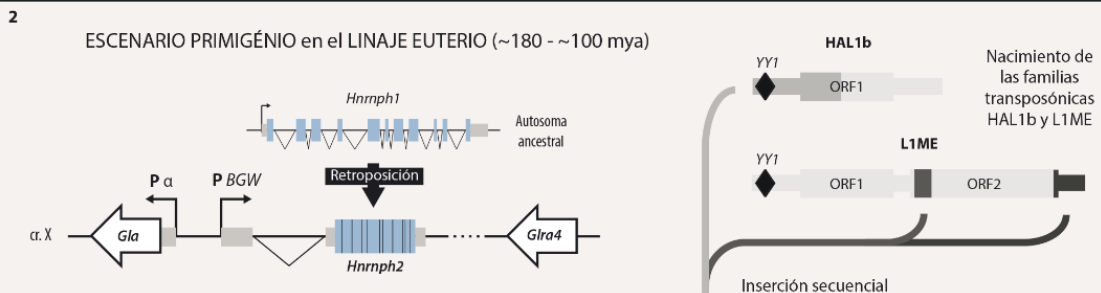
Los estudios de Shi y Garry<sup>99</sup> y de Iseki et al.<sup>100</sup> han mostrado que el motivo BGW y su *E-box* central parecen indispensables para la actividad transcripcional de estos genes. Sin embargo, quedaba por explorar su capacidad de dirigir la expresión tisular. Los resultados del ensayo de una selección de motivos BGW (fig. 14) inducen a pensar que los motivos BGW por sí solos no conferirían especificidad de tejido sino que serían secuencias altamente sensibles al estado regulatorio circundante. Por ejemplo, el motivo BGW (y regiones proximales) del *Tceal7* murino produjo un elevado número de embriones con patrones de expresión no consistentes, producto del *positional effect* (fig 14d). En cambio, esta misma región, expandida para incluir ~700 pb *upstream*, reprodujo el patrón de *Tceal7* observado durante el desarrollo embrionario (fig. 14e y fig. 15). Apoyando la hipótesis anterior, investigaciones previas han identificado sitios de unión funcionales para STAT3, c-Jun y p65 *upstream* del motivo BGW de *BEX2*<sup>137,150</sup>. Curiosamente, igual que ocurre entre los ORFs de *Bex1* y 2, *Tceal2* y 4, y *Tceal3*, 5 y 6 (fig. 9), las secuencias de los motivos BGW (y de los intrones) se homogeneizan mediante conversión génica

con relativa frecuencia<sup>96,99</sup>. Además, es muy chocante comprobar que este fenómeno no solo se da entre las regiones no codificantes de genes *Bex/Tceal*, sino también entre genes de familias distintas, como *BEX2* y *ARMCX4*<sup>96</sup>. El significado biológico de esta homogeneización de regiones reguladoras, si es que lo tuviera, permanece sin investigar, pero podría facilitar la nivelación de la expresión de los genes afectados. Por otro lado, tampoco se detectó actividad para el proto-motivo BGW *upstream* del gen *Gla* de koala (fig. 14a). En especies euterias, los motivos BGW muestran una inserción de 11 pb a 5' de la *E-box* (fig. 12). En esa zona yace un elemento de respuesta a AMP cíclico (*cAMP response element*, CRE), que parece necesario para la funcionalidad del promotor de *ARMCX1*<sup>110</sup>. Así, esta inserción específica de euterios podría haber resultado clave para la evolución del potencial regulador del motivo BGW. El motivo BGW, por lo tanto, es probablemente esencial para la transcripción de las familias génicas *Bex/Tceal* y *ArmcX*, pero señales mediadas por otras secuencias, cuya complejidad permanece por descifrar, parecen gobernar la expresión tisular de estos genes.

Otra de las observaciones relativas a la regulación de los genes *Bex/Tceal* y *ArmcX* es la conservación en la región intrónica de un sitio de unión para YY1 (fig. 13). Aunque no nos sea posible confirmarlo, la orientación y la situación con respecto al ORF, así como su capacidad de unir YY1 *in vitro*, indican una probable homología de esta secuencia y la del HAL1b ancestral a la familia *Bex/Tceal*. Un estudio de Bianchi et al. puede darnos pistas sobre la función de este sitio de unión. El único intrón del gen *Ubc* contiene un sitio de unión para YY1 en la hebra *minus*, que regula el *splicing* y potencia la transcripción<sup>151</sup>. Pese a que la literatura sobre transposones cooptados como *enhancers* es abundante<sup>54</sup>, hasta la fecha no se ha documentado ningún caso en que un TD retenga las secuencias reguladoras ancestrales propias. A falta de conocer su relevancia funcional, la conservación del sitio de unión para YY1 en los intrones de los genes *Bex/Tceal* podría constituir el primer caso.



ÚLTIMO ANCESTRO COMÚN de EUTERIOS y METATERIOS (~180 mya)



[Viene de la página anterior]

**Figura 20.** Escenario evolutivo de la formación del clúster BGW. (1) En el cromosoma X del último ancestro común de euterios y metaterios, *upstream* del promotor (P  $\alpha$ ) del gen de la alfa-galactosidasa (*Gla*) se encuentra un proto-motivo BGW. El gen *Gla4* yace inmediatamente *upstream* de ambos. (2) En el linaje euterio, un transcrito de *Hnrnp1* se retrocopia *upstream* de *Gla*, cooptando como promotor el proto-motivo BGW (P BGW), formando así un promotor bidireccional junto con P  $\alpha$ . Nacimiento de las familias de retrotransposones HAL1b y LIME. (3) P BGW', una secuencia con potencial para promover la transcripción, surge gracias a la duplicación de P BGW. Los retrotransposones HAL1b y LIME se transponen secuencialmente, dando lugar a una secuencia proto-*Bex/Tceal*. (4) La secuencia proto-*Bex/Tceal* coopta el P BGW' para transcribirse, lo que permite la formación del gen *Bex/Tceal* ancestral. Se conserva el sitio de unión para YY1 heredado del promotor del elemento HAL1b. (5) Se duplica una región que contiene el promotor P BGW' junto a un sitio de unión para YY1. El primer miembro de la familia *ArmcX* aparece cuando *Armc10* se retrocopia al cromosoma X y (6) coopta el P BGW'' para transcribirse. (7) Antes de la diversificación de los euterios actuales, las familias *Bex/Tceal* y *ArmcX* se expanden vía duplicación, dando lugar al clúster BGW. Las edades para el último ancestro común de euterio y metaterios y de los placentarios se obtuvieron de Meredith et al.<sup>241</sup> y Dos Reis et al.<sup>242</sup>.

Antes de este trabajo, solo se conocía la expresión durante el desarrollo de algunos genes *Bex/Tceal*. A la vista de diversos artículos recientes, este punto cobra especial relevancia. Ito et al. muestran que los genes *Bex1-4* de ratón tienen una expresión >400 veces más elevada en hepatoblastos (progenitores hepáticos) que en hepatocitos adultos<sup>123</sup>. Además, identifican a *Bex2* como un marcador de células progenitoras del epitelio pilórico adulto y del sistema hematopoyético. En E13.5 detectamos expresión de *Bex1-4* en el epitelio gástrico y en el primordio hepático (fig. 15 y fig. S2), ambos tejidos proliferativos<sup>152,153</sup>. Por su parte, De Angelis et al.<sup>122</sup> caracterizan la expresión epitelial de *Bex1* durante la organogénesis del esbozo pancreático, entre los estadios E9.5 y E12.5. Nosotros, en E13.5, también detectamos una potente expresión de *Bex1* en el epitelio pancreático (fig. S2). Importantemente, en este periodo del desarrollo pancreático (E9.5-E13.5) se produce una continua proliferación de las células progenitoras<sup>154</sup>. A su vez, *Bex1-3* muestran una alta expresión en progenitores neurales, que se atenúa tras la diferenciación<sup>121</sup>, y su expresión durante el desarrollo se extiende por todo el sistema nervioso central y periférico (fig. 15 y S2 y otros trabajos<sup>120,121</sup>). Por otro lado, *Tceal5* se expresa en la bolsa de Rathke en E13.5 (fig. S2), una evaginación del techo de la boca

en desarrollo que dará lugar al lóbulo anterior de la glándula pituitaria, es decir, a la adenohipófisis<sup>155</sup>. Recientemente, se ha reportado que *Tceal5* se expresa en progenitores de la glándula pituitaria promoviendo su proliferación<sup>130</sup>. De forma similar, *Tceal7* se expresa en mioblastos y regula su proliferación<sup>109</sup>, y durante el desarrollo es específico del linaje muscular (fig. 15). Otros tejidos en proliferación, como el epitelio renal<sup>156</sup> y pulmonar<sup>157</sup>, también muestran expresión de los genes *Bex/Tceal* (fig. 15 y S2). Además, *Bex2* se expresa en el epitelio del cristalino (fig. S2), y los ratones knockout para *Bex2* tienen la morfología de esta estructura alterada<sup>158</sup>. Por último, un dato sorprendente sobre la expresión de estos genes lo encontramos en un estudio de Alvarez et al.<sup>127</sup>. Al analizar los transcritos del mesencéfalo ventral en desarrollo de la rata, encontraron que los genes más comunes (el 12% de los *expressed-sequence tags*, ESTs) correspondían a miembros de la familia *Bex*. Por lo tanto, una vez contrastada la literatura con los patrones de expresión, es tentador sugerir que los genes *Bex/Tceal* regulan la proliferación y diferenciación en determinadas células progenitoras replicativas durante el desarrollo. Futuros estudios deberán terminar de caracterizar la expresión de los genes *Bex/Tceal* durante el desarrollo, así como esclarecer su relevancia funcional.

### **Generación y caracterización de ratones *Bex3* mutantes**

En este trabajo se presenta por primera vez la caracterización de una línea de ratones *Bex3* mutantes. Como se dijo en la sección de resultados, se generaron dos líneas, cada una con una deleción distinta. La línea en la que se llevaron a cabo los experimentos de fenotipado fue la L2, cuya mutación *in-frame* de 147 pb produce una deleción de 49 aminoácidos en el ORF (fig. 16). No se pudo descartar la presencia de la proteína en los ratones *Bex3*-L2 mediante *western blot* (WB), pese a que se probaron dos anticuerpos dirigidos a epítomos no delecionados. Según estudios anteriores, la proteína BEX3 es degradada rápidamente por el proteasoma, siendo el monómero especialmente sensible<sup>127,133</sup>. De hecho, en las líneas celulares PCNA y PC12, Mukai et al.

únicamente consiguieron detectar proteína tras tratar los cultivos con inhibidores del proteasoma<sup>133</sup>. En toda la literatura sobre *Bex3*, solo en el artículo de Calvo et al. se intenta y consigue (aunque de forma deficiente) detectar proteína nativa extraída de tejidos de animales vivos (neuronas del ganglio de la raíz dorsal)<sup>118</sup>. Dado que el anticuerpo utilizado en ese trabajo reconoce un epítipo que cae en la región delecionada, no es útil para nuestro estudio. Por otro lado, la proteína BEX3 humana también parece difícilmente detectable mediante WB. En *The Human Protein Atlas*, los análisis de proteína BEX3 basados en inmunohistoquímica muestran niveles elevados en cerebro y otros tejidos<sup>159</sup>. Sin embargo, los resultados del WB utilizando el mismo anticuerpo arrojan resultados no consistentes<sup>160</sup>. Por lo tanto, todo parece indicar que la detección de proteína BEX3 nativa con métodos convencionales es muy dificultosa. Una de las opciones a considerar sería tratar con inhibidores del proteasoma cultivos celulares establecidos a partir de tejidos de ratones WT y Bex3-L2, de forma similar al artículo de Mukai y colaboradores<sup>133</sup>. Otra de las alternativas consistiría en clonar en un plásmido de sobreexpresión los ORFs de ambas líneas más un *tag* amino o carboxi-terminal. De esta forma, podríamos comprobar si se genera una proteína más corta en el caso de L1, o si se aprovecha algún codón ATG alternativo en los L2. Por último, se podría analizar mediante espectrometría de masas la presencia de BEX3 en muestras Bex3-L2 y WT. De todos modos, si se confirmara que los ratones Bex3-L2 producen proteína mutante, el fenotipo observado hace pensar que la deleción de 49 aminoácidos provoca una pérdida de función en la proteína BEX3 resultante.

Los motivos principales que hacen de *Bex3* un gen atractivo son su patrón de expresión eminentemente neural, la participación en vías de señalización relevantes para el desarrollo, y los interactores que se le conocen (tabla 2).

**Tabla 2. Interactores de Bex3**

Interactor	Especie	Relevancia biológica
14-3-3ε <sup>161</sup>	Hsa, Mmu	14-3-3ε forma un complejo junto a BEX3 y p75 <sup>NTR</sup> . En células PC12nnr5 y HEK293, esta unión tiene un efecto dominante negativo sobre la inducción de la apoptosis en respuesta a NGF mediada por p75 <sup>NTR</sup> /BEX3.
BEX3 <sup>118</sup>	Mmu, Rno	El tratamiento con NGF estimula la acumulación de BEX3 dimérico en el núcleo. Una vez allí, se une al promotor de <i>trkA</i> y promueve su expresión, favoreciendo además el crecimiento neurítico.
eIF3h <sup>162</sup>	Hsa	-
p75 <sup>NTR</sup> <sup>133</sup>	Hsa, Mmu, Rno	En función de la presencia de NGF, BEX3 incrementa su interacción con el dominio intracelular de p75 <sup>NTR</sup> y parece mediar la apoptosis inducida por éste.
Smac <sup>163</sup>	Hsa, Mmu	La unión de BEX3 a Smac previene la ubiquitinación de Smac mediada por XIAP, promoviendo así la apoptosis inducida por TRAIL en células MCF-7.
TrkA <sup>118</sup>	Mmu, Rno	BEX3 se une a TrkA en células del ganglio de la raíz dorsal P0 de rata.
TrkB <sup>118</sup>	Rno	BEX3 se une a TrkB en células del ganglio de la raíz dorsal P0 de rata.
TSC1 <sup>136</sup>	Hsa, Mmu	TSC1 y 2 forman un complejo junto con BEX3 y p75 <sup>NTR</sup> . En células PC12h, TSC1 parece proteger a Bex3 de la degradación, lo que según los autores sugiere una participación de TSC1 en las señales apoptóticas mediadas por p75 <sup>NTR</sup> /BEX3.

Sobre BEX3 se ha estudiado especialmente su implicación en la señalización mediada por los receptores de neurotrofinas. El factor de crecimiento nervioso (*nerve growth factor*, NGF) es una neurotrofina que, uniéndose a los receptores p75<sup>NTR</sup> y TrkA, regula diversos aspectos de la estructura y función neuronal, tales como la supervivencia, la diferenciación, la plasticidad sináptica o la morfogénesis (*review* Reichardt, 2006). En ratón, BEX3 interacciona con el dominio citoplasmático de ambos receptores y, en función de la presencia de NGF, migra al núcleo en forma de dímero y modula la supervivencia, la diferenciación y el crecimiento neurítico<sup>118,133</sup>. Por otro lado, *Bex3* también está implicado en la vía NF-κB. Aunque se desconoce el mecanismo exacto, *BEX3* activa NF-κB al promover la desestabilización de IκBα<sup>164</sup>. IκBα secuestra en el citoplasma los heterodímeros formados por los factores de transcripción p65 y p50, que una vez translocados al núcleo se encargaran de activar el conjunto de genes *downstream* de NF-κB<sup>165</sup>.

Por último, BEX3 forma complejos con TSC1, TSC2 y p75<sup>NTR</sup><sup>136</sup>. El heterodímero TSC1/TSC2 participa en la inhibición de mTORC1, que junto con mTORC2 es uno de los dos complejos centrales de la vía mTOR

(*mechanistic target of rapamycin*)<sup>166</sup>. Puesto que p75<sup>NTR</sup> y TrkA regulan a NF- $\kappa$ B y mTOR, respectivamente<sup>167</sup>, las vías de señalización donde *Bex3* parece intervenir están estrechamente relacionadas.

En los análisis de comportamiento, los ratones *Bex3-L2* mostraron diversas alteraciones (fig. 17). Una de las más evidentes fue la disminución de la inmovilidad en las patas traseras típica de los ratones suspendidos por la cola. Aunque se hace necesaria una mejor caracterización de este fenotipo, los movimientos espasmódicos pueden ser atribuidos a un escaso control motor causado por defectos en el desarrollo de las motoneuronas, cuya supervivencia depende de una correcta transmisión de las señales neurotróficas<sup>168</sup>. Además de los movimientos espasmódicos, los ratones *Bex3-L2* mostraron comportamientos autísticos en la caja de preferencia social y en el reflejo de sobresalto (fig. 17c y 17d). Los trastornos del espectro autista tienen una base genética muy heterogénea, y se caracterizan por déficits en la socialización, dificultades en el lenguaje y la comunicación, y comportamientos repetitivos<sup>169</sup>. Las causas subyacentes al autismo se encuentran en un exceso o un defecto en el crecimiento, que se manifiesta a diferentes niveles, como pueden ser el tamaño del cerebro (micro o macrocefalia), y las conexiones neuronales (hipo o hiperconectividad)<sup>170</sup>. De manera importante, desregulaciones tanto en la señalización mediada por receptores de neurotrofinas como en la vía mTOR han sido implicadas en la etiología del autismo<sup>170</sup>. La vía mTOR integra señales ambientales (por ejemplo, nutrientes o factores de crecimiento) que regulan multitud de procesos esenciales para el crecimiento, como la síntesis proteica, el metabolismo y la homeostasis energética<sup>166</sup>. Debido a una descompensación de estos procesos, ratones knockout para inhibidores de esta vía, como PTEN o el complejo TSC1/TSC2, anteriormente mencionado, muestran un comportamiento autístico y macrocefalia, así como un incremento del tamaño del soma y de la densidad de espinas dendríticas<sup>171-174</sup>.

Excitantemente, el ratio p-S6/S6 (utilizado de manera rutinaria como indicador del estado de la vía mTOR) resultó estar significativamente aumentado en el encéfalo de ratones *Bex3-L2*, lo que indica una hiperactivación



de la vía mTOR (fig. X). Sin embargo, estudios anteriores parecen entrar en contradicción con estos resultados. Calvo et al. muestran que la disminución de la expresión de BEX3 en cultivos de neuronas del ganglio de la raíz dorsal reduce los niveles de fosforilación de Akt<sup>118</sup>. p-Akt se encarga, entre otras funciones, de impedir la inhibición de mTORC1 mediada por el complejo TSC1/TSC2<sup>166</sup>. Los experimentos que están en proceso pretenden esclarecer este punto mediante la caracterización en profundidad del estado de la vía mTOR en distintas regiones del encéfalo, como el cerebelo, el hipocampo y la corteza. Además, también se planea comprobar el tamaño y la citoarquitectura del cerebro, así como la presencia de alteraciones en ciertos tipos neuronales (como las neuronas de la capa V del córtex o las células de Purkinje) características de una hiperactivación de esta vía<sup>173,175</sup>.

Otra de las formas en que BEX3 puede estar regulando la señalización de la vía mTOR es mediante su interacción con eIF3h<sup>162</sup>. eIF3h es uno de los componentes nucleares del complejo eIF3 (*eukaryotic translation initiation factor 3*), esencial para el inicio de la traducción proteica<sup>176</sup>. eIF3 sirve de andamio para el acoplamiento de mTORC1, lo que permite la fosforilación inhibitoria de 4E-BP1 (*4E-binding protein 1*) y el avance del proceso de iniciación de la traducción<sup>177</sup>. Por lo tanto, la unión de BEX3 a uno de los componentes de eIF3 podría estar modulando la traducción de genes *downstream* de mTORC1. Curiosamente, la forma dimérica de BEX3 se une al promotor de *trkA* e influye diferencialmente los niveles de mRNA y los de proteína<sup>118</sup>.

Por otro lado, al ser diseccionados, los ratones Bex3-L2 muestran una mayor acumulación de tejido adiposo abdominal, así como una aparente esteatosis hepática (datos no mostrados). La activación de mTORC1 promueve la adipogénesis y la lipogénesis hepática<sup>166</sup>, con lo que, a falta de experimentos que las confirmen, estas observaciones preliminares respaldarían una hiperactivación generalizada de la vía mTOR. Finalmente, dado que por el momento no podemos descartar la presencia de proteína mutante, debemos considerar la posibilidad de que el fenotipo de los Bex3-L2 sea causado por una

forma constitutivamente activa de BEX3. Así, en un knockout completo (asumiendo la participación de BEX3 en esta vía), deberíamos observar una hipoactivación de mTOR. Sin embargo, los ratones *Bex3-L1*, que tienen una extensa delección que rompe el marco de lectura (fig. 16), también parecen sufrir obesidad. A la espera de nuevos resultados, la desregulación de la vía mTOR establece un punto de partida a partir del cual descifrar los mecanismos moleculares subyacentes al fenotipo de *Bex3-L2*, y es también, por lo tanto, un paso más hacia la comprensión de la función de *Bex3*.

NF- $\kappa$ B es otra de las vías dependiente de la señalización mediada por receptores de neurotrofinas en que *Bex3* parece estar involucrado<sup>164</sup>. De forma similar a mTOR, una gran variedad de señales extracelulares regulan la actividad de NF- $\kappa$ B. Esta vía ha sido involucrada en la regulación de la expresión de múltiples genes relacionados con el desarrollo neural, así como en la supervivencia neuronal y el crecimiento y la morfología neuríticas<sup>178</sup>. Aunque la relación entre el autismo y la activación de NF- $\kappa$ B no está clara<sup>179-181</sup>, estudios recientes aportan datos interesantes sobre este punto. Mutaciones en el factor de transcripción *MECP2* producen el síndrome de Rett, caracterizado por dificultades en la adquisición del lenguaje, déficits en el control motor y comportamientos autísticos<sup>182</sup>. Curiosamente, en animales *Mecp2*  $-/-$  los síntomas asociados al síndrome de Rett están mediados tanto por mTOR<sup>183,184</sup> como por NF- $\kappa$ B<sup>185</sup>. Por lo tanto, debido al papel de esta vía en los procesos mencionados y la presunta implicación de *Bex3* en su regulación, parte del fenotipo comportamental que observamos podría estar causado por alteraciones en la actividad de NF- $\kappa$ B. Como en el caso de mTOR, ya están en marcha experimentos enfocados a caracterizar la activación de la vía NF- $\kappa$ B.

En conclusión, tomando en conjunto lo que se conoce sobre *Bex3* y estos primeros resultados, consideramos atractiva la hipótesis que postula a *Bex3* como un mediador de las señales neurotróficas gracias a su capacidad de modular las vías efectoras mTOR y NF- $\kappa$ B.

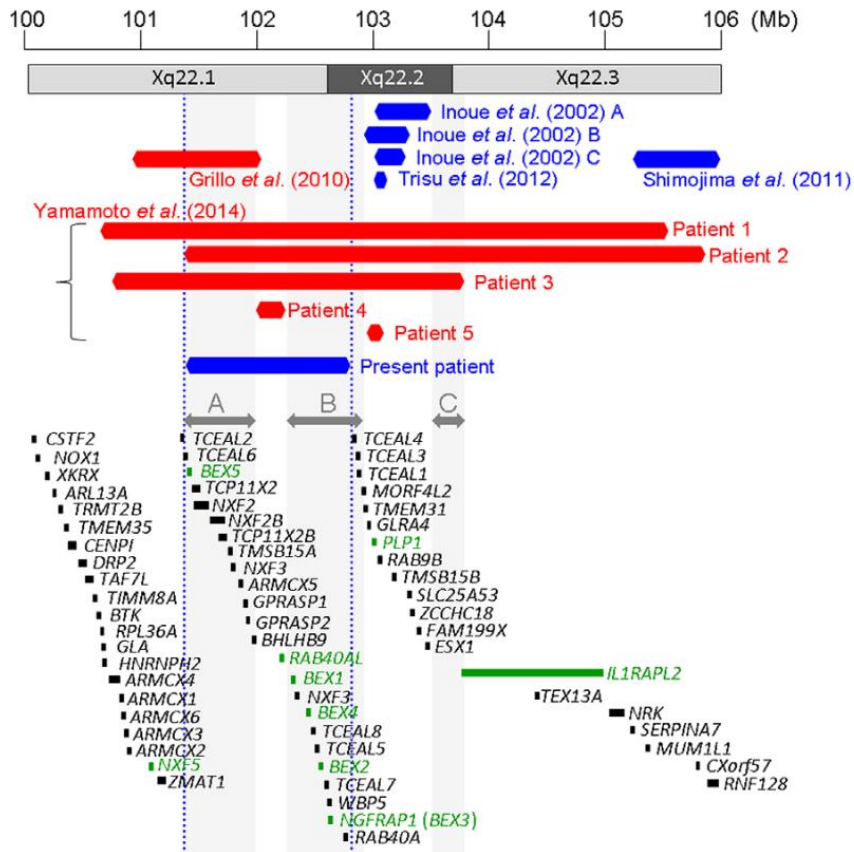
Sin embargo, tanto la influencia de estas vías como la expresión durante el desarrollo de *Bex3*, lejos de ser exclusivamente neural, se extiende a muchos

otros tejidos (fig. S2). Uno de ellos es el cartílago que originará el hueso, que en el estadio analizado (E13.5) se encuentra inmerso en el proceso de osificación endocondral<sup>186</sup>. Los ratones Bex3-L2 muestran una fusión de las costillas correspondientes a las vértebras C7 y T1, además de malformaciones craneales (fig. 19). Tanto los receptores de neurotrofinas como la vía NF-κB y el complejo mTORC1 han sido implicados en el desarrollo del sistema esquelético<sup>187-190</sup>. Sin embargo, las malformaciones craneales podrían deberse a un crecimiento anormal o excesivo del encéfalo que, como se ha comentado, es común en ratones con la vía mTOR hiperactiva. Con los datos disponibles, es por el momento muy difícil aventurar una explicación del fenotipo óseo observado en los ratones Bex3-L2.

Como último apunte, son interesantes los casos de microdeleciones cromosómicas en pacientes que abarcan partes del clúster BGW (fig. 21). Yamamoto et al. reportan cinco nuevos casos de mujeres, con distintas deleciones en heterocigosidad, que muestran severas deficiencias intelectuales, hipotonía, anomalías comportamentales y dismorfía facial<sup>191</sup>. Proponen como candidata para el fenotipo observado una región que incluye los genes *BEX1-4* y todos los *TCEAL* excepto *TCEAL2* y *TCEAL6*. El mismo grupo, hace escasos meses, ha documentado el primer caso de nulisomia de este locus en un paciente masculino, que incluye todos los genes *BEX/TCEAL* menos *TCEAL1-4*<sup>192</sup>. El paciente muestra microcefalia y graves discapacidades neurológicas y psicomotoras, y permanece encamado desde su nacimiento. El fenotipo mucho más severo de esta deleción indica que los genes implicados tendrían un importante papel en el desarrollo neurológico. Aunque estas deleciones también incluyen otros genes no pertenecientes a la familia *BEX/TCEAL*, entre ellos algunos *ARMCX*, las alteraciones en el comportamiento de la línea Bex3-L2 (fig. 17) y otras líneas knockout<sup>158</sup> inducen a pensar que parte del fenotipo observado en estos pacientes estaría causado por la ausencia los genes *Bex/Tceal*.

Finalmente, es necesario recalcar que hay que ser cuidadosos al extrapolar las funciones de BEX3 entre organismos. Aunque Mukai et al. 2000 detectaron

una interacción entre Bex3 y el dominio citoplasmático de p75NTR en células de ratón, rata y humano, estudios posteriores no han podido reproducir estos resultados<sup>127,162</sup>.



**Figura 21.** Representación de la región cromosómica Xq22. Los hexágonos azules y rojos indican las deleciones detectadas en los estudios especificados. Los rectángulos representan genes. Los rectángulos verdes son genes con una elevada expresión en el cerebro. Las regiones B y C han sido asociadas a las deficiencias neurológicas de los pacientes descritos por Yamamoto et al. (2014). *Present patient* hace referencia a la deleción del paciente estudiado por Shirai y colaboradores. Figura adaptada de Shirai et al.<sup>192</sup>.

## **Integración funcional de los genes *Bex/Tceal***

En los últimos años, empieza a acumularse evidencia que permite reconsiderar la naturaleza de los genes *Bex/Tceal*. Estos genes codificarían para un módulo de pequeñas proteínas adaptadoras con la capacidad de establecer múltiples y promiscuas interacciones, cuya función parece convergir en ciertas vías de señalización<sup>104</sup>. Así, se abre un muy estimulante espacio para la reflexión en torno a la evolución de los genes *Bex/Tceal* y su integración funcional.

Existen múltiples ejemplos de TDs para los que conocemos su función, y algunos de ellos han devenido esenciales para el organismo «huésped». Sin embargo, en referencia a los genes *de novo*, se han planteado muy acertadamente una serie de preguntas<sup>193</sup> que son igual de válidas —y es probable que igual de enigmáticas— en el caso de los TDs: ¿Cómo adquieren estos su funcionalidad? ¿Cómo devienen no solo funcionales, sino también esenciales? ¿Cómo se integran en la red de interacciones proteicas preexistente?

A diferencia de los genes duplicados, que pueden repartirse, es decir, subfuncionalizar la función ancestral, los TDs solo pueden integrarse neofuncionalizándose. En la fase de transición que enlaza los restos fragmentarios de transposones subsiguientemente domesticados con lo que podríamos considerar el TD como tal, el proceso de neofuncionalización pudo haber sucedido rápidamente mediante evolución adaptativa. De hecho, tanto la rama que lleva al grupo de genes *Bex3* como las de *Bex1-2* y *Bex4* evolucionaron bajo una intensa selección positiva<sup>99</sup>. Curiosamente, la mayoría de posiciones específicas que evolucionaron bajo selección adaptativa se encuentran en la región carboxi-terminal, que es donde mapean todas las interacciones conocidas de las proteínas BEX<sup>104</sup>. Todo ello, sumado a la difícil resolución de las relaciones filogenéticas entre los genes *Tceal* (fig. 9), hace pensar que, una vez originado el *Bex/Tceal* ancestral, una rápida expansión de la familia permitió que las distintas copias evolucionaran y probablemente diversificaran sus funciones. Como sucede con la familia *SIRH*<sup>63,194</sup>, posteriormente el set de genes *Bex/Tceal* se estableció —interesantemente, con

posteriores pérdidas y ganancias específicas de linaje— antes de la diversificación de los placentarios, y, salvo *Bex4*, evolucionaron desde entonces bajo selección purificadora (fig. 8 y 10). Sin embargo, al haber estudiado los efectos de la selección en el conjunto de la proteína, es posible que posiciones específicas en otras proteínas hayan evolucionado bajo selección adaptativa en ciertos linajes.

Durante el proceso de neofuncionalización, es necesario que los TDs se integren y remodelen las redes ancestrales de interacciones proteicas. Una característica de estas redes es la existencia de nodos (o *hubs*) altamente conectados, de los que depende la estabilidad del conjunto de interacciones: no es sorprendente, por lo tanto, la constatación que existe una relación entre conectividad y esencialidad<sup>195</sup>. En observaciones que no creemos arriesgado extrapolar a los TDs, se ha confirmado que los genes originados por transferencia horizontal y *de novo* tienden a localizarse en la periferia de las redes de interacciones y muestran una baja conectividad<sup>196,197</sup>. Sin embargo, ¿existe alguna característica que facilite esa integración y allane el camino para el desarrollo de funciones esenciales?

Interesantemente, según Zhang et al., las proteínas de genes jóvenes con una alta conectividad (proteínas *hub*) muestran regiones de baja complejidad e intrínsecamente desordenadas<sup>196</sup>. Así, una estructura desordenada facilitaría una rápida integración en las redes de interacciones y, por lo tanto, facilitaría el establecimiento de interrelaciones esenciales. Las regiones intrínsecamente desordenadas carecen de una estructura tridimensional fija, aunque tienen la capacidad de plegarse al formar complejos con otras proteínas<sup>198</sup>. A las proteínas con un alto porcentaje de este tipo de regiones se las llama proteínas intrínsecamente desordenadas (*intrinsically-disordered proteins*, IDPs) y, gracias a una flexibilidad estructural que les permite interactuar promiscuamente con otras proteínas, se encuentran sobrerrepresentadas entre las proteínas *hub*<sup>199</sup>. Recientemente, se ha demostrado que las proteínas BEX son IDPs con regiones *coiled-coil* en su extremo carboxi-terminal<sup>103,104</sup> y, con

el estudio estructural de las proteínas TCEAL, hemos podido expandir estos hallazgos a toda la familia BEX/TCEAL (fig. 11).

Los *coiled-coil* se consideran regiones intrínsecamente desordenadas, pues pueden adoptar una conformación «ordenada» al oligomerizarse o unirse a un interactor<sup>198</sup>. Curiosamente, el segmento de ORF1p del cual derivan los genes *Bex/Tceal* se corresponde con el inicio del *coiled-coil* característico de estas proteínas transposónicas (fig. 3). Se conocen varias interacciones con proteínas endógenas para la proteína ORF1p humana<sup>200-202</sup>, una de ellas mediada por la región *coiled-coil*<sup>203</sup>. Pese a la rápida tasa evolutiva a que está sometida la proteína ORF1p, y especialmente el *coiled-coil*<sup>141,142</sup>, no vemos ninguna razón por la que el ORF1p del elemento HAL1b activo en el ancestro de los placentarios no pudiera interactuar a su vez con algunas (quizás las mismas) proteínas endógenas. Por otro lado, tanto las ORF1p (gracias a la región *coiled-coil*) como ciertas proteínas BEX/TCEAL forman oligómeros<sup>103,204</sup>.

Por lo tanto, parece evidente que las propiedades estructurales de las proteínas BEX/TCEAL (y puede que algunos de sus interactores), claves para la funcionalidad de esta familia, han sido determinadas por las características de la proteína transposónica ancestral. Así, podemos concebir un escenario donde la integración y modulación de redes ancestrales, favorecida por unas características estructurales particulares, permitieron la acelerada evolución adaptativa y la expansión de la familia génica *Bex/Tceal*.

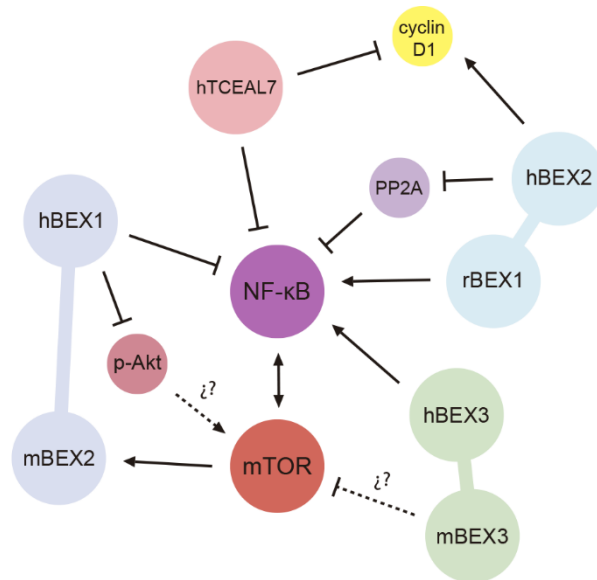
Si entendemos que un gen es esencial cuando resulta crítico para el desarrollo y la supervivencia de un organismo, no podemos decir que los genes *Bex/Tceal*, estudiados uno a uno, lo sean. Sin contar el presente estudio, entre los ya publicados y los que están siendo analizados por el IMPC (International Mouse Phenotyping Consortium), se han generado líneas de ratones knockout para *Bex1*<sup>116,117</sup>, *Bex2*<sup>123,158</sup>, *Bex4*<sup>205</sup>, *Tceal5*<sup>206</sup> y *Tceal9*<sup>207</sup>. Es necesario subrayar que el concepto de esencialidad es fácilmente extensible a otros procesos que, aunque no perturben críticamente el desarrollo de un organismo, sí comprometen de forma decisiva su viabilidad: por ejemplo, al interactuar con su entorno. Además, parece claro que el conjunto de la familia génica sí

sería esencial *sensu stricto*, vistos los severos efectos de su delección en pacientes humanos<sup>192</sup>.

Con el objetivo de comprender cómo los nuevos genes pueden devenir esenciales o funcionalmente relevantes, Chen et al. sugieren buscar la explicación en la renovación de vías de señalización ancestrales<sup>8</sup>. Pero, ¿en qué vías de señalización se integran las proteínas BEX/TCEAL? Fernández et al. proponen que los genes *Bex* formarían una familia de proteínas *hub* cuya extensa red de interacciones convergería en la vía de señalización NF- $\kappa$ B<sup>104</sup>. El hecho que, al revisar la literatura, confundan *TCEAL7* con *BEX4* hace incluso más interesante su hipótesis, pues enlaza la integración de los *Bex* y los *Tceal*. La discusión de la integración funcional de estos genes a la luz de este y anteriores estudios puede ayudarnos a comprender el papel evolutivo de esta nueva familia de transposones domesticados.

Los análisis a escala global del interactoma humano en distintos tejidos y líneas celulares parecen confirmar que las proteínas *Bex/Tceal* desempeñan el papel de proteínas *hub*. Por ejemplo, gracias a estos estudios se han identificado más de treinta interactores para BEX2, BEX3 y TCEAL1<sup>208-211</sup>. Por lo tanto, queda para el futuro un análisis exhaustivo de la red de interacciones de estos genes y su significado biológico. Sin embargo, tal y como apuntan Fernández et al., del estudio de la literatura existente parecen surgir ciertas líneas de convergencia en las vías de señalización moduladas por los genes *Bex/Tceal* (fig. 22). Como se ha comentado anteriormente, BEX3 interacciona con los receptores TrkA y p75<sup>NTR</sup>. En respuesta a NGF, migra al núcleo, donde promueve la diferenciación y el crecimiento neurítico<sup>118,133</sup>. Además, *BEX3* parece ser un activador de la vía NF- $\kappa$ B<sup>164</sup>. Curiosamente, en rata, *Bex1* muestra un comportamiento totalmente opuesto. Al tratar cultivos neuronales con NGF, BEX1 se transloca del núcleo al citoplasma, donde se une a p75<sup>NTR</sup> e inhibe la activación de la vía NF- $\kappa$ B<sup>121</sup>.





**Figura 22.** Esquema de la acción de los genes *Bex/Tceal* sobre la ciclina D1 y las vías de señalización mTOR y NF-κB. Los prefijos m (ratón), r (rata) y h (humano) indican la especie del ortólogo en cuestión. Debido a un error en la anotación, el ortólogo de *BEX1* humano es el gen *Bex2* de rata y ratón, y viceversa.

La anotación del gen *Bex1* murino en el locus humano de *BEX2* ha generado una gran confusión entre la comunidad científica que estudia la función de los genes *BEX1-2*. A pesar de ello, ambos han sido implicados en la regulación de NF-κB en humanos, inhibiéndola *BEX1*<sup>164</sup> y activándola *BEX2*<sup>212</sup>. Puesto que son proteínas cuya secuencia ha sido homogeneizada por conversión génica, creemos probable que exista una fina regulación de la expresión que compatibilice sus funciones, quizás mediada por el motivo BGW. De forma interesante, la expresión de *Bex1* depende de la fase del ciclo celular<sup>121</sup>. Por otro lado, *TCEAL7* también desempeña una función en esta vía, en su caso inhibiendo la activación de genes *downstream* de NF-κB<sup>131</sup>. Además, tanto *TCEAL7*<sup>213</sup> como *BEX2*<sup>214</sup> controlan la progresión del ciclo celular mediante la regulación de los niveles de ciclina D1.

Finalmente, la vía mTOR también parece estar bajo la influencia de los genes *Bex/Tceal*. Una cascada de señalización desencadenada por mTORC1 activa a NF-κB y media la sobreexpresión del gen *Bex2* murino, necesaria para la activación de genes *downstream*<sup>137</sup>. A su vez, la proteína BEX2 humana inhibe la acción de la fosfatasa PP2a, un conocido inhibidor de la vía mTOR<sup>150</sup>. Y por último, esta misma vía está alterada en los encéfalos de ratones *Bex3-L2* (fig. 18).

Las vías mTOR, NF-κB y NGF/p75<sup>NTR</sup>, así como la regulación del ciclo celular mediada por ciclinas, tienen un origen evolutivo muy anterior a la aparición de los euterios<sup>215-219</sup>. Por lo tanto, los genes *Bex/Tceal* regulan el ciclo celular, la apoptosis y la diferenciación gracias a su integración convergente en una serie de vías de señalización ancestrales. Sin embargo, quedan por explorar las consecuencias de su integración para la evolución de los euterios. Como hemos dicho, las vías de señalización mencionadas son esenciales para el desarrollo del sistema nervioso. En el artículo que describía por primera vez el clúster BGW, los autores proponían que estos genes podrían haber participado en la evolución adaptativa del neocórtex<sup>96</sup>, una innovación evolutiva de los mamíferos<sup>220</sup>. Tanto los mamíferos placentarios como marsupiales han reorganizado independientemente el neocórtex ancestral, desarrollando paralelamente un amplio rango de comportamientos y estilos de vida<sup>221</sup>. Sin embargo, en ciertos aspectos los placentarios muestran una mayor complejidad y plasticidad comportamental<sup>222</sup>. Además, aunque el tamaño cerebral relativo de ambos grupos es similar<sup>223</sup>, estudios recientes apuntan a que los marsupiales tendrían una densidad neuronal mucho menor<sup>224</sup>. Dadas las funciones conocidas de los genes *Bex/Tceal* y las deficiencias comportamentales observadas en ratones *Bex3-L2* (fig. 17) y knockout para *Bex2*<sup>158</sup>, es posible que un incremento en la complejidad de las vías de señalización relevantes para el desarrollo del sistema nervioso, posibilitada por la aparición de esta familia de proteínas, haya participado en la evolución de la etología y la citoarquitectura cerebral de los placentarios.

Además de su hipotético impacto sobre la evolución del ancestro de los placentarios, es posible que cambios en la regulación transcripcional, en las secuencias proteicas y en el número de genes de la familia *Bex/Tceal* hayan mediado innovaciones evolutivas en los distintos linajes de placentarios. Algunos genes *Bex/Tceal* muestran patrones de expresión específicos de linaje. Este es el caso de *Tceal7*: su expresión en ratón es muscular (fig. 15; Shi y Garry<sup>109</sup>), mientras que el ortólogo humano se expresa principalmente en el sistema nervioso<sup>114</sup>. Por otro lado, *Bex4* es el único gen *Bex/Tceal* sobre cuya secuencia proteica ha actuado, después de la radiación de los placentarios, una evidente relajación de las constricciones evolutivas (fig. 10). En cuanto a su función, *Bex4* es una *rara avis* dentro de su familia. En humano, BEX4 regula la unión de la sirtuina a la  $\alpha$ -tubulina, modulando así la acetilación y estabilidad de los microtúbulos<sup>100</sup>. Queda por ver si la selección positiva detectada es fruto de un proceso de neofuncionalización. Curiosamente, Tong y colaboradores encuentran que la proteína BEX3 humana se une a otra proteína asociada a los microtúbulos e involucrada en el transporte vesicular, la dinactina, aunque no van más allá en la caracterización de esta interacción<sup>162</sup>. Por lo tanto, es posible que la flexibilidad estructural de estas proteínas no solo haya participado en la remodelación de vías de señalización en el ancestro de los placentarios, sino que sea una cualidad recurrentemente cooptada por múltiples procesos biológicos en los linajes descendientes.

## **Conclusiones**



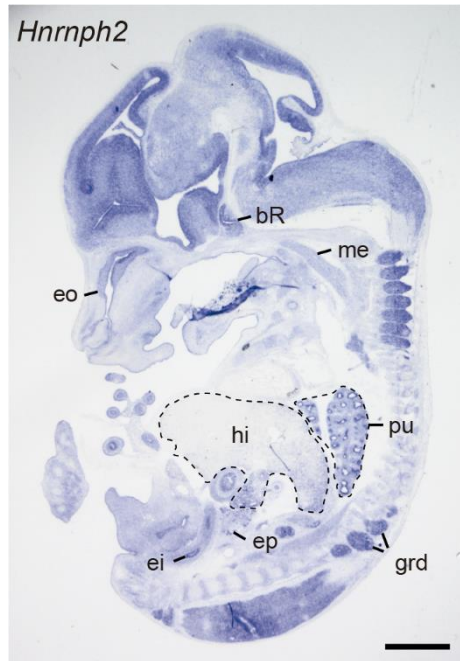
1. La combinación de criterios de búsqueda basados en la conservación de dominios y en el solapamiento con transposones, así como una cuidadosa anotación genómica, son necesarias para una óptima identificación de candidatos a TDs.
2. La familia *Bex/Tceal* tiene su origen en una domesticación molecular de fragmentos de retrotransposones L1 en el ancestro de los mamíferos placentarios. Esta es la segunda domesticación molecular de un retrotransposón sin LTRs en metazoos, y la primera en originar una familia génica.
3. La familia *Bex/Tceal* se expandió antes de la radiación de los placentarios, aunque algunos linajes han sufrido recurrentes pérdidas. Después de la diversificación de los mamíferos placentarios, la familia *Bex/Tceal* evolucionó, con la excepción de *Bex4*, bajo selección purificadora, probablemente para conservar las funciones adquiridas.
4. La estructura desordenada y con *coiled-coils* de los genes *Bex/Tceal* viene determinada por la estructura de la proteína de la cual derivan, es decir, el extremo amino-terminal de la proteína ORF1p de HAL1b. Esta estructura permite a las proteínas BEX/TCEAL una promiscuidad en sus interacciones que es clave para su función.
5. Al trazar el origen del motivo BGW a una secuencia *upstream* del gen *Gla* ya presente en el ancestro de los euterios, se ha resuelto la enigmática relación de las familias *Bex/Tceal* y *ArmcX* y del gen *Hnrnp2*. Además, el motivo BGW supone un ejemplo de cómo los TDs y los retrogenes son capaces de cooptar regiones reguladoras.
6. El motivo BGW parece no ser el responsable de conferir a la expresión transcripcional de los genes *Bex/Tceal* su especificidad de tejido, sino que más bien sería una secuencia con una alta capacidad de sensar las señales transcripcionales circundantes. Por lo tanto, la regulación de la expresión de los genes *Bex/Tceal* depende de *enhancers* distales.
7. En caso de confirmarse la funcionalidad y homología del sitio de unión para YY1 detectado en la región intrónica de los genes *Bex/Tceal*, supondría la primera vez que se reporta la conservación de regiones reguladoras del transposón ancestral en un TD.

8. Los genes *Bex/Tceal* parecen expresarse en células progenitoras proliferantes durante el desarrollo, probablemente regulando la proliferación y diferenciación celular.
9. Ratones con mutaciones en el gen *Bex3* muestran alteraciones en el comportamiento, algunas de ellas asociadas al autismo, y diversas malformaciones esqueléticas. Estos resultados pueden ayudar a comprender las bases genéticas de las deficiencias cognitivas y las malformaciones anatómicas de los pacientes afectados por las microdeleciones Xq22.1.
10. La hiperactivación de la vía mTOR detectada en el encéfalo de los ratones homocigotos para la mutación de la línea *Bex3-L2* ofrece una primera pista sobre la base molecular del fenotipo observado en estos ratones.
11. La estructura desordenada, la amplia red de interactores y la aparente convergencia funcional en ciertas vías de señalización relevantes para el desarrollo, convierten a las proteínas *BEX/TCEAL* en un interesante ejemplo de cómo los genes nuevos son capaces de integrarse y modular redes genéticas preexistentes. Queda para futuras investigaciones caracterizar en profundidad la integración funcional de estos genes, así como concretar su participación en las innovaciones evolutivas de los mamíferos euterios.

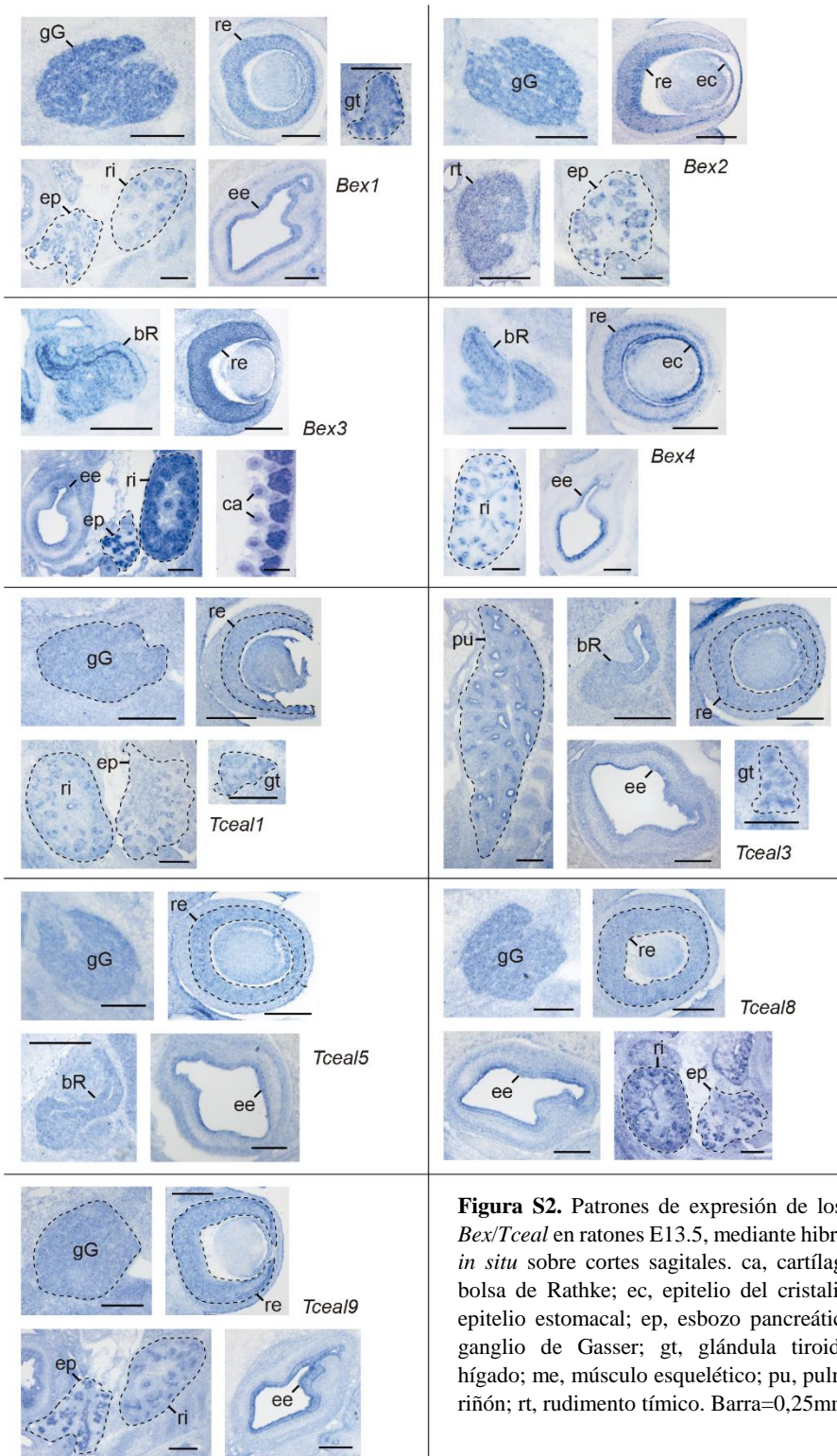
## **Material suplementario**







**Figura S1.** Patrón de expresión del gen *Hnrnp2* en ratones E13.5, mediante hibridación *in situ* sobre cortes sagitales. El patrón, especialmente intenso en SNC y SNP, es similar al de los genes *Bex1-3*. bR, bolsa de Rathke; ei, epitelio intestinal; eo, epitelio olfatorio; ep, esbozo pancreático; grd, ganglios de la raíz dorsal; hi, hígado; me, músculo esquelético; pu, pulmón. Barra=1 mm.



**Figura S2.** Patrones de expresión de los genes *Bex/Tceal* en ratones E13.5, mediante hibridación *in situ* sobre cortes sagitales. ca, cartílago; bR, bolsa de Rathke; ec, epitelio del cristalino; ee, epitelio estomacal; ep, esbozo pancreático; gG, ganglio de Gasser; gt, glándula tiroides; hi, hígado; me, músculo esquelético; pu, pulmón; ri, riñón; rt, rudimento túnico. Barra=0,25mm.

Tabla S1. Anotación de los patrones de expresión de los genes Bex/Tceal en ratones E13.5 mediante ISH

	SNC																
	Ep. resp.	Ep. inteañ.	Ep. olfat.	Ep. estom.	Ep. riñón	Esbozo pancreát.	B. Rathke	Gl. tiroid.	Rudimento tímico	Corazón	Higado	Músculo esquelético	GRD	Gang. Gass.	Retina	Méd. esp.	Encéfalo
<i>Tceal1</i>	+	-	+	-	+	+	+	+	-	-	-	+	+	+	+	+	+
<i>Tceal3</i>	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+
<i>Tceal5</i>	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+
<i>Tceal7</i>	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-	-	-
<i>Tceal8</i>	++	++	+	+	+	+	+	+	+	-	+	+	++	++	+	+	+
<i>Tceal9</i>	++	++	++	++	++	++	++	++	+	-	+	-	+	+	+	+	+
<i>Bex1</i>	++	++	+	+	+	++	+	+	+	-	+	++	++	++	++	++	++
<i>Bex2</i>	+	+	+	+	+	++	+	+	+	-	+	+	++	++	++	++	++
<i>Bex3</i>	++	++	+	+	++	++	+	+	+	+	+	+	++	++	++	++	++
<i>Bex4</i>	++	++	+	+	+	++	+	++	-	++	++	-	+	+	+	+	+

- no expresado; + expresado; ++ intensamente expresado



## **Referencias**



1. Muller. The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. *Genetics* **17**, (1935).
2. Ohno, S. *Evolution by Gene Duplication*. (Springer Berlin Heidelberg, 1970). doi:10.1007/978-3-642-86659-3
3. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Research* **20**, 1313–1326 (2010).
4. Force, A. *et al.* Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* **151**, 1531–45 (1999).
5. Tokuriki, N. & Tawfik, D. S. Protein Dynamism and Evolvability. *Science (80-. )*. **324**, 203–207 (2009).
6. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25–36 (2008).
7. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).
8. Chen, S., Krinsky, B. H. & Long, M. New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* **14**, 645–60 (2013).
9. Blomme, T. *et al.* The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* **7**, R43 (2006).
10. Hahn, M. W., Han, M. V. & Han, S.-G. Gene Family Evolution across 12 Drosophila Genomes. *PLoS Genet.* **3**, e197 (2007).
11. Hahn, M. W., Demuth, J. P. & Han, S.-G. Accelerated Rate of Gene Gain and Loss in Primates. *Genetics* **177**, 1941–1949 (2007).
12. Jaillon, O. *et al.* Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature* **431**, 946–957 (2004).
13. Dehal, P. & Boore, J. L. Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biol.* **3**, e314 (2005).
14. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–624 (2004).
15. Brunet, F. G. *et al.* Gene Loss and Evolutionary Rates Following Whole-Genome Duplication in Teleost Fishes. *Mol. Biol. Evol.* **23**, 1808–1816 (2006).
16. Session, A. M. *et al.* Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature* **538**, 336–343 (2016).
17. Gilbert, W. The exon theory of genes. *Cold Spring Harb. Symp. Quant.*



- Biol.* **52**, 901–5 (1987).
18. Wang, W., Zhang, J., Alvarez, C., Llopart, A. & Long, M. The origin of the Jingwei gene and the complex modular structure of its parental gene, yellow emperor, in *Drosophila melanogaster*. *Mol. Biol. Evol.* **17**, 1294–301 (2000).
  19. Crisp, A., Boschetti, C., Perry, M., Tunnacliffe, A. & Micklem, G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol.* **16**, 50 (2015).
  20. Li, F.-W. *et al.* Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 6672–7 (2014).
  21. Ropars, J. *et al.* Adaptive Horizontal Gene Transfers between Multiple Cheese-Associated Fungi. *Curr. Biol.* **25**, 2562–9 (2015).
  22. Yin, Z., Zhu, B., Feng, H. & Huang, L. Horizontal gene transfer drives adaptive colonization of apple trees by the fungal pathogen *Valsa mali*. *Sci. Rep.* **6**, 33129 (2016).
  23. Wijayawardena, B. K., Minchella, D. J. & DeWoody, J. A. Hosts, parasites, and horizontal gene transfer. *Trends Parasitol.* **29**, 329–338 (2013).
  24. McLysaght, A. & Guerzoni, D. New genes from non-coding sequence: the role of de novo protein-coding genes in eukaryotic evolutionary innovation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**, 20140332 (2015).
  25. de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLoS Genet.* **7**, e1002384 (2011).
  26. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
  27. Piégu, B., Bire, S., Arensbürger, P. & Bigot, Y. A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. *Mol. Phylogenet. Evol.* **86**, 90–109 (2015).
  28. McClintock, B. Cytogenetic Studies of Maize and *Neurospora*. *Carnegie Inst. Washingt. Yearb.* 146–152 (1947).
  29. Lerat, E. & Capy, P. Retrotransposons and retroviruses: analysis of the envelope gene. *Mol. Biol. Evol.* **16**, 1198–207 (1999).
  30. Casavant, N. C. *et al.* The end of the LINE?: lack of recent L1 activity in a group of South American rodents. *Genetics* **154**, 1809–17 (2000).

31. Cantrell, M. A., Scott, L., Brown, C. J., Martinez, A. R. & Wichman, H. A. Loss of LINE-1 activity in the megabats. *Genetics* **178**, 393–404 (2008).
32. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5280–5 (2003).
33. Tollis, M. & Boissinot, S. in *Genome dynamics* **7**, 68–91 (2012).
34. Martin, S. L. & Bushman, F. D. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* **21**, 467–75 (2001).
35. Januszyk, K. *et al.* Identification and solution structure of a highly conserved C-terminal domain within ORF1p required for retrotransposition of long interspersed nuclear element-1. *J. Biol. Chem.* **282**, 24893–904 (2007).
36. Khazina, E. & Weichenrieder, O. Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 731–6 (2009).
37. Mathias, S. L., Scott, A. F., Kazazian, H. H., Boeke, J. D. & Gabriel, A. Reverse transcriptase encoded by a human transposable element. *Science* **254**, 1808–10 (1991).
38. Feng, Q., Moran, J. V., Kazazian, H. H. & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–16 (1996).
39. Heidmann, T., Esnault, C. & Maestre, J. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**, 363–367 (2000).
40. Zhu, Z., Tan, S., Zhang, Y. & Zhang, Y. E. LINE-1-like retrotransposons contribute to RNA-based gene duplication in dicots. *Sci. Rep.* **6**, 24755 (2016).
41. Denli, A. M. *et al.* Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* **163**, 583–593 (2015).
42. Ivancevic, A. M., Kortschak, R. D., Bertozzi, T. & Adelson, D. L. LINEs between Species: Evolutionary Dynamics of LINE-1 Retrotransposons across the Eukaryotic Tree of Life. *Genome Biol. Evol.* **8**, 3301–3322 (2016).
43. Bao, W. & Jurka, J. Origin and evolution of LINE-1 derived ‘half-L1’ retrotransposons (HAL1). *Gene* **465**, 9–16 (2010).
44. Athanikar, J. N., Badge, R. M. & Moran, J. V. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic*

- Acids Res.* **32**, 3846–3855 (2004).
45. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
  46. Eddy, S. R. The C-value paradox, junk DNA and ENCODE. *Curr. Biol.* **22**, R898-9 (2012).
  47. Doolittle, W. F. Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5294–300 (2013).
  48. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mob. DNA* **7**, 9 (2016).
  49. Brunet, T. D. P. & Doolittle, W. F. Multilevel Selection Theory and the Evolutionary Functions of Transposable Elements. *Genome Biol. Evol.* **7**, 2445–57 (2015).
  50. Indjeian, V. B. *et al.* Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins. *Cell* **164**, 45–56 (2016).
  51. Ding, Y., Berrocal, A., Morita, T., Longden, K. D. & Stern, D. L. Natural courtship song variation caused by an intronic retroelement in an ion channel gene. *Nature* **536**, 329–332 (2016).
  52. Hof, A. E. van't *et al.* The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**, 102–105 (2016).
  53. Janoušek, V., Karn, R. C. & Laukaitis, C. M. The role of retrotransposons in gene family expansions: insights from the mouse *Abp* gene family. *BMC Evol. Biol.* **13**, 107 (2013).
  54. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
  55. Irimia, M. & Roy, S. W. Origin of Spliceosomal Introns and Alternative Splicing. *Cold Spring Harb. Perspect. Biol.* **6**, a016071–a016071 (2014).
  56. Huff, J. T., Zilberman, D. & Roy, S. W. Mechanism for DNA transposons to generate introns on genomic scales. *Nature* **538**, 533–536 (2016).
  57. Johnson, R. & Guigo, R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* **20**, 959–976 (2014).
  58. Britten, R. Transposable elements have contributed to thousands of

- human proteins. *Proc. Natl. Acad. Sci.* **103**, 1798–1803 (2006).
59. Miller, W. J., McDonald, J. F. & Pinsky, W. Molecular domestication of mobile elements. *Genetica* **100**, 261–70 (1997).
  60. Kidwell, M. G. & Lisch, D. R. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**, 1–24 (2001).
  61. Hoen, D. R. & Bureau, T. E. Discovery of Novel Genes Derived from Transposable Elements Using Integrative Genomic Analysis. *Mol. Biol. Evol.* **32**, 1487–1506 (2015).
  62. Carelli, F. N. *et al.* The life history of retrocopies illuminates the evolution of new mammalian genes. *Genome Res.* **26**, 301–14 (2016).
  63. Kokošar, J. & Kordiš, D. Genesis and regulatory wiring of retroelement-derived domesticated genes: a phylogenomic perspective. *Mol. Biol. Evol.* **30**, 1015–31 (2013).
  64. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022 (2011).
  65. Fenouil, R. *et al.* CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **22**, 2399–2408 (2012).
  66. Ploy, M.-C., Lambert, T., Couty, J.-P. & Denis, F. Integrons: an Antibiotic Resistance Gene Capture and Expression System. *Clin. Chem. Lab. Med.* **38**, 483–7 (2000).
  67. Fablet, M., Bueno, M., Potrzebowski, L. & Kaessmann, H. Evolutionary Origin and Functions of Retrogene Introns. *Mol. Biol. Evol.* **26**, 2147–2156 (2009).
  68. Kordiš, D. & Kokošar, J. What Can Domesticated Genes Tell Us about the Intron Gain in Mammals? *Int. J. Evol. Biol.* **2012**, 278981 (2012).
  69. Chorev, M. & Carmel, L. The function of introns. *Front. Genet.* **3**, 55 (2012).
  70. Huang, S. *et al.* Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* **166**, 102–114 (2016).
  71. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405 (2008).
  72. Sinzelle, L., Izsvák, Z. & Ivics, Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* **66**, 1073–1093 (2009).
  73. Chalopin, D., Naville, M., Plard, F., Galiana, D. & Volff, J.-N.

- Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **7**, 567–80 (2015).
74. Ray, D. A. *et al.* Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* **18**, 717–728 (2008).
  75. Lavialle, C. *et al.* Paleovirology of ‘syncytins’, retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120507–20120507 (2013).
  76. Kaneko-Ishino, T. & Ishino, F. The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front. Microbiol.* **3**, 262 (2012).
  77. Irie, M. *et al.* Cognitive Function Related to the Sirh11/Zcchc16 Gene Acquired from an LTR Retrotransposon in Eutherians. *PLoS Genet.* **11**, e1005521 (2015).
  78. Kapusta, A. *et al.* Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
  79. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
  80. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
  81. Gasteiger, E. *et al.* ExpASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–8 (2003).
  82. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
  83. Ronquist, F. *et al.* MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst. Biol.* **61**, 539–542 (2012).
  84. Abascal, F., Zardoya, R. & Posada, D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104–2105 (2005).
  85. Dodou, E., Xu, S.-M. & Black, B. L. mef2c is activated directly by myogenic basic helix-loop-helix proteins during skeletal muscle development in vivo. *Mech. Dev.* **120**, 1021–32 (2003).
  86. Ferran, J. L., Puelles, L. & Rubenstein, J. L. R. Molecular codes defining rostrocaudal domains in the embryonic mouse hypothalamus. *Front.*

- Neuroanat.* **9**, 46 (2015).
87. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–8 (2015).
  88. Mi, S. *et al.* Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**, 785–9 (2000).
  89. Blaise, S., de Parseval, N., Bénit, L. & Heidmann, T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 13013–8 (2003).
  90. Hayward, A., Ghazal, A., Andersson, G., Andersson, L. & Jern, P. ZBED evolution: repeated utilization of DNA transposons as regulators of diverse host functions. *PLoS One* **8**, e59940 (2013).
  91. Campillos, M., Doerks, T., Shah, P. K. & Bork, P. Computational characterization of multiple Gag-like human proteins. *Trends Genet.* **22**, 585–589 (2006).
  92. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
  93. Smit, A. F., Tóth, G., Riggs, A. D. & Jurka, J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* **246**, 401–417 (1995).
  94. Szak, S. T. *et al.* Molecular archeology of L1 insertions in the human genome. *Genome Biol.* **3**, research0052 (2002).
  95. Ostertag, E. M. & Kazazian, H. H. Twin Priming: A Proposed Mechanism for the Creation of Inversions in L1 Retrotransposition. *Genome Res.* **11**, 2059–2065 (2001).
  96. Winter, E. E. & Ponting, C. P. Mammalian BEX, WEX and GASP genes: coding and non-coding chimaerism sustained by gene conversion events. *BMC Evol. Biol.* **5**, 54 (2005).
  97. Foley, N. M., Springer, M. S. & Teeling, E. C. Mammal madness: is the mammal tree of life not yet resolved? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **371**, 20150140 (2016).
  98. Giordano, J. *et al.* Evolutionary History of Mammalian Transposons Determined by Genome-Wide Defragmentation. *PLoS Comput. Biol.* **3**, e137 (2007).
  99. Zhang, L. Adaptive evolution and frequent gene conversion in the brain expressed X-linked gene family in mammals. *Biochem. Genet.* **46**, 293–311 (2008).

100. Lee, J.-K. *et al.* Oncogenic microtubule hyperacetylation through BEX4-mediated sirtuin 2 inhibition. *Cell Death Dis.* **7**, e2336 (2016).
101. Martin, S. L., Branciforte, D., Keller, D. & Bain, D. L. Trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl. Acad. Sci.* **100**, 13815–13820 (2003).
102. Abrusan, G., Zhang, Y. & Szilagyi, A. Structure Prediction and Analysis of DNA Transposon and LINE Retrotransposon Proteins. *J. Biol. Chem.* **288**, 16127–16138 (2013).
103. Cabral, K. M. S. *et al.* Biophysical Studies on BEX3, the p75NTR-Associated Cell Death Executor, Reveal a High-Order Oligomer with Partially Folded Regions. *PLoS One* **10**, e0137916 (2015).
104. Fernandez, E. M., Díaz-Ceso, M. D. & Vilar, M. Brain expressed and X-linked (Bex) proteins are intrinsically disordered proteins (IDPs) and form new signaling hubs. *PLoS One* **10**, e0117206 (2015).
105. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 996–1010 (2010).
106. Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P. & Dunker, A. K. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins Struct. Funct. Bioinforma.* **61**, 176–182 (2005).
107. Obradovic, Z. *et al.* Predicting intrinsic disorder from amino acid sequence. *Proteins Struct. Funct. Genet.* **53**, 566–572 (2003).
108. Alva, V., Nam, S.-Z., Söding, J. & Lupas, A. N. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**, W410-5 (2016).
109. Shi, X. & Garry, D. J. Myogenic regulatory factors transactivate the *Tceal7* gene and modulate muscle differentiation. *Biochem. J.* **428**, 213–221 (2010).
110. Iseki, H. *et al.* Human Arm protein lost in epithelial cancers, on chromosome X 1 (ALEX1) gene is transcriptionally regulated by CREB and Wnt/ $\beta$ -catenin signaling. *Cancer Sci.* **101**, 1361–1366 (2010).
111. López-Doménech, G. *et al.* The Eutherian *Armex* genes regulate mitochondrial trafficking in neurons and interact with Miro and Trak2. *Nat. Commun.* **3**, 814 (2012).
112. Andrianaki, A., Siapati, E. K., Hirata, R. K., Russell, D. W. & Vassilopoulos, G. Dual transgene expression by foamy virus vectors

- carrying an endogenous bidirectional promoter. *Gene Ther.* **17**, 380–8 (2010).
113. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
  114. Tissue expression of TCEAL7 - Summary - The Human Protein Atlas. Available at: <http://www.proteinatlas.org/ENSG00000182916-TCEAL7/tissue>. (Accessed: 28th May 2017)
  115. Kothary, R. *et al.* Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707–14 (1989).
  116. Koo, J. H., Smiley, M. A., Lovering, R. M. & Margolis, F. L. Bex1 knock out mice show altered skeletal muscle regeneration. *Biochem. Biophys. Res. Commun.* **363**, 405–10 (2007).
  117. Khazaei, M. R. *et al.* Bex1 is involved in the regeneration of axons after injury. *J. Neurochem.* **115**, 910–20 (2010).
  118. Calvo, L. *et al.* Bex3 Dimerization Regulates NGF-Dependent Neuronal Survival and Differentiation by Enhancing trkA Gene Transcription. *J. Neurosci.* **35**, 7190–202 (2015).
  119. Jiang, C., Wang, J.-H., Yue, F. & Kuang, S. The brain expressed x-linked gene 1 (Bex1) regulates myoblast fusion. *Dev. Biol.* **409**, 16–25 (2016).
  120. Kim, A.-J., Lee, C.-S. & Schlessinger, D. Bex3 associates with replicating mitochondria and is involved in possible growth control of F9 teratocarcinoma cells. *Gene* **343**, 79–89 (2004).
  121. Vilar, M. *et al.* Bex1, a novel interactor of the p75 neurotrophin receptor, links neurotrophin signaling to the cell cycle. *EMBO J.* **25**, 1219–30 (2006).
  122. De Angelis, M. T. *et al.* Novel pancreas organogenesis markers refine the pancreatic differentiation roadmap of embryonic stem cells. *Stem Cell Rev.* **10**, 269–79 (2014).
  123. Ito, K. *et al.* Gene targeting study reveals unexpected expression of brain-expressed X-linked 2 in endocrine and tissue stem/progenitor cells in mice. *J. Biol. Chem.* **289**, 29892–911 (2014).
  124. Brown, A. L. & Kay, G. F. Bex1, a gene with increased expression in parthenogenetic embryos, is a member of a novel gene family on the mouse X chromosome. *Hum. Mol. Genet.* **8**, 611–9 (1999).
  125. Behrens, M., Margolis, J. W. & Margolis, F. L. Identification of members of the Bex gene family as olfactory marker protein (OMP)



- binding partners. *J. Neurochem.* **86**, 1289–96 (2003).
126. Koo, J. H., Saraswati, M. & Margolis, F. L. Immunolocalization of Bex protein in the mouse brain and olfactory system. *J. Comp. Neurol.* **487**, 1–14 (2005).
  127. Alvarez, E., Zhou, W., Witta, S. E. & Freed, C. R. Characterization of the Bex gene family in humans, mice, and rats. *Gene* **357**, 18–28 (2005).
  128. Akaishi, J. *et al.* Down-regulation of transcription elongation factor A (SII) like 4 (TCEAL4) in anaplastic thyroid cancer. *BMC Cancer* **6**, 260 (2006).
  129. Gylfe, A. E. *et al.* Identification of Candidate Oncogenes in Human Colorectal Cancers With Microsatellite Instability. *Gastroenterology* **145**, 540–543.e22 (2013).
  130. Gergics, P., Christian, H. C., Choo, M. S., Ajmal, A. & Camper, S. A. Gene Expression in Mouse Thyrotrope Adenoma: Transcription Elongation Factor Stimulates Proliferation. *Endocrinology* **157**, 3631–3646 (2016).
  131. Rattan, R. *et al.* TCEAL7, a putative tumor suppressor gene, negatively regulates NF- $\kappa$ B pathway. *Oncogene* **29**, 1362–1373 (2010).
  132. Kazi, J. U., Kabir, N. N. & Rönstrand, L. Brain-Expressed X-linked (BEX) proteins in human cancers. *Biochim. Biophys. Acta* **1856**, 226–33 (2015).
  133. Mukai, J. *et al.* NADE, a p75NTR-associated Cell Death Executor, Is Involved in Signal Transduction Mediated by the Common Neurotrophin Receptor p75NTR. *J. Biol. Chem.* **275**, 17566–17570 (2000).
  134. Mukai, J., Suvant, P. & Sato, T.-A. Nerve growth factor-dependent regulation of NADE-induced apoptosis. *Vitam. Horm.* **66**, 385–402 (2003).
  135. Sato, A. mTOR, a Potential Target to Treat Autism Spectrum Disorder. *CNS Neurol. Disord. Drug Targets* **15**, 533–43 (2016).
  136. Yasui, S. *et al.* The TSC1 gene product hamartin interacts with NADE. *Mol. Cell. Neurosci.* **35**, 100–8 (2007).
  137. Hu, Z. *et al.* Brain-expressed X-linked 2 Is Pivotal for Hyperactive Mechanistic Target of Rapamycin (mTOR)-mediated Tumorigenesis. *J. Biol. Chem.* **290**, 25756–25765 (2015).
  138. Nekrutenko, A. & Li, W. H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–21 (2001).

139. Wu, M., Li, L. & Sun, Z. Transposable element fragments in protein-coding regions and their contributions to human functional proteins. *Gene* **401**, 165–171 (2007).
140. Zdobnov, E. M., Campillos, M., Harrington, E. D., Torrents, D. & Bork, P. Protein coding potential of retroviruses and other transposable elements in vertebrate genomes. *Nucleic Acids Res.* **33**, 946–954 (2005).
141. Boissinot, S. & Furano, A. V. Adaptive evolution in LINE-1 retrotransposons. *Mol. Biol. Evol.* **18**, 2186–94 (2001).
142. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2005).
143. Amselem, J. *et al.* Evolution of the EKA family of powdery mildew avirulence-effector genes from the ORF 1 of a LINE retrotransposon. *BMC Genomics* **16**, 917 (2015).
144. McLaughlin, R. N. *et al.* Positive selection and multiple losses of the LINE-1-derived L1TD1 gene in mammals suggest a dual role in genome defense and pluripotency. *PLoS Genet.* **10**, e1004531 (2014).
145. Choi, Y. *et al.* Molecular structure of canine LINE-1 elements in canine transmissible venereal tumor. *Anim. Genet.* **30**, 51–3 (1999).
146. Martin, S. L., Li, W.-L. P., Furano, A. V. & Boissinot, S. The structures of mouse and human L1 elements reflect their insertion mechanism. *Cytogenet. Genome Res.* **110**, 223–228 (2005).
147. Meyer, T. J., Srikanta, D., Conlin, E. M. & Batzer, M. A. Heads or tails: L1 insertion-associated 5' homopolymeric sequences. *Mob. DNA* **1**, 7 (2010).
148. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005).
149. Janoušek, V., Laukaitis, C. M., Yanchukov, A. & Karn, R. C. The Role of Retrotransposons in Gene Family Expansions in the Human and Mouse Genomes. *Genome Biol. Evol.* **8**, 2632–2650 (2016).
150. Naderi, A., Liu, J. & Hughes-Davies, L. BEX2 has a functional interplay with c-Jun/JNK and p65/RelA in breast cancer. *Mol. Cancer* **9**, 111 (2010).
151. Bianchi, M. *et al.* Yin Yang 1 Intronic Binding Sequences and Splicing Elicit Intron-Mediated Enhancement of Ubiquitin C Gene Expression. *PLoS One* **8**, e65932 (2013).
152. Creamer, B., Sorter, R. G. & Bamforth, J. The turnover and shedding of epithelial cells. I. The turnover in the gastro-intestinal tract. *Gut* **2**, 110–

- 8 (1961).
153. Miyajima, A., Tanaka, M. & Itoh, T. Stem/Progenitor Cells in Liver Development, Homeostasis, Regeneration, and Reprogramming. *Cell Stem Cell* **14**, 561–574 (2014).
  154. Pan, F. C. & Wright, C. Pancreas organogenesis: From bud to plexus to gland. *Dev. Dyn.* **240**, 530–565 (2011).
  155. Rizzoti, K. & Lovell-Badge, R. Early Development of the Pituitary Gland: Induction and Shaping of Rathke’s Pouch. *Rev. Endocr. Metab. Disord.* **6**, 161–172 (2005).
  156. Lechner, M. S. & Dressler, G. R. The molecular basis of embryonic kidney development. *Mech. Dev.* **62**, 105–20 (1997).
  157. Rackley, C. R. & Stripp, B. R. Building and maintaining the epithelium of the lung. *J. Clin. Invest.* **122**, 2724–2730 (2012).
  158. Gene details for brain expressed X-linked 2 | International Mouse Phenotyping Consortium. Available at: <http://www.mousephenotype.org/data/genes/MGI:1338017>. (Accessed: 28th May 2017)
  159. Human Protein Atlas. Tissue expression of NGFRAP1 - Summary - The Human Protein Atlas. (2017). Available at: <http://www.proteinatlas.org/ENSG00000166681-NGFRAP1/tissue>.
  160. NGFRAP1 - Antibodies - The Human Protein Atlas. Available at: [http://www.proteinatlas.org/ENSG00000166681-NGFRAP1/antibody#assay\\_western\\_blot](http://www.proteinatlas.org/ENSG00000166681-NGFRAP1/antibody#assay_western_blot). (Accessed: 28th May 2017)
  161. Kimura, M. T. *et al.* 14-3-3 is involved in p75 neurotrophin receptor-mediated signal transduction. *J. Biol. Chem.* **276**, 17291–300 (2001).
  162. Tong, X., Xie, D., Roth, W., Reed, J. & Koeffler, H. P. NADE (p75NTR-associated cell death executor) suppresses cellular growth in vivo. *Int. J. Oncol.* **22**, 1357–62 (2003).
  163. Yoon, K., Jang, H. D. & Lee, S. Y. Direct interaction of Smac with NADE promotes TRAIL-induced apoptosis. *Biochem. Biophys. Res. Commun.* **319**, 649–54 (2004).
  164. Gewurz, B. E. *et al.* Genome-wide siRNA screen for mediators of NF- $\kappa$ B activation. *Proc. Natl. Acad. Sci.* **109**, 2467–2472 (2012).
  165. Hoesel, B. & Schmid, J. A. The complexity of NF- $\kappa$ B signaling in inflammation and cancer. *Mol. Cancer* **12**, 86 (2013).
  166. Laplante, M. & Sabatini, D. M. mTOR Signaling in Growth Control and Disease. *Cell* **149**, 274–293 (2012).

167. Reichardt, L. F. Neurotrophin-regulated signalling pathways. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1545–1564 (2006).
168. Domeniconi, M., Hempstead, B. L. & Chao, M. V. Pro-NGF secreted by astrocytes promotes motor neuron cell death. *Mol. Cell. Neurosci.* **34**, 271–279 (2007).
169. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).
170. Subramanian, M., Timmerman, C. K., Schwartz, J. L., Pham, D. L. & Meffert, M. K. Characterizing autism spectrum disorders by key biochemical pathways. *Front. Neurosci.* **9**, 313 (2015).
171. Kwon, C.-H., Zhu, X., Zhang, J. & Baker, S. J. mTor is required for hypertrophy of Pten-deficient neuronal soma in vivo. *Proc. Natl. Acad. Sci.* **100**, 12923–12928 (2003).
172. Kwon, C.-H. *et al.* Pten Regulates Neuronal Arborization and Social Interaction in Mice. *Neuron* **50**, 377–388 (2006).
173. Tsai, P. T. *et al.* Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice. *Nature* **488**, 647–651 (2012).
174. Sato, A. *et al.* Rapamycin reverses impaired social interaction in mouse models of tuberous sclerosis complex. *Nat. Commun.* **3**, 1292 (2012).
175. Huang, W.-C., Chen, Y. & Page, D. T. Hyperconnectivity of prefrontal cortex to amygdala projections in a mouse model of macrocephaly/autism syndrome. *Nat. Commun.* **7**, 13421 (2016).
176. Masutani, M., Sonenberg, N., Yokoyama, S. & Imataka, H. Reconstitution reveals the functional core of mammalian eIF3. *EMBO J.* **26**, 3373–83 (2007).
177. Ma, X. M. & Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nat. Rev. Mol. Cell Biol.* **10**, 307–18 (2009).
178. Gutierrez, H. & Davies, A. M. Regulation of neural process growth, elaboration and structural plasticity by NF- $\kappa$ B. *Trends Neurosci.* **34**, 316–25 (2011).
179. Malik, M. *et al.* NF- $\kappa$ B Signaling in the Brain of Autistic Subjects. *Mediators Inflamm.* **2011**, 1–10 (2011).
180. Naik, U. S. *et al.* A study of nuclear transcription factor-kappa B in childhood autism. *PLoS One* **6**, e19488 (2011).
181. Young, A. M. H., Campbell, E., Lynch, S., Suckling, J. & Powis, S. J. Aberrant NF-kappaB expression in autism spectrum condition: a

- mechanism for neuroinflammation. *Front. psychiatry* **2**, 27 (2011).
182. Zoghbi, H. Y. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).
  183. Ricciardi, S. *et al.* Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. *Hum. Mol. Genet.* **20**, 1182–1196 (2011).
  184. Tsujimura, K. *et al.* miR-199a Links MeCP2 with mTOR Signaling and Its Dysregulation Leads to Rett Syndrome Phenotypes. *Cell Rep.* **12**, 1887–1901 (2015).
  185. Kishi, N. *et al.* Reduction of aberrant NF- $\kappa$ B signalling ameliorates Rett syndrome phenotypes in Mecp2-null mice. *Nat. Commun.* **7**, 10520 (2016).
  186. Kozhemyakina, E., Lassar, A. B. & Zelzer, E. A pathway to bone: signaling molecules and transcription factors involved in chondrocyte development and maturation. *Development* **142**, 817–831 (2015).
  187. Wu, S., Flint, J. K., Rezvani, G. & De Luca, F. Nuclear Factor- B p65 Facilitates Longitudinal Bone Growth by Inducing Growth Plate Chondrocyte Proliferation and Differentiation and by Preventing Apoptosis. *J. Biol. Chem.* **282**, 33698–33706 (2007).
  188. Wu, S., Fadoju, D., Rezvani, G. & De Luca, F. Stimulatory Effects of Insulin-like Growth Factor-I on Growth Plate Chondrogenesis Are Mediated by Nuclear Factor- B p65. *J. Biol. Chem.* **283**, 34037–34044 (2008).
  189. Chen, J. & Long, F. mTORC1 signaling controls mammalian skeletal growth through stimulation of protein synthesis. *Development* **141**, 2848–2854 (2014).
  190. Wu, H. *et al.* Bone Size and Quality Regulation: Concerted Actions of mTOR in Mesenchymal Stromal Cells and Osteoclasts. *Stem cell reports* (2017). doi:10.1016/j.stemcr.2017.04.005
  191. Yamamoto, T. *et al.* An emerging phenotype of Xq22 microdeletions in females with severe intellectual disability, hypotonia and behavioral abnormalities. *J. Hum. Genet.* **59**, 300–6 (2014).
  192. Shirai, K., Higashi, Y., Shimojima, K. & Yamamoto, T. An Xq22.1q22.2 nullisomy in a male patient with severe neurological impairment. *Am. J. Med. Genet. A* **173**, 1124–1127 (2017).
  193. McLysaght, A. & Hurst, L. D. Open questions in the study of de novo genes: what, how and why. *Nat. Rev. Genet.* **17**, 567–78 (2016).

194. Lynch, C. & Tristem, M. A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. *Curr. Biol.* **13**, 1518–23 (2003).
195. Batada, N. N. *et al.* Stratus Not Altocumulus: A New View of the Yeast Protein Interaction Network. *PLoS Biol.* **4**, e317 (2006).
196. Zhang, W., Landback, P., Gschwend, A. R., Shen, B. & Long, M. New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* **16**, 202 (2015).
197. Lercher, M. J. & Pál, C. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* **25**, 559–67 (2008).
198. Cumberworth, A., Lamour, G., Babu, M. M. & Gsponer, J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem. J.* **454**, 361–9 (2013).
199. Haynes, C. *et al.* Intrinsic Disorder Is a Common Feature of Hub Proteins from Four Eukaryotic Interactomes. *PLoS Comput. Biol.* **2**, e100 (2006).
200. Zhu, Y. *et al.* L1-ORF1p, a Smad4 interaction protein, promotes proliferation of HepG2 cells and tumorigenesis in mice. *DNA Cell Biol.* **32**, 531–40 (2013).
201. Yang, Q. *et al.* LINE-1 ORF-1p functions as a novel HGF/ETS-1 signaling pathway co-activator and promotes the growth of MDA-MB-231 cell. *Cell. Signal.* **25**, 2652–60 (2013).
202. Goodier, J. L., Cheung, L. E. & Kazazian, H. H. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* **41**, 7401–7419 (2013).
203. Lu, Y. *et al.* LINE-1 ORF-1p functions as a novel androgen receptor co-activator and promotes the growth of human prostatic carcinoma cells. *Cell. Signal.* **25**, 479–89 (2013).
204. Anti-TCEAL7 antibody (ab122945) | Abcam. Available at: [http://www.abcam.com/tceal7-antibody-ab122945.html#description\\_images\\_1](http://www.abcam.com/tceal7-antibody-ab122945.html#description_images_1). (Accessed: 28th May 2017)
205. Gene details for brain expressed X-linked 4 | International Mouse Phenotyping Consortium. Available at: <http://www.mousephenotype.org/data/genes/MGI:3606746>. (Accessed: 28th May 2017)
206. Dickinson, M. E. *et al.* High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).

207. Gene details for WW domain binding protein 5 | International Mouse Phenotyping Consortium. Available at: <http://www.mousephenotype.org/data/genes/MGI:109567>. (Accessed: 28th May 2017)
208. Ewing, R. M. *et al.* Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3**, 89 (2007).
209. Wang, J. *et al.* Toward an understanding of the protein interaction network of the human liver. *Mol. Syst. Biol.* **7**, 536–536 (2014).
210. Rolland, T. *et al.* A Proteome-Scale Map of the Human Interactome Network. *Cell* **159**, 1212–1226 (2014).
211. Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425–40 (2015).
212. Naderi, A. *et al.* BEX2 is overexpressed in a subset of primary breast cancers and mediates nerve growth factor/nuclear factor-kappaB inhibition of apoptosis in breast cancer cell lines. *Cancer Res.* **67**, 6725–36 (2007).
213. Chien, J. *et al.* A role for candidate tumor-suppressor gene TCEAL7 in the regulation of c-Myc activity, cyclin D1 levels and cellular transformation. *Oncogene* **27**, 7223–34 (2008).
214. Naderi, A., Liu, J. & Bennett, I. C. BEX2 regulates mitochondrial apoptosis and G1 cell cycle in breast cancer. *Int. J. Cancer* **126**, n/a-n/a (2009).
215. van Dam, T. J. P., Zwartkruis, F. J. T., Bos, J. L. & Snel, B. Evolution of the TOR pathway. *J. Mol. Evol.* **73**, 209–20 (2011).
216. Wang, X. W., Tan, N. S., Ho, B. & Ding, J. L. Evidence for the ancient origin of the NF- $\kappa$ B cascade: Its archaic role in pathogen infection and immunity. *Proc. Natl. Acad. Sci.* **103**, 4204–4209 (2006).
217. Benito-Gutiérrez, È., Garcia-Fernández, J. & Comella, J. X. Origin and evolution of the Trk family of neurotrophic receptors. *Mol. Cell. Neurosci.* **31**, 179–192 (2006).
218. Bothwell, M. Evolution of the neurotrophin signaling system in invertebrates. *Brain. Behav. Evol.* **68**, 124–32 (2006).
219. Gunbin, K. V, Suslov, V. V, Turnaev, I. I., Afonnikov, D. A. & Kolchanov, N. A. Molecular evolution of cyclin proteins in animals and fungi. *BMC Evol. Biol.* **11**, 224 (2011).
220. Kaas, J. H. Neocortex in early mammals and its subsequent variations. *Ann. N. Y. Acad. Sci.* **1225**, 28–36 (2011).

221. Karlen, S. J. & Krubitzer, L. The functional and anatomical organization of marsupial neocortex: Evidence for parallel evolution across mammals. *Prog. Neurobiol.* **82**, 122–141 (2007).
222. Vaughan, T. A., Ryan, J. M. (James M. & Czaplewski, N. J. *Mammalogy*. (Jones and Bartlett Publishers, 2011).
223. Weisbecker, V. & Goswami, A. Brain size, life history, and metabolism at the marsupial/placental dichotomy. *Proc. Natl. Acad. Sci.* **107**, 16216–16221 (2010).
224. Seelke, A. M. H., Dooley, J. C. & Krubitzer, L. A. The cellular composition of the marsupial neocortex. *J. Comp. Neurol.* **522**, 2286–2298 (2014).
225. Esnault, C., Cornelis, G., Heidmann, O. & Heidmann, T. Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the EnvV syncytin, captured for a function in placentation. *PLoS Genet.* **9**, e1003400 (2013).
226. Kjeldbjerg, A. L., Villesen, P., Aagaard, L. & Pedersen, F. S. Gene conversion and purifying selection of a placenta-specific ERV-V envelope gene during simian evolution. *BMC Evol. Biol.* **8**, 266 (2008).
227. Best, S., Tissier, P. Le, Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* **382**, 826–829 (1996).
228. Murcia, P. R., Arnaud, F. & Palmarini, M. The Transdominant Endogenous Retrovirus enJS56A1 Associates with and Blocks Intracellular Trafficking of Jaagsiekte Sheep Retrovirus Gag. *J. Virol.* **81**, 1762–1772 (2007).
229. Yap, M. W., Colbeck, E., Ellis, S. A. & Stoye, J. P. Evolution of the Retroviral Restriction Gene Fv1: Inhibition of Non-MLV Retroviruses. *PLoS Pathog.* **10**, e1003968 (2014).
230. Stremlau, M., Perron, M., Welikala, S. & Sodroski, J. Species-Specific Variation in the B30.2(SPRY) Domain of TRIM5 Determines the Potency of Human Immunodeficiency Virus Restriction. *J. Virol.* **79**, 3139–3145 (2005).
231. Delorme-Axford, E., Sadovsky, Y. & Coyne, C. B. The Placenta as a Barrier to Viral Infections. *Annu. Rev. Virol.* **1**, 133–146 (2014).
232. Smit, A. F. & Riggs, A. D. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1443–8 (1996).
233. Fachinetti, D. *et al.* DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Dev. Cell* **33**, 314–327 (2015).



234. Casola, C., Hucks, D. & Feschotte, C. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol. Biol. Evol.* **25**, 29–41 (2008).
235. Mateo, L. & González, J. Pogo-like Transposases Have Been Repeatedly Domesticated into CENP-B-Related Proteins. *Genome Biol. Evol.* **6**, 2008–2016 (2014).
236. Cornelis, G. *et al.* Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E487-96 (2015).
237. Cost, G. J. & Boeke, J. D. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* **37**, 18081–93 (1998).
238. Cost, G. J., Feng, Q., Jacquier, A. & Boeke, J. D. Human L1 element target-primed reverse transcription in vitro. *EMBO J.* **21**, 5899–910 (2002).
239. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1872–7 (1997).
240. Tarver, J. E. *et al.* The Interrelationships of Placental Mammals and the Limits of Phylogenetic Inference. *Genome Biol. Evol.* **8**, 330–344 (2016).
241. Meredith, R. W. *et al.* Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. *Science (80-. ).* **334**, 521–524 (2011).
242. dos Reis, M. *et al.* Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proceedings. Biol. Sci.* **279**, 3491–500 (2012).

## **Anexos**

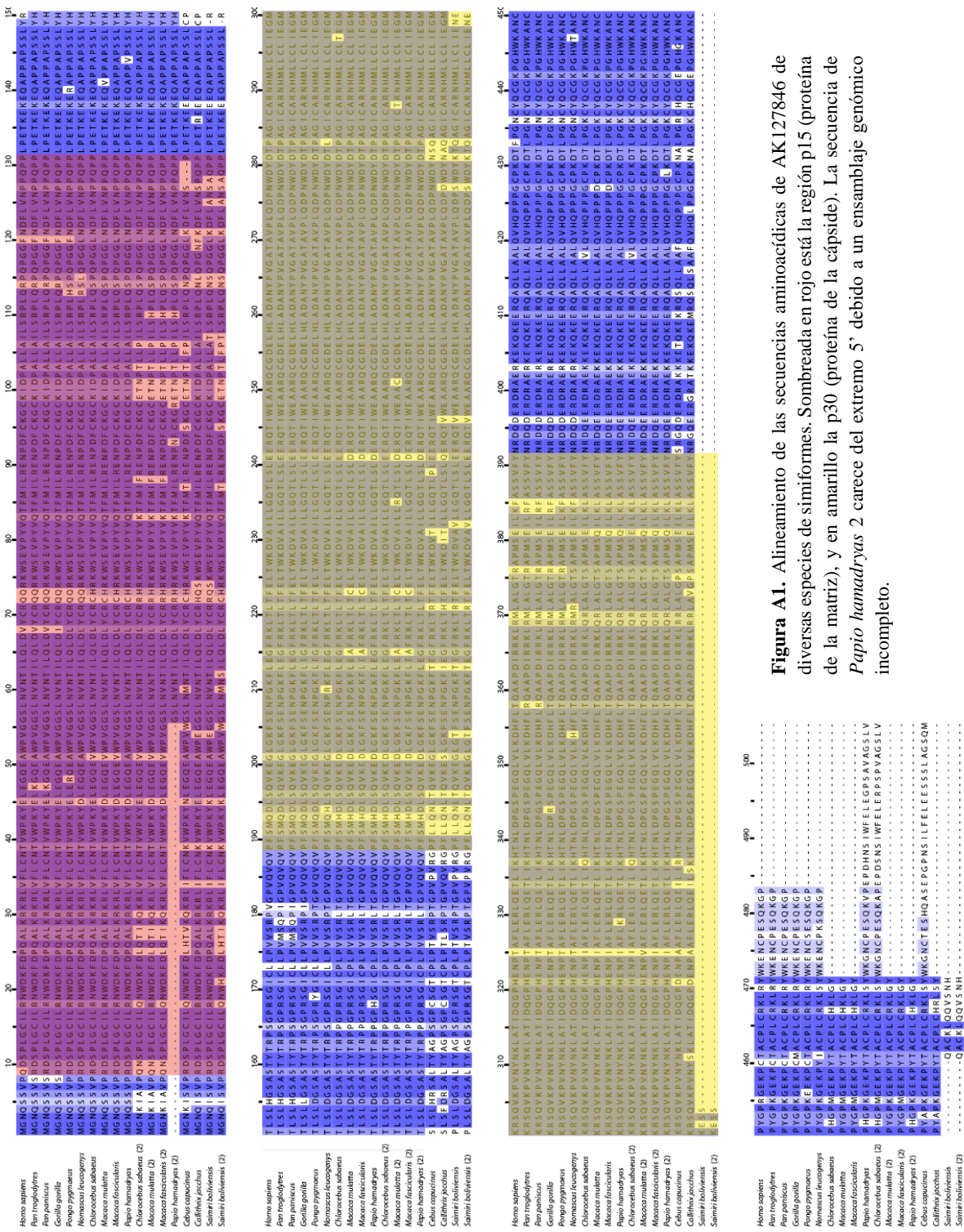


## Anexo I

Otro de los candidatos atractivos es *AK127846* (CTD-2620I22.3), derivado del gen *gag* de un ERV de tipo V<sup>1</sup>. Aunque está anotado como no codificante en GENCODE v24, el ORF de *AK127846* está conservado en los primates simiformes (fig A1). Curiosamente, el gen *env* del mismo provirus ERVV fue domesticado en el origen de los simios, dando lugar a *ERVV-1*<sup>2</sup>.

Se conocen solo dos casos análogos a *AK127846*, el gen *Fv1* en el género *Mus*<sup>3</sup> y enjS56A1 en oveja<sup>4</sup>. Ambos genes *gag* han devenido factores de restricción, esto es, genes del huésped con capacidad antiviral. Además de conservar las regiones p30 (proteína de la cápside) y p15 (proteína de la matriz), el extremo carboxi-terminal de *AK127846* es interespecíficamente variable (fig. S3), como sucede en los factores de restricción *Fv1* y *TRIM5 $\alpha$* <sup>5,6</sup>.

La única información que pudimos obtener sobre *AK127846* fueron sus *expressed sequence tags* (ESTs), todas derivadas de placenta humana, y un artículo donde se describe que su expresión es dependiente de *OVOL1*, un gen que promueve la diferenciación del sincitiotrofoblasto<sup>7</sup>. Debido a que la placenta, y más concretamente, el sincitiotrofoblasto forman una barrera frente a invasiones víricas<sup>7</sup>, decidimos investigar si *AK127846* podría ser un nuevo factor de restricción específico de los simiformes. En colaboración con el laboratorio de Frank Kirchhoff, del Instituto de Virología Molecular de la Universidad de Ulm, se realizaron unas pruebas de sobreexpresión de la proteína *AK127846* (acoplada a una secuencia de hemaglutinina) en células HEK293T (datos no mostrados). Debido a que no fue posible detectar proteína, se renunció a seguir con los experimentos. Sin embargo, estos primeros resultados negativos no descartan la posibilidad de que sí se produzca proteína de forma estable en otras líneas celulares (por ejemplo, en células CHM1htert, línea celular derivada de tejido trofoblástico).



**Figura A1.** Alineamiento de las secuencias aminoacídicas de AK127846 de diversas especies de simiiformes. Sombreada en rojo está la región p15 (proteína de la matriz), y en amarillo la p30 (proteína de la cápside). La secuencia de *Papio hamadryas* 2 carece del extremo 5' debido a un ensamblaje genómico incompleto.

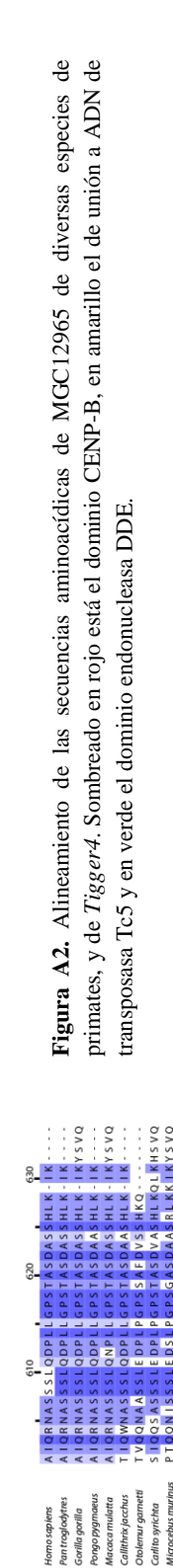
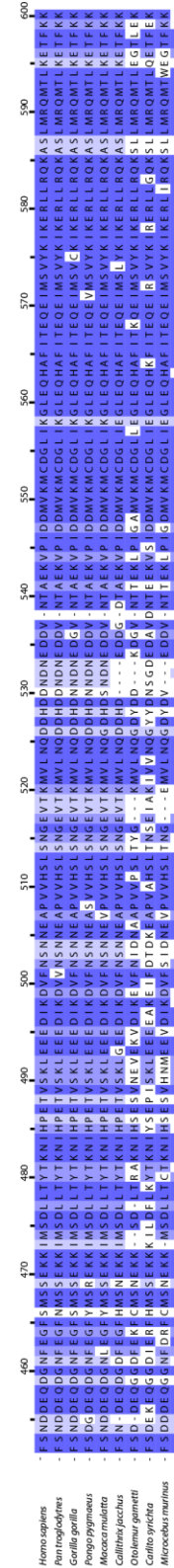
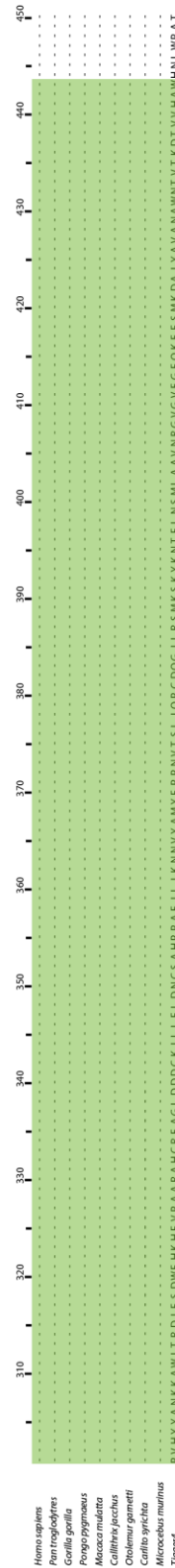
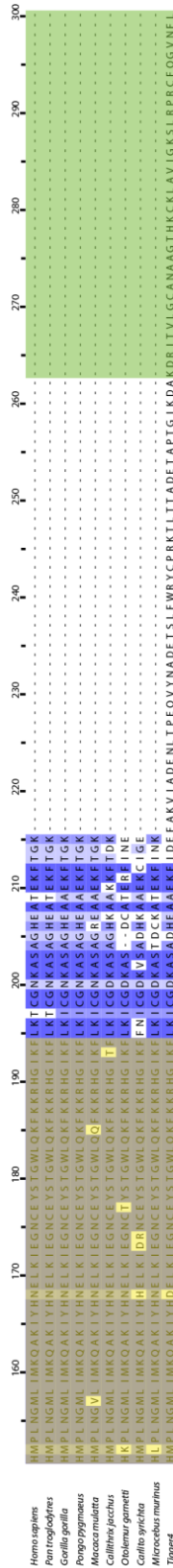
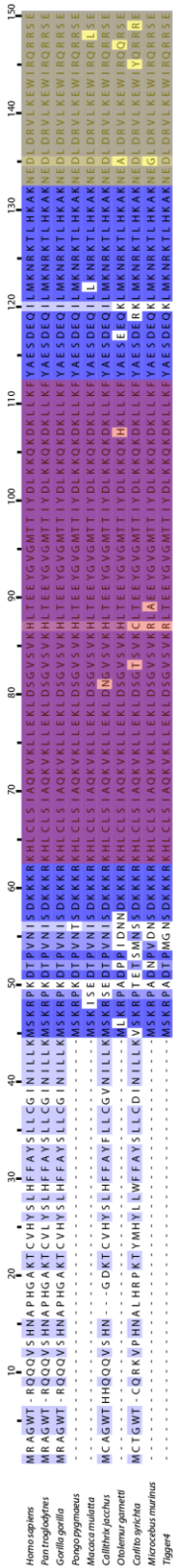
## Anexo II

*MGC12965* (AP001372.2) es otro candidato anotado como no codificante en GENCODEv24. A pesar de ello, tanto el ORF y la sintenia como los dominios amino-terminal de la proteína centromérica B (CENP-B) y de unión a ADN de la transposasa Tc5 están conservados en todos los primates (fig. A2 y A3). El alineamiento muestra numerosas inserciones y deleciones múltiples de tres, lo que indicaría una tendencia a mantener el ORF.

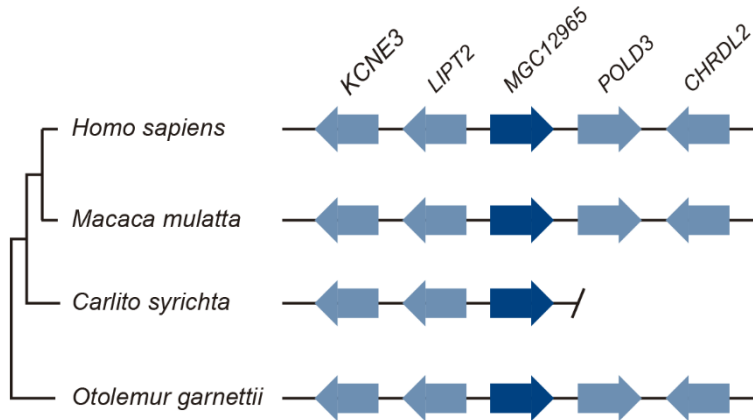
Uno de los primeros casos documentados de domesticación molecular fue el de *CENP-B*<sup>8</sup>. Este gen, domesticado en el origen de los mamíferos a partir de una transposasa *pogo*, se une a unas secuencias repetitivas centroméricas y permite una correcta segregación cromosómica<sup>9</sup>. Otros genes similares a *CENP-B* han resultado ser producto de la domesticación recurrente de transposasas *pogo* a lo largo de la evolución de los metazoos<sup>10,11</sup>.

El genoma humano, por ejemplo, codifica para al menos 9 genes relacionados con *CENP-B*, cuya función permanece prácticamente inexplorada: *JRK*, *JRKL* y *TIGD1-7*. *Tigger4*, el transposón del que deriva *MGC12965*, sólo se encuentra en primates.

Por lo tanto, *MGC12965* sería un nuevo miembro de la familia CENP-B específico de ese taxón, al que proponemos renombrar *TIGD8*.



**Figura A2.** Alineamiento de las secuencias aminoacídicas de MGC12965 de diversas especies de primates, y de *Tigge4*. Sombreado en rojo está el dominio CENP-B, en amarillo el de unión a ADN de transposasa Tc5 y en verde el dominio endonucleasa DDE.



**Figura A3.** Sintenia del locus de *MGC12965*. Debido a un ensamblaje genómico incompleto, no podemos determinar la sintenia *downstream* de *MGC12965* en el tarsero filipino (*Carlito syrichta*).

## Referencias de los anexos

1. Esnault, C., Cornelis, G., Heidmann, O. & Heidmann, T. Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the EnvV syncytin, captured for a function in placentation. *PLoS Genet.* **9**, e1003400 (2013).
2. Kjeldbjerg, A. L., Villesen, P., Aagaard, L. & Pedersen, F. S. Gene conversion and purifying selection of a placenta-specific ERV-V envelope gene during simian evolution. *BMC Evol. Biol.* **8**, 266 (2008).
3. Best, S., Tissier, P. Le, Towers, G. & Stoye, J. P. Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* **382**, 826–829 (1996).
4. Murcia, P. R., Arnaud, F. & Palmarini, M. The Transdominant Endogenous Retrovirus enJS56A1 Associates with and Blocks Intracellular Trafficking of Jaagsiekte Sheep Retrovirus Gag. *J. Virol.* **81**, 1762–1772 (2007).
5. Yap, M. W., Colbeck, E., Ellis, S. A. & Stoye, J. P. Evolution of the Retroviral Restriction Gene Fv1: Inhibition of Non-MLV Retroviruses. *PLoS Pathog.* **10**, e1003968 (2014).
6. Stremlau, M., Perron, M., Welikala, S. & Sodroski, J. Species-Specific Variation in the B30.2(SPRY) Domain of TRIM5 Determines the



- Potency of Human Immunodeficiency Virus Restriction. *J. Virol.* **79**, 3139–3145 (2005).
7. Renaud, S. J. *et al.* OVO-like 1 regulates progenitor cell fate in human trophoblast development. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6175-84 (2015).
  8. Delorme-Axford, E., Sadovsky, Y. & Coyne, C. B. The Placenta as a Barrier to Viral Infections. *Annu. Rev. Virol.* **1**, 133–146 (2014).
  9. Smit, A. F. & Riggs, A. D. Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1443–8 (1996).
  10. Fachinetti, D. *et al.* DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Dev. Cell* **33**, 314–327 (2015).
  11. Casola, C., Hucks, D. & Feschotte, C. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol. Biol. Evol.* **25**, 29–41 (2008).
  12. Mateo, L. & González, J. Pogo-like Transposases Have Been Repeatedly Domesticated into CENP-B-Related Proteins. *Genome Biol. Evol.* **6**, 2008–2016 (2014).







