



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Análisis de la regulación transcripcional del gen
RBFOX1 desde un enfoque evolutivo

Tesis presentada para optar al título de Doctora de la Universidad de
Buenos Aires en el área Ciencias Biológicas

Lic. Lara Berasain

Directora de Tesis: Dra. Lucia Franchini

Consejero de estudios: Dr. Marcelo Rubinstein

Instituto de Investigaciones en Ingeniería Genética y Biología Molecular
“Dr. Héctor N. Torres”

Buenos Aires, 2023

RESUMEN

Análisis de la regulación transcripcional del gen *RBFOX1* desde un enfoque evolutivo

Descifrar cómo los cambios en las redes regulatorias de la expresión génica contribuyen a los rasgos humano-específicos es uno de los grandes propósitos de la biología evolutiva humana. Identificar tales cambios y establecer su impacto funcional es un gran desafío, principalmente debido a las dificultades para predecir la función de las secuencias no codificantes. Sin embargo, recientemente se han identificado posibles impulsores genéticos subyacentes a las características biológicas humanas, entre los que se destacan las Regiones Aceleradas en Humanos (HARs). Estos elementos están altamente conservados en la mayoría de las especies de mamíferos, pero muestran un exceso significativo de cambios humano-específicos en sus secuencias. El alto grado de conservación sugiere que están sujetos a restricciones funcionales, mientras que el aumento de la tasa de sustituciones en humanos podría indicar la adquisición de una función diferencial en nuestra especie. Se ha demostrado que numerosos HARs actúan como *enhancers* transcripcionales durante el neurodesarrollo, y algunos muestran diferencias de expresión espaciotemporal en comparación con sus secuencias ortólogas.

A pesar de estos avances, la contribución de los HARs en la regulación génica *in vivo* sigue siendo poco conocida. En nuestro laboratorio hemos determinado que uno de los *locus* del genoma humano que acumula un alto número de secuencias HARs es el ocupado por el gen *RNA-binding Protein, fox-1 homolog (RBFOX1)*, un factor de *splicing* alternativo que actúa principalmente en células neuronales, músculo esquelético y corazón. Además, su disfunción ha sido asociada con varios trastornos del neurodesarrollo. En este trabajo empleamos ensayos de gen reportero en peces cebrá (*Danio rerio*) transgénicos para determinar si los *RBFOX1*-HARs son capaces de actuar como *enhancers* transcripcionales y comparamos su actividad con sus secuencias ortólogas de chimpancé. Con esta estrategia, descubrimos que 15 de los 19 *RBFOX1*-HARs son *enhancers* activos en peces cebrá y que 10 de estas regiones regulatorias exhiben una expresión diferencial en comparación con sus ortólogas de chimpancé.

Por otra parte, empleamos un enfoque de edición génica en ratones mediante CRISPR/Cas9 para eliminar, de manera independiente, dos HARs que resultaron de interés, y establecer el impacto funcional *in vivo* de cada delección. Mediante RT-qPCR observamos un aumento en la expresión de *Rbfox1* a los 14.5 días embrionarios en el mesencéfalo en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y a los 8 días postnatales en la corteza cerebral en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}. La secuenciación del ARNm reveló que un alto número de genes diferencialmente expresados en los ratones mutantes son genes regulados por *Rbfox1*. En cambio, casi no hubo cambios en el *splicing* alternativo, probablemente debido a una redundancia en las funciones de las proteínas de la familia Rbfox.

Nuestros hallazgos agregan a *RBFOX1* a la selecta lista de genes del neurodesarrollo que experimentaron evolución acelerada en el linaje humano con consecuencias en sus redes regulatorias de la expresión génica. Nuestro trabajo representa un paso más para llenar la brecha que existe entre los cambios genéticos humano-específicos y sus consecuencias fenotípicas.

Palabras clave: *RBFOX1*, regulación transcripcional, HARs, evolución del cerebro.

ABSTRACT

Transcriptional regulation analysis of *RBFOX1* gene from an evolutionary approach

A long-standing goal of evolutionary biology has been to decode how changes in gene regulatory networks contribute to human-specific traits. Identifying such changes and establishing their biological impact is a major challenge, mainly due to difficulties predicting the function of non-coding sequences. However, recent comparative studies have begun to reveal potential genetic drivers underlying human biological features, among which Human Accelerated Regions (HARs) stand out. These elements are highly conserved in most mammalian species but show a significant excess of human-specific sequence changes. The high degree of conservation of HARs suggests that HARs are subject to functional constraints, while the increased rate of substitutions in humans could indicate the acquisition of a differential function in our species. Numerous HARs have been shown to act as transcriptional enhancers during embryonic neurodevelopment, and some show changes in spatiotemporal expression compared to their orthologs sequences.

Despite these advances, the contribution of HARs in gene regulation *in vivo* remains poorly understood. In our laboratory we have determined that one of the locus of the human genome that accumulates a large number of HARs is the one occupied by the *RNA Binding Protein, fox-1 homolog (RBFOX1)* gene, a *trans*-acting factor that regulates alternative splicing mainly in neuronal cells, skeletal muscle, and heart. Besides, its dysfunction has been associated with various neurodevelopmental disorders. Here we employed transgenic reporter assays in zebrafish (*Danio rerio*) to determine if the *RBFOX1*-HARs are transcriptional enhancers and compare their activity with their chimpanzee orthologous sequences. With this approach, we uncovered that 15 out of 19 *RBFOX1*-HARs are active enhancers in zebrafish and that 10 of these regulatory regions exhibit differential expression compared to their chimpanzee orthologs.

Furthermore, we employed a genetic engineering approach in mice by CRISPR/Cas9 to independently remove two HARs of interest, and to establish the *in vivo* functional impact of each deletion. Using RT-qPCR, we found an increase in the expression of *Rbfox1* at 14.5 embryonic days in the midbrain in *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} mice and at 8 postnatal days in the cerebral cortex in *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} mice. RNA sequencing revealed that a high number of differentially expressed genes in the knockout mice are genes regulated by *Rbfox1*. In contrast, there were almost no changes in alternative splicing, probably due to a redundancy in the functions of the Rbfox family proteins.

Taken together, our findings adds *RBFOX1* to the limited list of genes involved in brain development that underwent accelerated evolution in the human lineage affecting its gene regulatory networks. Our work represents one step further to fill the gap between human-specific genetic changes and their phenotypic consequences.

Keywords: *RBFOX1*, transcriptional regulation, HARs, brain evolution.

AGRADECIMIENTOS

A mi familia, mamá, papá y Juli por el amor y la motivación.

A Manuel por el apoyo incondicional.

A la Dra. Lucía Franchini por haberme dado la oportunidad de formar parte de su grupo de investigación y acompañarme con gran dedicación durante todo el desarrollo del doctorado.

Al Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, toda su gente, becarios, investigadores, técnicos y personal de apoyo.

A la Universidad de Buenos Aires por la formación académica de excelencia.

Al CONICET y la Agencia por la financiación.

A los amigos que me dejó el doctorado: Lean, Ana, Mati, Sofi, Agus y Ale por la amistad y el apoyo que me brindaron siempre.

A la familia del 222 y del 128 por los almuerzos, los mates y las charlas sin sentido.

A Martita por la ayuda en la producción de ratones transgénicos.

A Paula por su aporte bioinformático imprescindible en este trabajo.

A los técnicos del acuario y del bioterio.

ÍNDICE

RESUMEN	2
ABSTRACT	3
AGRADECIMIENTOS	5
INTRODUCCIÓN	9
La historia de nuestra especie	9
Evolución del cerebro humano	11
¿Qué nos hace humanos?	12
<i>La hipótesis regulatoria</i>	13
<i>Validando a King y Wilson</i>	15
Regulación transcripcional	16
<i>Promotores</i>	17
<i>Enhancers y silenciadores</i>	18
<i>Aisladores</i>	19
<i>Predicción de enhancers</i>	19
<i>Dominios topológicamente asociados</i>	20
Evolución acelerada en el linaje humano	21
<i>Caracterización funcional de los HARs</i>	22
<i>RBFOX1</i> , un gen candidato	24
<i>Estructura génica y regulación transcripcional de RBFOX1</i>	25
<i>Impacto funcional de la pérdida de Rbfox1 en modelos murinos</i>	27
HIPÓTESIS Y OBJETIVOS	28
Hipótesis general	28
Objetivos generales	28
CAPÍTULO I	29
Hipótesis de trabajo	29
Objetivos específicos	29
Resultados	29
<i>Distribución de los HARs en RBFOX1</i>	29
<i>Análisis funcional de los RBFOX1-HARs en peces cebra transgénicos</i>	31
<i>Análisis de las sustituciones humano-específicas</i>	38
<i>Predicción de sitios de unión a factores de transcripción</i>	41
<i>Selección positiva codificante y no codificante</i>	44
<i>Caracterización de los patrones de expresión de los RBFOX1-HARs</i>	46
<i>Análisis de los promotores de RBFOX1 en peces cebra transgénicos</i>	50
Discusión de los resultados	54

<i>Ensayos funcionales en peces cebra transgénicos</i>	54
<i>Elección del modelo experimental</i>	55
<i>Necesidad de una tercera especie</i>	56
<i>Comparación con ensayos de MPRA</i>	57
<i>Análisis de los RBFOX1-HARs en humanos arcaicos</i>	59
<i>Factores de transcripción candidatos</i>	60
<i>Evolución codificante y no codificante</i>	60
<i>Promotor mínimo de RBFOX1</i>	61
CAPÍTULO II	62
Hipótesis de trabajo	62
Objetivos específicos	62
Resultados	62
<i>Generación de ratones genéticamente modificados</i>	62
<i>Análisis de los niveles de expresión de Rbfox1</i>	64
<i>Análisis de genes diferencialmente expresados</i>	67
<i>Análisis del uso diferencial de exones</i>	75
Discusión de los resultados	78
<i>Diferencias entre técnicas: RT-qPCR vs RNA-Seq</i>	78
<i>¿Es 2xHAR.321 un silenciador?</i>	79
<i>Genes regulados por Rbfox1</i>	81
<i>Comparación con otros ratones con deleciones en enhancers</i>	81
CONCLUSIONES	83
Principales resultados	83
Implicancias de los hallazgos	84
MATERIALES Y MÉTODOS	85
Análisis bioinformáticos	85
<i>Identificación de las regiones aceleradas en el linaje humano</i>	85
<i>Predicción de sitios de unión a factores de transcripción</i>	85
<i>Selección positiva en las secuencias no codificantes</i>	85
<i>Selección positiva en las secuencias codificantes</i>	86
Producción de peces cebra transgénicos	86
<i>Construcción de los transgenes</i>	86
<i>Transcripción in vitro del ARNm de la transposasa Tol2</i>	87
<i>Animales y acuario</i>	88
<i>Microinyección de embriones de peces cebra</i>	88
<i>Análisis del patrón de expresión de la proteína reportera EGFP</i>	89
Edición del genoma con CRISPR/Cas9 en ratones	90
<i>Diseño de las ARN guías</i>	90
<i>Transcripción in vitro del ARNm de la Cas9</i>	90

<i>Animales y bioterio</i>	90
<i>Microinyección pronuclear</i>	91
<i>Generación de la colonia de ratones modificados genéticamente</i>	92
Análisis de expresión de los ratones modificados genéticamente	92
<i>Obtención de embriones y congelado de muestras</i>	92
<i>Extracción de ARN y síntesis de ADN complementario</i>	93
<i>PCR cuantitativa en tiempo real</i>	93
<i>Secuenciación y análisis de ARN</i>	93
Histoquímica	94
<i>Congelado de muestras para cortes en criostato</i>	94
<i>Doble inmunohistoquímica de cortes histológicos de pez cebra</i>	95
Hibridación <i>in situ</i>	95
<i>Síntesis de las sondas de <i>rbfox1</i> para pez cebra</i>	95
<i>Síntesis de las sondas de <i>Rbfox1</i> para ratón</i>	96
<i>Hibridación in situ en cortes histológicos de pez cebra y ratón</i>	96
<i>Hibridación in situ en embriones enteros de pez cebra</i>	97
APÉNDICE I - Secuencia de los <i>primers</i> específicos	98
APÉNDICE II – Líneas de peces cebra transgénicos	100
APÉNDICE III – Genes diferencialmente expresados	104
APÉNDICE IV – Uso diferencial de exones	121
BIBLIOGRAFÍA	124

INTRODUCCIÓN

La historia de nuestra especie

La información disponible actualmente del registro fósil, junto con estudios genéticos, indican que los humanos modernos (*Homo sapiens*) y los chimpancés (*Pan troglodytes*) compartieron un último ancestro en común hace aproximadamente 5-7 millones de años (Ma) en África (Chen and Li, 2001; Brunet et al., 2002). Si bien *H. sapiens* es el único representante vivo del linaje *Hominina*, existió una gran diversidad de especies, muchas de las cuales coexistieron durante largos periodos de tiempo (Figura I) (Wood, 1996). La evolución de los rasgos modernos de *H. sapiens* no fue un proceso gradual, lineal y aditivo (Carroll, 2003; Tattersall, 2010). Por lo contrario, fue un patrón de evolución en mosaico, donde las diferentes estructuras morfológicas, como por ejemplo el sistema dentario, el sistema locomotor y el sistema nervioso, cambiaron a tasas diferentes y en momentos distintos (McHenry, 1975). Las relaciones de parentesco, orden cronológico, magnitud y ritmo de la evolución de los caracteres del linaje humano se encuentran constantemente sujetos a revisión a medida que ocurren nuevos hallazgos fósiles (Tattersall, 2010).

Los homínidos más antiguos que se conocen en la actualidad son *Sahelanthropus tchadensis* de Chad (Brunet et al., 2005) y *Orrorin tugenensis* de Kenia (Pickford and Senut, 2001), fechados en 7 Ma y 5-6 Ma atrás, respectivamente. Su tamaño cerebral, de aproximadamente 350 cm³, se encuentra dentro del rango que se observa en los chimpancés. Sin embargo, la posición y orientación del foramen magnum sugiere que *Sahelanthropus* presentaba signos de bipedalismo (Zollikofer et al., 2005), aunque esta afirmación continúa siendo materia de debate. Alrededor de 4 Ma atrás aparecieron los primeros miembros del género *Australopithecus*, homínidos bípedos facultativos (Johanson and White, 1980), cuyo cráneo tiene un tamaño que oscila entre 390 y 515 cm³, similar a los chimpancés y gorilas (Falk et al., 2000). Los primeros representantes del género *Homo*, homínidos bípedos que presentaban un aumento en el tamaño corporal, una mandíbula más pequeña con una reducción de las piezas dentarias y un incremento significativo del tamaño endocraneal, de entre 650 y 800 cm³ (Leakey et al., 1965), surgieron hace aproximadamente 2.5 Ma. Los miembros más antiguos del género, como *H. rudolfensis* y *H. habilis* eran relativamente similares a los australopitecinos gráciles, pero por primera vez fueron asociados con herramientas simples de piedra (Blumenschine et al., 2003). De hecho, la presencia de tecnología lítica es lo que definió históricamente a este nuevo género (Wood and Collard, 1999). Grupos de *Homo* tempranos dieron origen a otra especie de *Homo* en África, de cuerpo y cerebro más grande, *H. ergaster* (Wood, 1996). De acuerdo con el registro fósil, *H. ergaster* producía una variedad de herramientas de piedra que utilizaba para cortar carne y extraer médula ósea de los huesos. Además, probablemente aprendió a controlar el fuego y fue capaz de realizar artefactos para contener agua y alimentos. Equipado con estas nuevas habilidades *H. ergaster* fue el primer homínido en migrar fuera de África alrededor de 1.8 Ma atrás hacia Asia e Indonesia.

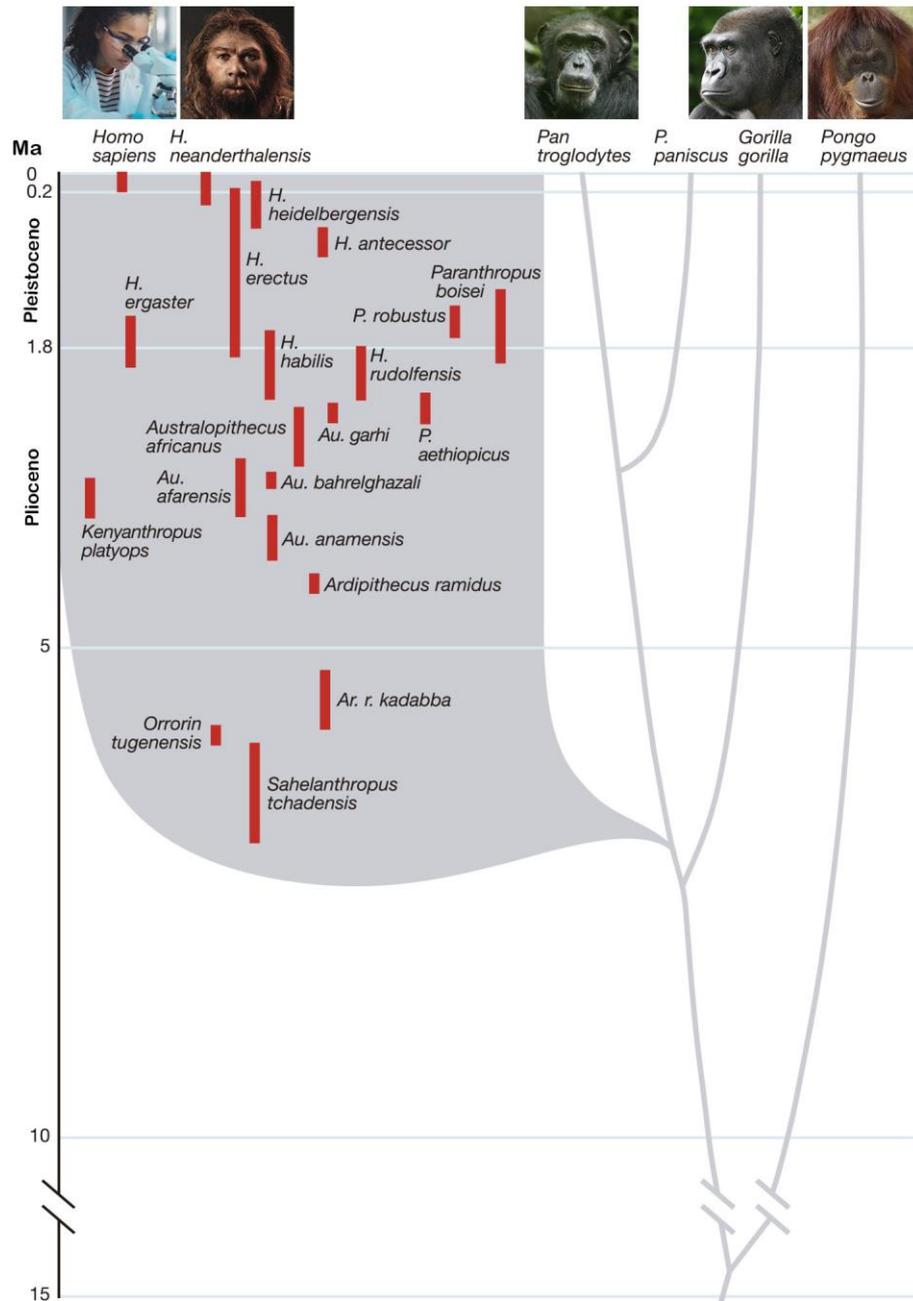


FIGURA I: Escala temporal y filogenia de los homínidos. Los tiempos aproximados de divergencia derivan de datos moleculares. Las relaciones filogenéticas entre homínidos son inciertas y se marcan como un sombreado gris. Las barras rojas denotan los intervalos de tiempo de ocurrencia de cada especie fósil. Modificada de (Carroll, 2003).

Fuera de África, *H. ergaster* dio origen a la especie que hoy reconocemos como *H. erectus* (Wood, 1996). La expansión global de *H. erectus* sugiere que eran ecológicamente flexibles, y que tenían las capacidades cognitivas para adaptarse y prosperar en entornos muy diferentes (Antón, 2003). Aquí se comienza a ver un gran aumento en el tamaño del cerebro, de entre 850 a 1100 cm³. Alrededor de 700 mil años atrás en África, *H. erectus* dio lugar a *H. heidelbergensis*, una especie muy similar a los humanos anatómicamente modernos en términos de proporciones corporales, adaptaciones dentales y capacidades cognitivas. *H. heidelbergensis*, a menudo referido como “humano

arcaico”, era un cazador activo que producía herramientas sofisticadas (Rightmire, 2009). Se cree que *H. neanderthalensis*, los homínidos adaptados al frío de comportamiento complejo y cerebros de tamaño similar al humano moderno, surgieron a partir de poblaciones de *H. heidelbergensis* en Europa hace al menos 250 mil años atrás (Hublin, 2009; Rightmire, 2009). Además, los estudios de ADN antiguo de fósiles de *H. neanderthalensis* sugieren que nuestro linaje podría haberse cruzado ocasionalmente con ellos (Green et al., 2010). La evidencia de fósiles y ADN sugiere que los humanos anatómicamente modernos, *H. sapiens*, evolucionaron en África hace 200 mil años, probablemente a partir de *H. heidelbergensis* (Relethford, 2008; Rightmire, 2009). La sofisticación conductual de *H. sapiens*, reflejada en el gran tamaño cerebral de 1400 cm³ y la evidencia arqueológica de un conjunto de herramientas más sofisticadas y técnicas de caza inteligentes, permitió que prosperara y se expandiera en el continente africano. Hace 100 mil años se desplazó hacia Eurasia, y eventualmente se expandió por todo el globo hacia Australia y las Américas (DeGiorgio et al., 2009).

Evolución del cerebro humano

El cerebro es quizás el órgano más emblemático del ser humano, pero no escapa a las reglas de la evolución y como tal presenta una neocorteza de seis capas típica de un cerebro mamífero, y las características típicas de un cerebro primate (Striedter, 2006), incluyendo una corteza visual bien desarrollada y una corteza prefrontal (Preuss, 2007; Kaas, 2013). En particular, la corteza prefrontal se ha relacionado con capacidades cognitivas de orden superior, como la toma de decisiones, la memoria de trabajo y la ejecución de funciones. Aunque los mamíferos no primates también tienen una corteza prefrontal, esta está compuesta únicamente por dos regiones, mientras que los primates presentan una región adicional denominada corteza prefrontal dorsolateral (Preuss, 1995).

Aunque el cerebro humano presenta una estructura típicamente primate, releva algunas características que lo distinguen y lo individualizan plenamente. En primer lugar, tienen el mayor número de neuronas entre los primates, 86 mil millones, en comparación con el número de neuronas de los chimpancés y gorilas, estimado en 28 y 33 mil millones, respectivamente (Azevedo et al., 2009). Además, las estimaciones de la cantidad de neuronas de las diferentes especies de homínidos extintos sugieren que, con 27 a 35 mil millones, el número de neuronas de los australopitecinos tempranos se encontraría dentro del rango de los grandes simios actuales. Para los *Homo* tempranos, se predice un número de neuronas de 50 a 60 mil millones, y de 62 mil millones en promedio para *H. erectus*. En cambio, para las especies de humanos arcaicos, *H. heidelbergensis* y *H. neanderthalensis* se predice un número de neuronas que caería dentro del rango de los humanos modernos (Herculano-Houzel, 2012). Si consideramos que el factor determinante para el desarrollo de las capacidades cognitivas depende principalmente del número de neuronas, estos resultados sugieren que el potencial cognitivo de los humanos arcaicos podría haber sido comparable a las halladas en los humanos anatómicamente modernos. En segundo lugar, el cerebro humano es aproximadamente tres veces más grande que el de los chimpancés (Stephan et al., 1981) y, aunque por mucho tiempo se utilizó al tamaño absoluto como predictor de las habilidades cognitivas, esto no resulta

así para todas las comparaciones, ya que muchas especies de grandes mamíferos, como cetáceos y elefantes, tienen cerebros de mayor tamaño que el humano (Roth and Dicke, 2005). Sin embargo, aunque el cerebro del elefante tiene cerca de 251 mil millones de neuronas, tan sólo 5.6 mil millones son corticales, es decir solo el 2.2%. En contraste, el 20.9% de las neuronas presentes en el cerebro humano son corticales, más del 10% que cualquier otro mamífero (Herculano-Houzel, 2012). Por último, algunos autores plantean que la singularidad del cerebro humano se encuentra ligada a factores más complejos que el aumento del tamaño cerebral y el número de neuronas, y que, en cambio, estarían relacionados con la organización y conectividad del cerebro (Rilling et al., 2008), la diversidad de tipos neuronales (Sherwood and Hof, 2007), el metabolismo neuronal y la plasticidad sináptica (Sherwood et al., 2008).

¿Qué nos hace humanos?

En un sentido estrictamente biológico podemos definir a la “humanidad” como el conjunto de características moleculares, celulares y comportamentales que nos distinguen de nuestros parientes primates más cercanos, los chimpancés. Algunos de estos rasgos son evidentes, como el aumento del tamaño de la corteza cerebral y los cambios en la estructura pélvica y en las extremidades que nos permitieron caminar erguidos (Falk et al., 1992; Antón et al., 2014). Otras diferencias son más difíciles de ver y de interpretar, y es probable que existan más que todavía no han sido descubiertas. Además, muchas características son sumamente complejas, por ejemplo, la evolución de una corteza más grande requirió de cambios en múltiples procesos durante el desarrollo cortical, incluida la proliferación y migración neuronal, el desarrollo axonal y posiblemente la aparición de nuevos tipos celulares (Geschwind and Rakic, 2013). Al mismo tiempo, nuestra comprensión del desarrollo neocortical y la morfogénesis de las extremidades todavía es incompleta.

A simple vista, identificar los orígenes evolutivos de los fenotipos humanos parece desalentador, sin embargo, a pesar de su complejidad biológica, todas las características humanas están, en última instancia, codificadas en un número finito de cambios genéticos que surgieron en el linaje humano desde la divergencia de los humanos con los chimpancés hace 5 Ma (Varki and Altheide, 2005; Bauernfeind et al., 2015). Las comparaciones del genoma humano con otros genomas de primates y vertebrados han revelado la ubicación de millones de cambios genéticos humano-específicos (Bailey and Eichler, 2006; Scally et al., 2012). Dichos cambios se suelen catalogar en tres clases generales: *a*) grandes alteraciones cromosómicas, *b*) inserciones y deleciones pequeñas y, *c*) sustituciones puntuales (Vallender et al., 2008; Franchini and Pollard, 2015).

Por un lado, las grandes alteraciones cromosómicas incluyen deleciones, duplicaciones y reordenamientos genómicos; son relativamente raras, pero como abarcan varios *loci* suelen tener un impacto mayor. Las primeras diferencias estructurales humano-específicas se analizaron mediante técnicas de tinción de las bandas de la cromatina, donde descubrieron la fusión de dos cromosomas ancestrales que dieron lugar al cromosoma 2, así como también bandas de heterocromatina constitutiva en los cromosomas 1, 9, 16 e Y, e inversiones pericéntricas en los cromosomas 1 y 18 (Yunis

and Prakash, 1982). En la actualidad se conocen más de 60 duplicaciones segmentarias humano-específicas (Goidts et al., 2006) y 152 genes con cambios en el número de copias (Fortna et al., 2004; Armengol et al., 2010). Muchas de estas variaciones estructurales han alterado la expresión génica de diversos genes del desarrollo (para una revisión ver (Franchini, 2021)). Por ejemplo, la inversión pericéntrica en el cromosoma 1 resultó en un aumento en el número de copias de varios genes del desarrollo, incluyendo *HYDIN* (*HYDIN axonemal central pair apparatus protein*) (Doggett et al., 2006), *SRGAP2* (*SLIT-ROBO Rho GTPase activating protein 2*) (Dennis et al., 2012) y genes que contienen el dominio proteico DUF1220, como la familia de genes NBPF (*Neuroblastoma breaking point family*) (Dumas and Sikela, 2009). Además, una variante estructural en el cromosoma 15q13-q14 resultó en la duplicación de varios genes, incluyendo *ARHGAP11B* (*Rho GTPase activating protein 11B*), una copia parcial de *ARHGAP11A* (*Rho GTPase activating protein 11A*) (Antonacci et al., 2014).

Por otro lado, las inserciones y deleciones pequeñas, también denominadas *indels*, abarcan aproximadamente 3.5% del genoma humano (Britten, 2002; Varki and Altheide, 2005) e involucran regiones genómicas de menos de 20 kpb. Los *indels* pueden tener efectos funcionales críticos, afectando regiones codificantes, variantes de *splicing* y marcos de lectura, o modificando regiones regulatorias conservadas. Por último, las sustituciones puntuales abarcan 1.2% del genoma humano, lo que equivale aproximadamente a 30 millones de sustituciones, la mayoría de las cuales se encuentran en regiones no codificantes (Varki and Altheide, 2005). El desafío actual consiste en determinar cuáles de estos cambios son relevantes, qué funciones moleculares codifican y, en última instancia, qué fenotipos produjeron.

La hipótesis regulatoria

La tarea de interpretar los cambios en las secuencias humano-específicas debe comenzar con una hipótesis que especifique las funciones moleculares que dichos cambios potencialmente alteraron durante la evolución humana. Entre los principales mecanismos genéticos se encuentran: *a*) la selección positiva en regiones codificantes del genoma que conducen a cambios en las secuencias de aminoácidos de las proteínas existentes, *b*) la ganancia de genes completamente nuevos y la pérdida de genes ancestrales en el linaje humano, y *c*) los cambios en regiones no codificantes del genoma, especialmente en las secuencias regulatorias en *cis* que dan lugar a la expresión de genes alterados (Sikela, 2006; O'Bleness et al., 2012).

Inicialmente, los trabajos que tenían como objetivo develar esos cambios se centraban en los primeros dos mecanismos, ya que los efectos funcionales de las sustituciones no sinónimas, o la ganancia y pérdida de genes son más fáciles de interpretar que las consecuencias de las sustituciones en las regiones no codificantes. De este modo pudieron detectarse varios genes del neurodesarrollo con evidencias de selección positiva en humanos (Sikela, 2006; O'Bleness et al., 2012). Algunos ejemplos incluyen al factor de transcripción *FOXP2* (*Forkhead box P2*), que ha sido asociado con el desarrollo del habla (Lai et al., 2001); *FOXP2* presenta dos sustituciones de aminoácidos con signos de selección positiva que se fijaron en el linaje humano luego de la divergencia con el linaje

de los chimpancés (Enard et al., 2002; Zhang et al., 2002). Otro ejemplo interesante es *WDR62* (*WD repeat domain 62*), una proteína asociada al centrosoma que se expresa en precursores neuronales y cuya ausencia está asociada con la microcefalia primaria (Nicholas et al., 2010; Yu et al., 2010); *WDR62* tiene seis aminoácidos con signos de selección positiva que fueron fijados en el linaje humano antes de la divergencia con los humanos arcaicos (Pervaiz and Abbasi, 2016).

A medida que estuvieron disponibles más genomas de vertebrados y se desarrollaron las herramientas bioinformáticas necesarias, fue posible implementar los modelos de evolución en las secuencias no codificantes. Durante mucho tiempo se había hipotetizado que los cambios en la regulación de la expresión génica contribuyen a las innovaciones evolutivas. Britten y Davidson propusieron por primera vez que las diferencias fenotípicas entre organismos podrían ser impulsadas principalmente por cambios regulatorios (Britten and Davidson, 1969, 1971). Poco tiempo después, King y Wilson ampliaron específicamente esta hipótesis para explicar las diferencias biológicas entre humanos y chimpancés (King and Wilson, 1975). En su clásico trabajo titulado “*Evolución a dos niveles en humanos y chimpancé*” describen por primera vez la gran similitud entre las proteínas de chimpancés y humanos mediante diversas técnicas bioquímicas. Sus hallazgos los llevaron a postular que el pequeño grado de divergencia molecular observado en las secuencias proteicas no podían explicar las grandes diferencias anatómicas y comportamentales observadas entre ambas especies. En cambio, especularon que los cambios evolutivos se basan más frecuentemente en modificaciones en los mecanismos que controlan la expresión génica, y que las mutaciones en las regiones regulatorias podrían explicar las principales diferencias biológicas entre humanos y chimpancés.

Evolutivamente, esta es una hipótesis atractiva porque los cambios en las regiones regulatorias permiten modificaciones significativas en la expresión génica sin alterar la función de los genes esenciales en procesos del desarrollo que están profundamente conservados. Por ejemplo, las innovaciones evolutivas en el cerebro adulto se originan a partir de cambios en la forma en que se desarrolla en el embrión (Rakic, 2009; Geschwind and Rakic, 2013). Es probable que tales cambios en el desarrollo impliquen diferencias en cuándo, dónde y cómo se expresa un conjunto de genes conservados (Carroll, 2005; Wray, 2007; Wittkopp and Kalay, 2011). Muchos de los genes necesarios para el desarrollo embrionario tienen efectos pleiotrópicos y, en consecuencia, las mutaciones en la secuencia codificante que alteran las proteínas afectan a múltiples tejidos (Figura II *a*). Por lo contrario, se sabe que la maquinaria regulatoria de muchos genes del desarrollo es modular, en el sentido de que el patrón completo de expresión génica está determinado por elementos regulatorios discretos que actúan en tejidos específicos y en momentos específicos (Visel et al., 2009). Los cambios en estos elementos regulatorios pueden alterar la expresión génica en una sola etapa o región durante el desarrollo (Figura II *b*). Por lo tanto, es probable que los cambios regulatorios produzcan modificaciones evolutivas en los programas de desarrollo sin interrumpirlos por completo.

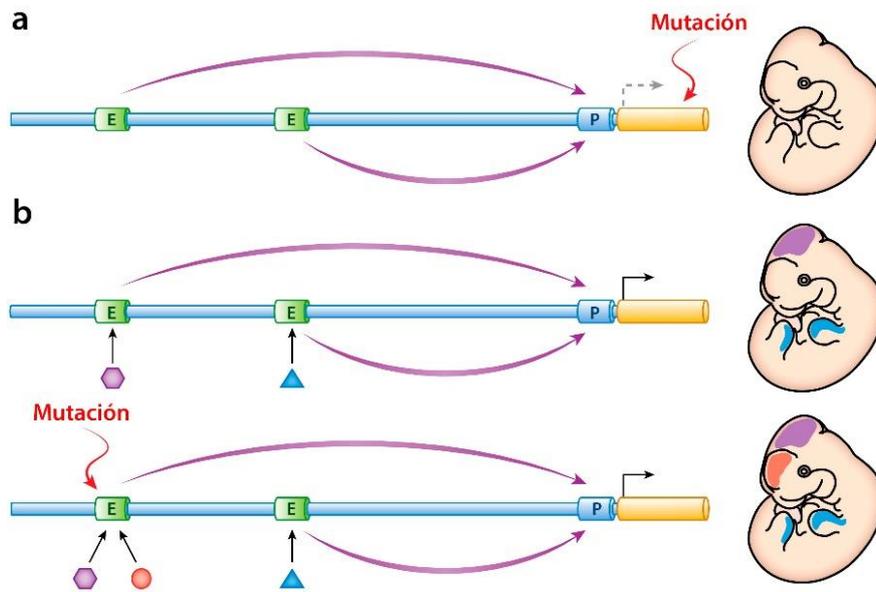


FIGURA II: Regulación génica como impulsor de la innovación evolutiva humana. (a) Las mutaciones en las secuencias codificantes de los genes pleiotrópicos tienen efectos generalizados. (b) Es probable que las mutaciones en las secuencias regulatorias tengan efectos tejido-específicos. En este simple ejemplo, una mutación convierte un *enhancer* del mesencéfalo en un *enhancer* con actividad tanto en el mesencéfalo como en el prosencéfalo, modificando la expresión del gen blanco en el cerebro sin afectar su expresión en las extremidades. Modificada de (Reilly and Noonan, 2016).

Validando a King y Wilson

Tuvieron que pasar varios años para comenzar a poner a prueba la hipótesis de King y Wilson, principalmente porque no fue hasta al menos 20 años después que llegó la era de la secuenciación de genomas. Este hito científico permitió la disponibilidad de secuencias genómicas de alta cobertura de humanos (Lander et al., 2001; The 1000 Genomes Project Consortium et al., 2015), primates (Chimpanzee Sequencing and Analysis Consortium, 2005; Gibbs et al., 2007), incluso homínidos extintos (Green et al., 2010; Prüfer et al., 2014), y muchos otros vertebrados, impulsando el descubrimiento de secuencias de ADN humano-específicas. Además, el desarrollo de las herramientas bioinformáticas necesarias para el análisis de dichas secuencias, así como también la aparición de enfoques transgénicos de alto rendimiento permitieron estudiar las secuencias genómicas funcionalmente relevantes (Visel et al., 2007; Hawkins et al., 2010).

El desafío actual, por ende, ha sido determinar cuáles de las sustituciones humano-específicas en las regiones no codificantes son las responsables de las características únicas de la biología del *H. sapiens*. Esta tarea resulta particularmente difícil por diversas razones. En primer lugar, el genoma no codificante es vasto, lo cual requiere de métodos que logren identificar y priorizar las mutaciones relevantes. Además, es difícil predecir las consecuencias moleculares, celulares y organizacionales de las mutaciones regulatorias, ya que no se conoce un código regulatorio que defina cómo la secuencia determina la función de los elementos regulatorios (Rubinstein and de Souza, 2013). Esto deja una ambigüedad mucho mayor para interpretar las tasas de cambio en estas regiones.

En segundo lugar, la mayoría de los rasgos distintivos de nuestra especie son sumamente complejos y, sin lugar a duda, son codificados por una combinación de mutaciones en diferentes *loci* genómicos. No obstante, avances recientes en la búsqueda de posibles impulsores genéticos han abierto un nuevo campo de genómica regulatoria comparada, permitiendo estudios experimentales de la evolución regulatoria en humanos.

Regulación transcripcional

El desarrollo de organismos complejos con muchos tipos celulares morfológica y funcionalmente diversos está determinado en gran medida por la información genética contenida en el ADN. Esta información incluye secuencias de genes que codifican proteínas y elementos regulatorios no codificantes que gobiernan cuándo, dónde y en qué nivel se expresa cada gen. Una expresión génica regulada es esencial para la integridad de todas las células y organismos eucariotas. La expresión de los genes se puede regular en varios pasos, incluidos el inicio y elongación de la transcripción, y el procesamiento, transporte, traducción y estabilidad del ARN mensajero (ARNm) (Maston et al., 2006). Sin embargo, la mayor parte de la regulación se produce a nivel de la iniciación de la transcripción. Normalmente existen dos familias distintas de elementos regulatorios que actúan en *cis*: *a*) un promotor, que se compone de un promotor mínimo y elementos regulatorios proximales, y *b*) elementos regulatorios distales, que pueden ser activadores (o *enhancers*), silenciadores y aisladores (o *insulators*) (Figura III) (Noonan and McCallion, 2010).

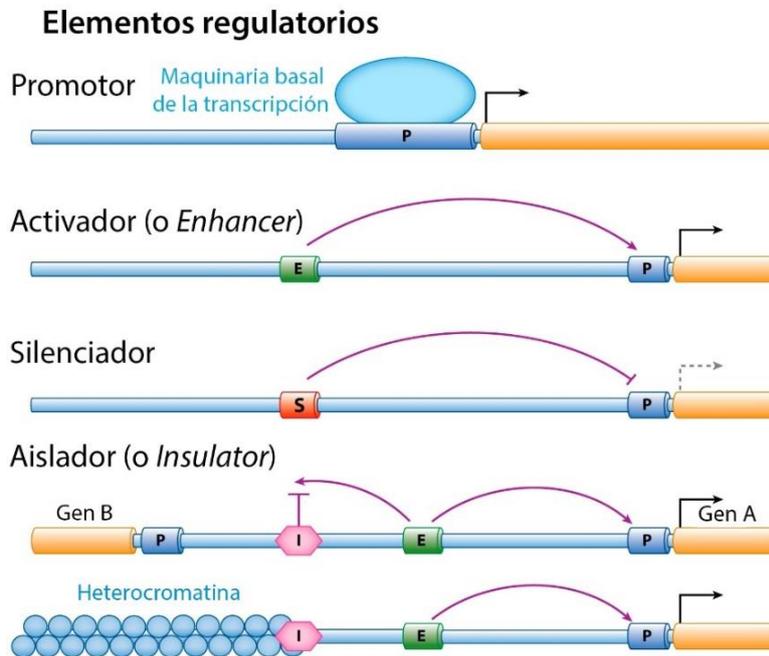


FIGURA III: Clasificación de los elementos regulatorios según su función. El promotor (P) es el sitio de acoplamiento para la maquinaria transcripcional basal. Los *enhancers* (E) y los silenciadores (S) activan o reprimen la transcripción por medio de la interacción con los promotores. Los *insulators* (I) bloquean la interacción *enhancer*-promotor, o actúan como barrera regulando el estado de la heterocromatina. Modificada de (Noonan and McCallion, 2010).

Promotores

El promotor es la región al comienzo de un gen que sirve como sitio de acoplamiento para la maquinaria transcripcional y el complejo de preiniciación, que consta de la ARN Polimerasa II y factores de transcripción generales (Noonan and McCallion, 2010). Sobre la base de diferentes propiedades que caracterizan a los promotores, que incluyen la distribución del sitio de inicio de la transcripción (TSS), los motivos en la secuencia, la configuración de la cromatina y la función del gen, se han propuesto tres tipos principales de promotores (Figura IV) (Haberle and Stark, 2018). En primer lugar, están los que presentan un solo TSS bien definido, motivos de ADN tipo caja TATA (*TATA box*), iniciador (Inr) y nucleosomas colocados de manera imprecisa con marcas de trimetilación de la lisina 4 de la histona H3 (H3K4me3) y acetilación de la lisina 27 de la histona H3 (H3K27ac). Se asocian preferentemente con genes altamente específicos con patrones de expresión restringidos y tienden a ser activos en células diferenciadas de tejidos adultos. En segundo lugar, se encuentran los promotores con múltiples TSSs estrechamente espaciados que se usan con una frecuencia similar, nucleosomas precisamente posicionados con marcas de H3K4me3 y H3K27ac que flanquean la región promotora. Suelen carecer de motivos definidos, pero se superponen con las islas CpG (CGI), regiones con un contenido elevado de GC y una alta densidad de dinucleótidos CpG (Carninci et al., 2006). Típicamente, esta clase de promotores es característico de los genes *housekeeping*. En tercer lugar, los promotores de los factores de transcripción del desarrollo se asemejan a los promotores de los genes *housekeeping*, pero están marcados de manera bivalente tanto con la marca de H3K4me3 como con la trimetilación de la lisina 27 de la histona H3 (H3K27me3), una modificación de la cromatina represiva. Presumiblemente, esto los prepara para la activación en el linaje celular correcto y para el silenciamiento en todas las demás células (Akalin et al., 2009).

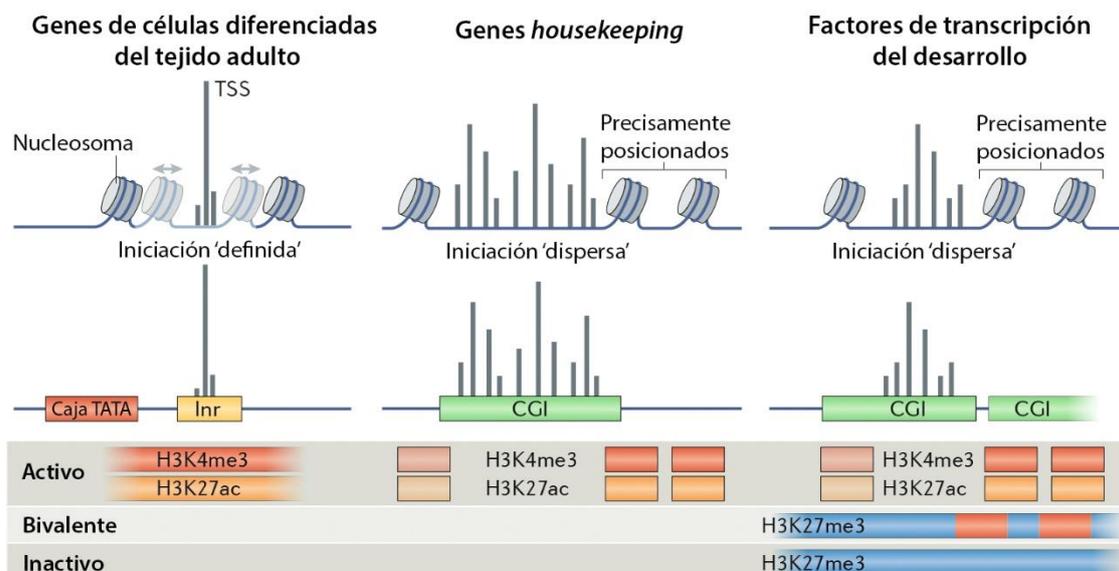


FIGURA IV: Tipos de promotores. Clasificación de los promotores según la distribución del inicio de la transcripción, los motivos en la secuencia, la configuración de la cromatina y la función del gen. Modificada de (Haberle and Stark, 2018).

La actividad promotora de una secuencia se suele medir en ensayos de gen reportero, ya sea evaluando la capacidad de impulsar la transcripción de manera autónoma, o en respuesta a la activación de la presencia de un *enhancer* (Figura V). En el primer caso la evaluación se realiza utilizando una construcción reportera que suele contener tanto el promotor mínimo como elementos regulatorios proximales, mientras que en el segundo caso la construcción sólo contiene el promotor mínimo (Haberle and Stark, 2018).

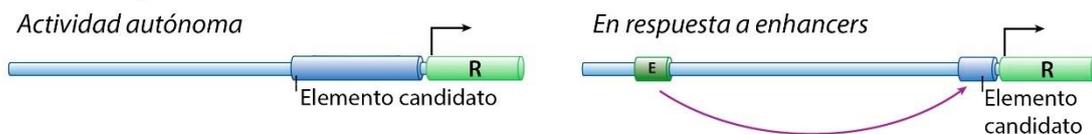
Enhancers y silenciadores

Los activadores (o *enhancers*) y los silenciadores son elementos regulatorios que independientemente de su orientación con respecto al promotor son capaces de activar o reprimir la transcripción, respectivamente (Noonan and McCallion, 2010). Además, pueden actuar a distancias variables, que se extienden desde unos pocos cientos de pares de bases hasta más de una mega base. Típicamente están conformados por secuencias de ADN que presentan sitios de unión para factores de transcripción que, al unirse al ADN cambian su conformación y son capaces de activar o reprimir la transcripción mediante contacto directo o indirecto con la maquinaria transcripcional.

La capacidad de un elemento de actuar como *enhancer* o silenciador se mide típicamente en ensayos reporteros (Figura V) (Arnone et al., 2004). Para evaluar la actividad *enhancer* se coloca el elemento candidato en una construcción compuesta por un promotor mínimo (necesario, pero no suficiente) y un gen reportero. Se considera que el elemento candidato es un *enhancer* cuando es capaz de dirigir la expresión de la proteína reportera de manera específica. En cambio, la inclusión del elemento candidato en una construcción reportera que ya tiene una expresión robusta, convierte el ensayo en una evaluación de la actividad silenciadora. Si el elemento candidato es capaz de disminuir la expresión de la proteína reportera, se considera silenciador.

Ensayos reporteros

Para evaluar promotores



Para evaluar elementos distales

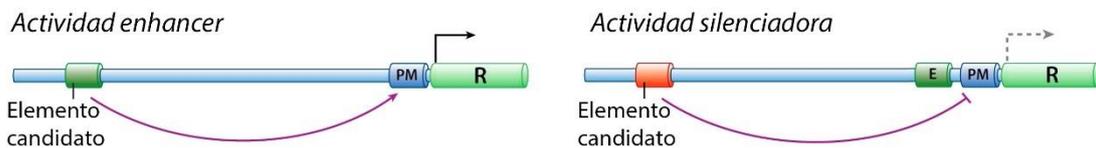


FIGURA V: Ensayos reporteros para la evaluación de elementos regulatorios. La actividad promotora se mide evaluando la capacidad de impulsar la transcripción del gen reportero de manera autónoma, o en respuesta a la entrada de *enhancers*. La actividad *enhancer* o silenciadora se mide evaluando la capacidad de impulsar o reprimir la transcripción del gen reportero, respectivamente. En el caso de los *enhancers*, se utiliza un promotor mínimo, que es necesario, pero no suficiente para iniciar la transcripción, mientras que para los silenciadores se utiliza un *enhancer* y un promotor mínimo con expresión robusta. Referencias: R, gen reportero. PM, promotor mínimo. E, *enhancer*.

Recientemente, se ha demostrado que los silenciadores tejido-específicos puede ser elementos regulatorios en *cis* bifuncionales que también actúan como *enhancers* en diferentes tipos celulares (para una revisión ver (Segert et al., 2021)). Los silenciadores se han tratado durante mucho tiempo como una clase de elementos regulatorios separada de los *enhancers*. Sin embargo, el conjunto actual de evidencias sugiere que la bifuncionalidad es sorprendentemente prevalente, donde más de la mitad de los silenciadores identificados muestran evidencias de actividad *enhancer* en otros estudios. Además, algunos factores de transcripción pueden actuar como activadores y represores en diferentes contextos, y esto podría contribuir a la bifuncionalidad de los elementos regulatorios. En el caso más simple, la bifuncionalidad de un elemento regulatorio podría reflejar la expresión diferencial de los factores de transcripción activadores frente a los represores en diferentes tejidos. La superposición de los sitios de unión también puede hacer que los represores expresados en un tipo celular compitan por la unión con los activadores, y los cofactores expresados diferencialmente pueden convertir activadores en represores.

Aisladores

Por último, los aisladores (o *insulators*) son un grupo de elementos regulatorios que se definen por dos propiedades (Gaszner and Felsenfeld, 2006). En primer lugar, pueden actuar como barrera para el estado de la cromatina. Esta actividad está definida por la capacidad de contener las marcas bioquímicas de la cromatina y evitar que las mismas se distribuyan hacia otras regiones del genoma. En segundo lugar, pueden bloquear la interacción entre un *enhancer* y un promotor, solo si el *insulator* se encuentra en el medio de ambos. Esta actividad aisladora es direccional y no afecta la capacidad del promotor o del *enhancer* de interactuar con otros elementos en la dirección opuesta.

Predicción de enhancers

Por definición, los *enhancers* contienen toda la información necesaria para reclutar a los factores de transcripción encargados de controlar la expresión espaciotemporal de sus genes blanco. Debido a que las secuencias consenso para la unión de factores de transcripción son altamente degeneradas, y están distribuidas por todo el genoma (Noonan and McCallion, 2010), la predicción de estos sitios no es suficiente para postular la existencia de un *enhancer*. En cambio, se utiliza una combinación de criterios. Por un lado, se utilizan métodos que detectan las características de la cromatina y de las proteínas de unión al ADN asociadas a la presencia de *enhancers*. Los *enhancers* típicamente se localizan en regiones donde hay una disrupción en la estructura de la cromatina. Esta disrupción puede observarse como una mayor sensibilidad a la digestión de la enzima DNasa I, que ocurre debido a que los nucleosomas son desplazados del ADN por la unión de proteínas específicas, quedando más expuesto a la digestión (Atchison, 1988). Además, presentan modificaciones específicas en las histonas de sus nucleosomas (Tafessu and Banaszynski, 2020). Dentro de este código se encuentra la presencia de la monometilación de la lisina 4 de la histona H3 (H3K4Me1), la ausencia de la trimetilación de la lisina 4 de la histona H3 (H3K4Me3) y la presencia de la acetilación de la lisina 27

de la histona H3 (H3K27ac). También se tiene en cuenta la asociación con proteínas específicas, como el coactivador p300, estrechamente asociado a los *enhancers* activos (Heintzman et al., 2007).

Por otro lado, la identificación sistemática a gran escala de *enhancers* fue posible por primera vez mediante estrategias genómicas comparativas. Estos enfoques se basan en que los elementos funcionales del genoma tienen tasas de evolución más bajas. Por lo tanto, se propuso que la conservación filogenética proporcionaría una forma de identificar *enhancers* dentro de la gran cantidad de secuencias no codificantes del genoma humano (Visel et al., 2009). Inicialmente, las comparaciones entre especies evolutivamente distantes, como humanos y peces, revelaron que muchas de estas secuencias que se habían conservado durante cientos de millones de años eran *enhancers* que impulsan la expresión en estructuras anatómicas altamente específicas durante el desarrollo embrionario (Aparicio et al., 1995; Pennacchio et al., 2006). Asimismo, los elementos ultraconservados entre humanos, ratones y ratas, están altamente enriquecidos en *enhancers* tejido-específicos (Bejerano et al., 2004; Visel et al., 2008). Sin embargo, aunque la conservación es indicativa de función, no es necesariamente indicativa de actividad *enhancer*, ya que existen muchos otros tipos de elementos funcionales que pueden tener tasas de conservación similares. Además, incluso cuando la conservación resulta de la función *enhancer*, la conservación no puede predecir cuándo y dónde está activo un *enhancer* en el organismo adulto o en desarrollo. Para todos los candidatos identificados, se necesitan estudios experimentales para descifrar las propiedades regulatorias de cada elemento.

Dominios topológicamente asociados

Si bien el conocimiento de la regulación transcripcional ha avanzado mucho en los últimos años, algunas respuestas están lejos de estar completas, incluyendo la capacidad de los elementos regulatorios de interactuar a distancia y de manera específica con sus genes blanco. El surgimiento de tecnologías relativamente recientes derivadas de la captura de la conformación de la cromatina (Dekker et al., 2002; Lieberman-Aiden et al., 2009), permiten detectar con una resolución sin precedentes la conformación tridimensional de la cromatina. De esta manera, ha surgido evidencia que indica que los genomas se organizan espacialmente en regiones discretas, denominadas dominios topológicamente asociados (TADs) (Dixon et al., 2012). Las interacciones entre los elementos dentro de un TAD son más frecuentes que las interacciones entre elementos de distintos TADs (Figura VI) (Symmons et al., 2014), en parte por la función de los TADs de acortar las distancias espaciales entre los elementos contenidos dentro de ellos. Además, los límites de los TADs están anclados por motivos de unión a la proteína arquitectónica CTCF (Phillips-Cremins et al., 2013). De acuerdo con el modelo de “extrusión de bucles”, las proteínas estructurales del complejo de cohesinas se unen al ADN e inician la formación de un bucle, y cuando encuentran una proteína CTCF unida al ADN, se detiene el crecimiento del bucle (Fudenberg et al., 2016). Aunque los motivos CTCF se encuentran distribuidos por todo el genoma, en los extremos de los TADs tienen una orientación convergente, con los motivos asimétricos enfrentados (Rao et al., 2014).

La inversión de la orientación del motivo CTCF puede ser suficiente para abolir o cambiar la posición de los límites de los TADs (de Wit et al., 2015; Guo et al., 2015).

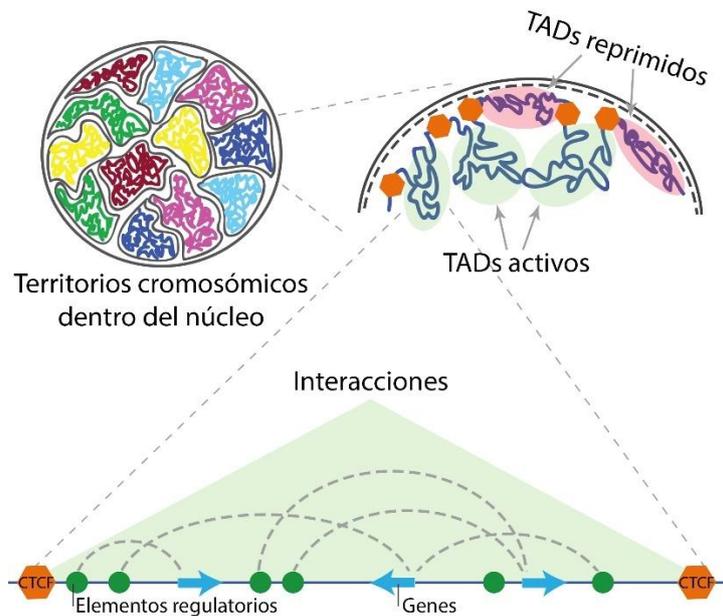


FIGURA VI: Organización tridimensional del genoma. Los genomas se organizan espacialmente en dominios asociados topológicamente (TADs). Las interacciones entre los elementos dentro de un TAD son más frecuentes que las interacciones entre elementos de distintos TADs. Modificado de (Matharu and Ahituv, 2015).

Evolución acelerada en el linaje humano

Dentro de los potenciales impulsores genéticos de la innovación evolutiva se encuentran las regiones aceleradas en humanos (HARs), secuencias evolutivamente conservadas que han acumulado más sustituciones que las esperadas al azar en el linaje humano (Figura VII a). La conservación de la secuencia sugiere que están sujetas a limitaciones funcionales, mientras que la mayor tasa de sustituciones en el linaje humano sugiere que su función pudo haber cambiado en nuestro linaje de manera específica (Levchenko et al., 2018). Curiosamente, la gran mayoría de los HARs son no codificantes. Motivados por el descubrimiento de King y Wilson en 1975 (King and Wilson, 1975), los autores que los describieron por primera vez plantearon la hipótesis de que la función de los HARs tenía que ser la regulación génica.

Varios trabajos, de manera independiente, han aplicado su propia metodología para identificar aceleración en el linaje humano (Pollard et al., 2006; Prabhakar et al., 2006; Bird et al., 2007; Bush and Lahn, 2008; Lindblad-Toh et al., 2011). En líneas generales, la metodología empleada consiste en identificar las regiones genómicas conservadas alineando un grupo de especies de mamíferos y/o vertebrados y aplicando un método estadístico para cuantificar el nivel de conservación. Luego se aplica otro método estadístico para detectar un subconjunto de elementos conservados que hayan sufrido cambios en su tasa de sustitución nucleotídica en el linaje humano. Por último, estas regiones son curadas eliminando aquellas que presentan errores en los alineamientos y/o que estuvieran presentes en regiones anotadas como pseudogenes. Las diferencias en las

opciones de análisis y los datos disponibles como, por ejemplo, las especies utilizadas en los alineamientos, los métodos utilizados para identificar elementos conservados, las pruebas de aceleración y los filtros bioinformáticos para eliminar artefactos, dieron como resultado una superposición modesta entre las distintas bases de datos de HARs identificados en los diferentes estudios (Kamm et al., 2013b). Estos trabajos proveen un conjunto de posibles elementos regulatorios que, dado su patrón de evolución molecular, son buenos candidatos a ser responsables de cambios en la regulación génica llevando eventualmente a la aparición de fenotipos humano-específicos.

Un análisis de la distribución de HARs en el genoma humano reveló que, de las 19.897 unidades transcripcionales anotadas en el genoma, solo 433 (2.5%) contiene al menos un HAR en sus regiones no codificantes, y de las unidades transcripcionales conteniendo HARs, 416 (96%) contienen 1, 2 o 3 elementos (Kamm et al., 2013b). Solo 7 genes humanos contienen 6 o más HARs, entre los que se destaca *NPAS3* (*Neuronal PAS domain protein 3*) con una cantidad excepcional de 14 HARs que fueron estudiados extensamente en el laboratorio durante las tesis doctorales de la Dra. Kamm y del Dr. Cinalli (Kamm et al., 2013b, 2013a). Dentro de este grupo de genes también se encuentra *RBFOX1* (*RNA binding protein, fox-1 homolog*) con 9 HARs. Este análisis se realizó sobre cuatro bases de datos de HARs (Pollard et al., 2006; Prabhakar et al., 2006; Bush and Lahn, 2008; Lindblad-Toh et al., 2011); si tenemos en cuenta una quinta base de datos (Bird et al., 2007) entonces *NPAS3* y *RBFOX1* acumulan 19 y 14 HARs, respectivamente. Además, sobre la base de que las interacciones entre elementos dentro del mismo TAD tienen más probabilidad de interactuar entre sí, un análisis de la distribución de HARs por TAD reveló que el TAD que contiene a *RBFOX1* está densamente poblado con 15 HARs. Este se encuentra dentro de los dominios con mayor densidad de HARs en todo el genoma, tanto en valor absoluto, es decir el número de HARs por TAD, como relativo a la longitud del TAD (Caporale et al., 2019). Estos dos análisis sugieren que el *locus* de *RBFOX1* pudo haber sido un sitio genómico donde se acumularon cambios moleculares a nivel de secuencias no codificantes conservadas desde la divergencia con el linaje de los chimpancés.

Caracterización funcional de los HARs

La primera estrategia para explorar la hipótesis de que los HARs cumplen funciones a nivel regulatorio fue asociar a los HARs con genes cercanos y hacer inferencias de asociación basadas en las funciones de estos genes, aprovechando la comprensión más detallada de las proteínas (Whalen and Pollard, 2022). Los análisis mostraron que los HARs están significativamente enriquecidos cerca de genes involucrados en la transcripción, la adhesión celular y el desarrollo, con un sesgo hacia tejidos del cerebro (Pollard et al., 2006; Prabhakar et al., 2006; Capra et al., 2013). Este patrón sugirió que los cambios de secuencia en los HARs durante la evolución humana podrían haber alterado la expresión de genes importantes en el cerebro. Sin embargo, la única evidencia que apoyaba esta hipótesis era la proximidad genómica; para poder verificar esta idea eran necesarios ensayos funcionales.

El primer enfoque que se utilizó para estudiar la función de los HARs como *enhancers* transcripcionales fueron los ensayos de gen reportero en ratones y peces cebra transgénicos (Figura VII b). Al integrar los resultados de estos estudios, encontramos que se han probado 86 HARs, observando actividad en al menos un tejido para 55 HARs (~64%), 24 de los cuales son *enhancers* activos en el cerebro (para una revisión ver (Whalen and Pollard, 2022)). Algunos ejemplos notables incluyen a HAR31 y HACNS369, dos *enhancers* intrónicos de *AUTS2* (*Autism susceptibility candidate 2*) (Oksenberg et al., 2013), un gen asociado con el trastorno de espectro autista (Sultana et al., 2002); 11 HARs en el *locus* de *NPAS3* (Kamm et al., 2013b), el gen con mayor cantidad de HARs en su unidad transcripcional que ha sido estrechamente asociado con esquizofrenia (Pickard et al., 2006); y 5 HARs en el *locus* de *FOXP2* (Caporale et al., 2019), un gen vinculado con el habla y el lenguaje en los humanos (Lai et al., 2001). Los tres estudios utilizaron ensayos reporteros en peces cebra transgénicos, un enfoque exitoso para capturar la actividad espaciotemporal *in vivo* de *enhancers* con alto rendimiento.

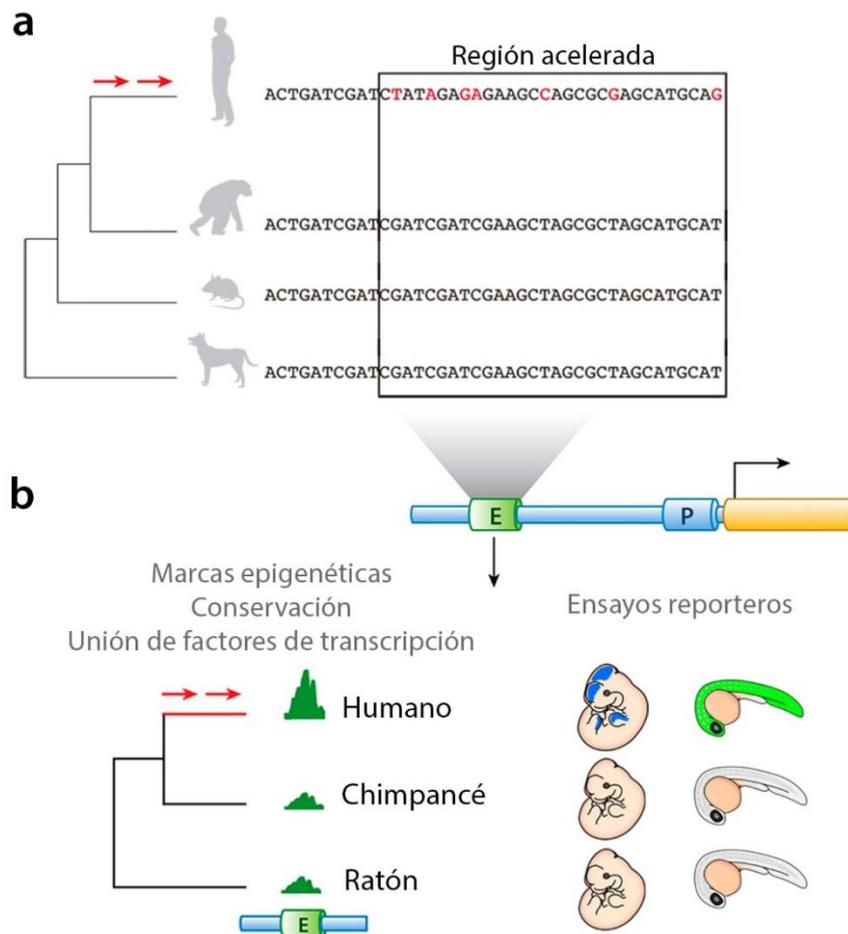


FIGURA VII: Evolución acelerada en humanos. (a) Las regiones aceleradas humanas (HARs) son secuencias evolutivamente conservadas que han acumulado más sustituciones que las esperadas al azar (rojo) en el linaje humano. (b) Caracterización funcional de los HARs como *enhancers* transcripcionales. Modificada de (Reilly and Noonan, 2016; Whalen and Pollard, 2022).

De los *enhancers* validados *in vivo*, en 33 se ha analizado comparativamente la actividad *enhancer* de la secuencia humana con la de chimpancé. Se observaron diferencias cualitativas de expresión entre las dos secuencias en 11 (~33%) de ellos. Algunos ejemplos incluyen a HAR2/HACNS1, un *enhancer* de *Gbx2* (*Gastrulation brain homeobox 2*) durante el desarrollo embrionario de las extremidades (Prabhakar et al., 2008; Dutrow et al., 2022); HARE5/ANC516, un *enhancer* de *Fzd8* (*Frizzled class receptor 8*) en células progenitoras neuronales que influye en la aceleración del ciclo celular y el tamaño del cerebro (Boyd et al., 2015); 2xHAR142, un *enhancer* que potencialmente regula a *Npas3* durante el desarrollo embrionario del cerebro anterior (Kamm et al., 2013a); y HACNS169, un *enhancer* que se expresa en el sistema nervioso durante el desarrollo embrionario, ubicado en la unidad transcripcional de *Foxp2* (Caporale et al., 2019). En todos los ejemplos mencionados, la secuencia humana del HAR tiene una actividad más extensa durante el desarrollo en comparación con su ortólogo de chimpancé en ensayos de gen reportero en ratones transgénicos.

***RBFOX1*, un gen candidato**

El gen de *RBFOX1* codifica para una proteína de unión al ARN (RBP) que se desempeña principalmente como un factor de *splicing* alternativo crítico en el desarrollo de múltiples tejidos, incluyendo el músculo esquelético, el músculo cardíaco y el cerebro (Jin et al., 2003; Nakahata and Kawamoto, 2005; Underwood et al., 2005). Tanto *RBFOX1* como los otros dos miembros de la familia RBFOX (*RBFOX2* y *RBFOX3*) se caracterizan por tener un único motivo de reconocimiento al ARN (RRM) altamente conservado que se une con gran especificidad a los motivos (U)GCAUG ubicados en los intrones que flanquean exones alternativos de los ARNm (Auweter et al., 2006). Típicamente actúan como activadores del *splicing* cuando se unen río abajo de un exón alternativo, y como represores cuando se unen río arriba (Figura VIII) (Jin et al., 2003; Nakahata and Kawamoto, 2005; Underwood et al., 2005; Kuroyanagi, 2009; Sun et al., 2012).

Existen varias redes de corregulación, tanto positivas como negativas, entre *RBFOX1* y otras RBP, como *MBN1* (Klinck et al., 2014), *NOVA1* (Zhang et al., 2010) y *PTBPI* (Li et al., 2015), que permiten coordinar la regulación del *splicing* alternativo en tejidos específicos. La unión cooperativa de múltiples factores que tienen patrones de expresión parcialmente superpuestos proporciona una mayor flexibilidad en la regulación del *splicing* que una simple interacción binaria entre *RBFOX1* y su motivo de ADN característico (Zhang et al., 2008). Además, *RBFOX1* forma parte de un gran complejo proteico junto con otros factores de *splicing* que afectan un conjunto de exones más amplio que el reconocido previamente debido a la interacción del complejo con los motivos de ADN de las otras RBP (Damianov et al., 2016). Por otro lado, el *splicing* alternativo es solo uno de los procesos metabólicos del ARN que están regulados por las RBP; en el citoplasma, *RBFOX1* se une a los motivos ubicados en las regiones 3' no traducidas (UTR), bloqueando la unión de micro-ARNs y, por ende, promoviendo la estabilidad y traducción de los ARNm blanco (Figura VIII) (Lee et al., 2016; Vuong et al., 2018).

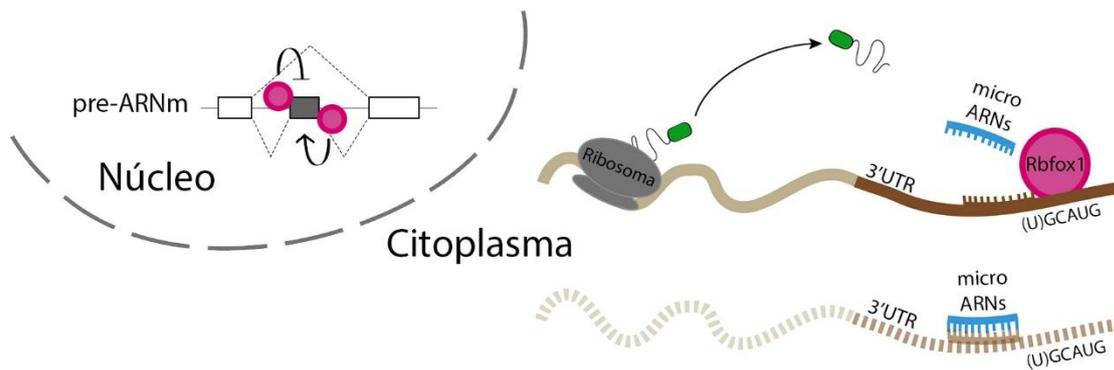


FIGURA VIII: *RBFOX1* tiene distintas funciones según su localización celular. En el núcleo, regula el *splicing* alternativo; cuando se une a los motivos (U)GCAUG ubicados río arriba de un exón alternativo actúa como represor del *splicing*, mientras que cuando se une río abajo actúa como activador, incluyendo al exón alternativo. En el citoplasma, se une a las regiones 3' UTR, bloqueando la unión de microARNs y promoviendo la traducción de los ARNm. Modificada de (Lipscombe and Lopez-Soto, 2018).

Finalmente, la evidencia genética sugiere que las aberraciones en los circuitos regulados por las proteínas RBFOX son factores de riesgo para múltiples trastornos humanos, especialmente trastornos del neurodesarrollo, como epilepsia (Bhalla et al., 2004; Lal et al., 2013a, 2015) y trastorno del espectro autista (TEA) (Martin et al., 2007; Sebat et al., 2007; Davis et al., 2012; Bill et al., 2013; Turner et al., 2016). En algunos casos se ha confirmado una expresión reducida del ARNm de *RBFOX1* en cerebros de pacientes con TEA, donde actúa como eje en la red transcripcional involucrada (Voineagu et al., 2011). Además, la mayoría de las deleciones reportadas eliminan exones codificantes, pero otras abarcan sólo secuencias intrónicas (Zhao, 2013). Estas deleciones podrían alterar la expresión de *RBFOX1* al eliminar promotores, interrumpir *enhancers* transcripcionales o afectar la estabilidad del ARNm y la eficiencia de la traducción.

Estructura génica y regulación transcripcional de RBFOX1

Comprender el rango de las funciones de *RBFOX1* requiere la apreciación de su compleja estructura génica y regulación transcripcional. La presencia de múltiples promotores alternativos, exones alternativos y sitios de *splicing* alternativos producen una diversidad de isoformas que difieren en sus dominios N- y C-terminal (Conboy, 2017). *RBFOX1* está codificado por un gen extremadamente grande de ~2.5 Mb con múltiples sitios de inicio de la transcripción en sus primeros exones alternativos que abarcan ~2.3 Mb, y que a menudo se describe como el 5' UTR (Damianov and Black, 2010); sin embargo, contiene exones codificantes evolutivamente conservados que mantienen un marco de lectura abierto que, en principio, pueden traducirse para generar proteínas con dominios N-terminal alternativos. En la región codificante, se genera mayor diversidad gracias a la inclusión de exones que producen dominios RRM activos frente a inactivos, exones que codifican isoformas cerebro- o músculo-específicas (Nakahata and Kawamoto, 2005), y exones que cambian el marco de lectura C-terminal para controlar la expresión de una señal de localización nuclear, generando isoformas nucleares o citoplasmáticas (Kuroyanagi, 2009).

Por otro lado, múltiples primeros exones, y presumiblemente promotores alternativos, se describieron previamente en ratón mediante PCR en tiempo real, encontrando que eran altamente específicos de tejido. Los primeros exones 1B y 1D.1 se expresan en corteza cerebral, estriado y cerebelo, mientras que 1C y 1C.1 se limitan al cerebelo y 1E se expresa exclusivamente en el corazón y músculo esquelético (Damianov and Black, 2010). Además, recientemente se identificaron dos primeros exones novedosos mediante secuenciación de isoformas (Iso-Seq), uno ubicado río arriba de 1B, denominado 1A, y otro ubicado río abajo de 1E, denominado 1F (Casanovas et al., 2020).

En cuanto a su regulación transcripcional, los intrones probablemente están poblados con múltiples *enhancers* que juegan un papel importante en la determinación de los patrones de expresión espaciotemporales, pero todavía no se ha descrito ninguno. En conjunto, estos procesos transcripcionales y postranscripcionales producen un conjunto de isoformas proteicas que difieren en función, distribución subcelular y expresión espaciotemporal (Figura IX).

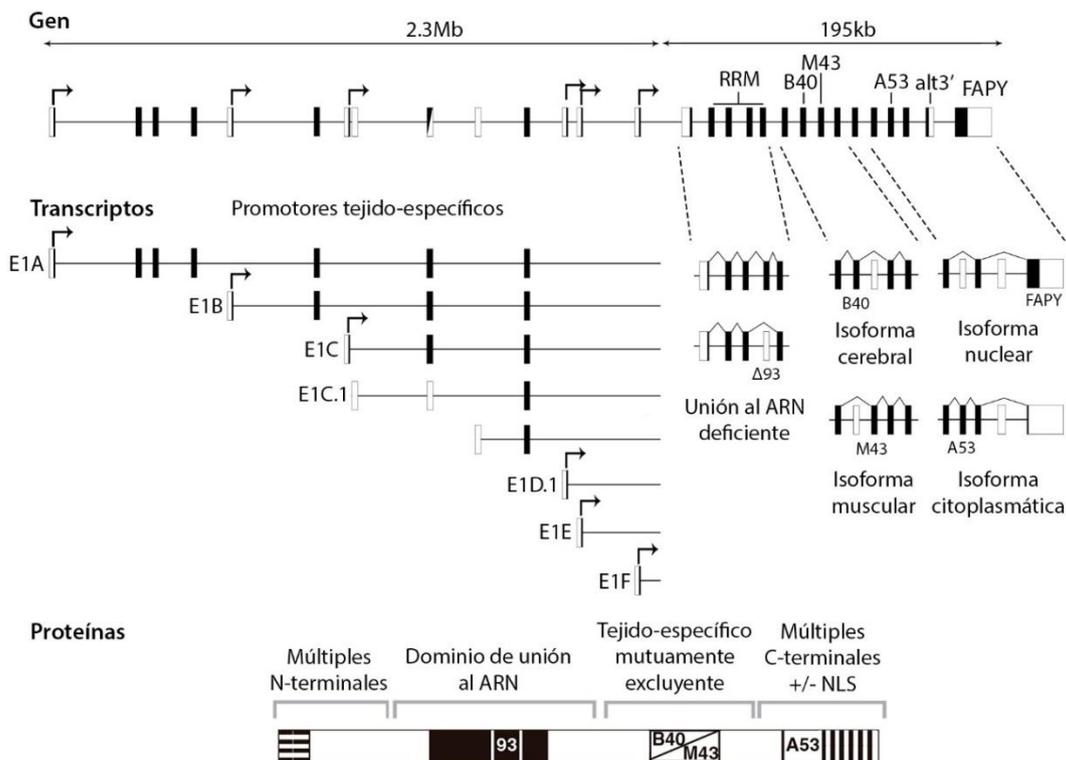


FIGURA IX: Estructura génica de *RBFOX1*. Múltiples promotores tejido-específicos, exones alternativos y sitios de splicing alternativos generan diversas isoformas de *RBFOX1*. La exclusión del tercer exón en la región RRM genera isoformas con unión al ARN deficiente; los exones mutuamente excluyentes B40 (cerebro-específico) y M43 (músculo-específico) generan isoformas que se expresan en tejidos específicos; la exclusión del exón A53 desplaza el marco de lectura y resulta en la pérdida de la señal de localización nuclear, generando isoformas con diferente localización celular. Las cajas blancas representan los exones no codificantes y las negras los codificantes. Cada inicio de la transcripción se indica con una flecha. Modificada de (Conboy, 2017).

Impacto funcional de la pérdida de Rbfox1 en modelos murinos

Varios estudios han caracterizado el papel de *Rbfox1* en el cerebro y el impacto funcional de su pérdida en modelos animales. Un ejemplo incluye la delección de *Rbfox1* específicamente en las células del sistema nervioso central y periférico en ratones *knock-out* condicionales (*Rbfox1^{loxp/loxp}/Nestin-Cre^{+/-}; Rbfox1^{-/-}*) (Gehman et al., 2011). Este modelo produce un fenotipo de hiperexcitabilidad neuronal en la circunvolución dentada y susceptibilidad a convulsiones espontáneas e inducidas con ácido kaínico, un agente neuroestimulador utilizado en estudios de epilepsia. Curiosamente, los cerebros de los ratones *Rbfox1^{-/-}* tienen un aumento del 60% en la expresión de *Rbfox2*, lo que sugiere una regulación cruzada que equilibra los niveles totales de las proteínas *Rbfox*. Sin embargo, los ratones *Rbfox2^{loxp/loxp}/Nestin-Cre^{+/-}* tienen un fenotipo neurológico distinto al de *Rbfox1^{-/-}*, sin hiperexcitabilidad neuronal ni susceptibilidad a convulsiones, evidenciando que *Rbfox1* y *Rbfox2* no son completamente redundantes en sus funciones neuronales (Gehman et al., 2012). Además, los ratones *Rbfox1^{-/-}* muestran relativamente pocos cambios en el *splicing* alternativo, probablemente debido a los cambios compensatorios en la expresión de *Rbfox2* y *Rbfox3* cuando *Rbfox1* está inactivo. Sin embargo, dichos cambios afectan múltiples componentes funcionales de la transmisión sináptica que podrían contribuir al fenotipo epiléptico, como canales iónicos, receptores de neurotransmisores y proteínas estructurales de la sinapsis. De particular interés, algunos transcritos alterados en los cerebros de los ratones *Rbfox1^{-/-}*, como *Gabrg2* (*Gamma-aminobutyric acid type A receptor subunit gamma 2*), *Grin1* (*Glutamate ionotropic receptor NMDA subunit 1*), *Scn8a* (*Sodium voltage-gated channel alpha subunit 8*) y *Snap25* (*Synaptosome associated protein 25*), han sido relacionados directamente con la epilepsia.

Además, el Consorcio Internacional de Fenotipos de Ratón (IMPC) es una entidad pública que se dedica a generar ratones *knock-out* de cada uno de los genes del genoma y analizar su consiguiente fenotipo (Dickinson et al., 2016). A diferencia de los ratones *knock-out* condicionales descritos en el apartado anterior, donde *Rbfox1* sólo está ausente en las células del sistema nervioso que expresan *Nestin*, los ratones del IMPC son mutantes nulos, es decir que la delección se da en cada una de las células de ratón. De esta manera, los ratones homocigotas *knock-out* de *Rbfox1* son letales antes del destete con penetrancia completa, lo que dificulta el correcto análisis del fenotipo en la adultez. Por otro lado, los ratones heterocigotas no presentan fenotipos graves en cuanto a la morfología de los tejidos ni al tamaño y peso corporal, pero tienen el largo de la tibia significativamente reducido (<https://www.mousephenotype.org/data/genes/MGI:1926224>).

HIPÓTESIS Y OBJETIVOS

Hipótesis general

La adquisición de nuevos patrones de expresión de genes relacionados con el desarrollo y la función cerebral en el linaje humano habría sido crítica para la evolución neuroanatómica del cerebro humano y de sus capacidades cognitivas diferenciales. Estos nuevos patrones de expresión estarían codificados en las sustituciones humano-específicas de las regiones regulatorias que actúan sobre dichos genes. Particularmente, en el presente trabajo se postula que el *locus* de *RBFOX1* podría haber experimentado evolución acelerada en sus regiones regulatorias en el linaje humano, lo que tuvo como consecuencia la aparición de nuevos patrones de expresión.

Objetivos generales

El objetivo principal de este trabajo es realizar un estudio detallado a nivel evolutivo y funcional de las regiones no codificantes de *RBFOX1* que evolucionaron de manera acelerada en la especie humana (*RBFOX1*-HARs). Para ello se propone: (i) estudiar si los *RBFOX1*-HARs pueden actuar como *enhancers* transcripcionales en peces cebra transgénicos, y analizar si presentan patrones de expresión diferenciales en comparación con sus secuencias ortólogas de chimpancé, y (ii) estudiar el impacto funcional *in vivo* de la delección de dos *RBFOX1*-HARs en la expresión de *Rbfox1* en ratones mutantes. Los puntos mencionados serán tratados en detalle en cada uno de los capítulos de este trabajo.

CAPÍTULO I

Hipótesis de trabajo

Postulamos que los HARs acumulados en el *locus* del gen *RBFOX1* tienen la capacidad de actuar como *enhancers* transcripcionales. Además, proponemos que el proceso evolutivo ocurrido en el linaje humano resultó en diferencias funcionales entre las secuencias humanas y sus ortólogas de chimpancé.

Objetivos específicos

El objetivo principal de este capítulo es determinar la función regulatoria de los *RBFOX1*-HARs. Para ello se propone:

- Analizar la actividad *enhancer* de manera comparada entre la secuencia humana y su ortóloga de chimpancé de los 19 HARs presentes en los intrones y cercanías de *RBFOX1* (*RBFOX1*-HARs) en ensayos de expresión de gen reportero en peces cebra transgénicos.
- Describir bioinformáticamente las sustituciones humano-específicas presentes en los *RBFOX1*-HARs utilizando bases de datos disponibles públicamente.
- Caracterizar la expresión de 2xHAR.321 y 2xHAR.17, dos HARs activos en el sistema nervioso mediante hibridación *in situ* e inmunofluorescencia.
- Analizar los patrones de expresión de los promotores alternativos de *RBFOX1* en ensayos de expresión de gen reportero en peces cebra transgénicos.

Resultados

Distribución de los HARs en RBFOX1

La unidad transcripcional de *RBFOX1* es una de las más grandes del genoma humano, ocupa aproximadamente 2.5 Mb en el brazo corto del cromosoma 16 (16p13.3) donde se agrupan 14 HARs (Figura 1a). Además, está organizada en dos TADs que no contienen ningún otro gen, excepto algunos pseudogenes y ARNs no codificantes (Figura 1b). Dentro de estos dos dominios se ubican 19 HARs, cuatro en el primer TAD (2xHAR.537, ANC.727, ANC.730 y ANC.737) y quince en el segundo TAD (ANC.217, 2xHAR.321, 2xHAR.300, ANC.739, HAR.135, 2xHAR.128, HACNS.177, 2xHAR.505, 2xHAR.437, 2xHAR.17, ANC.221, ANC.750, 2xHAR.204, 2xHAR.551 y HACNS.705). Los últimos cinco son intergénicos mientras que el resto son intrónicos. Los nombres de los HARs (HAR, 2xHAR, HACNS, ANC) derivan de los nombres asignados originalmente en los análisis que los reportaron (ver Materiales y Métodos).

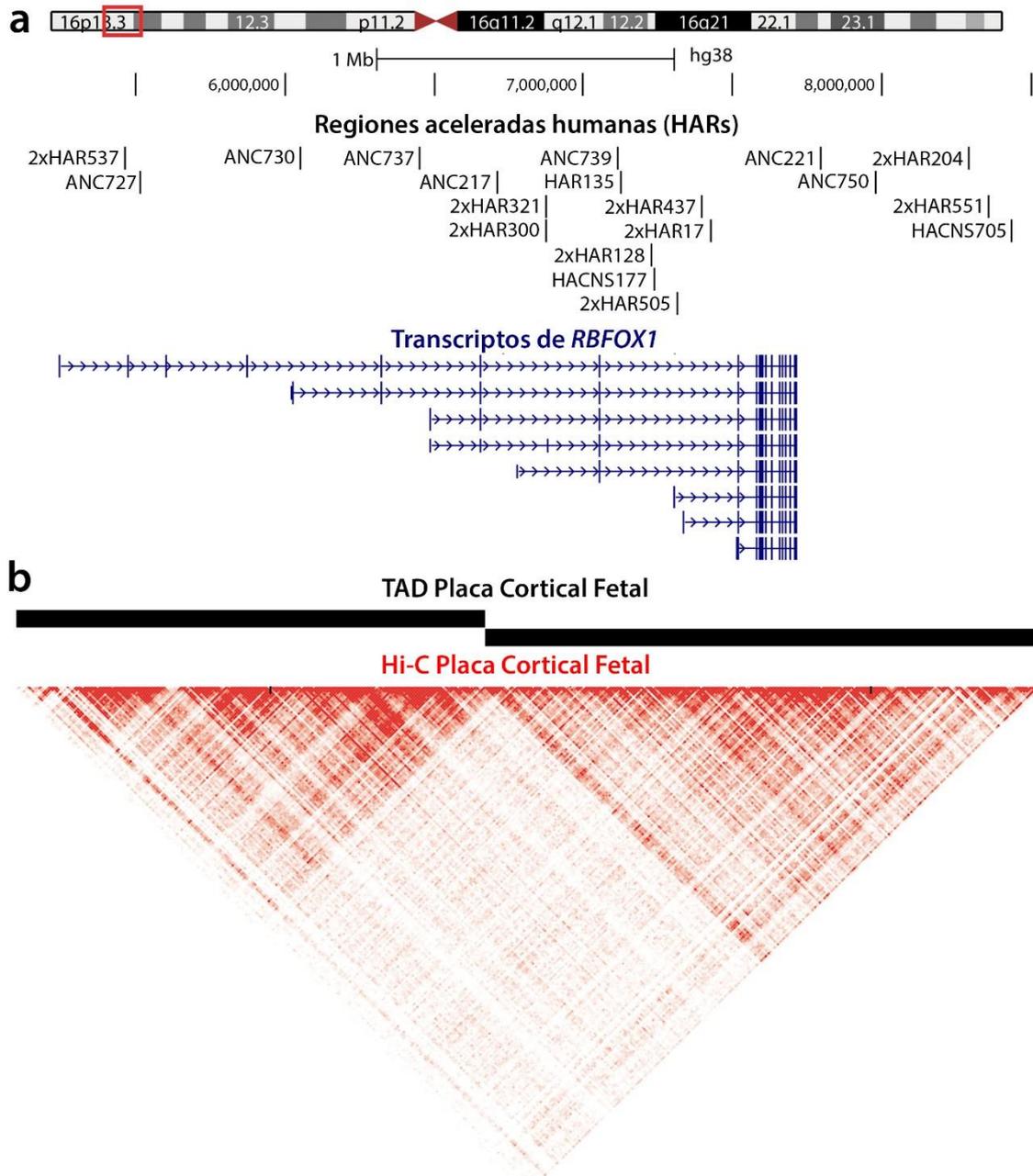


FIGURA 1: Localización genómica de los *RBFOX1*-HARs. (a) Ventana del UCSC *Genome Browser* donde se observan los distintos transcritos de *RBFOX1* y la localización genómica de los *RBFOX1*-HARs en el cromosoma 16 del genoma humano. Catorce HARs son intrónicos, mientras que cinco son intergénicos. (b) Visualización mediante un mapa de calor (*heatmap*) de las interacciones de la cromatina en la placa cortical fetal humana determinadas por Hi-C (Won et al., 2016). Cuanto más rojo el punto, mayor frecuencia de interacción.

Una pregunta que surge en primer lugar es si la acumulación de HARs se debe simplemente al gran tamaño de la región que ocupa *RBFOX1* en el genoma. De hecho, en un análisis de la distribución de los HARs en el genoma humano observaron que los genes con más de 800 kb (kilobase) tienen mayor probabilidad de tener al menos un HAR (Kamm et al., 2013b). Si consideramos el valor absoluto, es decir la cantidad total de HARs por unidad transcripcional, *RBFOX1* se ubica en segundo lugar detrás de *NPAS3*

(Figura 2a). Sin embargo, si calculamos la frecuencia de HARs/Mb (Megabase) normalizado al tamaño de la unidad transcripcional, *RBFOX1* cae al puesto 12 con aproximadamente 5,6 HARs/Mb (Figura 2b). Es evidente que la acumulación de HARs en *RBFOX1* puede explicarse parcialmente en función de su tamaño. Sin embargo, también es cierto que la mayoría de los genes con más de 800 kb no contiene HARs o contiene una frecuencia de hasta 2 HARs/Mb (81 de 143; ~56%). En cambio, *RBFOX1* se encuentran dentro del 10% con mayor frecuencia de HARs/Mb (Figura 2b). No es necesariamente un valor atípico, pero igualmente resulta una región genómica interesante para estudiar evolutivamente. Además, el segundo TAD que contiene a *RBFOX1* se encuentra dentro de los dominios más enriquecidos con HARs del genoma humano con ~8,1 HARs/Mb (Caporale et al., 2019). Dado que se ha demostrado que la mayoría de las interacciones ocurren dentro del mismo TAD, es probable que estos HARs interactúen con los promotores de *RBFOX1* para regular su expresión. De esta manera, nos propusimos caracterizar *in vivo* a los *RBFOX1*-HARs, con el fin de determinar si la aceleración en estas regiones se refleja en un cambio de función.

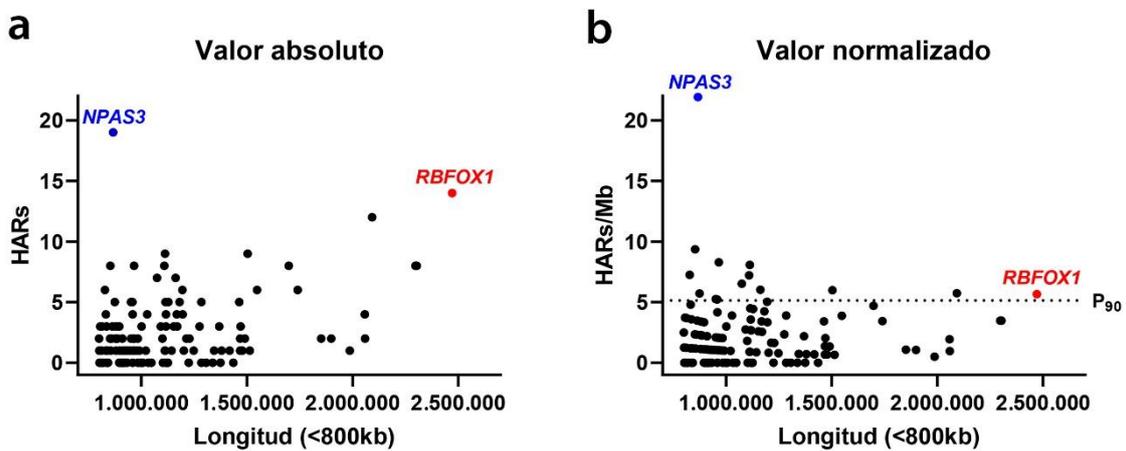


FIGURA 2: Distribución de los HARs en unidades transcripcionales mayores a 800 kb. (a) Cantidad total de HARs por unidad transcripcional. (b) Frecuencia de HARs/Mb normalizado al tamaño de la unidad transcripcional. La línea punteada representa el percentil 90avo (P₉₀); todos los valores por encima están dentro del 10% con mayor frecuencia de HARs/Mb.

Análisis funcional de los RBFOX1-HARs en peces cebra transgénicos

Basándonos en la hipótesis original de que los HARs actúan como *enhancers* transcripcionales, generamos peces cebra transgénicos para los 19 HARs presentes en los intrones y cercanías de *RBFOX1*. Además, para cada uno analizamos la expresión de la secuencia humana en comparación con su ortóloga de chimpancé, nuestro pariente primate vivo más cercano, para evaluar potenciales diferencias de expresión.

Brevemente, la estrategia consiste en clonar cada secuencia en un vector de expresión basado en el sistema Tol2 (Kawakami, 2007) río arriba del promotor mínimo de ratón *cFos* y la proteína reportera EGFP (Fisher et al., 2006b). Cada construcción se microinyecta en cigotos de pez cebra en estadio de una célula junto al ARN mensajero de una transposasa, la cual media la transgénesis en el genoma. Luego se crían hasta la madurez sexual y se cruzan con peces de tipo salvaje o *wild type* para obtener la

transmisión de la línea germinal. La expresión del reportero se analiza en la progenie a las 24, 48 y 72 horas post fecundación (hpf). Para evitar la influencia de los efectos posicionales, en todos los casos generamos tres líneas estables independientes (para todas las líneas transgénicas ver Apéndice II, Figura S1). Además, decidimos clonar toda la región conservada que contiene a cada HAR ya que consideramos que probablemente sea más representativa de una potencial unidad funcional. Se clasificaron los dominios de expresión de EGFP utilizando únicamente términos anatómicos generales y comparables entre vertebrados y se evitó la clasificación dentro de subdominios o tipos celulares.

De esta manera, observamos que la mayoría de los *RBFOX1*-HARs fue capaz de generar un patrón de expresión distintivo y reproducible del reportero. Además, en general observamos patrones prácticamente ubicuos durante los tres estadios analizados. Para facilitar la interpretación de los resultados, los dividimos en cuatro partes.

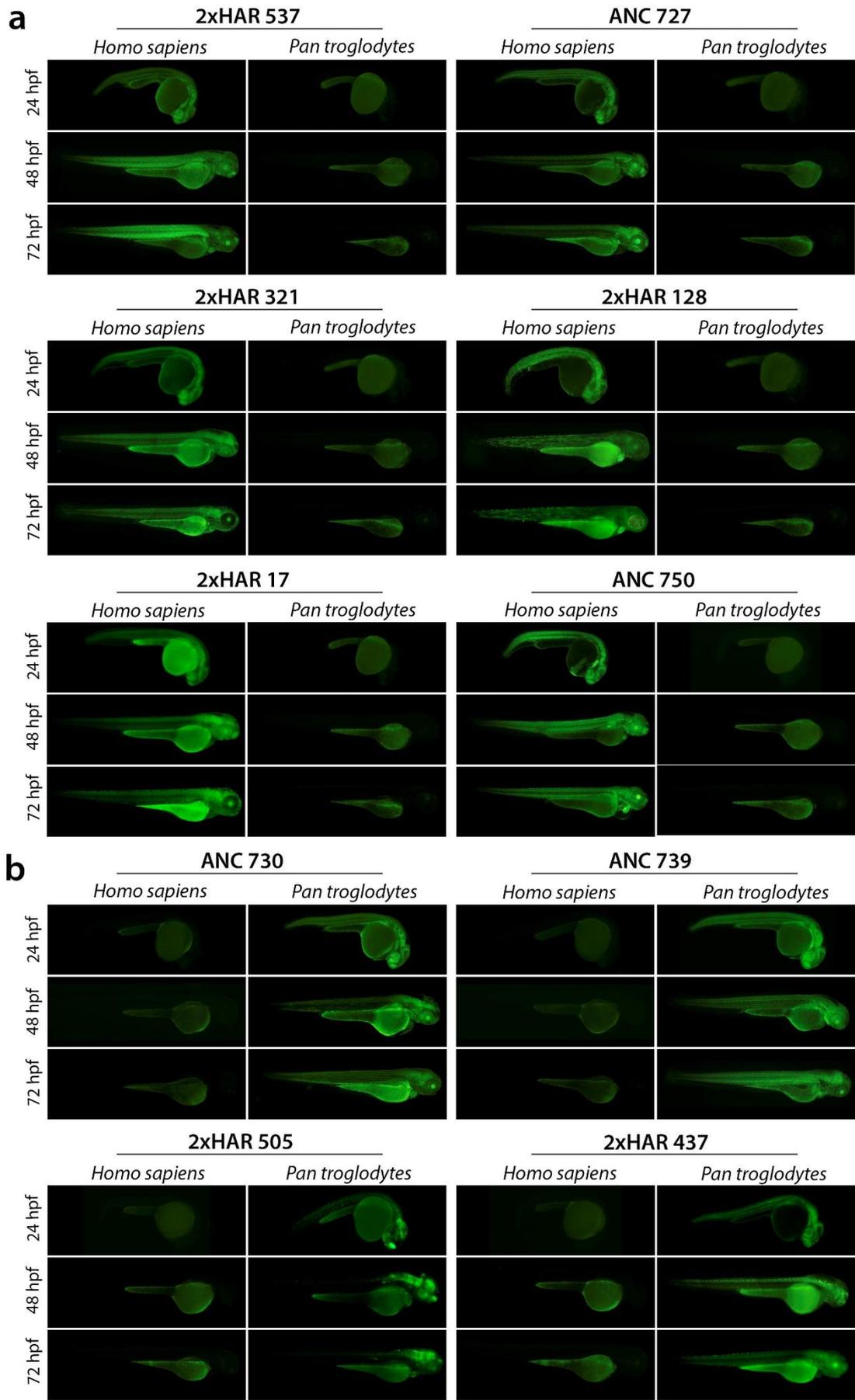
Por un lado, observamos que la secuencia humana de seis *RBFOX1*-HARs (2xHAR.537, ANC.727, 2xHAR.321, 2xHAR.128, 2xHAR.17 y ANC.750) actúa como *enhancer* mientras que la secuencia de chimpancé no (Figura 3a; Tabla I). En la mayoría (todos menos 2xHAR.128), observamos una fuerte expresión del reportero en el sistema nervioso en desarrollo, incluyendo el prosencéfalo, mesencéfalo, rombencéfalo y la médula espinal. También observamos expresión en el ojo y en el corazón, y en un caso (2xHAR.537) en el músculo esquelético. Además, 2xHAR.128 fue el único que mostró expresión restringida al corazón y al músculo esquelético. Por lo contrario, en el caso de otros cuatro *RBFOX1*-HARs (ANC.730, ANC.739, 2xHAR.505 y 2xHAR.437), mientras que la secuencia humana no fue capaz de impulsar la expresión del reportero, sus ortólogos de chimpancé generaron un amplio patrón de expresión (Figura 3b; Tabla I). En todos los casos, observamos expresión en todo el sistema nervioso, incluyendo el prosencéfalo, mesencéfalo, rombencéfalo y la médula espinal, y en el ojo y corazón. Por otro lado, otros cuatro *RBFOX1*-HARs (ANC.737, 2xHAR.300, HAR.135 y HACNS.177) no actuaron como *enhancers* ni la secuencia humana ni el ortólogo de chimpancé (Figura 3c; Tabla I). Por último, los patrones de expresión de la secuencia humana de cinco *RBFOX1*-HARs (ANC.217, ANC.221, 2xHAR.204, 2xHAR.551 y HACNS.705) fueron indistintos de sus ortólogos de chimpancé (Figura 3d; Tabla I).

TABLA I: Análisis de expresión en líneas estables de peces cebra transgénicos.

HAR	Localización (chr16)	Tamaño (pb)	Líneas con expresión	Dominios de expresión							
				P	M	R	M E	O	C	M u	
2xHAR.537	Hs	5,457,916-5,458,321	410	3/3	+	+	+	+	+	+	+
	Pt	5,526,361-5,526,764	408	0/3	-	-	-	-	-	-	-
ANC.727	Hs	5,511,846-5,512,070	229	3/3	+	+	+	+	+	+	-
	Pt	5,581,882-5,582,106	229	0/3	-	-	-	-	-	-	-
ANC.730	Hs	6,048,430-6,048,734	309	0/3	-	-	-	-	-	-	-
	Pt	6,122,214-6,122,518	309	3/3	+	+	+	+	+	+	-

ANC.737	Hs	6,446,900-6,447,452	557	0/3	-	-	-	-	-	-	-
	Pt	6,534,618-6,535,170	557	0/3	-	-	-	-	-	-	-
ANC.217	Hs	6,709,500-6,709,761	266	3/3	+	+	+	+	+	+	-
	Pt	6,797,501-6,797,762	266	3/3	+	+	+	+	+	+	-
2xHAR.321	Hs	6,870,345-6,870,950	606	3/3	+	+	+	+	+	+	-
	Pt	6,958,338-6,958,944	610	0/3	-	-	-	-	-	-	-
	Mm	6,672,274-6,672,926	653	3/3	+	+	+	+	+	+	-
2xHAR.300	Hs	6,872,793-6,873,646	854	0/3	-	-	-	-	-	-	-
	Pt	6,960,769-6,961,617	850	0/3	-	-	-	-	-	-	-
ANC.739	Hs	7,111,066-7,111,533	472	0/3	-	-	-	-	-	-	-
	Pt	7,203,515-7,203,982	472	3/3	+	+	+	+	+	+	-
HAR.135	Hs	7,124,816-7,124,969	154	0/3	-	-	-	-	-	-	-
	Pt	7,221,835-7,221,984	154	0/3	-	-	-	-	-	-	-
2xHAR.128	Hs	7,227,257-7,227,677	421	3/3	-	-	-	-	+	+	+
	Pt	7,329,289-7,329,786	503	0/3	-	-	-	-	-	-	-
HACNS.177	Hs	7,234,711-7,235,231	521	0/3	-	-	-	-	-	-	-
	Pt	7,336,704-7,337,217	507	0/3	-	-	-	-	-	-	-
2xHAR.505	Hs	7,308,354-7,310,174	1821	0/3	-	-	-	-	-	-	-
	Pt	7,412,193-7,414,009	1805	3/3	+	+	+	+	+	+	-
	Mm	7,043,904-7,045,657	1754	3/3	+	+	+	+	+	+	-
2xHAR.437	Hs	7,394,843-7,396,014	1172	0/3	-	-	-	-	-	-	-
	Pt	7,499,881-7,501,043	1167	3/3	+	+	+	+	+	+	-
2xHAR.17	Hs	7,422,050-7,423,055	1006	3/3	+	+	+	+	+	+	-
	Pt	7,786,312-7,787,190	883	0/3	-	-	-	-	-	-	-
ANC.221	Hs	7,792,466-7,792,615	154	3/3	+	+	+	+	+	+	-
	Pt	7,900,805-7,900,954	154	3/3	+	+	+	+	+	+	-
ANC.750	Hs	7,979,988-7,980,527	540	3/3	+	+	+	+	+	+	-
	Pt	8,093,444-8,093,983	540	0/3	-	-	-	-	-	-	-
2xHAR.204	Hs	8,291,250-8,291,561	316	3/3	+	+	+	+	+	+	-
	Pt	8,407,399-8,407,680	286	3/3	+	+	+	+	+	+	-
2xHAR.551	Hs	8,357,795-8,358,097	307	3/3	+	+	+	+	+	+	-
	Pt	8,472,819-8,473,069	255	3/3	+	+	+	+	+	+	-
HACNS.705	Hs	8,433,120-8,433,769	654	3/3	+	+	+	+	+	+	-
	Pt	8,551,005-8,551,654	654	3/3	+	+	+	+	+	+	-
TOTAL				66/120							

Referencias: Hs, *Homo sapiens*. Pt, *Pan troglodytes*. Mm, *Mus musculus*. P, prosencéfalo. M, mesencéfalo. R, rombencéfalo. ME, médula espinal. O, ojo. C, corazón. Mu, músculo. Con (+) y (-) se indica presencia o ausencia de expresión de EGFP, respectivamente.



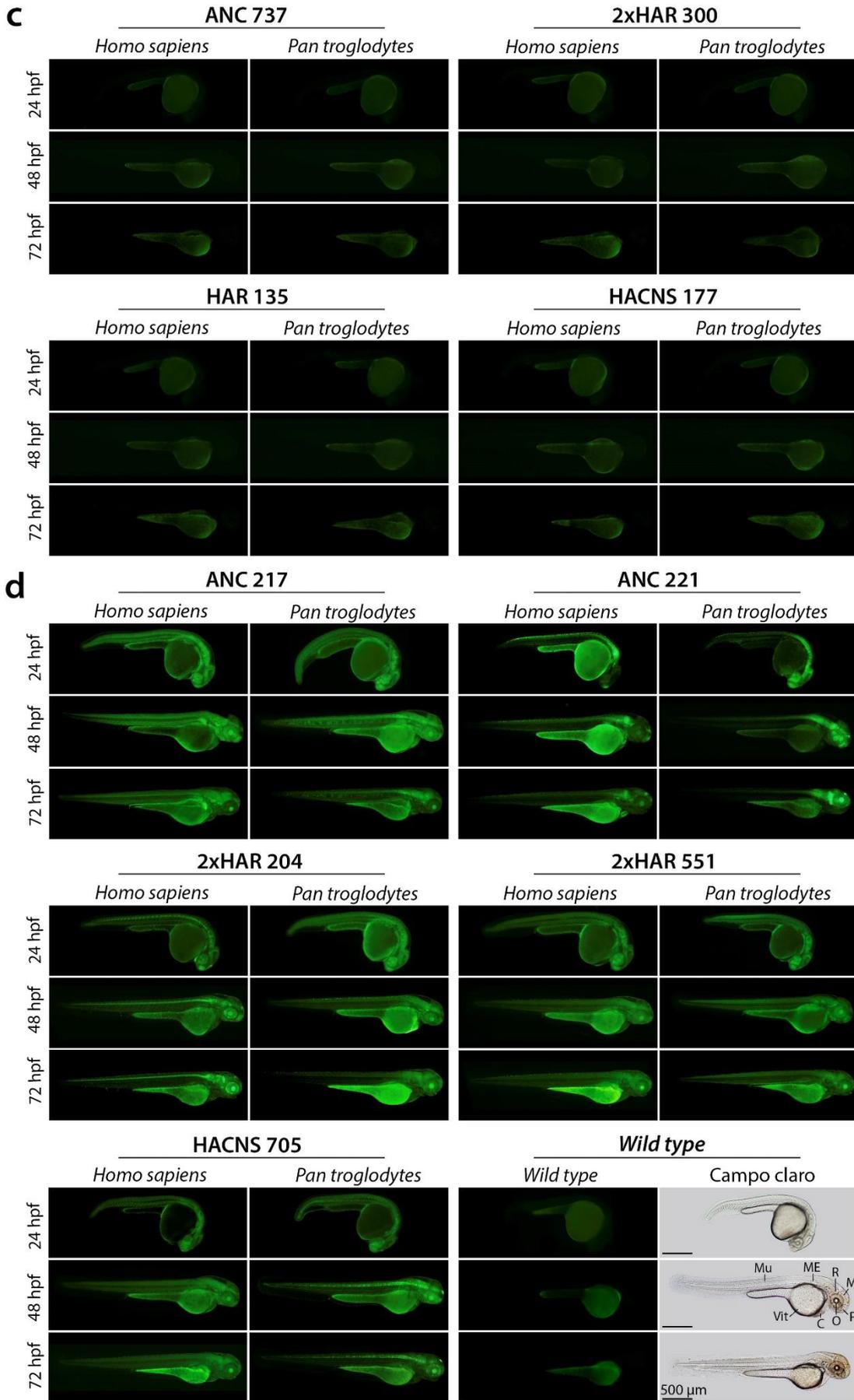


FIGURA 3: Ensayo de expresión en líneas estables de peces cebra transgénicos. Para cada *RBFOX1*-HARs se incluye una imagen representativa del patrón de expresión de la secuencia humana (*Homo sapiens*) y su ortóloga de chimpancé (*Pan troglodytes*) a 24, 48 y 72 hpf. Se divide en cuatro partes: (a) potenciales ganancias de función en el linaje humano, (b) potenciales pérdidas de función en el linaje humano, (c) HARs sin actividad *enhancer* y (d) HARs con patrones de expresión indistinguibles entre ortólogos. Se incluyen imágenes de referencia en campo claro y fluorescencia de peces cebra *wild-type*. El vitelo es autofluorescente. Referencias: P, prosencéfalo. M, mesencéfalo. R, rombencéfalo. ME, médula espinal. O, ojo. C, corazón. Mu, músculo. Vit, vitelo.

En resumen, descubrimos que 15 de los 19 *RBFOX1*-HARs son *enhancers* activos en peces cebra transgénicos y, en última instancia, que 10 HARs exhiben una expresión diferencial en comparación con sus ortólogos de chimpancé como consecuencia del proceso evolutivo de aceleración. Además, de los HARs con diferencia de expresión, 6 son potenciales ganancia de función y 4 potenciales pérdidas de función en el linaje humano. Para confirmar si efectivamente son ganancias o pérdidas de función, es necesario tener evidencia adicional proveniente de una tercera especie cercanamente emparentada para excluir la posibilidad de que el patrón de expresión observado para la secuencia de chimpancé sea el estado derivado. Para resolver este interrogante, en el laboratorio se decidió generar peces cebra transgénicos con las secuencias ortólogas de macaco Rhesus (*Macaca mulatta*) de los 10 *RBFOX1*-HARs diferencialmente expresados. El macaco Rhesus es una especie de primate catarrino de la familia Cercopithecidae, cuya divergencia con el linaje humano se estima hace 25 millones de años aproximadamente (Gibbs et al., 2007). El clonado de las secuencias fue llevado a cabo por Delia Christ, estudiante de la Universidad de Basilea, Suiza. La inyección de los transgenes y el análisis de los peces reporteros está siendo llevado a cabo actualmente por Abril Sileo, estudiante de Licenciatura en Biotecnología de la UADE en el marco de su tesis de licenciatura. Los resultados de las secuencias del macaco Rhesus no se muestran en este trabajo.

En cambio, extendimos el análisis a la secuencia ortóloga de ratón de dos *RBFOX1*-HARs que resultaron de interés. Además de actuar como tercera especie para confirmar la ganancia o pérdida de función de los HARs, la secuencia de ratón proporciona una base para los resultados del Capítulo II. De esta manera, seleccionamos a 2xHAR.321 y 2xHAR.505, dos HARs intrónicos de *RBFOX1*. Por un lado, 2xHAR.321 muestra una potencial ganancia de función en el linaje humano; la secuencia humana actúa como *enhancer* en el sistema nervioso en desarrollo, mientras que la secuencia de chimpancé no. Además, existen varias deleciones intrónicas que incluyen a la región comprendida por 2xHAR.321 reportadas en casos clínicos de diversos trastornos neurológicos, como trastorno del espectro autista (Turner et al., 2016), epilepsia, retraso en el habla y retraso del desarrollo global (Zhao, 2013). Para confirmar la ganancia de función en el linaje humano esperábamos que la secuencia de ratón sea incapaz de comportarse como *enhancer*, al igual que la secuencia de chimpancé. Sin embargo, observamos una fuerte expresión del reportero en los mismos dominios dirigidos por la secuencia humana (Figura 4a; Tabla I). Además, en los resultados preliminares de la secuencia de macaco

Rhesus también observamos expresión del reportero, lo que sugiere que el estado derivado no es el patrón de expresión dirigido por la secuencia humana, sino el de la secuencia de chimpancé. Una explicación de estos resultados es que en vez de ser una ganancia de función en el linaje humano, sea una pérdida de función en el linaje chimpancé. Por otro lado, también es posible que 2xHAR.321 sea un *enhancer* en mamíferos y otros primates, pero que el antepasado en común entre humanos y chimpancés haya perdido su función, y luego solo la rama de los humanos la haya vuelto a ganar. Se necesitan más análisis para confirmar o descartar estas explicaciones.

Por otro lado, la secuencia de 2xHAR.505 muestra una potencial pérdida de función en el linaje humano; la secuencia humana es incapaz de impulsar la expresión del reportero, mientras que su ortólogo de chimpancé se comporta como un fuerte *enhancer* en el sistema nervioso en desarrollo. Como era de esperarse, la actividad de la secuencia de ratón es igual a la de su ortólogo de chimpancé, confirmando la pérdida de función en el linaje humano (Figura 4b, Tabla I).

En total, en este trabajo generamos y analizamos 120 líneas de peces cebras transgénicos; 3 para cada una de las secuencias humanas y de chimpancé de los 19 *RBFOX1*-HARs, y 3 para cada una de las secuencias de ratón de 2xHAR.321 y 2xHAR.505.

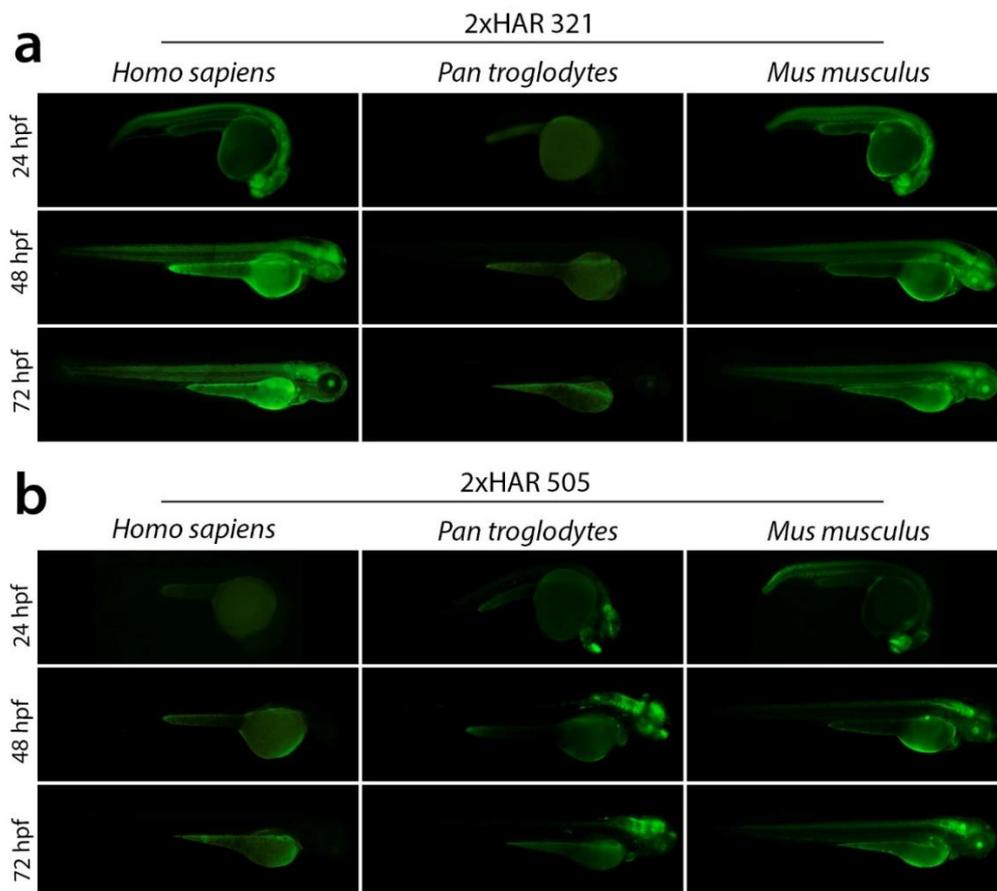


FIGURA 4: Análisis de expresión de una especie externa (*outgroup*) para confirmar la ganancia y pérdida de función en el linaje humano. Se incluye una imagen representativa del patrón de expresión de la secuencia humana (*Homo sapiens*) y sus ortólogos de chimpancé (*Pan troglodytes*) y ratón (*Mus musculus*) a 24, 48 y 72 hpf de (a) 2xHAR.321 y (b) 2xHAR.505.

Análisis de las sustituciones humano-específicas

Sobre la base de que las sustituciones humano-específicas presentes en las secuencias de los *RBFOX1*-HARs son las responsables de las diferencias de expresión que observamos entre las secuencias humanas y de chimpancé, decidimos realizar una descripción detallada de estas con bases de datos disponibles públicamente para complementar nuestros resultados *in vivo*.

En la Tabla II se describen cada una de las sustituciones humano-específicas de los HARs diferencialmente expresados, incluyendo la base humana acelerada, la base ancestral inferida a partir del alineamiento con genomas de otros vertebrados y la base de humanos arcaicos. Gracias a la publicación del primer genoma de *Homo neanderthalensis* de alta calidad (Green et al., 2010; Prüfer et al., 2014), es posible preguntarse en qué momento de la historia de nuestro linaje fueron fijadas las sustituciones humano-específicas. La mayoría de estas resultaron estar compartidas con el genoma de humanos arcaicos, y las que fueron anotadas como derivadas son polimórficas tanto en las poblaciones humanas actuales como en las poblaciones de humanos arcaicos, lo que sugiere que la mayoría de estos cambios ocurrieron antes de la divergencia entre *H. sapiens* y *H. neanderthalensis*.

Por otro lado, desde la publicación del genoma humano, ha habido un gran avance en el conocimiento de la genética de las poblaciones humanas y, por lo tanto, de la comprensión de la naturaleza de la diversidad genética. Haciendo uso de la base de datos del NCBI dbSNPs (Sherry et al., 2001), ubicamos las sustituciones que coinciden con polimorfismos comunes (*minor allele frequency*, MAF \geq 1%), ya que son menos probables de tener un efecto funcional humano-específico. Esta base de datos utiliza miles de genomas humanos provenientes de África, Europa, Asia del Este y América generando una colección de variación genéticas tales como polimorfismos de nucleótido único, *indels* y variantes estructurales. Por último, *phyloP* es un estimador base específico que permite inferir el grado de conservación de las sustituciones (Pollard et al., 2010); mientras más alto sea su valor, más conservada estará la base, siendo así más probable su funcionalidad. Para 2xHAR.321 y 2xHAR.505 también incluimos una figura con las sustituciones-humano específicas de cada región (Figura 5).

TABLA II: Análisis *in silico* de las sustituciones humano-específicas de los *RBFOX1*-HARs con expresión diferencial.

HAR	N° cambios	Posición (chr16)	Hs	Hn	Pt	A	<i>phyloP</i>	dbSNPs
2xHAR.537	6	5,458,097	C	C	G	G	2.07	C>G
		5,458,108	C	C	T	T	1.36	
		5,458,117	A	A	T	T	1.36	
		5,458,206	A	A	G	G	0.00	
		5,458,245	T	T	A	A	-1.08	
		5,458,246	G	G	A	A	-0.29	
ANC.727	4	5,511,899	T	T	C	C	0.51	C>T
		5,511,914	C	C	T	T	0.90	

		5,511,980	G	G	A	A	2.51	
		5,512,001	A	A	G	G	1.59	A>G
ANC.730	7	6,048,473	A	A	C	C	-0.16	
		6,048,481	C	C	A	A	0.08	
		6,048,508	T	T	C	C	1.80	
		6,048,592	C	C	T	T	0.86	
		6,048,621	C	C	G	G	1.41	
		6,048,640	T	T	C	C	1.96	
		6,048,646	G	G	A	A	-0.39	
2xHAR.321	11	6,870,383	A	A	G	C	0.26	
		6,870,411	T	T	A	A	0.00	
		6,870,456	A	A	G	G	0.13	A>G,T
		6,870,591	C	T>C	T	T	0.38	C>A,G,T
		6,870,650	C	C	T	T	1.31	
		6,870,717	G	G	A	C	0.29	
		6,870,740	C	C	T	T	0.89	
		6,870,803	G	G	A	A	4.04	
		6,870,880	A	A	G	G	1.31	
		6,870,882	T	T	A	A	1.74	
		6,870,940	G	G	A	A	0.89	
ANC.739	6	7,111,096	G	G	C	C/T	0.39	
		7,111,168	A	A	T	T	1.45	
		7,111,327	C	C	T	T	0.92	C>A,G,T
		7,111,330	C	C	T	T	1.45	
		7,111,356	G	A>G	A	A	-1.20	G>A
		7,111,482	A	A	T	T	-0.02	
2xHAR.128	5	7,227,429	G	G	C	C	1.87	
		7,227,460	G	G	T	A	0.59	
		7,227,479	G	G	A	A	0.23	
		7,227,494	G	G	A	A	1.59	
		7,227,509	G	G	C	C	0.66	
2xHAR.505	15	7,308,414	T	T	C	C	-0.15	
		7,308,738	G	G	A	A	0.44	
		7,308,767	G	G	A	A	-0.07	
		7,308,993	A	A	G	G	1.61	
		7,309,020	G	G	A	A	2.03	
		7,309,213	G	G	A	A	0.28	
		7,309,534	T	T	C	C	0.71	
		7,309,707	G	G	A	A	0.11	
		7,309,808	T	T	C	C	1.56	
		7,309,823	C	G>C	G	G	0.54	C>A,G,T
		7,309,841	G	G	T	T	0.03	
		7,309,887	C	C	A	A	0.45	
		7,309,918	A	A	G	G	1.13	
		7,309,935	G	G	A	A	0.37	
		7,310,167	C	C	G	G	-0.38	
2xHAR.437	12	7,394,915	T	T	C	C	0.34	T>C,G
		7,394,975	A	A	G	G	0.03	
		7,394,978	C	A/C	A	A	-2.58	C>A
		7,395,059	T	T	A	A	-0.93	T>A,G
		7,395,099	T	C>T	C	C	0.50	T>C,G
		7,395,263	A	A	T	T	2.90	
		7,395,575	C	C	A	A	0.72	

		7,395,658	C	C	A	A	2.45	
		7,395,704	T	T	C	C	-0.70	
		7,395,781	C	C	A	A	3.96	
		7,395,793	C	T>C	T	T	2.30	C>T
		7,395,867	A	A	G	G	0.95	
		7,422,517	G	G	C	C	3.60	
		7,422,552	A	A	G	G	6.18	
		7,422,553	A	A	C	C	4.66	
2xHAR.17	8	7,422,566	C	C	T	T	2.70	
		7,422,580	C	C	T	T	0.12	
		7,422,605	C	C	G	G	2.53	
		7,422,886	C	C	T	T	-0.28	
		7,422,965	T	T	A	A	1.45	
		7,980,061	T	T	G	G	-0.14	
ANC.750	3	7,980,187	T	T	C	C	0.38	
		7,980,213	T	T	A	A	2.71	

Referencias: Hs, *Homo sapiens*. Hn, *Homo neanderthalensis*. Pt, *Pan troglodytes*. A, ancestral. En naranja se destacan las sustituciones del *core* acelerado. En amarillo las sustituciones donde la base de chimpancé es distinta a la ancestral.

Teniendo en cuenta que en los ensayos *in vivo* observamos que la secuencia de chimpancé de 2xHAR.321 es el estado derivado, decidimos también evaluar las sustituciones chimpancé-específicas en dicha secuencia (Tabla III). De esta manera, observamos varias sustituciones que ocurrieron exclusivamente en el linaje chimpancé, dos de las cuales coinciden con sustituciones humano-específicas. Además, dado que la región acumula una gran cantidad de sustituciones chimpancé-específicas, analizamos si se trataba de una región acelerada en chimpancés previamente descriptas (CACNS) (Prabhakar et al., 2006), sin embargo dicho análisis resultó negativo.

TABLA III: Análisis *in silico* de las sustituciones chimpancé-específicas de 2xHAR.321.

HAR	N° cambios	Posición (chr16)	Hs	Pt	A	<i>phyloP</i>
		6,870,383	A	G	C	0.25
		6,870,425	A	T	A	0.76
		6,870,491	G	C	G	1.77
		6,870,497	T	C	T	-0.13
2xHAR.321	8	6,870,647	A	G	A	2.33
		6,870,661	A	C	A	1.14
		6,870,717	G	A	C	0.99
		6,870,785	T	C	T	1.73

Referencias: Hs, *Homo sapiens*. Pt, *Pan troglodytes*. A, ancestral. En azul se destacan las sustituciones del *core* acelerado. En amarillo las sustituciones chimpancé-específicas que coinciden con sustituciones humano-específicas.

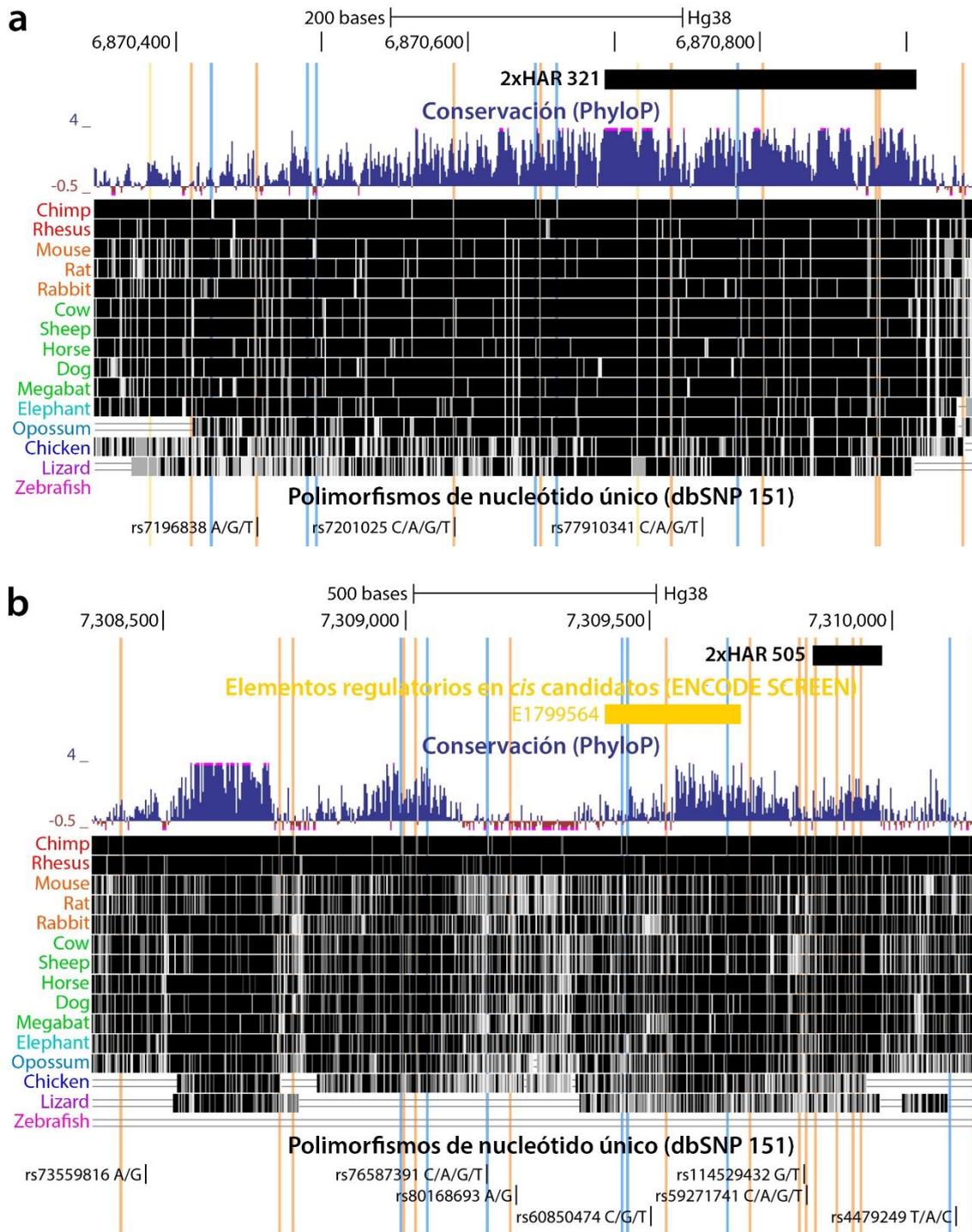


FIGURA 5: Sustituciones humano-específicas de (a) 2xHAR.321 y (b) 2xHAR.505. En naranja se destacan las sustituciones humano-específicas, en celeste las chimpancé-específicas y en amarillo las humano- y chimpancé-específicas. Para cada HAR se incluye el grado de conservación con un alineamiento múltiple, los polimorfismos de nucleótido único y los elementos regulatorios en *cis* candidatos.

Predicción de sitios de unión a factores de transcripción

A continuación, realizamos un análisis *in silico* para evaluar si las sustituciones humano-específicas en 2xHAR.321 y 2xHAR.505 pudieron haber modificado sitios de unión a

FIGURA 6: Análisis *in silico* de TFBS en la secuencia de 2xHAR.321. (a) Predicción de TFBS (JASPAR) para la secuencia humana y de ratón. En naranja se destacan las sustituciones humano-específicas, en celeste las chimpancé-específicas y en amarillo las humano- y chimpancé-específicas. Los TFBS de interés se destacan con un recuadro azul. (b) Detalle de los TFBS de interés. Para cada uno se incluye un alineamiento múltiple con las sustituciones marcadas y el perfil de unión del factor de transcripción.

FOXD3 es un miembro de la familia de factores de transcripción Forkhead que, como su nombre lo indica, se caracteriza por tener un dominio *forkhead* de unión al ADN (Wijchers et al., 2006). Juega un papel importante en el desarrollo del sistema nervioso en vertebrados, especialmente en la determinación, migración y diferenciación de las células de la cresta neural (Dottori et al., 2001; Cheung et al., 2005). Por otro lado, es posible que la secuencia de chimpancé de 2xHAR.321 haya ganado nuevos motivos de unión pero como JASPAR se limita al genoma humano y de ratón, no podemos evaluar los sitios específicos del genoma de chimpancé.

Con respecto a la secuencia de 2xHAR.505, buscamos TFBS que se hayan ganado o perdido exclusivamente en el genoma humano, que podrían explicar la pérdida de función de la secuencia humana en los ensayos *in vivo*. Encontramos dos motivos de unión presentes solamente en el genoma humano, uno para *TRPS1* (*Transcriptional repressor GATA binding 1*) en la posición 7,308,409-7,308,420 y otro para *ZBTB33* (*Zinc finger and BTB domain containing 33*) en la posición 7,308,757-7,308,771 (Figura 7a). Además, también encontramos un motivo de unión para *FOXBI* (*Forkhead box B1*) en la posición 7,309,013-7,309,023 que se encuentra presente en el genoma de chimpancé y de ratón, pero no en el humano (Figura 7a). Las tres regiones tienen una única sustitución humano-específica cada una; T/C en la posición 7,308,414 de *TRPS1*, G/A en la posición 7,308,767 de *ZBTB33* y G/A en la posición 7,309,020 de *FOXBI* (Figura 7b).

Por un lado, *TRPS1* es un represor transcripcional que se une específicamente a las secuencias consenso GATA y reprime la expresión de los genes regulados por la familia de factores de transcripción GATA (Malik et al., 2001; Chang et al., 2002). Se expresa en tejidos adultos del cerebro a niveles bajos, pero, hasta el momento, no se ha demostrado ningún papel en el sistema nervioso. Por otro lado, *ZBTB33*, también denominada *Kaiso*, es un represor transcripcional con actividad bimodal; se une a las islas CpG metiladas y a los motivos consenso *Kaiso* no metilados, y promueve la desacetilación de las histonas y la formación de estructuras represivas de la cromatina (Sun et al., 2021). Los ratones deficientes de *Zbtb33/Kaiso* tienen una reducción significativa en los volúmenes de los ventrículos laterales sin alteraciones significativas en los volúmenes del hipotálamo, estriado, tálamo y cuerpo caloso. Además, presentan varias alteraciones conductuales, como mayor locomoción y exploración en campo abierto (Kulikov et al., 2016), revelando un papel importante de *Zbtb33* en la regulación del sistema nervioso central. Por último, *FOXBI* es otro miembro de la familia de factores de transcripción Forkhead (Wijchers et al., 2006) que tiene un papel fundamental en el desarrollo del sistema nervioso, especialmente en el desarrollo del tálamo, el hipotálamo y la médula espinal (Wehr et al., 1997; Alvarez-Bolado et al., 2000).

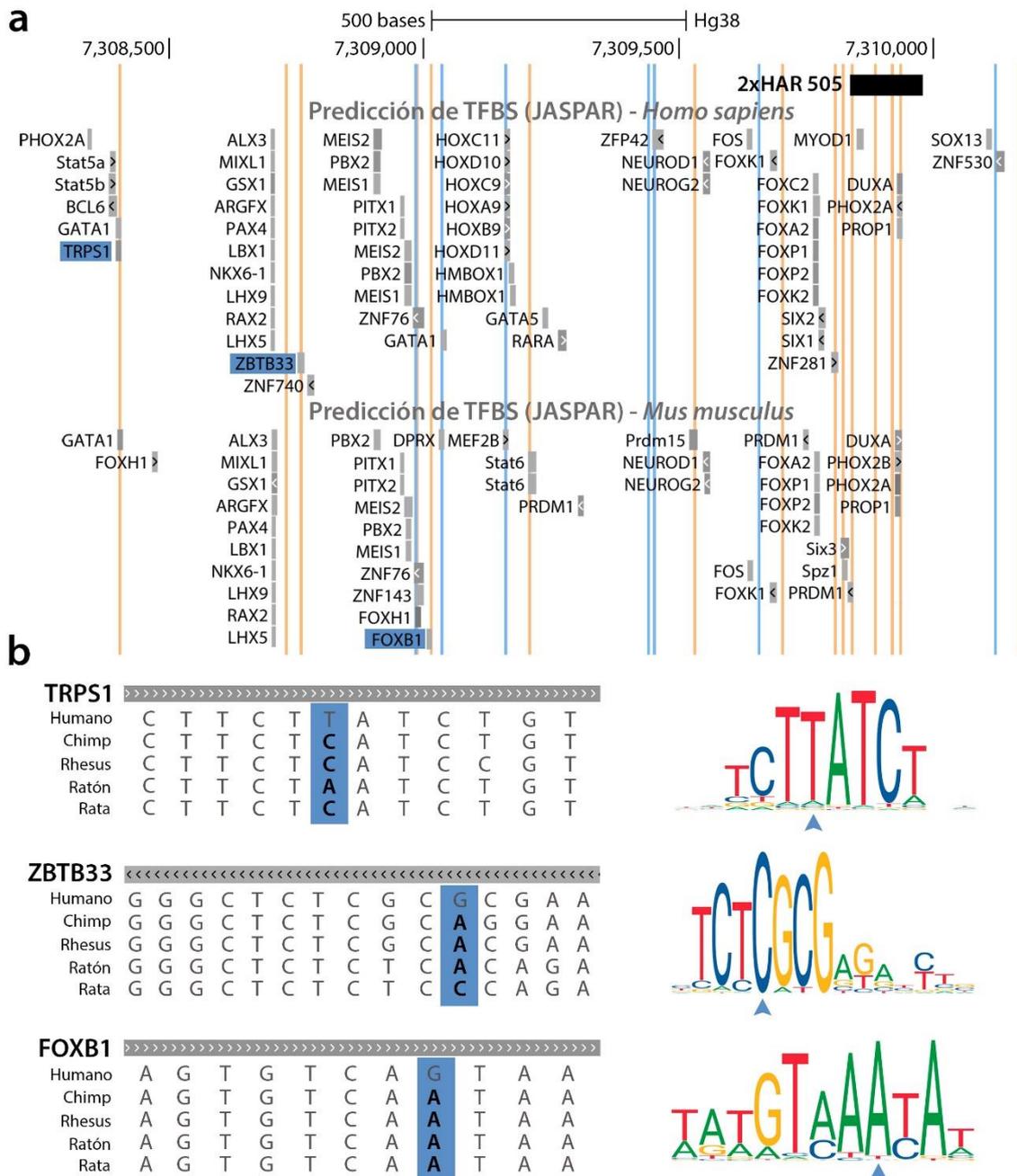


FIGURA 7: Análisis *in silico* de TFBS en la secuencia de 2xHAR.505. (a) Predicción de TFBS (JASPAR) para la secuencia humana y de ratón. En naranja se destacan las sustituciones humano-específicas y en celeste las chimpancé-específicas. Los TFBS de interés tienen un recuadro azul. (b) Detalle de los TFBS de interés. Para cada uno se incluye un alineamiento múltiple con las sustituciones marcadas y el perfil de unión del factor de transcripción.

Selección positiva codificante y no codificante

Una crítica común que surge de la utilización de los HARs como herramienta para identificar secuencias importantes en la evolución humana es que la selección positiva no es la única causa que puede producir señales de aceleración. Por ejemplo, la conversión de genes sesgada por GC (gBGC) es un mecanismo mutagénico causado por la preferencia en las enzimas que reparan el ADN por generar apareamiento entre bases

fuertes (GC) frente a bases débiles (AT) durante la recombinación meiótica (Duret and Galtier, 2009). Por lo tanto, puede producir un incremento en la probabilidad de fijación de los alelos de guanina y citosina, generando un rápido incremento en la tasa de sustitución en las zonas afectadas, independientemente de sus efectos sobre la selección.

Para evaluar este problema, uno de los grupos responsables de la generación de HARs creó un modelo integrado que utiliza una serie de pruebas de probabilidad linaje-específicas para evaluar los alineamientos de secuencias no codificantes en busca de evidencias de selección positiva, gBGC o ambas (Kostka et al., 2012). Haciendo uso de esta herramienta, realizamos una clasificación de los *RBFox1*-HARs y encontramos que la aceleración de 14 HARs (2xHAR.537, ANC.727, ANC.730, ANC.217, 2xHAR.321, HAR.135, 2xHAR.128, HACNS.177, 2xHAR.505, 2xHAR.437, 2xHAR.17, ANC.221, ANC.750 y HACNS.750) puede explicarse por selección positiva, mientras que solamente 3 HARs (ANC.737, 2xHAR.300 y ANC.739) se definen mejor por gBGC. Además, no pudimos analizar 2 HARs (2xHAR.204 y 2xHAR.551) porque tienen *gaps* en los alineamientos y, por ende, no tienen suficientes sitios informativos para ajustar los modelos neutrales.

Por otro lado, utilizando el programa PAML descubrimos que la secuencia codificante de *RBFox1* no experimentó ningún proceso de selección positiva ni en el linaje humano ni en el linaje primate, y en su lugar ha evolucionado bajo una fuerte selección purificadora (Tabla IV). Además, implementamos la misma metodología utilizada para generar la mayoría de los HARs a través del software *PhyloP* (Hubisz et al., 2011), que calcula los puntajes de aceleración y conservación, y no encontramos firmas de aceleración en las secuencias codificantes de *RBFox1*. De esta manera, nuestros resultados sugieren que la acción de la selección positiva podría haber jugado un papel importante en las secuencias no codificantes, pero no en las secuencias codificantes de *RBFox1*.

TABLA IV: Análisis de selección positiva en la secuencia codificante de *RBFox1*.

	Ln	LTR	<i>p</i>	Estimación de parámetros		
Primates						
H _a	-8,950.826958	0.000202	0.988600	<i>p</i> ₀ =0.96976	<i>p</i> ₁ =0.01464	<i>p</i> ₂ = 0.0156
H ₀	-8,950.827059			<i>ω</i> ₀ =0.04319	<i>ω</i> ₁ =1.00000	<i>ω</i> ₂ =1.18058
Humanos						
H _a	-8,950.870412	0.000000	1.000000	<i>p</i> ₀ =0.92425	<i>p</i> ₁ =0.01390	<i>p</i> ₂ = 0.06185
H ₀	-8,950.870412			<i>ω</i> ₀ =0.04330	<i>ω</i> ₁ =1.00000	<i>ω</i> ₂ =3.15537

Referencias: H_a, hipótesis alternativa. H₀, hipótesis nula. Ln, prueba de razón de verosimilitud. LTR = 2 x [Ln(H_a) – Ln(H₀)]. *p*, p-valor de la prueba de chi-cuadrado.

Caracterización de los patrones de expresión de los RBFOX1-HARs

A continuación, decidimos analizar en más detalle los patrones de expresión de los *RBFOX1*-HARs en el sistema nervioso en desarrollo. Pero antes, caracterizamos la expresión endógena de *rbfox1* durante el desarrollo de nuestro modelo animal, el pez cebra, mediante hibridación *in situ* a 24, 48 y 72 hpf, con una sonda que abarca desde el tercer hasta el octavo exón. A las 24 hpf, observamos expresión principalmente en el sistema nervioso, incluyendo el prosencéfalo, mesencéfalo y rombencéfalo, y en la médula espinal (Figura 8a). A las 48 y 72 hpf, continua la expresión en el sistema nervioso y fuera de este, en los ojos, corazón y, particularmente, en los arcos branquiales y aletas pectorales (Figura 8b y 8c). Además, en criosecciones coronales a 72 hpf observamos expresión en el tectum óptico, tálamo, retina y médula oblongada (Figura 8d y 8e).

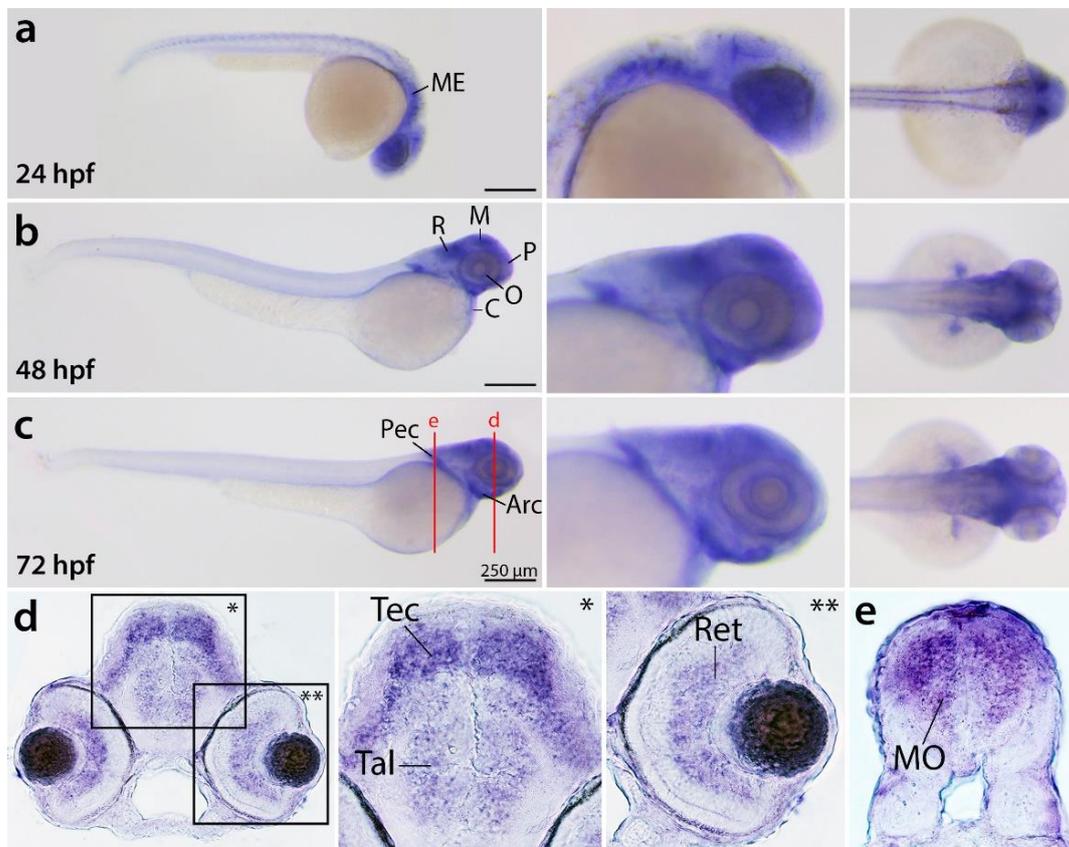


FIGURA 8: Patrón de expresión endógeno de *rbfox1*. Hibridación *in situ* en embriones de peces cebra a (a) 24, (b) 48 y (c) 72 hpf con vistas laterales y dorsales, y (d y e) en criosecciones coronales de 72 hpf. Referencias: P, prosencéfalo. M, mesencéfalo. R, rombencéfalo. ME, médula espinal. O, ojo. C, corazón. Pec, aleta pectoral. Arc, arcos branquiales. Tec, tectum óptico. Tal, tálamo. Ret, retina. MO, médula oblongada.

A partir de estos resultados, seleccionamos dos HARs activos en el sistema nervioso, 2xHAR.321 y 2xHAR.17, para caracterizar en detalle su expresión mediante doble inmunofluorescencia a 72 hpf en criosecciones coronales (Figura 9 y 10). Para tal fin utilizamos un anticuerpo policlonal anti-*Rbfox1* y un anticuerpo anti-EGFP validado en peces cebra con anterioridad en el laboratorio. El anticuerpo de *Rbfox1* fue diseñado contra los epítopes de ratón, pero en el laboratorio observamos que con frecuencia los

epítopes se encuentran conservados, y es posible que el anticuerpo también funcione en peces cebra. De esta manera, mediante microscopía confocal observamos que ambos HARs impulsan la expresión de EGFP a diferentes dominios anatómicos del sistema nervioso en desarrollo del pez cebra, incluyendo el tectum óptico (Figura 9c y 10c) y el tálamo (Figura 9d y 10d), así como también la retina (Figura 9e y 10e) y la médula oblongada (Figura 9f y 10f). Además, *rbfox1* se expresa ampliamente en todo el sistema nervioso y de manera nuclear.

Luego cuantificamos las células doblemente marcadas del total de células positivas para EGFP y observamos que ambos *enhancers* son capaces de impulsar la expresión del reportero a un alto porcentaje de células que expresan *rbfox1*, mayor al 80% (Figura 9g y 10g), lo que sugiere que contienen intrínsecamente la información necesaria para ser reconocidos por la maquinaria transcripcional en *trans* del pez cebra. Por otro lado, del total de células marcadas con *rbfox1*, solo una parte también expresa EGFP, evidenciando que existen otros elementos que regulan su expresión. Estos resultados son consistentes con resultados previos (Thisse and Thisse, 2014), y validan el uso de peces cebra como modelo animal para estudiar la expresión de potenciales elementos regulatorios de *RBFOX1*.

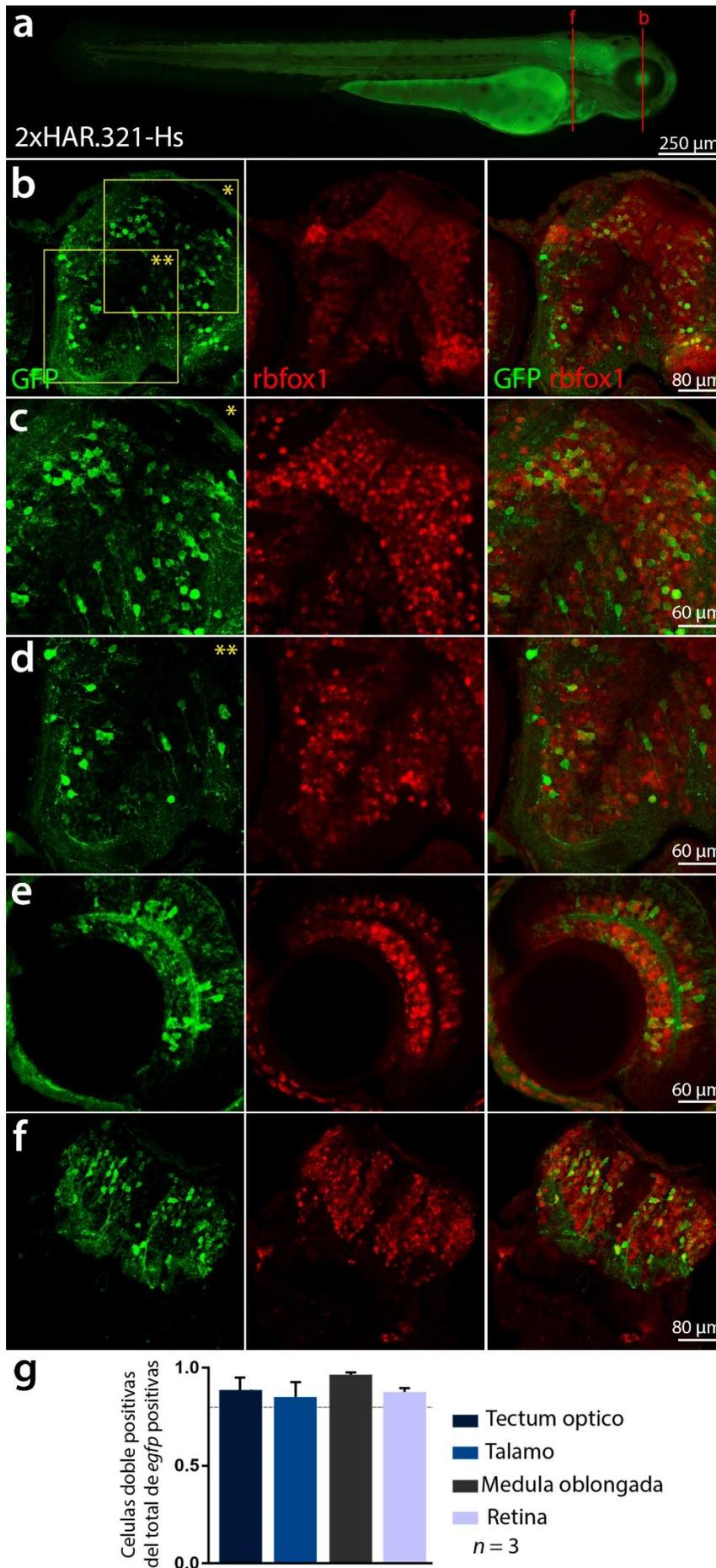


FIGURA 9:
Caracterización de la expresión dirigida por 2xHAR.321. (a) Imagen representativa del patrón de expresión de la secuencia humana (Hs, *Homo sapiens*) de 2xHAR.321 a 72 hpf. (b-f) Imágenes de microscopía confocal de la doble inmunofluorescencia en criosecciones coronales contra *rbfox1* y EGFP, haciendo énfasis en el (c) tectum óptico, (d) tálamo, (e) retina y (f) médula oblongada. (g) Cuantificación de las células doblemente marcadas del total de células marcadas con EGFP.

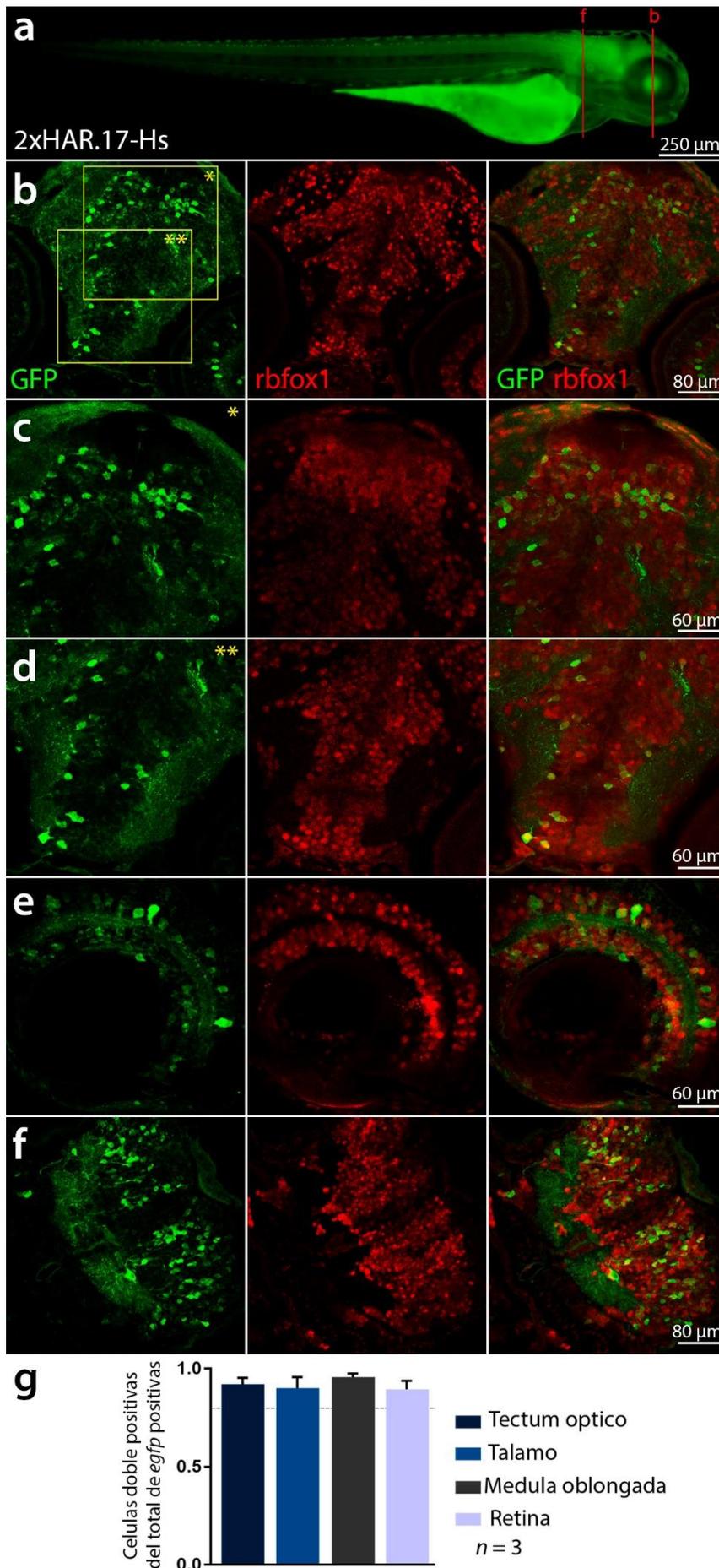


FIGURA 10:
Caracterización de la expresión dirigida por 2xHAR.17. (a) Imagen representativa del patrón de expresión de la secuencia humana (Hs, *Homo sapiens*) de 2xHAR.17 a 72 hpf. (b-f) Imágenes de microscopía confocal de la doble inmunofluorescencia en crio secciones coronales contra *rbfox1* y EGFP, haciendo énfasis en el (c) tectum óptico, (d) tálamo, (e) retina y (f) médula oblongada. (g) Cuantificación de las células doblemente marcadas del total de células marcadas con EGFP.

Análisis de los promotores de RBFOX1 en peces cebra transgénicos

Los ensayos publicados en peces cebra transgénicos que estudian la actividad *enhancer* de secuencias candidatas emplean una variedad de promotores mínimos, y aún no se ha aceptado ningún promotor mínimo como estándar. Además, teniendo en cuenta que los *enhancers* tienen mayor compatibilidad con ciertos promotores (Zabidi et al., 2015), decidimos evaluar la expresión de los promotores alternativos de *RBFOX1* con el objetivo de encontrar un promotor mínimo adecuado para futuras construcciones reporteras.

Para ello, primero examinamos los primeros exones en busca de evidencia de promotores utilizando una variedad de bases de datos disponibles públicamente. Estas incluyen islas CpG, sitios de inicio de la transcripción determinados por CAGE de FANTOM5 (Abugessaisa et al., 2017), marcas de histonas características de promotores del proyecto Roadmap y elementos regulatorios en *cis* candidatos (CREs) de ENCODE (ENCODE Project Consortium et al., 2020). Este último clasifica elementos regulatorios candidatos en base a la accesibilidad del ADN y marcas epigenéticas. En particular, los elementos candidatos tipo promotores se caracterizan por tener señales altas de DNAsa I y H3K4me3, y caen dentro de 200 pb de un TSS anotado por GENCODE. Además, la base de datos de promotores eucariotas (EPD) es una colección de elementos regulatorios validados experimentalmente (Meylan et al., 2020).

De esta manera, observamos que, excepto el primer exón correspondiente a la quinta isoforma, el resto tiene al menos una de las evidencias mencionadas anteriormente (Figura 11a). En particular, los primeros exones E1B y E1C/E1C.1 tienen perfiles de CAGE y señales de H3K4me3 exclusivamente en el cerebro, mientras que en E1E se limitan al músculo esquelético y corazón (Figura 11b). Además, en el primer exón E1D.1 los niveles de CAGE en cerebro son menores en comparación con los mencionados anteriormente, y se observan señales de H3K4me3 tanto en cerebro como en corazón, y en menor medida en músculo esquelético (Figura 11b). Por último, los recientemente identificados primeros exones E1A y E1F son los que menos evidencias poseen. Esta primera aproximación coincide con la información disponible de las secuencias de ratón (Damianov and Black, 2010; Casanovas et al., 2020), sugiriendo que los patrones de expresión de los promotores de *RBFOX1* se encuentran conservados.

A continuación, clonamos las secuencias candidatas en un vector de expresión basado en el sistema Tol2 río arriba del gen reportero EGFP, es decir sin el promotor mínimo de *cFos*, y generamos peces cebra transgénicos, con el objetivo de evaluar la actividad autónoma de cada potencial promotor. Inyectamos todas las secuencias candidatas, pero hasta el momento solamente analizamos la expresión de los promotores E1B y E1E (para todas las líneas transgénicas ver Apéndice II, Figura S2). Observamos que el promotor E1B se expresa principalmente en el sistema nervioso, incluidos el prosencéfalo, mesencéfalo y rombencéfalo, y la médula espinal. Por fuera de este, también se expresa en el ojo, corazón, arcos branquiales, aleta pectoral y músculo esquelético (Figura 11c). En conjunto, la expresión de la secuencia E1B es prácticamente ubicua en las tres edades analizadas. En cambio, la expresión del promotor E1E está restringida únicamente al músculo esquelético y corazón (Figura 11c). De esta manera, los patrones de expresión

observados *in vivo* reflejan la información obtenida a partir de las bases de datos disponibles públicamente.

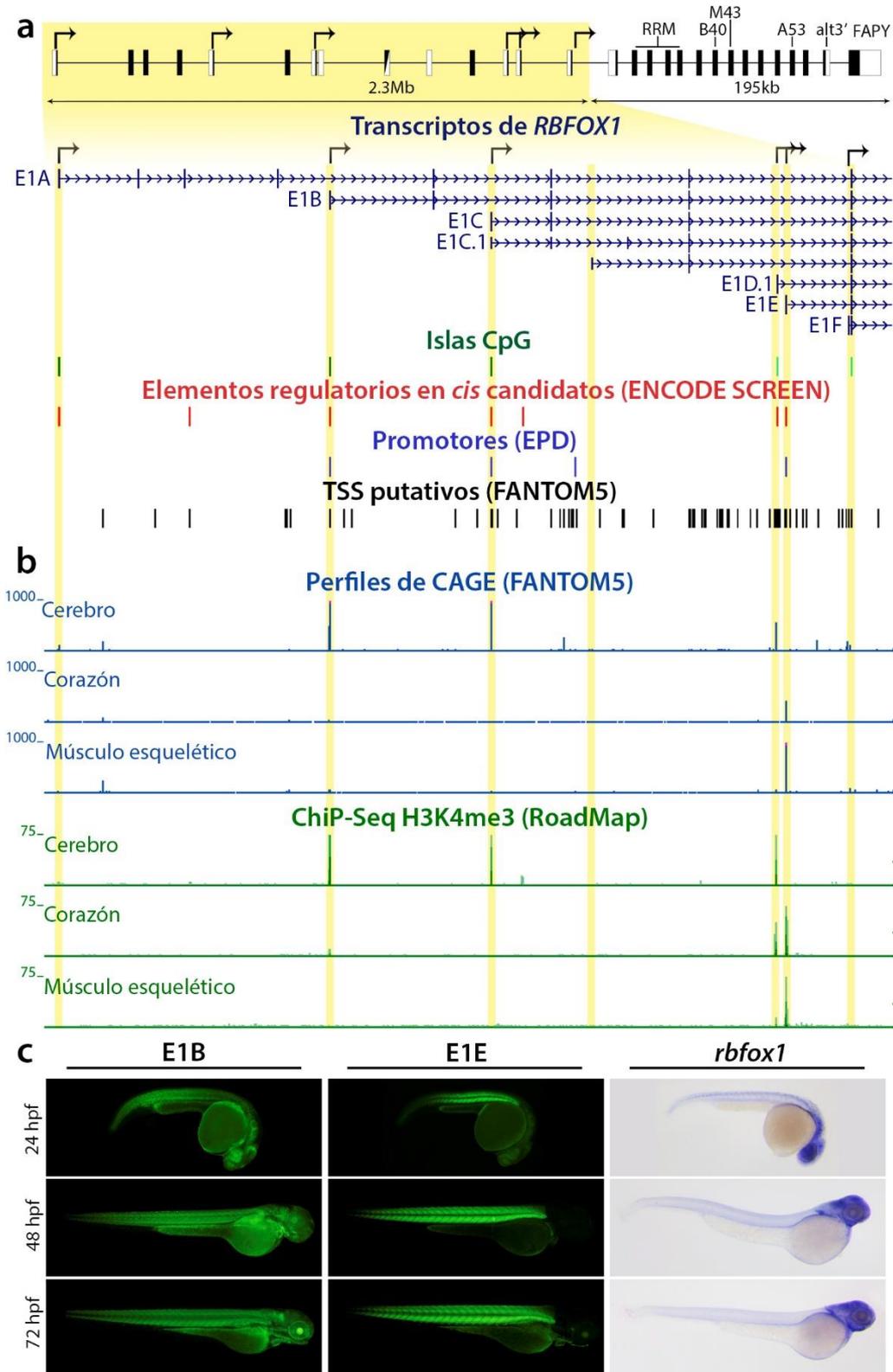


FIGURA 11: Análisis de promotores de *RBFOX1*. (a) Esquema de la estructura génica de *RBFOX1* en el genoma humano (Hg38). Las cajas blancas representan los exones no codificantes y las negras los codificantes. Cada inicio de la transcripción se indica con una flecha. Modificado de (Damianov and Black, 2010). Para cada primer exón se incluyen evidencias de promotores de

distintas bases de datos disponibles, incluyendo islas CpG, CREs de ENCODE, promotores de EPD y mapeo de TSSs de FANTOM5. (b) Perfiles de CAGE de FANTOM5 y señales de ChIP-Seq de H3K4me3 de Roadmap de cerebro, corazón y músculo esquelético. (c) Ensayo de expresión en líneas estables de peces cebra transgénicos de los primeros exones E1B y E1E a 24, 48 y 72 hpf y expresión endógena de *rbfox1* en peces cebra de tipo salvaje.

A partir de estos resultados, seleccionamos a la secuencia del primer exón E1B de *RBFOX1* (*RBFOX1*-E1B) para realizar un análisis exhaustivo. En primer lugar, observamos que la región coincide con una isla CpG, con un CRE tipo promotor (E1799220) y con un promotor validado mediante mapeo de TSSs (*RBFOX1*-2). Además, los análisis de CAGE muestran que posee múltiples TSSs estrechamente espaciados que se usan con una frecuencia similar, y que está marcado bivalentemente tanto con la marca de H3K4me3 como con la modificación de la cromatina represiva, H3K27me3 (Figura 12a). En base a estas características, podemos clasificarlo como un promotor del tercer tipo, típico de los factores de transcripción del desarrollo.

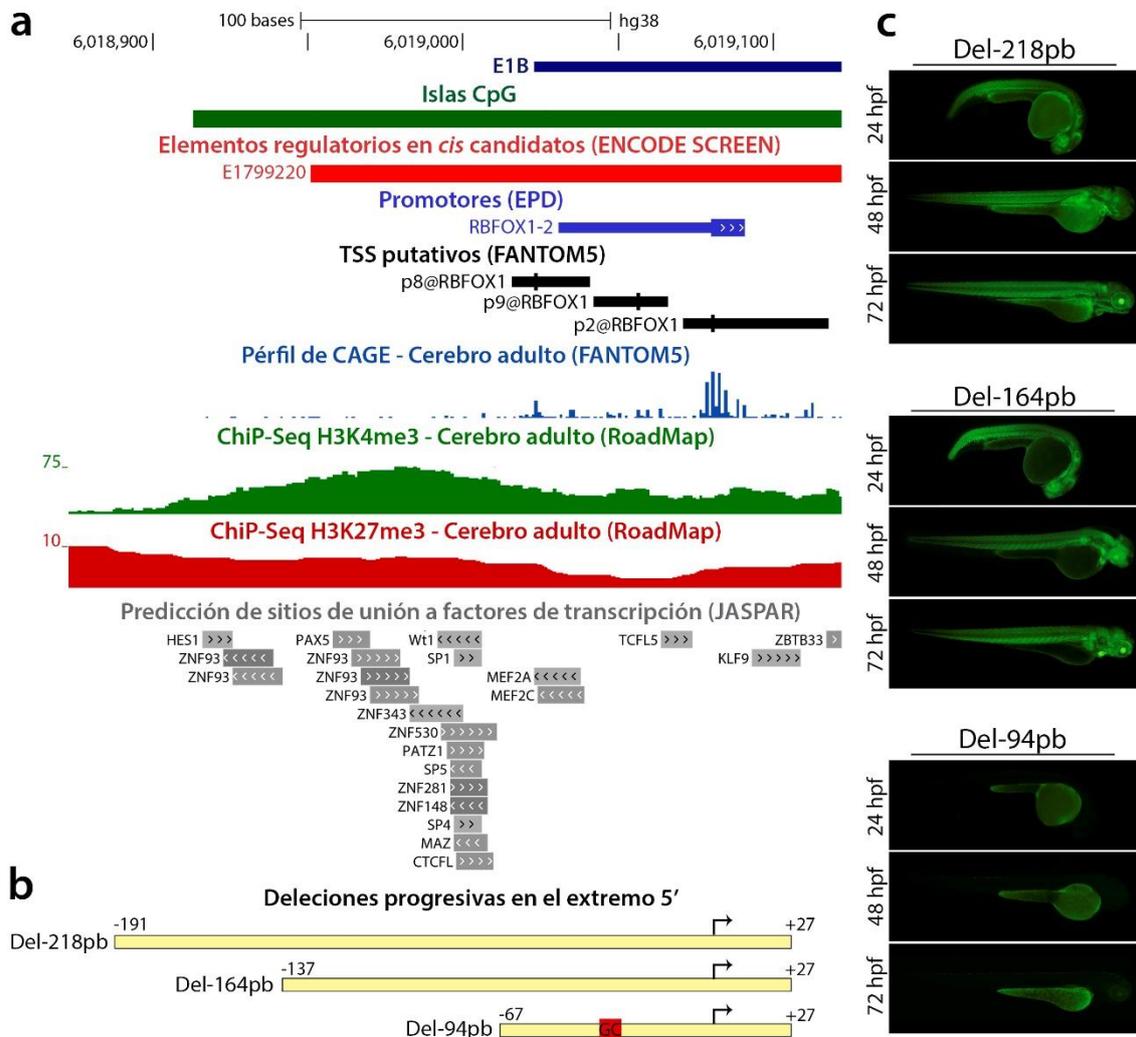


FIGURA 12: Análisis exhaustivo del primer exón *RBFOX1*-E1B. (a) Región genómica de *RBFOX1*-E1B, junto con evidencias de promotores de distintas bases de datos disponibles, incluyendo islas CpG, CREs (ENCODE), promotores (EPD), perfiles de CAGE y mapeo de TSSs (FANTOM5), marcas de histonas (Roadmap) y predicción de sitios de unión a factores de

transcripción (JASPAR). (b) Deleciones progresivas en el extremo 5'. La posición -35 posee una caja GC. (c) Ensayo de expresión en líneas estables de peces cebras transgénicos de las deleciones progresivas de *RBFOX1*-E1B a 24, 48 y 72 hpf.

Por otro lado, observamos que presenta varios sitios de unión a factores de transcripción predichos por JASPAR interesantes (Figura 12a). Por ejemplo, los factores de la familia SP se unen a motivos ricos en GC en las islas CpG, y los factores de la familia KLF se unen a las cajas GC en los promotores (Kaczynski et al., 2003). Además, los factores de la familia MEF2 (*Myocyte enhancer factor-2*) son activadores transcripcionales de la morfogénesis y miogénesis cardíaca (Edmondson et al., 1994; Naya and Olson, 1999), pero no solo intervienen en el desarrollo del músculo esquelético y cardíaco, sino también en la diferenciación y supervivencia neuronal (Heidenreich and Linseman, 2004). Mutaciones y deleciones en los genes de la familia MEF2 se han asociado con discapacidad cognitiva grave, epilepsia y malformación cerebral (Assali et al., 2019). Asimismo, *Rbfox1* regula el *splicing* alternativo de miembros de la familia de factores MEF2, y contribuye a un fenotipo de insuficiencia cardíaca mediante el *splicing* aberrante de un exón alternativo clave de *Mef2a* (Gao et al., 2016).

Para ser utilizados en ensayos reporteros, los promotores mínimos requieren poca o ninguna expresión ectópica para obtener resultados concluyentes. Es por ello que generamos una serie de reporteros con deleciones progresivas en el extremo 5' con el objetivo de determinar la pérdida de su función autónoma (Figura 12b). Los resultados indicaron que la región mínima del promotor se ubica entre las posiciones -67 a +27 (Figura 12c). Dicha región no posee ninguna caja TATA o CCAAT río arriba del TSS, pero sí una caja GC (p-valor = 0.01) en la posición -35.

Por último, intercambiamos el promotor mínimo de ratón *cFos* en el vector de expresión basado en el sistema Tol2 por el promotor mínimo *RBFOX1*-E1B y clonamos un par de *enhancers* para examinar la respuesta del promotor mínimo y la compatibilidad de los *enhancers* hacia el promotor. Como *enhancers* seleccionamos dos *RBFOX1*-HARs activos en distintos tejidos en desarrollo, 2xHAR.321 del sistema nervioso y 2xHAR.128 del músculo esquelético y corazón, para evaluar la respuesta del promotor en varios tejidos. De este modo, no solo observamos que la actividad *enhancer* con el promotor mínimo E1B es reproducible, sino que los patrones de expresión son prácticamente indistintos de los observados con el promotor mínimo de *cFos* (Figura 13). La única diferencia es que la expresión del reportero con el promotor mínimo E1B en el músculo esquelético parece ser más sólida. En conclusión, encontramos la región mínima del promotor *RBFOX1*-E1B, que no posee expresión ectópica por sí sola, pero actúa adecuadamente en respuesta a la presencia de *enhancers* tejido-específicos. De esta manera, podría utilizarse como promotor mínimo en futuras construcciones reporteras para analizar específicamente *enhancers* de *RBFOX1*. A su vez, estos resultados validan el uso del promotor mínimo de *cFos* en los ensayos de peces cebras transgénicos, ya que las diferencias de expresión entre los dos promotores no son drásticas.

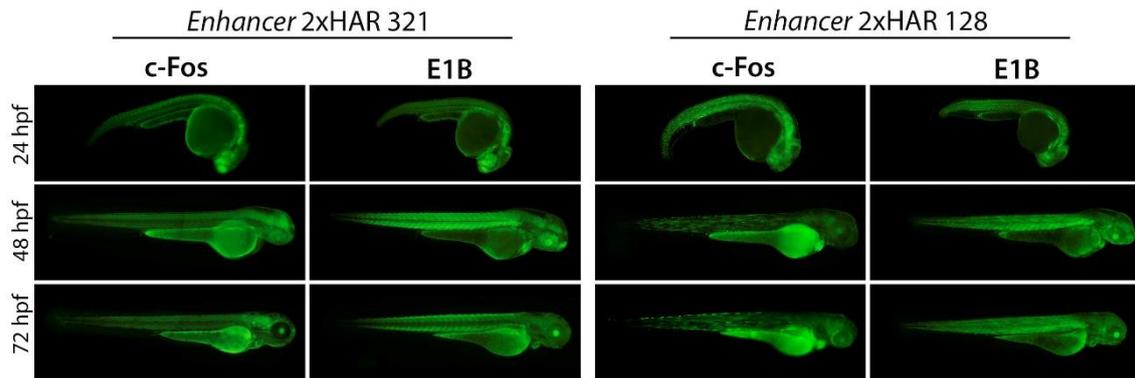


FIGURA 13: Respuesta del promotor mínimo *RBFOX1*-E1B a *enhancers*. Comparación de la expresión dirigida por un *enhancer* del sistema nervioso (2xHAR.321) y por un *enhancer* de músculo esquelético y corazón (2xHAR.128), con el promotor mínimo de ratón *cFos* y el promotor mínimo de E1B a 24, 48 y 72 hpf.

Discusión de los resultados

Ensayos funcionales en peces cebra transgénicos

En este capítulo postulamos que las regiones aceleradas humanas en *RBFOX1* tienen la capacidad de actuar como *enhancers* transcripcionales y que el proceso evolutivo ocurrido en el linaje humano resultó en diferencias funcionales en la actividad de las secuencias humanas. Para comprobar esta hipótesis, analizamos la actividad *enhancer* de manera comparada entre la secuencia humana y su ortóloga de chimpancé de los 19 *RBFOX1*-HARs en ensayos de expresión de gen reportero en peces cebra transgénicos.

Con este enfoque, descubrimos que la mayoría de los HARs fue capaz de generar un patrón de expresión distintivo y reproducible del reportero durante los estadios analizados. Los resultados más interesantes para nuestro objetivo fueron 10 *RBFOX1*-HARs que exhibieron expresión diferencial entre las secuencias humanas y de chimpancé; seis HARs (2xHAR.537, ANC.727, 2xHAR.321, 2xHAR.128, 2xHAR.17 y ANC.750) en los que la secuencia humana era activa y la de chimpancé inactiva, y cuatro HARs (ANC.730, ANC.739, 2xHAR.505 y 2xHAR.437) en los que la secuencia de chimpancé era activa y la humana inactiva. La mayoría resultaron ser *enhancers* fuertes del sistema nervioso en desarrollo, aunque también observamos expresión en otros tejidos, como músculo esquelético y corazón. Por otro lado, los patrones de expresión de cinco *RBFOX1*-HARs (ANC.217, ANC.221, 2xHAR.204, 2xHAR.551 y HACNS.705) fueron indistinguibles entre las secuencias humanas y de chimpancé, lo que sugiere que son regiones con funciones altamente conservadas.

Además, en cuatro *RBFOX1*-HARs (ANC.737, 2xHAR.300, HAR.135 y HACNS.177) ninguna de las dos especies fue capaz de dirigir la expresión del reportero. Hay varias razones que pueden explicar la ausencia de expresión observada en estas regiones. La más parsimoniosa es que la secuencia no contiene actividad *enhancer*, pudiendo o no albergar otras actividades regulatorias hasta ahora no analizadas, como *silencers* o *insulators*. También puede ser que la secuencia sea un *enhancer* pero que sea activo en

una etapa del desarrollo que no fue analizada. Por último, el sistema utilizado podría no ser el más adecuado para evaluar estas secuencias, es decir que la secuencia no puede actuar como *enhancer* porque la especie modelo en la que lo evaluamos carece de los factores de transcripción necesarios para interpretar la información en *cis*, o tienen diferentes patrones de expresión. Por lo tanto, los resultados negativos en los ensayos de gen reportero deben interpretarse con cuidado.

En resumen, descubrimos que 15 de los 19 *RBFOX1*-HARs son *enhancers* activos en peces cebra transgénicos y, en última instancia, que 10 HARs exhiben una expresión diferencial en comparación con sus ortólogos de chimpancé posiblemente como consecuencia del proceso evolutivo de aceleración. De los *enhancers* con expresión diferencial, 6 son potenciales ganancias de función y 4 potenciales pérdidas de función en el linaje humano. Estos resultados son consistentes con la hipótesis de que las regiones no codificantes de *RBFOX1* fueron blanco de una evolución acelerada a nivel regulatorio en el linaje humano, que tuvo como consecuencia la ganancia y pérdida de nuevas funciones regulatorias. Si estos resultaron en cambios en el patrón de expresión de *RBFOX1* en el linaje humano, confiriéndole algún rol en la evolución de nuestro cerebro, aún resta por ser investigado.

Por otra parte, es interesante destacar que la mayoría de los HARs analizados en las líneas de peces cebra transgénicos tienen patrones de expresión amplios que abarcan varios dominios anatómicos, entre los que se incluyen el cerebro, la médula espinal, el corazón y, en menor medida, el músculo esquelético. Todos ellos son lugares endógenos de expresión de *rbfox1* en peces cebra. Llamativamente, los dominios de expresión de estos elementos son altamente redundantes, ya que cada uno de los dominios de expresión mencionados anteriormente está representado por varios HARs. Existen múltiples ejemplos de genes, especialmente de aquellos activos durante el desarrollo, que tienen múltiples *enhancers* con actividades similares (Frankel et al., 2010). Sin embargo, la razón de esta aparente redundancia es aún material de estudio. Una de las hipótesis actuales sugiere que la superposición en los dominios de expresión sería necesaria para la canalización del fenotipo en aquellos estadios críticos del desarrollo, permitiendo amortiguar el impacto de cambios ambientales y confiriendo robustez a los fenotipos (Frankel et al., 2010; Lam et al., 2015; Osterwalder et al., 2018). Se requieren estudios adicionales para comprobar si este puede ser el rol que estén cumpliendo los múltiples *RBFOX1*-HARs con actividad redundante.

Elección del modelo experimental

Como cualquier modelo experimental, el uso de peces cebra para analizar *enhancers in vivo* tiene sus ventajas y desventajas. Por un lado, la generación de líneas reporteras transgénicas es un método eficiente para realizar una detección rápida de *enhancers* a gran escala permitiendo a su vez analizar los patrones de expresión espaciotemporales en el organismo completo durante múltiples etapas del desarrollo (Fisher et al., 2006b; Kawakami, 2007). Además, durante el desarrollo se puede inhibir la pigmentación de los peces fácilmente con el agregado de feniltiourea, permitiendo una clara observación de los tejidos internos, especialmente del sistema nervioso (Karlsson et al., 2001). Por otra

parte, la mayoría de los métodos actuales para la generación de reporteros *in vivo* implican la inserción al azar de los transgenes en el genoma del modelo experimental, exigiendo la necesidad de varias líneas independientes para evitar la influencia de los efectos posicionales. En el caso de los peces cebra, se pueden obtener varias líneas transgénicas con una sola inyección y en un lapso de solo 3 meses. También se ha comprobado que, aunque las secuencias candidatas no se encuentren conservadas en el genoma del pez cebra, aún son capaces de dirigir la expresión a tejidos específicos (Fisher et al., 2006a; Domené et al., 2013), sugiriendo que la información funcional se conserva a niveles por debajo del radar de alineamiento de secuencias genómicas a gran escala.

Sin embargo, también es necesario tener en cuenta las limitaciones de este modelo. Si bien existen trabajos que respaldan esta metodología (Kamm et al., 2013b; Oksenberg et al., 2013; Caporale et al., 2019), no se pueden descartar los cambios en *trans* ocurridos en el linaje de los peces cebra desde la divergencia con el linaje humano hace aproximadamente 450 Ma, es decir, a nivel de las proteínas encargadas de decodificar la información presente en los *enhancers* (Ritter et al., 2012). Se podrían emplear ensayos adicionales en otros modelos animales filogenéticamente más cercanos a la especie humana, como ratones transgénicos (Visel et al., 2007), pero este modelo funciona mejor cuando se evalúan pocas secuencias candidatas, ya que técnicamente es más difícil y costoso producir ratones transgénicos. Además, en la cuantificación de los ensayos de doble inmunofluorescencia observamos que dos *RBFOX1*-HARs fueron capaces de impulsar la expresión del reportero a un alto porcentaje de células que expresan *rbfox1*, lo que sugiere que contienen intrínsecamente la información necesaria para ser reconocidos por la maquinaria transcripcional en *trans* del pez cebra.

En definitiva, el uso de peces cebra es un método eficiente para la detección rápida de *enhancers* a gran escala, pero se debe mantener cierta cautela a la hora de extrapolar los resultados funcionales de los elementos regulatorios humanos, principalmente a nivel de los dominios de expresión de los *enhancers*. Estudios realizados comparando la actividad de *enhancers* humanos cuando se caracterizan en ratones o peces cebra transgénicos, mostraron que en aproximadamente 38% de los casos, los resultados obtenidos en peces cebra, a nivel de los dominios de expresión, no pudieron ser reproducidos en el modelo mamífero (Ritter et al., 2012). Más recientemente, estudios realizados comparando el patrón de expresión de *enhancers* humanos en líneas de ratones y peces transgénicos también reportaron discrepancias entre estos modelos (Ariza-Cosano et al., 2012).

Necesidad de una tercera especie

En los ensayos de gen reportero en peces cebra observamos 10 *RBFOX1*-HARs con expresión diferencial entre las secuencias humanas y de chimpancé, de las cuales 6 son potenciales ganancias de función y 4 son potenciales pérdidas de función en el linaje humano. Para confirmar si efectivamente son ganancias o pérdidas de función, es necesario tener evidencia adicional proveniente de una tercera especie cercanamente emparentada para excluir la posibilidad de que el patrón de expresión observado para la secuencia de chimpancé sea el estado derivado. Entonces, para dos HARs, uno con potencial ganancia de función (2xHAR.321) y otro con potencial pérdida de función

(2xHAR.505) en el linaje humano, decidimos extender el análisis a la secuencia ortóloga de ratón. Para el caso de 2xHAR.321 esperábamos que la secuencia de ratón sea incapaz de dirigir la expresión del reportero, al igual que la secuencia de chimpancé. Sin embargo, observamos una fuerte expresión del reportero en los mismos dominios dirigidos por la secuencia humana, lo que sugiere que el estado derivado no es el patrón de expresión dirigido por la secuencia humana, sino el de la secuencia de chimpancé. Este resultado reafirma la necesidad de evaluar una tercera especie en los análisis entre ortólogos. En este sentido, el trabajo en el laboratorio que hizo Delia Christ y está continuando Abril Sileo con el análisis de la secuencia ortóloga de macaco Rhesus, va a ser clave para la correcta interpretación de los resultados de dichas regiones, ya que no se puede descartar la posibilidad de que en el resto de los HARs con expresión diferencial, el estado derivado sea el de chimpancé.

Comparación con ensayos de MPRA

En los últimos años, los ensayos reporteros paralelos masivos (MPRA) han aumentado enormemente el rendimiento de la caracterización funcional de los HARs. Miles de construcciones reporteras, cada una con un código de barra único y un *enhancer* candidato, se analizan juntas en líneas celulares a través de secuenciación masiva del ARN (Inoue and Ahituv, 2015). A diferencia de los ensayos reporteros en animales transgénicos, la secuenciación en los ensayos de MPRA permite la cuantificación de la abundancia de cada constructo, pero carece de información espaciotemporal. Recientemente, se han publicado tres estudios independientes que comparan la actividad entre la secuencia humana y de chimpancé de los HARs (Girskis et al., 2021; Uebbing et al., 2021; Whalen et al., 2023). Motivados por el enriquecimiento de los HARs en genes relacionados con el desarrollo neuronal, todos los estudios utilizan líneas celulares neuronales, por lo que aún no tenemos una comprensión completa de la actividad de los HARs en otros tipos celulares y etapas del desarrollo. Por otro lado, dada la heterogeneidad de los enfoques de cada estudio, es decir, la elección del subconjunto de HARs, vectores, estrategias de *delivery*, líneas celulares, herramientas de análisis y umbrales estadísticos, la consistencia de los resultados entre los estudios es moderada.

A su vez, muy pocos de los resultados obtenidos en este trabajo coinciden con los resultados obtenidos en los ensayos de MPRA (Tabla V). Por un lado, las diferencias en los resultados de los HARs que fueron anotados como inactivos en MPRA y que en este trabajo observamos actividad en al menos una especie (ANC.217, ANC.739, ANC.221, 2xHAR.204, 2xHAR.551 y HACNS.705), se pueden deber al tamaño de las secuencias utilizadas. En este trabajo de tesis utilizamos toda la región conservada que contiene a cada HAR (*PhastCons*), ya que consideramos que probablemente sea más representativa de una potencial unidad funcional. En este caso, los tamaños de las secuencias analizadas tenían un promedio de 560 pb. En cambio, debido a las limitaciones en el tamaño en los ensayos estándar de MPRA, dos trabajos utilizaron únicamente el *core* acelerado (Uebbing et al., 2021; Whalen et al., 2023). Únicamente en un trabajo aplicaron una variación llamada CaptureMPRA, que permite evaluar secuencias de hasta 500 pb (Girskis et al., 2021). Por otro lado, los resultados que muestran mayor actividad en una

especie que en la otra en los estudios de MPRA y que en nuestro trabajo observamos igual expresión en ambas especies (ANC.750 y HACNS.705), podrían deberse a que los ensayos de MPRA son cuantitativos. Asimismo, varios *RBFox1*-HARs fueron anotados como activos en al menos una especie en los ensayos de MPRA, pero en este trabajo no observamos expresión (ANC.737, HAR.135 y HACNS.177). Estas diferencias posiblemente se deban a las limitaciones del modelo animal (ver Discusión, Elección del modelo experimental). Por último, algunos resultados obtenidos en este trabajo son totalmente diferentes a los de los ensayos de MPRA y, a su vez, tampoco coinciden entre ellos. Estas discrepancias pueden deberse a una variedad de razones, la mayoría de las cuales tienen que ver con los enfoques de cada trabajo. Por ejemplo, en el trabajo de Whalen et al. utilizan lentivirus como estrategia de *delivery* de la librería de MPRA en las células, de manera que los vectores se integran al azar en el genoma, al igual que la transgénesis en peces cebras. En cambio, en los otros dos trabajos, los vectores no se integran en el genoma, y la secuenciación se realiza sobre el ADN plasmídico. Además, como ya mencionamos, debido a las limitaciones en el tamaño de las secuencias que se pueden introducir en los vectores de MPRA, únicamente en el trabajo de Girsakis et al. se analizaron las secuencias completas de los HARs. Esto podría explicar porque en los otros dos trabajos hay tantos HARs aparentemente inactivos. Finalmente, posiblemente la mayor variación entre resultados se deba a los tipos celulares utilizados. En el trabajo de Girsakis et al. utilizaron la sublínea humana SH-SY5Y derivada de la línea celular de neuroblastoma SK-N-SH, y la línea celular murina Neuro2A (N2A). En el trabajo de Uebbing et al. usaron células madre neuronales (NSCs) derivadas de la línea celular humana H9. Y en el trabajo de Whalen et al. generaron células progenitoras neuronales (NPCs) a partir de células madre pluripotentes inducidas (iPSCs) de dos individuos humanos y dos individuos chimpancés. Aunque es indudable que utilizar células humanas para interrogar *enhancers* humanos tiene la ventaja de contar con la maquinaria transcripcional en *trans* necesaria para reconocer la información contenida en los *enhancers*, también hay que tener en cuenta las ventajas de los ensayos *in vivo* frente a los *in vitro*. Mientras que *in vitro* la medición de la actividad se realiza en un único tipo celular a un sólo estadio, *in vivo* es posible evaluar la actividad a distintos estadios en organismos embrionarios enteros, obteniendo una representación espaciotemporal completa de la actividad.

TABLA V: Comparación de los resultados con ensayos de MPRA.

HAR	Peces cebras	Girsakis et al.	Uebbing et al.	Whalen et al.
		3,171 HARs* CaptureMPRA 500 pb SH-SY5Y N2A	1,363 HARs** MPRA estándar 137 pb NSCs derivadas de H9	714 HARs*** lentiMPRA 171 pb NPCs derivadas de iPSCs
2xHAR.537	Activo solo Hs	Mayor actividad Pt	Inactivo	Inactivo
ANC.727	Activo solo Hs	Igual actividad Hs/Pt	NA	NA
ANC.730	Activo solo Pt	Mayor actividad Hs	NA	NA

ANC.737	Inactivo	Mayor actividad Pt	NA	NA
ANC.217	Activo Hs/Pt	Inactivo	NA	NA
2xHAR.321	Activo solo Hs	Igual actividad Hs/Pt	Inactivo	Inactivo
2xHAR.300	Inactivo	Inactivo	Inactivo	Igual actividad Hs/Pt
ANC.739	Activo solo Pt	Inactivo	NA	NA
HAR.135	Inactivo	Mayor actividad Pt	NA	NA
2xHAR.128	Activo solo Hs	Igual actividad Hs/Pt	Inactivo	Inactivo
HACNS.177	Inactivo	Mayor actividad Hs	Activo en Hs	NA
2xHAR.505	Activo solo Pt	Mayor actividad Hs	Inactivo	Inactivo
2xHAR.437	Activo solo Pt	Igual actividad Hs/Pt	Inactivo	Inactivo
2xHAR.17	Activo solo Hs	Igual actividad Hs/Pt	Inactivo	Inactivo
ANC.221	Activo Hs/Pt	Inactivo	NA	NA
ANC.750	Activo solo Hs	Mayor actividad Hs	NA	NA
2xHAR.204	Activo Hs/Pt	Inactivo	Inactivo	Inactivo
2xHAR.551	Activo Hs/Pt	Inactivo	Inactivo	Inactivo
HACNS.705	Activo Hs/Pt	Mayor actividad Hs	Inactivo	NA

Referencias: Hs, *Homo sapiens*. Pt, *Pan troglodytes*, NA, no analizado. *, HAR, 2xHAR, HACNS y ANC. **, 2xHAR y HACNS. ***, HAR y 2xHAR.

Análisis de los RBFOX1-HARs en humanos arcaicos

Si bien en la actualidad *Homo sapiens* es el único representante vivo de nuestro linaje directo, durante los últimos cinco millones de años coexistieron sobre la Tierra diferentes grupos de homínidos. La comparación del genoma humano con el genoma de humanos arcaicos, como *Homo neanderthalensis*, permite descubrir cuándo fueron fijadas las sustituciones humano-específicas en los HARs. Develar el momento histórico en el cual estas sustituciones tuvieron lugar podría brindar información muy valiosa. Por ejemplo, la datación de los sucesos de aceleración permitiría asociar estos eventos con el registro fósil de nuestro linaje y las características morfológicas que diferencian a los humanos modernos de aquellas poblaciones de homínidos arcaicos, y se podrían generar mejores hipótesis sobre el rol que pudieron haber tenido los HARs en la evolución humana (Burbano et al., 2012). La mayoría de las sustituciones humano-específicas en los RBFOX1-HARs resultaron estar compartidas con el genoma de *H. neanderthalensis*, y las que fueron anotadas como derivadas son polimórficas tanto en las poblaciones humanas actuales como en las poblaciones de humanos arcaicos, lo que sugiere que la evolución acelerada en estas regiones fue anterior a la divergencia entre *H. sapiens* y *H. neanderthalensis*. Las sustituciones compartidas con los homínidos arcaicos se consideran antiguas y es probable que hayan surgido antes de la divergencia entre las poblaciones de los ancestros de los humanos anatómicamente modernos y los ancestros de *H. neanderthalensis*, que se estima tuvo lugar hace 553.000 a 589.000 años atrás (Prüfer et al., 2014).

Factores de transcripción candidatos

En el análisis de los TFBS encontramos algunos factores de transcripción candidatos que podrían explicar las diferencias de expresión observadas en los ensayos *in vivo*. En primer lugar, las sustituciones en la secuencia de chimpancé de 2xHAR.321 tuvieron como resultado la pérdida de un sitio de unión para *FOXD3*, un factor de transcripción esencial en el neurodesarrollo de vertebrados (Dottori et al., 2001; Cheung et al., 2005). Por otra parte, las sustituciones humano-específicas en 2xHAR.505 resultaron en la ganancia de nuevos sitios de unión para *TRPS1* y *ZBTB33*, dos represores transcripcionales (Malik et al., 2001; Sun et al., 2021), y en la pérdida de un sitio de unión para *FOXB1*, un factor de transcripción fundamental en el desarrollo del tálamo, hipotálamo y médula espinal (Wehr et al., 1997; Alvarez-Bolado et al., 2000). Estos resultados podrían explicar la pérdida de función de la secuencia de chimpancé de 2xHAR.321 y la pérdida de función de la secuencia humana de 2xHAR.505 que observamos en los ensayos de peces cebra. Sin embargo, es importante señalar que el análisis *in silico* de los TFBS está basado únicamente en predicciones de secuencia y los motivos de unión sólo acarrean el potencial de unirse a sus proteínas correspondientes. Los programas de predicción como JASPAR pueden inferir el potencial de unión, pero no la funcionalidad de un sitio. Además, las secuencias consenso para la unión de factores de transcripción son, en general, altamente degeneradas, y están distribuidas por todo el genoma (Noonan and McCallion, 2010). Como resultado, la cantidad de falsos positivos puede enmascarar a las verdaderas coincidencias. En última instancia, la actividad funcional sólo puede ponerse a prueba mediante ensayos experimentales, como por ejemplo ChIP-Seq (Geertz and Maerkl, 2010). Asimismo, no hay evidencias experimentales que muestren que *FOXD3*, *TRPS1*, *ZBTB33* o *FOXB1* regulen a *RBFOX1*. De hecho, hay muy poca información acerca de la regulación transcripcional de *RBFOX1*. Este análisis predictivo representa un primer paso hacia la identificación de la red regulatoria que controla a este gen. Uno de los pocos factores de transcripción que se ha comprobado experimentalmente que regula a *RBFOX1* es *FOXP2* (Spiteri et al., 2007; Vernes et al., 2011), por lo que no es sorprendente que la secuencia de 2xHAR.505 tenga un potencial sitio de unión a *FOXP2* altamente conservado en mamíferos en la posición 7,309,767-7,309,777 del cromosoma 16.

Evolución codificante y no codificante

El incremento en la tasa de evolución molecular experimentada por las secuencias no codificantes contrasta notablemente con la conservación a nivel de la secuencia codificante de *RBFOX1*, ya que los análisis evolutivos de rama por sitio sugieren que la región codificante de *RBFOX1* ha evolucionado bajo una fuerte selección purificadora. Además, teniendo en cuenta que no se encontraron evidencias de evolución acelerada en el resto de las regiones no codificantes conservadas, entonces las únicas regiones con signos de aceleración en el *locus* de *RBFOX1* son los HARs. Estos datos sugieren que el proceso de aceleración ocurrió específicamente en ciertas regiones no codificantes conservadas en la región ocupada por *RBFOX1* y no como consecuencia de fenómenos evolutivos a gran escala que involucran grandes porciones del *locus*.

Por otro lado, una de las desventajas cuando se estudian cambios en la tasa de evolución molecular en las regiones no codificantes es que resulta muy difícil identificar los procesos evolutivos responsables de la aceleración. Se requieren pruebas adicionales para distinguir los procesos evolutivos subyacentes, sean éstos selección positiva, relajación de la selección negativa o eventos poblacionales. Adicionalmente, la aceleración puede ser producto de procesos que no involucran necesariamente cambios en el modo de selección, como la conversión sesgada por GC (Duret and Galtier, 2009; Kostka et al., 2012). En este capítulo realizamos una clasificación de los *RBFOX1*-HARs para distinguir los procesos evolutivos subyacentes a la aceleración, y encontramos que en la mayoría de los casos la aceleración puede explicarse mejor por selección positiva, mientras que solamente tres HARs se definen mejor por gBGC. Curiosamente, dos de estos últimos (ANC.737 y 2xHAR.300) no son *enhancers* en ninguna de las dos especies analizadas en los ensayos de gen reportero en peces cebra transgénicos.

Promotor mínimo de RBFOX1

Finalmente, el último objetivo de este capítulo fue estudiar la expresión de los promotores alternativos de *RBFOX1*. Originalmente, este objetivo no era parte del plan de estudios, pero surgió como una alternativa para poder generar construcciones reporteras que tengan el promotor mínimo de *RBFOX1*. Además, en un trabajo reciente evaluaron la eficiencia de seis promotores mínimos e interrogaron exhaustivamente su compatibilidad con *enhancers* del desarrollo, y descubrieron que el promotor mínimo de ratón de *cFos* tiene altas tasas de expresión ectópica en el corazón, ojo y aleta caudal (Begeman et al., 2022). Como observamos que el promotor *RBFOX1*-E1B se expresa en los mismos dominios de expresión que *rbfox1* en peces cebra, tal como determinamos mediante los ensayos de hibridización *in situ*, nos propusimos identificar la región de mínima actividad. Esta no posee expresión ectópica por sí sola, pero actúa adecuadamente en respuesta a la entrada de *enhancers* específicos de tejido. De esta manera, en esta tesis no solo pusimos a punto el análisis funcional de promotores en peces cebra transgénicos, sino que también construimos una nueva herramienta basada en el vector Tol2 que tiene el promotor mínimo E1B para estudiar *enhancers* específicos de *RBFOX1*. Este vector es útil para generar tanto peces como ratones transgénicos.

CAPÍTULO II

Hipótesis de trabajo

Proponemos que las regiones aceleradas 2xHAR.321 y 2xHAR.505 son elementos regulatorios críticos de *RBFOX1*, y que sus respectivas deleciones en el genoma de ratón tienen un impacto funcional en los niveles de expresión de *Rbfox1*.

Objetivos específicos

El objetivo principal de este capítulo es estudiar el impacto funcional de la deleción de las regiones aceleradas 2xHAR.321 y 2xHAR.505 en la expresión de *Rbfox1* en ratones genéticamente modificados y su consiguiente efecto fenotípico. Para ello se propone:

- Generar ratones mutantes nulos o *knock-out* que carezcan de la región acelerada 2xHAR.321 o 2xHAR.505 mediante CRISPR/Cas9.
- Analizar la expresión de *Rbfox1* en distintos tejidos del sistema nervioso central a dos estadios del desarrollo, uno embrionario y otro postnatal para determinar si existen diferencias de expresión entre los ratones mutantes y de tipo salvaje.
- Analizar los genes diferencialmente expresados y el uso diferencial de exones en los ratones mutantes mediante secuenciación del ARNm.

Resultados

Generación de ratones genéticamente modificados

En el capítulo anterior identificamos varios *RBFOX1*-HARs que actúan como *enhancers* transcripcionales en ensayos de gen reportero en peces cebra transgénicos. Sin embargo, la detección de elementos regulatorios mediante reporteros no permite evaluar la importancia relativa de cada *enhancer* en la transcripción *in vivo* de *RBFOX1*, lo que hace que la información espacial y temporal obtenida hasta el momento pueda no necesariamente ser la misma que se obtendría estudiando los mismos elementos regulatorios en su contexto endógeno. Por este motivo, decidimos generar, de manera independiente, ratones mutantes nulos que tengan delecionado dos HARs que resultaron de interés, 2xHAR.321 y 2xHAR.505. En ambos casos, en el capítulo anterior confirmamos que sus secuencias de ratón fueron capaces de actuar como *enhancers* transcripcionales en los ensayos de peces cebra transgénicos.

Para ello, implementamos la tecnología de edición génica CRISPR/Cas9 (Wang et al., 2013). Brevemente, la técnica consiste en la utilización de un complejo riboproteico conformado por la enzima endonucleasa Cas9 y una molécula de ARN guía (ARNg). La ARNg tiene una secuencia constante, necesaria para formar el complejo, y una secuencia variable, que se une por complementariedad de bases a la secuencia genómica que se

quiere editar de manera específica. Además, el complejo riboproteico sólo se unirá a aquellas secuencias que estén situadas inmediatamente de un sitio con secuencia NGG, denominado PAM. Una vez que la Cas9 realiza el corte en la doble cadena del ADN, se activan al menos dos mecanismos naturales de reparación del ADN, la unión de extremos no homólogos (NHEJ) o la reparación dirigida por homología (HDR). Por un lado, NHEJ es un mecanismo propenso a errores en donde dos extremos de ADN se unen, sin necesidad de homología, lo que a menudo da como resultado un grupo heterogéneo de inserciones y deleciones. En cambio, HDR es un mecanismo preciso que repara el ADN mediante recombinación homóloga con un ADN donador.

En contraposición a la generación de ratones mutantes nulos de regiones codificantes, donde se pueden conseguir efectos deletéreos en la función de las proteínas con un corrimiento en el marco de lectura con deleciones pequeñas e individuales, la pérdida de función de elementos no codificantes requiere la deleción de toda la región sospechada de actividad, que puede abarcar varios cientos de pares de bases. Por esta razón, decidimos emplear dos ARNg, una en cada uno de los extremos de la secuencia de interés. Entonces, para el mutante de 2xHAR.321 obtuvimos una deleción de 688 pb que abarca desde la posición 6,672,214 hasta 6,672,902 en el cromosoma 16 del genoma de ratón (Figura 14a), mientras que para el mutante de 2xHAR.505 obtuvimos una deleción de 866 pb en la posición 7,044,783-7,045,649 del cromosoma 16 (Figura 14b). Ambas deleciones fueron confirmadas por secuenciación.

Los ratones nacidos producto de la microinyección y posterior transferencia embrionaria son denominados F0, y en muchas ocasiones son animales quiméricos con múltiples deleciones. Por esta razón, es necesario realizar cruzas con animales de tipo salvaje o *wild-type* (*Rbfox1*^{+/+}) para seleccionar cuidadosamente las deleciones deseadas. La primera generación (F1) es portadora de un solo alelo parental mutado, por lo que se seleccionaron los ratones con la deleción deseada y se cruzaron con animales de tipo salvaje para iniciar la colonia de ratones mutantes. De la segunda generación (F2), producto de esta cruce, se seleccionaron ratones heterocigotas para la deleción deseada (*Rbfox1*^{+/Δ2xHAR.321} o *Rbfox1*^{+/Δ2xHAR.505}), y de la cruce de dos heterocigotas, es posible obtener los tres genotipos en la tercera generación (F3); heterocigotas (*Rbfox1*^{+/Δ2xHAR.321} o *Rbfox1*^{+/Δ2xHAR.505}), homocigotas *knock-out* (*Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} o *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}) y homocigotas *wild-type* (*Rbfox1*^{+/+}). A simple vista, los ratones mutantes tanto de 2xHAR.321 como de 2xHAR.505 son viables y fértiles, no presentan ningún fenotipo aparente y son prácticamente indistinguibles de sus compañeros de camada de tipo salvaje.

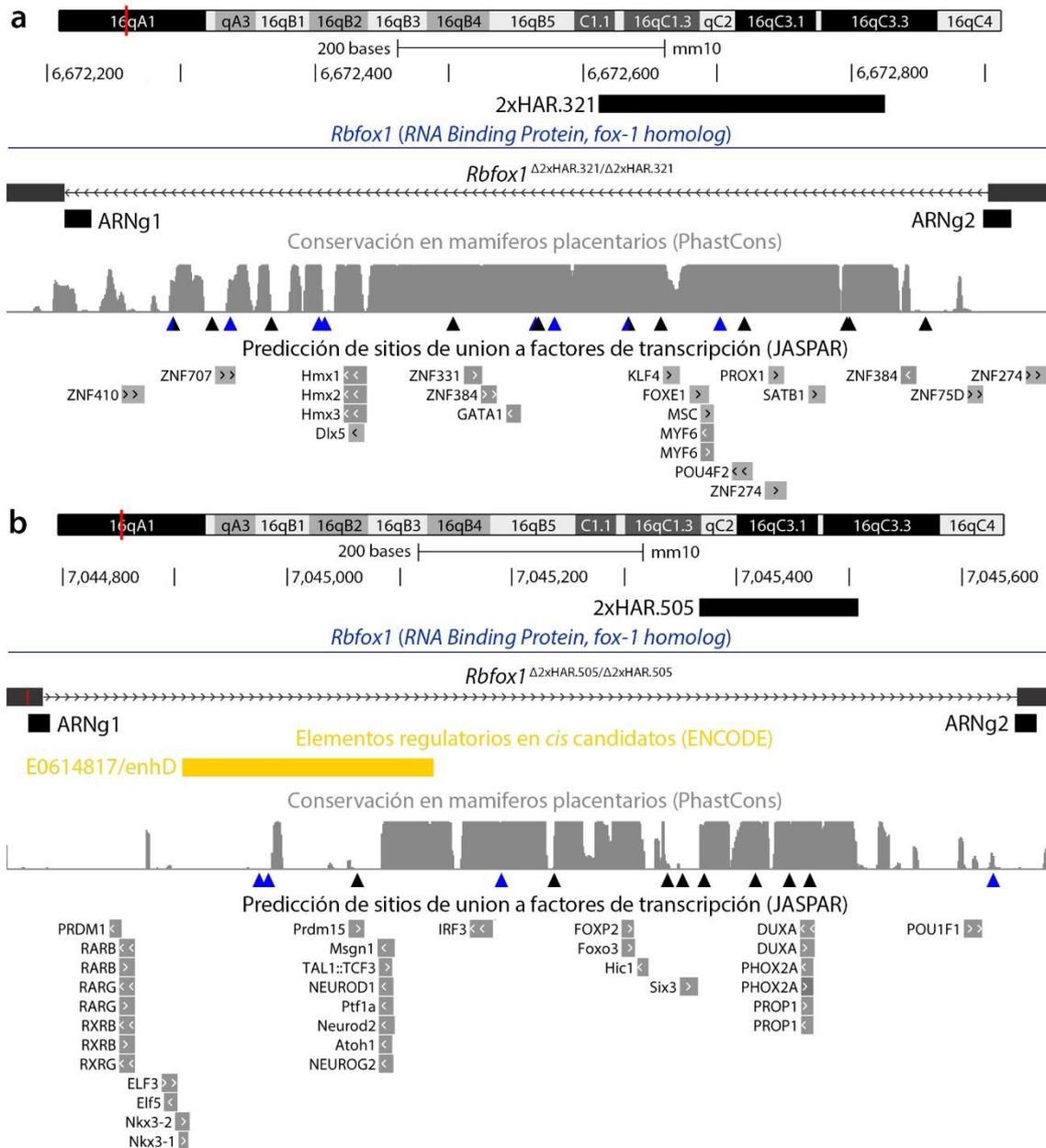


FIGURA 14: Generación de ratones mutantes nulos de (a) 2xHAR.321 y (b) 2xHAR.505. Para cada uno, se incluye la localización genómica de las ARN guías (ARNg) y el resultado de la secuenciación; la línea con flechas es la región que no pudo ser alineada por no estar presente. También se incluyen los elementos regulatorios en *cis* candidatos (ENCODE), el nivel de conservación en mamíferos placentarios (PhastCons) y una predicción de sitios de unión a factores de transcripción (JASPAR). Debajo de la conservación, las flechas indican la presencia de sustituciones; las flechas negras son las sustituciones humano-específicas y las flechas azules son las sustituciones chimpancé-específicas.

Análisis de los niveles de expresión de Rbfox1

A continuación, evaluamos si las deleciones tienen un impacto en los niveles de expresión de *Rbfox1* mediante RT-qPCR. Para ello, recolectamos muestras de tejido del sistema nervioso de ratones *knock-out* homocigotas y *wild-type* de la misma camada a 14.5 días embrionarios (E14.5) y a 8 días postnatales (P8). Para seleccionar los dominios

anatómicos de donde recolectar las muestras, primero caracterizamos la expresión endógena de *Rbfox1* mediante hibridización *in situ* a E14.5 (Figura 15), y utilizamos la información disponible de la expresión de *Rbfox1* en ratones juveniles (Hammock and Levitt, 2011). De esta manera, a E14.5 observamos que *Rbfox1* se expresa principalmente en el sistema nervioso, con una expresión robusta en la placa cortical del palio dorsal (Figura 15a/b), mesencéfalo (Figura 15c), rombencéfalo (Figura 15d) y médula espinal. Además, también observamos expresión en el estriado, tálamo, cerebelo y en los ganglios de la raíz dorsal. Por otro lado, al final de la segunda semana postnatal (P14), la expresión de *Rbfox1* se puede detectar en áreas discretas, como el bulbo olfatorio, el núcleo accumbens, el tálamo, la corteza, el hipocampo y el estriado (Hammock and Levitt, 2011). Además, en ratones adultos (P56) la expresión de *Rbfox1* se mantiene en los mismos dominios anatómicos (<https://mouse.brain-map.org/experiment/show/100145385>). Entonces, a E14.5 decidimos recolectar muestras de corteza cerebral, mesencéfalo y rombencéfalo, mientras que a P8 recolectamos muestras de corteza cerebral, mesencéfalo y cerebelo. De cada uno, aislamos el ARN y sintetizamos el ADNc, con seis réplicas biológicas de cada genotipo. Los *primers* específicos para *Rbfox1* se unen al dominio de unión al ARN que es común a todas las isoformas, y para la normalización de los datos utilizamos la expresión del gen *housekeeping* β -actina.

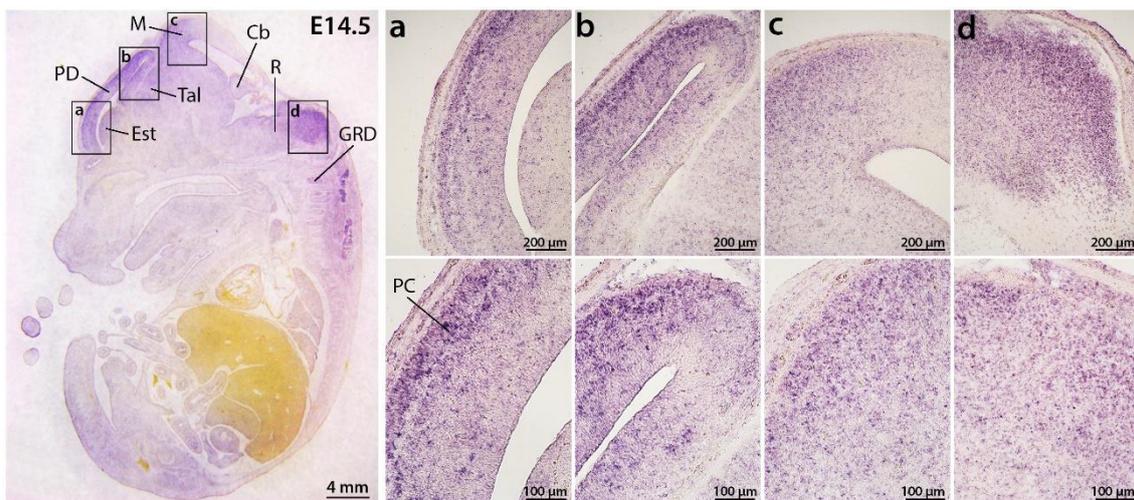


FIGURA 15: Expresión endógena de *Rbfox1*. Hibridización *in situ* en cortes sagitales de un embrión de ratón a E14.5 días. Se incluyen detalles de la expresión en el (a/b) palio dorsal, (c) mesencéfalo y (d) rombencéfalo. Debajo de cada uno se muestra una microfotografía a mayor aumento mostrando en detalle la expresión. Referencias: PD, palio dorsal. Est, estriado. Tal, tálamo. M, mesencéfalo. Cb, cerebelo. R, rombencéfalo. GRD, ganglios de la raíz dorsal. PC, placa cortical.

De esta manera, observamos que, en comparación con los ratones *Rbfox1*^{+/+}, los ratones *Rbfox1* ^{Δ 2xHAR.321/ Δ 2xHAR.321} presentan diferencias significativas en la expresión de *Rbfox1* únicamente en el mesencéfalo a los 14.5 días embrionarios (Figura 16b). Curiosamente, los ratones *Rbfox1* ^{Δ 2xHAR.321/ Δ 2xHAR.321} presentan un aumento aproximado del 15% en la expresión de *Rbfox1* (p-valor = 0.0040). El resto de los tejidos y estadios analizados no presentan diferencias significativas (Figura 16).

2xHAR.321 knock-out

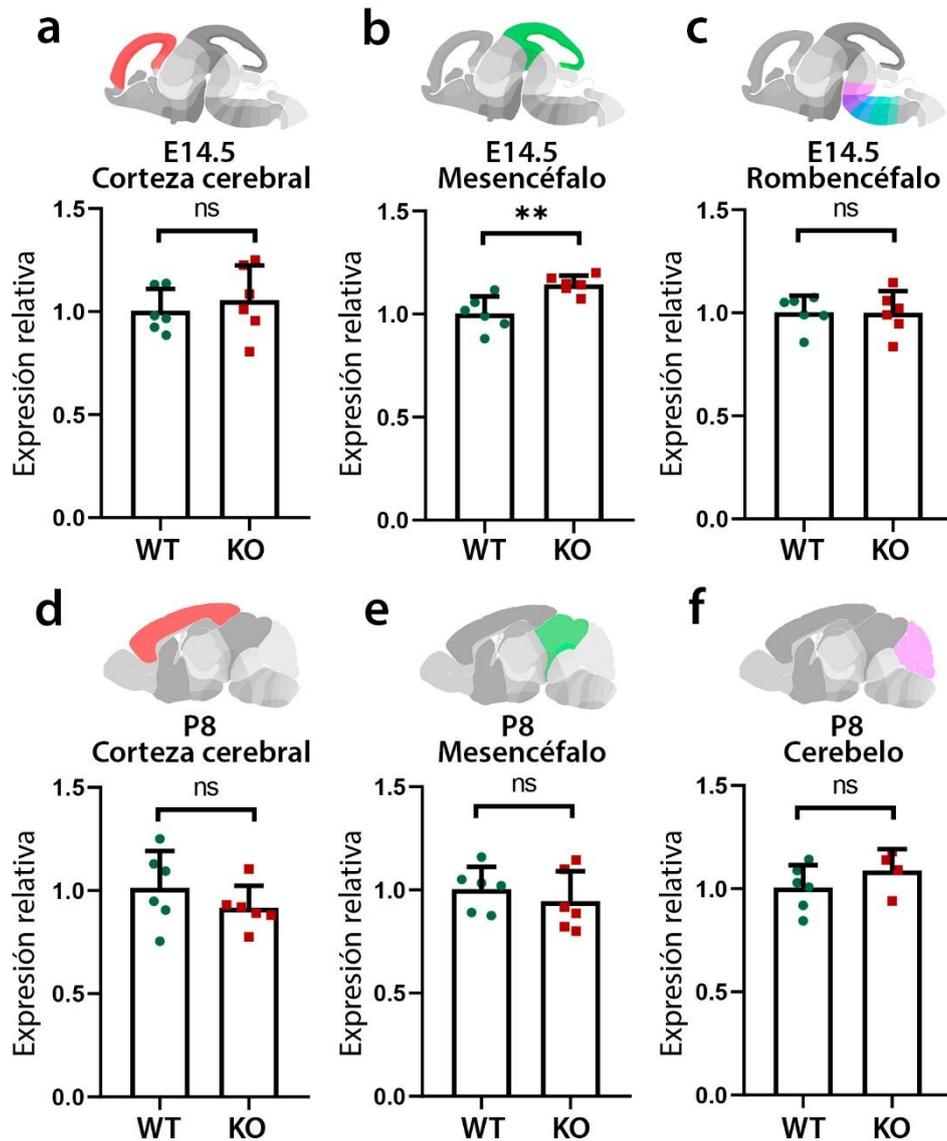


FIGURA 16: Análisis de la expresión de *Rbfox1* en ratones $Rbfox1^{\Delta 2xHAR.321/\Delta 2xHAR.321}$. Expresión relativa de *Rbfox1* a (a-c) E14.5 en (a) corteza cerebral, (b) mesencéfalo y (c) rombencéfalo, y a (d-f) P8 en (d) corteza cerebral, (e) mesencéfalo y (f) cerebelo. La cuantificación de la expresión se realizó con el método $2^{-\Delta\Delta Ct}$ con β -actina como gen de referencia, y la comparación entre ratones *wild-type* (WT) y *knock-out* (KO) se realizó con la prueba estadística t de Student. ns, no significativo; **, muy significativo (p-valor=0.0040).

Por el otro lado, observamos que los ratones $Rbfox1^{\Delta 2xHAR.505/\Delta 2xHAR.505}$ presentan diferencias significativas en comparación con los ratones $Rbfox1^{+/+}$ en la corteza cerebral a los 8 días postnatales (Figura 17d). En efecto, los ratones $Rbfox1^{\Delta 2xHAR.505/\Delta 2xHAR.505}$ presentan un aumento aproximado del 25% en la expresión de *Rbfox1* (p-valor = 0.0052). El resto de los tejidos y estadios analizados no presentan diferencias significativas (Figura 17).

2xHAR.505 knock-out

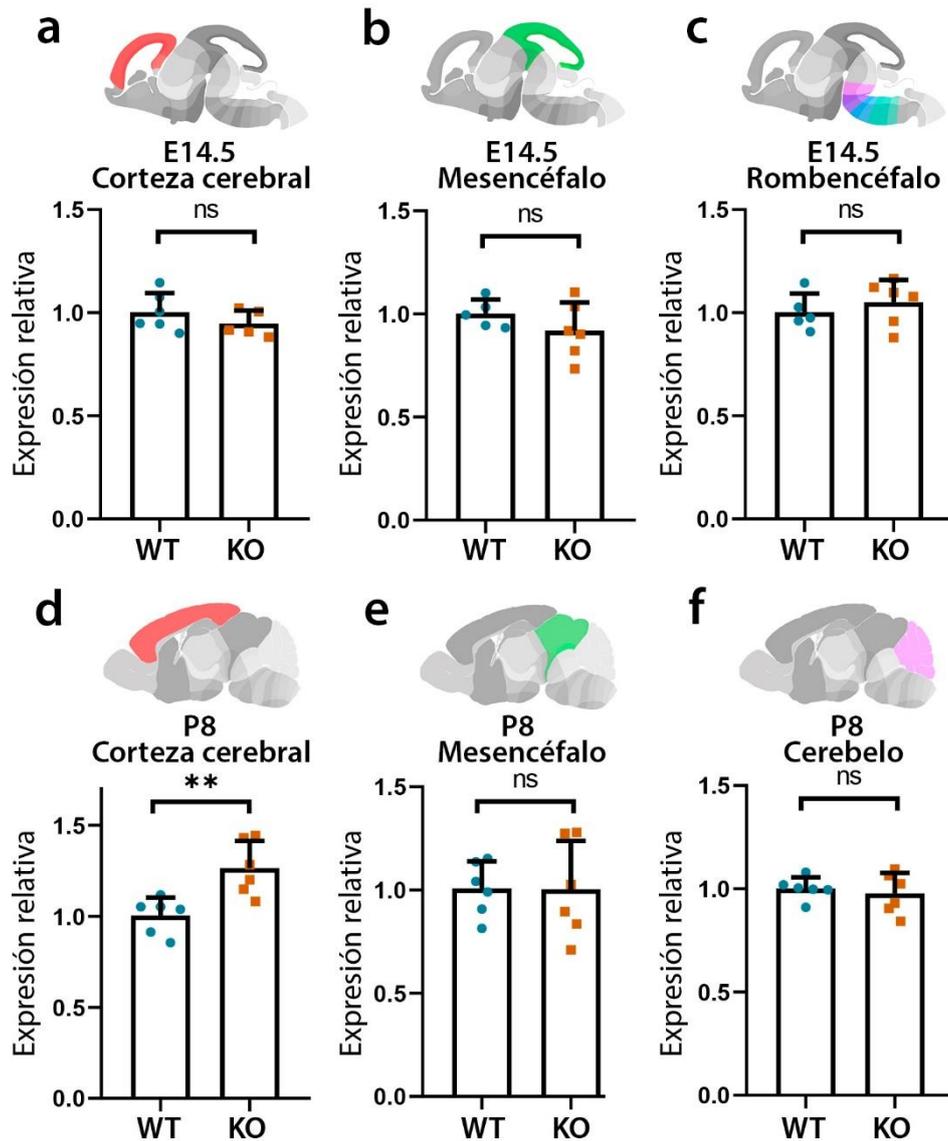


FIGURA 17: Análisis de la expresión de *Rbfox1* en ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}. Expresión relativa de *Rbfox1* a (a-c) E14.5 en (a) corteza cerebral, (b) mesencéfalo y (c) rombencéfalo, y a (d-f) P8 en (d) corteza cerebral, (e) mesencéfalo y (f) cerebelo. La cuantificación de la expresión se realizó con el método $2^{-\Delta\Delta Ct}$ con β -actina como gen de referencia, y la comparación entre ratones *wild-type* (WT) y *knock-out* (KO) se realizó con la prueba estadística t de Student. ns, no significativo; **, muy significativo (p-valor=0.0052).

Análisis de genes diferencialmente expresados

Considerando los resultados anteriores, a continuación decidimos explorar si las deleciones de los HARs tienen un impacto sobre la expresión de otros genes, especialmente los regulados por *Rbfox1*. Para ello, realizamos una secuenciación de alto rendimiento del ARNm en las edades y tejidos donde observamos que *Rbfox1* tiene diferencias significativas de expresión. De esta manera, observamos 272 genes diferencialmente expresados en el mesencéfalo a E14.5 en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} en comparación con los ratones *Rbfox1*^{+/+} con un p-valor

ajustado menor a 0.05 (Apéndice III, Tabla S1). De estos, 125 mostraron un aumento en los niveles de expresión y 147 una disminución. Las diferencias de expresión se miden mediante el cambio en el nivel de expresión o *fold-change* (FC), que describe cuánto cambia una cantidad entre dos mediciones. Se define como la relación entre las dos cantidades; por ejemplo, para las cantidades de expresión en los ratones de tipo salvaje (A) y los ratones mutantes (B), el cambio en el nivel de expresión de B con respecto de A es B/A . En bioinformática, a menudo se utiliza el logaritmo en base 2 del FC para el análisis y visualización de los cambios en el nivel de expresión, ya que es fácil de interpretar. Por ejemplo, una duplicación de la expresión es igual a un \log_2FC de 1, una cuadruplicación es igual a un \log_2FC de 2 y así sucesivamente. Además, la medida es simétrica cuando el cambio disminuye, es decir que una reducción a la mitad es igual a un \log_2FC de -1, una reducción de un cuarto es igual a un \log_2FC de -2 y así sucesivamente. De este modo, los veinte genes con mayor cambio en el nivel de expresión en los ratones *Rbfox1* ^{$\Delta 2xHAR.321/\Delta 2xHAR.321$} , tanto positivo como negativo, se representan en la Figura 18a. En todos los casos, los valores de \log_2FC son moderados, entre -0.65 y 0.65 (Figura 18b).

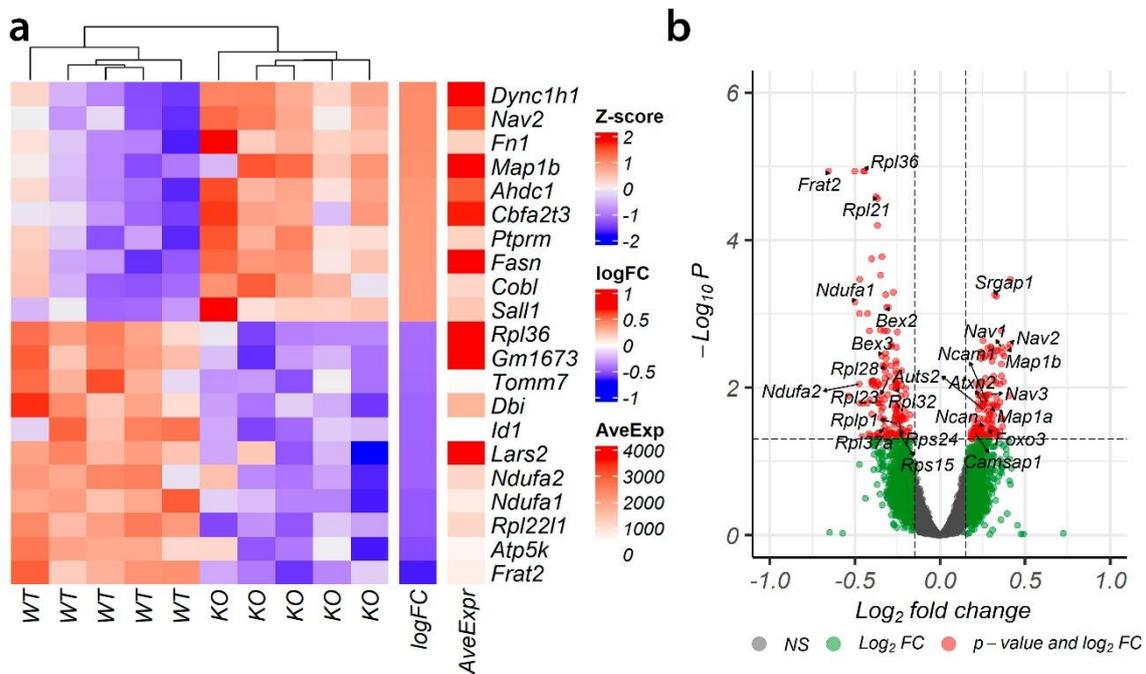


FIGURA 18: Genes diferencialmente expresados en los ratones *Rbfox1* ^{$\Delta 2xHAR.321/\Delta 2xHAR.321$} . (a) *Heatmap* de los genes con mayor cambio en el nivel de expresión con un p-valor ajustado por Benjamini y Hochberg < 0.05. Cada fila representa un gen y cada columna una muestra. Referencias: KO, *knock-out*. WT, *wild-type*. \log_2FC , logaritmo en base 2 del cambio en el nivel de expresión o *fold-change*. AveExpr, expresión promedio. (b) Gráfico de dispersión tipo volcán donde se diagrama la significancia ($-\log_{10}P$) versus el nivel de cambio (\log_2FC). Los puntos rojos representan los genes diferencialmente expresados con un p-valor ajustado < 0.05. Los genes con mayor expresión se encuentran a la derecha del centro, mientras que los genes con menor expresión se encuentran a la izquierda.

Algunos genes diferencialmente expresados en los ratones *Rbfox1* ^{$\Delta 2xHAR.321/\Delta 2xHAR.321$} que resulta interesante analizar en más detalle incluyen a *Atxn2* (*Ataxin 2*), un gen que codifica

una proteína de unión al ARN que controla distintos procesos postranscripcionales en las células de Purkinje del cerebelo (Lee et al., 2018) y que interactúa con *Rbfox1* (o *A2bp1*, *Ataxin-2 binding protein 1*) (Shibata et al., 2000); *Ncam1* (*Neuronal cell adhesion molecule 1*), un gen que codifica una proteína de adhesión celular implicada en las interacciones célula-célula, así como también célula-matriz durante el desarrollo y la diferenciación del sistema nervioso (Weledji and Assob, 2014); *Stxbp1* (*Syntaxin binding protein 1*) que desempeña un papel crucial en la liberación de neurotransmisores a través de la regulación de la syntaxina (Rizo and Südhof, 2012); y *Ncan* (*Neurocan*), un gen que codifica para un proteoglicano de sulfato de condroitina involucrado en la adhesión neuronal y en el crecimiento de neuritas durante el desarrollo (Friedlander et al., 1994). Además, también observamos varios genes con cambios en el nivel de expresión que pertenecen a la misma familia. Por ejemplo, *Map1a* y *Map1b* (*Microtubule associated protein 1*) son proteínas asociadas a microtúbulos involucradas en el ensamblaje de estos durante la neurogénesis (Halpain and Dehmelt, 2006); *Nav1*, *Nav2* y *Nav3* (*Neuronavigator*) son proteínas asociadas al citoesqueleto importantes para la migración neuronal, el crecimiento de neuritas y el desarrollo axonal (Powers et al., 2022); *Bex1*, *Bex2* y *Bex3* (*Brain expressed X-linked*) juegan un papel importante en la progresión del ciclo celular y la diferenciación neuronal (Carter, 2006); y *Hnrnpa0* y *Hnrnpa1* (*Heterogeneous nuclear ribonucleoprotein A*) son proteínas de unión al ARN que pueden regular el *splicing* alternativo tanto de forma independiente como en colaboración con otras RBPs, como *Rbfox1* (Krecic and Swanson, 1999).

A continuación, realizamos un análisis de enriquecimiento funcional con términos de ontología génica (GO) en los genes diferencialmente expresados de los ratones *Rbfox1* ^{Δ 2xHAR.321/ Δ 2xHAR.321}, y observamos que los genes que aumentan y que disminuyen su expresión están enriquecidos en términos distintos (Apéndice III, Tabla S2). Por un lado, los genes con expresión aumentada están relacionados con procesos biológicos involucrados en el desarrollo y la diferenciación celular, especialmente del sistema nervioso, así como también en la organización de la cromatina y la modificación de histonas (Figura 19a/c). En cambio, los genes con expresión disminuida están enriquecidos en términos relacionados con la traducción y la estabilidad proteica, la fosforilación oxidativa y la respiración aeróbica (Figura 19b/c). Con respecto a los componentes celulares, los genes que aumentan su expresión están enriquecidos en el citoplasma y el axón (Figura 19a/c), y los genes que disminuyen su expresión, en el ribosoma y el respirasoma (Figura 19b/c). Por último, los términos de la función molecular revelaron que los genes con expresión aumentada se relacionan con la unión a proteínas y a la cromatina (Figura 19a/c), mientras que los genes con expresión disminuida están enriquecidos en el constituyente estructural del ribosoma, y la unión al ARN (Figura 19b/c).

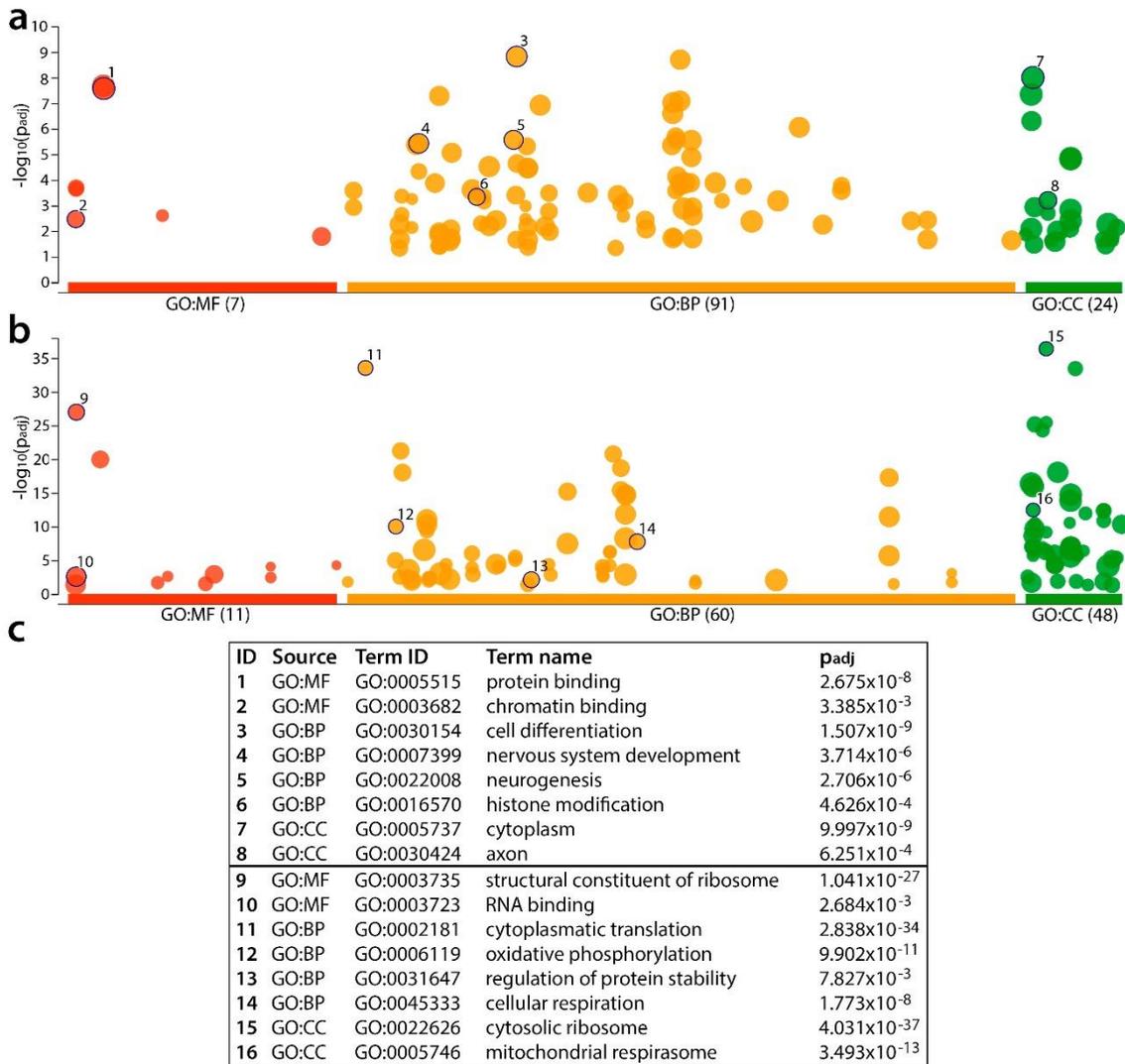


FIGURA 19: Análisis de enriquecimiento funcional de los genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321}. Representación de los valores $-\log_{10}(p_{adj})$ de los términos para (a) los genes con expresión aumentada y (b) los genes con expresión disminuida. (c) En la tabla se describen los términos principales con los correspondientes números.

Por otro lado, observamos 132 genes diferencialmente expresados en la corteza cerebral a P8 en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} en comparación con los ratones *Rbfox1*^{+/+} con un p-valor ajustado menor a 0.05 (Apéndice III, Tabla S3), 61 con expresión aumentada y 71 con expresión disminuida. Los veinte genes con mayor cambio en el nivel de expresión, tanto positivo como negativo, se representan en la Figura 20a. En este caso, los valores de log₂FC son un poco más altos que los observados en los genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} (Figura 20b).

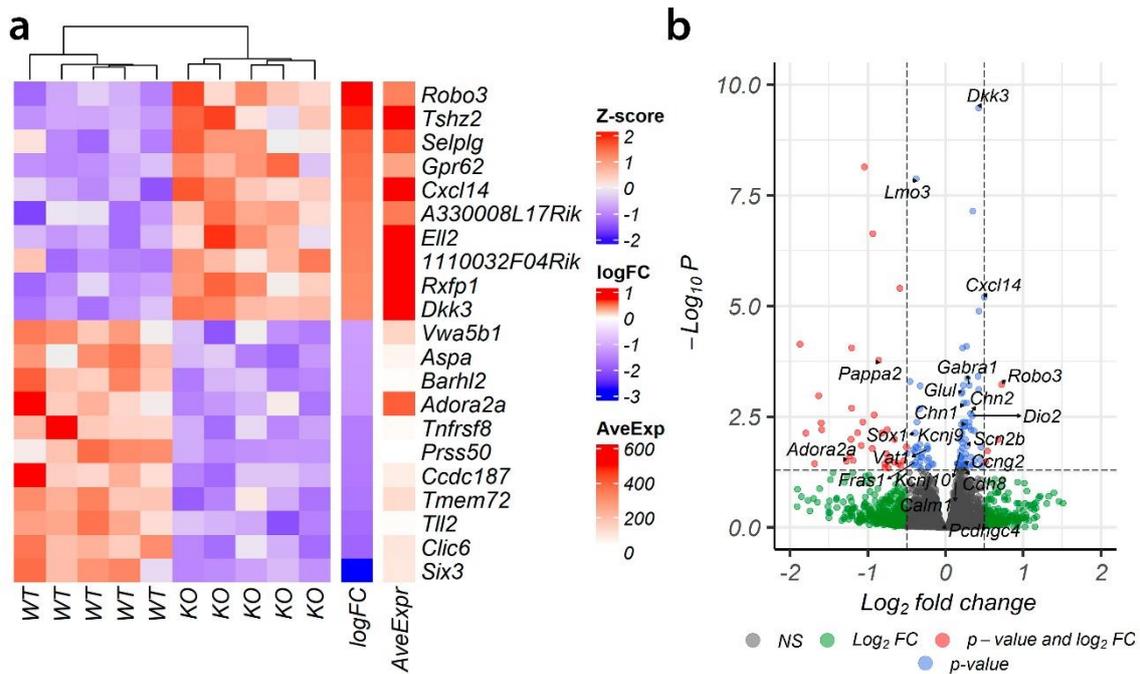
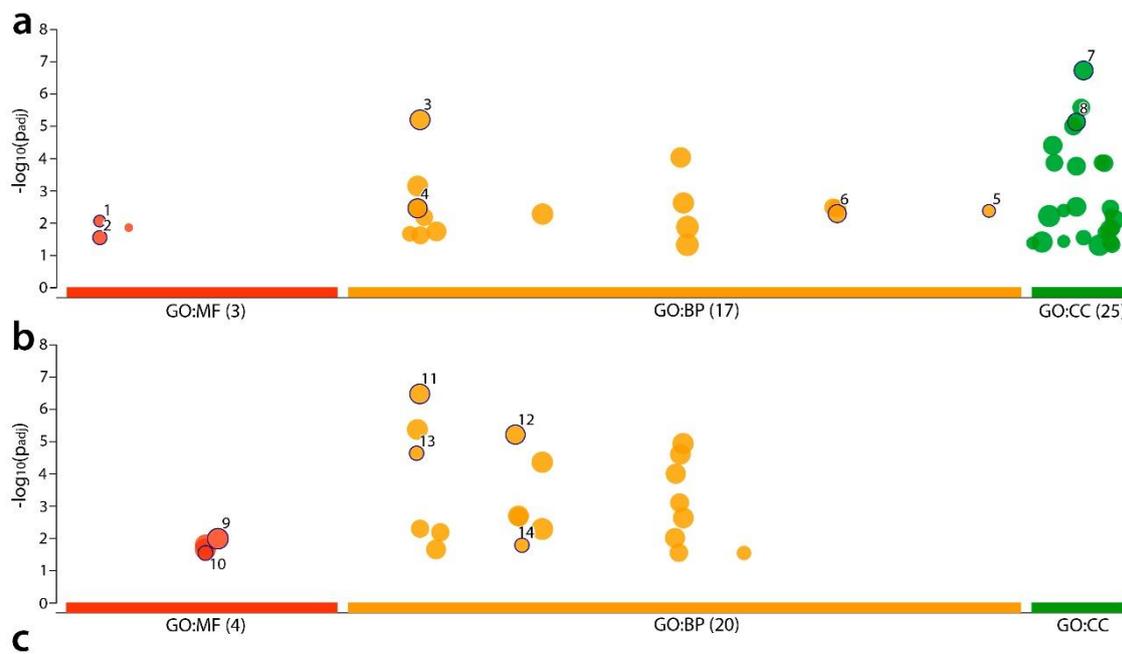


FIGURA 20: Genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}. (a) *Heatmap* de los genes con mayor cambio en el nivel de expresión con un p-valor ajustado por Benjamini y Hochberg < 0.05. Cada fila representa un gen y cada columna una muestra. Referencias: KO, *knock-out*. WT, *wild-type*. logFC, logaritmo en base 2 del cambio en el nivel de expresión o *fold-change*. AveExpr, expresión promedio. (b) Gráfico de dispersión tipo volcán donde se diagrama la significancia ($-\text{Log}_{10}P$) versus el nivel de cambio (Log_2FC). Los puntos rojos y azules representan los genes diferencialmente expresados con un p-valor ajustado < 0.05. Los genes con mayor expresión se encuentran a la derecha del centro, mientras que los genes con menor expresión se encuentran a la izquierda.

Algunos genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} que resulta interesante analizar en más detalle incluyen a *Gabra1* (*Gamma-aminobutyric acid receptor subunit alpha-1*), un gen que codifica un receptor GABA y que desempeña un papel importante en la formación de sinapsis inhibitorias funcionales (Fritschy and Panzanelli, 2014); *Scn2b* (*Sodium voltage-gated channel beta-2*), una subunidad del canal de sodio dependiente de voltaje que está implicada en la adhesión célula-célula y en la migración celular (Brackenbury and Isom, 2011); *Glul* (*Glutamate-ammonia ligase*), una glutamina sintetasa que convierte el glutamato neurotóxico en glutamina inofensiva (Zhou and Danbolt, 2014); y *Grm2* (*Glutamate metabotropic receptor 2*), un receptor metabotrópico de glutamato acoplado a la proteína G, que inhibe el vaciamiento de los contenidos vesiculares en la terminal presináptica de las neuronas glutamatérgicas (Ferraguti and Shigemoto, 2006). Además, encontramos que los genes *Kcnj3*, *Kcnj9* y *Kcnj10* (*Potassium inwardly rectifying channel subfamily J*) que son canales de potasio que controlan el potencial de membrana y el potencial de acción (Hibino et al., 2010); y los genes *Chn1* y *Chn2* (*Chimerin*) que son proteínas activadoras de GTPasa que juegan un papel importante en los mecanismos de transducción de señales neuronales (Lim, 1992), muestran diferencias de expresión entre los ratones mutantes y los de tipo salvaje.

El análisis de enriquecimiento funcional con términos de ontología génica (GO) de los genes diferencialmente expresados de los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} reveló que los genes que aumentan y que disminuyen su expresión están enriquecidos en términos vinculados con el sistema nervioso (Apéndice III, Tabla S4). En ambos casos, están relacionados con procesos biológicos involucrados en el desarrollo y la diferenciación del sistema nervioso. Además, los genes con expresión aumentada están enriquecidos en la señalización célula-célula, la señalización sináptica y el transporte de iones de potasio (Figura 21a/c), y los genes con expresión disminuida muestran enriquecimiento en términos relacionados a la neurogénesis, la vía de señalización de neuropéptidos y la regulación de la extensión del axón (Figura 21b/c). Por otro lado, respecto de los componentes celulares, los genes que aumentan su expresión tienen un enriquecimiento en la sinapsis y las células neuronales (Figura 21a/c). Por último, en los términos de la función molecular, los genes con mayor expresión están enriquecidos en la actividad de los canales de potasio dependiente de voltaje y activado por proteína G (Figura 21a/c), mientras que los genes con menor expresión se relacionan con la unión a iones metálicos y la unión a ácidos orgánicos (Figura 21b/c).



ID	Source	Term ID	Term name	p _{adj}
1	GO:MF	GO:0005242	inward rectifier potassium channel activity	8.869x10 ⁻³
2	GO:MF	GO:0005249	voltage-gated potassium channel activity	2.896x10 ⁻²
3	GO:BP	GO:0007399	nervous system development	6.450x10 ⁻⁶
4	GO:BP	GO:0007267	cell-cell signaling	3.575x10 ⁻³
5	GO:BP	GO:1990573	potassium ion import across plasma membrane	4.306x10 ⁻³
6	GO:BP	GO:0099536	synaptic signaling	5.229x10 ⁻³
7	GO:CC	GO:0045202	synapse	1.912x10 ⁻⁷
8	GO:CC	GO:0043025	neuronal cell body	7.567x10 ⁻⁶
9	GO:MF	GO:0046972	metal ion binding	1.054x10 ⁻²
10	GO:MF	GO:0043177	organic acid binding	2.927x10 ⁻²
11	GO:BP	GO:0007399	nervous system development	3.449x10 ⁻⁷
12	GO:BP	GO:0022008	neurogenesis	6.271x10 ⁻⁶
13	GO:BP	GO:0007218	neuropeptide signaling pathway	2.367x10 ⁻⁵
14	GO:BP	GO:0030516	regulation of axon extension	1.675x10 ⁻²

FIGURA 21: Análisis de enriquecimiento funcional de los genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}. Representación de los valores $-\log_{10}(p_{adj})$

de los términos para (a) los genes con expresión aumentada y (b) los genes con expresión reducida. (c) En la tabla se describen los términos principales con los correspondientes números.

Sorpresivamente, no observamos que *Rbfox1* este diferencialmente expresado en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} cuando evaluamos la expresión génica por secuenciación del ARNm (Apéndice III, Tabla S1 y S3). Sin embargo, a continuación decidimos evaluar si los genes diferencialmente expresados en los ratones mutantes son genes regulados por *Rbfox1*. Para ello, utilizamos información disponible de trabajos anteriores. Por un lado, cruzamos la lista de genes con datos de iCLIP-Seq, una variante de ChIP-Seq para ver interacciones entre proteínas y ARN, de dos trabajos donde evalúan las interacciones de *Rbfox1*, *Rbfox2* y *Rbfox3* en las regiones 3' UTR de los ARNm (Weyn-Vanhentenryck et al., 2014; Lee et al., 2016). De esta manera, observamos que miembros de la familia *Rbfox* se unen a las regiones 3'UTR de 107 (39,3%) y 59 (44,6%) genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} (Figura 22a; Apéndice III, Tabla S1) y *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} (Figura 22b; Apéndice III, Tabla S2), respectivamente.

Por otro lado, también utilizamos información de tres trabajos en donde se realizó una reducción de la expresión (*knock-down*) o la delección total (*knock-out*) de *Rbfox1*. En uno de ellos realizaron un doble *knock-down* de *Rbfox1* y *Rbfox3* mediante siARN (*small interfering ARN*) en cultivos de hipocampo preparados a partir de ratones P0 y evaluaron la expresión génica mediante microarreglos y secuenciación de ARNm (Lee et al., 2016). En otro trabajo del mismo grupo, diseccionaron hipocampos de ratones mutantes nulos de *Rbfox1* condicionales del sistema nervioso (*Rbfox1*^{loxp/loxp}/*Nestin-Cre*^{+/-}) y evaluaron la expresión génica por secuenciación de ARNm (Vuong et al., 2018). Y en el último trabajo, generaron ratones mutantes nulos de *Rbfox1* condicionales de interneuronas Somatostatina/SST+ (*Rbfox1*^{loxp/loxp}/*SST-Cre*^{+/-}) y Parvalbúmina/PV+ (*Rbfox1*^{loxp/loxp}/*Tac1-Cre*^{+/-}), para luego evaluar la expresión génica mediante secuenciación de ARNm en las interneuronas corticales (Wamsley et al., 2018). Con esta información, observamos que 40 (14,7%) y 55 (41,6%) genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} (Figura 22c; Apéndice III, Tabla S1) y *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} (Figura 22d; Apéndice III, Tabla S3), respectivamente, son genes regulados por *Rbfox1*.

Por último, de acuerdo con el modelo propuesto de la función citoplasmática de *Rbfox1*, este se une a los motivos (U)GCAUG ubicados en las regiones 3' UTR, bloqueando la unión de micro-ARNs y promoviendo la estabilidad de los ARNm. En uno de los trabajos mencionados, cruzaron los datos de iCLIP-Seq de *Rbfox1* con datos de iCLIP-Seq de Ago (*Argonaute*) para identificar pares de sitios *Rbfox1*-miARN que podrían competir por el sitio para regular la expresión de los ARNm (Lee et al., 2016). De los pares informados en el trabajo, tres (*Glul*, *Cxcl14* y *Pcdhgc4*) son genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} (Figura 22e).

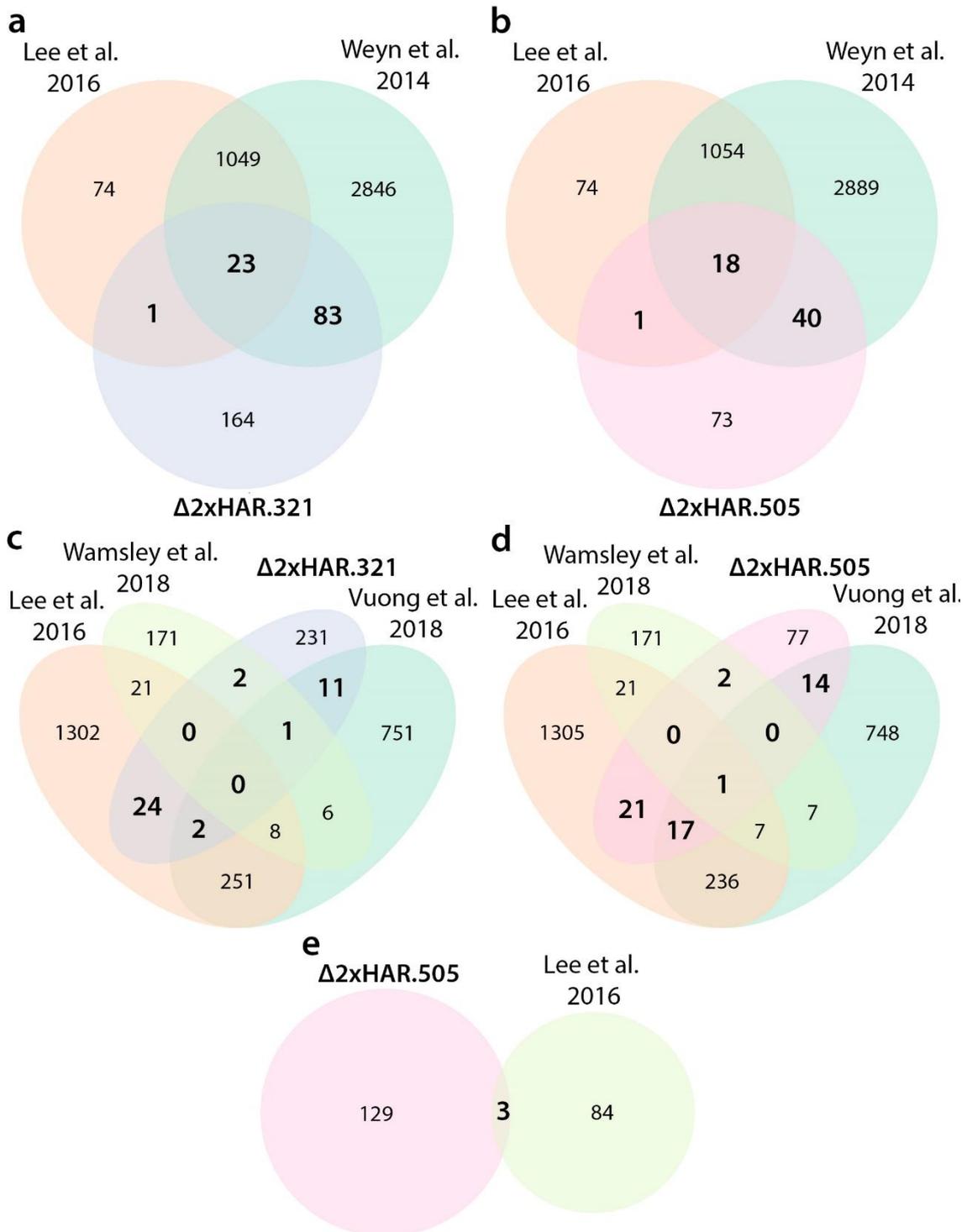


FIGURA 22: Solapamiento con información disponible de los genes regulados por *Rbfox1*. (a-b) Intersección de genes diferencialmente expresados en ratones (a) *Rbfox1* ^{$\Delta 2xHAR.321/\Delta 2xHAR.321$} y (b) *Rbfox1* ^{$\Delta 2xHAR.505/\Delta 2xHAR.505$} con datos de iCLIP-Seq de *Rbfox1/2/3*. (c-d) Intersección de genes diferencialmente expresados en ratones (c) *Rbfox1* ^{$\Delta 2xHAR.321/\Delta 2xHAR.321$} y (d) *Rbfox1* ^{$\Delta 2xHAR.505/\Delta 2xHAR.505$} con genes regulados por *Rbfox1* determinados en ensayos *knock-down* y *knock-out*. (e) Intersección de genes diferencialmente expresados en ratones *Rbfox1* ^{$\Delta 2xHAR.505/\Delta 2xHAR.505$} con pares de sitios *Rbfox1*-miARN determinados por iCLIP-Seq.

Análisis del uso diferencial de exones

Teniendo en cuenta que *Rbfox1* se desempeña principalmente como factor de *splicing* en el sistema nervioso, a continuación evaluamos el uso diferencial de exones en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} a partir de los datos obtenidos de la secuenciación del ARNm. Por uso diferencial de exones, nos referimos a cambios en el uso relativo de exones, que se define como el número de transcritos del gen que contiene el exón sobre el número total de transcritos del gen. Además, el uso diferencial de exones es un concepto más general que el *splicing* alternativo, ya que también incluye cambios en el uso de sitios alternativos del inicio de la transcripción y sitios de poliadenilación, lo que puede causar un uso diferencial de exones en los extremos 5' y 3' de los transcritos (Anders et al., 2012). De esta manera, observamos 100 diferencias en el uso de exones en 93 genes en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} con un p-valor ajustado por debajo de 0.05 (Apéndice IV, Tabla S5) y 15 diferencias en 13 genes en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} con un p-valor ajustado por debajo de 0.07 (Apéndice IV, Tabla S6). En ambos casos, la mayoría de las diferencias en el uso de exones se encuentran en las regiones 5' y 3' UTR de los transcritos, y solamente 18 (18%) y 3 (20%) de estas se encuentran en exones en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}, respectivamente.

Uno de los genes con uso diferencial de exones en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} es *Dnm1* (*Dynamín 1*), una GTPasa esencial para la liberación de neurotransmisores a través de la endocitosis de vesículas sinápticas (Blue et al., 2018). Típicamente tiene cinco dominios funcionales: un dominio de unión a GTP, un dominio medio que participa en la oligomerización, un dominio efector de GTPasa (GED), un dominio homólogo a Pleckstrina (PH) y un dominio rico en prolina (PRD) (Ferguson and De Camilli, 2012). Además, se han identificado mutaciones *de novo* en *DNM1* en pacientes con encefalopatías epilépticas infantiles, la mayoría de las cuales están localizadas en los dominios GTPasa y medio, y algunas en el dominio PH (von Spiczak et al., 2017). Dicho esto, observamos una disminución significativa en el exón 20, que forma parte del dominio GED (Figura 23). El *splicing* de un par de exones alternativos (10a y 10b) en *Dnm1* está bien documentado, pero no en el exón 20 que, de hecho, no es un exón alternativo, es decir que está presente en todas las isoformas.

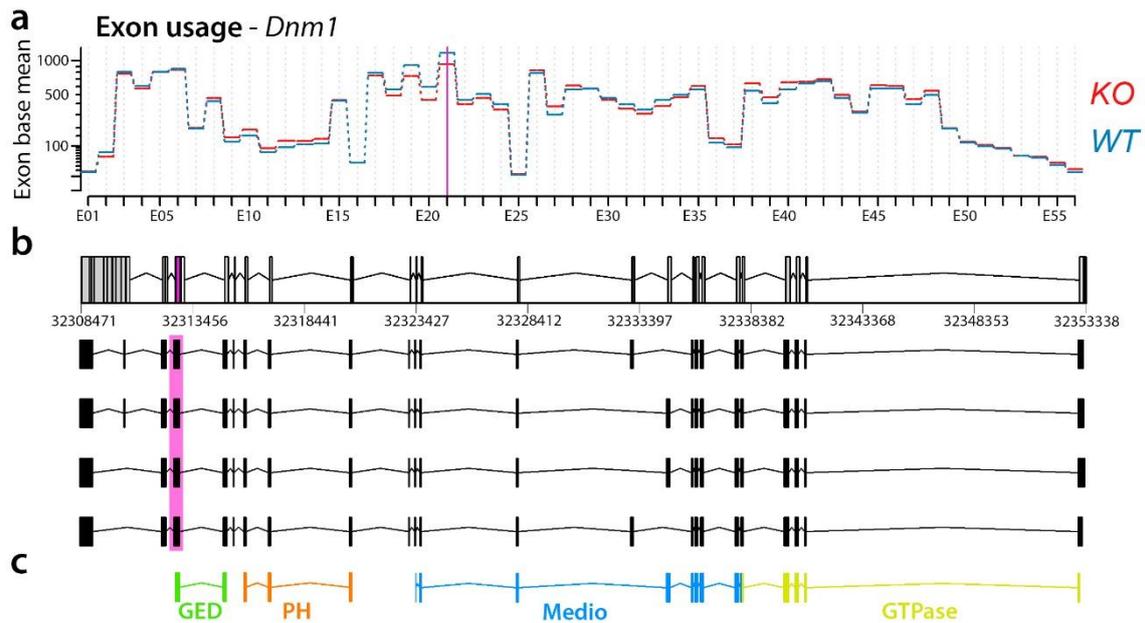


FIGURA 23: Uso diferencial de exones en *Dnm1*. (a) Visualización de los valores de expresión de cada uno de los exones del gen, separado por genotipo. (b) Algunos transcritos de *Dnm1*. (c) Dominios de la proteína.

En cuanto a los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}, uno de los genes con uso diferencial de exones es *Sipa1l2* (*Signal induced proliferation associated 1 like 2*), una proteína activadora de GTPasa (Spilker et al., 2008) involucrada en la autofagia neuronal (Andres-Alonso et al., 2019). En este caso observamos un aumento significativo en el exón 14 (Figura 24), que tampoco se ha documentado que sea un exón alternativo.

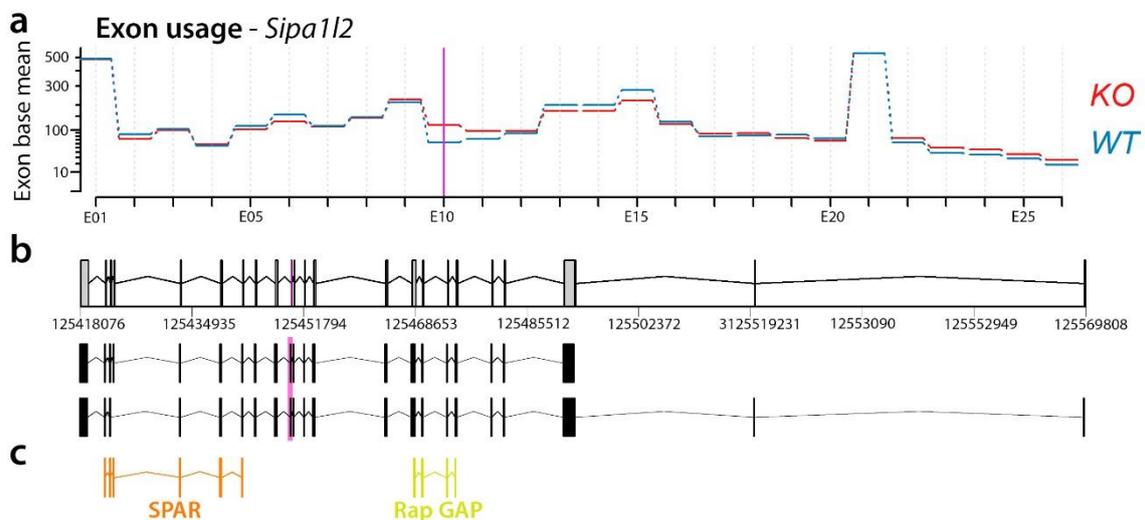


FIGURA 24: Uso diferencial de exones en *Sipa1l2*. (a) Visualización de los valores de expresión de cada uno de los exones del gen, separado por genotipo. (b) Algunos transcritos de *Sipa1l2*. (c) Dominios de la proteína.

Luego, evaluamos si las proteínas Rbfox se unen a los genes con uso diferencial de exones en los ratones mutantes. Para ello, utilizamos los datos de iCLIP-Seq de *Rbfox1*, *Rbfox2* y *Rbfox3* de uno de los trabajos mencionados anteriormente, pero esta vez utilizamos

únicamente la información de los sitios de unión en los intrones de los ARNm (Weyn-Vanhentenryck et al., 2014). De este modo, observamos que las proteínas Rbfox se unen en los intrones de 44 (47,3%) y 5 (38,4%) genes con uso diferencial de exones en los ratones *Rbfox1* ^{Δ 2xHAR.321/ Δ 2xHAR.321} (Figura 25a; Apéndice IV, Tabla S5) y *Rbfox1* ^{Δ 2xHAR.505/ Δ 2xHAR.505} (Figura 25b; Apéndice IV, Tabla S6), respectivamente.

Además, utilizamos la información de *splicing* disponible de la delección total de *Rbfox1* de tres trabajos, dos de los cuales mencionamos anteriormente. A manera de recordatorio, en uno de los trabajos evaluaron las diferencias de *splicing* en hipocampos de ratones mutantes nulos condicionales del sistema nervioso (*Rbfox1*^{*loxp/loxp*/Nestin-Cre^{+/-}) (Vuong et al., 2018) y en el otro en interneuronas corticales de ratones mutantes nulos condicionales de interneuronas SST+ (*Rbfox1*^{*loxp/loxp*/SST-Cre^{+/-}) y PV+ (*Rbfox1*^{*loxp/loxp*/Tac1-Cre^{+/-}) (Wamsley et al., 2018). En el tercer trabajo, también utilizaron los ratones mutantes nulos condicionales del sistema nervioso (*Rbfox1*^{*loxp/loxp*/Nestin-Cre^{+/-}) pero en cerebros enteros (Lovci et al., 2013). Con esta información, observamos que 16 (17,5%) y 2 (15,3%) genes con uso diferencial de exones en los ratones *Rbfox1* ^{Δ 2xHAR.321/ Δ 2xHAR.321} (Figura 25c; Apéndice IV, Tabla S5) y *Rbfox1* ^{Δ 2xHAR.505/ Δ 2xHAR.505} (Figura 25d; Apéndice IV, Tabla S6), respectivamente, son genes regulados por *Rbfox1*.}}}}

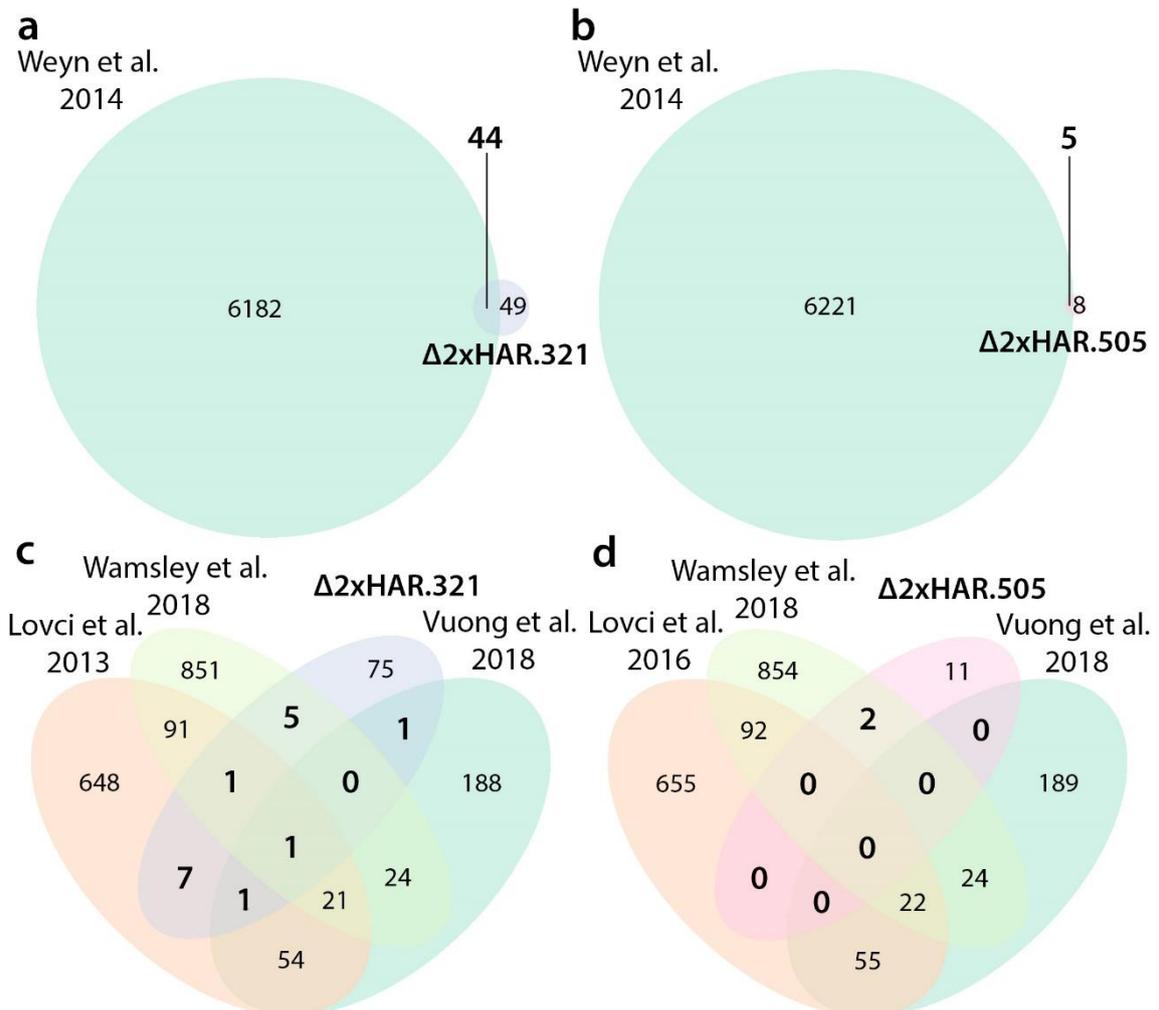


FIGURA 25: Solapamiento con información del *splicing* de los genes regulados por *Rbfox1*. (a-b) Intersección de genes con uso diferencial de exones en ratones (a) *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y (b) *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} con datos de iCLIP-Seq de *Rbfox1/2/3*. (c-d) Intersección de genes con uso diferencial de exones en ratones (c) *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y (d) *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} con genes regulados por *Rbfox1* determinados en ensayos *knock-out*.

Por último, no observamos solapamiento entre los genes diferencialmente expresados de los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}, así como tampoco en el uso diferencial de exones. Además, tampoco hubo solapamiento entre los genes con expresión diferencial y los genes con uso diferencial de exones en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}, mientras que en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} observamos que tres genes (*Rpl15*, *Taf10* y *Gm5148*) tuvieron diferencias tanto en la expresión como en el uso de exones.

Discusión de los resultados

Diferencias entre técnicas: RT-qPCR vs RNA-Seq

En este capítulo generamos ratones genéticamente modificados mediante CRISPR/Cas9 que carecen de los *enhancers* 2xHAR.321 o 2xHAR.505 para evaluar el impacto funcional *in vivo* de cada delección. Debido a la redundancia en los dominios de expresión de los *enhancers* que descubrimos en la primera parte del trabajo, así como también en las funciones de los parálogos de *Rbfox1*, sabíamos que era un objetivo ambicioso. Sin embargo, esta estrategia nos permite evaluar la importancia relativa de cada *enhancer* en la transcripción de *Rbfox1*, aunque el fenotipo resultante sea sutil.

En primer lugar, evaluamos los niveles de expresión de *Rbfox1* mediante RT-qPCR en distintos tejidos del sistema nervioso a los 14.5 días embrionarios y a los 8 días postnatales. La decisión de estas edades se debe a que a E14.5 la expresión de *Rbfox1* es robusta en el sistema nervioso (Hammock and Levitt, 2011) y va en aumento hasta alcanzar el pico de expresión a P8, donde se mantiene hasta las tres semanas postnatales para luego disminuir notablemente (Casanovas et al., 2020). Con esta estrategia, observamos que los ratones mutantes nulos de ambos *enhancers* no presentaron diferencias significativas en la mayoría de los tejidos y estadios analizados. Solamente observamos un aumento del 15% en la expresión de *Rbfox1* en el mesencéfalo a E14.5 en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} y un aumento del 25% en la corteza cerebral a P8 en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} en comparación con los ratones de tipo salvaje. Dados estos resultados, decidimos secuenciar el ARNm en las edades y tejidos donde observamos que *Rbfox1* tenía diferencias significativas de expresión para explorar el transcriptoma de cada genotipo en detalle. Aunque encontramos algunos genes diferencialmente expresados en ambos ratones mutantes, sorprendentemente *Rbfox1* no era uno de ellos, a pesar de haber observado diferencias significativas mediante RT-qPCR.

Entonces, ¿qué resultados son más confiables? Al comparar las tecnologías, ambas ofrecen una detección altamente sensible y confiable y solo se diferencian en el poder de descubrimiento, ya que con la qPCR solo se pueden detectar secuencias conocidas,

mientras que la secuenciación es un enfoque libre de hipótesis que no requiere conocimiento previo de la información de la secuencia. En cuanto a las posibles dificultades técnicas, para la qPCR utilizamos SYBR Green para la detección de la señal, un intercalante del ADN que detecta todo ADN doble cadena, incluidos los productos de reacción no específicos. Aunque es altamente improbable que este sea el problema, ya que diseñamos los *primers* de manera específica y para que no amplifique ADN genómico, se podría solucionar utilizando una sonda TaqMan, que permite la detección específica del producto deseado. Además, si este fuera el problema, hubiéramos observado diferentes curvas de *melting* para cada producto durante la amplificación. Por otro lado, las diferencias en los métodos de extracción del ARN (fenol-cloroformo vs columnas de sílice), así como la síntesis del ADNc (*primers* aleatorios vs captura poli-A) no deberían tener un impacto biológico. La realidad es que no es justo decir que una técnica es mejor que la otra. Ambas ofrecen una cuantificación sensible y confiable, y no hay ninguna razón para confiar más en un resultado que en el otro.

Recientemente se publicó un estudio exhaustivo donde se compararon cinco métodos de análisis de secuenciación con resultados de qPCR (Everaert et al., 2017; Coenye, 2021). Una de las principales conclusiones fue que, dependiendo del método de análisis, entre el 15 y 20% de los datos de secuenciación no coincidieron con los resultados obtenidos por qPCR. Este porcentaje de genes “no concordantes” consiste principalmente en genes con cambios en el nivel de expresión entre grupos relativamente bajos. Por ejemplo, el 93% de los genes “no concordantes” tienen un valor de log₂FC menor a 2, independientemente del método de análisis. Del pequeño grupo de genes con un log₂FC mayor a 2, la gran mayoría se expresa a niveles bajos en el primer cuartil de expresión, son significativamente más pequeños y tienen menos exones. En cambio, los genes con un log₂FC menor a 2 se distribuyen por igual en todos los cuartiles de expresión. Como las deleciones de 2xHAR.321 y 2xHAR.505 en los ratones mutantes generan diferencias en los niveles de expresión de *Rbfox1* moderadas en comparación con los ratones de tipo salvaje, entonces esto podría explicar las diferencias entre los resultados. Es probable que las diferencias de expresión observadas en los ensayos de RT-qPCR están fuera del umbral de detección de la secuenciación del ARNm. De hecho, todos los genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321} tienen un log₂FC menor a 1, y en el caso de los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505} el 97% de los genes tiene un log₂FC menor a 2, y los cuatro genes que tienen un log₂FC mayor (*Moap1*, *Lhx8*, *Six3* y *Otp*), tienen niveles de expresión extremadamente bajos.

¿Es 2xHAR.321 un silenciador?

A pesar de la creciente evidencia de la importancia de los silenciadores en el desarrollo, en general se han estudiado menos que los *enhancers*, ya sea por dificultades técnicas en los ensayos reporteros o porque la transcripción se regula principalmente a nivel de la activación. De hecho, la búsqueda de la palabra “*enhancers*” en PubMed lanza 1.8 millones de resultados frente a solo ~120 mil para “silenciadores”. Sin embargo, en los últimos años ha habido un gran avance hacia la descripción e identificación de los elementos silenciadores (para una revisión ver (Segert et al., 2021)). En particular, en un

trabajo reciente evaluaron miles de regiones genómicas accesibles que carecían de características típicas de promotores, *enhancers* e *insulators* y analizaron su capacidad de actuar como silenciadores, descubriendo ~3200 elementos con actividad silenciadora en ensayos MPRA en células K562 humanas (Doni Jayavelu et al., 2020). A su vez, estas regiones están enriquecidas en sitios de unión a factores de transcripción represores, como *REST* (*RE1 silencing transcription factor*), *FOSL2* (*FOS like 2, AP-1 transcription factor*), *JUN* (*Jun proto-oncogene, AP-1 transcription factor*), *BATF* (*Basic leucine zipper ATF-like transcription factor*) y miembros de la familia RFX (*Regulatory factor X*). Luego entrenaron un modelo computacional para predecir silenciadores, y descubrimos que dentro de la región conservada que contiene a 2xHAR.321 hay dos potenciales silenciadores. Además, mediante un análisis *in silico* específico de los factores de transcripción represores observamos que 2xHAR.321 contiene potenciales sitios de unión a *JUN*, *FOSL2*, *RFX6* y *BATF* (Figura 26).

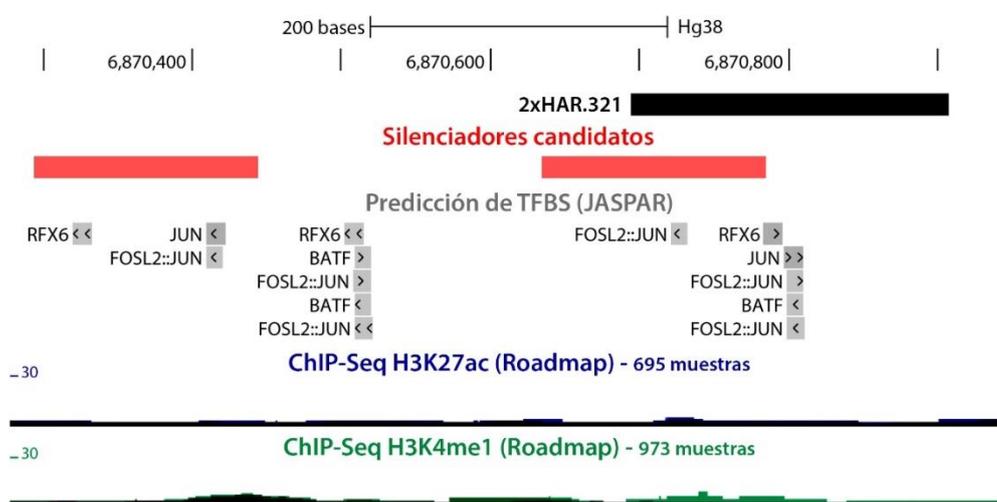


FIGURA 26: Elementos silenciadores candidatos. Se incluye la región acelerada 2xHAR.321, una predicción de los TFBS represores (JASPAR) y señales de CHIP-Seq de H3K27ac y H3K4me1 de una combinación de muestras de distintos tejidos y tipos celulares (Roadmap).

En el primer capítulo de este trabajo observamos que en peces cebra transgénicos, la secuencia humana y de ratón de 2xHAR.321 se comportan como *enhancers* fuertes del sistema nervioso, por lo que esperábamos observar una disminución en la expresión de *Rbfox1* en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321}. Sin embargo, observamos un aumento significativo de aproximadamente el 15% en el mesencéfalo a E14.5. Existen al menos dos razones que pueden explicar estos resultados. Por un lado, puede ser que en su contexto genómico estándar, 2xHAR.321 es efectivamente un elemento silenciador, pero en los ensayos reporteros, donde la región se encuentra fuera de contexto, predomina la activación de la expresión del reportero. Además, en los peces transgénicos generados con el sistema Tol2 se insertan en el genoma múltiples copias del transgén en distintos *loci* de los cuales no tenemos información. Es por esta razón que, de ser posible, es mejor realizar ensayos de transgénesis dirigidos de copia única. En el caso ideal, cuando el elemento se encuentra conservado en el modelo animal, lo más adecuado es realizar ensayos *knock-in*, donde se intercambia la secuencia original por la secuencia candidata sin modificar el contexto genómico. Si el elemento no se encuentra conservado, se podría

realizar una transgénesis dirigida a un *locus* neutral. Por ejemplo, en ratones, el *locus* ROSA26 se ha convertido en un sitio ampliamente utilizado para la integración de diversas construcciones reporteras (Zambrowicz et al., 1997), aunque, en este caso, se pierde la información del contexto genómico. Por otro lado, puede ser que 2xHAR.321 sea un elemento regulatorio bifuncional, es decir que actúa tanto como *enhancer* como silenciador en distintos tipos celulares. Aunque el conjunto actual de evidencias sugiere que la bifuncionalidad es más prevalente de lo que se creía, la secuencia de 2xHAR.321 no contiene las marcas epigenéticas características de *enhancers*, como H3K4me1 y H3K27ac (Figura 26), por lo que es probable que 2xHAR.321 no sea un elemento regulatorio bifuncional.

Genes regulados por Rbfox1

Gracias a la información disponible de trabajos anteriores, pudimos verificar que un alto número de los genes diferencialmente expresados en los ratones mutantes son genes regulados por *Rbfox1*. En primer lugar, los datos de iCLIP-Seq proporcionaron información sobre la interacción entre las proteínas Rbfox y los ARNm (Weyn-Vanhentenryck et al., 2014; Lee et al., 2016). En general, estos porcentajes fueron más altos, posiblemente porque es una combinación de la interacción de los tres parálogos. Por otro lado, ninguno de los tres trabajos que utilizamos para comparar los genes regulados por *Rbfox1* son totalmente comparables con nuestros resultados, especialmente los del mesencéfalo, ya que son distintos tejidos o tipos celulares y distintos estadios (Lee et al., 2016; Vuong et al., 2018; Wamsley et al., 2018). De todos modos, debido a que en nuestro trabajo realizamos deleciones de *enhancers* redundantes, que producen cambios en la expresión de *Rbfox1* sumamente sutiles, valoramos positivamente haber encontrado un alto número de genes regulados por *Rbfox1* en los ratones mutantes nulos.

Por otra parte, como mencionamos anteriormente, el uso diferencial de exones es un concepto más amplio que el *splicing* alternativo, ya que también incluye cambios en el uso de sitios alternativos del inicio de la transcripción y sitios de poliadenilación. Dicho esto, la mayoría de las diferencias que observamos en el uso de exones en los ratones mutantes se encontraba en las regiones UTR, y no en los exones alternativos como esperábamos dada la función de *Rbfox1* como factor de *splicing*. Es probable que la redundancia en los genes blanco de los parálogos de *Rbfox1*, que ya se ha descrito en los ratones mutantes condicionales del sistema nervioso de *Rbfox1* (Gehman et al., 2011), sea la razón de los casi nulos cambios en el *splicing* alternativo. Además, hay otros análisis que se enfocan en los eventos de *splicing* en vez del uso diferencial de exones que podríamos evaluar en el futuro.

Comparación con otros ratones con deleciones en enhancers

Existen varios trabajos en los que se generaron ratones mutantes nulos para distintos *enhancers*, cada uno con diferencias en la severidad del fenotipo. Por ejemplo, los ratones con deleciones individuales de 10 *enhancers* ubicados cerca de genes asociados con malformaciones congénitas de las extremidades, no exhiben cambios notables en la morfología de estas. En cambio, la combinación de estas deleciones con una expresión

basal reducida de los genes blanco, a partir de deleciones heterocigotas, genera fenotipos discernibles en las extremidades, similares en gravedad a los fenotipos de los ratones homocigotas (Osterwalder et al., 2018). Asimismo, los ratones que carecen de dos *enhancers* de *Tbx5* (*T-box transcription factor 5*) presentan una morfología normal de las extremidades, a diferencia de la deleción total de *Tbx5* que da como resultado la pérdida completa de las extremidades anteriores (Cunningham et al., 2018).

Por otro lado, hay mutantes en los que los fenotipos son más evidentes. Por ejemplo, los ratones que carecen de los *enhancers* ultraconservados ubicados cerca del factor de transcripción neuronal *Arx* (*Aristaless related homeobox*) son viables y fértiles, pero muestran anomalías neurológicas, incluyendo alteraciones en las poblaciones de neuronas y defectos cerebrales estructurales (Dickel et al., 2018). A su vez, la deleción de un *enhancer* distal de *Shh* (*Sonic hedgehog*) denominado ZRS da como resultado extremidades anteriores y posteriores truncadas (Sagai et al., 2005), y el intercambio de la secuencia ZRS de ratón por la secuencia ortóloga de cobra, que no posee extremidades, recapitula el fenotipo de la deleción de ZRS (Kvon et al., 2016). Además, las mutaciones puntuales en la secuencia ZRS en humanos causan malformaciones en las extremidades, más comúnmente polidactilia preaxial. La introducción de siete variantes humanas potencialmente patogénicas en la secuencia ZRS de ratón, resultó en la formación de dígitos adicionales, reproduciendo los fenotipos encontrados en humanos con polidactilia y confirmando la patogenicidad de las variantes (Kvon et al., 2020). Por último, la deleción simultánea de dos *enhancers* hipotalámicos de *Pomc* (*Proopiomelanocortin*), denominados nPE1 y nPE2, resulta en una reducción del ~90% de la expresión de *Pomc* y un fenotipo de obesidad extremo. Asimismo, la deleción de uno u otro da lugar a una reducción del ~75% de la expresión de *Pomc* durante el desarrollo embrionario temprano, lo que indica que ambos *enhancers* funcionan de manera sinérgica. En ratones adultos, la expresión de *Pomc* se controla de forma aditiva mediante ambos *enhancers*, siendo nPE1 el responsable del 80% y nPE2 del 20% de la transcripción. En consecuencia, los ratones deficientes de nPE1 presentan una obesidad leve, mientras que los ratones deficientes de nPE2 mantienen un peso corporal normal (Lam et al., 2015).

Las diferencias en la gravedad del fenotipo en los ratones mutantes se pueden deber a una variedad de razones, como por ejemplo la función del gen al que regulan, la cantidad de *enhancers* que participan en la regulación y la redundancia de estos. Por ejemplo, *Pomc* es un gen relativamente pequeño que tiene solamente dos exones codificantes y cuya unidad transcripción ocupa 5.7 kb. Además, hasta el momento sólo hay descritos dos *enhancers* que regulan a *Pomc* en el hipotálamo (de Souza et al., 2005). En cambio, *Rbfox1* es uno de los genes más grandes del genoma y posee decenas de regiones conservadas en sus regiones intrónicas que podrían ser *enhancers*. A su vez, en este trabajo encontramos varios *enhancers* con dominios de expresión redundantes. Entonces, considerando la falta de fenotipos notables en las deleciones de los *RBFOX1*-HARs, en el futuro podríamos aprovechar la estrategia mencionada en uno de los ejemplos (Osterwalder et al., 2018) y combinar las deleciones de los HARs con ratones que tengan un *background* genético con expresión reducida de *Rbfox1*, por ejemplo ratones mutantes heterocigotas.

CONCLUSIONES

Principales resultados

Descifrar cómo los cambios en las redes regulatorias de la expresión génica contribuyen a los rasgos humano-específicos es uno de los grandes propósitos de la biología evolutiva. Determinar tales cambios y establecer su impacto funcional es un gran desafío, principalmente debido a las dificultades para predecir la función de las secuencias no codificantes. Sin embargo, en los últimos años ha habido un gran avance hacia la identificación de potenciales impulsores genéticos de la innovación evolutiva humana. En particular, las Regiones Aceleradas en Humanos (HARs) son fuertes candidatos para codificar elementos regulatorios en *cis* cuya función podría haberse modificado en el linaje humano. En este trabajo postulamos que el factor de *splicing* neuronal *RBFOX1* (*RNA-binding Protein, fox-1 homolog*) podría haber sido blanco de una evolución acelerada a nivel regulatorio en el linaje humano, que tuvo como consecuencia la ganancia y/o pérdida de funciones regulatorias.

Mediante ensayos de gen reportero en peces cebra (*Danio rerio*) transgénicos evaluamos si los *RBFOX1*-HARs son capaces de actuar como *enhancers* transcripcionales y comparamos su actividad con sus secuencias ortólogas de chimpancé. Con esta estrategia descubrimos que 15 de los 19 HARs son *enhancers* activos durante el desarrollo en peces cebra. De estos, cinco tienen patrones de expresión indistinguibles entre las secuencias ortólogas, y los otros diez exhiben diferencias de expresión, seis con actividad únicamente en la secuencia humana y cuatro con actividad solamente en la secuencia de chimpancé. Además, seleccionamos dos *RBFOX1*-HARs para extender el análisis de expresión con una tercera especie como grupo externo. En el caso de 2xHAR.505 confirmamos la pérdida de función en el linaje humano, pero para 2xHAR.321 observamos que el estado derivado es el de la secuencia de chimpancé, de manera que no podemos descartar que sea una pérdida de función en el linaje chimpancé.

En la segunda parte, empleamos un enfoque de edición génica en ratones mediante CRISPR/Cas9 para eliminar, de manera independiente, la secuencia de 2xHAR.321 o 2xHAR.505, y establecer el impacto funcional *in vivo* de cada deleción. Mediante RT-qPCR observamos diferencias de expresión a los 14.5 días embrionarios en el mesencéfalo en los ratones *Rbfox1* ^{Δ 2xHAR.321/ Δ 2xHAR.321} y a los 8 días postnatales en la corteza cerebral en los ratones *Rbfox1* ^{Δ 2xHAR.505/ Δ 2xHAR.505}. Además, mediante secuenciación del ARNm observamos que un alto número de genes diferencialmente expresados en los ratones *knock-out* son genes regulados por *Rbfox1*. En cambio, casi no observamos cambios en el *splicing* alternativo, probablemente debido a una redundancia en las funciones de las proteínas de la familia Rbfox.

Implicancias de los hallazgos

Los cambios en las regiones regulatorias permiten modificaciones significativas en la expresión génica de *RBFOX1* sin alterar su función esencial en el sistema nervioso. *RBFOX1* integra la entrada de múltiples estímulos fisiológicos para modular la salida de isoformas estructural y funcionalmente diversas, y utiliza este conjunto de isoformas para ejecutar controles de gran alcance sobre los eventos de procesamiento del ARN (Conboy, 2017). Este último incluye no sólo la regulación del *splicing* alternativo (Jin et al., 2003), sino también la estabilidad del ARNm (Lee et al., 2016). De acuerdo con su complejidad celular y funcional, el sistema nervioso de los mamíferos hace un amplio uso de la regulación del *splicing* alternativo para generar isoformas especializadas que afectan a todos los aspectos del desarrollo y la función neuronal (Vuong et al., 2016). Aunque la delección neuronal de *Rbfox1* en los ratones mutantes condicionales genera relativamente pocos cambios en el *splicing* alternativo, afecta genes que codifican canales iónicos, receptores de neurotransmisores, proteínas estructurales del sistema nervioso y moléculas de señalización de calcio (Gehman et al., 2011). Cambios en la relación entre isoformas de estas proteínas pueden aumentar o disminuir el potencial de acción o perturbar el equilibrio de excitación-inhibición en los circuitos neuronales. Por ejemplo, *Snap25* utiliza un par de exones excluyentes (5a y 5b) para producir dos isoformas alternativas que tienen diferencias en la cinética de liberación de las vesículas sinápticas (Bark et al., 2004), y *Scn8a* tiene varios exones mutuamente excluyentes que pueden alterar la activación dependiente de voltaje (5a y 5n) o generar isoformas truncadas que carecen del dominio transmembrana (18a y 18n) (O'Brien et al., 2012). A su vez, a través de la unión a la región 3' UTR, *Rbfox1* promueve la estabilidad del ARNm de *Vamp1* (*Vesicle associated membrane protein 1*) en las neuronas inhibitorias, antagonizando la acción del microARN-9, lo que genera una disminución de la transmisión sináptica inhibitoria (Vuong et al., 2018). Además, la disfunción de *RBFOX1* en humanos se ha relacionado con epilepsia (Bhalla et al., 2004; Lal et al., 2013a, 2013b) y trastorno de espectro autista (Martin et al., 2007; Sebat et al., 2007; Voineagu et al., 2011), dos afecciones del sistema nervioso que a menudo son comórbidas. Entonces, así como cualquier interrupción en la regulación de *RBFOX1* puede generar un desbalance en los circuitos neuronales, la ganancia y pérdida de elementos regulatorios en el linaje humano podría haber generado nuevos patrones de expresión que hayan contribuido a la evolución y desarrollo del cerebro humano. En esta tesis encontramos al menos 10 regiones regulatorias con expresión diferencial en el linaje humano que potencialmente podrían haber resultado en cambios en el patrón de expresión de *RBFOX1*. Si esta ganancia y pérdida de funciones regulatorias resultaron en cambios en la expresión de *RBFOX1* en el linaje humano, confiriéndole algún rol en la evolución de nuestro cerebro, resta por ser investigado. Nuestros hallazgos agregan a *RBFOX1* a la selecta lista de genes del neurodesarrollo que experimentaron evolución acelerada en el linaje humano, con consecuencias en la regulación de la expresión génica. Aunque se necesitan más experimentos, este trabajo representa un paso más hacia el entendimiento de los cambios genéticos humano-específicos y de sus consecuencias fenotípicas.

MATERIALES Y MÉTODOS

Análisis bioinformáticos

Identificación de las regiones aceleradas en el linaje humano

Las regiones evolutivamente aceleradas en el linaje humano que se utilizan en este trabajo surgieron de una base de datos no redundante conteniendo 1629 elementos generada a partir de cinco bases de datos públicas que, de manera independiente y con algoritmos bioinformáticos desarrollados para tal fin, recopilan información acerca de la aceleración en el linaje humano (Kamm et al., 2013b). En las bases de datos originales los elementos acelerados se denominan *Human Accelerated Regions* (HAR) (Pollard et al., 2006), *Human Accelerated Conserved Non-Coding Sequences* (HACNS) (Prabhakar et al., 2006), *Human Terminal Branch Elements* (HTBE) (Bush and Lahn, 2008), *2nd Generation (2x) Human Accelerated Regions* (2xHAR) (Lindblad-Toh et al., 2011) y *Accelerated Conserved Non-Coding Sequences* (ANC) (Bird et al., 2007). Las localizaciones genómicas se visualizaron gráficamente en la versión humana Hg38, en la versión de chimpancé PanTro5 y en la versión de ratón Mm10 del UCSC *Genome Browser* (<https://genome.ucsc.edu/>).

Predicción de sitios de unión a factores de transcripción

El análisis en todo el genoma de los sitios de unión a factores de transcripción (TFBS) predichos se realizó con la base de datos de acceso público JASPAR (<https://jaspar.genereg.net/collection/core/>), que tiene una plataforma o *track* disponible (JASPAR2022) en el UCSC *Genome Browser* para la visualización en el genoma humano (Hg38) y de ratón (Mm10). La comparación de los TFBS en el genoma humano vs el genoma de ratón permitió seleccionar potenciales sitios que se pudieron haber ganado o perdido en el genoma humano. La vista predeterminada muestra sólo los sitios con puntuación de 400 o más (p-valor = 0,0001). Una vez identificados los TFBS de interés, se confirmó la presencia o ausencia del sitio en la secuencia de chimpancé con la herramienta “Escanear” de JASPAR disponible en línea. Para ello, primero se selecciona el perfil de unión del factor de transcripción candidato, se agrega al carrito y se incluye la secuencia que se desea escanear. Para que el análisis sea lo más sensible y específico posible, se utilizó 85% como umbral de puntuación relativa.

Selección positiva en las secuencias no codificantes

El análisis de selección positiva se basó en el modelo integrado que evalúa los alineamientos de secuencia en busca de evidencia de selección, sesgo GC o ambas (Kostka et al., 2012). Para el análisis de PhyloP de las regiones codificantes, se descargaron múltiples alineamientos anotados, en formato MAF, de Multiz20Way de goldenPath (<http://hgdownload.cse.ucsc.edu/goldenpath/hg38/multiz20way/>). El árbol filogenético también se descargó de goldenPath y el modelo de árbol local se generó con

el programa phyloFit (modelo REV) del paquete rphast usando múltiples alineamientos de sitios 4 veces degenerados (4d) extraídos del alineamiento de primates del bloque sintético 109, obtenido a través de SyntenyPortal (http://bioinfo.konkuk.ac.kr/synteny_portal/). Estos bloques son regiones genómicas que se conservan entre múltiples especies, donde los rasgos proporcionados por estos genes generalmente se heredan juntos. Para construir el modelo neutral, los sitios 4d se derivaron del conjunto de genes RefSeq de la versión del genoma humano Hg38 de esta región. El programa phastCons se usó para calcular regiones que se conservan en los alineamientos de especies de primates. Los parámetros utilizados fueron: longitud-esperada=45, cobertura-objetivo=0.3, rho=0.3 y opción de viterbi VERDADERO. A los elementos conservados predichos se les asignó una puntuación y se sometieron a análisis a través de phyloP. Este programa calculó los p-valores de aceleración para cada elemento en la rama humana. Las puntuaciones para la aceleración se produjeron utilizando una prueba de razón de probabilidad en cada columna de alineación.

Selección positiva en las secuencias codificantes

Se descargaron las secuencias codificantes de 57 especies de Ensembl (<https://ensembl.org/>) utilizando la herramienta de ortólogos y se curaron manualmente para completar los datos faltantes utilizando BLAST de NCBI (<https://blast.ncbi.nlm.nih.gov/>). Las secuencias se alinearon usando MUSCLE en MEGA7. El modelo A de sitio ramificado modificado de selección positiva se llevó a cabo en secuencias de mamíferos usando PAML. Las ramas focales fueron los linajes humanos y primates. Se probaron dos hipótesis anidadas para cada rama focal. En la hipótesis alternativa, se calcularon tres valores de ω : ω_0 , ω_1 y ω_2 que representan los codones bajo selección negativa, nula y positiva, respectivamente. En la hipótesis nula, solo se estimaron dos valores de ω (ω_0 y ω_1) y no se permitieron sitios seleccionados positivamente.

Producción de peces cebra transgénicos

Construcción de los transgenes

Las regiones conservadas *phastCons* que contienen a las secuencias aceleradas se amplificaron a partir de ADN genómico humano (Hs), de chimpancé (Pt) o de ratón (Mm) en una termocicladora para PCR (Dyad, BioRad) usando una polimerasa procesiva Taq (Invitrogen), 2.25 mM de MgCl₂, 1.5% de dimetilsulfóxido (DMSO) y *primers* específicos diseñados con Primer3 (<http://primer3.ut.ee/>) detallados en el Apéndice I. En el extremo 5' de los *primers forward* se agregaron cuatro bases nucleotídicas (CACC) que permiten el clonado direccionado de los insertos en el vector de entrada. El programa usado fue el siguiente:

- 1) Desnaturalización inicial: 95°C por 5 minutos.
- 2) Desnaturalización: 94°C por 30 segundos.
- 3) *Annealing*: 56-65°C por 40 segundos.
- 4) Extensión: 72°C por 1 minuto cada una kilobase.

- 5) Repetir pasos 2-4, 35 ciclos.
- 6) Extensión final: 72°C por 5 minutos.

Una vez amplificados, los fragmentos de PCR se aislaron en un gel de agarosa 1% y se purificaron con el kit comercial para extracción de gel *QIAquick® Gel Extraction Kit* (Qiagen). Los insertos purificados se clonaron direccionalmente por medio del sistema TOPO en el vector de entrada pENTRTM/D-TOPO (Invitrogen), que utiliza una Topoisomerasa I proveniente de *Vaccinia* para recombinar los insertos doble cadena a los extremos libres del vector. Se transformaron bacterias *Escherichia coli* químicamente competentes One ShotTM TOP10 (Invitrogen) mediante choque térmico y se seleccionaron colonias para pasar a un cultivo líquido de LB-kanamicina (50 µg/mL) como agente de selección. Finalmente, se aisló el vector con el kit comercial *Wizard® Plus SV Minipreps DNA Purification System* (Promega). Los vectores purificados se secuenciaron en Macrogen (<https://dna.macrogen.com/>) para verificar que no hayan existido mutaciones durante la amplificación.

Mediante el sistema Gateway (Invitrogen) se subclonaron los insertos desde el vector de entrada pENTRTM/D-TOPO (Invitrogen) hasta el vector destino pXIG-cFos (Fisher et al., 2006b). Se transformaron bacterias con ampicilina (100 µg/mL) como agente de selección y se aisló el vector como se explicó en el paso anterior. Brevemente, el sistema Gateway utiliza las enzimas recombinantes integrasa y excisionasa provenientes del bacteriófago λ, y el factor de integración de hospedador proveniente de *Escherichia coli*. Esta mezcla de enzimas, conocidas como clonasa, catalizan la recombinación *in vitro* entre el inserto del vector de entrada y el vector destino, direccionada entre los sitios attL presentes en el vector de entrada y los sitios attR presentes en el vector destino. El vector destino pXIG-cFos contiene el promotor mínimo del gen *cFos* de ratón río arriba de la secuencia codificante de la proteína reportera fluorescente *Enhanced Green Fluorescent Protein* (EGFP). Además, flanqueando la región que contiene el sitio de clonado Gateway, el promotor mínimo y el gen reportero, contiene las secuencias de reconocimiento del transposón Tol2 de Medaka (*Orzyas latipes*) que permite la generación de peces cebrá transgénicos con alta eficiencia (Kawakami, 2007).

Para el análisis de expresión de los promotores, se diseñó el *primer reverse* con un sitio de restricción de la enzima BamHI, ya que hay un sitio BamHI al final del promotor mínimo en el vector pXIG-cFos. De esta manera, primero se clonan direccionalmente en el vector de entrada pENTRTM/D-TOPO, se subclonan hasta el vector destino pXIG-cFos y se elimina el promotor mínimo del gen *cFos* mediante la digestión con BamHI y posterior circularización. Finalmente, se obtiene un vector que contiene al promotor que se quiere analizar seguido del gen reportero EGFP, flanqueado por las secuencias de reconocimiento del transposón Tol2.

Transcripción in vitro del ARNm de la transposasa Tol2

El ARNm de la transposasa Tol2 se preparó mediante una mini preparación de ADN plasmídico del vector pCS-Tp con el *HiSpeed® Plasmid Purification Kit* (Qiagen), se linealizó con la enzima de restricción NotI (New England Biolabs) y se purificó por

precipitación con etanol. La transcripción *in vitro* se realizó con el kit *mMESSAGE mMACHINE™ Sp6 Kit* (Invitrogen) y finalizada la reacción, el ADN molde se degradó con DNase I (Invitrogen). El ARN resultante se precipitó con cloruro de litio 0.4 M y 2 volúmenes de etanol frío. Finalmente, se resuspendió el ARNm en agua libre de nucleasas, se cuantificó y se llevó a una concentración de 175 ng/μL.

Animales y acuario

Los peces utilizados en este proyecto pertenecen a la cepa AB (Zebrafish International Resource Center, Universidad de Oregón, USA) y a la cepa exocriada obtenida del laboratorio de la Dra. Nora Calcaterra (IBR, Rosario). Los animales se crían en condiciones estándar en el acuario automatizado modelo ZF0601 (Aquatic Habitats) ubicado en el INGEBI-CONICET. Este sistema mantiene el agua en condiciones óptimas para la cría de los peces mediante aireación, desgaseado, filtración mecánica (*filter pad*), filtración biológica de amoníaco y nitritos (*Siporaz* y *Kaldnes*), control de temperatura, segunda filtración mecánica (50 μm *clartridge*), filtración química por medio de carbón activado y esterilización del agua por UV. Los animales se mantienen en peceras de policarbonato de 3 L de capacidad, a una densidad máxima de 6 peces adultos por tanque con un flujo de agua de 9 L/hora. Además, los animales se crían en un ciclo regular de 14 horas de luz y 10 horas de oscuridad, a 28±1°C, 7±1.5 mg/L de densidad de oxígeno, 600±100 μS de conductividad del agua y 7±1.5 de pH. Los peces juveniles y adultos reciben una dieta de alimento balanceado comercial (Larval AP100 y Adult Zebrafish Diet, Zeigler) dos veces al día suplementada con artemias frescas (INVE) una vez al día.

Microinyección de embriones de peces cebra

El día previo a la inyección, se armaron apareos en peceras más pequeñas donde hay dos compartimentos separados por una placa divisoria de plástico transparente. La base de estas peceras tiene una superficie con ranuras que permite separar los huevos fertilizados de los peces progenitores. Se colocan hembras y machos en cada uno de los compartimentos de la pecera de apareos, en una relación 2:1, dejándolos separados toda la noche por medio de la placa divisoria. El día de la inyección, poco después del inicio del ciclo de luz, se remueve la división que separa a los animales y dentro de los primeros 30 minutos de interacción, comienzan las primeras puestas de huevos. Las mismas están compuestas por aproximadamente 150 a 200 huevos cada una. Los huevos se colectan en una placa de Petri de 100 mm y se mantienen a 22°C en medio de embriones (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄, 0.00001% P/V de azul de metileno) hasta el momento de la inyección.

Los peces transgénicos se generaron microinyectando 1 nL de la solución de inyección (125 ng del transgén, 175 ng/μL del ARNm de la transposasa Tol2, 0.5M KCl, 1% Rojo fenol) en el vitelo de los embriones en estadio de 1-2 células (Figura M1). Se utilizaron pipetas preparadas a partir de capilares de borosilicato y estiradas con un estirador vertical de micropipetas (Narishige, Modelo PC-10). Con la ayuda de una pipeta Pasteur de vidrio, se colocaron los embriones ordenadamente sobre el canto de un portaobjetos apoyado sobre una placa de Petri. Las inyecciones se generaron con el microinyector Narishige-

IM300 (Narishige) conectado a un tanque de N₂ y utilizando un micromanipulador para manejar la pipeta de inyección bajo una lupa Leica Wild MC3 (Leica).

Luego de la inyección, los embriones se mantuvieron en placas de Petri con medio de embriones en una incubadora a 28°C. Al día 5, se trasladaron a peceras de cría sin flujo de agua. Se los alimentó dos veces al día con una dieta de larvas especialmente formulada para estadios larvales y juveniles. Al día 10, se conectó la pecera de cría al sistema de recirculado de agua bajo un régimen de flujo lento. Al día 13 se comenzó a suplementar la alimentación con artemias frescas (INVE). Pasado el día 15, se conectó el tanque de cría al sistema de recirculación de agua a tiempo completo. A los 21-30 días, se pasaron los animales juveniles a tanques para peces adultos. A los 2 meses, los animales entraron en régimen de alimentación de adultos. Y a los 3 meses, alcanzaron la madurez sexual, de manera que fueron separados para analizar la expresión de la proteína reportera EGFP.

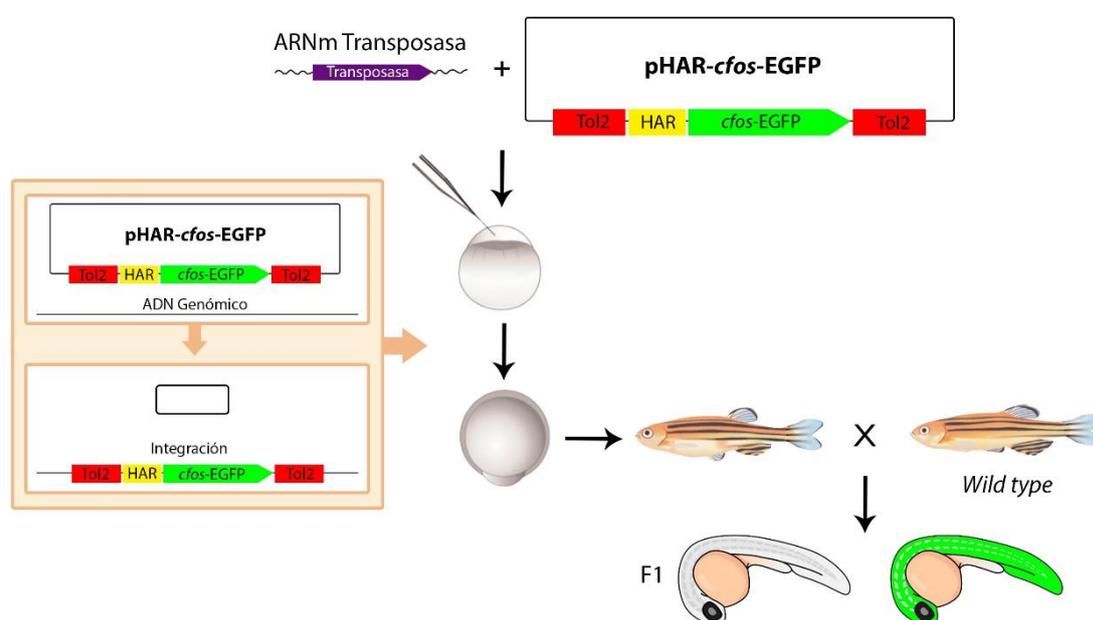


FIGURA M1: Generación de peces cebras transgénicos. Se inyecta el transgén y el ARNm de la transposasa en embriones de peces cebras en estadio de 1-2 células. La transposasa media la integración del transgén en el genoma. Se crían hasta los 3 meses cuando se puede analizar la expresión del reportero en la primera generación mediante cruces con animales *wild-type*.

Análisis del patrón de expresión de la proteína reportera EGFP

Una vez alcanzada la madurez sexual de los embriones inyectados con el transgén de interés, estos fueron cruzados con animales de tipo salvaje o *wild type* (WT) para analizar la expresión de la proteína reportera EGFP en la descendencia a 24, 48 y 72 hpf (horas post fertilización). Los embriones fueron descorionizados y tratados con 1-fenil-2-tiourea 0.1 mM (PTU; Sigma) a 24 hpf para prevenir la pigmentación de los animales en edades posteriores. Luego, se montaron en metilcelulosa 4% P/V (Sigma) y se observó la expresión de la proteína reportera EGFP usando un microscopio OLYMPUS BX51 acoplado a una cámara digital OLYMPUS DP72 y los filtros adecuados. Se consideró *enhancer* a todo transgén que produjera un patrón de expresión consistente en, al menos,

3 líneas independientes. Para los animales que no exhibieron expresión de la proteína reportera, se verificó que la transgénesis haya sido correcta extrayendo el ADN genómico a las 48 hpf con el kit comercial *Wizard® Genomic DNA Purification* (Promega). La presencia del transgén de interés fue detectada por medio de una PCR con *primers* específicos que se detallan en el Apéndice I para cada secuencia con el mismo programa de PCR mencionado anteriormente.

Edición del genoma con CRISPR/Cas9 en ratones

Diseño de las ARN guías

Las guías de ARN (ARNg) fueron diseñadas en el sitio CRISPR del MIT (<http://crispr.mit.edu/>). En el primer paso, se generó un molde mediante PCR con el vector px330 (Addgene) como andamio y *primers* adecuados. El *primer forward* es específico de cada guía y contiene la secuencia del promotor de la ARN Polimerasa T7 río arriba de la secuencia de la guía específica (N₂₀) y una secuencia complementaria al vector px330. El *primer reverse* es el mismo para todas las guías y corresponde a una secuencia complementaria al vector px330. La secuencia de cada guía se detalla en el Apéndice I. La síntesis de las ARNg se realizó con el kit de transcripción *MEGAShortScript T7* (Invitrogen) utilizando como molde el producto de PCR resultante del primer paso.

Transcripción in vitro del ARNm de la Cas9

El ARNm de la Cas9 se preparó mediante una mini preparación de ADN plasmídico del vector pMLM3613 con el *HiSpeed® Plasmid Purification Kit* (Qiagen), se linearizó con la enzima de restricción PmeI (New England Biolabs) y se purificó por precipitación de etanol. La transcripción *in vitro* se realizó con el kit *mMESSAGE mMACHINE™ T7 Kit* (Invitrogen) y finalizada la reacción, el ADN molde se degradó con DNase I (Invitrogen). El ARN resultante se precipitó con cloruro de litio 0.4 M y 2 volúmenes de etanol frío. Finalmente, se resuspendió el ARN en agua libre de nucleasas, se cuantificó y se llevó a una concentración de 175 ng/μL. La solución de microinyección (ARNm Cas9 100 μM, sgRNA-1 50 μM, sgRNA-2 50 μM) fue preparada el día anterior a la inyección de los cigotos de ratón.

Animales y bioterio

Los ratones utilizados en este proyecto pertenecen a la cepa FVB. Los animales se crían en condiciones estándar en el bioterio ubicado en el subsuelo del Complejo Obligado y son mantenidos por técnicos especializados de los institutos INGEBI e IBYME dependientes del CONICET. Los ratones se mantienen en jaulas independientes con un sistema de ventilación por inyección de aire filtrado al interior de cada jaula. El sistema de ventilación también extrae aire de las jaulas y lo elimina fuera de la sala del bioterio. Los animales están bajo un régimen de luz/oscuridad de 12 horas, cuentan con una dieta balanceada comercial y una temperatura de la sala regulada a 20-22°C. Todos los animales fueron tratados respetando los principios de la Guía para el cuidado y uso de animales de

laboratorio publicada por el Instituto Nacional de Salud de los Estados Unidos (NIH) y los protocolos experimentales fueron aprobados por el CICUAL del INGEBI.

Microinyección pronuclear

Los embriones aptos para microinyección se obtienen de la primera generación de la cepa híbrida FVB. Hembras FVB pre púberes de 4 y 5 semanas de edad son súper ovuladas mediante un tratamiento hormonal que consiste en una primera inyección de 5 UI i.p. de *Pregnant Mare Serum* (PMS; Sigma) a las 15 hs, cuatro días antes de la recolección de embriones. Dos días después, a las 13 hs, se realiza una segunda inyección de 5 UI i.p. de *Human Corionic Gonadotrophin* (hGC; Sigma). Inmediatamente después de la última inyección, las hembras súper ovuladas se ponen en jaulas de apareo con machos FVB. A la mañana siguiente, las hembras copuladas por el macho son sacrificadas por dislocación cervical. Se aíslan los oviductos y se colocan en una placa de Petri de 35 mm con medio de cultivo M2 (NaCl 88 mM, KCl 4.7 mM, CaCl₂·H₂O 1.7 mM, KH₂PO₄ 1.2 mM, MgSO₄·7H₂O 1.2 mM, glucosa 5.5 mM, piruvato de sodio 0.33 mM, lactato de sodio 19.8 mM, EDTA 0.1 mM, BSA 0.3%, penicilina 120 µL/mL, sulfato de estreptomicina 50 µg/mL, rojo fenol 0,001%, HEPES 25 mM y pH 7.4; Sigma). Mediante un par de agujas de 30 G y bajo una lupa, se desgarran los oviductos para liberar los ovocitos fecundados. Los mismos se aspiran con una pipeta Pasteur previamente estirada sobre llama. Los ovocitos junto con las células de granulosa que los rodean son colocados en una gota de 90 µL de medio M2 con 1% de hialuronidasa (Sigma). Una vez disgregadas las células de granulosa, los embriones se lavan mediante 5-6 pasajes en medio M2. Finalmente, los embriones listos para la microinyección se mantienen en grupos de 15 a 20 en medio M16 (NaCl 88 mM, KCl 4.7 mM, KH₂PO₄ 1.2 mM, NaHCO₃ 23 mM, glucosa 5.5 mM, MgSO₄·7H₂O 1.2 mM, ácido láctico 4.8 mM, piruvato de sodio 0.2 mM, lactato de sodio 19.8 mM, EDTA 0.1 mM, BSA 0.3%, penicilina 120 µL/mL, sulfato de estreptomicina 50 µg/mL, rojo fenol 0,001%, HEPES 25 mM y pH 7.4; Sigma) en un incubador a 37°C y 5% de CO₂.

Los ratones genéticamente modificados fueron generados por microinyección pronuclear de cigotas FVB utilizando micropipetas sujetadoras y de microinyección previamente preparadas a partir de capilares de borosilicato y estiradas con un estirador vertical de micropipetas modelo 720 (KOPF, David Kopf Instruments) y un micromanipulador (Leica). Los ovocitos fecundados fueron colocados en una gota de medio M2 ubicada en el medio de la depresión de un portaobjetos cóncavo. Se seleccionaron para inyección aquellos ovocitos que tuvieran ambos pronúcleos visibles utilizando un microscopio invertido con óptica de contraste de interferencia diferencial (FS Labovert, Leica). Sosteniendo el ovocito con la pipeta sujetadora, se inyectó 1 pL de solución de inyección estéril (10 mM Tris, pH 7.5, 0.1 mM EDTA, 100 mM NaCl) conteniendo la solución de microinyección (ARNm Cas9 100 µM, sgRNA-1 50 µM, sgRNA-2 50 µM). Luego de la microinyección, los embriones se mantuvieron en estufa a 37°C y 5% de CO₂ hasta el momento de la transferencia embrionaria.

La noche previa a la microinyección de las cigotas se armaron 12 apareos entre dos hembras jóvenes de entre 3 y 6 meses de edad y un macho vasectomizado. A la mañana

siguiente, se separaron las hembras pseudo preñadas, aquellas copuladas por los machos estériles. Dado que las hembras utilizadas para pseudo preñar estaban ciclando de forma natural con un ciclo estral que dura 4/5 días, se calcula que una hembra de cada 4/5 estaba en estro y, por lo tanto, en condiciones de ser copulada. Finalizada la microinyección, las hembras pseudo preñadas son anestesiadas con avertina (300 mg/kg i.p.). Por medio de una incisión lateral, se expone el oviducto. Luego de desgarrar la membrana que recubre el ovario y el oviducto, se transfieren 20 a 30 embriones microinyectados introduciendo en el ámpula de la hembra una pipeta Pasteur de vidrio adecuadamente estirada a llama cargada con los embriones en medio M2. Finalizada la maniobra se cierra la incisión con sutura.

La producción de embriones, microinyecciones y transferencias a hembras pseudo preñadas fueron realizadas por las técnicas Marta Treimun y Carolina Álvarez en el laboratorio del Dr. Marcelo Rubinstein (INGEBI- CONICET).

Generación de la colonia de ratones modificados genéticamente

Los ratones nacidos de las hembras transferidas fueron apartados de las madres sustitutas a los 20-21 días y separados por sexo. Se marcaron las orejas con un sacabocado para realizar una marcación numérica de los animales para su posterior identificación en el bioterio y del tejido de oreja se obtuvo el ADN genómico con el kit comercial *Wizard® Genomic DNA Purification* (Promega). La presencia de las deleciones fue detectada por medio de dos PCR independientes con *primers* específicos que se detallan en el Apéndice I. Una PCR sirve para detectar la deleción propiamente dicha (KO) y la segunda PCR sirve para distinguir entre animales heterocigotas y homocigotas (WT). La PCR KO fue purificada por precipitación de etanol y secuenciada en MacroGen para observar las deleciones generadas. Los animales inyectados mosaico (F0) se cruzaron con animales WT para seleccionar heterocigotas con la deleción deseada de la generación F1 para generar la colonia de mutantes. De la cruce de dos heterocigotas, se obtuvieron los tres genotipos (homocigota *wild type*, heterocigota y homocigota *knock-out*) para su posterior análisis.

Análisis de expresión de los ratones modificados genéticamente

Obtención de embriones y congelado de muestras

Como ninguno de los ratones mutantes resultó ser letal, generamos apareos entre animales homocigotas *knock-out* y entre animales homocigotas *wild type* (que son de la misma camada) para la recolección de muestras a dos edades. La presencia de un tapón vaginal de semen verifica la copulación y las hembras son separadas del macho hasta el día deseado de sacrificio. A los 14.5 días post coito (dpc), las hembras adultas fueron sacrificadas por dislocación cervical. Por medio de una incisión ventral, se aislaron los úteros y se colectaron los embriones para su posterior disección. De cada uno se recolectaron muestras de corteza cerebral, mesencéfalo y rombencéfalo. Además, a los 8 días post natal (dpn), se sacrificaron crías mediante decapitación con tijeras quirúrgicas para su posterior disección. De cada uno se recolectaron muestras de corteza cerebral,

mesencéfalo y cerebelo. Las muestras se guardaron en *RNA Later* (Invitrogen) a -80°C hasta el momento de uso.

Extracción de ARN y síntesis de ADN complementario

Los tejidos previamente diseccionados se descongelaron en hielo y se realizó una extracción de ARN con Trizol (Sigma). Brevemente, se agrega 1 mL de Trizol por cada 100 mg de tejido, se homogenizan y se agrega 0.2 mL de cloroformo cada mL de Trizol. Se centrifuga y se separa en dos fases, se toma la fase acuosa y se vierte en un nuevo tubo que contiene la misma cantidad de 2-propanol que el volumen final que se toma de la fase acuosa. Luego de centrifugar, se lava el precipitado de ARN con etanol 70% y se deja secar antes de resuspenderlo en agua libre de nucleasas. Todas las centrifugaciones se realizan a 4°C a máxima velocidad. Para comprobar la calidad de ARN se siembra en un gel de agarosa y se mide la concentración por NanoDrop. La síntesis de ADN complementario (ADNc) se realizó a partir de 1 μg de ARN previamente tratado con *DNAse I* (Invitrogen) con el kit *High Capacity cDNA Reverse Transcription* (Applied Biosystems) a 37°C con *primers* aleatorios. Para cada muestra se realizaron 6 réplicas biológicas. Se incluyó un control sin ARN y un control sin retrotranscriptasa (*MultiScribeTM Reverse Transcriptase*). Al finalizar, las muestras se guardaron a -20°C hasta el momento de su uso.

PCR cuantitativa en tiempo real

Luego de la obtención del ADNc se procedió a realizar las RT-qPCR mediante el kit *PowerUpTM SYBRTM Green Master Mix* (Applied Biosystems) según las instrucciones del proveedor en el equipo PeetLab. La eficiencia de los *primers* y la concentración óptima de las muestras se determinó con una curva de validación con una mezcla de todas las muestras. Cada muestra se realizó por triplicado para la detección tanto de *Rbfox1* como de β -actina como gen *housekeeping*. Las secuencias de los primers se detallan en Apéndice I. Además, se agregaron los controles mencionados en el punto anterior. La cuantificación de la expresión relativa se realizó mediante el método $2^{-\Delta\Delta\text{Ct}}$ que, suponiendo una eficiencia óptima, compara la variación de la expresión del gen de interés con la del gen de referencia (ΔCt) y la variación entre las muestras *knock-out* y las muestras *wild-type* ($\Delta\Delta\text{Ct}$) (Livak and Schmittgen, 2001). Luego, se aplicó una prueba t de Student, que compara las medias entre los grupos, para evaluar la significancia de la expresión de *Rbfox1* entre los animales *knock-out* y *wild-type*. Los datos con más de 1.5 veces el valor del rango intercuartílico, es decir, que se encuentran a esa distancia del primer y tercer cuartil, fueron considerados valores atípicos y excluidos del análisis estadístico. Las figuras de los resultados fueron obtenidas con la ayuda del programa *GraphPad Prism 6*.

Secuenciación y análisis de ARN

Se recolectaron muestras de mesencéfalo a 14.5 dpc de ratones *Rbfox1* ^{$\Delta 2x\text{HAR.321}/\Delta 2x\text{HAR.321}$} y *Rbfox1*^{+/+}, y de corteza cerebral a 8 dpn de ratones *Rbfox1* ^{$\Delta 2x\text{HAR.505}/\Delta 2x\text{HAR.505}$} y *Rbfox1*^{+/+}. Para cada uno se obtuvieron 5 réplicas biológicas. Se realizó una extracción

de ARN con Trizol como se explicó anteriormente, pero luego de la separación de fases, se tomó la fase acuosa y se añadió etanol para proporcionar las condiciones de unión apropiadas. Luego, se aplicó a una columna comercial de sílice, donde el ARN total se une a la membrana y el fenol y otros contaminantes se eliminan de manera eficiente. Para las muestras de mesencéfalo, se utilizaron columnas del kit *Direct-zol RNA Miniprep Kit* (Zymo Research) y para las muestras de corteza cerebral se utilizaron columnas del kit *RNeasy® Mini Kit* (Qiagen). El ARN se resuspendió en agua libre de nucleasas y se comprobó la calidad del ARN mediante geles de agarosa y NanoDrop. La secuenciación del ARNm se realizó en Novogene (<https://www.novogene.com/us-en/services/research-services/transcriptome-sequencing/mrna-sequencing/>) donde se utilizan plataformas de secuenciación de última generación (*Illumina NovaSeq*), que emplean una estrategia de secuenciación de 150 pb desde ambos extremos (*paired-end*). Previo a la secuenciación, se realizó un control de calidad de las muestras, donde los números de integridad del ARN (RIN) dieron 9.4 para las muestras del mesencéfalo y 9.8 para las muestras de corteza cerebral.

El control de calidad de las lecturas se realizó con el paquete Rfastp (Carroll, 2020). En el análisis de expresión diferencial, las lecturas se alinearon al genoma de ratón (Mm10) con HISAT2 (Kim et al., 2019) y Rsubread (Liao et al., 2019), se contaron con *featureCounts* (Liao et al., 2014) y *summarizeOverlaps* (Lawrence et al., 2013) y se analizaron con el paquete DESeq2 (Figura M2) (Love et al., 2014). Los p-valor de la prueba de Wald se corrigieron con el método de Benjamini y Hochberg (Benjamini and Hochberg, 1995) y se consideraron significativos los genes con un p-valor ajustado por debajo de 0.05 que hayan sido detectados en al menos dos métodos. En el análisis del uso de exones diferencial, las lecturas se alinearon con Rsubread, se contaron con *summarizeOverlaps* y se analizaron con DEXSeq (Figura M2) (Anders et al., 2012).

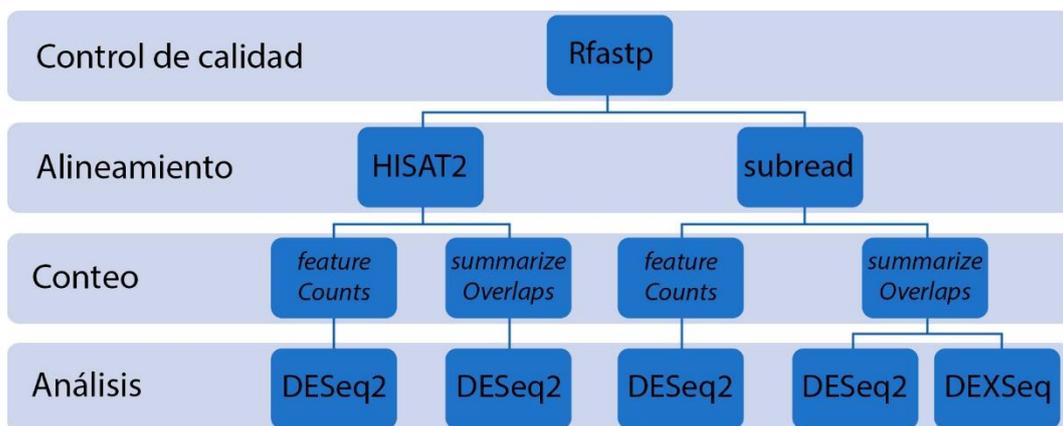


FIGURA M2: Flujo de trabajo del análisis de secuenciación de ARNm. Las lecturas se alinearon al genoma de ratón con HISAT2 o Rsubread, se contaron con *featureCounts* o *summarizeOverlaps* y se analizaron con DESeq2 o DEXSeq. Se consideraron diferencialmente expresados a los genes con un p-valor < 0.05 que hayan sido detectados al menos dos métodos.

Histoquímica

Congelado de muestras para cortes en criostato

Todos los embriones de pez cebra y ratón fueron congelados de la misma manera. Solo los peces se sacrificaron por inmersión en agua helada y ambos se fijaron en una solución de paraformaldehído 4% (PFA) en PBS 1X (Na₂HPO₄ 20 mM, NaCl 150 mM, pH 7.4) por 14-16 horas a 4°C. Luego, se realizaron lavados con PBS 1X y se equilibraron en una solución de sacarosa al 10% por 14-16 horas a 4°C. Al día siguiente, los embriones se embebieron en una solución de sacarosa 10%, gelatina bovina 10% en PBS 1X y se congelaron en isopentano enfriado a -55°C con hielo seco. El material se guardó a -80°C hasta el momento del corte. El tejido se cortó en un criostato (Leica) a 14 µm para el material de pez cebra y a 20 µm para el material de ratón. Los cortes obtenidos se adhirieron a un portaobjetos de vidrio *Superfrost Plus* (Fisher Brand) para su posterior uso.

Doble inmunohistoquímica de cortes histológicos de pez cebra

Los cortes histológicos previamente cortados en criostato se dejaron secar a temperatura ambiente por mínimo 30 minutos y se realizaron lavados con PBS 1X para eliminar los restos de gelatina. A continuación, se realizó el bloqueo de las uniones inespecíficas con una solución de 10% suero normal de burro (Jackson Immuno Research), 0.1% Triton X-100 (Sigma) en PBS 1X. Los anticuerpos primarios se incubaron juntos en una solución conteniendo 0.1% Triton X-100 en PBS 1X durante toda la noche a 4°C. Para la doble inmunohistoquímica se utilizó un anticuerpo policlonal anti-EGFP hecho en pollo (1/500, Aves Lab) y un anticuerpo policlonal anti-Rbfox1 hecho en conejo (1/100, Sigma). Al día siguiente, los cortes se lavaron en PBS 1X y se incubaron secuencialmente con los anticuerpos secundarios en una solución 0.1% Triton X-100 en PBS 1X por 1 hora a temperatura ambiente. Los anticuerpos secundarios fueron anti-IgG de pollo hecho en burro acoplado a Alexa Fluor 488 (1/1000, Jackson Immuno Research) y anti-IgG de conejo hecho en burro acoplado a Alexa Fluor 555 (1/1000, Jackson Immuno Research). Finalmente, los cortes se lavaron en PBS 1X y se montaron en *Vectashield* (Vector Laboratories) o *FluorSaveTM Reagent* (Millipore). Las imágenes fueron adquiridas en un microscopio OLYMPUS BX51 acoplado a una cámara digital OLYMPUS DP72.

Hibridación in situ

Síntesis de las sondas de rbfox1 para pez cebra

El plásmido pME18S-FL3 conteniendo la región codificante de *rbfox1* de *Danio rerio* fue amablemente provisto por el Dr. Bernard Thisse y la Dra. Christine Thisse de la Universidad de Medicina de Virginia, Charlottesville, USA. Como el plásmido pME18S-FL3 no contiene promotores para ninguna ARN Polimerasa, se decidió subclonar un fragmento de 522 pb por PCR con *primers* específicos que se detallan en el Apéndice I en un vector pGEM®-T Easy (Invitrogen) que se secuenció en Macrogen para chequear la direccionalidad del fragmento clonado.

Se preparó una midi preparación de ADN plasmídico del vector pGEM®-T Easy con el kit comercial *HiSpeed Plasmid Purification Kit* (Qiagen) y se linearizó, por separado, con las enzimas de restricción SacII (New England Biolabs) y PstI (New England Biolabs),

para producir la sonda *antisense* y *sense*, respectivamente. Las preparaciones de ADN se purificaron con el kit comercial *DNA Clean & CocentratorTM-5* (Zymo Research). La transcripción *in vitro* de la sonda *antisense* y *sense* se realizó con la ARN polimerasa Sp6 y T7 (Roche) respectivamente. El marcado de la sonda se realizó utilizando el kit *DIG RNA Labeling Mix* (Roche). Finalizada la reacción, el ADN molde se degradó con DNase I. Los productos de síntesis resultantes se analizaron en un gel de agarosa 1% y se cuantificaron en NanoDrop. La sonda se almacenó a -20°C hasta su uso.

Síntesis de las sondas de Rbfox1 para ratón

El plásmido pYX-Asc conteniendo la región codificante de *Rbfox1* de *Mus musculus* fue amablemente provisto por la Dra. Elizabeth Hammock de la Universidad de Medicina de Vanderbilt, Nashville, Tennessee, USA (Hammock and Levitt, 2011). Un fragmento de 505 pb se subclonó por PCR con *primers* específicos que se detallan en el Apéndice I en un vector pGEM®-T Easy (Invitrogen). La síntesis de las sondas *antisense* y *sense* se realizó como se explicó anteriormente para la sonda de *Danio rerio* con las mismas enzimas de restricción y las mismas ARN Polimerasas.

Hibridación in situ en cortes histológicos de pez cebra y ratón

Los cortes histológicos previamente cortados en criostato se dejaron secar a temperatura ambiente por 30 minutos mínimo. Posteriormente el tejido se fijó en una solución de PFA 4% en PBS 1X y se realizaron lavados en PBS 1X. Luego, se realizó un paso de acetilación de una solución de trietanolamina, ácido clorhídrico, anhídrido acético y agua libre de nucleasas con agitación suave. Posteriormente, el tejido se permeabilizó con una solución de Triton X-100 1% en PBS 1X. Paso seguido los cortes se lavaron en PBS 1X y se realizó un paso de pre-hibridización en una solución de hibridización (Formamida 50%, Dextran Sulfato 10%, ARNt de levaduras 1 mg/ml, Solución de Denhart's 1 mg/ml en PBS 1X preparado en agua libre de nucleasas) a temperatura ambiente por 4 horas en cámara húmeda. Luego, la solución de prehibridación se reemplazó por la solución de hibridación conteniendo 0.5 µg de la sonda de ARN marcada con digoxigenina. Los vidrios se cubrieron con un cubreobjeto y se colocaron en una cámara húmeda por 14-16 horas a 70 °C.

Al día siguiente, se realizó un primer lavado en *buffer* SSC 0.2X (NaCl 0.3 M y 30 mM de citrato de sodio, pH 7) a 72 °C y un segundo lavado en el mismo *buffer* a temperatura ambiente. Los cortes se equilibraron en solución B1 (0.1 M Tris, 0.15 M NaCl, Tritón-X100 0.1%, pH 7.5). El bloqueo se realizó con solución B1 con 10% de suero normal de cabra por 4 horas a temperatura ambiente en cámara húmeda. Luego, los vidrios se cubrieron con solución de B1 conteniendo 10% de suero normal de cabra y anticuerpo anti-digoxigenina acoplado a fosfatasa alcalina 1/2000 (Roche) y se incubó por 14-16 horas a temperatura ambiente.

Al día siguiente, se realizaron lavados en solución B1 y los cortes se equilibraron en solución B2 (0.1 M Tris, 0.1 M NaCl, 50 mM MgCl₂, Tween 20 0.1%, pH 9.5). La detección del producto coloreado se realizó reemplazando la solución B2 por la solución

B2 con los sustratos de la fosfatasa alcalina nitroblue tetrazodium (NBT; Roche) y 5-bromo-4-cloro-3-indolil fosfato (BCIP; Roche) (Schaeren-Wiemers and Gerfin-Moser, 1993). La reacción se finalizó sumergiendo los vidrios en PBS 1X. Los cortes se montaron en Mowiol (Polivinil alcohol, Sigma-Aldrich) y se fotografiaron en un microscopio OLYMPUS BX51 equipado con una cámara OLYMPUS DP72.

Hibridación in situ en embriones enteros de pez cebra

Se sacrificaron embriones de 24, 48 y 72 hpf mediante inmersión de agua helada. El tejido se fijó 14-16 horas en una solución de PFA 4% en PBS 1X a 4°C. Luego, los embriones se lavaron en PBT 1X (PBS 1X, 0.1% Tween 20) y se deshidrataron equilibrando los tejidos en soluciones de concentración creciente de metanol en PBT 1X (25%, 50% y 75%). Finalmente, los embriones se pasaron a metanol 100% y se almacenaron hasta el momento de ser usados a -20°C.

Los embriones se colocaron en placas multi-well y se rehidrataron pasándolos por soluciones de concentración decreciente de metanol en PBT 1X (75%, 50%, 25%). Se realizaron lavados con PBT 1X a temperatura ambiente. Los tejidos de 24, 48 y 72 hpf se permeabilizaron en PBT 1X con 10 µg/mL de proteinasa K por 15, 30 y 45 minutos, respectivamente. Se realizó una post fijación en PFA 4% en PBS 1X. Posteriormente, se realizaron lavados en PBT 1X. La prehibridación se realizó por 4 horas a 65°C en solución de hibridación (50% formamida, *buffer* SSC 5X, heparina 50 µg/mL, ARNt de tórula 0,5 mg/mL, Tween-20 0.1%). La hibridación se realizó por 14-16 horas a 65°C en 300 µL de solución de hibridación con 0.5 µg de sonda de ARN marcada con digoxigenina (DIG).

Al día siguiente, se realizaron una serie de lavados a 65°C con concentraciones decrecientes de formamida (50%, 25%) y *buffer* SSC 2X. Luego, se hicieron lavados a 65°C en *buffer* SSC (2X, 0.2X). Finalmente, se hicieron lavados a temperatura ambiente con concentraciones decrecientes de *buffer* SSC 0.2X (75%, 50%, 25%) en PBT 1X. El bloqueo se realizó por 4 horas a temperatura ambiente en solución de bloqueo (2% de suero normal de burro, 2 mg/mL de BSA en PBT 1X). Se reemplazó la solución anterior por solución de bloqueo fresca con 1/10000 de anticuerpo anti-digoxigenina acoplada a fosfatasa alcalina (Roche) y se incubó 14-16 horas a 4°C.

Al día siguiente, se hicieron lavados en PBT 1X y en *buffer* Tris (Tris HCl 100 mM pH 9.5, MgCl₂ 50 mM, NaCl 100 mM y Tween 20 0.1%). La detección del producto coloreado se realizó reemplazando el *buffer* Tris por solución PBT 1X con los sustratos de la fosfatasa alcalina *nitroblue tetrazodium* (NBT; Roche) y 5-bromo-4-cloro-3-indolil fosfato (BCIP; Roche). La reacción se finalizó sumergiendo los embriones en solución *stop* (PBS 1X pH 5.5, EDTA 1 mM y Tween 20 0.1%). Finalmente, se trataron con metanol 100%, se post fijaron con PFA 4% y se montaron en una solución de glicerol 50% (Thisse and Thisse, 2014). Los peces fueron fotografiados en un microscopio OLYMPUS BX51 equipado con una cámara OLYMPUS DP72.

APÉNDICE I - Secuencia de los *primers* específicos

Para el clonado de HARs

Nombre	Forward (5'→3')	Reverse (5'→3')
2xHAR537-Hs	<u>CACCAAAGCCATGACTGCCATTTTC</u>	TTCCCATGAGCTCTAATAAAATCA
ANC727-Hs	<u>CACCTGAGTTGTCACTCCCCAGCA</u>	CACCGTGTGCATCCATCTT
ANC730-Hs	<u>CACCTCTGCCCCTGAAACTTCATT</u>	GCTGTTGCAACTATCCTTCTCA
ANC737-Hs	<u>CACCCCATTTGCCAAACTGGTAGGT</u>	TCCACTTAACGCCTACCTCA
ANC217-Hs	<u>CACCTGTTGCAGCAGCATTTTCTC</u>	AGAGGGAATCAAACCCACACT
2xHAR321-Hs	<u>CACCTCCAAGATGCTCCTATTGGATT</u>	TGGACACTTACGGATAGAAATGG
2xHAR300-Hs	<u>CACCGCCTTCGAACAAAATGGTTA</u>	GAATCTCATGGCTGGTGTAAAA
ANC739-Hs	<u>CACCTTGAAAAGCCAAATAACTGTGG</u>	CCTTTTCTCACCACCCTTCC
HAR135-Hs	<u>CACCTCCTGTTTCTACTGTATGACTCTGA</u>	TCCAGTCCTGTGAATTGAGC
2xHAR128-Hs	<u>CACCACTGTCACATACCACACACACC</u>	AGGACCCATCCACTCAAATC
HACNS177-Hs	<u>CACCTTGCTTTTCACCAAATCCTTG</u>	TCCCAGAGCTATTCTCTCCTTTC
2xHAR505-Hs	<u>CACCATTTTCTCTGTTCTTACACCACCAG</u>	CTCATTTGGCTAGCTCACCTGTATT
2xHAR437-Hs	<u>CACCTCCGTCTACTTCCACGATCC</u>	AGAGAAAGCTCCCCTCCTTG
2xHAR17-Hs	<u>CACCAGGAGATGAAGCTGGGACTT</u>	ATCCCCGAAGGGAAGAGATA
ANC221-Hs	<u>CACCCCGATTCCCCTCCTCTACT</u>	TTCTCTCTTGCCAACCTTTG
ANC750-Hs	<u>CACCACTCTGGCACCCCTGA</u>	TGAGGTTTCCCAGAGTGTT
2xHAR204-Hs	<u>CACCCTGCATCCCAGACTGCTGTA</u>	GTTGGTGGTGTGCCTTCTTT
2xHAR551-Hs	<u>CACCTGACGCTTATTTTCTCCTCCA</u>	AAAGAAACCTGGTAAATATCGAAG
HACNS705-Hs	<u>CACCCGCCAAGCTCATACTACATAGA</u>	ATTGCTAAGCCTATGTCCGAGA
2xHAR537-Pt	<u>CACCAAAGCCATGACTGCCATTTTC</u>	TTCCCATGAGCTCTAATAAAATCA
ANC727-Pt	<u>CACCTGAGTTGTCACTCCCCAACA</u>	CACCGTGTGCATCCATCTT
ANC730-Pt	<u>CACCTCTGCCCCTGAAACTTCATT</u>	GCTGTTGCAACTATCCTTCTCA
ANC737-Pt	<u>CACCCCATTTGCCAAACTGGTAGGT</u>	TCCACTTAACGCCTACCTCA
ANC217-Pt	<u>CACCTGTTGCAGCAGCATTTTCTC</u>	AGAGGGAATCAAACCCACACT
2xHAR321-Pt	<u>CACCTCCAAGATGCTCCTATTGGATT</u>	TGGACACTTACGGATAGAAATGG
2xHAR300-Pt	<u>CACCGCCTTCGAACAAAATGGTTA</u>	GAATCTCATGGCTGGTGTAAAA
ANC739-Pt	<u>CACCTTGAAAAGCCAAATAACTGTGG</u>	CCTTTTCTCACCACCCTTCC
HAR135-Pt	<u>CACCTCCTGTTTCTACTGTATGACTCTGA</u>	TCCAGTCCTGTGAATTGAGC
2xHAR128-Pt	<u>CACCACTGTCACATACCACACACACC</u>	AGGACCCATCCACTCAAATC
HACNS177-Pt	<u>CACCTTGCTTTTCACCAAATCCTTG</u>	TCCCAGAGCTATTCTCTCCTTTC
2xHAR505-Pt	<u>CACCATTTTCTCTGTTCTTACACCACCAG</u>	CTCATTTGGCTAGCTCACCTGTATT
2xHAR437-Pt	<u>CACCTCCGTCTACTTCCACGATCC</u>	AGAGAAAGCTCCCCTCCTTG
2xHAR17-Pt	<u>CACCAGGAGATGAAGCTGGGACTT</u>	ATCCCCGAAGGGAAGAGATA
ANC221-Pt	<u>CACCCCGATTCCCCTCCTCTACT</u>	TTCTCTCTTGCCAACCTTTG
ANC750-Pt	<u>CACCACTCTGGCACCCCTGA</u>	TGAGGTTTCCCAGAGTGTT
2xHAR204-Pt	<u>CACCCTGCATCCCAGACTGCTGTA</u>	GTTGGTGGTGTGCCTTCTTT
2xHAR551-Pt	<u>CACCTGACGCTTATTTTCTCCTCCA</u>	AAAGAAACCTGGTAAATATCGAAG
HACNS705-Pt	<u>CACCCGCCAAGCTCATACTACATAGA</u>	ATTGCTAAGCCTATGTCCGAGA

2xHAR321-Mm	<u>CACCTGGACTTGAAGATGCCTC</u>	CCCCAAACTTAATGCACCTG
2xHAR505-Mm	<u>CACCTTCGGGTCTGTTCCATGA</u>	GCACTCAGCGTGGTAAAAGAGC

Para el clonado de promotores

Nombre	Forward (5' → 3')	Reverse (5' → 3')
E1B	<u>CACCTTTAGGCCAGGAGACCAG</u>	<u>GGATCCGTGTGTGTGTGCGTGTGTGT</u>
E1E	<u>CACCTCATAGCCTGGTTCCTCTG</u>	<u>GGATCCAAGCTGGGAGCCAAATCC</u>
E1A	<u>CACCAAGAGCAAAGACAAGGTG</u>	<u>GGATCCACTTGGGGAAGAGGAA</u>
E1C	<u>CACCAGAACAAGGGAAGGGA</u>	<u>GGATCCTTCATTGTGCGGCTGCA</u>
E1C.1	<u>CACCGAGCGAAGAAGACTCTA</u>	<u>GGATCCTCTCCAAGTGTCCAGA</u>
E1D.1	<u>CACCGGTTCTCGTGTACCTT</u>	<u>GGATCCGAGTGGGGACGCTTG</u>
E1F	<u>CACCTGAATGTTTGCCTCC</u>	<u>GGATCCTGTTCTTGTCCCACAC</u>
E1B-DEL-164	<u>CACCTGGAATCCAGGGCTGCAC</u>	<u>GGATCCGTGTGTGTGTGCGTGTGTGT</u>
E1B-DEL-94	<u>CACCTCGCGCACCAGATTATTT</u>	<u>GGATCCGTGTGTGTGTGCGTGTGTGT</u>

Para la generación de ratones modificados genéticamente

Nombre	ARNg (Secuencia de la T7 ARN Pol – N₂₀ – Secuencia complementaria)
2xHAR321-G1	TTAATACGACTCACTATAG-AATCTTCGTCGTGCTGACAA-GTTTTAGAGCTAGAAATAGC
2xHAR321-G2	TTAATACGACTCACTATAG-CTTAATGCACCTTGTAATTC-GTTTTAGAGCTAGAAATAGC
2xHAR505-G1	TTAATACGACTCACTATAG-ACGTGTCACCGTTCTGAAACA-GTTTTAGAGCTAGAAATAGC
2xHAR505-G2	TTAATACGACTCACTATAG-CAACTTACTTGCACTCAGCG-GTTTTAGAGCTAGAAATAGC
Primer reverse	AAAAGCACCGACTCGGTGCC

Para la genotipificación de ratones modificados genéticamente

Nombre	Forward (5' → 3')	Reverse (5' → 3')
2xHAR321-KO	GAGAAATTGGGGCATTCTGA	GTTGGGATCTGCTTCATGGT
2xHAR321-WT	TGCTTGTGGAGCATATTGGA	GGCAAAGAAACTTTTCTGAGATT
2xHAR505-KO	CTGTGGATTGGGGGTAGTCA	GATCCCTGGGTAAGCCATT
2xHAR505-WT	CTGTGGATTGGGGGTAGTCA	CACAGATGTGCCAACCTGAT

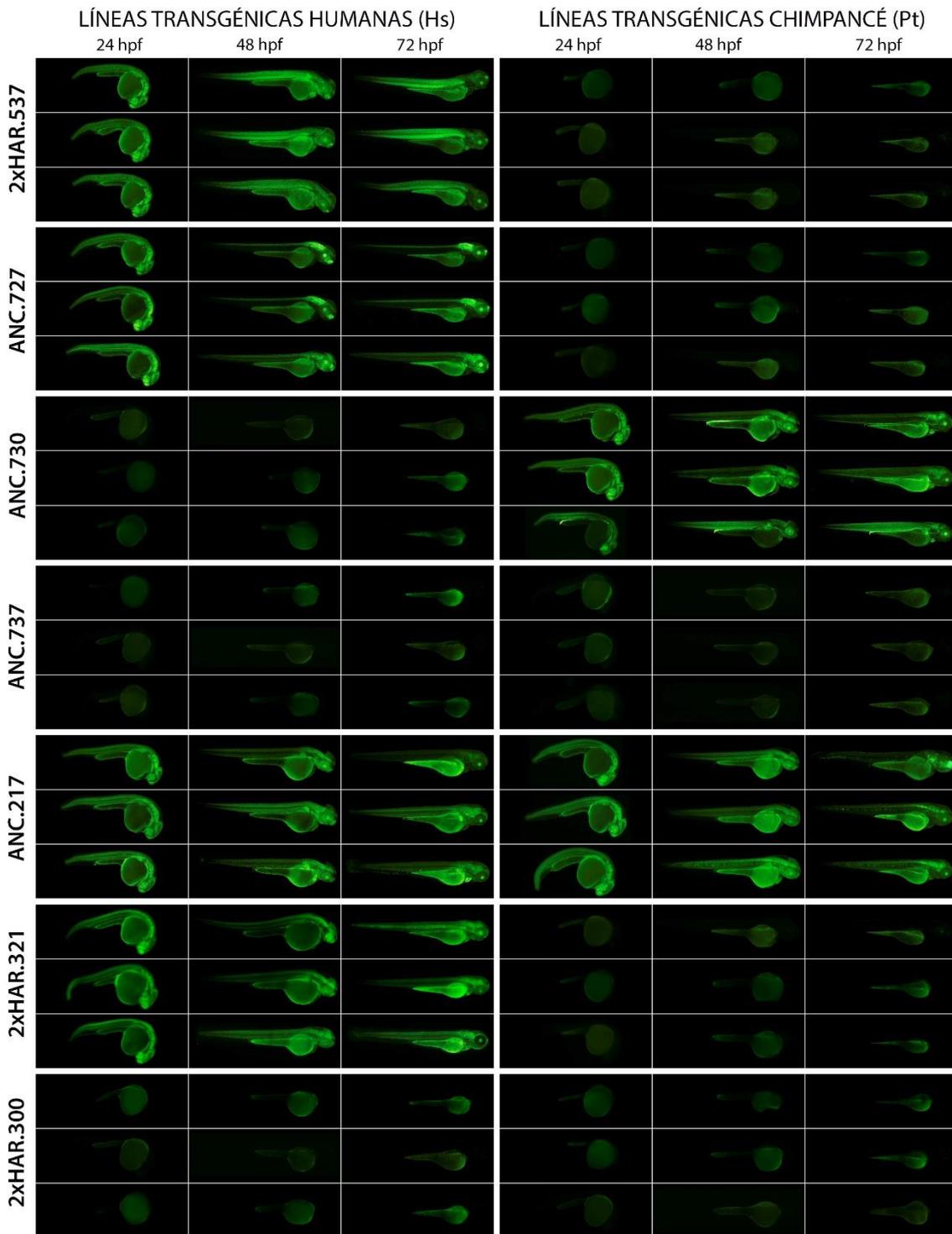
Para las RT-qPCR

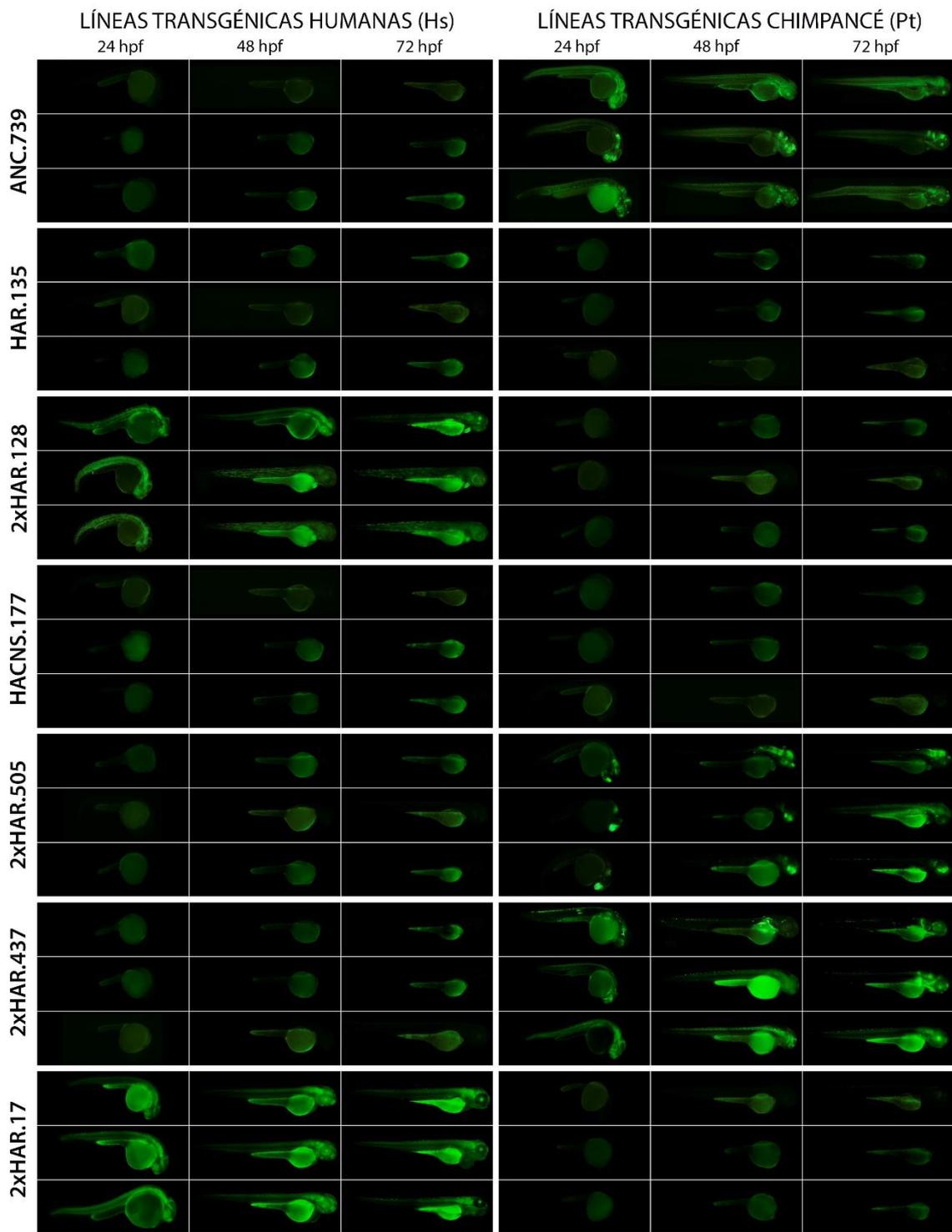
Nombre	Forward (5' → 3')	Reverse (5' → 3')
Rbfox1	GAGGCTCCAAGGGATTTGGTT	CGCGTGCTGTCGCATTATTA
β-actina	AGAGGGAAATCGTGCGTAC	CAATAGTGATGACCTGGCCGT

Para la generación de sondas para hibridización in situ

Nombre	Forward (5' → 3')	Reverse (5' → 3')
Rbfox1-Dr	GACAGACAGCCAACAGCAAA	CTCCATAGGCTGGGATGTGT
Rbfox1-Mm	GGGGGAATTACTCCCTCTTCTT	TGAGGTCTTTGAGCTGTTTC

APÉNDICE II – Líneas de peces cebra transgénicos





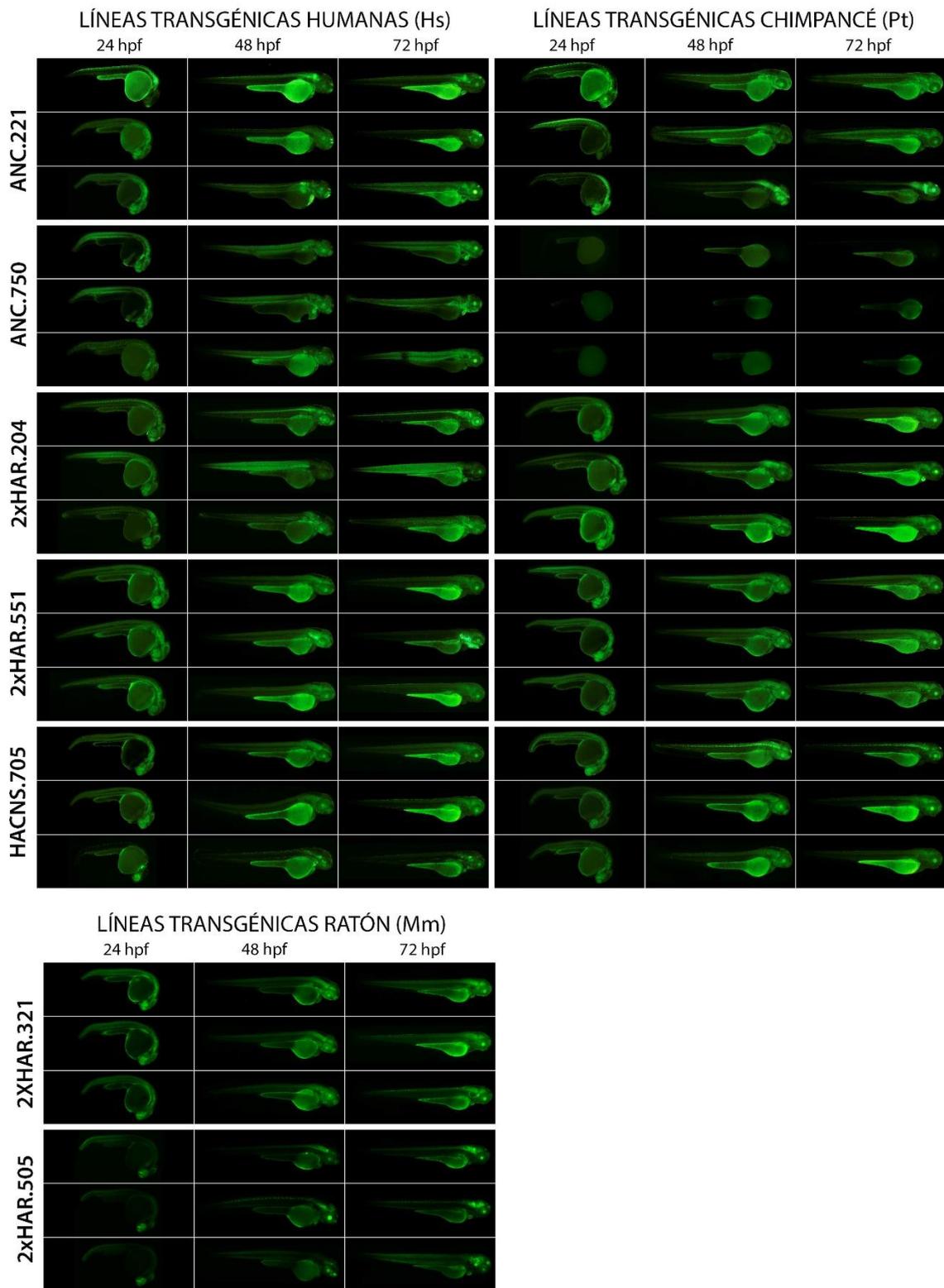


FIGURA S1: Líneas de peces cebra transgénicos de los *RBFOX1*-HARs. Se incluyen las 3 líneas reporteras generadas para cada *RBFOX1*-HAR a 24, 48 y 72 hpf para la secuencia humana (Hs), y sus ortólogos de chimpancé (Pt) y ratón (Mm).

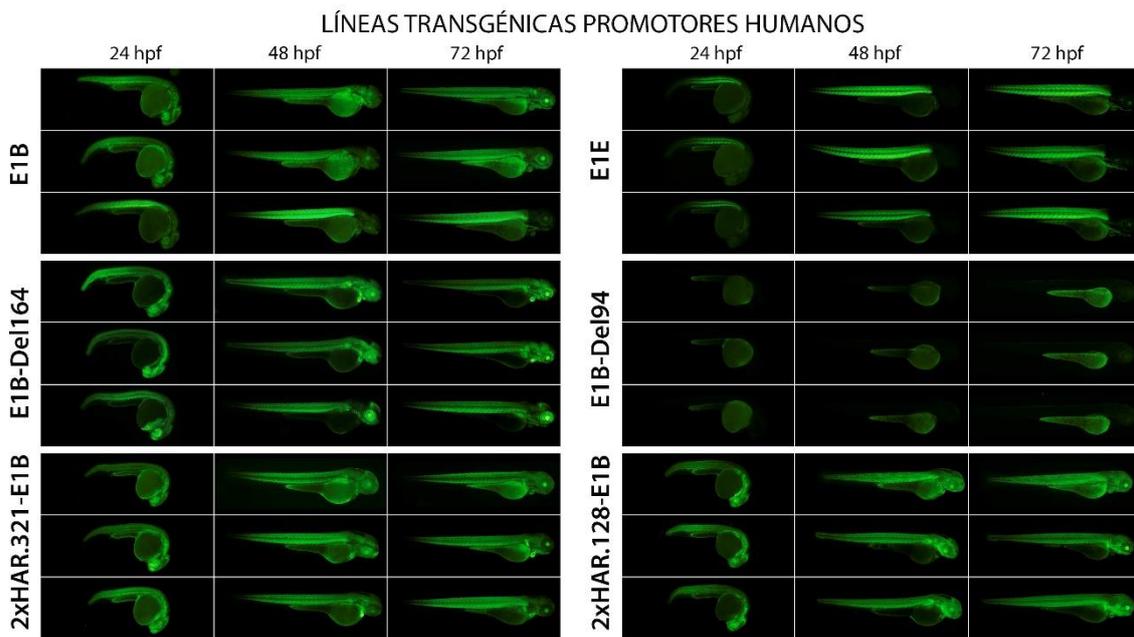


FIGURA S2: Líneas de peces cebra transgénicos de los promotores de *RBFOX1*. Se incluyen las 3 líneas reporteras generadas para cada promotor humano de *RBFOX1* a 24, 48 y 72 hpf.

APÉNDICE III – Genes diferencialmente expresados

TABLA S1: Genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321}

ENTREZ ID	Gene symbol	base Mean	log2 Fold change	lfcSE	stat	pvalue	padj	C L I E P
212398	Frat2	709	-0.65370	0.11141	-5.86765	4.42E-09	1.16E-05	+
68028	Rpl2211	1117	-0.50089	0.08476	-5.90979	3.43E-09	1.16E-05	
381633	Gm1673	3661	-0.44889	0.07558	-5.93912	2.87E-09	1.16E-05	
54217	Rpl36	6301	-0.44274	0.07398	-5.98444	2.17E-09	1.16E-05	
19933	Rpl21	4090	-0.37596	0.06600	-5.69594	1.23E-08	2.57E-05	
68197	Ndufc2	2143	-0.36759	0.06502	-5.65399	1.57E-08	2.74E-05	+
654359	Gm12338	1250	-0.36636	0.06525	-5.61484	1.97E-08	4.81E-05	
12867	Cox7c	2512	-0.36640	0.06686	-5.48035	4.24E-08	6.36E-05	
66492	Zmat2	2570	-0.34205	0.06476	-5.28170	1.28E-07	1.68E-04	+
67025	Rpl11	13602	-0.40130	0.07646	-5.24844	1.53E-07	1.79E-04	
381438	Gm5148	4285	-0.29848	0.05790	-5.15523	2.53E-07	2.50E-04	
67693	Hypk	3125	-0.34959	0.06810	-5.13335	2.85E-07	2.98E-04	
13167	Dbi	1686	-0.47087	0.09278	-5.07484	3.88E-07	3.40E-04	+
13424	Dync1h1	19362	0.41320	0.08143	5.07427	3.89E-07	3.40E-04	
19240	Tmsb10	47899	-0.27414	0.05502	-4.98258	6.27E-07	5.06E-04	+
117600	Srgap1	5507	0.31987	0.06476	4.93931	7.84E-07	5.48E-04	+
226562	Prrc2c	8256	0.33118	0.06736	4.91643	8.81E-07	5.77E-04	+
54405	Ndufa1	777	-0.49847	0.10236	-4.86970	1.12E-06	6.89E-04	+
12069	Bex2	6155	-0.31433	0.06517	-4.82301	1.41E-06	8.14E-04	+
66489	Rpl35	9385	-0.30403	0.06315	-4.81451	1.48E-06	8.14E-04	
102436	Lars2	19847	-0.47352	0.09955	-4.75672	1.97E-06	9.89E-04	+
101966	D8Ert738e	1152	-0.42636	0.08966	-4.75537	1.98E-06	9.89E-04	
20365	Serf1	2382	-0.35086	0.07560	-4.64072	3.47E-06	1.65E-03	+
28106	Mydgf	1299	-0.41509	0.09013	-4.60544	4.12E-06	1.70E-03	
54197	Rnf5	3199	-0.32782	0.07121	-4.60385	4.15E-06	1.70E-03	+
270106	Rpl13	22205	-0.31023	0.06740	-4.60258	4.17E-06	1.70E-03	
20020	Polr2a	4404	0.35946	0.07814	4.60039	4.22E-06	1.70E-03	
24075	Taf10	586	-0.41896	0.09010	-4.65023	3.32E-06	1.71E-03	
76900	Ssbp4	3133	-0.25084	0.05471	-4.58515	4.54E-06	1.76E-03	
16792	Laptm5	205	0.66497	0.14767	4.50294	6.70E-06	2.06E-03	
329540	Nol4l	6422	0.25295	0.05595	4.52082	6.16E-06	2.31E-03	+
66475	Rps23	8872	-0.28771	0.06413	-4.48652	7.24E-06	2.59E-03	
78286	Nav2	3005	0.41134	0.09179	4.48115	7.42E-06	2.59E-03	
66167	Tma7	3156	-0.28544	0.06406	-4.45624	8.34E-06	2.77E-03	+
330286	D630045J12Rik	5409	0.29841	0.06700	4.45361	8.44E-06	2.77E-03	+
15078	H3f3a	43476	-0.25646	0.05782	-4.43554	9.18E-06	2.83E-03	
215690	Nav1	7605	0.36137	0.08140	4.43954	9.02E-06	2.83E-03	+
328572	Ep300	4014	0.28050	0.06344	4.42146	9.80E-06	2.91E-03	
20112	Rps6ka2	1615	0.30257	0.06859	4.41143	1.03E-05	2.91E-03	+
17755	Map1b	66945	0.40087	0.09087	4.41157	1.03E-05	2.91E-03	+

18293	Ogdh	4458	0.33184	0.07571	4.38307	1.17E-05	3.15E-03	
230936	Phf13	1013	0.33895	0.07727	4.38645	1.15E-05	3.15E-03	+
75605	Kdm5b	8656	0.23937	0.05477	4.37064	1.24E-05	3.25E-03	
226412	R3hdm1	4660	0.24519	0.05640	4.34743	1.38E-05	3.40E-03	+
192159	Prpf8	10520	0.29790	0.06853	4.34678	1.38E-05	3.40E-03	
14104	Fasn	16996	0.36759	0.08460	4.34479	1.39E-05	3.40E-03	+
12859	Cox5b	2176	-0.32030	0.07382	-4.33916	1.43E-05	3.41E-03	+
12070	Bex3	5840	-0.34299	0.07960	-4.30916	1.64E-05	3.68E-03	+
74204	Xpo6	4093	0.26671	0.06190	4.30909	1.64E-05	3.68E-03	
217843	Unc79	3114	0.29433	0.06833	4.30753	1.65E-05	3.68E-03	
269700	Hectd4	10265	0.31571	0.07351	4.29476	1.75E-05	3.74E-03	+
230793	Ahdc1	2966	0.37637	0.08759	4.29687	1.73E-05	3.74E-03	+
56406	Ncoa6	4451	0.25543	0.05968	4.28026	1.87E-05	3.91E-03	+
18938	Ppp1r14b	6881	-0.31610	0.07405	-4.26889	1.96E-05	3.99E-03	+
231128	Fam193a	3202	0.23771	0.05571	4.26700	1.98E-05	3.99E-03	
13681	Eif4a1	3052	-0.27342	0.06310	-4.33308	1.47E-05	4.32E-03	
268449	Rpl23a	14968	-0.25882	0.06104	-4.24020	2.23E-05	4.34E-03	
214162	Kmt2a	6974	0.30476	0.07182	4.24355	2.20E-05	4.34E-03	+
269966	Nup98	1901	0.35994	0.08541	4.21440	2.50E-05	4.77E-03	
114641	Rpl31	10642	-0.25567	0.06075	-4.20838	2.57E-05	4.81E-03	
20068	Rps17	9417	-0.31894	0.07592	-4.20089	2.66E-05	4.89E-03	
14533	Bloc1s1	1065	-0.33530	0.07995	-4.19378	2.74E-05	4.96E-03	
21371	Tbca	2129	-0.28724	0.06859	-4.18772	2.82E-05	5.01E-03	
19943	Rpl28	9677	-0.32792	0.07855	-4.17451	2.99E-05	5.22E-03	
20104	Rps6	2571	-0.34276	0.08215	-4.17261	3.01E-05	5.45E-03	
66694	Uqcrfs1	4722	-0.22462	0.05421	-4.14378	3.42E-05	5.87E-03	+
17904	Myl6	6531	-0.26028	0.06290	-4.13784	3.51E-05	5.93E-03	
67840	Mrpl57	1022	-0.32229	0.07802	-4.13105	3.61E-05	6.01E-03	
320365	Fry	5278	0.36007	0.08801	4.09107	4.29E-05	6.93E-03	+
68926	Ubap2	4175	0.31400	0.07683	4.08677	4.37E-05	6.95E-03	
65019	Rpl23	6266	-0.29613	0.07260	-4.07903	4.52E-05	7.08E-03	
67945	Rpl41	16699	-0.32248	0.07932	-4.06579	4.79E-05	7.38E-03	
225929	Pat1	1907	0.26690	0.06576	4.05885	4.93E-05	7.49E-03	
16480	Jup	3241	0.31648	0.07832	4.04080	5.33E-05	7.98E-03	+
69871	Ppp1r35	871	-0.39622	0.09828	-4.03170	5.54E-05	8.06E-03	
69221	2410006H16Rik	755	-0.39221	0.09727	-4.03239	5.52E-05	8.06E-03	
108755	Lymr2	697	-0.40116	0.09985	-4.01755	5.88E-05	8.22E-03	
22272	Uqcrq	2188	-0.36858	0.09170	-4.01925	5.84E-05	8.22E-03	+
110323	Cox6b1	4376	-0.26371	0.06557	-4.02157	5.78E-05	8.22E-03	+
68544	Trir	3627	-0.24391	0.06084	-4.00886	6.10E-05	8.26E-03	+
170749	Mtmr4	5494	0.23654	0.05903	4.00705	6.15E-05	8.26E-03	+
12398	Cbfa2t3	3549	0.37356	0.09320	4.00820	6.12E-05	8.26E-03	
27370	Rps26	3172	-0.27008	0.06700	-4.03132	5.55E-05	8.34E-03	
23964	Tenm2	5980	0.32989	0.08256	3.99570	6.45E-05	8.56E-03	
59022	Edf1	2781	-0.32649	0.08185	-3.98882	6.64E-05	8.70E-03	
69094	Tmem160	565	-0.40113	0.10075	-3.98149	6.85E-05	8.82E-03	
19177	Psmb7	5435	-0.25500	0.06408	-3.97961	6.90E-05	8.82E-03	
103742	Mien1	1214	-0.36249	0.09134	-3.96840	7.24E-05	8.86E-03	

51902	Rnf24	4250	0.24236	0.06112	3.96520	7.33E-05	8.86E-03	+
76884	Cyfp2	9425	0.28226	0.07120	3.96457	7.35E-05	8.86E-03	+
245643	Frmpr3	1506	0.35120	0.08856	3.96556	7.32E-05	8.86E-03	
17991	Ndufa2	1131	-0.47477	0.11996	-3.95786	7.56E-05	8.91E-03	
216825	Usp22	22079	0.21372	0.05398	3.95949	7.51E-05	8.91E-03	+
66962	Swsap1	542	-0.39272	0.09941	-3.95039	7.80E-05	9.09E-03	
12868	Cox8a	4207	-0.25349	0.06424	-3.94634	7.94E-05	9.14E-03	+
240514	Ccdc85b	1415	-0.36247	0.09208	-3.93660	8.26E-05	9.32E-03	
11739	Slc25a4	17087	-0.22554	0.05728	-3.93756	8.23E-05	9.32E-03	
66915	Cops9	1201	-0.37544	0.09568	-3.92411	8.70E-05	9.71E-03	
56370	Tagln3	6310	-0.28364	0.07239	-3.91838	8.91E-05	9.73E-03	+
18245	Oaz1	1174	-0.33099	0.08346	-3.96564	7.32E-05	9.98E-03	
19951	Rpl32	10456	-0.25704	0.06584	-3.90416	9.46E-05	1.02E-02	+
170930	Sumo2	12548	-0.25918	0.06648	-3.89830	9.69E-05	1.03E-02	
15384	Hnrnpab	23964	-0.20241	0.05192	-3.89883	9.67E-05	1.03E-02	
54609	Ubqln2	11371	0.22829	0.05863	3.89404	9.86E-05	1.03E-02	+
50773	Nt5c	969	-0.30728	0.08000	-3.84093	1.23E-04	1.24E-02	
16765	Stmn1	68977	-0.20154	0.05249	-3.84001	1.23E-04	1.24E-02	+
22084	Tsc2	4324	0.27546	0.07173	3.84043	1.23E-04	1.24E-02	
260315	Nav3	3807	0.29221	0.07610	3.83986	1.23E-04	1.24E-02	
75560	Ep400	5228	0.25842	0.06743	3.83235	1.27E-04	1.27E-02	
100683	Trrap	6940	0.26034	0.06801	3.82792	1.29E-04	1.28E-02	+
386612	Thoc6	959	-0.34621	0.09062	-3.82065	1.33E-04	1.28E-02	
14004	Chchd2	7628	-0.23044	0.06035	-3.81841	1.34E-04	1.28E-02	+
328365	Zmiz1	15908	0.26808	0.07020	3.81869	1.34E-04	1.28E-02	+
14268	Fn1	1239	0.40867	0.10700	3.81914	1.34E-04	1.28E-02	
16764	Aff3	2888	0.22638	0.05934	3.81518	1.36E-04	1.29E-02	
11958	Atp5k	577	-0.53956	0.14162	-3.80985	1.39E-04	1.29E-02	
63958	Ube4b	6624	0.23568	0.06183	3.81166	1.38E-04	1.29E-02	+
79233	Zfp319	3369	0.27118	0.07137	3.79974	1.45E-04	1.33E-02	
14586	Gfra2	1031	0.33122	0.08743	3.78824	1.52E-04	1.38E-02	+
68011	Snrpg	1774	-0.36334	0.09622	-3.77632	1.59E-04	1.43E-02	
52838	Dnlz	1375	-0.32341	0.08567	-3.77494	1.60E-04	1.43E-02	
93760	Arid1a	9414	0.26611	0.07053	3.77295	1.61E-04	1.43E-02	+
67671	Rpl38	3243	-0.41123	0.10925	-3.76416	1.67E-04	1.47E-02	
56612	Pfdn5	1568	-0.25359	0.06677	-3.79764	1.46E-04	1.54E-02	+
214240	Disp2	6676	0.35292	0.09419	3.74688	1.79E-04	1.56E-02	+
15901	Id1	427	-0.47113	0.12641	-3.72687	1.94E-04	1.61E-02	
66169	Tomm7	505	-0.45289	0.12133	-3.73279	1.89E-04	1.61E-02	
75597	Ndufaf2	472	-0.42661	0.11458	-3.72308	1.97E-04	1.61E-02	
67281	Rpl37	5723	-0.36244	0.09710	-3.73247	1.90E-04	1.61E-02	
407785	Ndufs6	772	-0.36226	0.09727	-3.72432	1.96E-04	1.61E-02	
67062	Slc25a53	664	-0.34787	0.09330	-3.72844	1.93E-04	1.61E-02	
12933	Crmp1	42474	0.22323	0.05977	3.73473	1.88E-04	1.61E-02	+
17967	Ncam1	25262	0.27685	0.07432	3.72494	1.95E-04	1.61E-02	+
231887	Pdap1	4340	-0.24311	0.06541	-3.71690	2.02E-04	1.64E-02	+
20239	Atxn2	3723	0.25356	0.06829	3.71319	2.05E-04	1.65E-02	+
16971	Lrpl	13966	0.31053	0.08371	3.70974	2.07E-04	1.66E-02	+

72313	Fryl	3421	0.26207	0.07072	3.70569	2.11E-04	1.67E-02	
77264	Zfp142	1454	0.31743	0.08580	3.69964	2.16E-04	1.70E-02	+
100040970	Rpl14-ps1	1671	-0.27389	0.07417	-3.69288	2.22E-04	1.73E-02	
15161	Hcfc1	7143	0.23198	0.06290	3.68805	2.26E-04	1.74E-02	+
100604	Lrrc8c	801	0.33738	0.09144	3.68971	2.25E-04	1.74E-02	
110593	Prdm2	3727	0.25284	0.06859	3.68604	2.28E-04	1.74E-02	+
19336	Rab24	1692	-0.25267	0.06887	-3.66903	2.43E-04	1.84E-02	
17754	Map1a	4160	0.29976	0.08169	3.66969	2.43E-04	1.84E-02	+
19699	Reln	6199	0.33668	0.09182	3.66684	2.46E-04	1.84E-02	+
105171	Arrdc3	2842	-0.33570	0.09170	-3.66089	2.51E-04	1.87E-02	+
380773	Slirp	2243	-0.26214	0.07184	-3.64897	2.63E-04	1.94E-02	
67115	Rpl14	11073	-0.24748	0.06783	-3.64832	2.64E-04	1.94E-02	
53902	Rcan3	3853	0.22580	0.06201	3.64113	2.71E-04	1.98E-02	+
319974	Auts2	6664	0.25773	0.07084	3.63816	2.75E-04	1.99E-02	
66481	Rps21	4374	-0.34548	0.09519	-3.62927	2.84E-04	2.04E-02	
15382	Hnrnpa1	31634	-0.20614	0.05691	-3.62213	2.92E-04	2.08E-02	+
22401	Zmat3	1698	0.28224	0.07797	3.61972	2.95E-04	2.09E-02	+
791299	Gm10069	221	-0.46936	0.12865	-3.64846	2.64E-04	2.19E-02	
15194	Htt	3201	0.29393	0.08175	3.59573	3.23E-04	2.26E-02	+
242408	Fam221b	1190	-0.40025	0.11146	-3.59093	3.30E-04	2.29E-02	
52065	Mfhas1	2183	0.29038	0.08091	3.58916	3.32E-04	2.29E-02	+
12539	Cdc37	5588	-0.23455	0.06548	-3.58225	3.41E-04	2.32E-02	
100036537	Gm11149	3461	0.31410	0.08768	3.58230	3.41E-04	2.32E-02	
73247	Mrgbp	1367	-0.25979	0.07256	-3.58043	3.43E-04	2.32E-02	
22630	Ywhaq	22556	-0.17450	0.04876	-3.57852	3.46E-04	2.32E-02	+
22381	Tceal9	3584	-0.26089	0.07320	-3.56394	3.65E-04	2.44E-02	+
378702	Serf2	4950	-0.22553	0.06336	-3.55959	3.71E-04	2.46E-02	+
67210	Gatad1	2359	-0.22319	0.06291	-3.54788	3.88E-04	2.53E-02	+
72895	Setd5	5932	0.18794	0.05293	3.55091	3.84E-04	2.53E-02	
19173	Psmb5	2150	-0.26599	0.07522	-3.53639	4.06E-04	2.62E-02	
64050	Yeats4	4346	-0.23257	0.06584	-3.53229	4.12E-04	2.62E-02	
66480	Rpl15	21591	-0.19855	0.05624	-3.53048	4.15E-04	2.62E-02	
242466	Zfp462	9070	0.23050	0.06528	3.53121	4.14E-04	2.62E-02	+
58198	Sall1	1381	0.36252	0.10254	3.53531	4.07E-04	2.62E-02	
12808	Cobl	1149	0.36340	0.10298	3.52902	4.17E-04	2.62E-02	+
232440	H2aj	1049	-0.33117	0.09407	-3.52048	4.31E-04	2.64E-02	
20088	Rps24	16742	-0.24444	0.06942	-3.52097	4.30E-04	2.64E-02	+
18476	Pafah1b3	6929	-0.21180	0.06012	-3.52316	4.26E-04	2.64E-02	
18844	Plxna1	13970	0.25663	0.07290	3.52032	4.31E-04	2.64E-02	+
232975	Atp1a3	18432	0.28351	0.08059	3.51798	4.35E-04	2.65E-02	+
215194	Kri1	1071	-0.32884	0.09394	-3.50051	4.64E-04	2.76E-02	
75625	Mageh1	2591	-0.23629	0.06743	-3.50397	4.58E-04	2.76E-02	
11792	Apex1	3257	-0.22409	0.06403	-3.49969	4.66E-04	2.76E-02	
97130	C77080	1099	0.28253	0.08074	3.49939	4.66E-04	2.76E-02	
233210	Prr12	4131	0.30461	0.08700	3.50119	4.63E-04	2.76E-02	
94065	Mrpl34	1835	-0.26674	0.07653	-3.48538	4.91E-04	2.89E-02	
22166	Txn1	1344	-0.21723	0.06129	-3.54408	3.94E-04	2.92E-02	+
73158	Larp1	5126	0.22264	0.06414	3.47108	5.18E-04	3.02E-02	

227723	Prrc2b	27887	0.26408	0.07606	3.47196	5.17E-04	3.02E-02	+
56040	Rplp1	14447	-0.24553	0.07090	-3.46322	5.34E-04	3.08E-02	+
269152	Kif26b	4415	0.30455	0.08796	3.46252	5.35E-04	3.08E-02	+
59026	Huwe1	15990	0.20659	0.05970	3.46055	5.39E-04	3.09E-02	+
67186	Rplp2	8485	-0.26839	0.07775	-3.45210	5.56E-04	3.17E-02	
19274	Ptprm	1206	0.36784	0.10706	3.43583	5.91E-04	3.35E-02	
19981	Rpl37a	7069	-0.32696	0.09531	-3.43065	6.02E-04	3.39E-02	
224020	Pi4ka	5734	0.22002	0.06417	3.42853	6.07E-04	3.40E-02	
73748	Gad1l	3414	-0.25020	0.07182	-3.48344	4.95E-04	3.41E-02	+
330914	Arhgap32	4316	0.24573	0.07176	3.42454	6.16E-04	3.43E-02	+
50770	Atp11a	2886	0.20413	0.05971	3.41878	6.29E-04	3.49E-02	+
15354	Hmgb3	11274	-0.17931	0.05248	-3.41645	6.34E-04	3.49E-02	+
56484	Foxo3	1566	0.28367	0.08305	3.41569	6.36E-04	3.49E-02	
19716	Bex1	949	-0.29819	0.08604	-3.46574	5.29E-04	3.49E-02	+
67332	Snrpd3	2905	-0.23102	0.06772	-3.41154	6.46E-04	3.51E-02	+
13138	Dag1	3879	0.23931	0.07018	3.41012	6.49E-04	3.51E-02	
13004	Ncan	10960	0.24483	0.07183	3.40829	6.54E-04	3.51E-02	+
72056	1810055G02Rik	841	0.30317	0.08895	3.40832	6.54E-04	3.51E-02	
81702	Ankrd17	6828	0.19793	0.05818	3.40224	6.68E-04	3.57E-02	
14797	Tle5	5661	-0.19808	0.05753	-3.44285	5.76E-04	3.63E-02	+
108682	Gpt2	1511	0.29584	0.08712	3.39591	6.84E-04	3.64E-02	+
54563	Nup210	1560	0.31655	0.09332	3.39214	6.93E-04	3.67E-02	
69697	Camsap3	3502	0.22996	0.06783	3.39052	6.98E-04	3.67E-02	
66194	Pycl1	1497	-0.29879	0.08823	-3.38659	7.08E-04	3.69E-02	
66290	Atp6v1g1	5012	-0.21948	0.06480	-3.38700	7.07E-04	3.69E-02	+
72607	Usp13	768	0.34318	0.09880	3.47346	5.14E-04	3.69E-02	
213499	Fbxo42	2196	0.20477	0.06063	3.37716	7.32E-04	3.80E-02	
18637	Pfdn2	1667	-0.31582	0.09369	-3.37094	7.49E-04	3.85E-02	
211383	Amer3	2446	0.23425	0.06952	3.36947	7.53E-04	3.85E-02	+
330319	Wipf3	728	0.35037	0.10398	3.36948	7.53E-04	3.85E-02	+
17992	Ndufa4	4621	-0.19726	0.05864	-3.36397	7.68E-04	3.88E-02	+
54343	Atf7ip	6166	0.23090	0.06864	3.36391	7.68E-04	3.88E-02	+
18030	Nfil3	1128	0.32914	0.09786	3.36338	7.70E-04	3.88E-02	+
51812	Mcrs1	4344	-0.18519	0.05516	-3.35708	7.88E-04	3.95E-02	
142682	Zcchc14	7214	0.18146	0.05412	3.35278	8.00E-04	3.97E-02	+
280645	B3gat2	3614	0.20007	0.05969	3.35178	8.03E-04	3.97E-02	+
20910	Stxbp1	12936	0.23832	0.07105	3.35428	7.96E-04	3.97E-02	+
69638	Enho	3619	-0.23630	0.07065	-3.34463	8.24E-04	3.99E-02	+
76846	Rps9	17385	-0.23623	0.07061	-3.34550	8.21E-04	3.99E-02	
20340	Glg1	5954	0.19753	0.05900	3.34794	8.14E-04	3.99E-02	
64930	Tsc1	3888	0.20680	0.06179	3.34706	8.17E-04	3.99E-02	+
110521	Hivep1	3574	0.23721	0.07094	3.34387	8.26E-04	3.99E-02	
226169	Pprc1	634	0.31468	0.09388	3.35191	8.03E-04	4.00E-02	
208043	Setd1b	2936	0.33261	0.09956	3.34087	8.35E-04	4.02E-02	
76522	Lsm8	1051	-0.36137	0.10850	-3.33061	8.67E-04	4.11E-02	
227634	Camsap1	8906	0.20592	0.06181	3.33163	8.63E-04	4.11E-02	
68193	Rpl24	9533	-0.22217	0.06677	-3.32757	8.76E-04	4.13E-02	
19242	Ptn	17267	-0.27978	0.08418	-3.32349	8.89E-04	4.14E-02	+

217449	Trappc12	1545	0.25139	0.07563	3.32415	8.87E-04	4.14E-02	+
66493	Mrpl51	1143	-0.28896	0.08705	-3.31941	9.02E-04	4.15E-02	+
20054	Rps15	11971	-0.24157	0.07273	-3.32171	8.95E-04	4.15E-02	
114143	Atp6v0b	5079	-0.21814	0.06574	-3.31810	9.06E-04	4.15E-02	+
243529	H1f10	3410	-0.26707	0.08058	-3.31434	9.19E-04	4.15E-02	
107338	Gbf1	4360	0.20064	0.06053	3.31473	9.17E-04	4.15E-02	+
17979	Ncoa3	1993	0.25346	0.07645	3.31517	9.16E-04	4.15E-02	
67851	Mtres1	579	-0.36364	0.10989	-3.30917	9.36E-04	4.16E-02	
52530	Nhp2	1123	-0.28706	0.08673	-3.30987	9.33E-04	4.16E-02	
56531	Ylpm1	4342	0.23869	0.07212	3.30968	9.34E-04	4.16E-02	
76484	Kndc1	2743	0.31265	0.09438	3.31257	9.24E-04	4.16E-02	+
75202	Spaca6	472	-0.41899	0.12671	-3.30668	9.44E-04	4.18E-02	
15081	H3f3b	16055	-0.20654	0.06109	-3.38081	7.23E-04	4.18E-02	+
70439	Taf15	1326	-0.23963	0.07093	-3.37838	7.29E-04	4.19E-02	
66994	Cep19	879	-0.30536	0.09257	-3.29861	9.72E-04	4.26E-02	+
11772	Ap2a2	8063	0.18091	0.05493	3.29338	9.90E-04	4.32E-02	+
18221	Nudc	1211	-0.23272	0.06942	-3.35224	8.02E-04	4.33E-02	
77134	Hnrnpa0	14626	-0.22744	0.06778	-3.35583	7.91E-04	4.33E-02	+
17463	Psmc7	2011	-0.19177	0.05719	-3.35331	7.98E-04	4.33E-02	+
104082	Wdr7	4751	0.21972	0.06680	3.28910	1.01E-03	4.37E-02	+
67465	Sf3a1	3982	0.21138	0.06437	3.28378	1.02E-03	4.44E-02	+
17691	Sik1	1519	0.33689	0.10268	3.28090	1.03E-03	4.45E-02	
78294	Rps27a	11238	-0.20910	0.06386	-3.27451	1.06E-03	4.46E-02	+
71169	Nbas	1989	0.26793	0.08183	3.27410	1.06E-03	4.46E-02	
98366	Smap1	4619	0.19820	0.06057	3.27224	1.07E-03	4.46E-02	+
11899	Astn1	8199	0.21997	0.06731	3.26801	1.08E-03	4.46E-02	+
233863	Gtf3c1	4657	0.23622	0.07228	3.26835	1.08E-03	4.46E-02	+
111173	Erc1	7382	0.20362	0.06246	3.26008	1.11E-03	4.52E-02	
66414	Ndufa12	1231	-0.30796	0.09453	-3.25783	1.12E-03	4.52E-02	
67525	Trim62	1974	0.28751	0.08827	3.25705	1.13E-03	4.52E-02	+
231876	Lmtk2	2912	0.24299	0.07469	3.25319	1.14E-03	4.55E-02	+
18111	Nnat	67484	-0.15303	0.04595	-3.33023	8.68E-04	4.58E-02	+
66091	Ndufa3	826	-0.35942	0.11080	-3.24384	1.18E-03	4.63E-02	
14958	H1f0	6298	-0.25463	0.07858	-3.24063	1.19E-03	4.63E-02	+
67074	Mon2	3566	0.21438	0.06610	3.24308	1.18E-03	4.63E-02	+
381338	Lonrf2	8805	0.23793	0.07340	3.24144	1.19E-03	4.63E-02	+
66104	Tceal6	770	-0.32530	0.10049	-3.23695	1.21E-03	4.66E-02	
20184	Uimc1	4844	-0.22480	0.06781	-3.31539	9.15E-04	4.72E-02	
100502825	Rpl37	2384	-0.33918	0.10242	-3.31149	9.28E-04	4.72E-02	
76559	Atg2b	2094	0.23600	0.07320	3.22414	1.26E-03	4.76E-02	+
...	
268859	Rbfox1	5324	0.11349	0.07162	1.58441	1.13E-01	4.77E-01	

Referencias: baseMean, promedio del valor de conteo normalizado. log2FoldChange, logaritmo en base 2 del cambio en el nivel de expresión entre grupos. lfcSE, estimación del error estándar. stat, valor de estadística de prueba. pvalue, p-valor de la prueba de Wald. padj, p-valor ajustado por Benjamini y Hochberg. CLIP, solapamiento con datos de iCLIP-Seq. E, solapamiento con genes regulados por *Rbfox1*.

TABLA S2: Términos de GO de los genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.321/Δ2xHAR.321}.

Source	Term name	Term ID	padj	-log ₁₀ padj
Up-regulated genes				
GO:MF	binding	GO:0005488	2.10E-08	7.68
GO:MF	protein binding	GO:0005515	2.67E-08	7.57
GO:MF	transcription coregulator activity	GO:0003712	2.02E-04	3.69
GO:MF	transcription coactivator activity	GO:0003713	2.27E-04	3.64
GO:MF	nuclear receptor coactivator activity	GO:0030374	2.47E-03	2.61
GO:MF	chromatin binding	GO:0003682	3.38E-03	2.47
GO:MF	transcription regulator activity	GO:0140110	1.61E-02	1.79
GO:BP	cell differentiation	GO:0030154	1.51E-09	8.82
GO:BP	cellular developmental process	GO:0048869	2.00E-09	8.70
GO:BP	positive regulation of metabolic process	GO:0009893	5.22E-08	7.28
GO:BP	anatomical structure development	GO:0048856	8.25E-08	7.08
GO:BP	positive regulation of cellular process	GO:0048522	9.52E-08	7.02
GO:BP	developmental process	GO:0032502	1.20E-07	6.92
GO:BP	positive regulation of biological process	GO:0048518	2.54E-07	6.60
GO:BP	regulation of primary metabolic process	GO:0080090	8.76E-07	6.06
GO:BP	neuron development	GO:0048666	2.01E-06	5.70
GO:BP	neurogenesis	GO:0022008	2.71E-06	5.57
GO:BP	generation of neurons	GO:0048699	2.76E-06	5.56
GO:BP	regulation of nitrogen compound metabolic process	GO:0051171	2.82E-06	5.55
GO:BP	nervous system development	GO:0007399	3.71E-06	5.43
GO:BP	multicellular organism development	GO:0007275	4.24E-06	5.37
GO:BP	cell development	GO:0048468	4.42E-06	5.35
GO:BP	neuron projection development	GO:0031175	4.85E-06	5.31
GO:BP	positive regulation of macromolecule metabolic process	GO:0010604	8.63E-06	5.06
GO:BP	positive regulation of nitrogen compound metabolic process	GO:0051173	1.30E-05	4.89
GO:BP	neuron differentiation	GO:0030182	2.29E-05	4.64
GO:BP	regulation of metabolic process	GO:0019222	3.01E-05	4.52
GO:BP	regulation of cellular metabolic process	GO:0031323	3.41E-05	4.47
GO:BP	positive regulation of cellular metabolic process	GO:0031325	3.57E-05	4.45
GO:BP	axonogenesis	GO:0007409	4.65E-05	4.33
GO:BP	system development	GO:0048731	7.39E-05	4.13
GO:BP	neuron projection morphogenesis	GO:0048812	1.22E-04	3.91
GO:BP	regulation of cellular component organization	GO:0051128	1.29E-04	3.89
GO:BP	regulation of macromolecule metabolic process	GO:0060255	1.30E-04	3.89
GO:BP	regulation of cellular process	GO:0050794	1.33E-04	3.88
GO:BP	anatomical structure morphogenesis	GO:0009653	1.33E-04	3.88
GO:BP	plasma membrane bounded cell projection morphogenesis	GO:0120039	1.65E-04	3.78
GO:BP	axon development	GO:0061564	1.79E-04	3.75
GO:BP	cell projection morphogenesis	GO:0048858	1.85E-04	3.73
GO:BP	cellular component organization	GO:0016043	2.42E-04	3.62

GO:BP	cell morphogenesis	GO:0000902	2.61E-04	3.58
GO:BP	plasma membrane bounded cell projection organization	GO:0120036	2.61E-04	3.58
GO:BP	cell morphogenesis involved in neuron differentiation	GO:0048667	2.65E-04	3.58
GO:BP	protein modification process	GO:0036211	3.15E-04	3.50
GO:BP	cell part morphogenesis	GO:0032990	3.26E-04	3.49
GO:BP	histone acetylation	GO:0016573	3.43E-04	3.47
GO:BP	macromolecule modification	GO:0043412	3.91E-04	3.41
GO:BP	cell projection organization	GO:0030030	3.92E-04	3.41
GO:BP	internal peptidyl-lysine acetylation	GO:0018393	4.29E-04	3.37
GO:BP	internal protein amino acid acetylation	GO:0006475	4.29E-04	3.37
GO:BP	histone modification	GO:0016570	4.63E-04	3.33
GO:BP	negative regulation of microtubule depolymerization	GO:0007026	5.58E-04	3.25
GO:BP	developmental growth involved in morphogenesis	GO:0060560	6.52E-04	3.19
GO:BP	cellular component organization or biogenesis	GO:0071840	6.56E-04	3.18
GO:BP	regulation of cellular component biogenesis	GO:0044087	6.94E-04	3.16
GO:BP	peptidyl-lysine acetylation	GO:0018394	7.17E-04	3.14
GO:BP	protein acylation	GO:0043543	9.16E-04	3.04
GO:BP	regulation of microtubule depolymerization	GO:0031114	1.05E-03	2.98
GO:BP	positive regulation of RNA metabolic process	GO:0051254	1.14E-03	2.94
GO:BP	cell morphogenesis involved in differentiation	GO:0000904	1.17E-03	2.93
GO:BP	regulation of biological process	GO:0050789	1.32E-03	2.88
GO:BP	cellular component morphogenesis	GO:0032989	1.70E-03	2.77
GO:BP	protein acetylation	GO:0006473	2.26E-03	2.65
GO:BP	regulation of RNA metabolic process	GO:0051252	2.47E-03	2.61
GO:BP	histone H4 acetylation	GO:0043967	2.56E-03	2.59
GO:BP	negative regulation of microtubule polymerization or depolymerization	GO:0031111	3.41E-03	2.47
GO:BP	positive regulation of nucleic acid-templated transcription	GO:1903508	3.71E-03	2.43
GO:BP	positive regulation of DNA-templated transcription	GO:0045893	3.71E-03	2.43
GO:BP	protein metabolic process	GO:0019538	3.88E-03	2.41
GO:BP	positive regulation of RNA biosynthetic process	GO:1902680	3.97E-03	2.40
GO:BP	biological regulation	GO:0065007	4.17E-03	2.38
GO:BP	peptidyl-lysine modification	GO:0018205	5.02E-03	2.30
GO:BP	DNA-templated transcription	GO:0006351	5.52E-03	2.26
GO:BP	nucleic acid-templated transcription	GO:0097659	5.55E-03	2.26
GO:BP	regulation of nucleobase-containing compound metabolic process	GO:0019219	6.31E-03	2.20
GO:BP	RNA biosynthetic process	GO:0032774	6.65E-03	2.18
GO:BP	positive regulation of cellular biosynthetic process	GO:0031328	6.90E-03	2.16
GO:BP	microtubule depolymerization	GO:0007019	7.19E-03	2.14
GO:BP	positive regulation of nucleobase-containing compound metabolic process	GO:0045935	8.18E-03	2.09
GO:BP	positive regulation of macromolecule biosynthetic process	GO:0010557	8.18E-03	2.09
GO:BP	regulation of organelle organization	GO:0033043	1.07E-02	1.97
GO:BP	positive regulation of biosynthetic process	GO:0009891	1.08E-02	1.97
GO:BP	regulation of biosynthetic process	GO:0009889	1.37E-02	1.86

GO:BP	developmental growth	GO:0048589	1.75E-02	1.76
GO:BP	negative regulation of cellular process	GO:0048523	1.84E-02	1.74
GO:BP	regulation of protein metabolic process	GO:0051246	1.97E-02	1.71
GO:BP	regulation of macromolecule biosynthetic process	GO:0010556	1.98E-02	1.70
GO:BP	regulation of DNA-templated transcription	GO:0006355	2.08E-02	1.68
GO:BP	regulation of nucleic acid-templated transcription	GO:1903506	2.10E-02	1.68
GO:BP	hemopoiesis	GO:0030097	2.14E-02	1.67
GO:BP	regulation of cellular biosynthetic process	GO:0031326	2.17E-02	1.66
GO:BP	regulation of RNA biosynthetic process	GO:2001141	2.31E-02	1.64
GO:BP	regulation of gene expression	GO:0010468	2.42E-02	1.62
GO:BP	regulation of catabolic process	GO:0009894	3.77E-02	1.42
GO:BP	positive regulation of catabolic process	GO:0009896	3.96E-02	1.40
GO:BP	regulation of cellular catabolic process	GO:0031329	4.21E-02	1.38
GO:BP	regulation of protein-containing complex assembly	GO:0043254	4.56E-02	1.34
GO:BP	chromatin organization	GO:0006325	4.64E-02	1.33
GO:CC	cytoplasm	GO:0005737	1.00E-08	8.00
GO:CC	intracellular anatomical structure	GO:0005622	4.54E-08	7.34
GO:CC	nucleoplasm	GO:0005654	5.00E-07	6.30
GO:CC	intracellular organelle	GO:0043229	1.40E-05	4.86
GO:CC	organelle	GO:0043226	1.55E-05	4.81
GO:CC	axon	GO:0030424	6.25E-04	3.20
GO:CC	cytosol	GO:0005829	1.15E-03	2.94
GO:CC	membrane-bounded organelle	GO:0043227	1.47E-03	2.83
GO:CC	site of polarized growth	GO:0030427	1.98E-03	2.70
GO:CC	intracellular membrane-bounded organelle	GO:0043231	3.76E-03	2.43
GO:CC	cellular anatomical entity	GO:0110165	5.44E-03	2.26
GO:CC	transferase complex	GO:1990234	7.23E-03	2.14
GO:CC	neuron projection	GO:0043005	7.76E-03	2.11
GO:CC	nucleus	GO:0005634	8.03E-03	2.10
GO:CC	protein-containing complex	GO:0032991	9.93E-03	2.00
GO:CC	histone acetyltransferase complex	GO:0000123	1.34E-02	1.87
GO:CC	TSC1-TSC2 complex	GO:0033596	1.54E-02	1.81
GO:CC	intracellular protein-containing complex	GO:0140535	1.70E-02	1.77
GO:CC	postsynapse	GO:0098794	2.12E-02	1.67
GO:CC	acetyltransferase complex	GO:1902493	2.33E-02	1.63
GO:CC	protein acetyltransferase complex	GO:0031248	2.33E-02	1.63
GO:CC	nuclear lumen	GO:0031981	2.53E-02	1.60
GO:CC	Golgi apparatus	GO:0005794	3.29E-02	1.48
GO:CC	supramolecular complex	GO:0099080	3.52E-02	1.45

Down-regulated genes

GO:MF	structural constituent of ribosome	GO:0003735	1.04E-27	26.98
GO:MF	structural molecule activity	GO:0005198	1.09E-20	19.96
GO:MF	ubiquitin ligase inhibitor activity	GO:1990948	5.54E-05	4.26
GO:MF	ubiquitin-protein transferase inhibitor activity	GO:0055105	9.19E-05	4.04
GO:MF	protein-containing complex binding	GO:0044877	1.26E-03	2.90
GO:MF	nucleosomal DNA binding	GO:0031492	2.50E-03	2.60
GO:MF	RNA binding	GO:0003723	2.68E-03	2.57

GO:MF	ubiquitin-protein transferase regulator activity	GO:0055106	3.67E-03	2.43
GO:MF	rRNA binding	GO:0019843	2.12E-02	1.67
GO:MF	ribonucleoprotein complex binding	GO:0043021	3.01E-02	1.52
GO:MF	nucleic acid binding	GO:0003676	4.23E-02	1.37
GO:BP	cytoplasmic translation	GO:0002181	2.84E-34	33.55
GO:BP	translation	GO:0006412	5.87E-22	21.23
GO:BP	peptide biosynthetic process	GO:0043043	1.72E-21	20.76
GO:BP	amide biosynthetic process	GO:0043604	2.10E-19	18.68
GO:BP	peptide metabolic process	GO:0006518	9.46E-19	18.02
GO:BP	organonitrogen compound biosynthetic process	GO:1901566	5.34E-18	17.27
GO:BP	amide metabolic process	GO:0043603	3.93E-16	15.41
GO:BP	cellular macromolecule biosynthetic process	GO:0034645	6.63E-16	15.18
GO:BP	cellular macromolecule metabolic process	GO:0044260	1.65E-15	14.78
GO:BP	cellular nitrogen compound biosynthetic process	GO:0044271	2.15E-15	14.67
GO:BP	cellular biosynthetic process	GO:0044249	1.46E-12	11.83
GO:BP	organic substance biosynthetic process	GO:1901576	3.53E-12	11.45
GO:BP	biosynthetic process	GO:0009058	9.15E-12	11.04
GO:BP	macromolecule biosynthetic process	GO:0009059	4.98E-11	10.30
GO:BP	oxidative phosphorylation	GO:0006119	9.90E-11	10.00
GO:BP	aerobic respiration	GO:0009060	4.23E-10	9.37
GO:BP	cellular metabolic process	GO:0044237	5.65E-09	8.25
GO:BP	cellular respiration	GO:0045333	1.77E-08	7.75
GO:BP	cellular nitrogen compound metabolic process	GO:0034641	3.34E-08	7.48
GO:BP	metabolic process	GO:0008152	2.89E-07	6.54
GO:BP	mitochondrial ATP synthesis coupled electron transport	GO:0042775	5.39E-07	6.27
GO:BP	ATP synthesis coupled electron transport	GO:0042773	6.75E-07	6.17
GO:BP	energy derivation by oxidation of organic compounds	GO:0015980	9.72E-07	6.01
GO:BP	organonitrogen compound metabolic process	GO:1901564	2.13E-06	5.67
GO:BP	electron transport chain	GO:0022900	3.19E-06	5.50
GO:BP	respiratory electron transport chain	GO:0022904	8.76E-06	5.06
GO:BP	generation of precursor metabolites and energy	GO:0006091	1.03E-05	4.99
GO:BP	protein metabolic process	GO:0019538	3.81E-05	4.42
GO:BP	mitochondrial respiratory chain complex I assembly	GO:0032981	4.26E-05	4.37
GO:BP	NADH dehydrogenase complex assembly	GO:0010257	4.26E-05	4.37
GO:BP	proton motive force-driven mitochondrial ATP synthesis	GO:0042776	6.03E-05	4.22
GO:BP	aerobic electron transport chain	GO:0019646	9.32E-05	4.03
GO:BP	ribosome biogenesis	GO:0042254	1.00E-04	4.00
GO:BP	proton motive force-driven ATP synthesis	GO:0015986	1.27E-04	3.90
GO:BP	ribosomal large subunit biogenesis	GO:0042273	2.06E-04	3.69
GO:BP	nitrogen compound metabolic process	GO:0006807	2.82E-04	3.55
GO:BP	negative regulation of ubiquitin protein ligase activity	GO:1904667	8.25E-04	3.08
GO:BP	cellular process	GO:0009987	9.68E-04	3.01
GO:BP	rRNA metabolic process	GO:0016072	1.30E-03	2.89
GO:BP	primary metabolic process	GO:0044238	1.36E-03	2.87
GO:BP	ribosome assembly	GO:0042255	1.39E-03	2.86

GO:BP	mitochondrial respiratory chain complex assembly	GO:0033108	1.49E-03	2.83
GO:BP	ATP biosynthetic process	GO:0006754	2.08E-03	2.68
GO:BP	ribosomal small subunit biogenesis	GO:0042274	2.42E-03	2.62
GO:BP	rRNA processing	GO:0006364	3.23E-03	2.49
GO:BP	purine ribonucleoside triphosphate biosynthetic process	GO:0009206	4.07E-03	2.39
GO:BP	purine nucleoside triphosphate biosynthetic process	GO:0009145	4.31E-03	2.37
GO:BP	ribonucleoside triphosphate biosynthetic process	GO:0009201	5.70E-03	2.24
GO:BP	gene expression	GO:0010467	6.29E-03	2.20
GO:BP	regulation of protein stability	GO:0031647	7.83E-03	2.11
GO:BP	mitochondrion organization	GO:0007005	9.01E-03	2.05
GO:BP	organic substance metabolic process	GO:0071704	9.02E-03	2.04
GO:BP	negative regulation of ubiquitin-protein transferase activity	GO:0051444	9.83E-03	2.01
GO:BP	nucleoside triphosphate biosynthetic process	GO:0009142	1.01E-02	2.00
GO:BP	organelle organization	GO:0006996	1.13E-02	1.95
GO:BP	ribosomal large subunit assembly	GO:0000027	1.64E-02	1.79
GO:BP	regulation of ubiquitin protein ligase activity	GO:1904666	1.92E-02	1.72
GO:BP	regulation of ubiquitin-protein transferase activity	GO:0051438	2.56E-02	1.59
GO:BP	positive regulation of signal transduction by p53 class mediator	GO:1901798	3.39E-02	1.47
GO:BP	negative regulation of cellular catabolic process	GO:0031330	4.04E-02	1.39
GO:CC	cytosolic ribosome	GO:0022626	4.03E-37	36.39
GO:CC	ribosomal subunit	GO:0044391	3.52E-34	33.45
GO:CC	cytosolic large ribosomal subunit	GO:0022625	3.72E-26	25.43
GO:CC	ribosome	GO:0005840	6.86E-26	25.16
GO:CC	large ribosomal subunit	GO:0015934	5.13E-25	24.29
GO:CC	protein-containing complex	GO:0032991	8.80E-19	18.06
GO:CC	intracellular anatomical structure	GO:0005622	4.27E-17	16.37
GO:CC	cytoplasm	GO:0005737	1.07E-16	15.97
GO:CC	intracellular organelle	GO:0043229	1.87E-15	14.73
GO:CC	organelle	GO:0043226	1.40E-14	13.85
GO:CC	respiratory chain complex	GO:0098803	3.49E-13	12.46
GO:CC	mitochondrial respirasome	GO:0005746	3.49E-13	12.46
GO:CC	mitochondrial protein-containing complex	GO:0098798	4.54E-13	12.34
GO:CC	respirasome	GO:0070469	1.10E-12	11.96
GO:CC	inner mitochondrial membrane protein complex	GO:0098800	1.37E-11	10.86
GO:CC	polysomal ribosome	GO:0042788	2.79E-11	10.55
GO:CC	polysome	GO:0005844	3.66E-11	10.44
GO:CC	ribonucleoprotein complex	GO:1990904	4.60E-11	10.34
GO:CC	cytosol	GO:0005829	2.06E-10	9.69
GO:CC	cytosolic small ribosomal subunit	GO:0022627	6.85E-10	9.16
GO:CC	mitochondrion	GO:0005739	2.23E-09	8.65
GO:CC	mitochondrial envelope	GO:0005740	9.44E-08	7.03
GO:CC	intracellular non-membrane-bounded organelle	GO:0043232	1.01E-07	6.99
GO:CC	non-membrane-bounded organelle	GO:0043228	1.02E-07	6.99
GO:CC	mitochondrial inner membrane	GO:0005743	1.14E-07	6.94
GO:CC	small ribosomal subunit	GO:0015935	1.58E-07	6.80
GO:CC	intracellular membrane-bounded organelle	GO:0043231	2.47E-07	6.61

GO:CC	cytochrome complex	GO:0070069	4.10E-07	6.39
GO:CC	organelle inner membrane	GO:0019866	5.31E-07	6.28
GO:CC	mitochondrial respiratory chain complex IV	GO:0005751	8.45E-07	6.07
GO:CC	mitochondrial membrane	GO:0031966	9.38E-07	6.03
GO:CC	respiratory chain complex IV	GO:0045277	1.77E-06	5.75
GO:CC	mitochondrial respiratory chain complex I	GO:0005747	2.10E-06	5.68
GO:CC	respiratory chain complex I	GO:0045271	2.10E-06	5.68
GO:CC	NADH dehydrogenase complex	GO:0030964	2.10E-06	5.68
GO:CC	membrane-bounded organelle	GO:0043227	2.71E-06	5.57
GO:CC	oxidoreductase complex	GO:1990204	4.44E-06	5.35
GO:CC	catalytic complex	GO:1902494	7.44E-06	5.13
GO:CC	envelope	GO:0031975	4.84E-05	4.32
GO:CC	organelle envelope	GO:0031967	4.84E-05	4.32
GO:CC	cellular anatomical entity	GO:0110165	7.25E-05	4.14
GO:CC	nucleosome	GO:0000786	2.83E-03	2.55
GO:CC	protein-DNA complex	GO:0032993	1.31E-02	1.88
GO:CC	DNA packaging complex	GO:0044815	1.40E-02	1.85
GO:CC	membrane protein complex	GO:0098796	1.47E-02	1.83
GO:CC	nucleoplasm	GO:0005654	2.17E-02	1.66
GO:CC	U2-type precatalytic spliceosome	GO:0071005	4.72E-02	1.33
GO:CC	transmembrane transporter complex	GO:1902495	4.87E-02	1.31

Referencias: MF, función molecular. BP, proceso biológico. CC, componente celular. padj, p-valor ajustado por g:SCS (Set Counts and Sizes).

TABLA S3: Genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}

ENTREZ ID	Gene symbol	base Mean	log2 Fold change	lfcSE	stat	pvalue	padj	C L I E P
20473	Six3	94	-2.79177	0.35316	-7.90507	2.68E-15	4.25E-11	+
50781	Dkk3	4497	0.42641	0.05645	7.55433	4.21E-14	3.34E-10	+
18619	Penk	688	-1.04328	0.14718	-7.08837	1.36E-12	7.18E-09	+
109593	Lmo3	5855	-0.37895	0.05443	-6.96286	3.33E-12	1.32E-08	+
18032	Nfix	18476	0.35216	0.05265	6.68825	2.26E-11	7.17E-08	+
240725	Sulf1	472	-0.93607	0.14431	-6.48653	8.78E-11	2.32E-07	+
12561	Cdh4	2200	-0.58636	0.09741	-6.01934	1.75E-09	3.97E-06	+
57266	Cxcl14	2463	0.49914	0.08430	5.92121	3.20E-09	6.34E-06	+
381489	Rxfp1	2409	0.43007	0.07438	5.78173	7.39E-09	1.30E-05	+
209195	Clic6	98	-1.86999	0.34198	-5.46811	4.55E-08	7.22E-05	
26557	Homer2	4241	0.27013	0.04974	5.43103	5.60E-08	8.09E-05	+
383787	Ankrd63	169	-1.20647	0.22346	-5.39896	6.70E-08	8.77E-05	+
20743	Sptbn2	50555	0.22098	0.04102	5.38669	7.18E-08	8.77E-05	
23850	Pappa2	271	-0.85906	0.16347	-5.25527	1.48E-07	1.68E-04	+
114249	Npnt	4082	0.41977	0.08255	5.08496	3.68E-07	3.89E-04	+
22360	Nrsn1	4988	0.28152	0.05572	5.05238	4.36E-07	4.33E-04	+
239318	Plcx3	638	-0.45843	0.09151	-5.00934	5.46E-07	5.10E-04	+
19649	Robo3	337	0.72160	0.14523	4.96864	6.74E-07	5.95E-04	

14394	Gabra1	7239	0.30613	0.06195	4.94176	7.74E-07	6.15E-04	+	+
16911	Lmo4	7460	0.23265	0.04703	4.94687	7.54E-07	6.15E-04	+	
22348	Slc32a1	3672	-0.32742	0.06647	-4.92557	8.41E-07	6.36E-04	+	
329274	Fam163a	676	0.42618	0.08737	4.87765	1.07E-06	7.75E-04		+
56876	Nsmf	25191	0.21670	0.04475	4.84290	1.28E-06	8.83E-04		
14645	Glul	17405	0.20736	0.04302	4.81958	1.44E-06	9.52E-04	+	+
329366	Ccdc187	81	-1.63075	0.34055	-4.78864	1.68E-06	1.07E-03		
12450	Ccng1	2759	0.27566	0.05868	4.69775	2.63E-06	1.55E-03	+	
108699	Chn1	12807	0.23554	0.05010	4.70184	2.58E-06	1.55E-03	+	+
75718	Vwa5b1	141	-1.20715	0.26036	-4.63644	3.54E-06	2.01E-03	+	
330953	Hcn4	1493	-0.33487	0.07248	-4.62053	3.83E-06	2.10E-03	+	
69993	Chn2	2187	0.32153	0.07051	4.56025	5.11E-06	2.70E-03	+	+
380669	Lin28b	149	-0.91919	0.20247	-4.53999	5.63E-06	2.88E-03		
13371	Dio2	2579	0.34590	0.07649	4.52215	6.12E-06	3.04E-03	+	+
56710	Brinp1	7759	0.22824	0.05158	4.42496	9.65E-06	4.15E-03		
78896	Ecrq4	94	-1.06139	0.23870	-4.44654	8.73E-06	4.15E-03	+	
14805	Grik1	1166	-0.36771	0.08290	-4.43542	9.19E-06	4.15E-03		
16524	Kcnj9	3225	0.24658	0.05561	4.43433	9.24E-06	4.15E-03	+	+
66643	Lix1	2655	0.30431	0.06878	4.42450	9.67E-06	4.15E-03		
235631	Prss50	33	-1.60139	0.36329	-4.40797	1.04E-05	4.36E-03		
218194	Phactr1	15049	0.20290	0.04628	4.38399	1.17E-05	4.74E-03	+	+
12903	Crabp1	63	-1.47293	0.33367	-4.41428	1.01E-05	4.77E-03		
18124	Nr4a3	2806	0.30577	0.07018	4.35676	1.32E-05	5.24E-03		
13199	Ddn	19416	0.23600	0.05432	4.34491	1.39E-05	5.27E-03		+
105853	Mal2	3818	0.24673	0.05674	4.34886	1.37E-05	5.27E-03	+	+
13599	Ecel1	315	-0.75215	0.17515	-4.29430	1.75E-05	6.18E-03	+	
108068	Grm2	2435	0.33002	0.07683	4.29532	1.74E-05	6.18E-03	+	+
21941	Tnfrsf8	49	-1.59108	0.37023	-4.29750	1.73E-05	6.18E-03		
225642	Grp	803	0.36830	0.08614	4.27549	1.91E-05	6.58E-03		+
93896	Glp2r	124	0.69942	0.16308	4.28876	1.80E-05	6.92E-03		+
18387	Oprk1	120	-1.12927	0.26582	-4.24830	2.15E-05	7.28E-03		
212517	Cfap44	199	-0.81894	0.19383	-4.22504	2.39E-05	7.44E-03		
16601	Klf9	6163	0.21598	0.05106	4.23030	2.33E-05	7.44E-03	+	
20664	Sox1	760	-0.38886	0.09186	-4.23327	2.30E-05	7.44E-03		
24087	Tll2	44	-1.79396	0.42459	-4.22520	2.39E-05	7.44E-03		
16875	Lhx8	5	-4.92512	1.17352	-4.19686	2.71E-05	8.34E-03		
80892	Zfhx4	271	-0.65356	0.15751	-4.14928	3.34E-05	1.02E-02		+
228911	Tshz2	3220	0.67997	0.16407	4.14449	3.41E-05	1.02E-02	+	
11484	Aspa	60	-1.21649	0.29419	-4.13498	3.55E-05	1.02E-02		+
13383	Dlg1	3549	0.25385	0.06137	4.13624	3.53E-05	1.02E-02	+	
11539	Adora1	9735	0.18518	0.04483	4.13076	3.62E-05	1.03E-02	+	
16519	Kcnj3	3478	0.22381	0.05485	4.08047	4.49E-05	1.25E-02		+
58188	Vstm2b	2222	-0.32827	0.08085	-4.06020	4.90E-05	1.34E-02	+	+
69338	1700003D09Rik	176	-1.08414	0.26813	-4.04329	5.27E-05	1.42E-02		
72821	Scn2b	2782	0.27738	0.06879	4.03195	5.53E-05	1.46E-02	+	+
52906	Ahi1	7149	-0.23336	0.05794	-4.02784	5.63E-05	1.47E-02	+	
112407	Egln3	564	-0.50242	0.12524	-4.01166	6.03E-05	1.53E-02		+
26934	Racgap1	602	-0.38212	0.09529	-4.00983	6.08E-05	1.53E-02		+

192657	Ell2	505	0.45866	0.11458	4.00287	6.26E-05	1.55E-02	+
241568	Lrrc4c	4160	0.21743	0.05455	3.98617	6.71E-05	1.62E-02	
170768	Pfkfb3	1327	-0.36455	0.09139	-3.98908	6.63E-05	1.62E-02	
14696	Gnb4	5299	-0.23044	0.05801	-3.97261	7.11E-05	1.66E-02	
231290	Slc10a4	213	-0.94192	0.23701	-3.97413	7.06E-05	1.66E-02	+
26949	Vat1	2966	-0.23787	0.06005	-3.96100	7.46E-05	1.72E-02	+
270893	Tmem132e	936	-0.41819	0.10584	-3.95105	7.78E-05	1.76E-02	
17294	Mest	553	-0.39289	0.09931	-3.95631	7.61E-05	1.84E-02	+
20345	Selplg	416	0.54097	0.13758	3.93198	8.42E-05	1.88E-02	
16513	Kcnj10	7022	0.19031	0.04849	3.92451	8.69E-05	1.92E-02	+
18208	Ntn1	705	-0.39289	0.10042	-3.91261	9.13E-05	1.98E-02	
20357	Sema5b	1070	-0.29826	0.07627	-3.91052	9.21E-05	1.98E-02	+
100042856	Gvin2	199	-0.76168	0.19600	-3.88609	1.02E-04	2.16E-02	
110304	Gla3	575	-0.49268	0.12706	-3.87754	1.06E-04	2.19E-02	
217082	Hlf	2017	0.25818	0.06661	3.87624	1.06E-04	2.19E-02	+
93707	Pcdhgc4	1025	-0.21601	0.05587	-3.86599	1.11E-04	2.37E-02	+
104382	Barhl2	43	-1.23531	0.32115	-3.84652	1.20E-04	2.44E-02	+
18549	Pcsk2	8385	0.19045	0.04965	3.83562	1.25E-04	2.52E-02	+
20346	Sema3a	1437	-0.30653	0.08014	-3.82505	1.31E-04	2.57E-02	+
23963	Tenm1	4122	-0.31819	0.08321	-3.82406	1.31E-04	2.57E-02	+
245038	Dclk3	675	-0.36159	0.09482	-3.81358	1.37E-04	2.59E-02	+
68203	Diras2	11649	0.19908	0.05216	3.81646	1.35E-04	2.59E-02	+
237553	Trhde	1184	0.32536	0.08531	3.81380	1.37E-04	2.59E-02	+
210573	Tmem151b	9185	0.20278	0.05327	3.80633	1.41E-04	2.63E-02	+
329427	Chn1os3	2759	0.25126	0.06618	3.79681	1.47E-04	2.71E-02	
19714	Rev3l	7488	-0.21847	0.05763	-3.79108	1.50E-04	2.74E-02	+
27220	Cartpt	281	-0.74018	0.19615	-3.77356	1.61E-04	2.90E-02	
68725	1110032F04Rik	515	0.44584	0.11856	3.76050	1.70E-04	3.03E-02	
100042056	9130019P16Rik	656	0.36903	0.09857	3.74372	1.81E-04	3.12E-02	
14275	Folr1	86	-1.18600	0.31691	-3.74240	1.82E-04	3.12E-02	
66425	Pcp4l1	553	-0.53288	0.14239	-3.74236	1.82E-04	3.12E-02	+
270097	Vat1l	4410	-0.23375	0.06248	-3.74152	1.83E-04	3.12E-02	+
11540	Adora2a	404	-1.27591	0.34204	-3.73030	1.91E-04	3.23E-02	+
67971	Tppp3	1595	-0.32486	0.08729	-3.72148	1.98E-04	3.31E-02	+
436090	Gpr62	262	0.51834	0.13945	3.71695	2.02E-04	3.33E-02	
231470	Fras1	1116	-0.38938	0.10496	-3.70976	2.07E-04	3.40E-02	+
60344	Fign	195	-0.65538	0.17683	-3.70625	2.10E-04	3.41E-02	
217124	Ppp1r9b	25397	0.17552	0.04752	3.69366	2.21E-04	3.55E-02	+
12453	Ccni	14982	0.15140	0.04118	3.67637	2.37E-04	3.63E-02	+
227325	Dner	15875	-0.17964	0.04881	-3.68034	2.33E-04	3.63E-02	+
18597	Pdha1	11445	0.15675	0.04254	3.68451	2.29E-04	3.63E-02	+
218772	Rarb	268	-0.60557	0.16480	-3.67455	2.38E-04	3.63E-02	
319776	Tmem72	123	-1.68324	0.45829	-3.67284	2.40E-04	3.63E-02	
11906	Zfhx3	176	-0.78339	0.21298	-3.67832	2.35E-04	3.63E-02	
14199	Fhl1	4743	-0.15831	0.04325	-3.66024	2.52E-04	3.74E-02	+
231605	Galnt9	2396	0.29955	0.08183	3.66056	2.52E-04	3.74E-02	+
76441	Daam2	1944	0.24801	0.06793	3.65110	2.61E-04	3.79E-02	+
170676	Peg10	1127	-0.42899	0.11753	-3.64995	2.62E-04	3.79E-02	+

216049	Zfp365	5760	0.22584	0.06185	3.65124	2.61E-04	3.79E-02	+
12452	Ccng2	1700	0.24982	0.06850	3.64686	2.65E-04	3.80E-02	+
192650	Cabp7	297	0.42655	0.11483	3.71461	2.04E-04	3.83E-02	+
64113	Moap1	3	5.08236	1.36990	3.71003	2.07E-04	3.83E-02	
259172	Mfrp	11	-1.97739	0.53308	-3.70933	2.08E-04	3.83E-02	
109205	Sobp	9322	0.16396	0.04506	3.63846	2.74E-04	3.89E-02	
235320	Zbtb16	1061	0.35532	0.09784	3.63172	2.82E-04	3.92E-02	+
19739	Rgs9	494	-0.57381	0.15863	-3.61736	2.98E-04	4.07E-02	
381373	Sp9	765	-0.34190	0.09480	-3.60647	3.10E-04	4.21E-02	
234624	A330008L17Rik	350	0.46434	0.12914	3.59561	3.24E-04	4.27E-02	
12564	Cdh8	4427	0.23463	0.06532	3.59211	3.28E-04	4.27E-02	+
21923	Tnc	3431	-0.21947	0.06104	-3.59540	3.24E-04	4.27E-02	+
72585	Lypd1	7169	-0.22096	0.06166	-3.58329	3.39E-04	4.38E-02	+
231842	Amz1	349	-0.36238	0.09952	-3.64126	2.71E-04	4.38E-02	+
232174	Cyp26b1	890	-0.73389	0.20587	-3.56487	3.64E-04	4.59E-02	+
18420	Otp	12	-2.57891	0.71231	-3.62046	2.94E-04	4.62E-02	
21802	Tgfa	1634	-0.31208	0.08768	-3.55942	3.72E-04	4.65E-02	+
67608	Narf	1654	0.19191	0.05317	3.60927	3.07E-04	4.73E-02	+
...	
268859	Rbfox1	7380	-0.06398	0.06802	-0.94052	3.47E-01	8.90E-01	

Referencias: baseMean, promedio del valor de conteo normalizado. log2FoldChange, logaritmo en base 2 del cambio en el nivel de expresión entre grupos. lfcSE, estimación del error estándar. stat, valor de estadística de prueba. pvalue, p-valor de la prueba de Wald. padj, p-valor ajustado por Benjamini y Hochberg. CLIP, solapamiento con datos de iCLIP-Seq. E, solapamiento con genes regulados por *Rbfox1*.

TABLA S4: Términos de GO de los genes diferencialmente expresados en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}.

Source	Term name	Term ID	padj	-log10 padj
Up-regulated genes				
GO:MF	inward rectifier potassium channel activity	GO:0005242	8.87E-03	2.05
GO:MF	G-protein activated inward rectifier potassium channel activity	GO:0015467	1.42E-02	1.85
GO:MF	voltage-gated potassium channel activity	GO:0005249	2.90E-02	1.54
GO:BP	nervous system development	GO:0007399	6.45E-06	5.19
GO:BP	system development	GO:0048731	9.49E-05	4.02
GO:BP	multicellular organism development	GO:0007275	7.31E-04	3.14
GO:BP	anatomical structure development	GO:0048856	2.43E-03	2.61
GO:BP	anterograde trans-synaptic signaling	GO:0098916	3.35E-03	2.47
GO:BP	chemical synaptic transmission	GO:0007268	3.35E-03	2.47
GO:BP	cell-cell signaling	GO:0007267	3.57E-03	2.45
GO:BP	trans-synaptic signaling	GO:0099537	3.62E-03	2.44
GO:BP	potassium ion import across plasma membrane	GO:1990573	4.31E-03	2.37
GO:BP	synaptic signaling	GO:0099536	5.23E-03	2.28
GO:BP	developmental process	GO:0032502	5.40E-03	2.27

GO:BP	behavior	GO:0007610	6.70E-03	2.17
GO:BP	regulation of cellular process	GO:0050794	1.35E-02	1.87
GO:BP	anatomical structure morphogenesis	GO:0009653	1.86E-02	1.73
GO:BP	potassium ion transport	GO:0006813	2.20E-02	1.66
GO:BP	central nervous system development	GO:0007417	2.45E-02	1.61
GO:BP	regulation of biological process	GO:0050789	4.84E-02	1.32
GO:CC	synapse	GO:0045202	1.91E-07	6.72
GO:CC	cell body	GO:0044297	2.70E-06	5.57
GO:CC	neuronal cell body	GO:0043025	7.57E-06	5.12
GO:CC	somatodendritic compartment	GO:0036477	1.02E-05	4.99
GO:CC	cell junction	GO:0030054	4.04E-05	4.39
GO:CC	synaptic membrane	GO:0097060	1.38E-04	3.86
GO:CC	dendrite	GO:0030425	1.40E-04	3.85
GO:CC	dendritic tree	GO:0097447	1.43E-04	3.84
GO:CC	neuron projection	GO:0043005	1.77E-04	3.75
GO:CC	cell projection	GO:0042995	3.22E-03	2.49
GO:CC	postsynapse	GO:0098794	3.59E-03	2.44
GO:CC	neuron projection membrane	GO:0032589	4.19E-03	2.38
GO:CC	parallel fiber to Purkinje cell synapse	GO:0098688	5.51E-03	2.26
GO:CC	membrane	GO:0016020	6.21E-03	2.21
GO:CC	plasma membrane bounded cell projection	GO:0120025	8.11E-03	2.09
GO:CC	glutamatergic synapse	GO:0098978	1.47E-02	1.83
GO:CC	plasma membrane region	GO:0098590	1.57E-02	1.80
GO:CC	glial cell projection	GO:0097386	1.97E-02	1.71
GO:CC	postsynaptic membrane	GO:0045211	2.90E-02	1.54
GO:CC	dendrite membrane	GO:0032590	3.79E-02	1.42
GO:CC	plasma membrane	GO:0005886	3.97E-02	1.40
GO:CC	presynapse	GO:0098793	4.04E-02	1.39
GO:CC	cyclin-dependent protein kinase holoenzyme complex	GO:0000307	4.25E-02	1.37
GO:CC	neuron to neuron synapse	GO:0098984	4.78E-02	1.32
GO:CC	cell periphery	GO:0071944	4.99E-02	1.30

Down-regulated genes

GO:MF	metal ion binding	GO:0046872	1.05E-02	1.98
GO:MF	cation binding	GO:0043169	1.63E-02	1.79
GO:MF	ion binding	GO:0043167	2.21E-02	1.66
GO:MF	organic acid binding	GO:0043177	2.93E-02	1.53
GO:BP	nervous system development	GO:0007399	3.45E-07	6.46
GO:BP	multicellular organism development	GO:0007275	4.35E-06	5.36
GO:BP	neurogenesis	GO:0022008	6.27E-06	5.20
GO:BP	anatomical structure development	GO:0048856	1.19E-05	4.93
GO:BP	neuropeptide signaling pathway	GO:0007218	2.37E-05	4.63
GO:BP	system development	GO:0048731	2.56E-05	4.59
GO:BP	developmental process	GO:0032502	4.49E-05	4.35
GO:BP	animal organ development	GO:0048513	1.02E-04	3.99
GO:BP	generation of neurons	GO:0048699	8.09E-04	3.09
GO:BP	cell differentiation	GO:0030154	2.07E-03	2.68
GO:BP	neuron differentiation	GO:0030182	2.26E-03	2.65

GO:BP	cellular developmental process	GO:0048869	2.39E-03	2.62
GO:BP	central nervous system development	GO:0007417	5.17E-03	2.29
GO:BP	multicellular organismal process	GO:0032501	5.28E-03	2.28
GO:BP	animal organ morphogenesis	GO:0009887	6.66E-03	2.18
GO:BP	cell development	GO:0048468	9.95E-03	2.00
GO:BP	regulation of axon extension	GO:0030516	1.67E-02	1.78
GO:BP	anatomical structure morphogenesis	GO:0009653	2.26E-02	1.65
GO:BP	neuron development	GO:0048666	2.87E-02	1.54
GO:BP	regulation of extent of cell growth	GO:0061387	2.93E-02	1.53

Referencias: MF, función molecular. BP, proceso biológico. CC, componente celular. padj, p-valor ajustado por g:SCS (Set Counts and Sizes).

APÉNDICE IV – Uso diferencial de exones

TABLA S5: Uso diferencial de exones en los ratones *Rbfox1*^{Δ2xHAR.321/ Δ2xHAR.321}.

ENTREZ ID	Gene symbol	Feature ID	Exon base Mean	padj	log2 Fold change	chr	start	end	C L I P	S
17165	Mapkapk5	17165_43	134	0.001	1.224	5	121545483	121545905	+	
18007	Neol	18007_39	443	0.001	0.552	9	59035966	59036425	+	+
59044	Rnf130	59044_3	423	0.001	0.517	11	50025348	50025647	+	
59126	Nek6	59126_20	1939	0.001	0.395	2	38587481	38594606		
67088	Cand2	67088_15	1313	0.001	0.388	6	115803637	115805557	+	
381438	Gm5148	381438_3	8278	0.001	0.103	3	37714326	37715135		+
24075	Taf10	24075_1	3468	0.012	-0.195	7	105739393	105742895		
19045	Ppp1ca	19045_23	1572	0.013	0.325	19	4195105	4195419		
54151	Cyhr1	54151_12	156	0.013	1.047	15	76656717	76656920	+	
69256	Zfp397	69256_1	115	0.013	0.724	18	23954705	23954858		
230908	Tardbp	230908_34	954	0.013	0.408	4	148618076	148618367		
232976	Zfp574	232976_1	103	0.013	1.004	7	25072567	25072645		
12988	Csk	12988_27	146	0.015	0.814	9	57645565	57645617		
14660	Gls	14660_52	313	0.015	0.751	1	52232812	52233223	+	
76798	Srrm4os	76798_5	175	0.015	-0.363	5	116462395	116465487		
11777	Ap3s1	11777_3	105	0.018	0.703	18	46741913	46741935	+	
52480	Snhg14	52480_36	361	0.018	-0.721	7	59696212	59696230		
21833	Thra	21833_5	282	0.019	0.478	11	98741811	98741872	+	
23805	Apc2	23805_26	792	0.019	-0.266	10	80308961	80309175		
71769	Bbs10	71769_2	204	0.019	-0.108	10	111299213	111301727		
74450	Pank2	74450_24	1378	0.019	-0.236	2	131296692	131299186	+	
100504178	Dhrs13os	100504178_1	334	0.020	-0.258	11	78030012	78031931		
11777	Ap3s1	11777_4	106	0.021	0.695	18	46741936	46741941	+	
14325	Ftl1	14325_2	1424	0.021	0.539	7	45457945	45457964		
20637	Snrnp70	20637_7	965	0.021	0.339	7	45376714	45377145		+
69256	Zfp397	69256_3	137	0.021	0.571	18	23954862	23955023		
319655	Podxl2	319655_2	181	0.021	0.852	6	88842565	88842578	+	
11771	Ap2a1	11771_20	440	0.022	0.423	7	44903741	44903884		
14950	H13	14950_21	1121	0.022	0.28	2	152704622	152705036	+	+
53379	Hnrnpa2b1	53379_19	1981	0.022	0.335	6	51465179	51465241		
74200	Khdc4	74200_29	142	0.022	0.712	3	88710853	88711084	+	+
99982	Kdm1a	99982_39	898	0.022	0.547	4	136602309	136602687	+	+
100041480	Myef2l	100041480_1	4862	0.022	-0.054	3	10088110	10089618		
18190	Nrxn2	18190_42	684	0.024	0.462	19	6497939	6498089	+	+
11938	Atp2a2	11938_30	505	0.025	0.697	5	122501896	122502174		
12468	Cct7	12468_21	567	0.025	0.437	6	85468283	85468475	+	
13669	Eif3a	13669_22	804	0.025	-0.349	19	60777939	60778147	+	
13669	Eif3a	13669_21	741	0.025	-0.364	19	60776626	60776797	+	
14660	Gls	14660_51	246	0.025	0.496	1	52232592	52232811		
14955	H19	14955_19	341	0.025	0.441	7	142577413	142578095		
19211	Pten	19211_2	1099	0.025	0.337	19	32757595	32758523	+	
21976	Top3b	21976_17	154	0.025	0.643	16	16877832	16878016	+	

28185	Tom70a	28185_1	1035	0.025	0.354	16	57121703	57122194		
66494	Prelid1	66494_1	1334	0.025	0.287	13	55320500	55322018		
71472	Usp19	71472_51	410	0.025	-0.41	9	108498720	108498993	+	+
77531	Anks1b	77531_84	132	0.025	0.709	10	90970093	90970212	+	+
80752	Fam20c	80752_1	248	0.025	0.566	5	138754514	138755082		+
110157	Raf1	110157_40	512	0.025	0.386	6	115676236	115676635	+	
231093	Agbl5	231093_51	315	0.025	0.487	5	30906668	30906965		
100504178	Dhrs13os	100504178_2	262	0.025	0.303	11	78032157	78032665		
19056	Ppp3cb	19056_9	743	0.032	-0.327	14	20509416	20509448	+	
67480	Cwc25	67480_2	630	0.032	-0.237	11	97745615	97747469	+	
16950	Lox13	16950_22	302	0.033	-0.377	6	83050800	83052562		
26374	Cop1	26374_2	492	0.033	0.304	1	159232326	159232647	+	
27886	Ess2	27886_1	715	0.033	-0.332	16	17900709	17902323		
664968	Tmem238	664968_5	83	0.033	0.725	7	4789074	4789656		
19712	Rest	19712_5	597	0.034	0.162	5	77280703	77286432		
70661	Sik3	70661_18	207	0.034	0.538	9	46202071	46202885	+	
72656	Ints8	72656_41	231	0.034	0.486	4	11253981	11254258	+	+
78798	Eml4	78798_4	145	0.034	0.765	17	83350993	83351214	+	
80912	Pum1	80912_23	372	0.034	-0.331	4	130753942	130754178	+	
192651	Zfp286	192651_5	515	0.034	0.294	11	62779605	62780729		
231464	Cnot6l	231464_9	82	0.034	-0.736	5	96086142	96086293	+	
268510	Mgat5b	268510_1	508	0.034	0.338	11	116918863	116919543	+	
108934	Smim13	108934_2	2265	0.035	-0.145	13	41272566	41276577		
68695	Hddc3	68695_29	475	0.036	-0.341	7	80350121	80350547		
414066	BC037032	414066_4	562	0.037	0.152	15	4026019	4027393		
108943	Trmt10a	108943_21	223	0.038	0.393	3	138156705	138159818		
12778	Ackr3	12778_5	549	0.039	0.264	1	90213834	90216751		
14269	Fnbp1	14269_14	137	0.039	0.718	2	31036542	31036663	+	
68205	Urm1	68205_16	128	0.039	-0.666	2	29844065	29845002		+
13429	Dnm1	13429_21	1013	0.041	0.304	2	32312710	32312889	+	+
22017	Tpmt	22017_2	253	0.041	-0.521	13	47023543	47025178	+	
97287	Mtmr14	97287_3	105	0.042	0.701	6	113237897	113238121		+
11991	Hnrnpd	11991_28	500	0.044	0.454	5	99978577	99978624	+	
15505	Hsph1	15505_22	585	0.044	-0.317	5	149629787	149629983		
16848	Lfng	16848_12	232	0.044	0.603	5	140614675	140615545		
21915	Dtymk	21915_6	174	0.044	-0.494	1	93798776	93799001	+	
24075	Taf10	24075_8	435	0.044	0.227	7	105743830	105743984		
26394	Lypla2	26394_1	819	0.044	0.428	4	135968224	135968700		
54196	Pabpn1	54196_5	399	0.044	0.524	14	54894308	54894390		
56471	Stmn4	56471_22	2432	0.044	0.419	14	66361075	66361648	+	+
58202	Nelfb	58202_22	427	0.044	-0.336	2	25209854	25210074		
66480	Rpl15	66480_5	1124	0.044	0.215	14	18269304	18269316		
75826	Senp2	75826_3	141	0.044	0.587	16	22009561	22009652	+	
84113	Ptov1	84113_11	122	0.044	0.603	7	44863789	44863973		
107831	Adgrb1	107831_50	1473	0.044	0.216	15	74586757	74587457	+	+
216792	Iba57	216792_1	239	0.044	0.304	11	59155369	59158659		
230598	Nrd1	230598_36	527	0.044	-0.301	4	109050901	109050989	+	
241944	Zfp267	241944_7	1660	0.044	0.225	3	36164056	36170335		
270097	Vat1l	270097_3	312	0.044	0.352	8	114205634	114205942	+	

319801	Tigar	319801_1	506	0.044	-0.298	6	127085116	127087880	+
329942	Csm2	329942_80	681	0.044	0.392	4	128565020	128567656	
20742	Sptbn1	20742_12	703	0.046	0.319	11	30109871	30109947	+
21923	Tnc	21923_1	737	0.046	0.395	4	63959785	63960543	+
22644	Rnf103	22644_9	1237	0.046	-0.261	6	71508869	71510763	
67792	Rgs8	67792_18	241	0.047	0.393	1	153690898	153691000	+
24075	Taf10	24075_10	703	0.049	0.235	7	105744087	105744342	
52666	Arhgef25	52666_37	166	0.049	0.823	10	127188295	127188772	
110074	Dut	110074_9	119	0.049	-0.759	2	125257749	125258608	

Referencias: Exon baseMean, promedio del valor de conteo normalizado. padj, p-valor ajustado por Benjamini y Hochberg. log2FoldChange, logaritmo en base 2 del cambio en el nivel de expresión entre grupos. chr, cromosoma. start, inicio del exón. end, fin del exón. CLIP, solapamiento con datos de iCLIP-Seq. S, solapamiento con genes regulados por *Rbfox1*.

TABLA S6: Uso diferencial de exones en los ratones *Rbfox1*^{Δ2xHAR.505/Δ2xHAR.505}.

ENTREZ ID	Gene symbol	Feature ID	Exon base Mean	padj	log2 Fold change	chr	start	end	C L I P S
100039596	Tcf24	100039596_1	36	0.013	0.613	1	9960163	9963521	
76872	Ccdc116	76872_6	70	0.014	-1.039	16	17142331	17142720	
27370	Rps26	27370_7	343	0.014	-0.416	10	128626476	128626524	+
12626	Cetn3	12626_16	99	0.018	-0.857	13	81792964	81793668	+
241576	Ldlrad3	241576_1	609	0.018	0.368	2	101950203	101953175	+
79555	BC005537	79555_2	449	0.027	-0.384	13	24802020	24802160	+
18141	Nup50	18141_11	698	0.031	0.295	15	84940008	84942963	
18971	Pold1	18971_43	17	0.031	2.417	7	44542765	44542878	
71870	Cfap45	71870_9	10	0.037	-2.351	1	172529810	172529952	
58801	Pmaip1	58801_6	36	0.041	-0.893	18	66463270	66464080	
27370	Rps26	27370_6	44	0.041	1.852	10	128626241	128626424	+
209011	Sirt7	209011_23	7	0.045	3.655	11	120622212	120622802	
244668	Sipa112	244668_12	86	0.054	-0.815	8	125449881	125449984	+
17756	Map2	17756_1	14	0.060	-4.389	1	66175273	66175312	+
17756	Map2	17756_2	16	0.060	-4.568	1	66175313	66175315	+

Referencias: Exon baseMean, promedio del valor de conteo normalizado. padj, p-valor ajustado por Benjamini y Hochberg. log2FoldChange, logaritmo en base 2 del cambio en el nivel de expresión entre grupos. chr, cromosoma. start, inicio del exón. end, fin del exón. CLIP, solapamiento con datos de iCLIP-Seq. S, solapamiento con genes regulados por *Rbfox1*.

BIBLIOGRAFÍA

- Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., et al. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. *Sci. Data* 4, 170107.
- Akalín, A., Fredman, D., Amer, E., Dong, X., Bryne, J. C., Suzuki, H., et al. (2009). Transcriptional features of genomic regulatory blocks. *Genome Biol.* 10, R38.
- Alvarez-Bolado, G., Zhou, X., Voss, A. K., Thomas, T., and Gruss, P. (2000). Winged helix transcription factor Foxb1 is essential for access of mammillothalamic axons to the thalamus. *Development* 127, 1029–1038.
- Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Res.* 22, 2008–2017.
- Andres-Alonso, M., Ammar, M. R., Butnaru, I., Gomes, G. M., Acuña Sanhueza, G., Raman, R., et al. (2019). SIPA1L2 controls trafficking and local signaling of TrkB-containing amphisomes at presynaptic terminals. *Nat. Commun.* 10, 5448.
- Antón, S. C. (2003). Natural history of Homo erectus. *Am. J. Phys. Anthropol.* Suppl 37, 126–170.
- Antón, S. C., Potts, R., and Aiello, L. C. (2014). Human evolution. Evolution of early Homo: an integrated biological perspective. *Science* 345, 1236828.
- Antonacci, F., Dennis, M. Y., Huddleston, J., Sudmant, P. H., Steinberg, K. M., Rosenfeld, J. A., et al. (2014). Palindromic GOLGA8 core duplicons promote chromosome 15q13.3 microdeletion and evolutionary instability. *Nat. Genet.* 46, 1293–1302.
- Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., et al. (1995). Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1684–1688.
- Ariza-Cosano, A., Visel, A., Pennacchio, L. A., Fraser, H. B., Gómez-Skarmeta, J. L., Irimia, M., et al. (2012). Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression. *BMC Genomics* 13, 713.
- Armengol, G., Knuutila, S., Lozano, J.-J., Madrigal, I., and Caballín, M.-R. (2010). Identification of human specific gene duplications relative to other primates by array CGH and quantitative PCR. *Genomics* 95, 203–209.
- Arnone, M. I., Dmochowski, I. J., and Gache, C. (2004). Using reporter genes to study cis-regulatory elements. *Methods Cell Biol.* 74, 621–652.
- Assali, A., Harrington, A. J., and Cowan, C. W. (2019). Emerging roles for MEF2 in brain development and mental disorders. *Curr. Opin. Neurobiol.* 59, 49–58.
- Atchison, M. (1988). Enhancers: Mechanisms of action and cell specificity. *Annu. Rev. Cell Dev. Biol.* 4, 127–153.
- Auweter, S. D., Fasan, R., Reymond, L., Underwood, J. G., Black, D. L., Pitsch, S., et al. (2006). Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J.* 25, 163–173.

- Azevedo, F. A. C., Carvalho, L. R. B., Grinberg, L. T., Farfel, J. M., Ferretti, R. E. L., Leite, R. E. P., et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J. Comp. Neurol.* 513, 532–541.
- Bailey, J. A., and Eichler, E. E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* 7, 552–564.
- Bark, C., Bellinger, F. P., Kaushal, A., Mathews, J. R., Partridge, L. D., and Wilson, M. C. (2004). Developmentally regulated switch in alternatively spliced SNAP-25 isoforms alters facilitation of synaptic transmission. *J. Neurosci.* 24, 8796–8805.
- Bauernfeind, A. L., Soderblom, E. J., Turner, M. E., Moseley, M. A., Ely, J. J., Hof, P. R., et al. (2015). Evolutionary divergence of gene and protein expression in the brains of humans and chimpanzees. *Genome Biol. Evol.* 7, 2276–2288.
- Begeman, I. J., Emery, B., Kurth, A., and Kang, J. (2022). Regeneration and developmental enhancers are differentially compatible with minimal promoters. *Dev. Biol.* 492, 47–58.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., et al. (2004). Ultraconserved elements in the human genome. *Science* 304, 1321–1325.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300.
- Bhalla, K., Phillips, H. A., Crawford, J., McKenzie, O. L. D., Mulley, J. C., Eyre, H., et al. (2004). The de novo chromosome 16 translocations of two patients with abnormal phenotypes (mental retardation and epilepsy) disrupt the A2BP1 gene. *J. Hum. Genet.* 49, 308–311.
- Bill, B. R., Lowe, J. K., Dybuncio, C. T., and Fogel, B. L. (2013). Orchestration of neurodevelopmental programs by RBFOX1: implications for autism spectrum disorder. *Int. Rev. Neurobiol.* 113, 251–267.
- Bird, C. P., Stranger, B. E., Liu, M., Thomas, D. J., Ingle, C. E., Beazley, C., et al. (2007). Fast-evolving noncoding sequences in the human genome. *Genome Biol.* 8, R118.
- Blue, R. E., Curry, E. G., Engels, N. M., Lee, E. Y., and Giudice, J. (2018). How alternative splicing affects membrane-trafficking dynamics. *J. Cell Sci.* 131, jcs216465.
- Blumenschine, R. J., Peters, C. R., Masao, F. T., Clarke, R. J., Deino, A. L., Hay, R. L., et al. (2003). Late Pliocene Homo and hominid land use from Western Olduvai Gorge, Tanzania. *Science* 299, 1217–1221.
- Boyd, J. L., Skove, S. L., Rouanet, J. P., Pilaz, L.-J., Bepler, T., Gordân, R., et al. (2015). Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* 25, 772–779.
- Brackenbury, W. J., and Isom, L. L. (2011). Na channel β subunits: Overachievers of the ion channel family. *Front. Pharmacol.* 2, 53.
- Britten, R. J. (2002). Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13633–13635.
- Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science* 165, 349–357.
- Britten, R. J., and Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q. Rev. Biol.* 46, 111–138.

- Brunet, M., Guy, F., Pilbeam, D., Lieberman, D. E., Likius, A., Mackaye, H. T., et al. (2005). New material of the earliest hominid from the Upper Miocene of Chad. *Nature* 434, 752–755.
- Brunet, M., Guy, F., Pilbeam, D., Mackaye, H. T., Likius, A., Ahounta, D., et al. (2002). A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* 418, 145–151.
- Burbano, H. A., Green, R. E., Maricic, T., Lalueza-Fox, C., de la Rasilla, M., Rosas, A., et al. (2012). Analysis of human accelerated DNA regions using archaic hominin genomes. *PLoS One* 7, e32877.
- Bush, E. C., and Lahn, B. T. (2008). A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol. Biol.* 8, 17.
- Caporale, A. L., Gonda, C. M., and Franchini, L. F. (2019). Transcriptional enhancers in the FOXP2 locus underwent accelerated evolution in the human lineage. *Mol. Biol. Evol.* 36, 2432–2450.
- Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. R., and Pollard, K. S. (2013). Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20130025.
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., et al. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* 38, 626–635.
- Carroll, S. B. (2003). Genetics and the making of Homo sapiens. *Nature* 422, 849–857.
- Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS Biol.* 3, e245.
- Carroll, W. W. J.-D. L. (2020). *Rfastp*. Bioconductor doi: 10.18129/B9.BIOC.RFASTP.
- Carter, B. D. (2006). A Bex-cycle built for two. *EMBO Rep.* 7, 382–384.
- Casanovas, S., Schlichtholz, L., Mühlbauer, S., Dewi, S., Schüle, M., Strand, D., et al. (2020). Rbfox1 is expressed in the mouse brain in the form of multiple transcript variants and contains functional E boxes in its alternative promoters. *Front. Mol. Neurosci.* 13, 66.
- Castro-Mondragon, J. A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R. B., Turchi, L., Blanc-Mathieu, R., et al. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 50, D165–D173.
- Chang, G. T. G., van den Bemd, G.-J. C. M., Jhamai, M., and Brinkmann, A. O. (2002). Structure and function of GC79/TRPS1, a novel androgen-repressible apoptosis gene. *Apoptosis* 7, 13–21.
- Chen, F. C., and Li, W. H. (2001). Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 68, 444–456.
- Cheung, M., Chaboissier, M.-C., Mynett, A., Hirst, E., Schedl, A., and Briscoe, J. (2005). The transcriptional control of trunk neural crest induction, survival, and delamination. *Dev. Cell* 8, 179–192.
- Chimpanzee Sequencing and Analysis Consortium (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69–87.
- Coenye, T. (2021). Do results obtained with RNA-sequencing require independent verification? *Biofilm* 3, 100043.
- Conboy, J. G. (2017). Developmental regulation of RNA processing by Rbfox proteins. *Wiley Interdiscip. Rev. RNA* 8, e1398.

- Cunningham, T. J., Lancman, J. J., Berenguer, M., Dong, P. D. S., and Duester, G. (2018). Genomic knockout of two presumed forelimb Tbx5 enhancers reveals they are nonessential for limb development. *Cell Rep.* 23, 3146–3151.
- Damianov, A., and Black, D. L. (2010). Autoregulation of Fox protein expression to produce dominant negative splicing factors. *RNA* 16, 405–416.
- Damianov, A., Ying, Y., Lin, C.-H., Lee, J.-A., Tran, D., Vashisht, A. A., et al. (2016). Rbfox proteins regulate splicing as part of a large multiprotein complex LASR. *Cell* 165, 606–619.
- Davis, L. K., Maltman, N., Mosconi, M. W., Macmillan, C., Schmitt, L., Moore, K., et al. (2012). Rare inherited A2BP1 deletion in a proband with autism and developmental hemiparesis. *Am. J. Med. Genet. A* 158A, 1654–1661.
- de Souza, F. S. J., Santangelo, A. M., Bumashny, V., Avale, M. E., Smart, J. L., Low, M. J., et al. (2005). Identification of neuronal enhancers of the proopiomelanocortin gene by transgenic mouse analysis and phylogenetic footprinting. *Mol. Cell. Biol.* 25, 3076–3086.
- de Wit, E., Vos, E. S. M., Holwerda, S. J. B., Valdes-Quezada, C., Verstegen, M. J. A. M., Teunissen, H., et al. (2015). CTCF binding polarity determines chromatin looping. *Mol. Cell* 60, 676–684.
- DeGiorgio, M., Jakobsson, M., and Rosenberg, N. A. (2009). Out of Africa: modern human origins special feature: explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16057–16062.
- Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. *Science* 295, 1306–1311.
- Dennis, M. Y., Nuttle, X., Sudmant, P. H., Antonacci, F., Graves, T. A., Nefedov, M., et al. (2012). Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* 149, 912–922.
- Dickel, D. E., Ypsilanti, A. R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B. J., et al. (2018). Ultraconserved enhancers are required for normal development. *Cell* 172, 491–499.e15.
- Dickinson, M. E., The International Mouse Phenotyping Consortium, Flenniken, A. M., Ji, X., Teboul, L., Wong, M. D., et al. (2016). High-throughput discovery of novel developmental phenotypes. *Nature* 537, 508–514.
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380.
- Doggett, N. A., Xie, G., Meincke, L. J., Sutherland, R. D., Mundt, M. O., Berbari, N. S., et al. (2006). A 360-kb interchromosomal duplication of the human HYDIN locus. *Genomics* 88, 762–771.
- Domené, S., Bumashny, V. F., de Souza, F. S. J., Franchini, L. F., Nasif, S., Low, M. J., et al. (2013). Enhancer turnover and conserved regulatory function in vertebrate evolution. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20130027.
- Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R. D. (2020). Candidate silencer elements for the human and mouse genomes. *Nat. Commun.* 11, 1061.
- Dottori, M., Gross, M. K., Labosky, P., and Goulding, M. (2001). The winged-helix transcription factor Foxd3 suppresses interneuron differentiation and promotes neural crest cell fate. *Development* 128, 4127–4138.
- Dumas, L., and Sikela, J. M. (2009). DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb. Symp. Quant. Biol.* 74, 375–382.

- Duret, L., and Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* 10, 285–311.
- Dutrow, E. V., Emera, D., Yim, K., Uebbing, S., Kocher, A. A., Krenzer, M., et al. (2022). Modeling uniquely human gene regulatory function via targeted humanization of the mouse genome. *Nat. Commun.* 13, 304.
- Edmondson, D. G., Lyons, G. E., Martin, J. F., and Olson, E. N. (1994). Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development* 120, 1251–1263.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* 418, 869–872.
- ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shores, N., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 583, 699–710.
- Everaert, C., Luypaert, M., Maag, J. L. V., Cheng, Q. X., Dinger, M. E., Hellemans, J., et al. (2017). Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Sci. Rep.* 7, 1559.
- Falk, D., Aiello, L., and Dean, C. (1992). An introduction to human evolutionary anatomy. *Man (Lond.)* 27, 410.
- Falk, D., Redmond, J. C., Jr, Guyer, J., Conroy, C., Recheis, W., Weber, G. W., et al. (2000). Early hominid brain evolution: a new look at old endocasts. *J. Hum. Evol.* 38, 695–717.
- Ferguson, S. M., and De Camilli, P. (2012). Dynamin, a membrane-remodelling GTPase. *Nat. Rev. Mol. Cell Biol.* 13, 75–88.
- Ferraguti, F., and Shigemoto, R. (2006). Metabotropic glutamate receptors. *Cell Tissue Res.* 326, 483–504.
- Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., and McCallion, A. S. (2006a). Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* 312, 276–279.
- Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L., Urasaki, A., Kawakami, K., et al. (2006b). Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat. Protoc.* 1, 1297–1305.
- Fortna, A., Kim, Y., MacLaren, E., Marshall, K., Hahn, G., Meltesen, L., et al. (2004). Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol.* 2, E207.
- Franchini, L. F. (2021). Genetic mechanisms underlying cortical evolution in mammals. *Front. Cell Dev. Biol.* 9, 591017.
- Franchini, L. F., and Pollard, K. S. (2015). Genomic approaches to studying human-specific developmental traits. *Development* 142, 3100–3112.
- Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F., and Stern, D. L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* 466, 490–493.
- Friedlander, D. R., Milev, P., Karthikeyan, L., Margolis, R. K., Margolis, R. U., and Grumet, M. (1994). The neuronal chondroitin sulfate proteoglycan neurocan binds to the neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits neuronal adhesion and neurite outgrowth. *J. Cell Biol.* 125, 669–680.

- Fritschy, J.-M., and Panzanelli, P. (2014). GABAA receptors and plasticity of inhibitory neurotransmission in the central nervous system. *Eur. J. Neurosci.* 39, 1845–1865.
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L. A. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038–2049.
- Gao, C., Ren, S., Lee, J.-H., Qiu, J., Chapski, D. J., Rau, C. D., et al. (2016). RBFOX1-mediated RNA splicing regulates cardiac hypertrophy and heart failure. *J. Clin. Invest.* 126, 195–206.
- Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat. Rev. Genet.* 7, 703–713.
- Geertz, M., and Maerkl, S. J. (2010). Experimental strategies for studying transcription factor-DNA binding specificities. *Brief. Funct. Genomics* 9, 362–373.
- Gehman, L. T., Meera, P., Stoilov, P., Shiue, L., O'Brien, J. E., Meisler, M. H., et al. (2012). The splicing regulator Rbfox2 is required for both cerebellar development and mature motor function. *Genes Dev.* 26, 445–460.
- Gehman, L. T., Stoilov, P., Maguire, J., Damianov, A., Lin, C.-H., Shiue, L., et al. (2011). The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* 43, 706–711.
- Geschwind, D. H., and Rakic, P. (2013). Cortical evolution: judge the brain by its cover. *Neuron* 80, 633–647.
- Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., Mardis, E. R., et al. (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316, 222–234.
- Girskis, K. M., Stergachis, A. B., DeGennaro, E. M., Doan, R. N., Qian, X., Johnson, M. B., et al. (2021). Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* 109, 3239–3251.e7.
- Goidts, V., Cooper, D. N., Armengol, L., Schempp, W., Conroy, J., Estivill, X., et al. (2006). Complex patterns of copy number variation at sites of segmental duplications: an important category of structural variation in the human genome. *Hum. Genet.* 120, 270–284.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., et al. (2010). A draft sequence of the Neandertal genome. *Science* 328, 710–722.
- Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D. U., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900–910.
- Haberle, V., and Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* 19, 621–637.
- Halpain, S., and Dehmelt, L. (2006). The MAP1 family of microtubule-associated proteins. *Genome Biol.* 7, 224.
- Hammock, E. A. D., and Levitt, P. (2011). Developmental expression mapping of a gene implicated in multiple neurodevelopmental disorders, A2bp1 (Fox1). *Dev. Neurosci.* 33, 64–74.
- Hawkins, R. D., Hon, G. C., and Ren, B. (2010). Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11, 476–486.
- Heidenreich, K. A., and Linseman, D. A. (2004). Myocyte enhancer factor-2 transcription factors in neuronal differentiation and survival. *Mol. Neurobiol.* 29, 155–166.

- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc. Natl. Acad. Sci. U. S. A.* 109 Suppl 1, 10661–10668.
- Hibino, H., Inanobe, A., Furutani, K., Murakami, S., Findlay, I., and Kurachi, Y. (2010). Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol. Rev.* 90, 291–366.
- Hubisz, M. J., Pollard, K. S., and Siepel, A. (2011). PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* 12, 41–51.
- Hublin, J. J. (2009). Out of Africa: modern human origins special feature: the origin of Neandertals. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16022–16027.
- Inoue, F., and Ahituv, N. (2015). Decoding enhancers using massively parallel reporter assays. *Genomics* 106, 159–164.
- Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., et al. (2003). A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* 22, 905–912.
- Johanson, D. C., and White, T. D. (1980). On the status of *Australopithecus afarensis*. *Science* 207, 1104–1105.
- Kaas, J. H. (2013). The evolution of brains from early mammals to humans. *Wiley Interdiscip. Rev. Cogn. Sci.* 4, 33–45.
- Kaczynski, J., Cook, T., and Urrutia, R. (2003). Sp1- and Krüppel-like transcription factors. *Genome Biol.* 4, 206.
- Kamm, G. B., López-Leal, R., Lorenzo, J. R., and Franchini, L. F. (2013a). A fast-evolving human NPAS3 enhancer gained reporter expression in the developing forebrain of transgenic mice. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20130019.
- Kamm, G. B., Pisciotto, F., Kliger, R., and Franchini, L. F. (2013b). The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Mol. Biol. Evol.* 30, 1088–1102.
- Karlsson, J., von Hofsten, J., and Olsson, P. E. (2001). Generating transparent zebrafish: a refined method to improve detection of gene expression during embryonic development. *Mar. Biotechnol. (NY)* 3, 522–527.
- Kawakami, K. (2007). Tol2: a versatile gene transfer vector in vertebrates. *Genome Biol.* 8 Suppl 1, S7.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915.
- King, M. C., and Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116.
- Klinck, R., Fourier, A., Thibault, P., Toutant, J., Durand, M., Lapointe, E., et al. (2014). RBFOX1 cooperates with MBNL1 to control splicing in muscle, including events altered in myotonic dystrophy type 1. *PLoS One* 9, e107324.

- Kostka, D., Hubisz, M. J., Siepel, A., and Pollard, K. S. (2012). The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. *Mol. Biol. Evol.* 29, 1047–1057.
- Krecic, A. M., and Swanson, M. S. (1999). hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* 11, 363–371.
- Kulikova, A. V., Korostina, V. S., Kulikova, E. A., Fursenko, D. V., Akulov, A. E., Moshkin, M. P., et al. (2016). Knockout Zbtb33 gene results in an increased locomotion, exploration and pre-pulse inhibition in mice. *Behav. Brain Res.* 297, 76–83.
- Kuroyanagi, H. (2009). Fox-1 family of RNA-binding proteins. *Cell. Mol. Life Sci.* 66, 3895–3907.
- Kvon, E. Z., Kamneva, O. K., Melo, U. S., Barozzi, I., Osterwalder, M., Mannion, B. J., et al. (2016). Progressive loss of function in a limb enhancer during snake evolution. *Cell* 167, 633–642.e11.
- Kvon, E. Z., Zhu, Y., Kelman, G., Novak, C. S., Plajzer-Frick, I., Kato, M., et al. (2020). Comprehensive in vivo interrogation reveals phenotypic impact of human enhancer variants. *Cell* 180, 1262–1271.e15.
- Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., and Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523.
- Lal, D., Pernhorst, K., Klein, K. M., Reif, P., Tozzi, R., Toliat, M. R., et al. (2015). Extending the phenotypic spectrum of RBFOX1 deletions: Sporadic focal epilepsy. *Epilepsia* 56, e129–33.
- Lal, D., Reinthaler, E. M., Altmüller, J., Toliat, M. R., Thiele, H., Nürnberg, P., et al. (2013a). RBFOX1 and RBFOX3 mutations in rolandic epilepsy. *PLoS One* 8, e73323.
- Lal, D., Trucks, H., Møller, R. S., Hjalgrim, H., Koeleman, B. P. C., de Kovel, C. G. F., et al. (2013b). Rare exonic deletions of the RBFOX1 gene increase risk of idiopathic generalized epilepsy. *Epilepsia* 54, 265–271.
- Lam, D. D., de Souza, F. S. J., Nasif, S., Yamashita, M., López-Leal, R., Otero-Corchon, V., et al. (2015). Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold. *PLoS Genet.* 11, e1004935.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9, e1003118.
- Leakey, L. S. B., Tobias, P. V., and Napier, J. R. (1965). A new species of genus homo from Olduvai gorge. *Curr. Anthropol.* 6, 424–427.
- Lee, J., Kim, M., Itoh, T. Q., and Lim, C. (2018). Ataxin-2: A versatile posttranscriptional regulator and its implication in neural function. *Wiley Interdiscip. Rev. RNA* 9, e1488.
- Lee, J.-A., Damianov, A., Lin, C.-H., Fontes, M., Parikshak, N. N., Anderson, E. S., et al. (2016). Cytoplasmic Rbfox1 regulates the expression of synaptic and autism-related genes. *Neuron* 89, 113–128.
- Levchenko, A., Kanapin, A., Samsonova, A., and Gainetdinov, R. R. (2018). Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. *Genome Biol. Evol.* 10, 166–188.
- Li, Y. I., Sanchez-Pulido, L., Haerty, W., and Ponting, C. P. (2015). RBFOX and PTBP1 proteins regulate the alternative splicing of micro-exons in human brain transcripts. *Genome Res.* 25, 1–13.

- Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Liao, Y., Smyth, G. K., and Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* 47, e47.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Lim, L. (1992). n-Chimaerin and neuronal signal transduction mechanisms. *Biochem. Soc. Trans.* 20, 611–614.
- Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M. F., Parker, B. J., Washietl, S., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 478, 476–482.
- Lipscombe, D., and Lopez-Soto, E. J. (2018). Protected by a fox. *Neuron* 98, 3–5.
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25, 402–408.
- Lovci, M. T., Ghanem, D., Marr, H., Arnold, J., Gee, S., Parra, M., et al. (2013). Rbfox proteins regulate alternative mRNA splicing through evolutionarily conserved RNA bridges. *Nat. Struct. Mol. Biol.* 20, 1434–1442.
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Malik, T. H., Shoichet, S. A., Latham, P., Kroll, T. G., Peters, L. L., and Shivdasani, R. A. (2001). Transcriptional repression and developmental functions of the atypical vertebrate GATA protein TRPS1. *EMBO J.* 20, 1715–1725.
- Martin, C. L., Duvall, J. A., Ilkin, Y., Simon, J. S., Arreaza, M. G., Wilkes, K., et al. (2007). Cytogenetic and molecular characterization of A2BP1/FOX1 as a candidate gene for autism. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 144B, 869–876.
- Maston, G. A., Evans, S. K., and Green, M. R. (2006). Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* 7, 29–59.
- Matharu, N., and Ahituv, N. (2015). Minor loops in major folds: Enhancer-promoter looping, chromatin restructuring, and their association with transcriptional regulation and disease. *PLoS Genet.* 11, e1005640.
- McHenry, H. M. (1975). Fossils and the mosaic nature of human evolution. *Science* 190, 425–431.
- Meylan, P., Dreos, R., Ambrosini, G., Groux, R., and Bucher, P. (2020). EPD in 2020: enhanced data visualization and extension to ncRNA promoters. *Nucleic Acids Res.* 48, D65–D69.
- Nakahata, S., and Kawamoto, S. (2005). Tissue-dependent isoforms of mammalian Fox-1 homologs are associated with tissue-specific splicing activities. *Nucleic Acids Res.* 33, 2078–2089.
- Naya, F. J., and Olson, E. (1999). MEF2: a transcriptional target for signaling pathways controlling skeletal muscle growth and differentiation. *Curr. Opin. Cell Biol.* 11, 683–688.
- Nicholas, A. K., Khurshid, M., Désir, J., Carvalho, O. P., Cox, J. J., Thornton, G., et al. (2010). WDR62 is associated with the spindle pole and is mutated in human microcephaly. *Nat. Genet.* 42, 1010–1014.

- Noonan, J. P., and McCallion, A. S. (2010). Genomics of long-range regulatory elements. *Annu. Rev. Genomics Hum. Genet.* 11, 1–23.
- O’Bleness, M., Searles, V. B., Varki, A., Gagneux, P., and Sikela, J. M. (2012). Evolution of genetic and genomic features unique to the human lineage. *Nat. Rev. Genet.* 13, 853–866.
- O’Brien, J. E., Drews, V. L., Jones, J. M., Dugas, J. C., Barres, B. A., and Meisler, M. H. (2012). Rbfox proteins regulate alternative splicing of neuronal sodium channel SCN8A. *Mol. Cell. Neurosci.* 49, 120–126.
- Oksenberg, N., Stevison, L., Wall, J. D., and Ahituv, N. (2013). Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genet.* 9, e1003221.
- Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243.
- Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502.
- Pervaiz, N., and Abbasi, A. A. (2016). Molecular evolution of WDR62, a gene that regulates neocortico genesis. *Meta Gene* 9, 1–9.
- Phillips-Cremins, J. E., Sauria, M. E. G., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S. K., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295.
- Pickard, B. S., Pieper, A. A., Porteous, D. J., Blackwood, D. H., and Muir, W. J. (2006). The NPAS3 gene—emerging evidence for a role in psychiatric illness. *Ann. Med.* 38, 439–448.
- Pickford, M., and Senut, B. (2001). The geological and faunal context of Late Miocene hominid remains from Lukeino, Kenya. *C. R. Acad. Sci. (Ser. 2a) (Sci. Terre Planete/Earth Planet. Sci.)* 332, 145–152.
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121.
- Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., et al. (2006). Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2, e168.
- Powers, R. M., Hevner, R. F., and Halpain, S. (2022). The Neuron Navigators: Structure, function, and evolutionary history. *Front. Mol. Neurosci.* 15, 1099554.
- Prabhakar, S., Noonan, J. P., Pääbo, S., and Rubin, E. M. (2006). Accelerated evolution of conserved noncoding sequences in humans. *Science* 314, 786.
- Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., et al. (2008). Human-specific gain of function in a developmental enhancer. *Science* 321, 1346–1350.
- Preuss, T. M. (1995). Do rats have prefrontal cortex? The rose-woolsey-akert program reconsidered. *J. Cogn. Neurosci.* 7, 1–24.
- Preuss, T. M. (2007). “Primate Brain Evolution in Phylogenetic Context,” in *Evolution of Nervous Systems* (Elsevier), 1–34.
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505, 43–49.

- Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nat. Rev. Neurosci.* 10, 724–735.
- Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680.
- Reilly, S. K., and Noonan, J. P. (2016). Evolution of gene regulation in humans. *Annu. Rev. Genomics Hum. Genet.* 17, 45–67.
- Relethford, J. H. (2008). Genetic evidence and the modern human origins debate. *Heredity (Edinb.)* 100, 555–563.
- Rightmire, G. P. (2009). Out of Africa: modern human origins special feature: middle and later Pleistocene hominins in Africa and Southwest Asia. *Proc. Natl. Acad. Sci. U. S. A.* 106, 16046–16050.
- Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428.
- Ritter, D. I., Dong, Z., Guo, S., and Chuang, J. H. (2012). Transcriptional enhancers in protein-coding exons of vertebrate developmental genes. *PLoS One* 7, e35202.
- Rizo, J., and Südhof, T. C. (2012). The membrane fusion enigma: SNAREs, Sec1/Munc18 proteins, and their accomplices—guilty as charged? *Annu. Rev. Cell Dev. Biol.* 28, 279–308.
- Roth, G., and Dicke, U. (2005). Evolution of the brain and intelligence. *Trends Cogn. Sci.* 9, 250–257.
- Rubinstein, M., and de Souza, F. S. J. (2013). Evolution of transcriptional enhancers and animal diversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20130017.
- Sagai, T., Hosoya, M., Mizushima, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* 132, 797–803.
- Scally, A., Dutheil, J. Y., Hillier, L. W., Jordan, G. E., Goodhead, I., Herrero, J., et al. (2012). Insights into hominid evolution from the gorilla genome sequence. *Nature* 483, 169–175.
- Schaeren-Wiemers, N., and Gerfin-Moser, A. (1993). A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: in situ hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* 100, 431–440.
- Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., et al. (2007). Strong association of de novo copy number mutations with autism. *Science* 316, 445–449.
- Segert, J. A., Gisselbrecht, S. S., and Bulyk, M. L. (2021). Transcriptional silencers: Driving gene expression with the brakes on. *Trends Genet.* 37, 514–527.
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311.
- Sherwood, C. C., and Hof, P. R. (2007). “The evolution of neuron types and cortical histology in apes and humans,” in *Evolution of Nervous Systems* (Elsevier), 355–378.
- Sherwood, C. C., Subiaul, F., and Zawidzki, T. W. (2008). A natural history of the human mind: tracing evolutionary changes in brain and cognition. *J. Anat.* 212, 426–454.

- Shibata, H., Huynh, D. P., and Pulst, S. M. (2000). A novel protein with RNA-binding motifs interacts with ataxin-2. *Hum. Mol. Genet.* 9, 1303–1313.
- Sikela, J. M. (2006). The jewels of our genome: the search for the genomic changes underlying the evolutionarily unique capacities of the human brain. *PLoS Genet.* 2, e80.
- Spilker, C., Acuña Sanhueza, G. A., Böckers, T. M., Kreutz, M. R., and Gundelfinger, E. D. (2008). SPAR2, a novel SPAR-related protein with GAP activity for Rap1 and Rap2. *J. Neurochem.* 104, 187–201.
- Spiteri, E., Konopka, G., Coppola, G., Bomar, J., Oldham, M., Ou, J., et al. (2007). Identification of the transcriptional targets of FOXP2, a gene linked to speech and language, in developing human brain. *Am. J. Hum. Genet.* 81, 1144–1157.
- Stephan, H., Frahm, H., and Baron, G. (1981). New and revised data on volumes of brain structures in insectivores and primates. *Folia Primatol. (Basel)* 35, 1–29.
- Striedter, G. F. (2006). Précis of principles of brain evolution. *Behav. Brain Sci.* 29, 1–12; discussion 12–36.
- Sultana, R., Yu, C.-E., Yu, J., Munson, J., Chen, D., Hua, W., et al. (2002). Identification of a novel gene on chromosome 7q11.2 interrupted by a translocation breakpoint in a pair of autistic twins. *Genomics* 80, 129–134.
- Sun, J., Yang, J., Miao, X., Loh, H. H., Pei, D., and Zheng, H. (2021). Proteins in DNA methylation and their role in neural stem cell proliferation and differentiation. *Cell Regen. (Lond.)* 10, 7.
- Sun, S., Zhang, Z., Fregoso, O., and Krainer, A. R. (2012). Mechanisms of activation and repression by the alternative splicing factors RBFOX1/2. *RNA* 18, 274–283.
- Symmons, O., Uslu, V. V., Tsujimura, T., Ruf, S., Nassari, S., Schwarzer, W., et al. (2014). Functional and topological characteristics of mammalian regulatory domains. *Genome Res.* 24, 390–400.
- Tafessu, A., and Banaszynski, L. A. (2020). Establishment and function of chromatin modification at enhancers. *Open Biol.* 10, 200255.
- Tattersall, I. (2010). Human evolution and cognition. *Theory Biosci.* 129, 193–201.
- The 1000 Genomes Project Consortium, Auton, A., Abecasis, G. R., Altshuler, D. M., (Co-Chair), Durbin, R. M., (Co-Chair), Abecasis, G. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
- Thisse, B., and Thisse, C. (2014). In situ hybridization on whole-mount zebrafish embryos and young larvae. *Methods Mol. Biol.* 1211, 53–67.
- Turner, T. N., Hormozdiari, F., Duyzend, M. H., McClymont, S. A., Hook, P. W., Iossifov, I., et al. (2016). Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am. J. Hum. Genet.* 98, 58–74.
- Uebbing, S., Gockley, J., Reilly, S. K., Kocher, A. A., Geller, E., Gandotra, N., et al. (2021). Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2007049118.
- Underwood, J. G., Boutz, P. L., Dougherty, J. D., Stoilov, P., and Black, D. L. (2005). Homologues of the *Caenorhabditis elegans* Fox-1 protein are neuronal splicing regulators in mammals. *Mol. Cell. Biol.* 25, 10005–10016.
- Vallender, E. J., Mekel-Bobrov, N., and Lahn, B. T. (2008). Genetic basis of human brain evolution. *Trends Neurosci.* 31, 637–644.

- Varki, A., and Altheide, T. K. (2005). Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* 15, 1746–1758.
- Vernes, S. C., Oliver, P. L., Spiteri, E., Lockstone, H. E., Puliyadi, R., Taylor, J. M., et al. (2011). Foxp2 regulates gene networks implicated in neurite outgrowth in the developing brain. *PLoS Genet.* 7, e1002145.
- Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L. A. (2007). VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35, D88–92.
- Visel, A., Prabhakar, S., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., et al. (2008). Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* 40, 158–160.
- Visel, A., Rubin, E. M., and Pennacchio, L. A. (2009). Genomic views of distant-acting enhancers. *Nature* 461, 199–205.
- Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384.
- von Spiczak, S., Helbig, K. L., Shinde, D. N., Huether, R., Pendziwiat, M., Lourenço, C., et al. (2017). DNMI1 encephalopathy: A new disease of vesicle fission. *Neurology* 89, 385–394.
- Vuong, C. K., Black, D. L., and Zheng, S. (2016). The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* 17, 265–281.
- Vuong, C. K., Wei, W., Lee, J.-A., Lin, C.-H., Damianov, A., de la Torre-Ubieta, L., et al. (2018). Rbfox1 regulates synaptic transmission through the inhibitory neuron-specific vSNARE Vamp1. *Neuron* 98, 127–141.e7.
- Wamsley, B., Jaglin, X. H., Favuzzi, E., Quattrocchio, G., Nigro, M. J., Yusuf, N., et al. (2018). Rbfox1 mediates cell-type-specific splicing in cortical interneurons. *Neuron* 100, 846–859.e7.
- Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F., et al. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153, 910–918.
- Wehr, R., Mansouri, A., de Maeyer, T., and Gruss, P. (1997). Fkh5-deficient mice show dysgenesis in the caudal midbrain and hypothalamic mammillary body. *Development* 124, 4447–4456.
- Weledji, E. P., and Assob, J. C. (2014). The ubiquitous neural cell adhesion molecule (N-CAM). *Ann. Med. Surg. (Lond.)* 3, 77–81.
- Weyn-Vanhentenryck, S. M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 6, 1139–1152.
- Whalen, S., Inoue, F., Ryu, H., Fair, T., Markenscoff-Papadimitriou, E., Keough, K., et al. (2023). Machine learning dissection of human accelerated regions in primate neurodevelopment. *Neuron*. doi: 10.1016/j.neuron.2022.12.026.
- Whalen, S., and Pollard, K. S. (2022). Enhancer function and evolutionary roles of human accelerated regions. *Annu. Rev. Genet.* 56. doi: 10.1146/annurev-genet-071819-103933.
- Wijchers, P. J. E. C., Burbach, J. P. H., and Smidt, M. P. (2006). In control of biology: of mice, men and Foxes. *Biochem. J.* 397, 233–246.

- Wittkopp, P. J., and Kalay, G. (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69.
- Won, H., de la Torre-Ubieta, L., Stein, J. L., Parikshak, N. N., Huang, J., Opland, C. K., et al. (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* 538, 523–527.
- Wood, B. (1996). Human evolution. *Bioessays* 18, 945–954.
- Wood, B., and Collard, M. (1999). The human genus. *Science* 284, 65–71.
- Wray, G. A. (2007). The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* 8, 206–216.
- Yu, T. W., Mochida, G. H., Tischfield, D. J., Sgaier, S. K., Flores-Sarnat, L., Sergi, C. M., et al. (2010). Mutations in WDR62, encoding a centrosome-associated protein, cause microcephaly with simplified gyri and abnormal cortical architecture. *Nat. Genet.* 42, 1015–1020.
- Yunis, J. J., and Prakash, O. (1982). The origin of man: a chromosomal pictorial legacy. *Science* 215, 1525–1530.
- Zabidi, M. A., Arnold, C. D., Scherhuber, K., Pagani, M., Rath, M., Frank, O., et al. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* 518, 556–559.
- Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA β geo 26 gene trap strain leads to widespread expression of β -galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl. Acad. Sci. U. S. A.* 94, 3789–3794.
- Zhang, C., Frias, M. A., Mele, A., Ruggiu, M., Eom, T., Marney, C. B., et al. (2010). Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 329, 439–443.
- Zhang, C., Zhang, Z., Castle, J., Sun, S., Johnson, J., Krainer, A. R., et al. (2008). Defining the regulatory network of the tissue-specific splicing factors Fox-1 and Fox-2. *Genes Dev.* 22, 2550–2563.
- Zhang, J., Webb, D. M., and Podlaha, O. (2002). Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* 162, 1825–1835.
- Zhao, W.-W. (2013). Intragenic deletion of RBFOX1 associated with neurodevelopmental/neuropsychiatric disorders and possibly other clinical presentations. *Mol. Cytogenet.* 6, 26.
- Zhou, Y., and Danbolt, N. C. (2014). Glutamate as a neurotransmitter in the healthy brain. *J. Neural Transm. (Vienna)* 121, 799–817.
- Zollikofer, C. P. E., Ponce de León, M. S., Lieberman, D. E., Guy, F., Pilbeam, D., Likius, A., et al. (2005). Virtual cranial reconstruction of *Sahelanthropus tchadensis*. *Nature* 434, 755–759.