

Tesis Doctoral

Bases genéticas de la evolución del cerebro humano: estudio evolutivo y funcional del gen NPAS3

Kamm, Gretel Betiana

2014-04-24

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Kamm, Gretel Betiana. (2014-04-24). Bases genéticas de la evolución del cerebro humano: estudio evolutivo y funcional del gen NPAS3. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Kamm, Gretel Betiana. "Bases genéticas de la evolución del cerebro humano: estudio evolutivo y funcional del gen NPAS3". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2014-04-24.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

Bases genéticas de la evolución del cerebro humano: estudio evolutivo y funcional del gen *NPAS3*

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

Gretel Betiana Kamm

Directora de tesis: Dra. Lucía Florencia Franchini
Consejero de estudios: Dr. Marcelo Rubinstein

Lugar de trabajo: Instituto de Investigaciones en Ingeniería Genética y Biología Molecular Dr. Héctor N. Torres.

Buenos Aires, 2014

RESUMEN

Comprender el origen y la evolución del cerebro humano es uno de los desafíos más sobresalientes que enfrenta la ciencia. Los cambios genéticos que llevaron a la adquisición de las capacidades distintivas del cerebro humano están codificados en nuestro genoma. La disponibilidad de más de cincuenta genomas de vertebrados permite hoy reconstruir nuestra historia evolutiva. Nuestra hipótesis es que 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 de las capacidades cognitivas excepcionales de nuestro cerebro. Haciendo uso de bases de datos públicas de secuencias no codificantes conservadas con evidencia de evolución acelerada en el linaje humano (denominados HAEs por *human accelerated elements*), se encontró que el factor de transcripción *neuronal PAS domain-containing protein 3 (NPAS3)* contiene 14 HAEs, el mayor número de HAEs para un solo gen en todo el genoma humano. Usando un ensayo de expresión en peces cebra transgénicos se demostró que 11 de los 14 HAEs son capaces de activar la expresión de la proteína reportera EGFP durante el desarrollo embrionario, particularmente en el sistema nervioso. Además, utilizando ratones transgénicos, se realizó un análisis comparativo estudiando los patrones de expresión de uno de los HAEs de *NPAS3* y secuencias ortólogas de chimpancé y ratón. El *enhancer* humano muestra una extensión del patrón de expresión en el telencéfalo. Este cambio humano específico pudo haber contribuido con la evolución de alguna de las características de nuestro cerebro.

ABSTRACT

Understanding the origin and evolution of the human brain is one of the greatest challenges that today science faces. The genetic changes that led to the acquisition of the distinctive capacities of the human brain are encoded in our genome. The availability of more than fifty vertebrate genomes allows unraveling our evolutionary history. Our hypothesis is that the acquisition of new expression patterns in the human lineage of genes involved with the development and functioning of the brain would have been critical for the evolution of our unique cognitive capacities. Using public databases of human accelerated conserved non coding sequences (HAEs or *human accelerated elements*), we found that the transcription factor *neuronal PAS domain-containing protein 3* (*NPAS3*) contains 14 HAEs, the largest number detected for a human gene. Using an enhancer transcription assay in transgenic zebrafish we show that 11 out of the 14 HAEs activated the expression of the reporter gen EGFP during zebrafish development in the central nervous system. In addition, using transgenic mice we performed an expression pattern comparative analysis of human, chimpanzee and mouse ortholog sequences of a selected HAE. We found that the human enhancer shows an extended expression pattern in the forebrain. This human-specific change could have contributed to the evolution of some of the distinctive capacities of our brain.

“La realidad no nos habla objetivamente, y no hay científico que pueda verse libre de las coerciones de la psique y la sociedad. El mayor impedimento para la innovación científica suele ser una cerrazón conceptual, no una falta de hechos”.

S. J. Gould. La vida maravillosa. 1999.

AGRADECIMIENTOS

A la Universidad de Buenos Aires por la educación gratuita de excelencia que me brindó a lo largo de todos estos años, permitiéndome acceder a una formación académica y profesional de altísima calidad.

A las autoridades del Instituto de Ingeniería Genética y Biología Molecular (INGEBI) y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y a la Agencia Nacional de Promoción Científica y Tecnológica por haber hecho posible la realización de esta tesis garantizando la financiación e infraestructura necesarias.

A la Dra. Lucía Florencia Franchini por abrir las puertas de su laboratorio, y haber confiado en mí para llevar adelante este proyecto. Agradezco especialmente la dedicación de mi directora de tesis que, en todo momento, mostró un enorme compromiso con el desarrollo de este trabajo y con mi formación profesional.

Al Dr. Marcelo Rubinstein por brindar un apoyo incondicional a nuestro trabajo. Muchas gracias también por los invaluable consejos y por mostrar una gran vocación científica que es transmitida a todos sus estudiantes.

A mis compañeros de los laboratorios de la Dra. Franchini y del Dr. Rubinstein, que me han acompañado durante estos años, por aportar valiosas ideas y consejos. Mi particular agradecimiento a Francisco por sus contribuciones valiosas a este trabajo, a Rodrigo por haber dado una mano invaluable en este proyecto, a Sabina por haber guiado mis primeros pasos en el acuario, a Jéssica y Ramiro por su generosa ayuda con producción de ratones transgénicos, a Jéssica y Vanina por la asistencia permanente con la colonia de ratones enseñándonos siempre como mejorar nuestras técnicas en el bioterio, a Daniela, Laura y Milena por colaborar en el trabajo diario de mantener el acuario y a Sofía por su apoyo y amistad.

Al Departamento de Ecología, Genética y Evolución de esta Facultad y especialmente a todo el cuerpo docente de Evolución. Muchas gracias por haber permitido disfrutar junto a ustedes del placer de la docencia y por haber sido un pilar fundamental de mi formación.

Al personal de apoyo del INGEBI que incondicionalmente garantiza las condiciones necesarias para desarrollar la tarea científica.

A mi mamá y a mi papá, por su apoyo constante.

A Juan, por tu infinita paciencia y generosa ayuda.

ÍNDICE

RESUMEN	2
ABSTRACT	3
AGRADECIMIENTOS	4
ABREVIATURAS	9
INTRODUCCIÓN GENERAL	12
Conócete a ti mismo.....	12
La herencia primate.....	14
Historia evolutiva de nuestro linaje.....	18
Las bases genéticas de la evolución humana.....	21
1) Cambios genómicos de gran escala.....	23
2) Cambios genéticos pequeños.....	23
a) Cambios en el número de copias de un gen.....	24
b) Cambios en la secuencia codificante de un gen.....	25
c) Pérdida de genes.....	27
d) Cambios en las regiones regulatorias en <i>cis</i>	28
Los aportes del campo de la evolución del desarrollo.....	29
Los mecanismos que subyacen la evolución morfológica.....	30
El estudio de regiones regulatorias.....	32
El código regulatorio del genoma.....	32
Los elementos regulatorios.....	33
Predicción de <i>enhancers</i>	36
Genes blanco de los <i>enhancers</i>	39
Ensayos funcionales para regiones regulatorias.....	41
Origen de los <i>enhancers</i> y exaltación.....	44
HIPÓTESIS GENERAL	46
OBJETIVO GENERAL	46
Capítulo I	47
INTRODUCCIÓN ESPECÍFICA	48
La evolución regulatoria de nuestro genoma.....	48
Detección de evolución acelerada en el genoma humano.....	50
HIPÓTESIS DE TRABAJO	53
OBJETIVOS ESPECÍFICOS	53
RESULTADOS	54
Análisis a escala genómica del agrupamiento de regiones no codificantes con evidencia de evolución acelerada en el genoma humano.....	54
Estudio de agrupamiento de regiones aceleradas en el genoma humano.....	54
Baja redundancia entre las bases de datos de HAEs.....	56
<i>NPAS3</i> se localiza en el <i>locus</i> con más HAEs del genoma humano.....	58
Estudio de las regiones codificantes de <i>NPAS3</i>	61
Polimorfismos en los HAEs presentes en intrones de <i>NPAS3</i>	62
Capítulo II	66
INTRODUCCIÓN ESPECÍFICA	67
Caracterización funcional de los HAEs.....	67
Del genoma al estudio de un gen específico.....	68
<i>Tracheiless</i> , el gen ortólogo de <i>NPAS3</i> en moscas.....	70
Patrón de expresión de <i>NPAS3</i> en el sistema nervioso.....	72
<i>NPAS3</i> y el análisis de ratones mutantes.....	72
<i>NPAS3</i> , un gen cuya disfunción está asociada al desarrollo de esquizofrenia y otras patologías del sistema nervioso.....	73
HIPÓTESIS DE TRABAJO	76
OBJETIVOS ESPECÍFICOS	76

RESULTADOS.....	78
Análisis del patrón de expresión de <i>NPAS3</i> en pez cebra.....	78
Caracterización funcional de las regiones aceleradas de <i>NPAS3</i> en peces transgénicos.....	80
Comparación de los patrones de expresión transiente de los elementos HAR21, HAR89, HAR96 y HAR202 y sus ortólogos en chimpancé en peces transgénicos.....	84
Análisis de la actividad <i>enhancer</i> de los HAEs en líneas transgénicas en pez cebra.....	85
Análisis comparativo de los elementos HAR21, HAR89 y HAR202 en líneas de peces transgénicos.....	93
El elemento acelerado no codificante 2xHAR142 se comporta como un <i>enhancer</i> activo durante el desarrollo embrionario en ratones transgénicos...	94
El elemento 2xHAR142 dirige la expresión de <i>LacZ</i> a un dominio que se superpone con el dominio de expresión del gen <i>NPAS3</i>	100
Origen evolutivo de las regiones aceleradas.....	107
El elemento acelerado no codificante 2xHAR142 contiene una secuencia derivada de un tranposón de la familia MIR.....	111
Capítulo III.....	113
INTRODUCCIÓN ESPECÍFICA.....	114
Morfolinos.....	114
HIPÓTESIS DE TRABAJO.....	116
OBJETIVO ESPECÍFICO.....	116
RESULTADOS.....	117
Análisis del patrón de expresión de <i>NPAS3</i> en pez cebra	117
Fenotipo de larvas de pez cebra tratadas con morfolinos contra <i>NPAS3</i>	118
DISCUSIÓN DE RESULTADOS.....	128
Agrupamiento de HAEs en el genoma humano.....	128
Análisis del patrón evolutivo en regiones codificantes y no codificantes en el <i>locus</i> de <i>NPAS3</i>	128
Análisis de las sustituciones exclusivas de homínidos en los HAEs de <i>NPAS3</i>	130
Ensayos funcionales en peces cebra transgénicos.....	132
Ensayos funcionales en ratones transgénicos.....	136
Origen evolutivo del elemento 2xHAR142 y otras regiones con signos de exaptación en el genoma humano.....	137
Comparación del patrón de expresión de <i>NPAS3</i> entre vertebrados.....	138
Ensayos funcionales con morfolinos contra el ARNm de <i>NPAS3</i>	140
<i>NPAS3</i> posiblemente forme parte de un módulo genético muy estable y antiguo asociado a la morfogénesis por ramificación.....	143
<i>NPAS3</i> , ¿Un gen con una compleja red regulatoria?.....	145
Dinámica de los elementos no codificantes conservados en el <i>locus</i> de <i>NPAS3</i>	148
MATERIALES Y MÉTODOS.....	152
Análisis de agrupamiento de regiones aceleradas en linaje humano.....	152
Generación de la base de datos de HAEs.....	152
Generación de las bases de datos de intervalos genómicos.....	152
Agrupamiento de HAEs en el genoma humano.....	153
Análisis de agrupamiento de HAEs en genes gigantes.....	153
Análisis de elementos conservados <i>PhastCons</i> en el genoma humano.....	154
Análisis de la evolución de las regiones no codificantes conservadas del gen <i>NPAS3</i>	155
Análisis de solapamiento de las cuatro bases de datos de HAEs.....	155
Producción de peces cebra transgénicos.....	155

Construcción de los transgenes para pez cebra.....	155
Transcripción <i>in vitro</i> del ARN que codifica para la transposasa <i>Tol2</i>	159
Animales y acuario.....	160
Obtención de embriones para inyección.....	160
Inyección de los embriones con la transposasa.....	161
Análisis del patrón de expresión de la proteína reportera <i>EGFP</i>	162
Generación de líneas de peces cebra transgénicos.....	163
Producción de ratones transgénicos.....	163
Construcción de transgenes para ratón.....	163
Preparación de los transgenes para microinyección pronuclear.....	165
Animales y bioterio.....	166
Producción de embriones para microinyección.....	166
Microinyección pronuclear.....	167
Transferencia de los embriones inyectados a hembras pseudo preñadas....	168
Genotipificación de los ratones transgénicos.....	168
Extracción de ADN genómico.....	168
Análisis de la expresión de <i>LacZ</i> en ratones transgénicos.....	169
Hibridación <i>In Situ</i> en cortes de crióstato.....	170
Extracción de ARN de embriones de pez cebra o cerebro de ratón.....	170
Síntesis de ADN copia.....	171
Amplificación y clonado de las secuencias molde para las sondas de <i>In Situ</i>	171
Síntesis de la sonda.....	172
Congelado y corte del tejido.....	173
Hibridación <i>In situ</i> en cortes histológicos.....	173
Hibridación <i>In situ</i> fluorescente en cortes histológicos.....	174
Hibridación <i>In situ</i> fluorescente e inmunohistoquímica en cortes histológicos de embriones de ratón.....	175
Doble inmunohistoquímica en cortes de pez cebra.....	175
Hibridación <i>In Situ</i> en embriones completos de pez cebra.....	176
Preparación del tejido.....	176
Hibridación <i>In Situ</i> en embriones completos.....	176
Cultivo primario de fibroblastos y preservación del cultivo celular.....	177
Extracción de ADN genómico de cultivos celulares.....	178
Inyección de morfolinos.....	179
Tinción de alcian blue.....	180
Tinción de hematoxilina-eosina.....	180
APÉNDICE.....	182
Análisis evolutivo en la región codificante del gen <i>NPAS3</i>	182
Identificación de sitios probables de pegado de factores de transcripción	183
BIBLIOGRAFÍA.....	186

ABREVIATURAS

2xHARs: 2x Human Accelerated Regions.

A

ADN: ácido desoxirribonucleico.
ADNc: ácido desoxirribonucleico copia.
AhR: aryl hydrocarbon receptor.
AKAP6: A-kinase anchor protein 6.
ARN: ácido ribonucleico.
ARNm: ácido ribonucleico mensajero.
ARNPolIII: ARN polimerasa tipo II.
ARNT: aryl hydrocarbon receptor nuclear translocator.
ASPM: abnormal spindle-like microcephaly-associated.
ATP: adenosine trifosfato.
APOB: apolipoprotein B.

B

BCL6: B-Cell Lymphoma 6.
bHLH-PAS: basic helix-loop-helix, period/aryl.
BMP4: bone morphogenetic protein 4.
BNC2: basonuclin 2.
Bnl: Branchless.
Btl: Breathless.

C

CBP: CREB-binding protein.
CDK5RAP2: CDK5 regulatory-subunit-associated protein 2.
CENPJ: centromeric protein J.
CMAH: cytidine monophospho-N-acetylneuraminic acid hydroxyase.
CNSs: conserved noncoding sequences.
CNTNAP2: contactin-associated protein-like 2.

D

DACH: dachshund homolog 1.
DCX: doublecortin.
dNTP: deoxynucleotide triphosphates.
DNase I: Deoxyribonuclease I.
Dpc: días pos coitum.

E

EGFP: enhanced green fluorescent protein.
EGLN3 Egl Nine Homolog 3
ET: elemento transponible.

F

FGF: Fibroblast growth factor
Fgf8: Fibroblast growth factor 8.
FGFR: Fibroblast growth factor receptor
FHIT: fragile histidine triad proteína.

G

gBGC: *GC-biased gene conversion*

GCN5: *general control of amino-acid synthesis protein-like 5*

H

H3K4Me1: monometilación de la histona H3 en su lisina 4.

H3K4Me3: trimetilación de la histona H3 en su lisina 4.

HACNS: *human accelerated conserved non-coding sequence*.

HAE: *human accelerated elements*.

HIF: *hypoxia-inducible factor*.

HAR: *human accelerated regions*.

HAT: histonas acetil transferasa.

hGC: *human chorionic gonadotrophin*.

Hpf: horas post fecundación.

HTBE: *human terminal branch elements*.

Hs: *Homo sapiens*.

HS6ST3: *heparan sulfate 6-O-sulfotransferase 3*.

HSP68: *Heat shock protein 68*.

I

ISH: *in situ hybridization*.

IHQ: inmunohistoquímica.

K

Kpb: kilo pares de bases.

L

LMBR1: *limb region 1*.

LINs: *long interspersed nucleotide element*.

LTR: *long terminal repeat*.

M

M: molar

Mb: mega pares de bases.

MCPH: *microcephaly primary autosomal recessive*.

MIR: *mammalian interspersed repetitive*.

Mm: *Mus musculus*.

MO: morfolino.

MYH16: *myosin heavy chain 16*.

N

NPAS3: *neuronal PAS domain-containing protein 3*.

NTM: *neurotrimin*.

P

Pb: pares de bases.

PBS: *phosphate buffered saline*.

PCAF: *P300/CBP-associated factor*.

PCR: *Polymerase Chain Reaction*.

PMS: *pregnant mare serums*

PNI: precursor neural indiferenciado.

POMC: Pro-opiomelanocortin.

Pt: *Pan troglodytes*

R

RBFOX1: RNA binding protein, fox-1 homolog.

RefSeq: Reference Sequence.

RT-PCR: reverse-transcription PCR.

S

Shh: sonic hedgehog.

SIM: single minded. *SINE*: short interspersed nucleotide element.

S.I.: sin inyectar.

S.MO: sin morfolino.

SNC: sistema nervioso central.

SOX2: Sex determining Region Y-box 2.

Spry: Sprouty.

SRGAP2: Slit-Robo Rho GTPase-activating protein 2.

STAT5: signal transducer and activator of transcription 5.

T

TA: temperature ambiente.

TADs: transcriptional associates domains

TFBS: transcription factor binding site.

Tgo: Tango.

Trh: Trachealess.

TRRAP: transformation/transcription domain-associated protein.

Tuj: β -tubulina III.

U

UCSC: University of California, Santa Cruz.

UI: unidades de inyección.

UT: unidad transcripcional.

UCNE: ultraconserved non-coding elements.

W

Wt: wild type.

WWOX: WW domain-containing protein.

“En 1758, Linneo se enfrentó con la difícil decisión de cómo clasificar a su propia especie en la edición definitiva de su Systema Naturae. ¿Debía limitarse a incluir a Homo sapiens entre todos los demás animales o debía acaso crear para nosotros una condición separada? Linneo buscó una solución intermedia. Nos situó dentro de su clasificación (cerca de los monos y los murciélagos), pero nos dejó aparte por medio de su descripción. Definió a nuestros parientes por caracteres mundanos y distintivos de tamaño, forma y número de dedos de las manos y de los pies. Para Homo sapiens escribió tan sólo el imperativo socrático: «Nosce te ipsum», «conócete a ti mismo». Para Linneo, Homo sapiens era, a la vez, especial y no especial.”

SJ Gould. Desde Darwin. 2010.

INTRODUCCIÓN GENERAL

Conócete a ti mismo

Todos los organismos están situados en algún lugar del patrón de formas y entidades dejadas por los procesos del cambio evolutivo que tuvieron lugar desde el origen común de la vida, al menos tal como hoy existe en la Tierra (Mayr 2001). Este es uno de los aportes más trascendentes de la teoría evolutiva al brindar una idea de continuidad entre las formas de vida (Gould 2006; Darwin 2007). En palabras de S.J. Gould:

”Sabemos que la evolución debe ser la razón fundamental del orden de la vida porque no hay otra explicación que pueda coordinar los datos dispares de la embriología, la biogeografía, el registro fósil, los órganos vestigiales, las relaciones taxonómicas, etc.”

SJ Gould. La Vida Maravillosa. 2006.

Dentro de este enorme árbol de la vida, existe una rama que siempre despertó un interés particular, agitando todo tipo de pasiones: el origen de nuestra propia especie. Los debates que llevan a situar a nuestra especie como una rama más del árbol de la vida fueron largos y profundos, debido, fundamentalmente, a una resistencia a la idea de continuidad entre nuestra especie y las demás formas de vida (Foley 2004). Muestra de ello, por ejemplo, a lo largo de la historia del estudio de la evolución humana se propusieron muchas características como únicas de nuestra especie, pero la gran mayoría de ellas resultaron ser menos únicas de lo sugerido en un principio (Striedter 2005). Entre ellas, la noción que solo los humanos pueden recordar episodios específicos de sus vidas se puso en duda cuando se descubrió que ciertos pájaros que esconden comida pueden recordar no solo donde y qué escondieron sino también cuando (Clayton y Dickinson 1998; Clayton, Bussey y Dickinson 2003). De forma similar, la idea del hombre como el “fabricante de herramientas” fue cuestionada al encontrarse que los chimpancé también fabrican y usan herramientas (Weir et al

2002). Estos hallazgos llevaron a plantear que no existe una discontinuidad cualitativa entre los humanos y el resto de los animales (Cartmill 1990). Sin embargo, al observar algunas de las características de nuestra especie (como ser el lenguaje complejo, la cognición avanzada o la conciencia) se fomenta la noción que los humanos son, de alguna forma, “especiales” (Striedter 2005). Es una paradoja que yace en el corazón del estudio de la evolución humana: como simultáneamente reconciliar esta idea de brecha que separa a los humanos del resto de los primates y la continuidad inherente del proceso evolutivo (Foley 2004; Gould 2004).

Es cierto que los sucesos que dieron origen a nuestra especie fueron únicos e irrepetibles. Siguiendo la historia evolutiva de nuestro linaje se encontrará un conjunto de acontecimientos singulares y de novedades exclusivas que hacen a nuestra especie. Pero, todo evento de especiación es, en esencia, la suma de eventos únicos y singulares, todos ellos enmarcados dentro de los procesos de cambio evolutivo (Foley 2004). Son todas aquellas características que hacen de las especies, linajes evolutivos independientes. En este sentido, los humanos no son más especiales que ningún otro grupo de organismos. La configuración de las características de nuestra especie (más allá de lo sorprendentes o complejas que estas puedan ser consideradas) se forjó bajo los mismos principios evolutivos que subyacen la dinámica de la vida sobre nuestro planeta (Foley 2004). Desconocer esto ha ido de la mano con interpretaciones teleológicas del origen del hombre (Gould 2004). Dejando de lado el debate sobre la dualidad del término “único” aplicado a la trayectoria evolutiva de *Homo sapiens*, no se puede dejar de reconocer que los humanos presentan ciertos rasgos particulares. Estas características podrán ser minoritarias dentro del conjunto de características que hacen, por ejemplo, de nuestra especie un primate, un mamífero, un cordado y un eucariota, etc. (sino ¡qué poco se podría decir de nosotros mismos al estudiar como modelo a otros animales!). Pero el impacto que tienen sobre la forma de entender y relacionarnos con el mundo que nos rodea tan solo un ápice de nuestras características humanas es de tal magnitud que ha llevado a invertir mucho esfuerzo para tratar de develar sus orígenes (Gould 2010b).

Por otro lado, el estudio de la evolución humana atraviesa, tal vez como pocas áreas del conocimiento, una multiplicidad de facetas de nuestra sociedad y encara un desafío fascinante y complejo: comprender nuestro propio origen. Los aspectos de la sociedad que pueden verse transformados por los descubrimientos que puedan provenir de este campo son múltiples. Claramente, hay un interés académico por develar los procesos subyacentes en la evolución humana. Pero también hay un interés médico en responder cuales fueron los cambios genéticos y fisiológicos que ocurrieron y cómo esos cambios impactan en la biología actual de nuestra especie. Por ejemplo, la identificación de genes responsables de nuestras capacidades cognitivas complejas

podrá ser crucial para comprender alteraciones moleculares vinculadas con el desarrollo de enfermedades neurológicas y psiquiátricas tales como esquizofrenia, dislexia, autismo y déficit de atención con hiperactividad (ADHD)(Miller 2006; Crespi 2007; Moalic 2010). Es por ello que, avances en el descubrimiento de genes responsables de las características distintivas de nuestra especie tendrán un impacto extraordinario no solo en biología sino también en medicina (Varki 2003; Moalic 2010). Finalmente, en el plano ético, el plano político y el filosófico, comprender quienes somos y de dónde venimos impactará de manera irreversible en la forma de entender al hombre como entidad biológica y social al mismo tiempo (Lamb 2013).

Develar las características específicas de los humanos y cómo estas surgieron implica interactuar con los diferentes niveles de complejidad que comprenden estos cambios: desde las diferencias genómicas, las redes genéticas que controlan el desarrollo, la fisiología, el comportamiento, la estructura social y hasta la forma en que los humanos establecen relaciones con el resto de las especies. Es por eso que, solo será posible entender de forma íntegra cómo surgieron los rasgos específicos de los humanos integrando la información proveniente de campos tan diversos como la genética, la paleontología y la antropología, entre otras disciplinas; recordando que inevitablemente se construirá una narración de la evolución del hombre desde una mirada humana.

La herencia primate

Homo sapiens es una de las aproximadamente 200 especies vivientes de primates que colectivamente constituyen el orden Primates (Purvis 1995). Este grupo se originó hace alrededor de 65 millones de años atrás (Tavare et al. 2002). Los primates se clasifican en dos grandes subórdenes: Strepsirrhini y Haplorrhini. Los strepsirrhinos se conocen comúnmente como prosimios e incluyen a los lemures, lorinos y gálagos. Por otro lado, los haplorrhinos consisten en los tarseros y los antropoideos. Los antropoideos comprenden a los platirrinios (o monos del nuevo mundo, como el mono araña y el mono tití) y los catarrhinos. Dentro de este último grupo se agrupan los cercopitécidos (o monos del viejo mundo como los macacos, los colobos y los babuinos) y los grandes simios u hominoideos (gibones, orangutanes, gorilas, chimpancés y humanos) (Purvis 1995; Begun 2003) (figura 1).

Tanto los primates actuales como los fósiles constituyen un conjunto diverso de especies. Algunas están entre las especies que retuvieron más características ancestrales dentro de la clase de los mamíferos; otras muestran especializaciones morfológicas y de comportamiento completamente novedosas dentro de los mamíferos (Martin 1986; Foley 2004; Fleagle 2013). Históricamente, se han estudiado las características de los primates tratando de identificar las autapomorfias de los

humanos. Sin embargo, muchas características a veces consideradas específicas de los humanos como ser la capacidad de caminar erguidos, gran inteligencia y una gran sociabilidad, más que representar una discontinuidad con el resto de primates, son una extensión de las características de estos como grupo (Martin 1990).

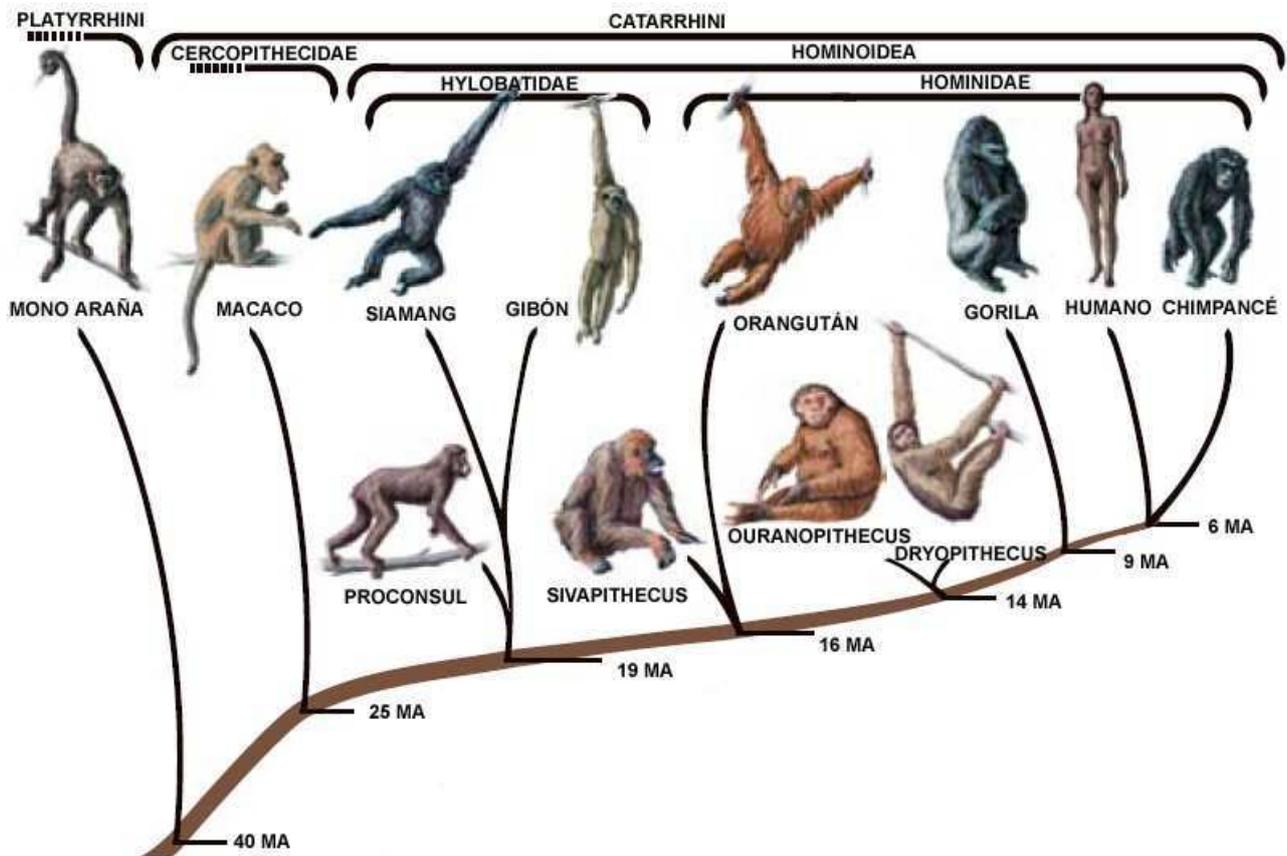


Figura 1: Árbol filogenético que muestra las relaciones de parentesco probables entre distintos grupos de primates actuales y extintos. Abreviatura: MA, millones de años atrás. Modificado de (Begun 2003).

Tal vez sea el cerebro el órgano más emblemático de *Homo sapiens*, sede principal de gran parte de las capacidades más destacadas de los humanos. Sin embargo, muchas de estas características se encuentran en mayor o menor medida en los cerebros de todos los primates. Por un lado, comparando con el conjunto de los mamíferos, los primates (particularmente los diurnos) se caracterizan por presentar un sentido de la visión muy desarrollado, mientras que el sentido del olfato está más reducido (Fleagle 2013). Esto está acompañado por una tendencia evolutiva dentro de este clado a una mayor diurnalidad (Striedter 2005) e incluso en los catarrinos, existió una reinención de la visión tricrómica (Jacobs 1993; Dulai et al. 1999). Posiblemente como producto de la predominancia del sentido de la visión dentro del grupo, los cerebros de los primates son relativamente más grandes a los encontrados en otros mamíferos de peso similar (Stephan, Frahm y Baron 1981; Baron, Frahm y Stephan 1988; Striedter 2005). Esta característica se pone

particularmente de manifiesto al comparar las reglas de escalado específicas de clado entre, por ejemplo, roedores y primates de tamaños corporales similares. Por un lado, los cerebros primates tienen una mayor densidad de neuronas en sus cerebros que los cerebros de roedores de tamaño similar (por ejemplo 40.000 neuronas/mg en la corteza del mono aotus *Aotus trivirgatus* contra 12.000 neuronas/mg en el roedor agutí *Dasyprocta primnolopha*) mientras que la densidad de células no neuronales en los cerebros de estos dos grupos es comparable (Herculano-Houzel, Mota y Lent 2006; Herculano-Houzel et al. 2007). El mayor número de neuronas por unidad de volumen presumiblemente podría conferir a los primates con mayores capacidades computacionales que roedores de tamaños equivalentes (Herculano-Houzel 2011). De hecho, los primates no humanos son extraordinariamente inteligentes. En la naturaleza, los individuos dedican mucho tiempo a establecer una red de amistades y observar las alianzas de sus congéneres. Este tipo de interacciones sociales son muy complejas así como las habilidades cognitivas necesarias para procesar este tipo de información (Foley 2004). De hecho, la evolución progresiva de la inteligencia como forma de vida ha sido considerada como una de las características más destacadas de los primates como grupo (Jolly 1985).

Un aspecto importante de todos los cerebros primates es que tienen cortezas cerebrales inusualmente grandes para el tamaño absoluto del cerebro. Por ejemplo, la corteza cerebral de los humanos es la más grande en tamaño relativo (con un 75,5% - 84,0% de la masa total del cerebro) (Frahm, Stephan y Stephan 1982; Hofman 1988; Rilling y Insel 1999). Pero, otros primates e incluso, otros mamíferos no están muy atrás. La corteza cerebral representa el 73,0% de la masa total del cerebro en los chimpancés (Stephan, Frahm y Baron 1981), 74,5% en el caballo y 73,4% en las ballenas (Hofman 1985). Además, con un promedio de 86 mil millones de neuronas y 85 mil millones de células no neuronales (Azevedo et al. 2009), el cerebro humano tiene tantas neuronas como se espera para un cerebro primate genérico de su tamaño, ajustándose a lo predicho por las reglas de escalado del cerebro primate (Gabi et al. 2010; Herculano-Houzel 2012a). También las densidades neuronales en la corteza y el cerebelo se ajustan a los valores esperables para un cerebro primate (Azevedo et al. 2009; Herculano-Houzel 2012a). En resumen, el cerebro humano se ajusta a un cerebro primate típico, en lo que respecta a número de células cerebrales totales, distribución de neuronas entre la corteza y el cerebelo y proporción de células neuronales y no neuronales, al menos (Herculano-Houzel 2012).

Dentro de los antropoideos, los cerebros de los chimpancés, gorilas y orangutanes tienen menor número neto de neuronas en sus cerebros, con 28, 33 y 28 mil millones de neuronas respectivamente (Herculano-Houzel y Kaas 2011), en comparación con las halladas en el cerebro

humano; aún cuando los gorilas y los orangutanes tienen tamaños corporales similares o incluso más grandes que el humano (Marino 1998). Aun así, los cerebros de estos grandes simios tienen tantas neuronas como las esperadas para un cerebro primate genérico de su tamaño, satisfaciendo nuevamente las reglas de escalado de los cerebros primates (Herculano-Houzel y Kaas 2011). Por otro lado, se encontró que la relación tamaño del cuerpo-tamaño del cerebro se mantuvo constante dentro de la mayor parte de los grupos de primates (incluyendo a los humanos), donde el cerebro representa aproximadamente el 2% de la masa total del cuerpo (Marino 1998; Azevedo et al. 2009). Sin embargo, dentro de los grandes simios, particularmente gorilas y orangutanes, se incrementó no isométricamente el tamaño del cuerpo respecto del tamaño del cerebro llegando a representar el 0,5% y 1% del peso total del cuerpo respectivamente (Marino 1998). Sin embargo, esta tendencia al incremento del tamaño del cuerpo frente al tamaño del cerebro no se observa dentro del linaje de los humanos y chimpancés. Es por ello que la relación tamaño del cerebro-tamaño del cuerpo en los humanos y chimpancés es similar a la hallada en grupos que divergieron tempranamente en el linaje de los primates y antes de la divergencia de los grandes simios (Herculano-Houzel y Kaas 2011). Estos resultados son consistentes con la hipótesis que sugiere que el tamaño corporal y el tamaño del cerebro estuvieron sujetos a diferentes presiones selectivas dentro de los primates (Montgomery et al. 2010).

Estas evidencias sugieren que la noción que los humanos modernos desarrollaron cerebros excesivamente grandes para su tamaño corporal sería incorrecta (Marino 1998). En su lugar, mientras que la regla de escalado de los cerebros primates en términos de números de neuronas respecto del tamaño del cuerpo estaría ampliamente conservada dentro de este orden, ciertos linajes de grandes simios (particularmente gorilas y orangutanes) se apartan de esta tendencia general del grupo, habiendo desarrollado tamaños corporales excesivamente grandes para su masa cerebral (Herculano-Houzel y Kaas 2011). Entonces, las capacidades cognitivas sofisticadas halladas en los humanos podrían haber surgido simplemente del hecho que los cerebros humanos tienen un número neto más alto de neuronas frente a los cerebros de otros primates y/o mamíferos. Esto pudo haber sido posible, al menos en parte, gracias las altas densidades de neuronas que contienen los cerebros primates; densidades probablemente superiores a las halladas en otros grupos de mamíferos (como roedores, cetáceos y/o elefantes) (Herculano-Houzel 2009; Herculano-Houzel 2012a; Herculano-Houzel 2012b). En concordancia con esta hipótesis, se observó que las habilidades cognitivas de los primates no humanos están correlacionadas directamente con el tamaño absoluto del cerebro y no con el tamaño relativo (Deaner et al. 2007).

Por otro lado, si se compara el cerebro primate con el de las especies no primates más estrechamente relacionada también se observan varias diferencias estructurales y de organización. La mayoría de estas diferencias están relacionadas con el desarrollo del sistema visual (Striedter 2005). Por ejemplo, las neuronas de cada retina de los primates proyectan bilateralmente al cerebro medio (colículo superior). Estos, junto con la localización de los ojos en posición más frontal, pudieron haber favorecido la visión en profundidad de los primeros primates (Striedter 2005). Además, se observa una expansión de la corteza somatosensorial dando lugar a nuevas áreas que no tienen homología obvia en otros mamíferos no primates y aparición de ciertas especializaciones en el sistema motor. Estos cambios pudieron estar relacionados con un incremento de la coordinación motora (Striedter 2005).

Historia evolutiva de nuestro linaje

Los procesos evolutivos responsables del patrón de cambio a lo largo del linaje humano son materia de debate. Seguramente, la gran mayoría de los rasgos específicos de nuestro linaje hayan sido el resultado de las presiones selectivas sobre nuestro fenotipo y la historia demográfica de nuestro linaje ocurridos desde la divergencia del linaje *Pan* (O'Bleness et al. 2012). Sin embargo, no se puede excluir a los eventos contingentes como fuente de una, aun no determinada, cantidad de características hoy observables (Gould 2006). El escenario evolutivo más aceptado para el origen de nuestro linaje sugiere que este ocurrió en África, en un ambiente marcado por frecuentes cambios climáticos (Foley 2004).

Al recompilar la información suministrada por el registro fósil actualmente disponible surgen dos patrones importantes en la historia evolutiva de nuestro linaje. Por un lado, si bien en este momento *Homo sapiens* es el único representante vivo de nuestro linaje, desde el comienzo de la divergencia con *Pan*, existió una gran diversidad de formas, muchas de las cuales coexistieron durante largos períodos de tiempo (Foley 2004). Sin embargo, de estos otros homínidos (subfamilia Homininae, nombre dado a todos los representantes de nuestro linaje directo) (Carroll 2003; Foley 2004) no quedan representantes vivos en la actualidad. Esta gran diversidad de formas morfológicas halladas en el registro fósil de nuestro linaje fue interpretada como múltiples especies, muchas de las cuales coexistieron, y le imprime al árbol filogenético de nuestro linaje el aspecto de un arbusto (Wood 1996; Tattersall 2000). Sin embargo, esta interpretación arbustiva de la evolución humana fue desafiada recientemente (Lordkipanidze et al. 2013). Evidencia fósil hallada en el sitio arqueológico Dmanisi (Georgia) sugiere que existió una gran diversidad morfológica dentro y entre grupos de homínidos (principalmente entre los primeros integrantes del género *Homo*). Estos resultados permitirían sostener la hipótesis de existencia de un único linaje evolutivo con continuidad

filogeográfica entre continentes (Lordkipanidze et al. 2013). Por otro lado, las características morfológicas que diferencian a *H. sapiens* de los chimpancés u otros primates no se desarrollaron simultáneamente en algún momento durante la historia de nuestro linaje. Por el contrario, se observa un patrón de evolución en mosaico, donde las diferentes estructuras morfológicas (como ser el sistema dentario, el sistema locomotor y/o el sistema nervioso) cambiaron a tasas diferentes y en momentos distintos (McHenry 1975).

Las estimaciones hechas en base a datos biogeográficos, la sistemática de los primates actuales y los aportes de la genética molecular, sostienen que los primeros homínidos pudieron haber divergido de los primeros representantes del linaje de los chimpancés en algún momento durante el Mioceno tardío o Plioceno temprano (esto es, entre 8 y 4 millones de años atrás) en África; muy probablemente, en la parte más central de África (Leakey et al. 1998; Brunet et al. 2002; Foley 2004; Kumar et al. 2005). El aspecto más importante de la divergencia de los homínidos tempranos, de los cuales el grupo más conocido son los australopitecinos, era que eran bípedos facultativos (White 1980; Aiello 1994; Striedter 2005). En un ejemplo de evolución en mosaico, *Australopithecus afarensis*, una de las especies de australopitecinos más estudiada, parece tener características esencialmente simiescas del cuello para arriba (volumen endocraneal similar al de los chimpancés modernos, prognatismo pronunciado y caninos relativamente grandes con presencia de diastema) junto con otras características postcraniales que lo unen al linaje de los homínidos (por ejemplo, no poseían pulgar del pie oponible) (Johanson y White 1979; Wood 1996).

Hace aproximadamente 2,5 millones de años atrás existió una radiación importante de un grupo de australopitecinos. Esta involucró el surgimiento de homínidos con una mandíbula más pequeña y una reducción en el tamaño de las piezas dentarias (Foley 2004). Además, este grupo mostró un incremento significativo del tamaño del endocráneo, un pequeño aumento del tamaño corporal y presencia de bipedalismo completo (Leakey, Tobias y Napier 1964). Por primera vez, herramientas simples de piedra pudieron ser definitivamente asociadas a grupos de homínidos (Leakey, Tobias y Napier 1964). Era el surgimiento de los primeros representantes del género *Homo*. De hecho, la presencia de tecnología lítica es lo que definió históricamente a este nuevo género (Wood y Collard 1999). Este género estuvo compuesto por miembros bastante diversos. Algunos grupos de *Homo* tempranos, como *Homo rudolfensis* y *H. habilis*, eran relativamente similares a los australopitecinos gráciles (Leakey, Tobias y Napier 1964) por lo que su inclusión actual dentro del género *Homo* resulta controversial (Wood 1996). Sin embargo, hace aproximadamente 2 millones de años atrás, grupos de *Homo* tempranos dieron origen a otra especie de *Homo*, de cuerpo y cerebro más grande en África: *H. ergaster* (Wood 1996; Lewin 1997). De acuerdo con el registro fósil, *H. ergaster*

producía una variedad de herramientas de piedra que utilizaba para cortar carne y extraer la médula ósea de los huesos. Además, este homínido probablemente aprendió a controlar el fuego y fue capaz de realizar artefactos para contener agua y/o alimentos. Equipado con estas nuevas habilidades, *H. ergaster* fue el primer homínido en migrar fuera de África alrededor de 1.8 millones de años atrás hacia Asia e Indonesia. Ya fuera de África, *H. ergaster* dio origen a la especie que hoy reconocemos como *H. erectus*. (Wood 1996; Foley 2004; Striedter 2005).

Existen varias hipótesis en torno del origen de los humanos modernos (Foley 2004). En un extremo, una hipótesis sostiene que la transformación ocurrió gradualmente dentro de todas las poblaciones de *H. erectus*, donde fuera que estas estuvieran. Esto condujo al surgimiento de múltiples poblaciones de humanos modernos en África y Eurasia. Bajo esta mirada, las raíces genéticas de las poblaciones modernas de *H. sapiens* son profundas y se encuentran en el origen de las primeras poblaciones de *H. erectus* (hace aproximadamente 2 millones de años atrás) (Thorne y Wolpoff 1992; Wolpoff 1999). En el otro extremo, una hipótesis alternativa plantea un origen único para *H. sapiens* (en África) seguido por una expansión poblacional hacia el resto del viejo mundo con reemplazo de las poblaciones de humanos arcaicos (*H. ergaster*, *H. erectus*, *H. heidelbergensis* y *H. neardenthalensis*) previamente establecidas en estos territorios. En este escenario, las raíces genéticas de las poblaciones modernas de *H. sapiens* son recientes y datan de aproximadamente 200.000 - 100.000 años (Cann, Stoneking y Wilson 1987; Stringer y Andrews 1988; Wilson y Cann 1992; Templeton 2002). Es importante notar que, no hay un consenso claro respecto de cuantas especies del género *Homo* existieron (Thorne y Wolpoff 1992; Wood 1996) y que reconocer poblaciones humanas como especies distintas o no depende de la interpretación que se esté considerando. Obviamente, estas son dos visiones extremas y existen posiciones intermedias a ambas teorías (Foley 2004). Cualquiera sea el caso, la evidencia muestra que hace aproximadamente 250.000-200.000 años atrás, *H. erectus* desapareció y apareció un nuevo grupo de homínido con características consideradas anatómicamente modernas (como ser un cráneo redondeado, mayor gracilidad y presencia de mentón, por mencionar algunas) (Lahr y Foley 1998). Este grupo somos nosotros, *H. sapiens* (Striedter 2005).

Por otro lado, respecto a la evolución del cerebro en el linaje de los homínidos (al menos en lo que a tamaño respecta), se observa que el patrón de cambio no fue gradual, sino que hubo algunos momentos de aumento abrupto (estimado a partir del volumen del endocráneo) (Striedter 2005). Por un lado, los australopitecinos tempranos tenían un volumen del endocráneo de aproximadamente 400 cm³, un volumen similar al hallado en los chimpancés y gorilas actuales. Considerando que el tamaño del cuerpo de los australopitecinos era más pequeño que el de los chimpancés y gorilas

modernos, esto implica que el surgimiento de nuestro linaje pudo estar acompañado por un incremento modesto (de aproximadamente 30%) del tamaño relativo del cerebro (Foley 2004; Striedter 2005). Sin embargo, el surgimiento del género *Homo*, hace aproximadamente 2 - 2,5 millones de años atrás, estuvo acompañado por un incremento abrupto del tamaño de cerebro. Por ejemplo, los rangos de volumen del endocráneo de *H. habilis* y *H. rudolfensis* se estiman entre los 650 - 800 cm³ y el de *H. ergaster/H. erectus* en 850 - 1100 cm³ (McHenry 1982; Hawks et al. 2000). Este dramático evento de expansión del cerebro en el género *Homo* involucró también un aumento del tamaño corporal (Striedter 2005). Finalmente, la aparición de *H. sapiens* (y posiblemente otros humanos arcaicos) está asociada con un tercer evento de incremento en el tamaño del cerebro, alcanzando un volumen de entre 1100 cm³ a más de 1400 cm³. En este caso, la expansión no estuvo acompañada por un aumento del tamaño corporal significativo (Hofman 1983; Striedter 2005). Recientemente, y tomando como hipótesis que las reglas de escalado de los cerebros primates están conservadas, se estimaron los números de neuronas de diferentes especies de homínidos (Herculano-Houzel y Kaas 2011). Estas estimaciones 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 (excluidos los humanos). Por otro lado, para los homos tempranos, se predice un número de neuronas de 50 a 60 mil millones, y de 62 mil millones en promedio para *H. erectus*. Llamativamente, para las especies de humanos arcaicos (*H. heidelbergensis* y *H. neanderthalensis*) se predice un número de neuronas que caerían dentro del rango de los humanos modernos (Azevedo et al. 2009; Herculano-Houzel 2011; Herculano-Houzel y Kaas 2011). Si el factor determinante para el desarrollo de capacidades cognitivas sofisticadas depende principalmente del número de neuronas, estos resultados sugieren que el potencial cognitivo de estas especies de humanos arcaicos podrían haber sido comparables a las halladas en los humanos modernos (Herculano-Houzel y Kaas 2011).

Si bien es posible que con el surgimiento de los humanos anatómicamente modernos hace 100.000-200.000 años atrás (Tattersall 2009) se hayan sumado algunos cambios genéticos adicionales relacionados con el incremento en las capacidades cognitivas o con la aparición del lenguaje complejo, los cambios anatómicos de gran escala característicos de nuestro cerebro probablemente ya debieron haberse establecido para ese momento (Foley 2004). Entonces, tal vez, los eventos más salientes de la evolución del cerebro humano pudieron haber ocurrido en tiempos profundos, para los cuales, la detección de procesos de evolución adaptativa por detección de polimorfismos actuales no es aplicable (Vallender 2008).

Las bases genéticas de la evolución humana

Dado que no existen otros representantes vivos de nuestro género, se recurre frecuentemente a la comparación con los simios más cercanamente emparentados con nuestra especie para definir cuales son nuestras características diferenciales como especie. Esto es particularmente cierto en los estudios que se basan en la comparación de secuencias genómicas (O'Bleness 2012). Como se mencionó anteriormente, para entender cómo surgieron estas características diferenciales, se pueden encarar estudios desde perspectivas muy diversas, todas ellas capaces de brindar información valiosa. Situándose a nivel genético/molecular, una pregunta fundamental del campo y que se remonta al descubrimiento del ADN como sustrato de la herencia es: ¿Cuáles son las secuencias genómicas que sustentaron la trayectoria evolutiva única tomada por nuestra especie?

El estudio de la evolución humana desde la perspectiva de la genética molecular procura hallar los cambios a nivel genómico y genético que potencialmente pudieron haber contribuido con la adquisición de las características particulares de nuestra especie. La tarea comprende desde el desarrollo de herramientas bioinformáticas para la descripción de los patrones de cambio evolutivo a nivel genómico, la identificación de las regiones candidatas, la predicción de cómo esos cambios pudieron impactar a nivel molecular con la estructura y actividad de una determinada región de ADN y los ensayos funcionales necesarios para sostener o no las hipótesis generadas. Finalmente, uno de los objetivos últimos del campo es poder correlacionar los cambios genéticos con cambios fenotípicos, mejorando así la capacidad de comprender los procesos evolutivos subyacentes y cómo esos cambios impactan en la biología actual de nuestra especie (O'Bleness 2012).

En la última década ocurrieron avances importantes en el campo gracias a una mayor disponibilidad de secuencias genómicas. La posibilidad de acceder a los genomas de múltiples especies junto con el desarrollo de un gran número de herramientas bioinformáticas para el análisis de dichas secuencias, hizo que se potenciara formidablemente la capacidad para examinar y encontrar los cambios genéticos específicos de los humanos. En la actualidad ya se dispone de los genomas secuenciados de nueve primates (incluyendo a todos los grandes simios) más de cuarenta mamíferos y otros tantos vertebrados (Lander et al. 2001; Waterston et al. 2002; The_Chimpanzee_Sequencing_and_Analysis_Consortium 2005; Scally et al. 2012). Incluso, se dispone de las secuencias genómicas provenientes de múltiples individuos humanos y datos genómicos de dos linajes de homínidos ancestrales: los Neandertales y los Denisovanos, un grupo de humanos arcaicos relacionados con los primeros (Noonan et al. 2006; Abecasis et al. 2010; Reich et al. 2010; Meyer et al. 2012).

Más allá de todas las dificultades que todavía existen (como puede ser hallar modelos animales adecuados para poner a prueba las hipótesis, mejores predicciones bioinformáticas, limitado conocimiento de cómo se codifica información compleja en el ADN, etc.), diversos grupos de investigación se embarcaron en la formidable tarea de estudiar diferentes cambios genéticos que pudieron haber contribuido con la adquisición de las características particulares de nuestro linaje. Estos trabajos son fuente de valiosa información y ponen de manifiesto las complejidades y limitaciones con las cuales se enfrenta esta área de estudio.

Los trabajos citados a continuación son solo una fracción de los hallazgos realizados hasta el momento, pero ponen de manifiesto los diferentes abordajes encarados, algunos de los avances obtenidos y los alcances del campo. Dependiendo de cómo y donde fue el cambio, se encontraron:

1) Cambios genómicos de gran escala: estos cambios involucran modificaciones estructurales en los cromosomas y fueron los primeros en ser estudiados (O'Bleness 2012). Desde mediados de los años 70 y haciendo uso de técnicas de tinción de bandas de cromatina, se descubrió que tuvo lugar un cambio en el número haploide de cromosomas entre los humanos y los primates más cercanamente emparentados. Hubo una reducción en el número haploide de 24 a 23 por fusión de dos cromosomas ancestrales dando lugar al cromosoma humano 2 (Yunis 1980). Recientemente, se estimó el momento en que ocurrió esta fusión en 4 a 5 millones de años atrás (Ventura et al. 2012). También se describieron bandas de heterocromatina constitutiva específicas de los humanos en los cromosomas 1, 9, 16 e Y e inversiones pericéntricas en los cromosomas 1 y 18 (Yunis y Prakash 1982). Sin embargo, al momento no se pudo asociar conclusivamente ninguno de estos eventos citológicos con características fenotípicas específicas de los humanos (O'Bleness et al. 2012).

2) Cambios genéticos pequeños: estos cambios involucran eventos que ocurrieron en regiones del genoma de menos de 20 kpb y que no pueden ser identificados por métodos citológicos convencionales (O'Bleness et al. 2012). Entre estos cambios se encuentran las sustituciones nucleotídicas, las inserciones, las deleciones y el cambio en el número de copias de genes. A su vez, estas diferencias pueden afectar regiones codificantes, regiones no codificantes regulatorias o el contenido de regiones repetitivas (O'Bleness et al. 2012). Para poner en contexto lo que implica analizar este tipo de variaciones, la divergencia nucleotídica entre los genomas de los chimpancés y los humanos debido a sustituciones nucleotídicas es aproximadamente del 1,2% (alrededor de 18 millones de cambios nucleotídicos) (Consortium 2005). Sin embargo, este resultado es una subestimación dado que los cálculos se realizan exclusivamente en aquellas regiones donde es posible hacer alineamientos genómicos confiables y se excluyen regiones altamente repetitivas, o

regiones donde hay variaciones en el número de copias de genes (Consortium 2005). Los cambios genéticos pequeños pueden tener diferentes consecuencias funcionales dependiendo de dónde y cómo se adquirieron estas diferencias. Se encontraron:

a) Cambios en el número de copias de un gen: existe evidencia creciente que la aparición de nuevos genes pudo haber contribuido con la evolución fenotípica humana (Lander et al. 2001; Samonte y Eichler 2002). La introducción de nuevos genes en los genomas ocurre, por lo general, gracias a un evento de duplicación, habitualmente dentro de familias génicas que presentan predisposición para tales eventos. En estos casos, es factible que ocurra una relajación de las restricciones funcionales sobre las variantes duplicadas. Esto a su vez puede favorecer eventos de neofuncionalización o subfuncionalización (Ohno 1970). La neofuncionalización consiste en la adquisición de una nueva función por parte de una o ambas copias del gen recientemente duplicado. Por otro lado, la subfuncionalización implica la segmentación o partición de la función ancestral entre los parálogos (Vallender, Mekel-Bobrov y Lahn 2008). Utilizando ADN de cinco especies de grandes simios (incluido *Homo sapiens*), se identificaron 140 genes con cambios en el número de copias en humanos en comparación con estas especies (134 genes con ganancia de copias y 6 con pérdidas). La mayoría de estos cambios ocurrieron en genes que se encuentran en regiones pericentroméricas, subteloméricas o regiones ricas en regiones repetitivas. Sin embargo, aún no se logró establecer cual fue el impacto, si alguno, de estos eventos de variación en el número de copias de genes en la biología de nuestra especie (Fortna et al. 2004).

Por otro lado, un estudio genómico posterior a gran escala identificó un caso extremo de variación en el número de copias entre humanos y otros grandes simios. Este evento involucró la amplificación del dominio proteico DUFF1220. Mientras que se encontraron más de 200 copias de este dominio distribuidas a lo largo del genoma humano, solo se encontraron 34, 30 y 1 copias en los genomas del chimpancé, mono rhesus y ratón respectivamente. Adicionalmente, se encontró evidencia que estos dominios pudieron estar sujetos a selección positiva reciente. Además, las proteínas que contienen dominios DUF1220 se expresan exclusivamente en neuronas en humanos (Popesco et al. 2006). Paralelamente, se describió que diferentes deleciones en el *locus* 1q21.1 (altamente enriquecido en dominios DUF1220), correlacionaban con microcefalia y/o esquizofrenia, y que duplicaciones en esta región estaban asociadas con macrocefalia y/o autismo en humanos. En estos casos, las deleciones y duplicaciones contenían y/o flanqueaban dominios DUF1220. Por lo tanto, duplicaciones en el dominio DUF1220 (a nivel genómico) podrían estar vinculadas con el incremento del tamaño del cerebro experimentado en el linaje humano, y

variaciones en el número de copias de este dominio en el *locus* 1q21.1 podrían estar involucradas en diferentes patologías neurológicas (Dumas y Sikela 2009).

Recientemente se identificó otro evento de duplicación génica que pudo haber sido importante como fuente de variación fenotípica en el linaje humano. *Slit-Robo Rho GTPase-activating protein 2 (SRGAP2)* es un gen altamente conservado entre los mamíferos que participa del desarrollo de la corteza regulando la diferenciación y la migración neuronal. Se encontró que ocurrieron tres eventos de duplicación génica de este gen exclusivamente en el linaje humano. La primera duplicación de la copia ancestral (llamada *SRGAP2A* en humanos) fue incompleta y abarcó el promotor y los primeros nueve exones de este gen. Este evento tuvo lugar hace aproximadamente 3,4 millones de años y dio origen a *SRGAP2B*. Posteriormente, ocurrieron dos duplicaciones de *SRGAP2B* generando los genes parálogos *SRGAP2C* y *SRGAP2D* hace aproximadamente 2,4 y 1 millón de años, respectivamente. Evidencia funcional sugiere que las copias incompletas son capaces de dimerizar con la copia completa de *SRGAP2* pero carecen de otros dominios funcionales. La naturaleza incompleta de la duplicación de *SRGAP2B* sugiere un modelo de neofuncionalización instantánea donde las nuevas copias funcionan como dominantes negativos de la copia ancestral. Presumiblemente, la nueva función de las copias de *SRGAP2* pudo haber contribuido con cambios neoténicos específicos. Este ejemplo pone de manifiesto una nueva función para los parálogos de *SRGAP2* con impacto potencial en la evolución fenotípica de nuestro linaje (Dennis et al. 2012).

b) Cambios en la secuencia codificante de un gen: hoy se sabe que las mutaciones que tienen lugar en las regiones codificantes de proteínas pueden generar cambios en la secuencia de aminoácido de las mismas y ser fuente de novedades a nivel evolutivo (Ohno 1970). Una de las ventajas que posee el estudio de regiones codificantes es que existen métodos potentes para detectar indicios de procesos selectivos (Tajima 1989; Goldman y Yang 1994; Suzuki y Gojobori 1999; Nielsen y Huelsenbeck 2002; Yang, Wong y Nielsen 2005; Zhang, Nielsen y Yang 2005). Además, se cuenta con abundante información sobre las propiedades fisicoquímicas de los aminoácidos y la estructura de las proteínas. Esto hace más sencillo poner en contexto un cambio en la secuencia de una proteína (Vallender 2008).

Existen numerosas diferencias en la secuencia de proteínas entre los humanos y otros mamíferos (Kosiol et al. 2008). Comparando las secuencias de humano y chimpancé, la mayoría de las proteínas muestran al menos un aminoácido de diferencia (Consortium 2005). Es probable que gran parte de estas sustituciones tengan muy pocas consecuencias funcionales, y que por lo tanto, no

sean relevantes para la evolución fenotípica de los humanos. Sin embargo, algunos de estos cambios pudieron tener efectos funcionales importantes pudiendo constituir la base molecular de ciertos rasgos específicos de nuestra especie (Vallender, Mekel-Bobrov y Lahn 2008).

Existen varios genes candidatos relacionados al desarrollo del cerebro y sus funciones que muestran evidencias de selección positiva en humanos (Vallender 2008). Un ejemplo es el caso de los genes asociados a microcefalia primaria autosómica y recesiva (o microcefalia verdadera), una condición congénita caracterizada por una reducción severa del tamaño del cerebro que está acompañada generalmente de retraso mental (Woods, Bond y Enard 2005). Si bien la organización macroscópica de las diferentes estructuras del cerebro está conservada en los pacientes afectados por esta enfermedad, la reducción de la corteza cerebral es mucho más severa que la reducción de otras regiones del cerebro (Woods, Bond y Enard 2005). La causa genética de esta enfermedad es heterogénea y hasta el momento fueron identificadas siete regiones del genoma humano asociadas a esta condición. Estos *loci* fueron llamados *microcephaly primary autosomal recessive* 1 a 7 (o *MCPHI-7* por sus siglas en inglés) (Kaindl et al. 2010). Además, dentro de estas siete regiones se lograron identificar genes candidatos asociados a microcefalia. Entre ellos se encuentran: *microcephalin 1* (conocido como *MCPHI*) (Jackson et al. 1998; Jackson et al. 2002), *CDK5 regulatory-subunit-associated protein 2* (*CDK5RAP2*; *MCPH3*) (Bond et al. 2005), *centrosomal protein 152* (*CEP152*; *MCPH4*) (Guernsey et al. 2010), *abnormal spindle-like, microcephaly-associated* (*ASPM*; *MCPH5*) (Pattison et al. 2000; Shen et al. 2005) y *centromeric protein J* (*CENPJ*; *MCPH6*) (Leal et al. 2003; Bond et al. 2005) y *SCL/TAL1-interrupting locus* (*STIL*; *MCPH7*) (Kumar et al. 2009).

Dado que a lo largo de la evolución de nuestro linaje se observa un incremento pronunciado del volumen del cerebro (Striedter 2005), se realizaron estudios sobre el conjunto de genes asociados a microcefalia con el objetivo de detectar signos de evolución positiva que pudieran sugerir un rol de estos genes en la evolución del cerebro humano. Por un lado, el gen *ASPM* presenta signos de evolución positiva en el linaje que conduce a los humanos, incluyendo al linaje que separa a los grandes simios de los monos del nuevo mundo (Evans et al. 2004b). Por otro lado, el gen *microcephalin 1* muestra una señal muy fuerte de selección positiva en el linaje que lleva a los grandes simios (Evans et al. 2004a). Por otro lado, los genes *CDK5RAP2* y *CENPJ* muestran altas tasas de sustitución no sinónima en primates y roedores y particularmente *CDK5RAP2* muestra además altas tasas en las ramas terminales de humanos y chimpancés (Evans et al. 2005; Mekel-Bobrov et al. 2005; Evans, Vallender y Lahn 2006). Finalmente, el gen *CEP152* muestra signos de selección positiva en el linaje humano (Guernsey et al. 2010). En la actualidad se sabe que los cinco

genes antes mencionados participan del control del ciclo celular (Guernsey et al. 2010; Kaindl et al. 2010). Se especula que los cambios evolutivos experimentados por este grupo de genes pudieron haber impactado en el potencial proliferativo de los precursores neurales durante la embriogénesis humana dando como resultado un incremento en el número de divisiones celulares de los progenitores neurales y llevando a la expansión del cerebro humano y del cerebro primate (Vallender, Mekel-Bobrov y Lahn 2008). Sin embargo, la ausencia de estudios funcionales impide hasta el momento verificar esta hipótesis.

Otro caso interesante de cambios en regiones codificantes con consecuencias potenciales en la evolución de rasgos específicos en los humanos se identificó sobre la secuencia del gen *Forkhead box P2 (FOXP2)*. Por un lado, mutaciones en este gen fueron asociadas con defectos severos del habla en familias humanas (Lai et al. 2001; MacDermot et al. 2005). Adicionalmente, estudiando el patrón de cambio molecular de este gen en los primates, se identificaron dos sustituciones aminoacídicas que se fijaron en el linaje humano después de la divergencia con el linaje de los chimpancés y para las cuales se encontró evidencia de selección positiva (Enard et al. 2002; Zhang, Webb y Podlaha 2002). Estudios realizados sobre ratones transgénicos donde se introdujeron las dos sustituciones específicas de los humanos en el gen *FoxP2* endógeno de ratón mostraron que estas modificaciones produjeron alteraciones en las vocalizaciones de las crías y ocasionaron defectos neurológicos en los circuitos de los ganglios córtico-basales (Enard et al. 2009), (circuitos relevantes para el control del habla y producción del lenguaje) (Ullman 2001; Lieberman 2002). Entonces, las sustituciones que tuvieron lugar en la secuencia codificante del gen *FOXP2* en el linaje humano podrían estar vinculadas con cambios en los circuitos de los ganglios córtico-basales, que participan de la producción del lenguaje (Enard et al. 2009).

c) Pérdida de genes. No todas las alteraciones en los genes producen variantes funcionales. De hecho, los cambios estructurales producen frecuentemente variantes no funcionales. Se conoce como pseudogenización al proceso por el cual un gen previamente funcional adquiere una alteración en su secuencia que lo vuelve inactivo, pese a que la mayor parte de la estructura del gen permanece intacta y reconocible en el genoma (Jacq C 1977; Mighell et al. 2000). Hasta el momento, en el genoma humano se identificaron aproximadamente 20.000 pseudogenes (Deloukas et al. 2001; Torrents et al. 2003; Khurana et al. 2010; Pei et al. 2012). Adicionalmente, se identificaron 80 pseudogenes no procesados (es decir, que no provienen de eventos de retrotranscripción) exclusivos del genoma humano, es decir, genes que sufrieron el proceso de pseudogenización después de la divergencia con el linaje de los chimpancés. Entre estos pseudogenes, las funciones (antes de la pseudogenización) de quimiorrecepción (olfato y gusto) y

respuesta inmune están sobrerrepresentadas, sugiriendo posibles cambios fisiológicos a estos dos niveles en el linaje humano (Wang, Grus y Zhang 2006). Por otro lado, estas dos categorías funcionales contienen varias familias multigénicas conocidas por su tasa alta de evolución molecular y por presentar repertorios de genes linaje-específicos (Consortium 2005; Nei y Rooney 2005).

Si bien el rol de la pérdida de la función de la mayor parte de estas proteínas permanece sin ser estudiada, en algunos casos hay evidencia de un posible impacto funcional de este proceso en la evolución de nuestro linaje. Uno de estos ejemplos lo constituye el gen de la *cytidine monophosphate-N-acetylneuraminic acid hydroxylase (CMAH)*. Una diferencia importante entre los humanos y otros mamíferos (incluidos los chimpancés y bonobos), es que los humanos son incapaces de sintetizar el ácido *N*-glicolil-neuramínico (Neu5Gc) (Muchmore, Diaz y Varki 1998). Esta deficiencia es producto de la inactivación del gen *CMAH* por inserción de un elemento móvil de la familia Alu que corre el marco de lectura (Hayakawa et al. 2001). La inactivación de este gen en humanos causó una reconfiguración de la superficie celular por modificación de los azúcares de superficie de millones de moléculas (Brinkman-Van der Linden et al. 2000). Estudios posteriores sugieren que la inactivación del gen *CMAH* ocurrió hace aproximadamente 2,1 - 2,2 millones de años, posiblemente en el momento de la gran expansión del cerebro humano que tuvo lugar con el origen del género *Homo* (Chou et al. 2002). Es interesante especular sobre el rol de esta inactivación y la evolución del cerebro humano dado que los cerebros mamíferos contienen bajos niveles de Neu6Gc, posiblemente por baja expresión del gen *CMAH* (Kawano et al. 1995). La ausencia constitutiva de expresión de *CMAH* en los cerebros humanos, entonces, pudo haber tenido alguna ventaja frente a la regulación de la expresión de *CMAH* que presentan otros mamíferos (Chou et al. 2002).

La pseudogenización del gen de la cadena pesada de miosina 16 (*MYH16*) constituye un segundo ejemplo. Este gen se expresa en la musculatura de la mandíbula en los primates. Sin embargo, la secuencia de este gen se encuentra inactivada en los humanos por una mutación que corre el marco de lectura. Esta mutación tuvo lugar después de la divergencia del linaje de los humanos y el linaje de los chimpancés y bonobos. Se propone que la pérdida de este filamento de miosina en humanos pudo haber alterado la musculatura y morfología craneofacial (Stedman et al. 2004).

d) Cambios en las regiones regulatorias en *cis*. Una forma de cambio evolutivo involucra alteraciones en la expresión de los genes, es decir cambios que modifican el dominio espacial, temporal y/o el nivel de expresión de un determinado gen (Ohno 1970; King 1975; Carroll 2005). Las causas moleculares de estos cambios son diversos: cambios que afectan regiones promotoras,

cambios en regiones que promueven o inhiben la expresión génica e incluso cambios en elementos regulatorios de función aún no descripta. En un trabajo reciente, se encontraron 510 deleciones genómicas exclusivas de humanos, de las cuales, la gran mayoría se encuentra en regiones no codificantes altamente conservadas entre el resto de los mamíferos (McLean et al. 2011). En particular se describió la pérdida en el linaje humano de una región regulatoria cuya posible función en otros mamíferos es la de dirigir la expresión del gen que codifica para el receptor de andrógenos a las vibrisas sensoriales y a las espinas de queratina del pene. Llamativamente, tanto las vibrisas sensoriales como las espinas de keratina en el pene están ausentes en los humanos. Se especula que la ausencia de esta región regulatoria en humanos pudo haber contribuido con la adquisición de estas características en nuestro linaje al modificar la sensibilidad a estas hormonas sexuales en estos tejidos (McLean et al. 2011).

Los aportes del campo de la evolución del desarrollo

Los rasgos característicos de los humanos como ser el lenguaje complejo, el bipedalismo o la capacidad para inventar y usar herramientas sofisticadas tienen su origen en cambios morfológicos, como por ejemplo el incremento del tamaño del cerebro, cambios en la forma de los miembros y modificaciones del aparato fonador que facilitan la producción de sonidos y palabras (Carroll 2003). Todas las estructuras morfológicas distintivas presentes en los individuos adultos de nuestra especie se originan durante el desarrollo embrionario. Es por ello que aquellas preguntas enfocadas al origen evolutivo de dichos cambios morfológicos deben ser abordadas desde la perspectiva de la biología evolutiva del desarrollo (Hall 2003).

Según propusiera S.J. Gould, la evolución tienen lugar cuando la ontogenia se altera ya sea por introducción de nuevos caracteres en cualquier estadio del desarrollo, o cuando ocurren cambios heterocrónicos (Gould 2010a). Los cambios heterocrónicos tienen lugar cuando hay un cambio de la velocidad de desarrollo de caracteres ya presentes (Gould 2010a). Por otro lado, los cambios heterocrónicos como los nuevos caracteres pueden afectar cualquiera de los niveles de organización de las redes y procesos del desarrollo, como por ejemplo la transferencia del control genómico maternal al cigótico, las interacciones célula-célula, la diferenciación y migración celular, diversas interacciones a nivel tejido y orgánico, etc (Gilbert 2005). Se propone que todos estos procesos, en principio, podrían ser fuente de variación morfológica a nivel filético (Hall 2003). Por ejemplo, se encontró que las bases genéticas de la variación morfológica del ancho y largo de los picos entre diferentes especies de los llamados “pinzones de Darwin” (Fringillidae, Passeriformes) pueden ser explicadas por cambios en la expresión de dos genes *bone morphogenetic protein 4 (BMP4)* y calmodulina durante el desarrollo (Abzhanov et al. 2006). A su vez, se propone que la morfología

adulta refleja una especialización a diferentes tipos de dietas poniendo de manifiesto cómo diferentes presiones selectivas pueden modificar los patrones de desarrollo (Abzhanov et al. 2006).

Por otro lado, es de esperar que en el desarrollo de la mayor parte de los caracteres complejos, como por ejemplo el tamaño del cerebro, la morfología craneofacial o el desarrollo de las áreas corticales involucradas en el control del lenguaje, intervengan múltiples genes (Lamb 2013). En estos casos, la variación de estos caracteres a nivel evolutivo probablemente se deba a cambios en muchos genes con efecto pequeño sobre el fenotipo y uno o pocos genes con efectos mayores (Carroll 2003). Entonces, si una adaptación involucra muchos genes, se debe considerar cómo la actividad de toda la red del desarrollo y sus interacciones pueden evolucionar mediante la selección de mutaciones que ocurren en genes individuales (Lamb 2013). Nuevas revisiones en materia de la evolución de desarrollo hacen hincapié que, incluso, en la mayoría de los casos, el gen singular no sería la unidad de selección, porque sus efectos sobre el fenotipo (en última instancia lo que será seleccionado) dependen de las redes presentes y en promedio los efectos de los diferentes alelos de un gen aislado serían selectivamente neutros (Lamb 2013). Otras de las ventajas de estudiar directamente las redes del desarrollo sería la posibilidad de abarcar procesos claves como la plasticidad (la capacidad de ajustarse a las condiciones) y la canalización (estabilidad de un fenotipo frente a perturbaciones genéticas o ambientales) ya que tanto la plasticidad como la canalización son, generalmente, propiedades de las redes y no de los genes particulares (Waddington 1942; Lamb 2013).

Los mecanismos que subyacen la evolución morfológica

Un parámetro clave que afecta la contribución relativa de los diferentes mecanismos genéticos y/o procesos del desarrollo a la variación anatómica es el grado de pleiotropía que tengan posibles mutaciones o modificaciones sobre estos sistemas (Carroll 2005). Particularmente a nivel de genes puntuales, se espera que mutaciones con efectos pleiotrópicos importantes causen mayores efectos deletéreos en la aptitud biológica de los organismos y sean por lo tanto una fuente menor de variación morfológica que las mutaciones con efectos más acotados. De esta forma, el grado de pleiotropía impondría una restricción al tipo de cambio que puede tener lugar en la evolución morfológica (Carroll 2005; Prud'homme, Gompel y Carroll 2007). Teniendo esto en cuenta, es importante estudiar los mecanismos genéticos para comprender como una mutación puede afectar a los diferentes niveles que conforman un organismo y entender su potencial cómo fuente de variabilidad morfológica en las poblaciones (Britten y Davidson 1969; Carroll 2005). Dentro de este marco teórico, se propone que son fundamentalmente alteraciones en la expresión temporal y espacial de genes del desarrollo (debido a cambios en sus regiones regulatorias en *cis*) y no sustituciones aminoacídicas en los propios genes, el mecanismo principal por el cual se produce la

evolución morfológica (Jacob y Monod 1961; Monod y Jacob 1961; Britten y Davidson 1969; Carroll 2003; Prud'homme, Gompel y Carroll 2007; Wray 2007).

La idea que los cambios en regiones regulatorias podrían influenciar significativamente la evolución morfológica, no es nueva y data de los primeros estudios de Jacob y Monod sobre la regulación de la expresión génica en bacterias (Jacob y Monod 1961; Monod y Jacob 1961). Esta hipótesis se basa en una concepción esencialmente conservadora y modular del proceso de desarrollo que considera que los organismos están contruidos de unidades modulares (Gilbert 2005). La ventaja de las unidades modulares es que permiten que ciertas partes del cuerpo cambien sin interferir con las funciones de las otras partes (Gilbert 2005). Entre ellas, las regiones que forman parte de la regulación de los genes son modulares. Por ejemplo, es muy probable que cambios drásticos en la actividad de un factor de transcripción (que a su vez es responsable de la regulación de muchos genes) tenga efectos pleiotrópicos deletéreos. Por otro lado, modificaciones en la actividad de un elemento regulatorio en *cis* discreto que controla la expresión de una única proteína en un tejido determinado es probable que resulte en cambios precisos del patrón de expresión espacial y/o temporal, produciendo cambios fenotípicos específicos sin alterar la mayor parte de las vías del desarrollo (Carroll 2005; Gilbert 2005). Entonces, si un gen particular pierde un elemento regulatorio, sufre una modificación en él o lo gana, el organismo que contiene este alelo particular expresará este gen en diferentes lugares, en diferentes tiempos o a diferentes tasas en relación con aquellos organismos que conservan el alelo original pudiendo derivar en el desarrollo de diferentes morfologías (Gilbert 2005).

En la actualidad aún existen pocos trabajos en donde se estudia detalladamente la influencia de cambios en la regulación de la expresión génica con cambios fenotípicos a nivel evolutivo (de Souza, Franchini y Rubinstein 2013). Entre estos ejemplos, se encontró que cambios en un *enhancer* que controla la expresión del gen homeobox *Prx1* produjeron la expansión del dominio de expresión de este gen en los miembros anteriores en el murciélago *Carollia perspicillata* (respecto del patrón de expresión de su ortólogo en ratones). Este cambio habría contribuido con la elongación de los miembros anteriores de esta especie (Cretkos et al. 2008). En otro ejemplo, se encontró que la ausencia de espinas en la aleta pélvica de varias poblaciones de agua dulce de peces espinosos (familia Gasterosteidae) está causada por la delección del *enhancer* que dirige la expresión del factor de transcripción *Pitx1* a la aleta pélvica. El desarrollo de las espinas pélvicas, que serían un elemento de defensa contra predadores en las poblaciones marinas de peces espinosos, depende de la expresión de *Pitx*. Adicionalmente, se encontró que la delección del *enhancer* pélvico de este gen ocurrió independientemente en las diferentes poblaciones de agua dulce de los peces espinosos

(Shapiro et al. 2004; Chan et al. 2010). Por otro lado, alteraciones en regiones regulatorias son las responsables del cambio de color de la corola en la especie de planta con flor *Phlox drummondii* (Hopkins y Rausher 2011). Esta modificación fenotípica contribuiría a reducir los cruzamientos interespecíficos con la especie simpátrica *P. cuspidata* al generar una preferencia de polinizador, postulando a la selección natural como el proceso conductor del cambio. El mecanismo molecular involucra mutaciones en dos regiones regulatorias en *cis* que modulan la expresión de genes que codifican para enzimas de la vía de biosíntesis de antocianinas. De esta forma, se generan variaciones en los niveles de expresión de estos genes que finalmente derivan en cambios en la coloración de las flores (Hopkins y Rausher 2011).

La contribución que cambios en regiones regulatorias tuvieron sobre el desarrollo de rasgos específicos de los humanos es actualmente materia de intenso estudio. La idea que las bases moleculares principales de la evolución morfológica en los humanos se debe a cambios en regiones regulatorias de la expresión génica tampoco es nueva. En el trabajo clásico, “*Evolución a dos niveles en Humanos y Chimpancé*”, King y Wilson (King 1975) fueron unos de los primeros investigadores en comparar un gran número de secuencias proteicas entre chimpancé y humanos. Ellos encontraron que la distancia genética entre ambas especies, calculada en base a diferencias en secuencias proteicas y otros métodos bioquímicos, era extremadamente pequeña. La distancia genética estimada era tan pequeña como aquella encontrada entre pares de especies hermanas y muy por debajo de lo esperado para un par de especies congénicas. Sus hallazgos los llevaron a postular que las diferencias en secuencias regulatorias, más que en las secuencias proteicas, serían las responsables de las diferencias biológicas entre chimpancé y humanos. (King 1975). Este hallazgo fue considerado paradójico y propuso un rol predominante a los cambios en regiones regulatorias como posibles responsables del cambio morfológico entre nuestras especies. Sin embargo, por muchos años no existieron herramientas experimentales para poner a prueba esta hipótesis.

El estudio de regiones regulatorias

El código regulatorio del genoma

Mientras que es relativamente sencillo inferir la función de una proteína a partir de su secuencia codificante en el ADN (gracias al relativamente rígido código genético), en general no es posible descifrar las propiedades funcionales de una región no codificante solamente analizando su secuencia. Esto se debe principalmente a que aún no se conoce en detalle el código regulatorio del genoma (Rubinstein y de Souza 2013). Este conocimiento implicaría comprender las relaciones

complejas que tienen lugar entre las regiones regulatorias del ADN, factores de transcripción, ARNs regulatorios, vías de señalización y marcas epigenéticas, las que en su conjunto, determinan la expresión de un gen (Rubinstein y de Souza 2013).

Este relativo desconocimiento del código regulatorio produjo un sesgo dentro de disciplinas tales como la genómica comparativa y la genética evolutiva hacia el análisis y detección de eventos en regiones codificantes (como pueden ser los eventos de duplicación génica y evolución de la secuencia de proteínas). Mientras que, los cambios en regiones no codificantes tendieron a ser ignorados (Carroll 2005). Sin embargo, secuenciado el genoma humano, hoy se sabe que aproximadamente 1,5% de nuestro genoma está compuesto por secuencias que codifican para proteínas y un 45% está compuesto por ADN repetitivo (principalmente secuencias remanentes de retrotransposones). Aproximadamente el 50% restante está compuesto por ADN no codificante antiguamente conocido como ADN basura (Lander et al. 2001). En la actualidad, existe evidencia creciente que buena parte de estas secuencias no codificantes contienen los elementos regulatorios responsables de generar los diferentes programas de expresión génica (Heintzman y Ren 2009). La clave para comprender las bases moleculares que sustentan la evolución morfológica y fisiológica estará en dilucidar cómo actúan los mecanismos encargados de controlar la expresión génica (Britten y Davidson 1969; Carroll 2005; Wray 2007). Además, este conocimiento será muy importante para conocer cómo estos mecanismos participan en procesos patológicos y defectos del desarrollo, los cuales cada vez más están siendo asociados a mutaciones en regiones regulatorias (Rubinstein y de Souza 2013).

Los elementos regulatorios

Entre los elementos regulatorios mejor caracterizados se encuentran los promotores y las secuencias regulatorias que actúan a distancia como ser: las secuencias activadoras de la expresión o *enhancer*, los inhibidores de la expresión génica o *silencers* y las secuencias aisladoras o *insulators* (Wittkopp y Kalay 2011). La transcripción comienza en las regiones promotoras donde se reclutan las proteínas de la maquinaria basal de transcripción (Lee y Young 2000; Butler y Kadonaga 2002). Se postula que los promotores definen la orientación y el sitio exacto de inicio de la transcripción (Smale y Kadonaga 2003).

Por otro lado, los *enhancers* actúan reclutando factores de transcripción determinados, coactivadores y complejos modificadores de la cromatina que en su conjunto, promueven la transcripción desde los promotores blanco (Glass y Rosenfeld 2000; Ong y Corces 2011). Por otro lado, los *silencers* unen proteínas y/o complejos de remodelación de la cromatina que inhiben la

expresión de un gen (Glass y Rosenfeld 2000). Finalmente, las secuencias aisladoras proveen un nivel de regulación adicional, previniendo el avance de regiones de heterocromatina sobre regiones de cromatina activa, y/o restringiendo el alcance de un *enhancer* exclusivamente a su/s promotor/es blanco (West, Gaszner y Felsenfeld 2002) (figura 2).

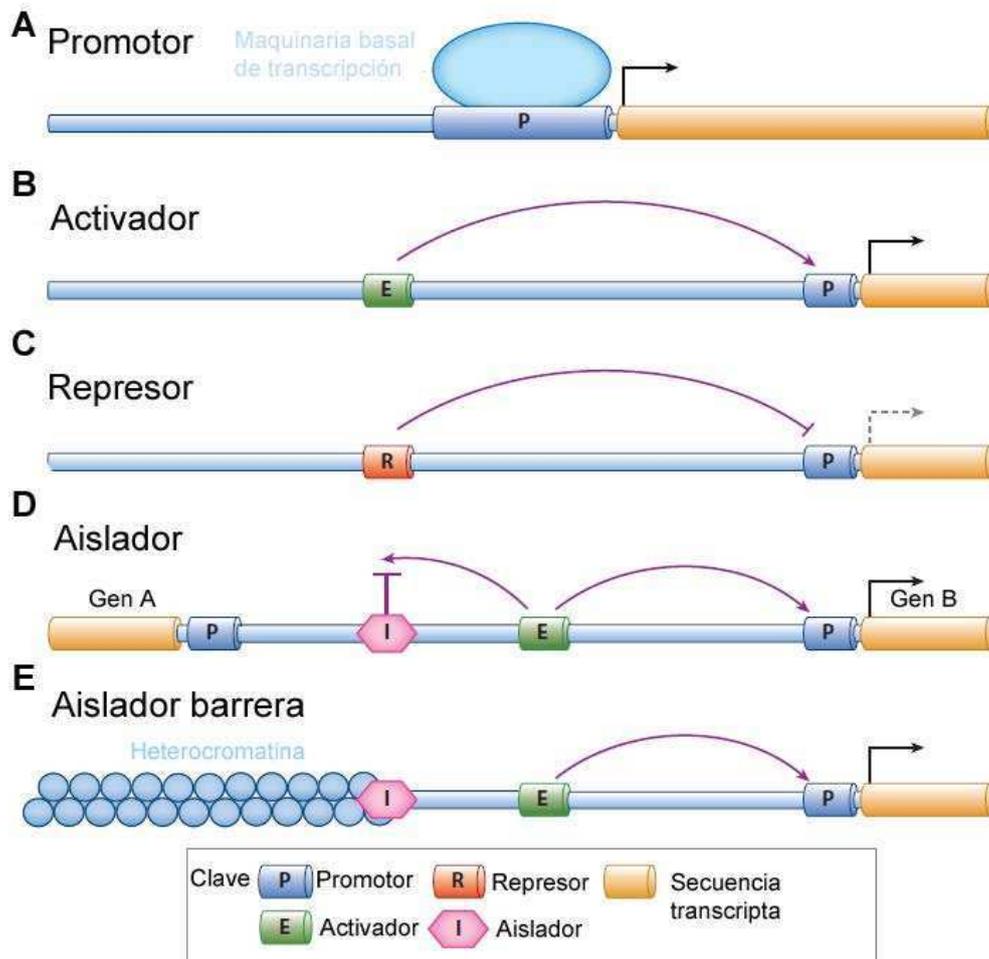


Figura 2: Tipos de elementos regulatorios conocidos. (A) Las secuencias promotoras es el sitio donde se unen los elementos de la maquinaria basal de la transcripción. Los activadores o *enhancers* (B) y los represores o *silencers* (C) median efectos positivos y negativos sobre la transcripción de sus genes blanco por medio de la interacción con los promotores. Los elementos aisladores que median el control del alcance de los *enhancers* (D) bloquean la activación de la transcripción promiscua por parte de un *enhancer* acotándola exclusivamente a su gen blanco. Existen regiones con función aisladora que funcionan como barrera al avance de las zonas de heterocromatina (E) permitiendo la normal expresión de los genes que controlan. Modificado de (Noonan y McCallion 2010).

A las secuencias *enhancers* y *silencers* se las definió clásicamente como módulos discretos de ADN capaces de ejercer su actividad independientemente de su orientación respecto del promotor o su posición al sitio de inicio de transcripción (Blackwood y Kadonaga 1998). Los *enhancers*

promueven la expresión génica reclutando factores de transcripción específicos, los cuales a su vez, promueven la descondensación de la cromatina y/o facilitan el ensamblaje de la maquinaria de transcripción del complejo de la ARN polimerasa II (ARNPolII) en los promotores (Atchison 1988; Kadonaga 1998; Maston, Evans y Green 2006). El genoma humano codifica para 1700 – 1900 factores de transcripción distintos (Vaquerizas et al. 2009). Estas proteínas usualmente contienen dos dominios distintos, un responsable de reconocer secuencias determinadas sobre el ADN (dominio de unión al ADN) y otro encargado de mediar la actividad regulatoria (generalmente conocido como dominio de transactivación). Uno de los mecanismos principales por los cuales los factores de transcripción actúan es reclutando cofactores encargados de traer proteínas de la maquinaria de remodelación de la cromatina (Kadonaga 1998; Lemon y Tjian 2000). Esta maquinaria es capaz de reposicionar nucleosomas específicos sobre los promotores de forma tal de exponer las secuencias necesarias para el ensamblaje del complejo de la ARNPolII. Por ejemplo, uno de los complejos remodeladores más estudiados es el SWI/SNF. Este complejo es capaz de modificar de forma no covalente y dependiente de ATP la cromatina, deslizando los nucleosomas sobre el ADN (Kingston y Narlikar 1999). Por otro lado, otro tipo de coactivadores que pueden ser reclutados por los factores de transcripción para promover la expresión génica son las histonas acetil transferasa (HATs). Estos cofactores son capaces de acetilar el dominio N-terminal de las histonas H3 y H4 en residuos específicos. Esta modificación sobre las histonas produce que los nucleosomas se descondensen y colabora con el reclutamiento de factores necesarios para la transcripción (Strahl y Allis 2000). Entre los complejos de HATs más conocidos se encuentran: PCAF, CBP, p300, GCN5, TRRAP, entre otros (Rosenfeld, Lunyak y Glass 2006). Adicionalmente, los mediadores son un tercer tipo de cofactores reclutados por los factores de transcripción. Estas proteínas facilitan la transcripción por medio de la interacción directa con la ARNPol II (Hatzis y Talianidis 2002) (figura 3).

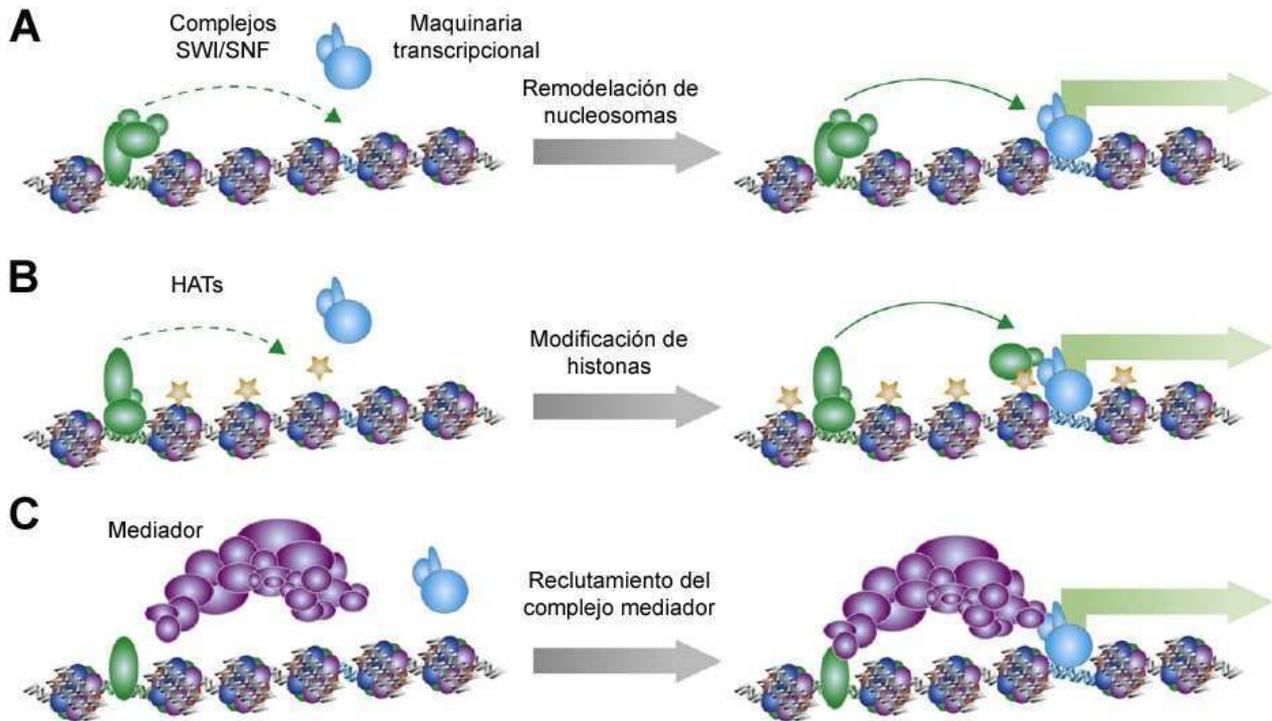


Figura 3: Tres mecanismos clásicos por los cuales pueden actuar los *enhancers*. (A) Los factores de transcripción reclutan maquinaria de remodelación de la cromatina conteniendo los complejos como SWI/SNF (óvalos verdes). Estos complejos son capaces de deslizar los nucleosomas sobre el ADN de los promotores (ADN azul) exponiendo secuencias necesarias para el ensamblaje de la maquinaria de transcripción (óvalos azules). (B) Los factores de transcripción pueden también reclutar complejos de coactivadores con actividad de HAT capaces de acetilar residuos específicos en las histonas (indicado como estrellas amarillas sobre la cromatina). Esta modificación en la cromatina puede facilitar la unión de más factores promotores de la transcripción (óvalos verdes adicionales) o descondensar la cromatina promoviendo la unión de la maquinaria transcripcional (óvalos azules). (C) Los factores de transcripción pueden reclutar complejos mediadores (óvalos violetas) los cuales, a su vez, pueden interactuar directamente con la ARNPolIII para activar la expresión. Modificado de (Heintzman y Ren 2009).

Predicción de *Enhancers*

El modelo actual de regulación génica propone que los *enhancers* contienen en sus secuencias la información necesaria para reclutar los factores de transcripción encargados de controlar temporal y espacialmente la actividad de sus genes blanco. Sin embargo, las secuencias consenso para unión de los factores de transcripción son, en general, altamente degeneradas y se las puede encontrar extensamente distribuidas por todo el genoma (Matys et al. 2003), por lo que la sola predicción de estos sitios es insuficiente para postular la existencia de un *enhancer* (Noonan y McCallion 2010; Wittkopp y Kalay 2011). No obstante, en la actualidad se cuenta con dos grandes aproximaciones metodológicas por las cuales se puede predecir la existencia de una región regulatoria (Heintzman y

Ren 2009). Por un lado están los métodos que detectan características en la cromatina y en las proteínas unidas al ADN que, a su vez, están asociadas a la presencia de *enhancers*. Paralelamente, están los métodos que se basan en la conservación de las secuencias de ADN a nivel filogenético (Tagle et al. 1988; Bulger y Groudine 2011).

Siguiendo la primera aproximación, los *enhancers* típicamente se localizan en regiones donde hay una disrupción en la estructura de la cromatina. Esta disrupción puede ser puesta en evidencia como una mayor sensibilidad a la digestión con la enzima DNAsa I. Esta hipersensibilidad a la digestión con DNAsa I ocurre debido a que los nucleosomas son desplazados del ADN por la unión de proteínas específicas (como por ejemplo, los factores de transcripción). De esta forma, el ADN está más expuesto a la digestión (Atchison 1988). Por otro lado, el ADN de regiones *enhancers* presenta modificaciones específicas en las histonas de sus nucleosomas. Dentro de este código se encuentran la monometilación de la histona H3 en su lisina 4 (H3K4Me1) y ausencia de trimetilación (H3K4Me3) (Bulger y Groudine 2011). Otro tipo de evidencia experimental que sugiere la existencia de un *enhancer* es la asociación de regiones del ADN con proteínas específicas (Heintzman y Ren 2009). Una de las asociaciones más estudiadas es el caso de la proteína p300, un coactivador estrechamente asociado a los *enhancers* activos (Heintzman et al. 2007). Por ejemplo, utilizando una estrategia que consistía en inmunoprecipitar ADN utilizando un anticuerpo contra la proteína p300 (marcador de *enhancers* activos), seguido de secuenciación masiva, se encontraron entre 500 y 2500 sitios de unión para p300 en tejido de cerebro de embriones de ratón de 11.5 *días post coitum* (dpc). En casi todos los casos, las regiones halladas resultaron funcionar como *enhancers* en ensayos con ratones transgénicos (Visel et al. 2009). Más recientemente, el Proyecto ENCODE identificó aproximadamente 400.000 elementos con características cromatínicas asociadas con la presencia de *enhancers* en diferentes tipos celulares humanos (Bernstein et al. 2012). La ventaja de este tipo de aproximaciones es que, en teoría, permite simultáneamente identificar un *enhancer* junto con la información sobre el tejido y estadio del desarrollo donde el regulador está activo (Visel et al. 2009). Sin embargo, varios autores consideran que las aproximaciones basadas en el mapeo a gran escala de características de la cromatina proveen indicios sobre la presencia de regiones con potencial capacidad regulatoria pero que deben hacerse necesariamente estudios funcionales adicionales para comprobar esta función (de Souza, Franchini y Rubinstein 2013; Graur et al. 2013).

Por otro lado, otro criterio utilizado para detectar posibles *enhancers* (y otras regiones funcionales en el genoma) es la conservación filogenética en regiones no codificantes (Tagle et al. 1988). La hipótesis en la que se basan estos estudios comparativos asume que aquellas regiones funcionales

del genoma (tanto que codifican para proteínas como no codificantes) tendrán una tasa de evolución molecular más baja, comparada con la tasa neutra de evolución (que es la tasa a la cual, en teoría, evolucionan las regiones no funcionales). Esto se debe a que solo una pequeña proporción de los cambios que tienen lugar en las regiones con restricciones funcionales pasará a la próxima generación (cambios que resulten selectivamente neutros y/o ventajosos) (Kimura 1968; Tagle et al. 1988). De acuerdo con la teoría neutralista o del equilibrio mutación-deriva, estos nuevos alelos neutros o ventajosos se fijarán en las poblaciones ya sea por deriva genética o por selección natural, acumulándose, y formarán parte de la divergencia que se analizará en los estudios comparativos. Sin embargo, la gran parte de las sustituciones que ocurren en las regiones funcionales del genoma resultan deletéreas, siendo el destino más probable de estos alelos la eliminación por selección purificadora. Estos cambios no quedan representados en las poblaciones, no aportando a la divergencia. Por otro lado, las mutaciones que ocurren en sitios sin restricciones funcionales tienen por lo general efectos despreciables sobre la aptitud (es decir, son selectivamente neutros) y están libres para fijarse por deriva génica, de allí la mayor tasa de evolución molecular en estas regiones (Kimura 1968; Kimura 1983).

En consecuencia, las regiones genómicas sin restricciones funcionales entre dos especies acumulan sustituciones a tasa neutras y la divergencia aumenta a medida que la distancia evolutiva entre las especies se incrementa. Mientras que, las secuencias ortólogas sujetas a restricciones funcionales en ambas especies también cambian pero a una tasa menor. Entonces, la conservación filética es un parámetro extremadamente útil para predecir la presencia de un *enhancer* u otro tipo de región funcional en secuencias no codificantes (Tagle et al. 1988). Sin embargo este método no permite predecir dónde y cuándo ese posible regulador estará activo. También, es importante notar que la conservación entre secuencias ortólogas es un indicador de conservación de la función pero no lo garantiza (Noonan y McCallion 2010). Por ejemplo, secuencias conservadas entre dos especies pueden tener en cada una funciones distintas y a la inversa, funciones conservadas entre dos especies pueden estar codificadas en secuencias con muy poca conservación (Noonan y McCallion 2010). Por ejemplo, se encontró que secuencias regulatorias presentes en el *locus RET* humano y que no presentan niveles de conservación manifiesta con elementos presentes en el genoma del pez cebra (*Danio rerio*) son capaces de dirigir la expresión a dominios específicos del gen *RET* endógeno cuando se evalúan en peces cebra transgénicos. Una de las explicaciones para este hallazgo es que existen secuencias ortólogas en pez cebra a las humanas ensayadas pero estas secuencias divergieron más allá del nivel de detección ya sea por pequeños cambios en los sitios de unión para factores de transcripción, rearrreglos de estos sitios u otros cambios evolutivos (Fisher et al. 2006a). En un ejemplo contrario al anterior, se reportó que los elementos regulatorios ar-D, ar-A

y ar-C del gen *Sonic hedgehog* (*Shh*) en pez cebra dirigen la expresión a dominios distintos del animal cuando se los compara con sus elementos ortólogos en ratón. Esta diferencia en la actividad de estos *enhancers* es llamativa si se toma en cuenta que las secuencias de estos elementos están altamente conservadas entre estas especies (con aproximadamente un 60% de identidad de secuencia) (Ertzer et al. 2007).

La identificación de regiones funcionales en el genoma por conservación de secuencia se aceleró sustancialmente gracias a que ahora se cuenta con múltiples genomas secuenciados para comparar y métodos sofisticados para evaluar el nivel de conservación de las secuencias (Noonan y McCallion 2010). Los primeros estudios realizados comparando el genoma humano con el de otros vertebrados encontraron miles de secuencias altamente conservadas no codificantes. Estas fueron llamadas *conserved noncoding sequences* (*CNSs* por sus siglas en inglés) (Siepel et al. 2005; Venkatesh et al. 2006). Estos estudios hallaron que los *CNSs* no están distribuidos al azar, agrupándose preferencialmente cerca de genes del desarrollo y en desiertos génicos. La gran representación de *CNSs* asociados a genes del desarrollo puede indicar que, en oposición a genes controlados menos dinámicamente, estos genes requieren un mayor número de regiones regulatorias para poder modular de forma precisa la expresión durante el desarrollo. Además, se encontró que las agrupaciones de *CNSs* correlacionan con regiones donde la sintenia está conservada entre los vertebrados. Esto podría deberse a que estos elementos regulatorios deben permanecer asociados con sus genes blanco, siendo estas regiones genómicas más resistentes a los rearrreglos (Boffelli, Nobrega y Rubin 2004; Ahituv et al. 2005; Ovcharenko et al. 2005).

Más recientemente, un estudio mapeó en gran detalle las secuencias de ADN bajo selección purificadora en los mamíferos, utilizando para ello las secuencias completas de 29 genomas, incluido *Homo sapiens*. Este análisis mostró que 4,2% de las secuencias presentes en los genomas de estos mamíferos están conservadas filogenéticamente. Dentro de esta fracción, 68% de estas secuencias son intrónicas e intergénicas y, al menos, un 30% de las secuencias conservadas totales se encuentra en *loci* con marcas de cromatina típicas de *enhancers* (Lindblad-Toh et al. 2011). Esto indicaría que, aproximadamente dos tercios de todas las secuencias bajo selección purificadora (al menos dentro de los mamíferos) del genoma humano son no codificantes. Entonces, cualquier trabajo que quiera encontrar las bases genéticas de la diversidad morfológica debería considerar a las regiones no codificantes como parte importante de su campo de estudio (Carroll 2005).

Genes blanco de los *enhancers*

Otro de los grandes desafíos cuando se trabaja con *enhancers*, una vez identificados, es determinar cuáles son sus genes blanco. Se puede visualizar lo complejo de este objetivo si se tiene en cuenta que los *enhancers* se pueden encontrar en posición proximal o distal, actuando incluso a millones de pares de bases de distancia del promotor blanco (Lettice et al. 2003). Además, se pueden hallar dentro de intrones u exones de genes vecinos e incluso puede activar genes blanco que se localizan en otros cromosomas (Lomvardas et al. 2006; van Heyningen y Bickmore 2013). No se conoce aun en detalle el mecanismo preciso por el cual estos reguladores pueden actuar sobre tales distancias. Uno de los modelos más populares propone que los elementos regulatorios distales son capaces de reclutar proteínas que interactúan a su vez con la maquinaria de transcripción vía formación de un lazo de cromatina. De esta forma, regulador y promotor quedan en estrecho contacto (Krivega y Dean 2012).

Sin embargo, como primera aproximación para identificar los genes blanco se suele asignar un *enhancer* al gen que tenga físicamente más próximo. No obstante, esta hipótesis se vuelve más sólida si el patrón de expresión del gen candidato y el *enhancer* se superponen (al menos en parte del dominio espacial y temporal de expresión del gen) (Heintzman y Ren 2009). La demostración definitiva requiere de ensayos delecionales o de mutagénesis (ya sea en mutantes naturales o recurriendo a modelos animales) donde se pueda evaluar el impacto de cambios en las regiones regulatorias sobre el patrón de expresión de su gen blanco.

Uno de los ejemplos más llamativos de *enhancer* distales es el regulador del factor de transcripción *Sonic hedgehog* (*SHH*) que controla la expresión de este gen en los miembros posteriores en desarrollo de mamíferos. Este *enhancer* se localiza a más de 1 Mb del sitio de inicio de la transcripción de *SHH*, en medio de un intrón del gen *Limb Region 1* (*LMBR1*), gen que se encuentra río arriba de *SHH* (Lettice et al. 2003). La disrupción de este *enhancer* en ratón ocasiona la degeneración de los miembros posteriores por falta de un correcto establecimiento del eje antero-posterior del miembro en desarrollo (Sagai et al. 2005).

El ejemplo anterior pone de manifiesto la importancia de los *enhancer* distales durante el desarrollo y lo complejo de asociar *a priori* *enhancers* y genes blanco. Incluso, ya se empieza a vislumbrar un nivel de complejidad adicional a la hora de estudiar *enhancers*. Tradicionalmente, se consideró que cada *enhancer* controlaba un único dominio de expresión de su gen blanco. Sin embargo, existe evidencia creciente sobre la gran redundancia de elementos regulatorios que habría en los metazoos, principalmente aquellos asociados a genes del desarrollo (Barolo 2012; Frankel 2013). Por ejemplo,

un estudio llevado a cabo en moscas *Drosophila melanogaster* transgénicas, demostró que dos genes blanco del factor de transcripción *Dorsal*, los genes *brinker* y *sog*, tienen cada uno, al menos, dos *enhancers* capaces de reclutar a *Dorsal*. Estos dos pares de *enhancers* están activos durante los mismos estadios del desarrollo larval del animal y dirigen la expresión a dominios aparentemente idénticos (Hong, Hendrix y Levine 2008; Perry et al. 2010). Se postula que esta redundancia de función podría estar asociada a una mayor estabilidad fenotípica (permitiendo resistir a cambios deletéreos en las regiones regulatorias) (Hong, Hendrix y Levine 2008; Frankel et al. 2010). Por otro lado, también se sugiere que los *enhancers* con actividad regulatoria parcialmente redundante podrían ser fuente de novedades evolutivas. En este tipo de sistemas, las mutaciones que tuvieran lugar en uno de estos elementos podrían ser mejor tolerados, pudiendo, al mismo tiempo, ser fuente de nuevos patrones de expresión (Hong, Hendrix y Levine 2008). No obstante, a nivel experimental complejizan aun más los ensayos funcionales ya que, la ausencia de un fenotipo por delección o mutación no necesariamente implica la ausencia de función de ese elemento *in vivo*.

Ensayos funcionales para regiones regulatorias

Como se detalló anteriormente, en la actualidad se cuenta con métodos refinados para predecir regiones regulatorias. Sin embargo, siempre se deben realizar ensayos *in vivo* para confirmar la presencia de un elemento regulatorio (de Souza, Franchini y Rubinstein 2013). La mayoría de los trabajos sobre regiones funcionales no codificantes se concentran en el estudio de regiones *enhancer*. Esto es así, principalmente, porque no existen aproximaciones experimentales robustas y aplicables a gran escala para probar regiones, por ejemplo, con función *silencer* y/o *insulator* en vertebrados (Noonan 2009).

Todos los ensayos funcionales para regiones regulatorias requieren de la construcción de transgenes y la posterior introducción de los mismos en los organismos o cultivos celulares adecuados (Arnone 2004). Idealmente, en los casos que se quiera evaluar cambios regulatorios a nivel filético, los dos elementos candidatos deben ser probados individualmente en las dos especies de las cuales estas regiones provienen. Estos experimentos se pueden realizar construyendo pares recíprocos de organismos modificados genéticamente (Wittkopp y Kalay 2011).

Para estudiar funcionalmente regiones *enhancer*, se requiere, además, de un promotor para mediar su actividad y un gen blanco sobre los cuales actuar. Estos elementos son fundamentales y deben ser tenidos en cuenta a la hora de construir los transgenes. Los transgenes se pueden diseñar de forma tal que contengan al *locus* genómico completo. En estos casos, el transgén contiene el gen blanco (conteniendo o no un gen reportero o etiqueta en su interior), su propio promotor y el

enhancer en su ubicación nativa (Wittkopp y Kalay 2011). También se pueden construir transgenes quiméricos en donde el *enhancer* candidato suele ubicarse río arriba de un gen reportero (generalmente en posición proximal). En este segundo caso, se puede utilizar el promotor mínimo del gen blanco (si este se conoce). Por el contrario, si el gen blanco de un *enhancer* no se conoce o no se cuenta con la secuencia del promotor mínimo, se recurren a promotores mínimos que derivan de otros genes e incluso de especies distintas que las que proviene el *enhancer*. Los promotores mínimos son elementos incapaces de activar la expresión de los genes reporteros en ausencia de un *enhancer* funcional y solo brindan las secuencias necesarias para el ensamblaje del complejo de la ARNPolIII (Wittkopp y Kalay 2011). Es importante remarcar que muchos *enhancer* conservan su patrón de expresión aun cuando son removidos de su contexto genómico normal. Esta autonomía implica que los *enhancers* pueden ensamblar un conjunto completo de actividades bioquímicas que son lo suficientemente robustas como para activar la expresión génica de forma definida (Swanson, Evans y Barolo 2010).

Una vez construido el transgén, éste debe ser introducido en el modelo experimental adecuando. Existen diferentes métodos para introducir los transgenes dependiendo del modelo con el que se trabaje. Por ejemplo, transfección de células en cultivo (por ejemplo, líneas celulares humanas), transfecciones transientes en animales transgénicos (como en peces), introducción de material genético extracromosomas (en nematodos), integración al azar en el genoma (como en plantas, peces cebra y ratones) e integración dirigida a *loci* endógenos determinados (como en moscas). El último nivel de resolución es el reemplazo de *loci* endógenos y esto puede ser realizado en levaduras o en ratones (Wittkopp y Kalay 2011; Ahituv 2012)

Una de las metodologías más novedosas que se dispone para evaluar la actividad *enhancer in vivo* consiste en un ensayo basado en el uso del transposón Tol2 de Medaka (*Orzias latipes*) para mediar de forma eficiente la transgénesis en pez cebra (*Danio rerio*) (Kawakami y Shima 1999). Esta metodología hace uso de la alta eficiencia de integración de este transposón para movilizar los transgenes al genoma (Kawakami, Shima y Kawakami 2000). Para evaluar la actividad regulatoria, los elementos candidatos se clonan en un vector universal que contiene las secuencias necesarias para el reconocimiento de la transposasa y un gen reportero (generalmente *EGFP*) bajo la regulación de un promotor mínimo (como por ejemplo el promotor mínimo del gen *cFos* de ratón). Este vector se co-inyecta con el ARNm que contienen la secuencia codificante para la transposasa en embriones de pez cebra en el estadio de 1 a 2 células (Kawakami et al. 2004; Fisher et al. 2006b). Dentro de las ventajas que cuenta este sistema está la alta tasa de transgénesis alcanzada y la relativamente baja laboriosidad del método (comparado con la producción de ratones

transgénicos). Además, permite evaluar la actividad *enhancer* en un modelo de vertebrado y a lo largo del desarrollo embrionario (Fisher et al. 2006b). Otra de las ventajas es que los embriones de pez cebra pueden ser tratados con inhibidores químicos para obtener peces transparentes, facilitándose así el análisis de la expresión del gen reportero en embriones intactos (Karlsson, von Hofsten y Olsson 2001). Con este método fue posible evaluar la actividad regulatoria de elementos humanos con éxito (Fisher et al. 2006a).

Cuando se estudian cambios regulatorios a nivel evolutivo que involucran secuencias humanas, nunca es posible hacer los experimentos recíprocos mencionados anteriormente. La única alternativa para evaluar los elementos en un contexto celular nativo consiste en realizar ensayos en líneas celulares humanas (Wittkopp y Kalay 2011). Sin embargo, esta metodología presenta serias limitaciones. Por un lado, debe conocerse el patrón de expresión del *enhancer* para utilizar una línea celular donde ese elemento esté activo. Por el otro, los trabajos sobre cultivos celulares no pueden dar información sobre el patrón temporal de expresión de los elementos analizados en un organismo completo (Ahituv 2012) como es necesario, por ejemplo, cuando se estudian elementos activos durante el desarrollo embrionario. Es por ello que la mejor alternativa en estos casos es realizar estudios donde las diferentes secuencias ortólogas sean evaluadas en una tercera especie modelo (Vallender, Mekel-Bobrov y Lahn 2008). Por ejemplo, existen numerosos trabajos que analizan la expresión de secuencias regulatorias humanas recurriendo a la utilización en ratones (Pennacchio et al. 2006; Visel et al. 2007) o peces transgénicos (Fisher et al. 2006a; Visel et al. 2007). En estos casos, se trabaja sobre animales modelos de mamíferos y vertebrados respectivamente, pudiéndose registrar la actividad de un *enhancer* a lo largo del desarrollo embrionario del animal.

Sin embargo, también estos ensayos tienen sus limitaciones. Los estudios realizados con ratones transgénicos son muy costosos y laboriosos. Es por ello que en general, se analiza uno o unos pocos estadios del desarrollo (típicamente a 11.5 dpc). Esto limita la posibilidad de encontrar regiones activadores de la expresión génica ya que aquellas regiones activas en etapas anteriores o posteriores a la/s ensayada/s serán consideradas inactivas (Noonan 2009). Por otro lado, los estudios en animales transgénicos requieren que las secuencias humanas sean funcionales en contextos celulares heterólogos, como los brindados por ratones o peces. Estos modelos podrían no brindar el entorno necesario para que se recapitule la actividad endógena de un regulador (Noonan 2009). Por ejemplo, diferencias en la afinidad de un factor de transcripción por una secuencia blanco o cambios en el patrón temporal o espacial de un factor de transcripción podrían resultar en un patrón de expresión de un *enhancer* humano en ratón o en pez cebra que no reproduce el patrón endógeno humano (Ritter et al. 2010; Ariza-Cosano et al. 2012).

Incluso, recientemente se descubrió una propiedad muy llamativa de la regulación génica que podría complejizar aun más el análisis de las regiones regulatorias. El gen *Fgf8* codifica un factor clave para el desarrollo embrionario (Crossley y Martin 1995). Este gen está regulado por múltiples *enhancers*, muchos de ellos con actividades tejido-específicas superpuestas, distribuidos a lo largo de una región genómica de 220 kpb (Beermann et al. 2006; Komisarczuk, Kawakami y Becker 2009). Incluso, se hallaron *enhancers* de *Fgf8* ubicados dentro de intrones de otros genes no relacionados (evolutiva y/o funcionalmente) (Marinic et al. 2013). Llamativamente, se encontraron por medio de ensayos de expresión de genes reporteros en ratones transgénicos que dentro de estas 220 kpb existe un gran número de *enhancers* adicionales cuyas actividades regulatorias no se traducen de forma obvia al patrón de expresión endógeno de *Fgf8* (es decir, dirigen el patrón de expresión de la proteína reportera a dominios de expresión diferentes de los de *Fgf8* u otros genes vecinos). Este estudio sugiere que puede existir una gran diferencia entre la capacidad intrínseca de un *enhancer* (aquella que se manifiesta cuando se inserta al azar en el genoma en los ensayos de genes reporteros) y su verdadera función (la contribución real a la expresión de su gen blanco en su posición genómica normal). Adicionalmente, sugiere que la estructura espacial del *locus* de *Fgf8* podrían tener una capacidad adicional de filtrar la influencia de ciertos *enhancers* locales de forma tal que estas actividades no se reflejen en el patrón de expresión endógeno (Marinic et al. 2013).

Origen de los enhancers y exaptación

Existen varios mecanismos genéticos que pueden dar lugar al surgimiento de nuevos *enhancers*. Por ejemplo, pueden originarse por mutaciones al azar que generan agrupamientos de sitios de unión para factores de transcripción o por rearrreglos cromosómicos que promueven la adopción de nuevos *enhancers* por parte de genes blanco. Un tercer mecanismo posible involucra la inserción de elementos móviles o elementos transponibles (ET) (Rubinstein y de Souza 2013). Los ET son regiones discretas de ADN que se pueden mover de un sitio a otro dentro de los genomas y, a veces, incluso entre genomas (Rouleux-Bonnin 2011). En mamíferos, un 30-50% de estos genomas están compuestos por ET y estos elementos se pueden clasificar en: 1) elementos autónomos como los *long interspersed nucleotide elements* (LINEs por sus siglas en inglés), 2) los *short interspersed nucleotide elements* (SINEs por sus siglas en inglés), que dependen de los LINEs para su propagación, 3) retrovirus endógenos con largos elementos terminales repetitivos (o LTRs, por *long terminal repeats*), 4) transposones de ADN. Los tres primeros tipos de ET son los más abundantes en mamíferos, se propagan por medio de un ARN intermediario y se llaman retrotransposones, mientras que los transposones de ADN utilizan una transposasa para cortarse y pegarse en otros *loci* del genoma (Bohne et al. 2008).

Si bien se sugirió que los ET formaban parte del ADN no funcional o se los consideró parásitos que explotaban los recursos de las células para su propagación (Doolittle y Sapienza 1980; Orgel y Crick 1980), existe evidencia que los ET tendrían un importante rol en la evolución funcional y estructural de los genomas (Kazazian 2004; Rouleux-Bonnin 2011). Los casos donde elementos identificables en el ADN, como ser los ET, con una capacidad potencial de tornarse funcionales durante el curso de la evolución (aunque su origen haya sido no adaptativo), resultan eventualmente cooptados para una determinada función, se conocen como exaptaciones (Brosius y Gould 1992). Resultados experimentales recientes obtenidos en diferentes modelos indican que los ET son capaces de movilizar conjuntos preexistentes de sitios para la unión de factores de transcripción, pudiendo, de esta forma, actuar como una fuente de elementos regulatorios novedosos, especialmente linaje-específicos (Bourque et al. 2008; Smith et al. 2008; Studer et al. 2011; Romero, Ruvinsky y Gilad 2012). Esta evidencia refuerza la hipótesis que la exaptación de ET podría haber contribuido con el surgimiento de novedades evolutivas promoviendo cambios en las redes de regulación génica en mamíferos y otros vertebrados (Feschotte 2008; Bourque 2009). Sin embargo, solo en algunos casos se pudo demostrar funcionalmente que la actividad regulatoria fue una innovación mediada por la exaptación de un elemento móvil (de Souza, Franchini y Rubinstein 2013).

HIPÓTESIS GENERAL

La hipótesis general de este trabajo es que la adquisición de nuevos patrones de expresión de genes relacionados con el desarrollo y funcionamiento del cerebro en el linaje humano, habría sido crítica para la evolución neuroanatómica de nuestro cerebro y sus capacidades cognitivas diferenciales. Estos nuevos patrones de expresión se originarían por cambios en la secuencia de ADN de regiones regulatorias de genes que se expresan en el sistema nervioso central.

OBJETIVO GENERAL

El objetivo principal de este proyecto es identificar las bases genéticas de la evolución del cerebro humano. Para ello se propone realizar un estudio detallado a nivel evolutivo y funcional de regiones genómicas no codificantes, altamente conservadas y que cambiaron su tasa de evolución molecular exclusivamente en el linaje humano. Estas regiones podrían contener elementos regulatorios de la expresión génica que cambiaron su actividad exclusivamente en nuestro linaje. Este estudio comprende los siguientes puntos: (1) analizar la distribución de estas regiones en el genoma humano, identificando *loci* particularmente enriquecidos en estos elementos, (2) estudiar si estas regiones no codificantes y conservadas pueden actuar como elementos regulatorios de la expresión génica, (3) evaluar si ocurrieron cambios a nivel filogenético en la actividad de estas regiones regulatorias e identificar los posibles genes blanco, (4) estudiar las funciones de los potenciales genes blanco en un marco comparativo. Los puntos mencionados anteriormente serán tratados en detalle en cada uno de los tres capítulos de esta tesis. Finalmente, se espera que de la integración de los resultados obtenidos en cada uno de los puntos anteriores se pueda relacionar cambios a nivel regulatorio con cambios a nivel fenotípico dentro de nuestro linaje.

Capítulo I

Estudio de agrupamiento genómico de regiones de evolución acelerada en el linaje humano

INTRODUCCIÓN ESPECÍFICA

La evolución regulatoria de nuestro genoma

Tuvieron que pasar muchos años para que se pudiera comenzar a poner a prueba la hipótesis de King y Wilson y emprender el estudio de cambios específicos de los humanos en la regulación de la expresión de los genes (King 1975). La hipótesis principal en estos trabajos sostiene que una parte significativa de características exclusivas de nuestra especie debieron haber surgido por medio de modificaciones a nivel regulatorio de la arquitectura génica del desarrollo; es decir, como variaciones específicas de los humanos del programa general, altamente conservado, del desarrollo de los mamíferos (Britten y Davidson 1969; King 1975; Carroll 2003; Noonan 2009).

No obstante, identificar regiones regulatorias activas durante el desarrollo embrionario que hayan cambiado su actividad exclusivamente en el linaje humano no es una tarea sencilla. Como los programas genéticos del desarrollo tienden a estar altamente conservados (Gilbert 2005), es esperable que gran parte de las secuencias encargadas de la regulación de estos genes también estén conservadas entre nosotros y el resto de los primates u otros mamíferos. Entonces, para llevar a cabo esta identificación se requieren métodos específicos con los cuales detectar, a nivel de las secuencias de ADN no codificante, las evidencias dejadas por los procesos evolutivos (Pollard et al. 2006a; Prabhakar et al. 2006; Bird et al. 2007; Lindblad-Toh et al. 2011)

Sin embargo, existen grandes limitaciones a la hora de aplicar al estudio de las regiones no codificantes los métodos más usados para la detección a nivel molecular de procesos de selección positiva. Por ejemplo, en las regiones genómicas no codificantes no es sencillo predecir si una sustitución nucleotídica determinada resulta sinónima o no sinónima (uno de los principios utilizados para la detección de evolución adaptativa en regiones que codifican para proteínas) (Kostka et al. 2012). Esto se debe a que no se conoce en detalle como se codifica la información en los elementos regulatorios del genoma. Sumando a lo anterior, es muy probable que los aspectos más importantes de la evolución de los rasgos específicos de nuestro linaje a nivel genético (incluidos varias instancias de expansión de nuestro cerebro) hayan ocurrido en tiempos profundos de la historia evolutiva de nuestra especie (más allá de los 2 millones de años) (Striedter 2005; Vallender, Mekel-Bobrov y Lahn 2008). Para estos tiempos, la identificación a nivel genético de procesos de evolución adaptativa por detección de polimorfismos actuales no es aplicable (Tattersall 2000). En su conjunto, estas características impiden o limitan enormemente el uso de los algoritmos más frecuentemente usados para la detección de procesos de evolución adaptativa en los

genomas (sean estos los basados en la comparación de las tasas de sustitución sinónima y no sinónima en el primer caso o de estudio de frecuencias de polimorfismos actuales, en el segundo).

Dadas las limitaciones anteriormente mencionadas, una forma de predecir los sitios más probables donde pudieron haber ocurrido cambios a nivel de la regulación de los genes es recurrir a la detección de modificaciones significativas en la tasa de evolución molecular en regiones no codificantes, muy conservadas evolutivamente y en un linaje determinado (Pollard et al. 2006; Prabhakar et al. 2006; Kostka et al. 2012). Aplicando esta metodología se pueden poner en evidencia eventos de evolución acelerada; entendiendo por esta al patrón observado en determinadas regiones genómicas para un linaje específico donde ocurrieron más sustituciones nucleotídicas que las que se hubieran esperado a partir de la tasa de evolución de fondo de esa región (Noonan 2009; O'Bleness et al. 2012). Detectar evolución acelerada sobre regiones conservadas del genoma tiene varias ventajas. Por un lado, la conservación es evidencia de restricción funcional y en último lugar, de importancia biológica. Por el otro lado, el poder de detección de cambios en la tasa de evolución es mayor cuando se compara con una tasa de evolución de fondo baja que cuando se compara con una tasa cercana a la neutra y además, se reducen los costos computacionales ya que sólo se analiza una porción acotada (típicamente del 5-10%) del genoma (Kostka et al. 2012). Sin embargo, esta metodología tiene sus desventajas. Como en estos casos lo que se estudia son cambios en la tasa de evolución molecular es muy difícil identificar los procesos evolutivos responsables de la aceleración. Por ejemplo, en las regiones con evidencia de evolución aceleradas la tasa de sustitución puede ser significativamente más alta que la esperada dada la extrema conservación de esa secuencia en otras especies, pero puede llegar a no exceder la tasa neutral de evolución y entonces podría reflejar un proceso de relajación de la selección purificadora más que un proceso de evolución adaptativa (Kostka et al. 2012). Entonces, se requieren test adicionales para distinguir los procesos evolutivos subyacentes, como por ejemplo, comparar la tasa de evolución molecular en la región acelerada con la tasa neutra de evolución de esa región, entre otros (Pollard et al. 2006a; Lindblad-Toh et al. 2011; Kostka et al. 2012).

Por otro lado, la selección positiva y la relajación de la selección negativa son algunas de las causas que pueden producir un patrón de evolución acelerado, pero no las únicas. Por ejemplo, los cambios en el tamaño o estructuración de las poblaciones pueden llevar a una fijación rápida de los alelos segregantes y producir modificaciones en la tasa de evolución molecular (Kreitman 2000). Adicionalmente, la aceleración de la tasa de sustitución también puede ser el producto de procesos que no involucran necesariamente cambios en el modo de selección o eventos poblacionales (Pollard et al. 2006). Por ejemplo, el proceso evolutivo selectivamente neutro conocido como GC-

biased gene conversion (gBGC por sus siglas en inglés) es un mecanismo mutagénico causado por la preferencia en las enzimas que reparan el ADN por generar apareamientos entre bases fuertes (GC) frente a los de bases débiles (AT) en sitios donde se hayan generado heteroduplex de ADN durante la recombinación meiótica (Strathern, Shafer y McGill 1995). Por lo tanto, el proceso de gBGC puede producir un incremento en la probabilidad de fijación de alelos C y G frente a alelos A y T, generando en las zonas afectadas un rápido incremento en tasa de sustitución, comparable en magnitud al producido por un proceso de selección positiva. Sin embargo, hay una diferencia: se observa un gran sesgo en las sustituciones de bases débiles a bases fuertes. Una vez que el proceso de gBGC alcanza un equilibrio decrece la tasa de sustitución (Eyre-Walker y Hurst 2001; Kostka et al. 2012). Es importante destacar que, si bien el proceso de gBGC se considera selectivamente neutro, existen trabajos que muestran que el incremento del contenido GC de secuencias transcritas puede incrementar su nivel de expresión, lo que podría traer asociado algún valor adaptativo (Kudla et al. 2006). Sin embargo, el rol del proceso de gBGC en la evolución está aún en discusión.

Detección de evolución acelerada en el genoma humano

Hasta el presente cinco grupos han llevado a cabo estudios genómicos a gran escala con el objetivo de detectar regiones genómicas filogenéticamente conservadas que hayan cambiado su tasa de evolución molecular exclusivamente en el linaje humano (Pollard et al. 2006; Prabhakar et al. 2006; Bird et al. 2007; Bush y Lahn 2008; Lindblad-Toh et al. 2011). Cada uno de estos grupos, de forma independiente, aplicó su propia metodología produciendo como resultado una base de datos de elementos acelerados. El esquema general para la construcción de estas bases de datos se puede resumir en tres pasos:

- 1- **Identificación de regiones genómicas conservadas.** Estas regiones pudieron abarcar regiones codificantes y no codificantes (Pollard et al. 2006) o no codificantes exclusivamente (Prabhakar et al. 2006; Bush y Lahn 2008; Lindblad-Toh et al. 2011). La conservación filogenética se estudió eligiendo un grupo de especies de mamíferos y/o vertebrados (pudiendo o no incluir a *Homo sapiens* en este paso), alineando sus genomas y aplicando algún método estadístico para cuantificar el nivel de conservación. Finalmente, de esos alineamientos, cada grupo eligió recuperar aquellas regiones genómicas que superaran algún valor de conservación determinado. De esta forma, se distinguieron regiones con una señal de selección purificadora fuerte y buenos candidatos a ser regiones funcionales donde las sustituciones podrían tener efectos biológicos.

2- Aplicación de un método estadístico para detectar regiones del genoma humano que hayan cambiado su tasa de sustitución nucleotídica. En esta segunda etapa se analizaron exclusivamente los alineamientos que en la etapa anterior resultaron satisfacer el nivel de conservación establecido. Los métodos y tests estadísticos utilizados para la identificación de las regiones con evidencia de evolución acelerada en el genoma humano fueron diversos. En algunos trabajos se calculó cuan probable era obtener los alineamientos observados para cada región dado un modelo de evolución por selección negativa (o de conservación, donde la tasa de sustitución era la misma, y baja, para todos los linajes) o de evolución acelerada exclusivamente en el linaje humano. Luego, se compararon los ajustes a ambos modelos y se determinó estadísticamente cuanto más probable eran los alineamiento observados dado uno u otro modelo (Pollard et al. 2006a; Lindblad-Toh et al. 2011). En otro trabajo, se estimó estadísticamente cuan sorprendente era observar la secuencia humana (posiciones conservadas y no conservadas) dado la tasa de evolución neutra (o de fondo) de esa región y el nivel de conservación filogenético de cada posición en el alineamiento (Prabhakar et al. 2006). Finalmente, un quinto trabajo estimó que cuatro sustituciones específicas de humano en las regiones conservadas de más de 100 pares de bases era estadísticamente significativo como señal de evolución acelerada (Bird et al. 2007). En resumen, el resultado general de este segundo paso fue la identificación de regiones genómicas con un cambio significativo en la tasa de evolución molecular exclusivamente en el linaje humano.

3- Curado de las regiones identificadas. Una vez identificadas las regiones genómicas con una tasa de evolución molecular significativamente acelerada en el linaje humano, se curaron (manualmente o recurriendo a filtros específicos) para eliminar aquellas regiones que presentaran errores en los alineamientos y/o que estuvieran presentes en regiones anotadas como pseudogenes (entre algunos de los filtros aplicados en estos trabajos). De esta forma, se obtuvieron listas de elementos acelerados altamente curados y ordenados por su valor de verosimilitud (o chi-cuadrado).

Es importante notar que estas metodologías excluyen del análisis a las regiones potencialmente regulatorias que hayan surgido *de novo* en el linaje humano o que no puedan ser detectadas por conservación, ya que una de las condiciones exigidas para todos los elementos es la conservación filogenética. Además, también quedan excluidas regiones no codificantes conservadas donde haya ocurrido relativamente poco cambio (no lo suficiente como para generar un cambio significativo en la tasa de evolución molecular de esa región), sin que esto implique que esos cambios no fueran importantes para la evolución morfológica de nuestra especie. Es por ello que estos trabajos son, en

teoría, capaces de recuperar una fracción del universo de regiones regulatorias con importancia potencial en la evolución regulatoria de nuestro genoma.

HIPÓTESIS DE TRABAJO

Teniendo en cuenta que solo un tercio de todas las regiones conservadas en el genoma humano codifican para proteínas y que existe evidencia que cambios linaje-específico en regiones regulatorias del genoma pueden ser los responsables de cambios morfológicos a nivel evolutivo, el estudio de regiones no codificantes con signos de evolución acelerada exclusivamente en el linaje humano podría proveer información importante para comprender las bases genéticas responsable de la evolución fenotípica de nuestra especie. En particular, la hipótesis de trabajo en este primer capítulo es que los *loci* con mayor densidad de elementos no codificantes con signos de evolución acelerada en el genoma humano tienen más probabilidades de haber contribuido con el surgimiento de novedades evolutivas exclusivas de nuestro linaje.

OBJETIVOS ESPECÍFICOS

- ✓ Realizar un análisis de la distribución genómica de regiones aceleradas exclusivamente en el linaje humano e identificar *loci* altamente enriquecidos con estos elementos. Para ello se hará uso de bases de datos públicas de elementos con signos de cambios en su tasa de evolución molecular exclusivamente en nuestro linaje (Pollard et al. 2006; Prabhakar et al. 2006; Bush y Lahn 2008; Lindblad-Toh et al. 2011).
- ✓ Estudiar el entorno genómico en el cual se encuentra la región con mayor densidad de elementos acelerados del genoma humano. Para ello, se analizará tanto los patrones evolutivos en regiones que codifican para proteínas, como también secuencias no codificantes presentes en este *locus*.
- ✓ Identificar si las sustituciones específicas de los humanos (en comparación con otros primates) presentes en el *locus* con mayor densidad de elementos acelerados del genoma humanos surgieron en tiempos profundos de la historia evolutiva de nuestro linaje (anteriores a la divergencia entre humanos modernos y Neandertales y Denisovanos), o son posteriores a este evento de divergencia. Para ello, se hará uso de los genomas actualmente disponibles de humanos arcaicos (Reich et al. 2010; Meyer et al. 2012), así como también bases de datos de polimorfismos actuales en las poblaciones humanas (Abecasis et al. 2010; Abecasis et al. 2012).

RESULTADOS

Análisis a escala genómica del agrupamiento de regiones no codificantes con evidencia de evolución acelerada en el genoma humano

Estudio de agrupamiento de regiones aceleradas en el genoma humano

Con el objetivo de identificar si existe alguna región genómica y/o unidad transcripcional altamente enriquecida en elementos acelerados en el linaje humano, se estudió la distribución a nivel genómico de estas regiones. En este estudio, se hizo uso de cuatro bases de datos públicamente disponibles de elementos acelerados en el linaje humano obtenidas con algoritmos bioinformáticos desarrollados para tal fin: 202 *Human Accelerated Regions* (HAR por sus siglas en inglés) (Pollard et al. 2006), 992 *Human Accelerated Conserved Non-Coding Sequences* (HACNS por sus siglas en inglés) (Prabhakar et al. 2006), 63 *Human Terminal Branch Elements* (HTBE por sus siglas en inglés) (Bush y Lahn 2008) y 543 *2x Human Accelerated Regions* (2xHAR por sus siglas en inglés) (Lindblad-Toh et al. 2011).

Existe una quinta base de datos de elementos no codificantes de evolución rápida en el linaje humano (Bird et al. 2007). Sin embargo, estos elementos no se incluyeron en el análisis de agrupamiento dado que, la metodología empleada para su identificación difiere sustancialmente de la metodología empleada por el resto de los grupos. La diferencia fundamental consiste en que no se aplicaron test estadísticos para la identificación de cambios en la tasa de evolución molecular de una región. Alternativamente, estos autores imponen un valor fijo de cambio a sus alineamientos (de cuatro sustituciones por 100 nucleótidos en regiones con un valor de conservación dado) y sostienen que este valor es suficiente para detectar aceleración en la tasa de evolución molecular (Bird et al. 2007).

Antes de comenzar con el estudio de agrupamiento, los intervalos genómicos originalmente expresados en coordenadas correspondientes al ensamblado del genoma humano en su versión NCBI35/hg17 (HARs y HACNS) se convirtieron manualmente a las coordenadas NCBI36/hg18, utilizando la herramienta BLAT disponible en www.genome.ucsc.edu. Esta conversión debió realizarse debido a que las diferentes versiones de ensamblado del genoma humano difieren en la localización de las regiones genómicas. De aquí en adelante, siempre se trabajó en coordenadas genómicas expresadas en NCBI36/hg18 de acuerdo con el *UCSC Genome Browser* (www.genome.ucsc.edu). Luego, los 1820 elementos acelerados identificados por los estudios antes

mencionados se concatenaron en una única tabla, donde los elementos solapados (total o parcialmente) provenientes de dos o más bases de datos fueron convertidos en un único intervalo genómico. De esta forma, se obtuvo una nueva base de datos con 1629 elementos no redundantes. Se nombró a estos elementos *Human Accelerated Elements* (HAEs por sus siglas en inglés), independientemente de la base de datos original de la cual provenía el elemento.

Con el objetivo de estudiar la distribución de los HAEs en el genoma humano, se adoptó un doble abordaje. Por un lado, se analizó si existe algún/as unidad/es transcripcional/es humana/s particularmente enriquecida/s en secuencias aceleradas (lo cual la/s convertiría/n en buen/as candidata/s para posteriores análisis funcionales). Para ello, se obtuvo la base de datos de genes humanos altamente curados *RefSeq* desde el sitio www.genome.ucsc.edu (Pruitt, Tatusova y Maglott 2005). Para eliminar redundancias, todos los intervalos de las unidades transcripcionales *RefSeq* parcial o totalmente solapados fueron convertidos en un único intervalo genómico. De esta forma, se generó una base de datos conteniendo 19.897 unidades transcripcionales humanas no redundantes (UT, intervalo genómico desde el primer al último exón de cada *RefSeq*). Sin embargo, la base de datos de unidades transcripcionales no redundantes así generada excluye a las regiones intergénicas. Reconociendo que podrían existir *loci* particularmente enriquecidos en HAEs en regiones no exclusivamente génicas (por ejemplo, incluyendo más de un gen o en desiertos génicos), el segundo abordaje consistió en analizar el agrupamiento de HAEs por mega base (Mb; 1.000.000 pares de bases) de genoma humano. Para ello, se generaron cuatro bases de datos (aquí llamadas “ventanas”) conteniendo intervalos genómicos de 1 Mb de largo que abarcan al genoma humano completo (~ 3000 intervalos genómicos por ventana). Cada una de estas ventanas se generó de forma tal que los intervalos sucesivos estuvieran desplazados 250 kpb respecto del mismo intervalo entre ventanas, generando solapamientos de 750 kpb entre bases de datos contiguas. Se fijó el largo de los intervalos genómicos en 1 Mb porque las regiones regulatorias de los genomas de los vertebrados pueden actuar sobre grandes regiones, incluso alejadas 1-2 Mb de su gen blanco (Vavouri y Lehner 2009). Entonces, se consideró que esta distancia sería potencialmente adecuada para identificar agrupamientos de regiones no codificantes en el genoma humano con probable actividad regulatoria junto con su/s gen/es blanco/s.

Utilizando las herramientas bioinformáticas públicamente disponibles en el sitio *Galaxy* (<http://main.g2.bx.psu.edu/>), se asignó cada HAE a/los intervalo/s genómico/s de 1 Mb o UT que lo contenía total o parcialmente. Finalmente, se generaron dos tipos de histogramas. Por un lado, se obtuvo la distribución de frecuencias de HAEs por Mb para cada una de las cuatro ventanas de intervalos genómicos. Adicionalmente, se obtuvo la distribución de frecuencias de HAEs por

unidad transcripcional. En la figura 4, se resume la metodología aplicada para estudiar la distribución de HAEs en el genoma humano.

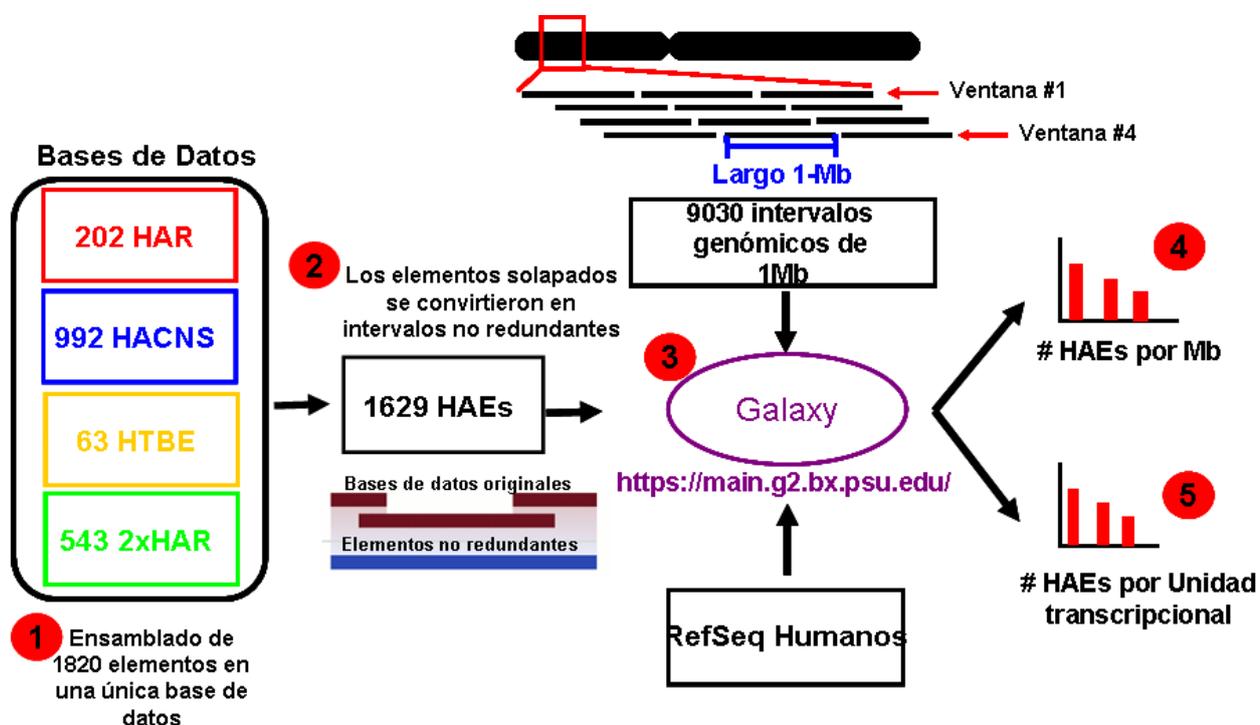


Figura 4: Agrupamiento de HAEs en el genoma humano. Se muestra un esquema representativo de los 5 pasos metodológicos utilizados para obtener el agrupamiento de HAEs en el genoma humano. (1) Concatenamiento de las bases de datos de elementos acelerados en el genoma humano. (2) Eliminación de redundancias entre bases de datos y generación de la base de datos de elementos acelerados en el genoma humano no solapados (HAEs). (3) Cruce de la base de datos de HAEs con los intervalos *RefSeq* y con ventanas genómicas de 1 Mb de longitud. Generación del histograma de HAEs por (4) Mb o (5) *RefSeq*.

Baja redundancia entre las bases de datos de HAEs

Al generar la base de datos de HAEs, la primera observación fue la baja redundancia entre las bases de datos originales de elementos acelerados aquí utilizados (nótese que, de 1820 elementos totales, solamente 191 elementos aparecieron en dos o más bases de datos; figura 5, A). En particular, se halló que un solo elemento acelerado fue recuperado por los cuatro trabajos (figura 5, A). Existen varias causas que pueden originar esta baja superposición de elementos entre estos trabajos; aun cuando el objetivo principal detrás de estos estudios fue el mismo: encontrar secuencias altamente conservadas en el genoma humano con signos de cambio en su tasa de evolución molecular exclusivamente en el genoma humano (Pollard et al. 2006; Prabhakar et al. 2006; Bush y Lahn 2008; Lindblad-Toh et al. 2011). Como se mencionó en la introducción, cada grupo utilizó su propia metodología. Entre las principales diferencias metodológicas se encuentran: la profundidad

de conservación de las secuencias requerida (que a su vez depende de las especies incluidas en el análisis) (Boffelli, Nobrega y Rubin 2004), los filtros aplicados para el curado de los alineamientos y las pruebas estadísticas empleadas para detectar cambios en la tasa de evolución molecular, entre otros (Capra et al. 2013a). La influencia de estas diferencias metodológicas es particularmente evidente cuando se compara la superposición entre las bases de datos de HARs y 2xHARs, generadas por el mismo grupo de investigación pero utilizando diferente número de especies para definir conservación (cuatro frente a veintinueve especies respectivamente) (Pollard et al. 2006a; Lindblad-Toh et al. 2011). En este caso, solamente 37 elementos fueron recuperados por ambos estudios (figura 5, A). No obstante, las diferencias metodológicas antes mencionadas son solo algunas de las causas que pueden explicar la baja redundancia, pero no las únicas. Sin embargo, hasta al momento, no hay razones para sostener que los resultados obtenidos por uno u otro trabajo son mejores en cuanto a su capacidad de predecir regiones regulatorias del genoma que cambiaron su actividad exclusivamente en el linaje humano.

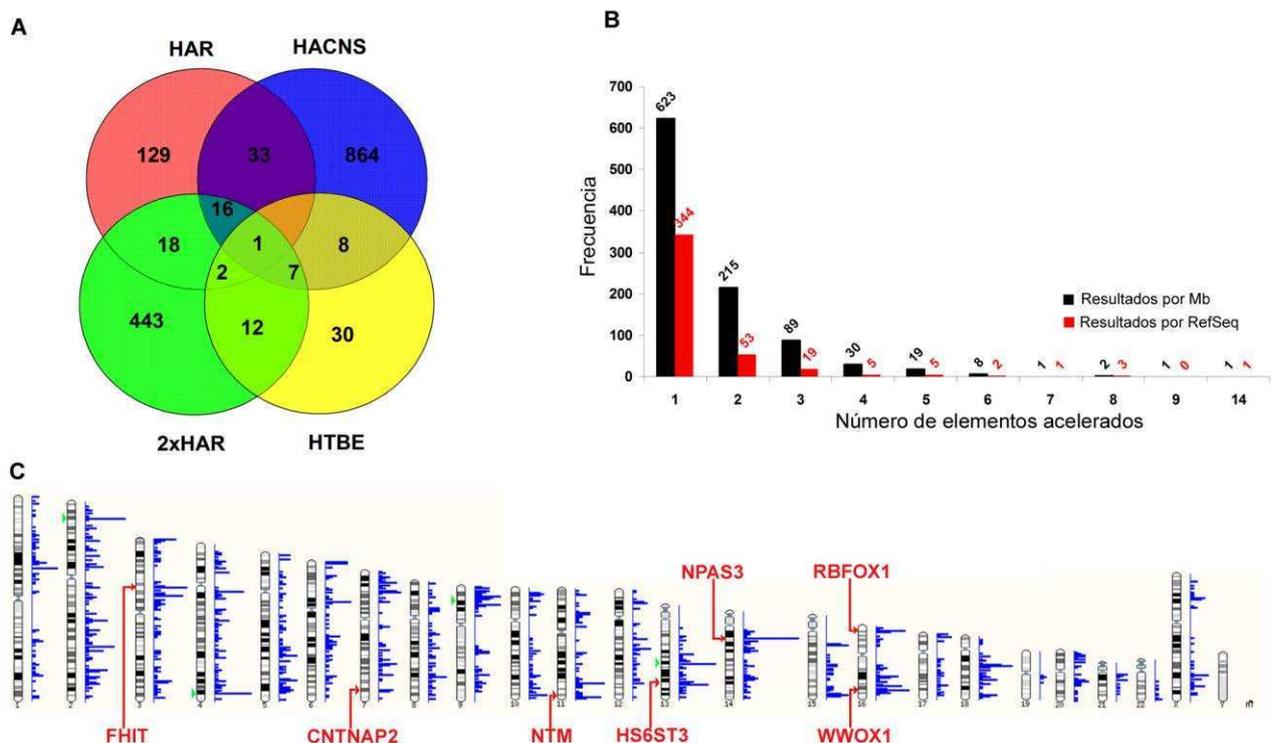


Figura 5: Agrupamiento de HAEs en el genoma humano. (A) Diagrama de Venn mostrando el solapamiento de los HAEs provenientes de las diferentes bases de datos originales HARs (Pollard, Salama, King, et al. 2006; Pollard, Salama, Lambert, et al. 2006); HACNS (Prabhakar et al. 2006); HTBE (Bush y Lahn 2008); y 2xHARs (Lindblad-Toh et al. 2011). Adicionalmente, hay 65 elementos compartidos entre las bases de datos 2XHAR y HACNS y 3 elementos compartidos por las bases de datos de HARs y HTBE que no se muestran en el diagrama. El número total de HACNS y 2XHARs expresados en el diagrama es mayor al número de elementos publicados en las bases de datos

originales debido a solapamientos múltiples en los *loci* HACNS56/2, HAR20 y HACNS11. (B) Histograma de frecuencia mostrando la distribución de HAEs por gen humano o por intervalos genómicos de 1 Mb (solo se muestran los resultados para la ventana 2). (C) Esquema donde se muestra con barras azules la densidad de HAEs a lo largo de los cromosomas humanos generado con el *density track* disponible en el sitio Ensembl (www.ensembl.org). Con flechas rojas se muestra la ubicación de las UT con 6 o más HAEs. Con flechas verdes se indica la ubicación de los intervalos genómicos de 1 Mb con 8 o más HAEs en caso que no hayan sido indicados anteriormente con flechas rojas.

***NPAS3* se localiza en el *locus* con más HAEs del genoma humano**

Por otro lado, el análisis de agrupamiento antes descrito reveló que, la gran mayoría de las UT o intervalos genómicos de 1 Mb no contienen HAEs. De los 19.897 UT no redundantes anotadas en el genoma humano, solo 433 (2.5%) contienen al menos un HAE (figura 5, B). De las UT conteniendo HAEs, la gran mayoría contienen 1, 2 o 3 elementos (416 UT de 433, 96%, figura 5, B). Solo 7 genes humanos contienen 6 o más HAEs (figura 5, B y C). De forma similar, promediando los resultados obtenidos en cada ventana de análisis se encontró que aproximadamente 0,5% de los intervalos genómicos de 1 Mb contienen 6 o más HAEs (figura 5, B y C; tabla 1). Llamativamente, se encontró que una única UT, la correspondiente al gen *neuronal PAS domain-containing protein 3* (*NPAS3*), contiene la cantidad excepcional de 14 HAEs; el mayor agrupamiento de HAEs para un solo gen en el genoma humano (figura 5, B y C; tabla 1). *NPAS3* es seguido por los genes *WW domain-containing protein* (*WWOX*; chr16: 76691052–77804065), *fragile histidine triad proteína* (*FHIT*; chr3: 59710076–61212173), y *RNA binding protein, fox-1 homolog* (*RBFOX1*; chr16: 6009133–7703341), cada uno con 8 HAEs (figura 5, C; tabla 1).

Como las unidades transcripcionales de *NPAS3*, *WWOX*, *FHIT* y *RBFOX1* son particularmente grandes, abarcando 0.9, 1.1, 1.5, y 1.7 Mb respectivamente, se estudió si los genes que poseen unidades transcripcionales grandes están enriquecidos en HAEs. Aquí se considera como “genes gigantes” a aquellos genes cuyas UT abarcan más de 800 kpb. De los 92 genes gigantes presentes en el genoma humano, 63 (68%) contienen al menos un HAE. Esto indica que los genes gigantes tienen una mayor probabilidad de contener un elementos acelerado (tabla 2, test exacto de fisher $P < 0,00001$). De hecho, el número de pares de bases conservadas (definidas por el alineamiento de 17 genomas de vertebrados utilizando el programa PhastCons) (Siepel et al. 2005) por par de bases de UT depende del largo del gen (tabla 2, test exacto de Fisher $P < 0,0001$), sugiriendo que los genes gigantes acumulan más pares de bases conservadas no exónicas por par de bases de gen.

Tabla 1: Genes y regiones genómicas conteniendo 6 o más HAEs

# HAEs	Posición RefSeq (NCBI36/hg18)	Símbolo Gen	HAEs por kpb
14	chr14:32,478,210-33,340,702	<i>NPAS3</i>	1,62
8	chr3:59,710,076-61,212,173	<i>FHIT</i>	0,53
	chr16:6,009,133-7,703,341	<i>RBFOX1</i>	0,47
	chr16:76,691,052-77,804,065	<i>WWOX</i>	0,72
7	chr11:130,745,581-131,711,926	<i>NTM</i>	0,72
6	chr7:145,444,386-147,749,021	<i>CNTNAP2</i>	0,26
	chr13: 95,541,093-96,289,813	<i>HS6ST3</i>	0,80

# HAEs	Posición Mb (NCBI36/hg18)	Genes en el intervalo genómico	HAEs por kpb
14	chr14:32,250,032-33,250,032	<i>AKAP6, NPAS3</i>	1,4
12	chr14:32,500,032-33,500,032	<i>NPAS3, EGLN3</i>	1,2
9	chr14:32,000,032-33,000,032	<i>AKAP6, NPAS3</i>	0,9
	chr9:16,000,016-17,000,016	<i>BNC2</i>	0,9
	chr2:21,250,021-22,250,021	-	0,9
	chr14:32,750,032-33,750,032	<i>NPAS3, EGLN3</i>	0,9
8	chr2:21,000,021-22,000,021	<i>APOB, LOC645949</i>	0,8
	chr4:182,000,182-183,000,182	<i>LINC00290</i>	0,8
	chr4:182,250,182-183,250,182	<i>LINC00290</i>	0,8
	chr13:70,500,070-71,500,070	<i>DACHI</i>	0,8
	chr16:6,500,006- 7,500,006	<i>RBFOX1</i>	0,8
	chr2:21,500,021-22,500,021	<i>LOC645949</i>	0,8
	chr9:15,750,015-16,750,015	<i>C9orf93, BNC2</i>	0,8
	chr13:95,750,095-96,750,095	<i>HS6ST3, OXGRI, MBNL2</i>	0,8
	chr16:6,750,006-7,750,006	<i>RBFOX1</i>	0,8
	chr3:6,0250,060-61,250,060	<i>FHIT</i>	0,8

Por otro lado, el análisis de agrupamiento de HAEs por Mb mostró que el intervalo genómico que contiene mayor número de HAEs, con 14 elementos, es aquel que contiene parte de la UT de gen *NPAS3* (tabla 1). Además, el segundo intervalo genómico con más elementos acelerados, en este caso con 12 HAEs, también incluye parte de la UT de *NPAS3*. Adicionalmente, hay cuatro intervalos genómicos con 9 HAEs cada uno y dos de ellos nuevamente contienen parte de la UT de *NPAS3*. Entre los restantes dos intervalos, uno está ubicado en el cromosoma 9 (chr9: 16000016–17000016) y contiene al gen *Basomucilin 2 (BNC2)* y la cuarta región está ubicada en una zona sin genes en el cromosoma 2 (figura 5, C; tabla 1). Finalmente, hay 10 intervalos de 1 Mb que contienen cada uno 8 HAEs (figura 5, B y C; tabla 1).

Para analizar en más detalle la densidad de HAEs por gen, se dividió el número de HAEs por el largo de la unidad transcripcional de cada gen conteniendo HAEs. El análisis indicó que *NPAS3* contiene 1,6 HAEs por cada 100 kpb de largo, siendo la región genómica con mayor densidad de elementos acelerados entre aquellos genes con más número de elementos acelerados del genoma humano (tabla 1).

Tabla 2: Análisis de HAEs y elementos conservados en diferentes regiones del genoma humano.

	Regiones intergénicas	UT	UT <800kpb	UT >800kpb
Número intervalos	19.921	19.897	19.805	92
Número total de pares de bases	1.875.206.043	1.204.963.940	1.097.300.376	107.663.564
Número total de pares de bases no exónicas^{&}	1.875.206.043	1.181.170.047	1.073.892.604	107.277.443
Número de regiones conservadas no exónicas^{&}	180.939	136.593	119.631	16.962
Número de pares de bases en regiones conservadas no exónicas^{&}	35.334.189	24.643.651	21.501.333	3.142.318
% de pares de bases conservadas no exónicas	1,88	2,09	2,00	2,92
Número de intervalos genómicos con HAEs	630	433	370	63
Número de intervalos genómicos sin HAEs	19.291	19.464	19.435	29
Número de HAEs	1020	609	446	163

[&]Las regiones conservadas que solaparan con exones se filtraron utilizando el *track spliced EST* presente en el sitio de UCSC (www.genome.ucsc.edu); [§]Definidas por el alineamiento de 17 genomas de vertebrados (Siepel et al. 2005) y con un valor de conservación mayor a 400; En todos los casos se trabajó con unidades transcripcionales (UT) no redundantes (NCBI36/hg18).

Este estudio de agrupamiento de HAEs indica que *NPAS3* se localiza en la región genómica de mayor densidad de HAEs de todo el genoma humano. En particular, se encontró que la UT de *NPAS3* contiene 14 HAEs no redundantes. De ellos, 6 son HARs (HAR21, HAR89, HAR96, HAR173, HAR189 y HAR202), 5 son HACNS (HACNS96, HACNS221, HACNS490, HACNS553 y HACNS658) y 4 son 2xHAR (2xHAR122, 2xHAR142, 2xHAR157 y 2xHAR2). Sin embargo, los elementos HACNS490 y HAR189 se encuentran solapados, reduciendo el número de elementos independientes de 15 a 14 (figura 6, B). Adicionalmente, se encuentra un 15^{to} elemento, el

HAR144, localizado aprox. 200 kpb río arriba de *NPAS3*, en un intrón del gen *A kinase anchor protein 6 (AKAP6)* (figura 6, B y C). Todos los HAEs presentes en *NPAS3* son elementos no codificantes presentes en los intrones de este gen y están altamente conservados (tabla suplementaria 3). Estas características sugieren que los HAEs antes mencionados podrían funcionar como elementos regulatorios de la expresión génica.

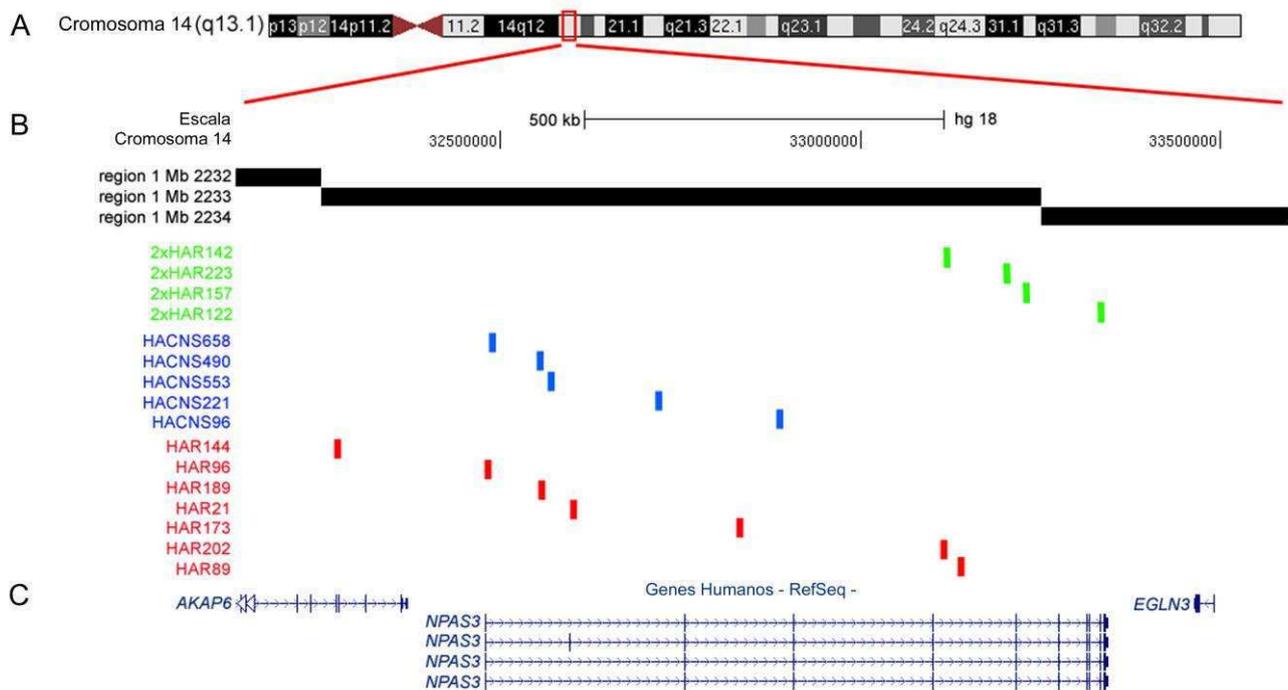


Figura 6: (A). Localización del *locus* de *NPAS3* en el cromosoma 14 del genoma humano (NCBI36/hg18). (B) Ventana de 1.1 Mb de longitud de acuerdo con el *UCSC Genome Browser* (www.genome.ucsc.edu) mostrando la localización de los intervalos genómicos de 1 Mb correspondientes a la ventana 2 y los HAR, HACNS y 2xHARs presentes en *NPAS3*. (C) Genes de la base de datos RefSeq en la ventana de 1.1 Mb.

Estudio de las regiones codificantes de *NPAS3*

Con el objetivo de estudiar el contexto genómico en el cual el gen *NPAS3* está inmerso y analizar el patrón evolutivo de este gen en el linaje de los mamíferos, el Lic. Francisco Pisciotano, miembro del laboratorio, realizó un profundo análisis comparativo de las regiones codificantes y no codificantes de este gen. Estos estudios demostraron que, si bien los intrones de *NPAS3* contienen una acumulación inusual de elementos conservados con signos de rápido cambio en el linaje humano, la secuencia de proteína codificada por este gen está muy conservada entre los mamíferos. En promedio, la identidad de secuencia de *NPAS3* entre especies de mamíferos es del 95% y este porcentaje se incrementa dentro de los primates donde alcanza un 98,7% (tabla suplementaria s1).

En particular, la identidad de la proteína entre chimpancé y humanos es del 99,8%, con solo dos sustituciones aminoacídicas (A559P y A787G) en 940 residuos. Además, se estudió las secuencias codificantes de este gen con el objetivo de indentificar signos de procesos de evolución acelerada o positiva en el linaje humano y de los chimpancés utilizando el programa PAML. Los resultados mostraron que el patrón evolutivo hallado en las secuencias codificantes de *NPAS3* es mejor explicado por un proceso de evolución purificadora, que por procesos de evolución acelerada o evolución positiva (tabla suplementaria 2).

Adicionalmente, se estudiaron comparativamente las regiones codificantes de *NPAS3*, las regiones no codificantes conservadas presentes en este *locus* (obtenidas al comparar los genomas de 17 especies de vertebrados) (Siepel et al. 2005) y las regiones conservadas con signos de evolución acelerada (HAEs) usando el programa phyloP (<http://compugen.bsccb.cornell.edu/phast/>). En este estudio se consideraron todas las secuencias presentes entre la última base del gen *AKAP6* y la primera base del gen *EGNL3* (los genes río arriba y río debajo de *NPAS3* respectivamente). Luego, se generaron alineamientos con las secuencias de los primates humano (*Homo sapiens*), chimpancé (*Pan troglodytes*), orangután (*Pongo pygmaeus*) y mono rhesus (*Macaca mulatta*) a lo largo de cuatro particiones de este *locus*: secuencias codificantes, regiones no codificantes conservadas, regiones HAEs y regiones no codificantes no conservadas. Este análisis mostró que las secuencias codificantes están altamente conservadas ($P = 0,001$ en la rama humana y $P=3,54 \times 10^{-28}$ en las especies chimpancé, orangután y mono rhesus) y sin signos de evolución acelerada en el linaje humano ($P=0,99$), coincidiendo con los resultados obtenidos con el programa PAML. Por otro lado, las regiones HAEs tienen una mayor probabilidad de ajustar a un modelo de evolución acelerada en el linaje humano ($P=1,47 \times 10^{-4}$) que a un modelo de evolución purificadora ($P=0,99$). Por otro lado, estas mismas regiones muestran signos de alta conservación en los linajes de primates no humanos ($P=1,16 \times 10^{-65}$). Además, las regiones conservadas que no contienen HAEs tienen una probabilidad más alta de ajustar a un modelo de evolución purificadora que a un modelo de evolución acelerada para todo el grupo de primates ($P=2,29 \times 10^{-97}$). Por último, las regiones no codificantes no conservadas no se ajustan a un modelo de conservación ni en el linaje humano ($P=0,82$) ni en los linajes de primates no humanos ($P=0,99$), mostrando signos de aceleración ($P=7,5 \times 10^{-7}$) y evolución neutral ($P=0,187$) en el linaje humano y de primates no humanos respectivamente.

Polimorfismos en los HAEs presentes en intrones de *NPAS3*

Gracias a la secuenciación de múltiples genomas humanos fue posible detectar sitios polimórficos de alta frecuencia en diferentes poblaciones humanas. Por ejemplo, el proyecto 1000 genomas secuenció, en su primera fase, 1.092 genomas humanos provenientes de 14 poblaciones de África,

Europa, Este de Asia y América (Abecasis et al. 2010; Abecasis et al. 2012). Gracias a este análisis se detectaron sitios polimórficos presentes en al menos una población humana en todos los HAEs de *NPAS3* (considerando también las secuencias conservadas que contienen a los HAEs) (tabla 3). En particular, existen al menos cuatro regiones HAEs de *NPAS3* que tienen polimorfismos registrados para los sitios con sustituciones humano específicas (HAR189, HAR202, HACNS658 y 2xHAR142) (tabla 3). De los 5 sitios polimórficos en poblaciones humanas ubicados en las posiciones registradas como variantes específicas de humanos en estos 4 elementos, en dos de esos polimorfismos todas las variantes halladas son derivadas respecto a la base homóloga ancestral inferida a partir de los alineamientos con genomas de otros mamíferos (G/T en humanos y A ancestral en polimorfismo rs145548775 del HAR189 y C/G en humanos y T ancestral en el polimorfismo rs114570257 del 2xHAR142). Para los otros tres sitios polimórficos en bases específicas de humanos en los HAEs intrónicos de *NPAS3*, una de las variantes polimórficas corresponde con la base ancestral inferida (C/T en humanos y T ancestral en el polimorfismo rs10134006 del 2xHAR142, A/T en humanos y T ancestral en el polimorfismo rs79570848 del HAR202 y T/C en humanos y C ancestral en el polimorfismo rs10149308 del HACNS658). Las restantes sustituciones específicas de los humanos presentes en los HAEs de *NPAS3* resultan ser no polimórficas.

Adicionalmente, la reciente publicación del primer genoma de Neandertal de alta calidad, permitió revisar la datación del origen de las sustituciones específicas de humanos en los HAEs (Prufer et al. 2014). La gran mayoría de las sustituciones específicas de los humanos resultaron estar compartidas con los genomas de los humanos arcaicos disponibles en la actualidad (Neandertales y Denisovanos). Sin embargo, se encontraron algunas excepciones. Tres pares de bases presentes en HAEs de *NPAS3* fueron anotadas como derivadas en humanos (diferentes a las bases ortólogas en Neandertales y Denisovanos). Dos de estas son polimórficas en las poblaciones humanas actuales (rs10149308 en HACNS658 y rs79570848 en HAR202) y una tercera es fija (base chr14:34160070 en 2xHAR157).

Tabla 3: Polimorfismos actuales en HAEs de *NPAS3* y secuencias en genomas de humanos arcaicos.

HAE	Localización genómica	Largo (bp)	SP	Sust Hs	D/A	PHs	Número polimorfismo	Humanos Arcaicos	Estado
HAR96	chr14: 324,86,645-32,487,168	524	7	2	G/A	-		G	
					T/A	-		T	
HACNS658	chr14: 32,487,503-32,487,837	335	6	6	C/T	-	rs10149308	C	P
					T/C	T/C		T/C	
					G/A	-		G	
					A/T	-		A	
					T/C	-		T	
C/T	-	C							

HAR189/ HACNS490	chr14: 32,558,123-32,558,864	742	10	6	G/C	-	rs145548775	G	
					C/T	-		C	
					C/T	-		C	
					G/T	-		G	
					G/A	G/T		G	
A/G	-	A							
HACNS553	chr14: 32,566,093-32,566,346	254	5	4	G/A	-	G		
					G/T	-	G		
					T/A	-	T		
					G/A	-	G		
HAR21	chr14: 32,596,637-32,596,862	226	2	3	T/G	-	T		
					G/A	-	G		
					T/C	-	T		
HACNS221	chr14: 32,718,813-32,719,042	230	3	5	T/C	-	T		
					G/T	-	G		
					G/T	-	G		
					T/C	-	T		
					C/T	-	C		
HAR173	chr14: 32,830,590-32,830,773	184	2	2	C/A	-	C		
					G/A	-	G		
HACNS96	chr14: 32,888,897-32,889,211	315	6	6	G/A	-	G		
					C/A	-	C		
					C/T	-	C		
					G/A	-	G		
					C/T	-	C		
					G/T	-	G		
HAR202	chr14: 33,115,156-33,115,595	440	8	5	C/T	-	C		
					A/T	A/T	rs79570848	T	F
					G/A	-	G		
					T/C	-	T		
					G/T	-	G		
2xHAR142	chr14: 33,118,315-33,118,577	263	4	6	G/A	-	G		
					C/T	C/G	rs114570257	C	
					T/G	-	T		
					G/A	-	G		
					G/A	-	G		
					C/T	C/T	rs183390399	C	
HAR89	chr14: 33,135,162-33,135,937	776	7	3	G/A	-	G		
					G/A	-	G		
					A/G	-	A		
2xHAR223	chr14: 33,199,815-33,200,914	1100	13	8	T/C	-	T		
					C/T	-	C		
					C/T	-	C		
					G/A	-	G		
					C/T	-	C		
					C/G	-	C		
					C/T	-	C		
					T/G	-	T		
2xHAR157	chr14: 33,229,602-33,230,232	631	5	10	C/T	-	C		
					C/T	-	C		
					C/T	-	C		
					G/A	-	G		
					A/C	-	C	F	
					G/C	-	G		
					C/A	-	C		
					A/C	-	A		
					T/C	-	T		
G/C	-	G							
2xHAR122	chr14: 33,331,419-33,331,887	469	9	4	C/T	-	C		
					C/T	-	C		
					G/A	-	G		
					G/A	-	G		

Nota: Las coordenadas genómicas, el largo y el número de sustituciones humanas específicas corresponden a la región conservada completa que contiene al HAE (Siepel et al. 2005). Abreviaturas: SP, número de sitios polimórficos en los HAEs en poblaciones humanas actuales de acuerdo al 1000

genomes project phase 1 (track disponible en www.ucsc.edu/) (Abecasis et al. 2010; Abecasis et al. 2012). Sust Hs, número de sustituciones específicas de los humanos dentro de los HAEs. D/A, base derivada /base ancestral inferida a partir del alineamiento de múltiples genomas de mamíferos en las posiciones que cambiaron exclusivamente en el linaje humano. PHs, alelos en sitios polimórficos en poblaciones humanas actuales en las posiciones que cambiaron exclusivamente en el linaje humano. En humanos arcaicos se indica la base correspondiente inferida en el genoma de humanos arcaicos a partir de la secuenciación de múltiples genomas de Neandertales y Denisovanos (Reich et al. 2010; Meyer et al. 2012; Prufer et al. 2014). En recuadros amarillos se indica cuando la base en esta posición fue anotada como derivada en humanos modernos respecto a la base inferida en los humanos arcaicos. En condición se indica si la base derivada en humanos fue detectada en condición polimórfica (P) o fija (F) en los genomas de humanos arcaicos (Prufer et al. 2014).

Capítulo II

Análisis funcional de regiones de evolución acelerada en el linaje humano

INTRODUCCIÓN ESPECÍFICA

Caracterización funcional de los HAEs

Uno de los desafíos cuando se trabaja con HAEs es la caracterización funcional de los mismos. Este objetivo es particularmente complejo dado que la gran mayoría de estos elementos se encuentran en regiones del genoma no caracterizadas funcionalmente y que la conservación por si misma es incapaz de predecir una función (Noonan y McCallion 2010). Sin embargo, varios grupos de investigación han tratado de estudiar estos elementos a nivel funcional.

Dado que los HAEs son secuencias altamente conservadas entre los mamíferos y que mayoritariamente se ubican en regiones que no codifican para proteínas, una de las hipótesis es que actuarían como elementos regulatorios (Capra et al. 2013a). Por ejemplo, estudios funcionales realizados sobre el HACNS1 (también denominado HAR2), el HAE con la mayor tasa de evolución molecular del genoma humano (Prabhakar et al. 2006), encontraron que este elemento es capaz de actuar como un *enhancer* activo durante el desarrollo embrionario. Adicionalmente, encontraron que el HACNS1 tiene un dominio de expresión más amplio comparado con los elementos ortólogos de chimpancé y mono rhesus (ganando expresión en los miembros en desarrollo) (Prabhakar et al. 2008). Sin embargo, aún se desconoce el gen blanco del HACNS1 y/o el impacto funcional del cambio en el patrón de expresión dirigido por este *enhancer*.

Por otro lado, también se realizaron estudios funcionales sobre el HAR1. Este elemento resultó ser un gen de ARN que se expresa durante el desarrollo, principalmente en las neuronas de Cajal-Retzius en la neocorteza. Además, se encontró que el HAR1 (HAR1F) se co-expresa con la proteína reelina, una glicoproteína de matriz extracelular secretada por las neuronas de Cajal-Retzius y que es fundamental para la especificación de las capas de la corteza (Pollard et al. 2006b).

En la actualidad existe evidencia adicional que indica que muchas de las regiones no codificantes con una tasa de sustitución alta en el linaje humano exclusivamente tienen la capacidad de actuar como *enhancers* durante el desarrollo embrionario. Por ejemplo, un estudio reciente mostró que 24 HAEs (de entre 29 analizados) resultaron ser capaces de activar la expresión de un gen reportero en ensayos con ratones transgénicos. Además, 5 de estos *enhancers* mostraron una actividad diferencial entre los elementos humano y chimpancé, sugiriendo que podrían haber participado en la adquisición de nuevos dominios de expresión en alguno (u ambos) de estos linajes (Capra et al. 2013a).

Estos resultados sugieren que los HAEs serían una rica colección de secuencias en las cuales buscar las bases genéticas de cambios específicos de los humanos en la regulación de los genes (Capra et al. 2013a). Adicionalmente, la alta densidad de HAEs en el *locus* de *NPAS3* sugiere que este gen pudo haber adquirido cambios regulatorios en el linaje de los humanos. Los caracteres morfológicos que pudieron verse afectados por la alta tasa de evolución experimentada por los HAEs presentes en esta región dependerá fuertemente de la función desarrollada por este gen.

Del genoma al estudio de un gen específico

Como adelanta el título, esta tesis se concentra en el estudio del gen *NPAS3*. *NPAS3* es un gen que codifica para un factor de transcripción de la familia *basic helix-loop-helix, period/aryl hydrocarbon receptor/single minded* (bHLH-PAS) (Brunskill et al. 1999). Las proteínas miembro de esta familia son todos reguladores críticos de la expresión génica, capaces de responder frente a diversos estímulos ambientales (incluida luz, oxígeno, voltaje y potencial redox) y de cumplir múltiples roles durante el desarrollo del sistema nervioso, las oscilaciones circadianas y el desarrollo de las tráqueas, entre otras (Massari y Murre 2000; Kewley, Whitelaw y Chapman-Smith 2004).

Se pueden reconocer dos grandes familias de factores de transcripción bHLH/PAS: los miembros de la clase I son factores que no pueden ni homodimerizar ni heterodimerizar con otros factores de la clase I. Entre este grupo se encuentran *NPAS3*, las proteínas *aryl hydrocarbon receptor (AhR)*, factores respondedores a hipoxia (*HIF; HIF-1, HIF-2, y HIF-3*) y proteínas de la familia *single minded (SIM1 y SIM2)*. Para formar complejos activos, esta primera clase de proteínas debe dimerizar con proteínas de la clase II. Por otro lado, la clase II de las bHLH/PAS pueden homo y heterodimerizar promiscuamente. Uno de los miembros de esta familia es la proteína *carbon receptor nuclear translocator (ARNT)* (Lindebro 1995; Wang et al. 1995; Ema et al. 1996) (figura 7,A).

Todos los miembros de la familia de proteínas bHLH/PAS se caracterizan por tener un dominio N-terminal rico en aminoácidos básicos por donde la proteína reconoce las secuencias blanco y se unen al ADN (*basic*) adyacente a un dominio de dimerización primario (HLH). Tanto el dominio básico como el HLH son necesarios para la formación de complejos activos sobre el ADN (Bacsi 1996). Además, estos factores de transcripción tienen una o dos secuencias repetitivas degeneradas de aproximadamente 130 aminoácidos cada una (PAS-A y PAS-B) (figura 7, A). Todas las proteínas de la superfamilia bHLH/PAS deben dimerizar para formar complejos funcionales sobre el ADN. En este punto, el dominio PAS juega un rol fundamental. Por un lado, los dominios PAS

son capaces de censar cambios ambientales y traducir estos cambios en modificaciones en el patrón de interacción de la proteína. De esta forma, los dominios PAS brindan una interfase de interacción secundaria y determinan con que pareja podrá dimerizar. Finalmente, hacia el extremo C-terminal, los miembros de esta familia pueden contener diversos motivos de transactivación responsables de la modulación de la expresión de sus genes blanco (Kewley, Whitelaw y Chapman-Smith 2004; Pickard et al. 2006) (figura 7, B).

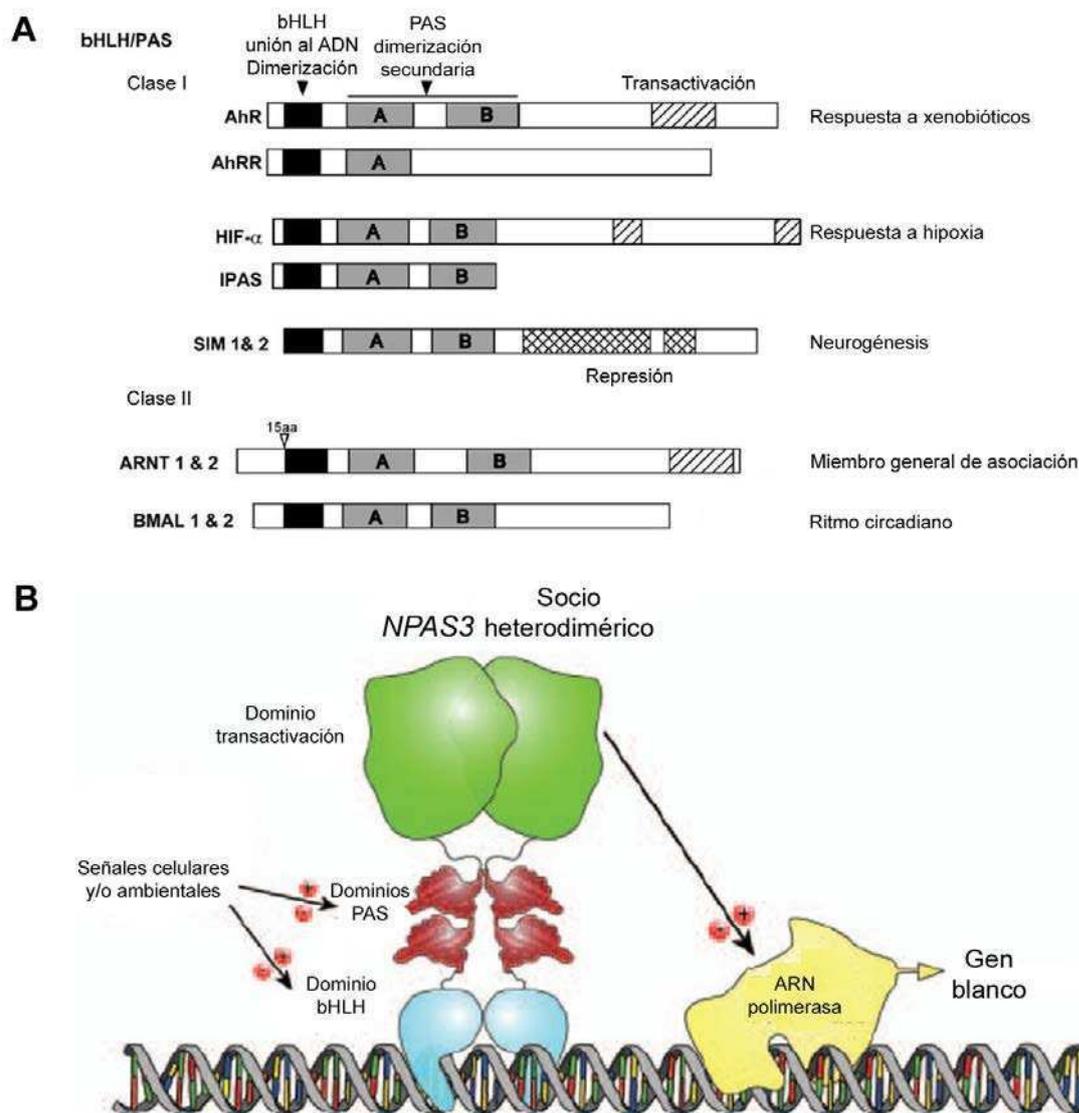


Figura 7: (A) Representación esquemática de la estructura de algunos de los factores de transcripción de la familia bHLH-PAS. (B) Esquema que representa el conocimiento actual de la función de *NPAS3*. En el dibujo se muestra a *NPAS3* unida a su secuencia blanco sobre el ADN formando un heterodímero con su proteína socio. Ambas proteínas modulan conjuntamente la expresión de sus genes blancos a través de la interacción con la polimerasa de ARN tipo II. El esquema hace hincapié en la capacidad potencial de *NPAS3* de responder frente a ciertos estímulos ambientales aún no identificados. Modificado de (Kewley, Whitelaw y Chapman-Smith 2004; Pickard et al. 2006).

Actualmente, aparte de *NPAS3*, se conocen otros tres miembros de la subfamilia NPAS: *NPAS1*, un gen cuya disfunción se asocia con el desarrollo de esquizofrenia (Erbel-Sieler et al. 2004), *NPAS2*, que participa de procesos relacionados con la memoria y las oscilaciones circadianas (Garcia et al. 2000; Reick et al. 2001) y *NPAS4*, que participa del desarrollo del sistema nervioso (Lin et al. 2008). La comparación de las secuencias aminoacídicas de los miembros de la familia de proteínas bHLH/PAS muestra que *NPAS1* y *NPAS3* están íntimamente relacionados, mostrando la mayor identidad de secuencia dentro de los integrantes del subgrupo de proteínas NPAS (Yan et al. 2014) (figura 8). Sin embargo, una diferencia notable entre estos dos genes está en el largo de sus unidades transcripcionales: los once exones de *NPAS1* están localizados en un segmento de 20 kpb en el cromosoma 7 del ratón, mientras que los exones de *NPAS3* están distribuidos en un *locus* genómico que abarca 791 kpb del cromosoma 12 de ratón (correspondiente al cromosoma 14 humano) (Erbel-Sieler et al. 2004).

***Trachealless*, el gen ortólogo de *NPAS3* en mosca**

Por otro lado, los genes parálogos en vertebrados *NPAS3* y *NPAS1* serían aparentemente ortólogos del gen *Trachealless* (*Trh*) de *Drosophila melanogaster*, compartiendo 45% y 42% de identidad de secuencia aminoacídica con *Trh*, respectivamente (Brunskill et al. 1999; Pieper et al. 2005) (figura 8, A y B). A diferencia de los genes *NPAS3* y *NPAS1*, se conoce relativamente bien la función de *Trh*. Este gen, que abarca aproximadamente 30 kpb del cromosoma 3L de *D. melanogaster* (Marygold et al. 2013), es un factor de transcripción de la familia bHLH-PAS esencial para el desarrollo del sistema respiratorio de *D. melanogaster* (Wilk, Weizman y Shilo 1996; Chung, Chavez y Andrew 2011). Este sistema está compuesto por un conjunto de tubos interconectados (o tráqueas) que surgen desde las placas traqueales, un grupo de células ectodérmicas que se forman a cada lado de diez segmentos embrionarios (abarcando desde el segundo segmento torácico hasta el octavo segmento abdominal). Las células de cada placa se invaginan y dan lugar a la formación del árbol traqueal. La estructura básica se asemeja a la de los pulmones, con tubos formados por células epiteliales que se ramifican por los cuales se transportan oxígeno directamente a los órganos internos (Chung, Chavez y Andrew 2011).

En embriones de mosca, la expresión de *Trh* se puede detectar inicialmente en los núcleos de las células que darán origen a las tráqueas, luego en las células en migración que formarán los tubos del sistema traqueal y finalmente en las tráqueas ya formadas (estadios embrionarios 11-15). Adicionalmente, se observa expresión en las células de los ductos de las glándulas salivares (estadio 13) (Wilk, Weizman y Shilo 1996; Chung, Chavez y Andrew 2011). Estudios realizados en

embriones de mosca mutantes mostraron que, en ausencia de *Trh*, no se forman las placas de la tráquea y no ocurre la invaginación desde los primordios de las tráqueas conduciendo a una completa ausencia del sistema respiratorio de la mosca (Wilk, Weizman y Shilo 1996). Por otro lado, la expresión ectópica de *Trh* tiene la capacidad de inducir la formación de placas traqueales ectópicas. El inicio de la expresión de *Trh* depende de estímulos externos, pero una vez iniciada su expresión, entra en un circuito de autorregulación que garantiza el mantenimiento de los niveles de expresión de este factor de transcripción (Wilk, Weizman y Shilo 1996). Adicionalmente, *Trh* regula la expresión de todos los genes conocidos que intervienen en la formación de tráqueas en la mosca (entre ellos el gen ortólogo en mosca de los *Fibroblast growth factor receptors* de vertebrados, *Breathless*) (Chung, Chavez y Andrew 2011). Por otro lado, entre los estadios embrionarios 13-16 comienza la expresión de *Trh* en sistema nervioso central. En particular, se detecta expresión del ARN mensajero en el cordón nervioso ventral y en el cerebro de las moscas (Tomancak et al. 2002; Tomancak et al. 2007). Sin embargo, se desconoce la función de este gen a nivel del sistema nervioso en desarrollo de la mosca.

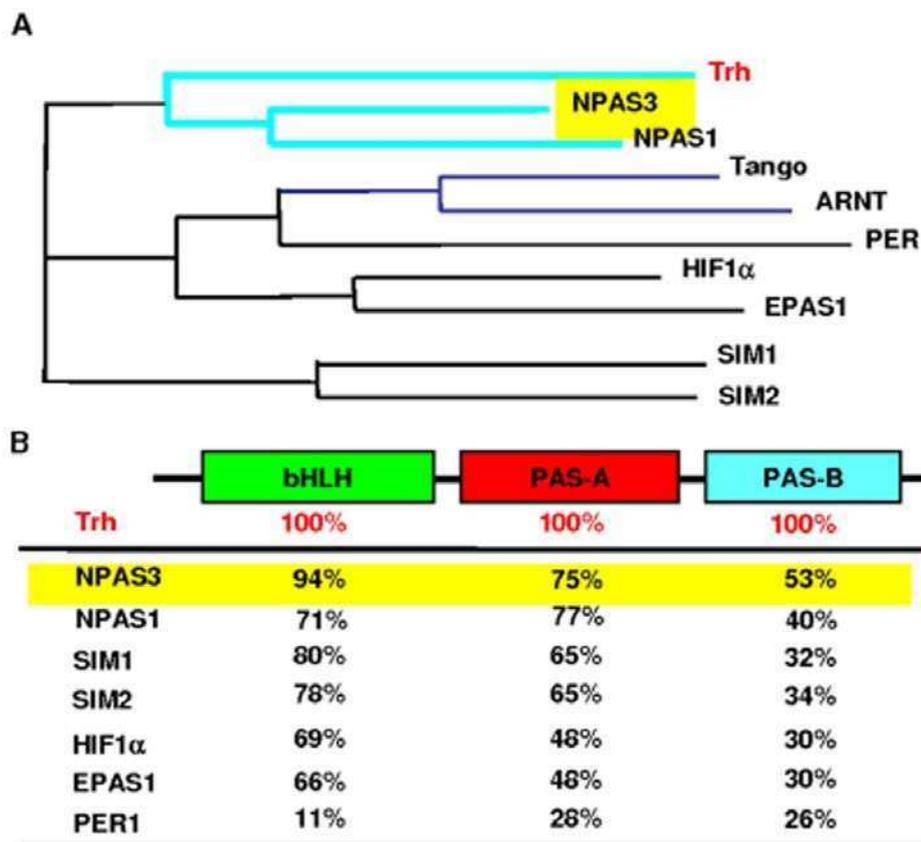


Figura 8: *NPAS3* y *NPAS1*, los homólogos murinos del gen de *Drosophila Trh*. (A) Árbol de genes: los genes humanos *NPAS3* y *NPAS1* están más relacionados con el gen *Trh* de *Drosophila* mientras que la proteína *ARNT* está más relacionado con el gen de *Drosophila Tango*. Adicionalmente, se muestran las relaciones con otras proteína de la familia de bHLH-PAS como ser *PER*, *HIF1 α* , *EPAS1* y *SIM1/2*. (B)

Alineamiento de secuencia: *NPAS3* comparte mayor similitud de secuencia aminoacídica con *Trh* en los dominios bHLH, PAS-A y PAS-B. Modificado de (Zhou et al. 2009).

Patrón de expresión de *NPAS3* en el sistema nervioso de mamíferos

En mamíferos *NPAS3* se expresa abundantemente en el sistema nervioso en desarrollo tanto en ratones como en humanos (Brunskill et al. 1999; Gould y Kamnasaran 2011), donde se observan altos niveles de expresión en el tubo neural, neuroepitelio y, en etapas posteriores, en la capa neopallial de la corteza. Fuera del sistema nervioso, se observa expresión de *NPAS3* en pulmones, corazón, miembros, piel y riñones (Brunskill et al. 1999; Zhou et al. 2009). Ya en adultos, se observa expresión de *NPAS3* en los núcleos de pequeñas interneuronas inhibitorias uniformemente distribuidas a lo largo de todo el sistema nervioso central. Dobles marcaciones muestran que *NPAS3* se expresa casi exclusivamente en neuronas que sintetizan el neurotransmisor inhibitorio ácido gamma-aminobutírico (GABA) y no hay superposición de marca con marcadores de células gliales. Por otro lado, el número de células que expresan *NPAS3* excede, casi en un orden de magnitud, al de las células que expresan *NPASI* (Erbel-Sieler et al. 2004).

***NPAS3* y el análisis de ratones mutantes**

Estudios realizados sobre ratones mutantes donde se truncó la expresión de *Npas3* (*Npas3*^{-/-}) mostraron diversos defectos en el sistema nervioso y en otros sistemas (Erbel-Sieler et al. 2004; Brunskill et al. 2005; Zhou et al. 2009). Por un lado, los ratones *Npas3*^{-/-} presentan una baja viabilidad. Dentro de las primeras horas después del nacimiento, las crías *Npas3*^{-/-} muestran un intenso jadeo y se vuelven cianóticas. Para las 48 horas de vida, el 95% de estas crías muere, mientras que el 100% de sus hermanos heterocigotas para la versión truncada del gen u homocigotas para el *locus* no afectado, sobreviven. La causa de la alta letalidad de las crías *Npas3*^{-/-} se atribuye principalmente a defectos severos en el desarrollo de los alvéolos con un posterior desarrollo de bronquiectasia (una condición donde ocurre una dilatación anormal e irreversible del árbol bronquial) (Zhou et al. 2009). Por otro lado, en las camadas que fueron cuidadas por madres mutantes *Npas3*^{-/-} también se observó una alta mortalidad de las crías (independientemente de su genotipo en el *locus* de *Npas3*) debido a carencias en el cuidado parental (Erbel-Sieler et al. 2004; Brunskill et al. 2005).

Entre otra de las características, los ratones *Npas3*^{-/-} pesan un 20% menos que sus hermanos de genotipo salvaje o heterocigotas y esta diferencia en el peso se mantiene durante toda la vida (Erbel-Sieler et al. 2004; Brunskill et al. 2005). Entre las anomalías morfológicas halladas en los ratones *Npas3*^{-/-} se incluye alteraciones en el patrón de foliación del cerebelo, hipoplasia y

defectos de conectividad del cuerpo calloso, ensanchamiento del tercer y cuarto ventrículo y del ventrículo lateral, reducción de tamaño del hipocampo anterior, entre otros (Brunskill et al. 2005). Con respecto a los defectos comportamentales, se encuentra que los ratones *Npas3*^{-/-} desarrollan comportamientos motores estereotipados (Erbel-Sieler et al. 2004), presentan hiperactividad, defectos en la coordinación motora y en la memoria de reconocimiento y muestran una respuesta alterada frente a situaciones de ansiedad (Brunskill et al. 2005). También se encontraron perturbaciones en las vías de señalización dopaminérgicas, glutamatérgicas y serotoninérgicas (Brunskill et al. 2005). Estos estudios ponen de manifiesto el rol esencial de *Npas3* tanto a nivel estructural como funcional del cerebro.

***NPAS3*, un gen cuya disfunción está asociada al desarrollo de esquizofrenia y otras patologías del sistema nervioso.**

Actualmente, diferentes estudios asocian ciertas variantes genéticas presentes en el *locus* que contiene a *NPAS3* con una mayor susceptibilidad a desarrollar esquizofrenia y otras enfermedades neurológicas en humanos (Pickard et al. 2006). La primera evidencia de dicha asociación se reportó al hallar que una translocación recíproca entre los cromosomas 9 y 14 que ocasionaba la disrupción la secuencia de *NPAS3* entre los exones 3-5 pero no afectaba a ningún gen sobre el cromosoma 9, estaba presente en una familia afectada de esquizofrenia severa (figura 9, A). La madre y la hija afectadas tenían la translocación t(9;14)(q34;q13) pero no estaba presente en el padre no afectado. De expresarse, la predicción del producto proteico de este gen truncado sería una proteína conteniendo el dominio de interacción con el ADN bHLH y de dimerización PAS (Pickard et al. 2006) pero carente del dominio C-terminal de transactivación (y posiblemente de alguno de los dominios PAS también) (Kamnasaran et al. 2003; Pickard et al. 2005) (figura 9, B). Como la haploinsuficiencia de *NPAS3* no tiene evidentes consecuencias funcionales sobre el fenotipo de ratones mutantes a nivel del sistema nervioso (Erbel-Sieler et al. 2004), una de la hipótesis que surgen es que el cerebro humano, a diferencia del de ratón, podría ser más sensible a la disminución en la carga génica de *NPAS3*. Sin embargo, sí se observaron defectos en la estructura alveolar de ratones heterocigotas para el alelo truncado de *Npas3* (Zhou et al. 2009). Una segunda posibilidad sostiene que la translocación observada en la familia afectada de esquizofrenia podría generar la expresión de una versión de *NPAS3* capaz de actuar como dominante negativo. Dado que los miembros de la familia bHLH-PAS deben homo u heterodimerizar para poder ejercer su función, entonces la expresión de una proteína capaz de unirse al ADN y dimerizar (por medio de sus dominios básicos y HLH-PAS, respectivamente) pero carente de sus dominios regulatorios podría ocasionar el secuestro de los miembros proteicos necesarios para la dimerización formando complejos no funcionales sobre el ADN (Erbel-Sieler et al. 2004; Pickard et al. 2006).

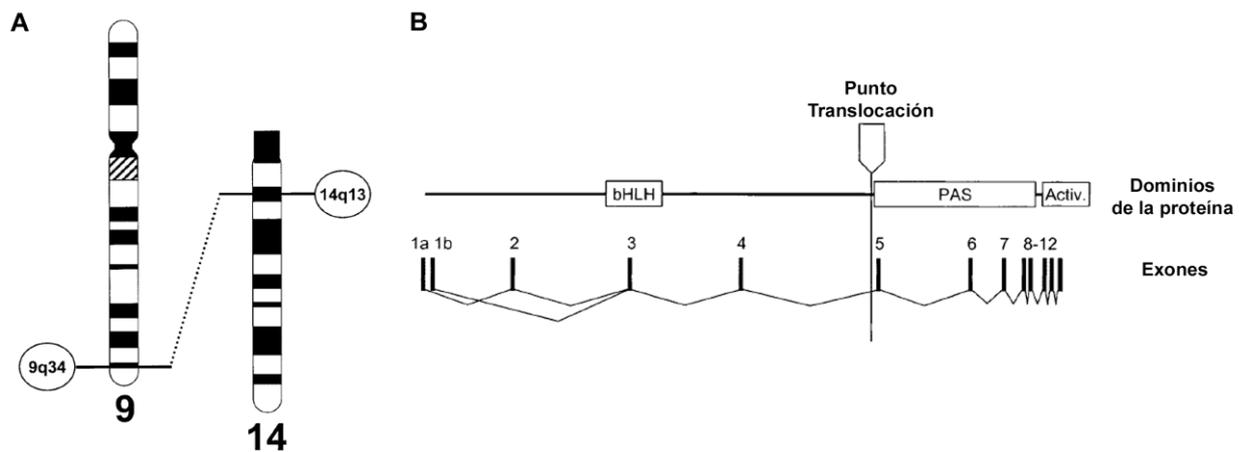


Figura 9: Rearreglo genómico en el gen *NPAS3* presente en el cromosoma humano 14. (A) Ideograma representando la translocación recíproca $t(9;14)(q34.11;q13.1)$ observada en la madre e hija afectadas de esquizofrenia. (B) Esquema indicando el sitio del rearreglo cromosómico afectando al *locus* de *NPAS3*. Adicionalmente, se muestran los dominios funcionales codificados por los exones de *NPAS3*. Modificado de (Pickard et al. 2005).

Por otro lado, recientemente, se identificó una mutación puntual (un cambio nucleotídico G por A) en la región codificante de *NPAS3* que co-segregaba con esquizofrenia en una familia humana pequeña. La mutación identificada está en el exón 8 del gen y da lugar a un cambio aminoacídico valina por isoleucina en la posición 304 (V304I). Esta mutación estaba presente en los tres miembros afectados de esquizofrenia de esta familia y ausente en los miembros restantes. Ensayos funcionales mostraron que, cuando se expresó la proteína *NPAS3*-V304I en neuronas primarias de ratón hubo una disminución significativa del crecimiento de las neuritas respecto de la expresión de la proteína salvaje, brindando evidencia del impacto funcional de esta mutación podría tener (Yu et al. 2013). Adicionalmente, un reporte reciente describe la presencia de una delección intragénica de 0,111 Mb dentro del *locus* de *NPAS3* hallada en una niña de 8 años con signos de retraso del desarrollo. La niña afectada caminó a los 2,5 años, comenzó a hablar a los 4 años de edad y presentó adicionalmente retrasos en el desarrollo de habilidades motoras generales y finas. Entre otras afecciones se encontraron incontinencia urinaria, siringomielia (formación de quistes) en la médula espinal, obesidad, astigmatismo y macrocefalia. Estudios adicionales mostraron que la delección fue heredada de la madre, la cual también presentó problemas de aprendizaje (C. Armour 2013).

Adicionalmente, existen varios trabajos que reportan la asociación de diversos polimorfismos dentro del *locus* de *NPAS3* con enfermedades psiquiátricas, principalmente esquizofrénica y desorden bipolar (Huang et al. 2010; Macintyre et al. 2010). Por ejemplo, un estudio de caso-

control encontró asociación entre múltiples alelos presentes en intrones de *NPAS3* con una mayor probabilidad de desarrollar esquizofrenia y desorden bipolar (Pickard et al. 2009). Más recientemente, un estudio de asociación de escala genómica (GWAS, por *genome-wide association studies*) de variación con desorden bipolar, encontró que 10 polimorfismos de sitio único presentes en intrones de *NPAS3* estaban positivamente asociados con esta condición (Weber et al. 2011).

HIPÓTESIS DE TRABAJO

La hipótesis de trabajo de este capítulo sostiene que las regiones HAEs que se acumularon en el *locus* del gen *NPAS3* funcionan como *enhancers* transcripcionales de este gen. Además, se postula que el proceso evolutivo ocurrido en el linaje humano en estas regiones resultó en diferencias funcionales entre los *enhancers* humanos y los *enhancers* ortólogos de nuestro pariente primate más cercano, el chimpancé (*Pan troglodytes*).

OBJETIVOS ESPECÍFICOS

- ✓ Investigar si los 14 HAEs presentes en los intrones de *NPAS3* tienen la capacidad de actuar como reguladores de la expresión génica. En particular, se analizará si estos elementos pueden funcionar como *enhancers* transcripcionales. Para ello, se generarán peces transgénicos portadores de una construcción conteniendo las potenciales regiones regulatorias, fusionadas río arriba del gen reportero EGFP.
- ✓ Estudiar si algunos de estos 14 HAEs tienen patrones de expresión que sugieran diferencias de actividad entre los elementos humanos y regiones ortólogas de chimpancé. Con este objetivo, se realizará un análisis comparativo de expresión en peces cebra transgénicos llevando regiones ortólogas de humano y chimpancé.
- ✓ Analizar el patrón de expresión del elemento 2xHAR142, uno de los HAEs presentes en el 5to intrón de *NPAS3*, en un modelo mamífero. Para ello, se generarán ratones transgénicos conteniendo la secuencia de este elemento fusionada río arriba de un promotor mínimo dirigiendo la expresión de la proteína reportera *LacZ*.
- ✓ Estudiar cambios evolutivos en la actividad de este posible *enhancer* comparando la actividad producida por la secuencias humana del *enhancer* en relación con las regiones ortólogas de otro primate (chimpancé, *Pan troglodytes*) y de un roedor (ratón, *Mus musculus*).
- ✓ Analizar si el patrón de expresión producido por este *enhancer* se solapa con el patrón de expresión del gen *Npas3* endógeno. Para ello se realizarán ensayos de inmunohistoquímica (IHQ) e hibridación *in situ* sobre muestras de embriones de ratones transgénicos.

- ✓ Estudiar el origen evolutivo de los HAEs, particularmente de aquellos presentes en el *locus* de *NPAS3*.

RESULTADOS

Análisis del patrón de expresión de *NPAS3* en pez cebra

En la segunda parte de este trabajo, se abordó el estudio funcional de los HAEs presentes en *NPAS3*. Para ello, se decidió estudiar si los 14 elementos acelerados presentes en *NPAS3* tienen la capacidad de activar la expresión génica utilizando un ensayo de gen reportero en peces transgénicos. Si bien el sistema de ensayo de *enhancers* en peces cebra transgénicos mostró ser una técnica eficiente y confiable para estudiar la capacidad de activar la expresión de potenciales regiones regulatorias humanas (Fisher et al. 2006a; Fisher et al. 2006b), antes de comenzar con este análisis se determinó si *NPAS3* se expresa en el cerebro del pez cebra y durante el desarrollo embrionario. Para ello, se detectó la expresión de *NPAS3* por *reverse-transcription* PCR (RT-PCR) sobre muestras de ARNm obtenidas de embriones de pez cebra de 24 horas post fecundación (hpf) y 48 hpf y sobre muestras de ARNm extraídas de cerebros adultos utilizando para ello dos juegos de *primers* específicos. Se encontró que *NPAS3* se expresa en los tres estadios analizados (figura 10, A). Adicionalmente, se estudió el patrón de expresión de *NPAS3* en embriones de pez cebra de aproximadamente 24 y 48 hpf. Para ello se realizaron hibridizaciones *in situ* (ISH) utilizando dos sondas independientes diseñadas para detectar este mensajero. Se seleccionaron estos dos estadios embrionarios ya que, al nivel morfológico y del transcriptoma comparado de los vertebrados, 24 hpf constituye el estadio más conservado entre embriones de ratón (*Mus musculus*), pollo (*Gallus gallus*), rana (*Xenopus laevis*) y pez cebra (*Danio rerio*) (Irie y Kuratani 2010). Se considera que los embriones de pez cebra de esta edad representan el estadio filotípico de los vertebrados, donde poseen las características clásicas del plan corporal de los vertebrados (Ballard 1981). Por otro lado, 24 y 48 hpf constituirían estadios homólogos a embriones de ratón de entre 10.5 y 16.5 *días post coitum* (Irie y Kuratani 2010).

Los ensayos de hibridizaciones *in situ* mostraron que *NPAS3* se expresa ubicuamente a 24 hpf (figura 10, B). Un día después, el patrón de expresión es más restringido, incluyendo diferentes zonas del cerebro en desarrollo y la médula espinal. En el cerebro, se observó altos niveles de expresión en el telencéfalo, diferentes estructuras diencefálicas incluidas el tálamo y el hipotálamo y también en el techo óptico. Fuera del sistema nervioso central (SNC), se observó expresión en los ojos, el corazón y particularmente, altos niveles de expresión en los arcos branquiales, aletas pectorales y el techo de la boca (figura 10, B y C). Estas observaciones indicarían que el patrón general de expresión de *NPAS3* está conservado entre los vertebrados (Brunskill et al. 1999; Gould y Kamnasaran 2011) y alentaron el uso del pez cebra como modelo en donde estudiar posibles regiones regulatorias de *NPAS3*.

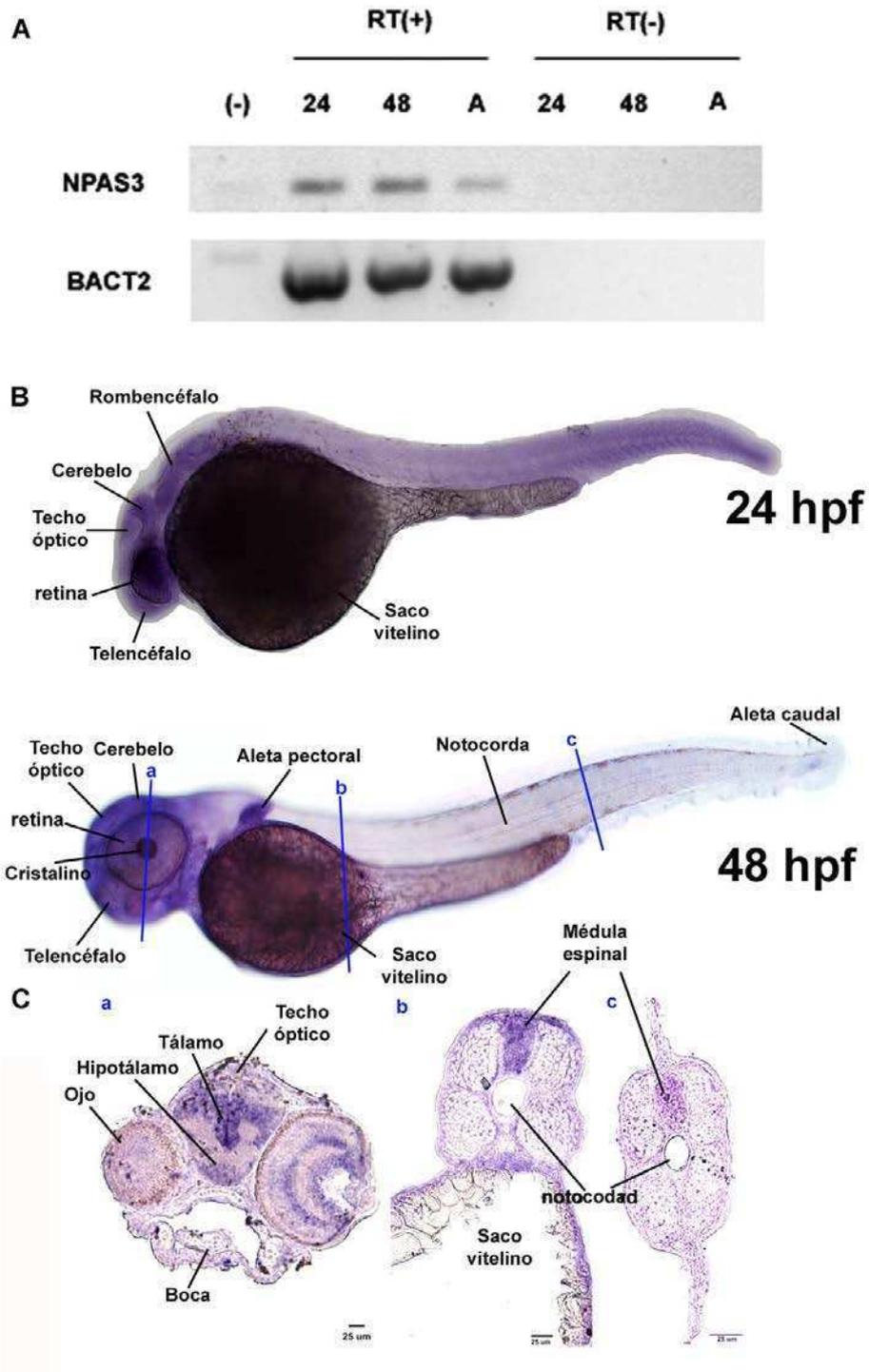


Figura 10: Expresión del gen *NPAS3* durante el desarrollo del pez cebra a 24 y 48 hpf (A) Análisis de la expresión de *NPAS3* por RT-PCR sobre muestras de ARNm obtenidas de embriones de 24 y 48 hpf y de cerebro adulto. Se muestran los productos de PCR del gen *NPAS3* y del gen control β -actina 2. Abreviaturas: (-) control negativo. 24/48, 24 y 48 hpf. A, cerebro adulto con (+) o sin (-) agregado de enzima transcriptasa reversa. (B) Patrón de expresión de *NPAS3* a 24 (arriba) y 48 (abajo) hpf revelado por hibridación *in situ* sobre embriones completos. (C) Hibridación *in situ* sobre cortes coronales de crióstato mostrando detalles del patrón de expresión de *NPAS3* a 48 hpf.

Caracterización funcional de las regiones aceleradas de *NPAS3* en peces transgénicos

Posteriormente, las 14 regiones conservadas (detectadas a través del alineamiento de los genomas de múltiples especies de vertebrados usando el programa PhastConst) (Siepel et al. 2005) que contienen los HAEs presentes en intrones de *NPAS3* se clonaron río arriba del promotor mínimo del gen *cFos* de ratón regulando la expresión de la proteína reportera EGFP. Este vector contiene las secuencias Tol2 que son reconocidas por una transposasa de Medaka, la cual media la transgénesis en el genoma de pez cebra (Kawakami et al. 2004; Fisher et al. 2006b). En todos los casos, se decidió clonar toda la región conservada que contiene al HAE porque probablemente sea esta más representativa de la unidad funcional potencialmente contenida en estos elementos (frente a la estricta descripción bioinformática de los HAEs) (Prabhakar et al. 2008).

Los embriones de pez cebra inyectados F0 son animales transgénicos mosaicos. Esto significa que algunos tejidos del embrión tendrán el transgén inserto en el genoma, mientras otros tejidos no (Nusslein-Volhard 2002). Las razones de este fenómeno no se conocen con detalle. Sin embargo, esta característica hace que sea necesario evaluar múltiples embriones F0 para poder reconstruir el patrón de expresión de los *enhancers* analizados (Ostrander 2000). Para estudiar la actividad *enhancer* de los HAEs se recurrió a la metodología estandarizada para capturar la actividad *in vivo* de regiones conservadas no codificantes en pez cebra propuesta por el proyecto CONDOR (detalles disponibles <http://condor.nimr.mrc.ac.uk/methods.html#assay>) (McEwen et al. 2006; Woolfe et al. 2007). Este protocolo fue utilizado con éxito para la detección de *enhancers* en peces cebra transgénicos mosaico (Woolfe et al. 2005; McEwen et al. 2006). Brevemente, se analizó el patrón de expresión de EGFP a 24 y 48 hpf en 30-100 embriones F0 por edad y por transgén proviniendo de, al menos, dos días de inyección independientes. Se clasificó los dominios de expresión de EGFP utilizando únicamente términos anatómicos generales y comparables entre vertebrados (como ser “médula espinal” o “prosencéfalo”) y se evitó la clasificación dentro de subdominios o tipos celulares (Ariza-Cosano et al. 2012). Los resultados de cada ensayo se sumaron y se consideró como *enhancer* a todo elemento que produjera > 20% de embriones F0 con expresión consistente de EGFP (ver materiales y métodos para más detalles). Como control de la expresión basal de EGFP en este sistema, se inyectó el vector vacío en las mismas condiciones experimentales. En los ensayos de expresión transiente en pez cebra se considera que el patrón de expresión hallado consistentemente en 10-20% de los embriones F0 analizados será representativo de la expresión dirigida por el mismo *enhancer* después de la transmisión germinal (animales no mosaico) (Fisher et al. 2006b; Lee et al. 2011; Oksenberg et al. 2013).

El análisis de expresión de los HAEs de *NPAS3* a 24 y 48 hpf en peces transgénicos mostró que 11 de los 14 elementos analizados son capaces de activar de forma distintiva y reproducible la expresión de EGFP a estas edades; observándose principalmente expresión en diferentes dominios de SNC en desarrollo (figura 11; tabla 4). Llamativamente, no se encontró correlación entre la profundidad de conservación filogenética de los HAEs y la capacidad de activar la expresión de EGFP (figura 11; tabla 4; tabla suplementaria s3). De los 11 elementos capaces de dirigir la expresión de forma consistente en el ensayo de expresión el pez cebra, solo 4 de ellos están conservados en los vertebrados incluido pez cebra (tabla suplementaria s3). Además, 2 de los elementos que no fueron capaces de activar la expresión de EGFP en este ensayo están conservados en pez cebra (tabla 4; tabla suplementaria s3). Por otro lado, el vector vacío es incapaz de generar un patrón de expresión reproducible, observándose esporádicamente algunos animales con arreglos irregulares de células verdes principalmente en corazón y músculo somático (tabla 4). De hecho, otros grupos usando el mismo sistema de expresión reportaron sitios similares de expresión basal de EGFP (Oksenberg et al. 2013). Por último, estos resultados muestran que gran parte de los HAEs presentes en *NPAS3* podrían actuar como reguladores de la expresión génica *in vivo*, desarrollándose como *enhancers* durante el desarrollo del sistema nervioso. Por otro lado, estos resultados están en línea con observaciones previas que sugieren que la información funcional (en este caso de los *enhancers*) podría estar conservada entre vertebrados, más allá del nivel de detección de la conservación entre secuencias ortólogas *per se* (Fisher et al. 2006a).

Por otro lado, existen diversas razones que pueden explicar los resultados negativos obtenidos en los ensayos de expresión transiente para tres de los HAEs de *NPAS3* (HAR202Hs, HACNS658Hs y 2xHAR223Hs). Por un lado, estos elementos podrían no contener una actividad *enhancer* pudiendo o no albergar otra/s actividad/es regulatoria/s hasta ahora no analizadas (por ejemplo *silencers*, *insulator*, etc). Este hallazgo sugiere una hipótesis particularmente interesante: las sustituciones acumuladas en estos elementos en el linaje que dio lugar a los humanos podrían haber llegado a disrumpir la actividad *enhancer*, por ejemplo, modificando sitios de unión para factores de transcripción. Trabajo realizado por la Dra. Lucía Franchini permitió mostrar que, las sustituciones específicas de los humanos presentes en los HAEs en *NPAS3* producen, para todos los elementos, ganancias y pérdidas de secuencias predichas de unión para factores de transcripción (tabla suplementaria s3). Este hallazgo refuerza la idea de que el proceso de evolución acelerada experimentada por los HAEs de *NPAS3* podría haber generado cambios en el patrón espacial y temporal de expresión de estos elementos contribuyendo con la adquisición de algunas de las características específicas de nuestro cerebro.

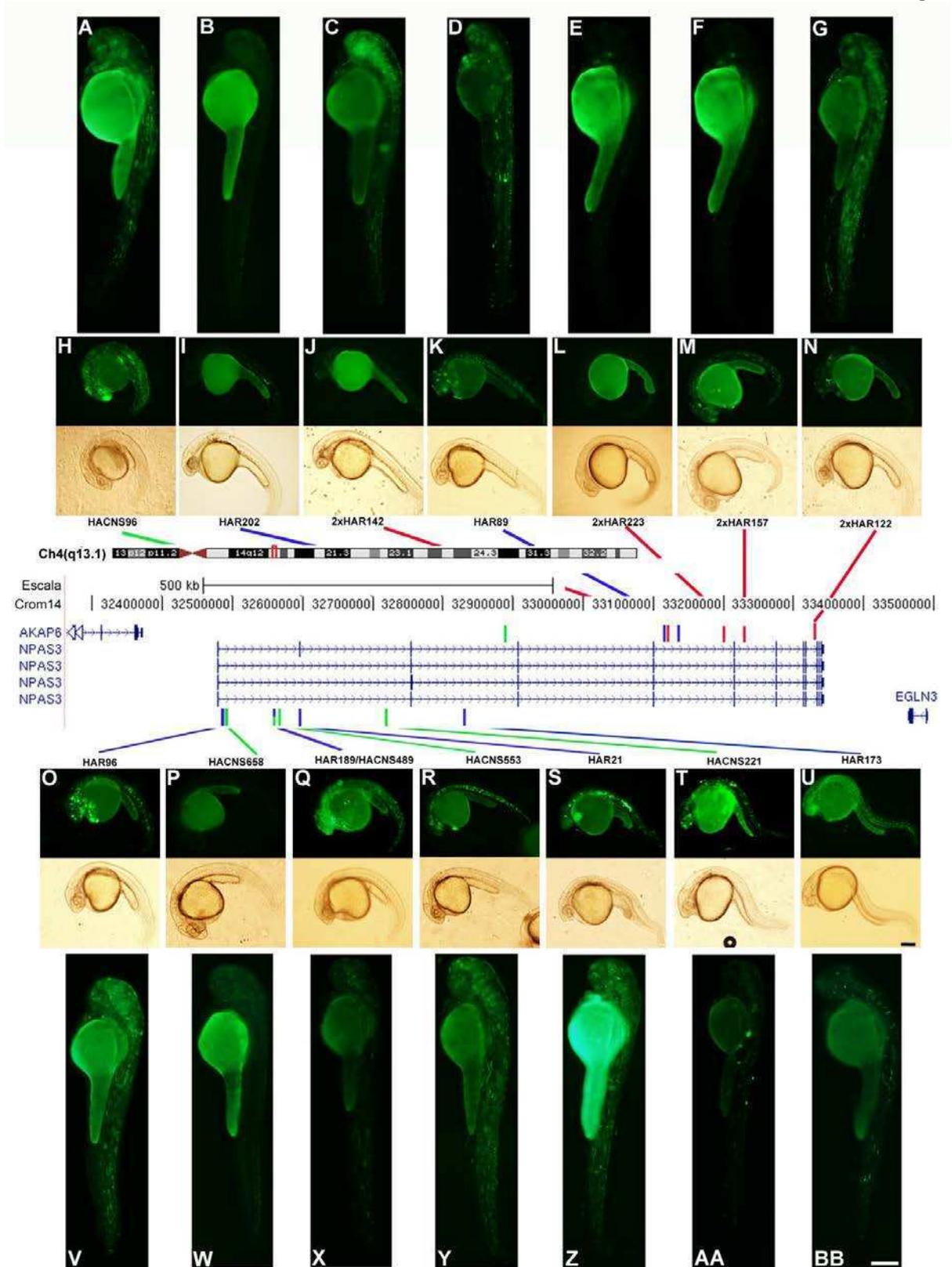


Figura 11: Ensayo de expresión transientes en peces cebrá transgénicos. En el panel central se muestra la localización del *locus* de *NPAS3* en el cromosoma 14 del genoma humano (NCBI36/hg18) en una ventana de 1.1 Mb de longitud de acuerdo a con el *UCSC Genome Browser* (www.genome.ucsc.edu). Abajo en barras coloreadas se indica la localización de los 14 HAEs y debajo las variantes de *splicing* de *NPAS3*. Las barras azules indican los HAR (Pollard et al. 2006), las verdes representan los HACNS (Prabhakar et al. 2006) y las rojas muestran los 2xHARs (Lindblad-Toh et al. 2011). Los elementos solapados se muestran con barras de dos colores (azul y verde). Arriba y debajo del panel central se

muestran micrografías de peces transgénicos con el patrón representativo de la expresión mosaico de EGFP para cada elemento a 24 hpf (H,I,J,K,L,M,N,O,P,Q,R,S,T yU) y 48 hpf (A,B,C,D,E,F,G,V,W,X,Y,Z y AA). Escala: 250 µm.

Tabla 4: Análisis de expresión en peces cebra transgénicos a 24 y 48 hpf.

Nombre del Transgén	Localización genómica (NCBI36/hg18)	Largo (bp)	N	%(+)	%C	Dominios de expresión consistente							
						T	M	R	O	ME	AF	C	MS
HAR96Hs	chr14: 32486645-32487168	524	87/ 72	82/ 83	39/21	+	+	+	+	-/+	+/-	-	-
HAR96Pt	chr14:31,834,352-31,834,876	525	71/ 75	73/ 85	30/28	+	+	+	+	-/+	-/+	-/+	-/+
HACNS658Hs	chr14: 32487503-32487837	335	116/ 94	25/ 34	-	-	-	-	-	-	-	-	-
HAR189Hs/ HACNS489Hs	chr14: 32558123-32558864	742	71/ 67	80/ 82	42/18	+/-	-/+	-/+	-	-/+	-	-	-/+
HACNS553Hs	chr14: 32566093-32566346	254	94/ 108	63/ 82	50/32	+	+	+	+	+	-	-/+	-/+
HAR21Hs	chr14: 32596637-32596862	298	71/ 64	84/ 86	38/33	+	+	+	+	+	+/-	-/+	+
HAR21Pt	chr14: 32596637-32596862	298	80/ 64	57/ 74	34/37	+	+	+	+	+	+	+	+
HACNS221Hs	chr14: 32718813-32719042	230	62/ 56	77/ 68	34/37	+/-	+	+	+/-	+/-	-	-	+/-
HAR173Hs	chr14: 32830590-32830773	184	79/ 69	75/ 74	28/37	+/-	+	-/+	+/-	-/+	-	-	-
HACNS96	chr14: 32888897-32889211	315	95/ 100	78/ 84	26/17	+	+	+	+	+	+/-	-/+	+
HAR202Hs	chr14: 33115156-33115595	440	69/ 75	32/ 17	-	-	-	-	-	-	-	-	-
HAR202Pt	chr14: 32605631-32606070	440	64/ 57	80/ 67	28/31	+	+	+	+	+	-	-	+
2xHAR142Hs	chr14: 33118315-33118577	263	94/ 40	81/ 70	27/30	+	-/+	+	+	+	-/+	-	-/+
HAR89Hs	chr14: 33135162-33135937	776	64/ 28	83/ 86	22/36	+	+	+	+/-	+	+/-	-	+
HAR89Pt	chr14:32,490,096-32,490,869	774	79/ 80	66/ 72	24/29	+/-	+	+	+/-	+	-	-	+
2xHAR223Hs	chr14: 33199815-33200914	1100	86/ 33	42/ 36	-	-	-	-	-	-	-	-	-
2xHAR157Hs	chr14: 33229602-33230232	631	58/ 32	90/ 44	36/37	+/-	+/-	+/-	+/-	+/-	-	-	+/-
2xHAR122Hs	chr14: 33331419-33331887	469	62/ 31	80/ 84	53/45	+	+	+	+	+	-	-	-/+
<i>cFos</i>		-	64/ 114	30/ 33	-	-	-	-	-	-	-	-	-

Nota: Las celdas en verde indican los HAEs que no mostraron actividad *enhancers* y el control. En celeste, se indica el ensayo realizado con la secuencia de chimpancé. Las coordenadas genómicas, el largo y el número de sustituciones humanas específicas corresponden a la región conservada completa que contiene al HAE y la que fue usada para generar los transgenes. Referencias: N, número de animales analizados. %(+), porcentaje de embriones con expresión de EGFP. %C, porcentaje de embriones con expresión consistente de EGFP. T, telencéfalo. M, mesencéfalo. R, rombencéfalo. O, ojo. ME, médula espinal. AF, arcos faríngeos. C, corazón, MS, músculo somático. Cuando los resultados son diferentes entre las dos edades, entre barras se indican los resultados a 24/48 hpf e (+) y (-) indica presencia o ausencia de expresión consistente en la estructura respectivamente.

Alternativamente, el sistema utilizado para analizar la capacidad de activar la expresión génica puede no ser el más adecuado para evaluar estos tres elementos. Si bien se cuenta con abundante

literatura que respalda el uso de peces cebras transgénicos como una metodología rápida y confiable para estudiar *enhancers* humanos *in vivo* (Fisher et al. 2006a; Ghiasvand et al. 2011; Kenyon et al. 2011), existen ciertas limitaciones que deben ser tenidas en cuenta. Entre estas, no se pueden descartar que cambios en *trans* ocurridos desde la divergencia entre el linaje del pez cebra y los humanos (aproximadamente hace 450 millones de años), hagan irreconocibles *enhancers* humanos cuando son estudiados en el sistema de pez cebra (Ritter et al. 2010). Serían necesarios ensayos adicionales en otros modelos (por ejemplo, ratones transgénicos) para confirmar o descartar esta última hipótesis.

Por último, es interesante destacar que, no se encontraron diferencias entre los elementos generados por las tres bases de datos aquí analizadas experimentalmente (HARs, HACNS y 2xHARs). Las tres metodologías aplicadas a la predicción de secuencias de evolución acelerada en linaje humano fueron capaces de detectar elementos con actividad *enhancer* y no se encuentran hasta aquí razones para sostener que las predicciones generadas por alguna de las bases de datos sean mejores.

Comparación de los patrones de expresión transiente de los elementos HAR21, HAR89, HAR96 y HAR202 y sus ortólogos en chimpancé en peces transgénicos

Con el objetivo de estudiar si las sustituciones acumuladas en los HAEs en el linaje humano pudieron haber llegado a modificar la actividad de una región *enhancer*, se analizó comparativamente el patrón de expresión de EGFP de varios HAEs seleccionados y sus secuencias ortólogas de chimpancé en peces cebras transgénicos.

Por un lado, el análisis comparativo de expresión de los elementos humanos HAR21, HAR89 y HAR96 (HAR21Hs, HAR89Hs y HAR96Hs) y secuencias ortólogas de chimpancé (HAR21Pt, HAR89Pt y HAR96Pt) no mostró indicios de cambio en la actividad de estos *enhancers*. Estos seis elementos analizados fueron capaces de activar de forma distintiva y reproducible la expresión de EGFP a 24 y 48 hpf; observándose principalmente expresión en diferentes dominios de SNC en desarrollo (tabla 4).

Adicionalmente, se clonó la región ortóloga en chimpancé del elemento HAR202 (HAR202Pt) y se analizó su capacidad de activar la expresión de EGFP a 24 y 48 hpf. Llamativamente, se encontró que, mientras el elemento HAR202Hs es incapaz de producir un patrón reproducible de EGFP, el elemento HAR202Pt funciona como un potente *enhancer* dirigiendo la expresión de EGFP principalmente al SNC en desarrollo del pez cebra (tabla 4). Este resultado se sumaría a la evidencia creciente que los HAEs podrían ser regiones regulatorias que cambiaron su función en el linaje

humano (Prabhakar et al. 2008; Capra et al. 2013). Si bien el elemento HAR202 experimentó un aumento en su tasa de evolución molecular exclusivamente en el linaje humano, comparado con la tasa de evolución más baja presente en el linaje de los chimpancés y otros primates (Pollard et al. 2006), sin evidencia adicional proveniente de un tercer linaje cercanamente emparentado, no se puede excluir la posibilidad de que el patrón de expresión observado para el elemento de chimpancé represente el estado derivado.

Análisis de la actividad *enhancer* de los HAEs en líneas transgénicas en peces cebra

Con el objetivo de validar la reproducibilidad de los ensayos de expresión transiente y analizar en más detalle la actividad de estos posibles *enhancers*, se generaron líneas de peces transgénicos llevando las construcciones HAR21Hs, HAR21Pt, HAR89Hs, HAR89Pt, HAR96Hs, HAR202Hs, HAR202Pt, HACNS96Hs y HACNS221Hs. En todos los casos, los animales F0 inyectados con cada construcción se dejaron llegar a la madurez sexual y se cruzaron con animales de genotipo salvaje (o *wild type*; wt). Se establecieron al menos dos líneas independientes por elemento y se analizó el patrón de expresión de EGFP a 48 hfp en los animales F1. Se registró como expresión consistente de EGFP aquella hallada en dos o más líneas independientes.

Analizando los patrones de expresión dirigidos por los elementos HAR21Hs, HAR21Pt, HAR89Hs, HAR89Pt, HAR96Hs, HAR202Pt, HACNS96Hs y HACNS221Hs, se encontró que la expresión observada en los ensayos transientes es ampliamente consistente con el patrón de expresión detectado en las líneas estables, principalmente a nivel del SNC (figura 12; figura 13; tabla 4). Este resultado está en línea con observaciones previas que sugieren que la reconstrucción del patrón de expresión en animales mosaicos por observación de un gran número de embriones será representativo del patrón observable en líneas estables (Fisher et al. 2006b). En todos los casos, estos transgenes fueron capaces de dirigir la expresión de EGFP al sistema nervioso en desarrollo incluyendo diferentes estructuras del cerebro y la médula espinal. Por fuera del sistema nervioso, también se observó abundante expresión en corazón, arcos branquiales y aleta pectoral (estos últimos también dominios de expresión de *NPAS3*) (figura 12; figura 13; tabla 5).

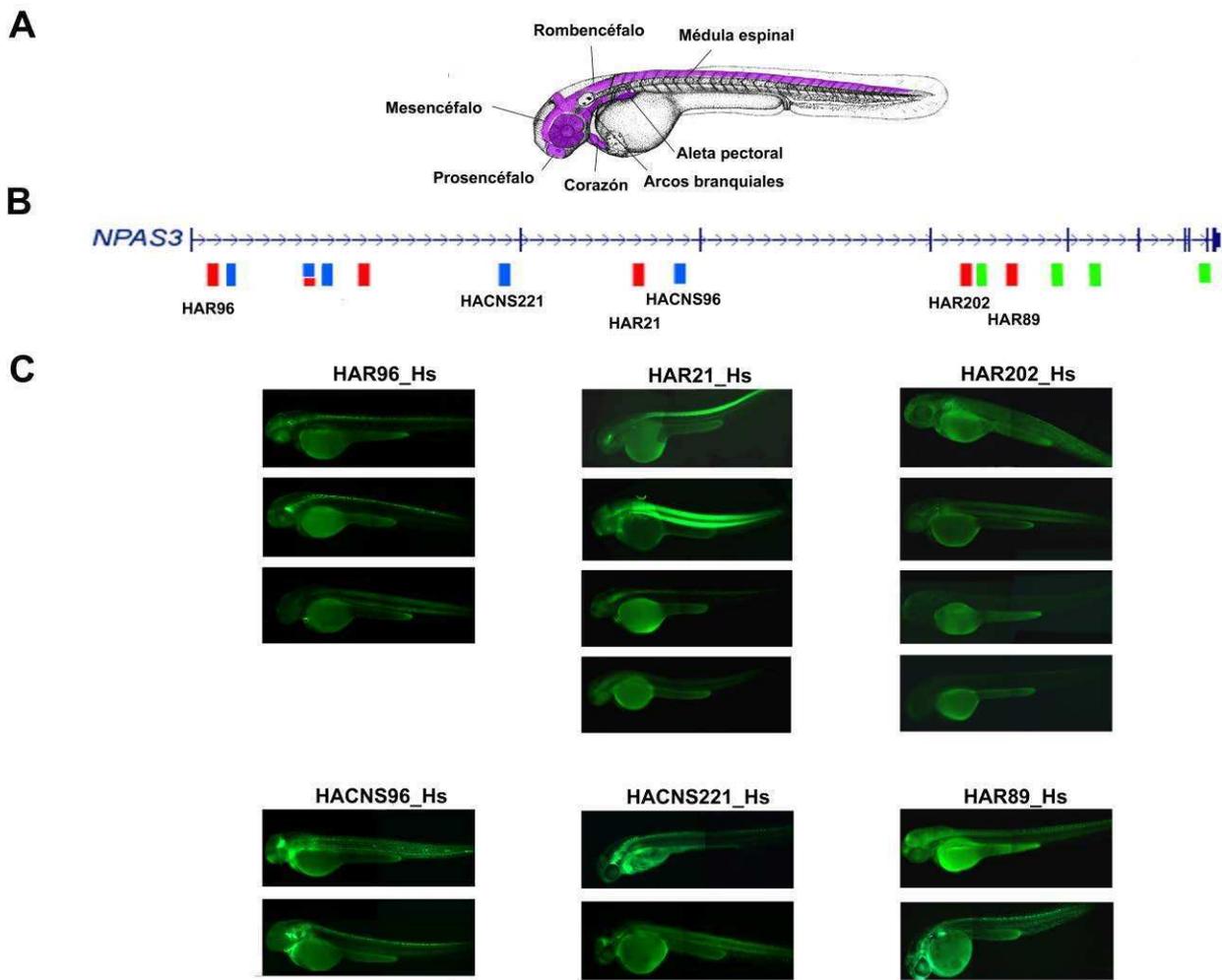


Figura 12: (A) Esquema representativo de la expresión de *NPAS3* en pez cebra a 48 hpf generado a partir de ensayos de hibridación ISH en embriones completos. (B) Esquema del *locus* de *NPAS3* de acuerdo con el *UCSC Genome Browser* (www.genome.ucsc.edu). En barras coloreadas se indica la localización de los 14 HAEs. Las barras rojas indican los HAR (Pollard et al. 2006), las azules representan los HACNS (Prabhakar et al. 2006) y las verdes muestran los 2xHARs (Lindblad-Toh et al. 2011). Los elementos solapados se muestran con barras de dos colores (roja y azul). Se indican los nombres de los elementos que fueron estudiados en líneas transgénicas de pez cebra. (C) Micrografías mostrando la expresión de EGFP hallada en las líneas transgénicas generadas con los elementos HAR21Hs, HAR89Hs, HAR96, HAR202Hs, HACNS96Hs y HACNS221Hs. Cada foto corresponde con una línea independiente. Escala 500µm.

Tabla 5: Análisis de expresión de líneas de peces cebra transgénicos a 48 hpf.

Nombre del transgén	# de líneas con expresión de EGFP	Dominios de expresión presentes en 2 o más líneas transgénicas								
		P	M	R	ME	C	AF	AP	Ojo	
HAR21Hs	5/5*	+	+	+	+	+	+	+	+	+
HAR21Pt	2/2	+	+	+	+	+	-	+	+	+
HAR96Hs	3/3	-	+	+	+	+	-	+	+	+
HACNS96Hs	3/3**	-	+	+	+	+	+	-	-	-
HAR202Hs	2/4	-	-	-	-	-	-	-	-	-
HAR202Pt	4/4	+	+	+	+	-	+	+	+	+
HAR89Hs	2/2	+	+	+	+	+	+	+	+	+
HAR89Pt	2/2	+	+	+	+	+	+	+	+	+
HACNS221Hs	2/2	+	+	+	+	-	+	-	-	-

Nota: Las celdas en celeste indican los resultados con la secuencia de chimpancé y las verdes, aquellos transgenes que no tuvieron expresión reproducible entre 2 o más líneas. Referencias: P, prosencéfalo. M, mesencéfalo. R, rombencéfalo. ME, médula espinal. C, corazón. AB, arcos faríngeos. AP, aleta pectoral, O, ojo. Con (+) y (-) se indica presencia/ausencia de expresión de EGFP. *Una de estas líneas transgénicas no se muestra en la figura 13 debido a que tiene un patrón de expresión muy divergente mostrando expresión de EGFP exclusivamente en los arcos branquiales.**Una de estas líneas transgénicas se excluyó del análisis por mostrar un nivel de expresión de EGFP muy bajo.

Adicionalmente, con el objetivo de estudiar en más detalle el patrón de expresión dirigido por algunos de los HAEs, se realizaron ensayos de inmunohistoquímica en cortes transversales de embriones de pez cebrá donde se codetectó la proteína reportera EGFP y el marcador de neuronas postmitóticas HuC/D (Park et al. 2000). Se escogió este marcador debido a que existía evidencia que HuC/D y *NPAS3* colocalizan parcialmente en el sistema nervioso de ratones en desarrollo (datos no mostrados). Los e HAR21Hs, HAR96Hs, HACNS96 y HAR202Pt son capaces de dirigir la expresión, al menos parcialmente, a neuronas posmitóticas marcadas con HuC/D. Por fuera del SNC, la mayoría de los elementos mostraron actividad en corazón, arcos faríngeos y aletas pectorales (figura 14; figura 15; figura 16; figura 17).

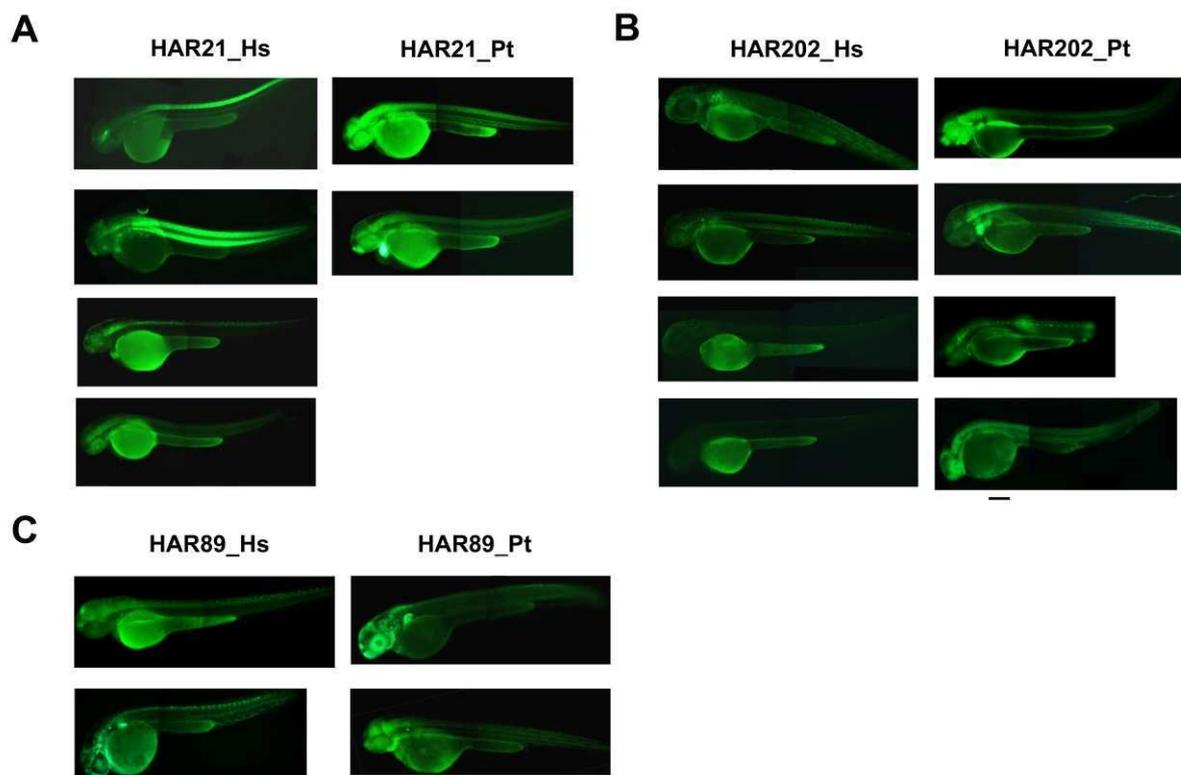


Figura 13: Comparación del patrón de expresión de EGFP en líneas de peces transgénicos a 48 hpf. (A) Micrografías mostrando la expresión de EGFP hallada en las líneas transgénicas generadas con los elementos HAR21Hs y HAR21Pt. (B) Micrografías mostrando la expresión de EGFP hallada en las líneas transgénicas generadas con los elementos HAR202Hs y HAR202Pt. (C) Micrografías mostrando la expresión de EGFP hallada en las líneas transgénicas generadas con los elementos HAR89Hs y HAR89Pt. Cada foto corresponde con una línea independiente. Escala 500µm.

Sin embargo, se observaron algunas diferencias entre los patrones de expresión dirigidos por cada elemento. Por un lado, el elemento HACNS96Hs dirige la expresión de EGFP al techo óptico y la parte más lateral del tegumento donde se codetecta la expresión de EGFP y HuC/D (figura 14, K-M). Adicionalmente, se observa codetección de EGFP y HuC/D en la parte más ventral de la médula oblonga del rombencéfalo (figura 14, Q-S) y en médula espinal (figura 14, T-V). Por otro lado, los elementos HAR202Pt (figura 15, P-U) y HAR96Hs (figura 16, M-R) dirigen la expresión a neuronas postmitóticas en la parte lateral y ventral del rombencéfalo y la médula espinal. Tanto el elemento HAR202Pt (figura 15, J-L) como el elemento HAR96Hs (figura 16, J-K) dirigen la expresión de EGFP al techo óptico. Finalmente, el elemento HAR21Hs dirige la expresión de EGFP a varios dominios del SNC incluidos el prosencéfalo, mesencéfalo, rombencéfalo y la médula espinal (figura 17, F y G).

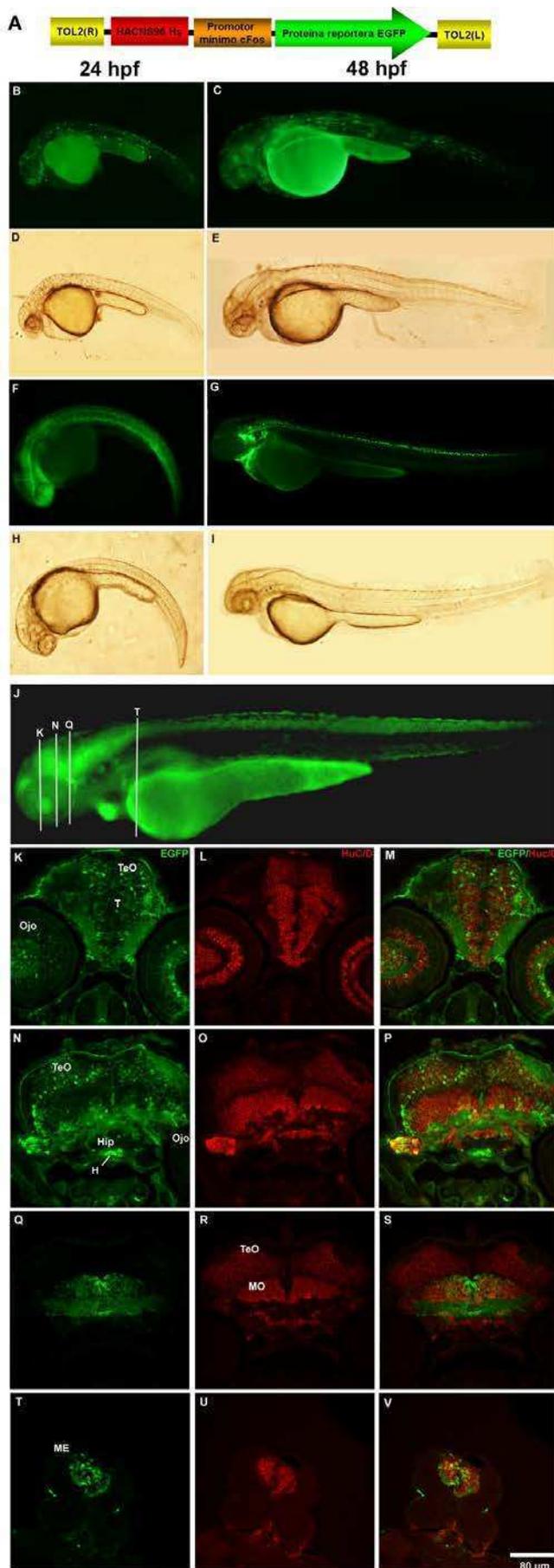


Figura 14: El elemento HACNS96 dirige la expresión de EGFP a diferentes lugares del sistema nervioso en desarrollo en peces cebrá mosaico y líneas estables. (A) Esquema de la construcción basada en el sistema Tol2 utilizada para generar peces transgénicos conteniendo el elemento HACNS96. (B-I) Microscopías fluorescentes (B, C, F y G) y campo claro (D, E, H e I) de embriones transgénicos estables (F-I) y mosaicos (B-E) a 24 y 48 hpf. (J) Expresión de EFGP en línea transgénica representativa a 72 hpf. (K-V) Co-detección de EGFP y el marcador HuC/D por inmunohistoquímica en cortes coronales sobre la línea representativa mostrada en (J) a 72 hpf. TeO, techo óptico. Hip, hipotálamo. H, hipófisis. MO, médula oblonga. ME, médula espinal.

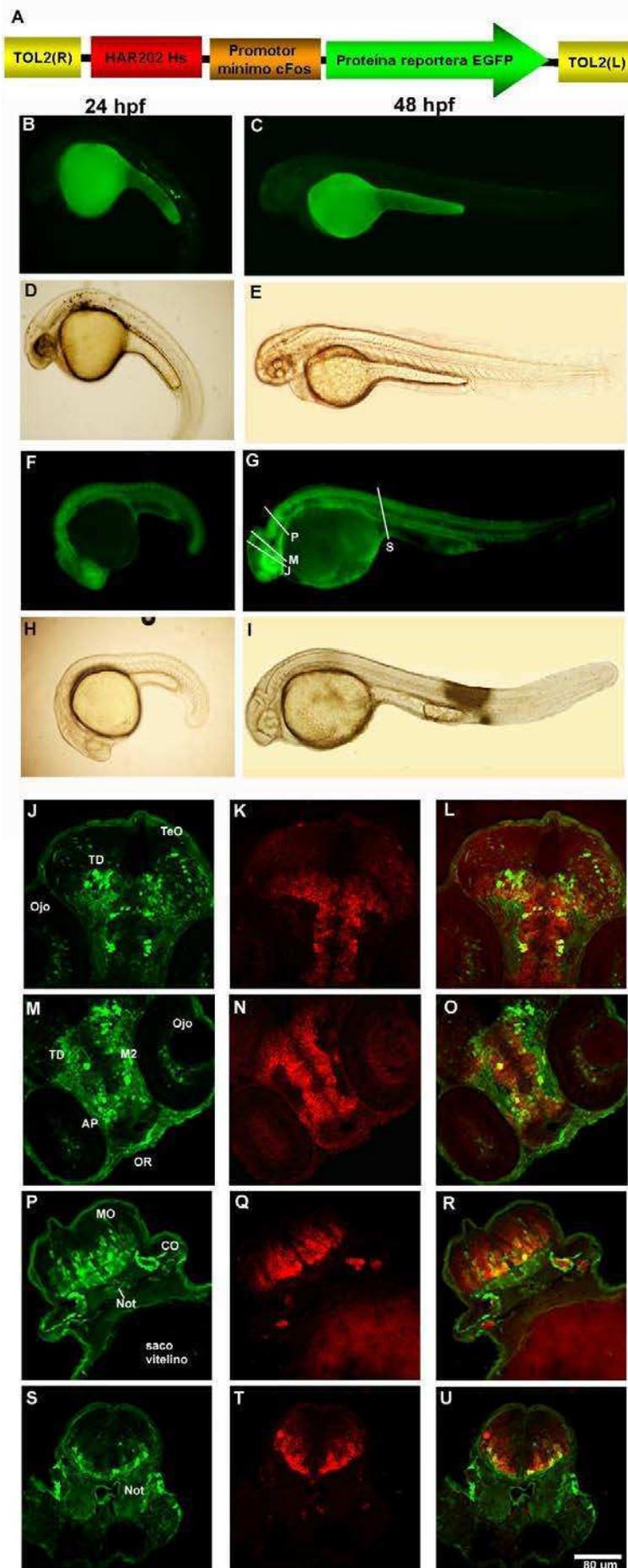


Figura 15: Diferencias en el patrón de expresión dirigido por los elementos HAR202Hs y HAR202Pt en peces cebra transgénicos. (A) Esquema de la construcción basada en el sistema Tol2 utilizada para generar peces transgénicos conteniendo el elemento HAR202. (B-I) Microscopías fluorescentes (B, C, F y G) y campo claro (D, E, H e I) de embriones transgénicos estables para el elemento HAR202Pt (F-I) y mosaicos para el elemento HAR202Hs (B-E) a 24 y 48 hpf. (J-U) Co-detección de EGFP y el marcador HuC/D por inmunohistoquímica en cortes coronales sobre línea representativa del elemento HAR202Pt a 48 hpf. TeO, techo óptico. TD, tegumento dorsal. Not, notocorda. AP, área preóptica. CO, cápsula ótica. MO, médula oblonga.

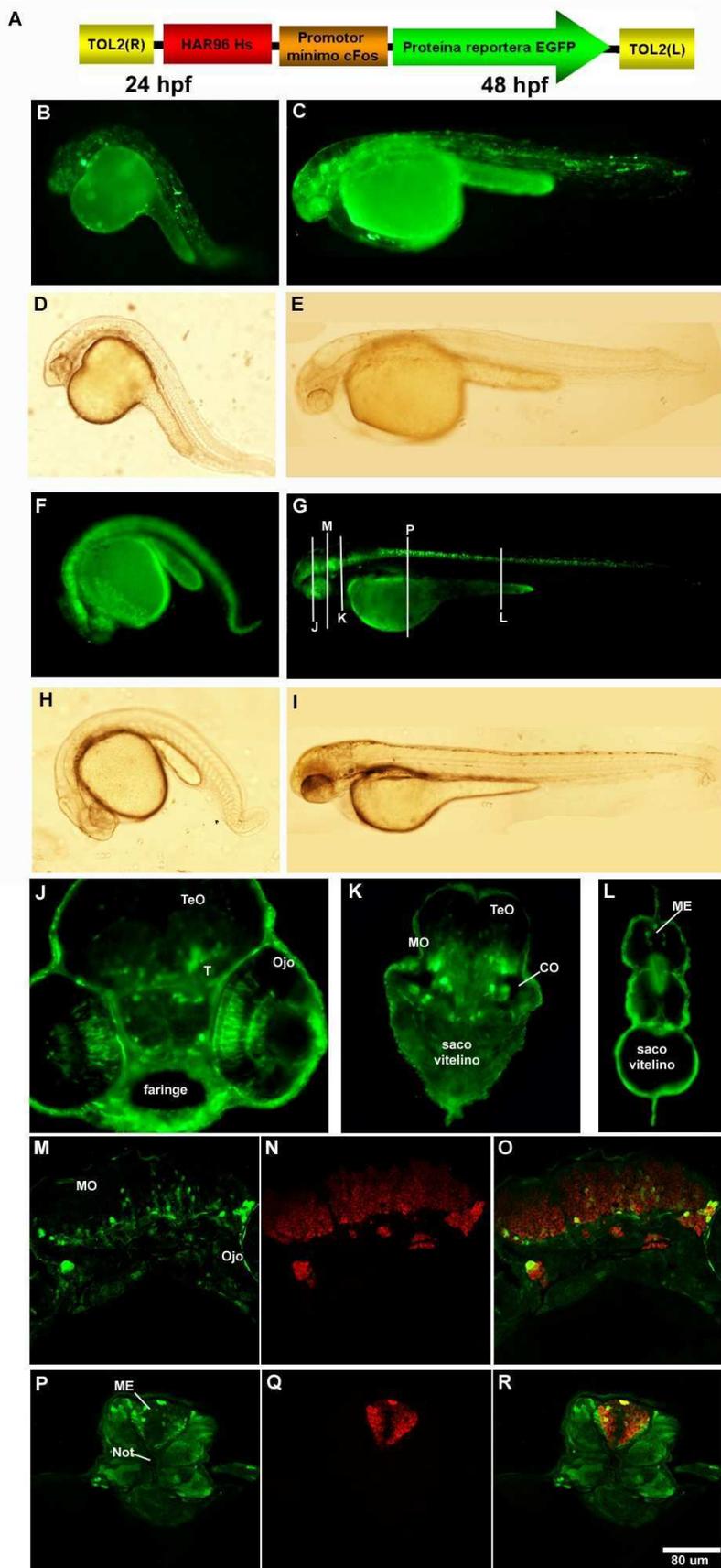


Figura 16: El elemento HAR96 dirige la expresión de EGFP a diferentes lugares del sistema nervioso en desarrollo en peces cebra mosaico y líneas estables. (A) Esquema de la construcción basada en el sistema Tol2 utilizada para generar peces transgénicos conteniendo el elemento HAR96. (B-I) Microscopías fluorescentes (B, C, F y G) y campo claro (D, E, H e I) de embriones transgénicos estables (F-I) y mosaicos (B-E) a 24 y 48 hpf. (J-V) Co-detección de EGFP y el marcador HuC/D por inmunohistoquímica en cortes coronales sobre la línea representativa mostrada en (I) a 48 hpf. TeO, techo óptico. T, tálamo. Not, notocorda. CO, cápsula ótica. MO, médula oblonga. ME, médula espinal.

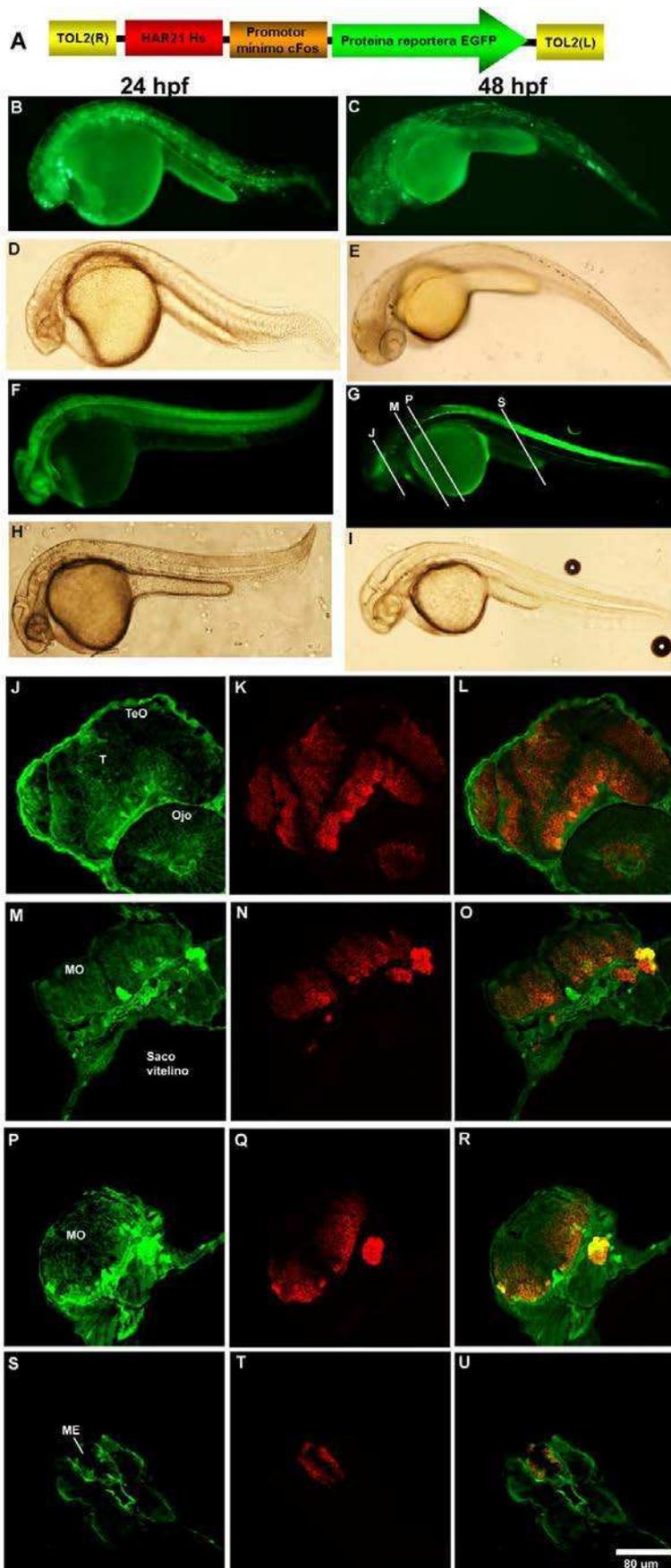


Figura 17: El elemento HAR21 dirige la expresión de EGFP a diferentes lugares del sistema nervioso en desarrollo en peces cebra mosaico y líneas estables. (A) Esquema de la construcción basada en el sistema Tol2 utilizada para generar peces transgénicos conteniendo el elemento HAR21. (B-I) Microscopías fluorescentes (B, C, F y G) y campo claro (D, E, H e I) de embriones transgénicos estables (F-I) y mosaicos (B-E) a 24 y 48 hpf. (J-U) Co-detección de EGFP y el marcador HuC/D por inmunohistoquímica en cortes coronales sobre la línea representativa mostrada en (G) a 48 hpf. TeO, techo óptico. T, tálamo. MO, médula oblonga. ME, médula espinal.

Análisis comparativo de los elementos HAR21, HAR89 y HAR202 en líneas de peces transgénicos

Con el objetivo de estudiar en más detalle posibles cambios evolutivos en la actividad en alguno de los *enhancers* recientemente descubiertos, se realizó un estudio comparativo de expresión de los elementos HAR21, HAR89 y HAR202, utilizando para ello líneas de peces transgénicos.

Por un lado, se establecieron cinco líneas de peces transgénicos llevando el elemento HAR21Hs y dos líneas llevando la región ortóloga en chimpancé. Comparando los resultados obtenidos, no se observaron cambios apreciables en el patrón de expresión a 48 hfp entre estos elementos (Figura 13, A; tabla 5). Ambos *enhancers* muestran expresión consistente en el cerebro y médula espinal. Además, estos elementos también dirigen la expresión de EGFP a corazón y aleta pectoral. La única diferencia de expresión detectada estuvo a nivel de los arcos branquiales donde tres de las cuatro analizadas llevando el elemento HAR21Hs mostraron expresión mientras que solo expresó allí una de las dos líneas generadas conteniendo el elemento HAR21Pt (Figura 13, A; tabla 5). No obstante, los ensayos de expresión transiente sugieren que probablemente este sea un dominio de expresión compartido por ambos elementos (tabla 4).

Por otro lado, se generaron dos líneas de peces transgénicos llevando el elemento HAR89Hs y dos líneas llevando la región ortóloga en chimpancé. Nuevamente en este caso, no se observaron cambios en el patrón de expresión a 48 hpf entre estos elementos (figura 13, C; tabla 5). Ambos *enhancers* muestra expresión consistente en el SNC, corazón, arcos branquiales y aleta pectoral (figura 13, C; tabla 5). Si bien hasta el momento no existe evidencia que la evolución acelerada experimentada por los elementos HAR21 y HAR89 en el linaje humano haya conducido a modificaciones en la actividad de estos *enhancers*, estos resultados no excluyen la posibilidad que los patrones de expresión de los elementos ortólogos se diferencien en otros estadios del desarrollo aquí no estudiados. Adicionalmente, ensayos funcionales realizados en otros modelos experimentales (como ser ratones transgénicos) podrán mostrar cambios sutiles de expresión no apreciables en el modelo de pez cebra.

Por otro lado, se generaron cuatro líneas de peces transgénicos llevando el elemento HAR202Hs y cuatro líneas llevando la región ortóloga en chimpancé. Se observó alta variabilidad en el patrón de expresión de las líneas generadas con el elemento HAR202Hs. En una línea, se observó expresión de EGFP en corazón y aleta pectoral pero no en SNC. En la segunda se detectaron bajos niveles de expresión en rombencéfalo y médula espinal. Otras dos líneas adicionales resultaron negativas para

la expresión de EFGP. Por otro lado, las cuatro líneas llevando el elemento HAR202Pt mostraron un extenso y prominente patrón de expresión en SNC, así como también en arcos faríngeos y aleta pectoral (figura 13, B; tabla 5). Estos resultados son consistentes con los encontrados en el ensayo de expresión transiente (figura 15, tabla 4). Mientras el elemento HAR202Pt es capaz de activar fuerte y consistentemente la expresión de EGFP, el elemento humano no tiene esta misma capacidad, mostrando gran variabilidad entre línea y línea.

Para tratar de explicar esta posible diferencia en la actividad de los elementos HAR202Hs y HAR202Pt, se estudiaron los sitios de unión para factores de transcripción en sus secuencias. Se encontró que, comparado con el elemento HAR202Pt, el elemento HAR202Hs perdió dos sitios: un sitio para *STAT5* y otro para *BCL6* (tabla suplementaria s3). Ambos factores de transcripción se expresan en el cerebro en desarrollo (Markham, Schuurmans y Weiss 2007; Leamey et al. 2008; Tiberi et al. 2012). Por un lado, *BCL6* codifica para un factor proneurogénico necesario para una correcta neurogénesis en la corteza cerebral del ratón (Tiberi et al. 2012). Por otro lado, *STAT5* juega un rol clave en la generación de interneuronas corticales y en el posicionamiento de los axones en la médula espinal y el telencéfalo durante el desarrollo de SNC (Markham, Schuurmans y Weiss 2007). Serán necesarios más experimentos para confirmar este posible caso de pérdida de función del elemento HAR202Hs (o alternatively, ganancia de función del elemento HAR202Pt en el linaje del chimpancé) y para entender cuales serían las causas moleculares de este fenómeno.

El elemento acelerado no codificante 2xHAR142 se comporta como un *enhancer* activo durante el desarrollo embrionario en ratones transgénicos

Hasta ahora, en este trabajo se mostró que el gen *NPAS3* es un caso notable en la evolución humana. Por un lado, se encontró que este gen contiene el agrupamiento más grande de HAEs en el genoma humano. Adicionalmente, se mostró que, al menos 11 de los 14 HAEs presentes en intrones de *NPAS3* son capaces de funcionar como *enhancers* en pez cebra, particularmente durante el desarrollo del sistema nervioso. Sin embargo, la distancia evolutiva que existe entre el pez cebra y humanos limita la extrapolación a *Homo sapiens* de los resultados obtenidos en el modelo vertebrado. Por esta razón, con el objetivo de explorar el impacto que los cambios filogenéticos experimentados en los HAEs de *NPAS3* pudieron tener sobre la evolución del cerebro humano, el siguiente paso fue estudiar en más detalle la función y origen evolutivo de uno de estos elementos en un modelo mamífero (particularmente, en ratón). De entre los 14 HAEs de *NPAS3*, se seleccionó el elemento 2xHAR142. Este elemento comprende 165 pb localizadas en el 5to intrón de *NPAS3* (chr14: 33,118,402-33,118,566, NCBI36/hg18; figura 6). Adicionalmente, el elemento 2xHAR142 es capaz de actuar como un *enhancer* en peces transgénicos (tabla 4).

Con el objetivo de estudiar si el elemento 2xHAR142 tiene la capacidad de actuar como un *enhancer* activo durante el desarrollo embrionario en un modelo mamífero, se amplificaron por PCR 502 pb conteniendo completamente las 165 pb el elemento humano 2xHAR142 en el contexto de la región conservada que lo contiene (definida por el programa PhastCons a partir de alineamientos entre los genomas de 17 especies de vertebrados) (Siepel et al. 2005). Esta región se denominó PC_2xHAR142Hs y se clonó río arriba del promotor mínimo del gen HSP68 de ratón regulando la expresión de la proteína reportera *LacZ* (Kothary et al. 1989) (Figura 18, A). El promotor mínimo usando en estos ensayos, en ausencia de elementos regulatorios adicionales, es incapaz de promover la expresión del gen reportero durante el desarrollo embrionario de ratón (Kothary et al. 1989). Además, este sistema ha sido usando exitosamente para estudiar la actividad de *enhancers* clonados río arriba de su secuencia (Poulin et al. 2005).

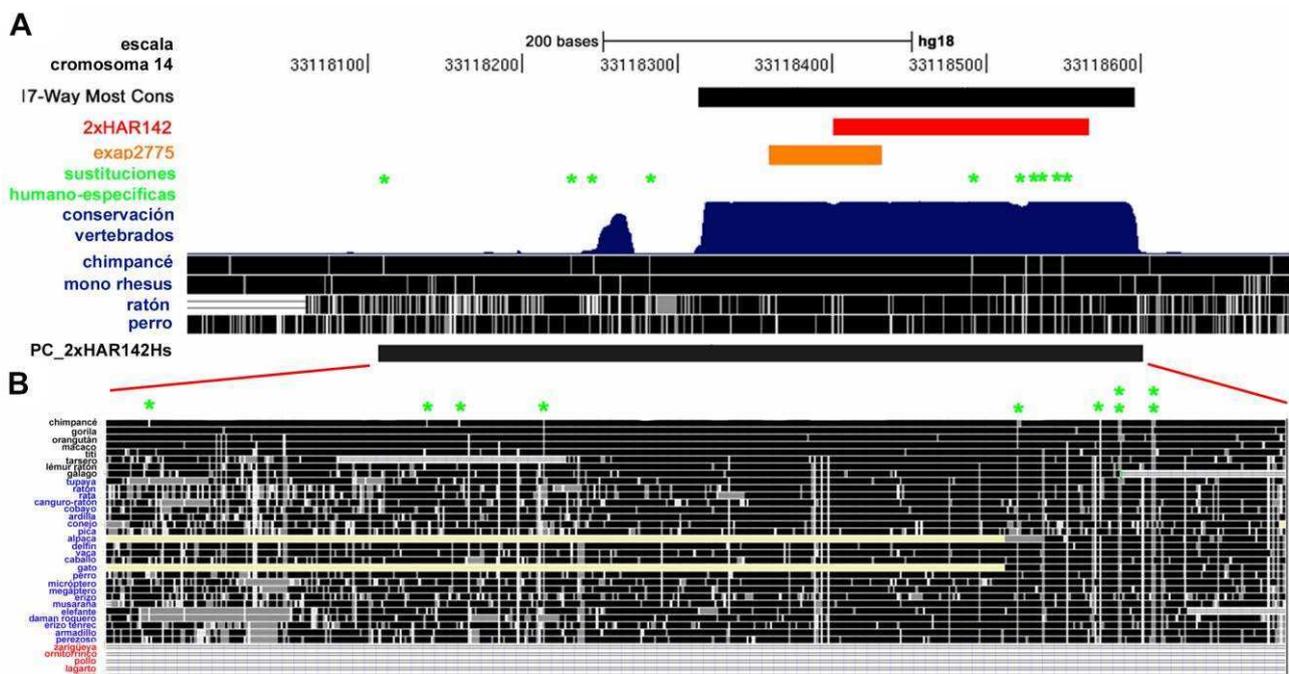


Figura 18: Diagrama y alineamiento de la secuencia del *enhancer* PC_2xHAR142Hs. (A) Esquema de la región clonada en el transgén PC_2xHAR142Hs conteniendo la región conservada descrita por el programa PhastCons, el elemento 2xHAR142 y el elemento móvil exaptado exapt2775. (B) Conservación entre vertebrados de la secuencia del elemento PC_2xHAR142Hs de acuerdo al USCS Genome Browser. Con asteriscos verdes se indican las sustituciones humano-chimpancé.

Se generaron 6 líneas independientes de ratones transgénicos conteniendo esta construcción y se estudió la expresión de la proteína reportera a tres edades distintas del desarrollo embrionario del ratón: 10.5, 12.5 y 14.5 dpc (figura 19). A pesar que se observó cierta variabilidad entre línea y línea (probablemente debido a diferencias en el número y posición de inserción del transgén en el

genoma) (figura 21), se encontró que el elemento PC_2xHAR142Hs tiene la capacidad intrínseca de actuar como un *enhancer*, activando la expresión de *LacZ* de forma consistente en varios dominios de expresión durante el desarrollo embrionario del ratón. Particularmente, en las 6 líneas generadas se observó expresión consistente de la proteína reportera *LacZ* en diferentes lugares del sistema nervioso en desarrollo, incluyendo el cerebro y la médula espinal, y en las tres edades analizadas (Figura 19; figura 21). A 10.5 y 12.5 dpc se observó expresión de *LacZ* en el rombencéfalo, mesencéfalo y prosencéfalo (Figura 19, A y B). Particularmente, a 12.5 dpc y 14.5 dpc dentro del prosencéfalo, la expresión está principalmente localizada en la corteza en desarrollo, en el neuroepitelio del hipocampo, el tálamo ventral y el hipotálamo (Figura 19, B y C). A estas mismas edades y en el mesencéfalo, se observó expresión en el tegumento de los pedúnculos cerebrales. En el rombencéfalo, se observó expresión prominente en el neuroepitelio del cerebelo y en el fascículo, entre otros. Fuera del sistema nervioso, se observó expresión de *LacZ* en los primordios de miembros, ojos, oído interno en desarrollo, primordios de las vértebras, arcos faríngeos, entre otros (Figura 19, B y C). Estos resultados demuestran que el elemento PC_2xHAR142Hs tiene la capacidad intrínseca de actuar como un *enhancer* activo durante el desarrollo embrionario.

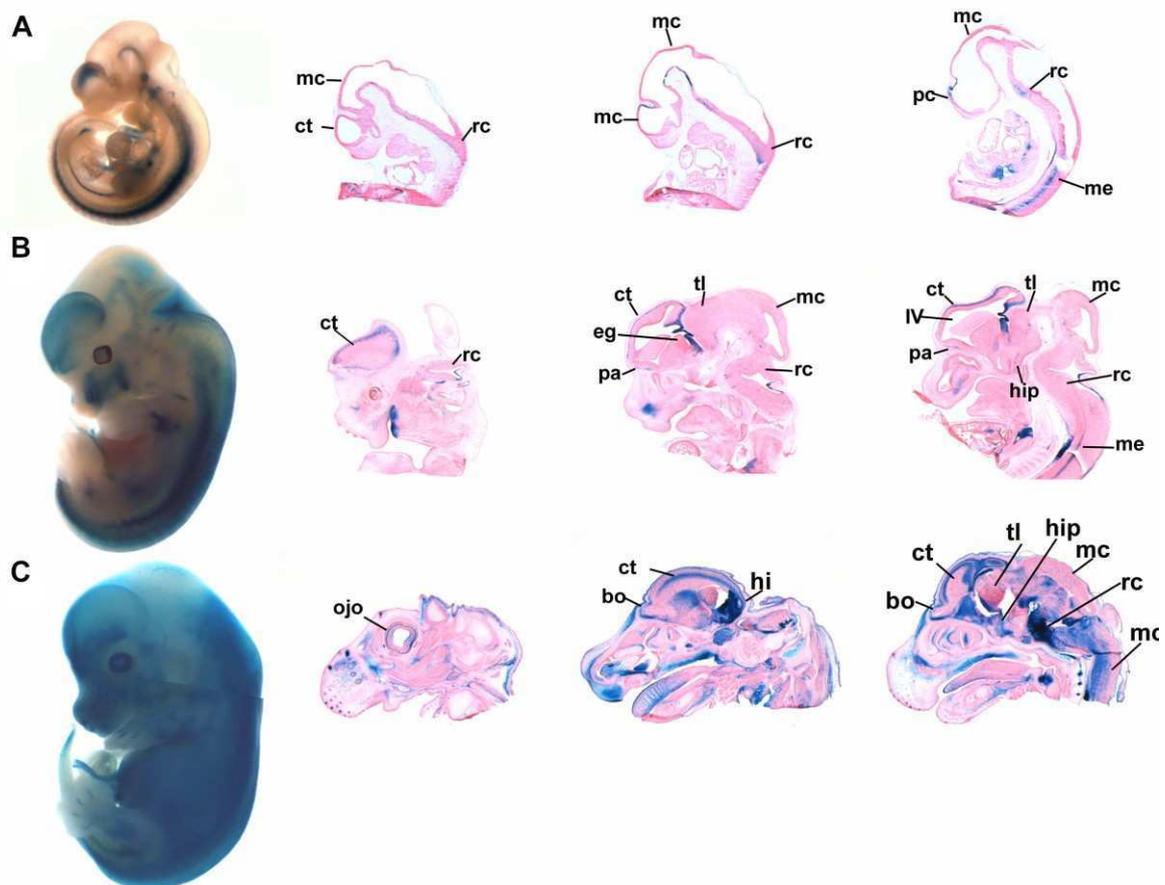


Figura 19: Línea de ratón transgénico mostrando el patrón de expresión de *LacZ* representativo el elemento PC_2xHAR142Hs en embriones completos y en cortes sagitales a (A) 10.5 dpc, (B) 12.5 dpc y (C) 14.5 dpc. Referencias: (ct) corteza en desarrollo, (mc) mesencéfalo, (rc) rombencéfalo, (me) médula

espinal, (pa) área preóptica, (eg) eminencia ganglionar, (tl) tálamo, (IV) cuarto ventrículo, (hip) hipotálamo, (bo) bulbo olfatorio, (hi) hipocampo.

Con el objetivo de determinar si la evolución acelerada en el linaje humano experimentada por el elemento 2xHAR142 humano (2xHAR142Hs) pudo llevar a un cambio a nivel filogenético en el patrón de expresión del elemento PC_2xHAR142Hs, se comparó el patrón de expresión producido por este *enhancer* con los generados por regiones genómicas que contienen a los elementos ortólogos en chimpancé y ratón del elemento 2xHAR142Hs. Para ello, se generaron y analizaron a 12.5 dpc tres líneas independientes de ratones transgénicos conteniendo la región ortóloga de 502 pb en chimpancé del *enhancer* PC_2xHAR142Hs. Este elemento, que lleva el elemento ortólogo en chimpancé del HAE 2xHAR142 en el contexto de la región conservada que lo contiene, se llamó PC_2xHAR142Pt y se clonó río arriba del promotor mínimo HSP68 de ratón, dirigiendo la expresión de la proteína reportera *LacZ*. Adicionalmente, se generaron y analizaron a 12.5 dpc tres líneas independientes de ratones transgénicos donde la proteína reportera *LacZ* quedó bajo la regulación de una región de 1.5 kpb que tiene el elemento 2xHAR142Mm junto con la región conservada que lo contiene. Esta región, llamada PC_2xHAR142Mm, se clonó nuevamente río arriba del promotor mínimo del gen HSP68 de ratón.

Analizando el patrón de expresión del elemento PC_2xHAR142Pt a 12.5 dpc se encontró que esta construcción produce un patrón de *LacZ* con varias similitudes con el patrón observado para el elemento PC_2xHAR142Hs (figura 20; figura 21). Por ejemplo, como ya fuera hallado para el elemento humano, los dominios de expresión del elemento PC_2xHAR142Pt incluyen la médula espinal, el rombencéfalo y el mesencéfalo (aunque este último con niveles de expresión más bajos a los hallados para el elemento PC_HAR142Hs) (figura 20, A y B; figura 21). Por fuera del SNC, se encontró que el elemento PC_2xHAR142Pt dirige la expresión a los primordios de los miembros y de las vértebras, estando también estos dominios de expresión compartidos entre los dos *enhancers* primates. Sin embargo, a nivel del prosencéfalo, en los ratones llevando la construcción PC_2xHAR142Pt sólo se observó expresión en el área preóptica; mientras que en la corteza en desarrollo no se detectó expresión de *LacZ* (figura 20, B; figura 21). Este patrón contrasta con la abundante expresión de *LacZ* desarrollada por el elemento PC_2xHAR142Hs en la corteza en desarrollo a 12.5 dpc (figura 19, B; figura 20 A y B; figura 21).

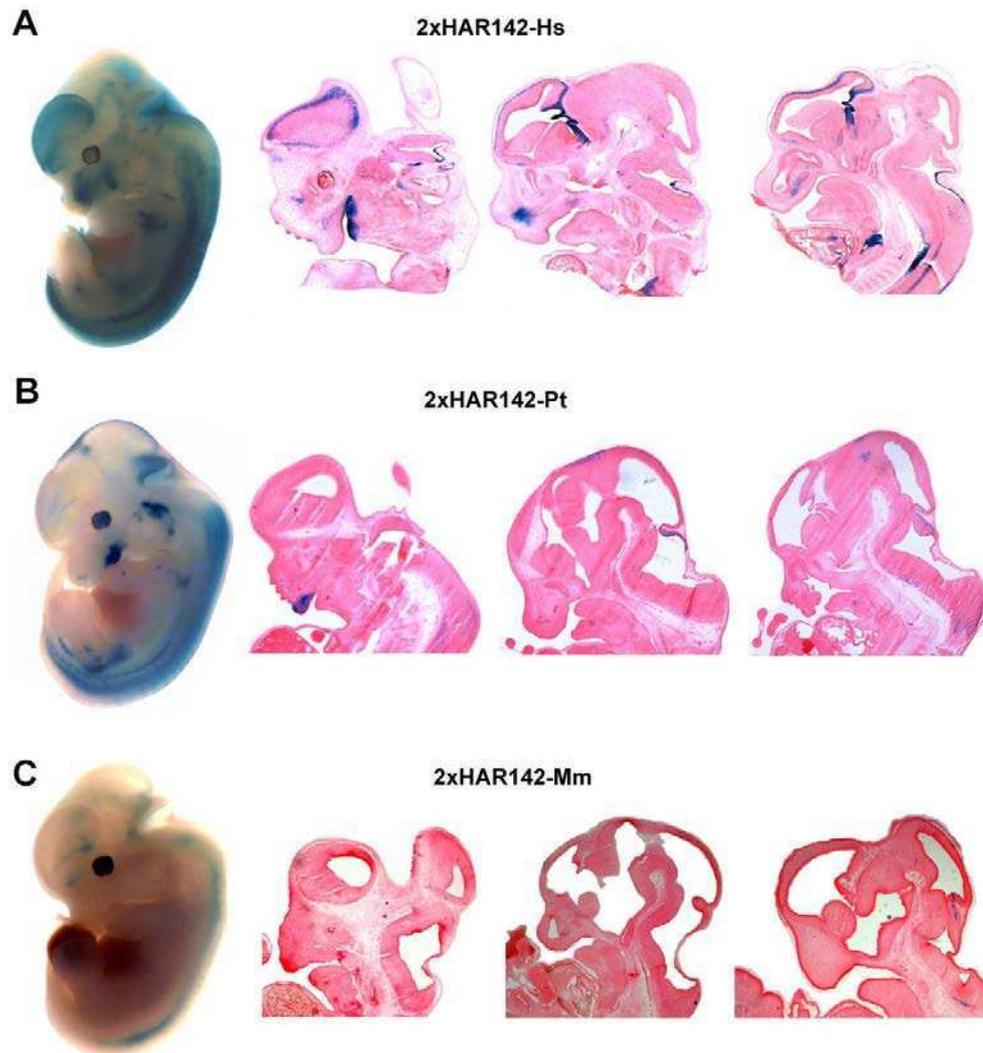
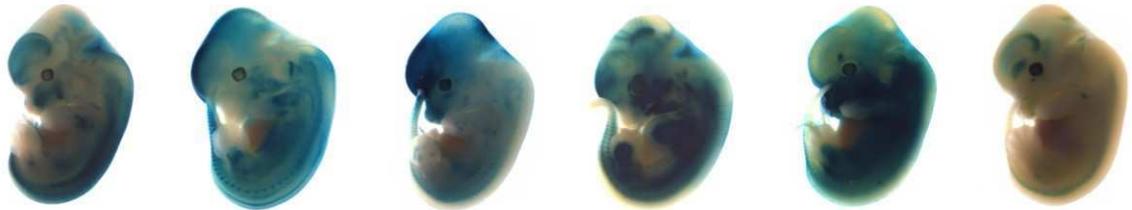


Figura 20: Análisis de expresión comparativo entre los elementos PC_2xHAR142Hs, PC_2xHAR142Pt y PC_2xHAR142Mm en ratones transgénicos. Patrones de expresión de *LacZ* representativos de los elementos (A) PC_2xHAR142Hs, (B) PC_2xHAR142Pt y (C) PC_2xHAR142Mm en embriones completos y en cortes sagitales de ratón a 12.5 dpc.

Por otro lado, se encontró que 3 líneas independientes de ratones conteniendo la región PC_2xHAR142Mm expresaban *LacZ* consistentemente en un dominio más restringido respecto del dominio de expresión observado para los *enhancers* PC_2xHAR142Hs y PC_2xHAR142Pt. Por un lado, el elemento PC_2xHAR142Mm dirigió la expresión de *LacZ* a la médula espinal, al rombencéfalo y, en bajo niveles, al mesencéfalo (figura 20, C; figura 21), estando estos dominios de expresión compartidos por los tres *enhancers* (figura 20; figura 21). Sin embargo, no se observó expresión de *LacZ* en los primordios de las vértebras y/o miembros, como sí fuera observado para los *enhancers* primates (figura 20; figura 21). Se necesitará más trabajo para determinar si estos dos últimos dominios de expresión son una innovación surgida en el linaje de los primates o si se

perdieron en el linaje de los ratones. Por otro lado, en el prosencéfalo, se observó expresión en el área preóptica pero no se detectó expresión en la corteza en desarrollo, nuevamente en gran contraste con lo observado para el *enhancer* PC_2xHAR142Hs (figura 20; figura 21). Estos resultados sugieren que el proceso de evolución acelerada experimentada por el elemento 2xHAR142 en el linaje humano pudo conducir a un cambio funcional en la actividad de este *enhancer*, particularmente en la corteza en desarrollo.

A PC_2xHAR142Hs



B PC_2xHAR142Pt



C PC_2xHAR142Mm



Figura 21: Análisis de expresión comparativo entre los elementos PC_2xHAR142Hs, PC_2xHAR142Pt y PC_2xHAR142Mm en ratones transgénicos. Líneas independientes mostrando los patrones de expresión de *LacZ* de los elementos (A) PC_2xHAR142Hs, (B) PC_2xHAR142Pt y (C) PC_2xHAR142Mm en embriones completos de ratón a 12.5 dpc.

Estudiando la región conservada del *enhancer* PC_2xHAR142Hs se encontró que tienen un 98% de identidad de secuencia con la región ortóloga presente en el elemento PC_2xHAR142Pt (6 sustituciones específicas de los humanos en la región altamente conservada definida por el programa PhastCons, figura 18, A y B) y 85% de identidad con la secuencia ortóloga en ratón (con 37 sustituciones a lo largo de la región conservada PhastCons) (figura 18, A y B). Esta información sugiere que las sustituciones acumuladas en esta secuencia después de la separación de los linajes que dieron origen a los chimpancés y a los humanos podrían ser responsables de la adquisición de

nuevos territorios de expresión por parte del elemento 2xHAR142Hs, particularmente en la corteza en desarrollo. Analizando los cambios en los sitios predichos de unión para factores de transcripción (tabla suplementaria 3), se observa que entre las secuencias ortólogas de chimpancé y humano del elemento 2xHAR142 existen múltiples cambios (ya sea ganancias o pérdidas) en los sitios probables de unión para estas proteínas. Se requerirá trabajo adicional para estudiar cual/es de estos sitios (u otros aún no identificados) son las responsables del cambio heterotópico en el patrón de expresión de estos dos *enhancers*.

Es importante destacar que existe una diferencia de tamaño en el largo de los *enhancers* PC_2xHAR142Hs/Pt y PC_2xHAR142Mm (con 502 pb y 1.5 kpb vs. respectivamente). Sin embargo, estas tres construcciones (PC_2xHAR142Hs, PC_2xHAR142Pt y PC_2xHAR142Mm) llevan las regiones conservada ortólogas (definida a partir de los genomas de 17 especies de vertebrados con el programa PhastCons)(Siepel et al. 2005) que contiene a los elementos 2xHAR142Hs, 2xHAR142Pt y 2xHAR142Mm respectivamente y solo se diferencian en regiones no conservadas (figura 18). La interpretación de los resultados donde el *enhancer* PC_2xHAR142Hs cambió su actividad en el linaje que condujo a los humanos asume que las regiones no conservadas presentes en los elementos PC_2xHAR142Hs, PC_2xHAR142Pt y PC_2xHAR142Mm serían no funcionales (aunque no se puede descartar que existan regiones funcionales adicionales en estos *enhancers* no detectables por conservación entre especies de mamíferos).

El elemento 2xHAR142 dirige la expresión de *LacZ* a un dominio que se superpone con el dominio de expresión del gen *NPAS3*.

Como se detalló anteriormente, uno de los grandes desafíos cuando se trabaja con elementos regulatorios del genoma (como ser los *enhancers*), consiste en identificar sus genes blanco. Una primera aproximación consiste en asignar como probable blanco de un *enhancer* al gen que tiene más próximo (Heintzman y Ren 2009). Sin embargo, esta hipótesis se vuelve más firme si se confirma que el *enhancer* y el probable gen blanco comparten, al menos en parte, sus dominios de expresión (el dominio endógeno del gen y el de la proteína reportera, por ejemplo).

A continuación, se compararon los patrones de expresión de los *enhancers* PC_2xHAR142Mm y PC_2xHAR142Pt con el patrón de expresión endógeno del gen *Npas3* durante el desarrollo del sistema nervioso central del ratón con el objetivo de determinar si existía algún solapamiento entre los dominios de expresión del gen y de los *enhancers*. Se eligieron estudiar específicamente estos dos *enhancers* porque el primero proviene de la misma especie que el gen endógeno bajo estudio (las condiciones óptimas para ensayar los dominios de expresión de regulador y gen blanco) y el

segundo porque produce un patrón de expresión similar al *enhancer* de ratón pero sus líneas transgénicas expresan mayores niveles de la proteína reportera *LacZ* (facilitando la detección por medio de anticuerpos) (figura 20, figura 21). Adicionalmente, como hasta ahora el patrón de expresión del gen *NPAS3* solo fue parcialmente descrito (Brunskill et al. 1999; Gould y Kamnasaran 2011), se propuso producir simultáneamente un análisis detallado de los dominios de expresión de *Npas3* durante el desarrollo embrionario de ratón. Con el objetivo de detectar la proteína *Npas3* se probaron diferentes anticuerpos comerciales (Santa Cruz Biotechnology, cat#sc-50255 y Sigma-Aldrich cat#HPA002892) y protocolos de inmunohistoquímica (IHQ) en cortes histológicos de cerebro de ratón en desarrollo. Lamentablemente, ninguno de estos anticuerpos y/o protocolos resultó eficiente en la detención de este factor de transcripción, al menos durante las edades embrionarias ensayadas. Debido a lo anterior, se siguió la expresión de *Npas3* recurriendo a ensayos de hibridación *in situ* (ISH) utilizando una sonda diseñada para detectar el ARNm de este gen (Brunskill et al. 1999). Se realizaron ensayos de ISH a tres edades del desarrollo embrionario de ratón (10.5, 12.5 y 14.5 dpc) sobre líneas transgénicas llevando los *enhancers* PC_2xHAR142Mm y PC_2xHAR142Pt. Por otro lado, también se estudió los tipos celulares en los cuales se expresa *Npas3* utilizando una combinación de ensayos de ISH y codetección por IHQ de diferentes marcadores celulares.

A 10.5 dpc se observó que *Npas3* se expresa todo a lo largo del neuroepitelio del rombencéfalo, mesencéfalo y prosencéfalo (figura 22, A, B y C; figura 23, A y F). También se codetectó la expresión de *Npas3* y del marcador de precursor neural indiferenciado (PNI) *SOX2* observándose que numerosas células expresan conjuntamente *SOX2* y *Npas3* en prosencéfalo y rombencéfalo (figura 23, B, C, D, E; G, H, I y J). Por otro lado, en estas mismas zonas, se detectó menor densidad de células expresando *Npas3* en las capas más externas del neuroepitelio, donde se expresa el marcador de neuronas postmitóticas *DCX* (figura 23, A y F).

A 12.5 y 14 dpc, se observó abundante expresión de *Npas3* en el sistema nervioso en desarrollo, incluidos el rombencéfalo, mesencéfalo, prosencéfalo y médula espinal, así como también en otras áreas como ser la dermis, miembros en desarrollo, oído interno, el ojo y el epitelio olfatorio (figura 22, D - K; figura 23, K, P, V, AA y FF). A 12.5 dpc, en el rombencéfalo, *Npas3* se expresa todo a lo largo del neuroepitelio encontrándose especialmente altos niveles de expresión en la zona ventricular (zona de alta proliferación celular y rica en progenitores neuronales indiferenciados) (Rakic 2000). Esta región además está densamente poblada por células que expresan el marcador de PNI *SOX2* y es una zona en la cual también se detecta expresión de *Npas3* (figura 23, V, W, X, Y y Z). Adicionalmente, se observó menor expresión de *Npas3* por fuera de la zona ventricular y en

territorios donde se expresa conjuntamente los marcadores de neuronas postmitóticas *DCX* y β -*tubulina III* o *Tuj* (figura 23, P, Q, R, S, T y AA, BB, CC, DD y EE). Similares resultados se observaron en la médula espinal (figura 23, K, L, M, N, O) y la corteza en desarrollo (figura 23 FF, GG, HH, II y JJ).

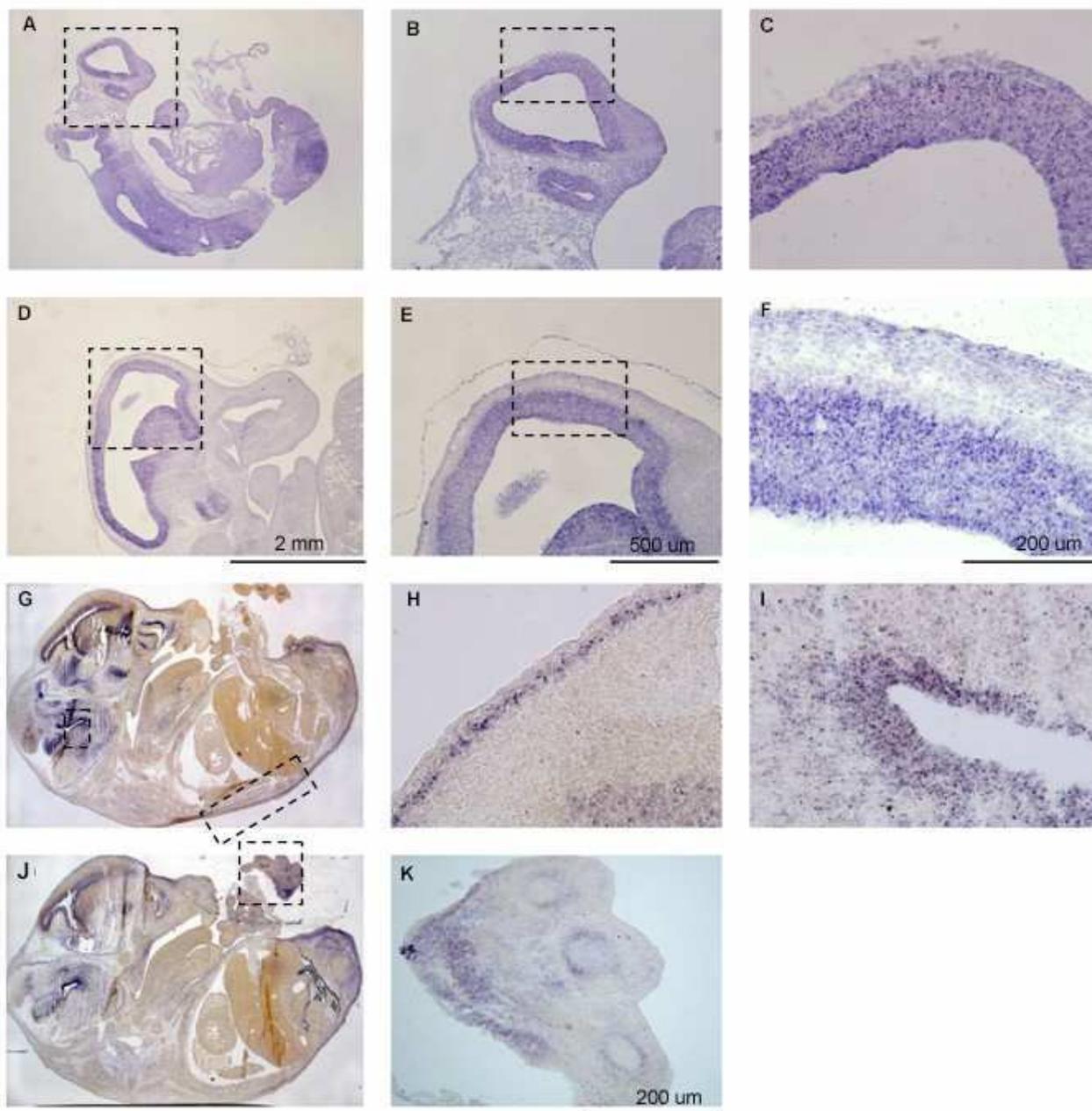


Figura 22: Micrografías en campo claro mostrando la expresión del ARNm de *Npas3* en cortes sagitales durante el desarrollo embrionario de ratón. (A-C) 10.5 dpc, (D-F) 12.5 dpc y (G-K) 14.5 dpc. Detalle de la expresión de *Npas3* en (B-C y E-F) telencéfalo, (H) dermis, (I) rombencéfalo, (K) miembro anterior. Con recuadros negros se indican las zonas que se muestran en más detalle. Las microscopías G y J tienen 1 cm de lado.

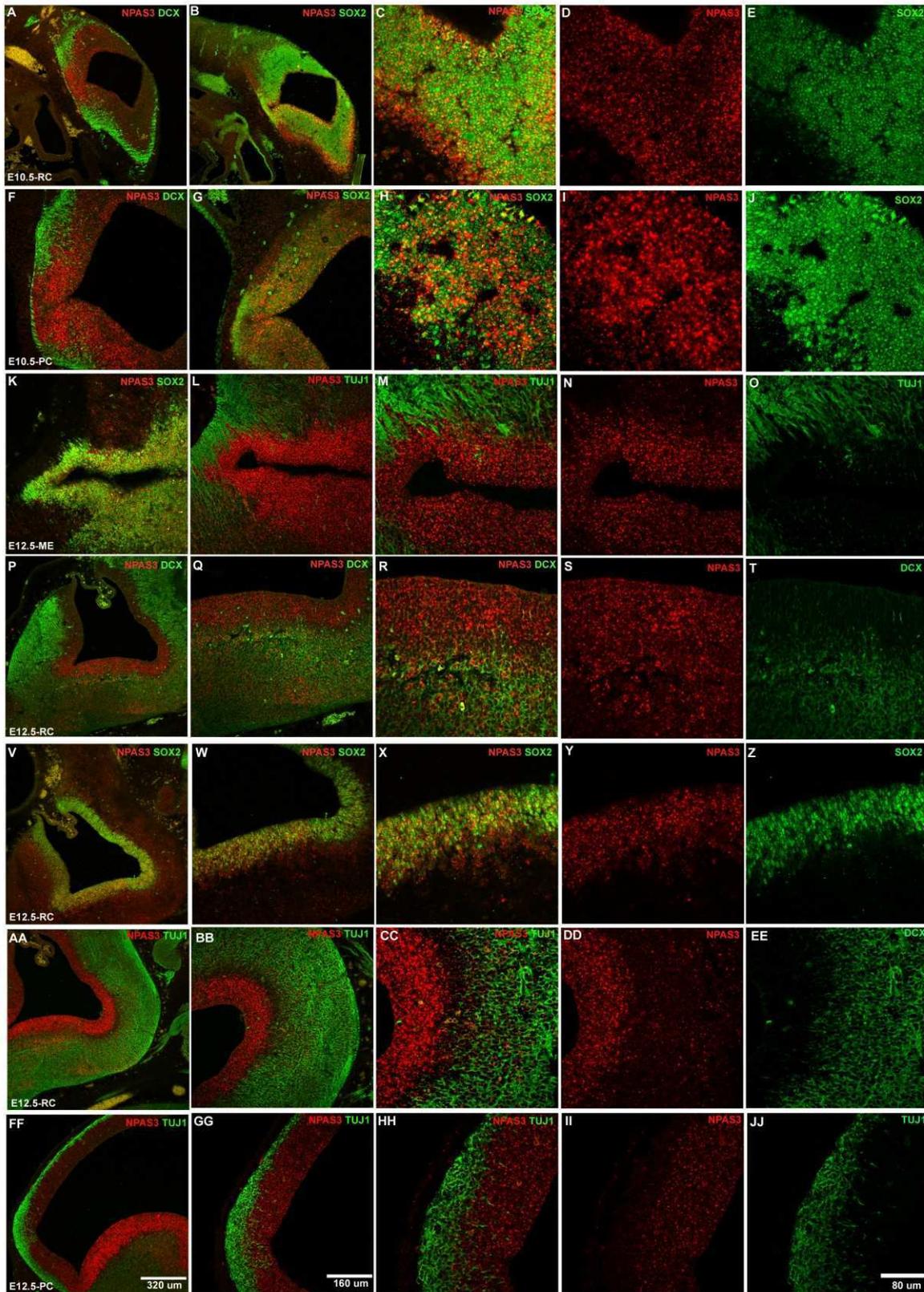


Figura 23: Expresión de *Npas3* en el sistema nervioso en desarrollo. ISH fluorescente contra el ARNm de *Npas3* combinado con IHQ contra los marcadores tempranos de neuronas *DCX* y β -*tubulina III* y el marcador de PNI *SOX2* en cortes sagitales de embriones de ratón a 10.5 y 12.5 dpc. Expresión de *Npas3* a 10.5 dpc en el rombencéfalo (A-E) y en el prosencéfalo (F-J) en combinación con los marcadores *DCX* (A, F) y *SOX2* (B-E y G-J). A 12.5, la expresión de *Npas3* se muestra en la médula espinal (K-O), en el rombencéfalo (P-EE) y en el prosencéfalo (FF-JJ). En la médula espinal, la

expresión de *Npas3* se muestra en combinación con los marcadores *SOX2* (K) y *TUJ1* en baja (L) y alta magnificación (M-O). En el rombencéfalo, la expresión de *Npas3* se muestra en combinación con los marcadores *DCX* (P-T), *SOX2* (V-Z) y *TUJ1* (AA-EE). En el prosencéfalo, la expresión de *Npas3* se muestra en combinación con el marcador *TUJ1* a baja (FF y GG) y alta magnificación (HH y JJ). Referencias: (RC) rombencéfalo, (PC) prosencéfalo, (ME) médula espinal, (CT) corteza en desarrollo. Las referencias de los marcadores utilizados se indican en color sobre las micrografías.

El paso siguiente fue detectar simultáneamente el ARNm de *Npas3* por ISH y la proteína reportera *LacZ* por IHQ. Lamentablemente, nuevamente se encontraron varias dificultades técnicas para codetectar *Npas3* y *LacZ* ya que no se consiguieron anticuerpos anti-*LacZ* capaces de reconocer a la proteína luego del protocolo de ISH (que resulta particularmente drástico para los tejidos, posiblemente desnaturalizando muchos epítopes). Es por esta razón que se estudiaron los dominios de expresión de *Npas3* y los producidos por los *enhancers* PC_2xHAR142Mm y PC_2xHAR142Pt combinando ISH para *Npas3* e IHQ contra diferentes marcadores celulares y *LacZ*.

Se observó que la expresión de *LacZ* dirigida por los elementos PC_2xHAR142Mm y 2xHAR142Pt a 12.5 dpc se localiza ampliamente en los dominios de expresión de los marcadores *DCX* y *TUJ1* en el rombencéfalo y médula espinal (figura 24, E-L; figura 25, C-F, K-N, D-R, S-V y W-Z). Sin embargo, se encontró mínima superposición entre los dominios de *LacZ* y el marcador *SOX2* en esos lugares (figura 24, A-D y M-P; figura 25, G-J y AA-DD). Por otro lado, el patrón de expresión de *Npas3* a 12.5 dpc incluye el neuroepitelio completo de la médula espinal y el rombencéfalo (figura 23, K-N, P-S, V-Y, AA-DD; figura 25 E, F, I, J, M, N, CC y DD). El dominio de expresión de *Npas3* se superpone ampliamente al dominio de expresión de *SOX2* (figura 23, K y V-Z; figura 25 I, J, CC y DD) pero también alcanza los dominios de expresión de los marcadores *DCX* y *TUJ1* (figura 23, L-O, P-T y AA-EE; figura 25 E, F, M, N, W y X). En su conjunto, estos resultados indican que los *enhancers* 2xHAR142Mm y 2xHAR142Pt podrían dirigir la expresión de *LacZ* a un subdominio del dominio de expresión del gen *Npas3*. Esto refuerza la hipótesis que estos elementos podrían ser verdaderos *enhancers* de *Npas3*, participando de la regulación de la expresión de este gen. Los resultados conclusivos podrán ser aportados por estudios adicionales donde se deleccione específicamente el *enhancer* PC_2xHAR142Mm del genoma murino y se evalúe el impacto de la pérdida de este elemento sobre la expresión endógena de *Npas3*.

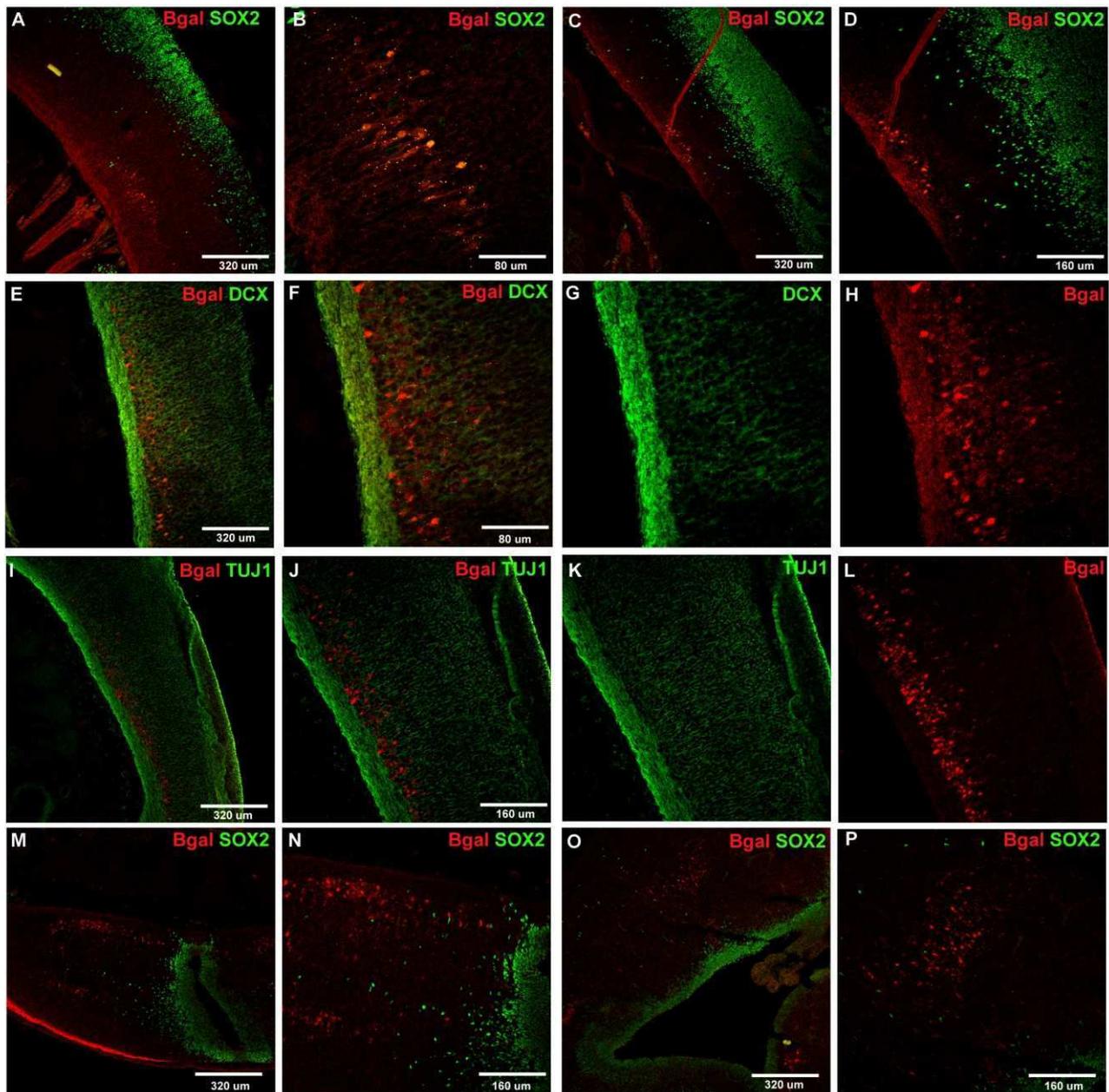


Figura 24: Patrón de expresión de *LacZ* (Bgal) a 12.5 dpc en diferentes tipos celulares en el sistema nervioso en desarrollo. Microfotografías de alta resolución mostrando el patrón de expresión de la proteína reportera *LacZ* bajo el control del *enhancer* PC_2xHAR142Mm a diferentes niveles de la médula espinal (A -N) y el rombencéfalo (O y P). La expresión de *LacZ* se muestra en combinación con los marcadores tempranos de linaje neuronal *DCX* (E, F y G) y *TUJ1* (I, J y K) y el marcador de PNI *SOX2* (A-D y M-P) en el sistema nervioso central. Las referencias de los marcadores utilizados se indican en color sobre las micrografías.

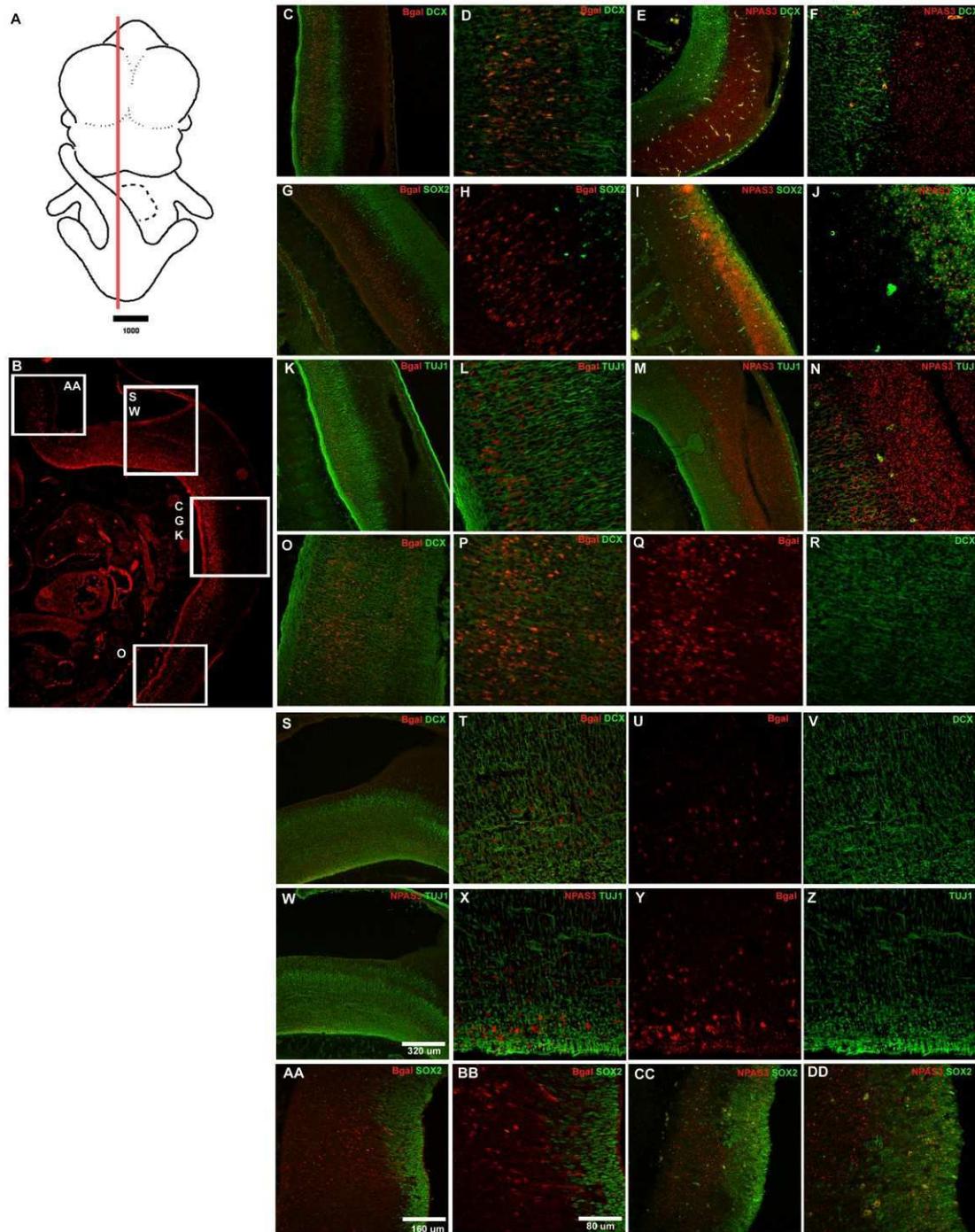


Figura 25: Patrón de expresión de *LacZ* a 12.5 dpc en diferentes tipos celulares en el sistema nervioso en desarrollo. (A) Esquema de un embrión completo de ratón a 12.5 dpc. La línea roja indica aproximadamente la altura de los cortes presentados a continuación. (B) Imagen de baja resolución de un corte histológico de un ratón transgénico expresando *LacZ* bajo el control del *enhancer* PC_2xHAR142Pt en diferentes zonas del rombencéfalo y la médula espinal. Los recuadros blancos sobre esta imagen indican aproximadamente la localización de las imágenes de alta resolución mostradas a continuación. (C-DD) Microfotografías de alta resolución mostrando el patrón de expresión de la proteína reportera *LacZ* bajo el control del *enhancer* PC_2xHAR142Pt a diferentes niveles de la médula espinal (C, D, G, H, K, L, P y O) y el rombencéfalo (S-Z y AA-BB). La expresión

de *LacZ* se muestra en combinación con los marcadores tempranos de linaje neuronal *DCX* (C, D, E, F, O, P, R, S, T, y V) y *TUJI* (K, L, M, N, W, X y Z) y el marcador de PNI *SOX2* (G, H, I, J, AA, BB, CC y DD) en el sistema nervioso central. Las referencias de los marcadores utilizados se indican en color sobre las micrografías.

Origen evolutivo de las regiones aceleradas

Como se detalló anteriormente, uno de mecanismos genéticos que pueden dar lugar al surgimiento de nuevos *enhancers* involucra la inserción de elementos transponibles (ET) (Rubinstein y de Souza 2013). Por la capacidad de movilizar a diferentes lugares del genoma secuencias con sitios para la unión de factores de transcripción, los ET pueden volverse elementos exaptados y, por ejemplo, adquirir una función como nuevos *enhancers* para aquellos genes próximos al sitio de inserción (Bourque et al. 2008; Smith et al. 2008; Studer et al. 2011; Romero, Ruvinsky y Gilad 2012).

Reconociendo la potencialidad de los ET de ser exaptados en regiones regulatorias, la siguiente actividad consistió en estudiar si existían casos de elementos con signos de evolución acelerada en el genoma humano para los cuales hubiera evidencia que provenían de un evento de exaptación de un elemento móvil (evidenciado por la alta conservación de la secuencia del ET, atribuible a un proceso de evolución purificadora luego de su inserción). Para realizar este análisis se recurrió a una base de datos pública de elementos derivados de ET en el genoma humano, no exónicos y altamente conservados en mamíferos, aquí llamados *exapted repeats 2007* (Lowe, Bejerano y Haussler 2007).

Para ello, se cruzaron la base de datos de HAEs y la de elementos móviles con evidencia de exaptación *exapted repeats 2007*. Se encontró que de los 1629 HAEs presentes en el genoma humano, solo 31 de ellos solapan con alguno de los 10.402 elementos móviles altamente conservados descritos por Lowe *et. al.* (Lowe, Bejerano y Haussler 2007). De entre ellos, 28 corresponden a *HACNS* y 3 a *2xHARs*. En particular, analizando el *locus* de *NPAS3*, se encontraron 10 elementos derivados de ET y conservados en sus intrones. De entre estos 10 elementos, existe un único elemento con evidencia de exaptación que se superponen con uno de los HAEs presentes en este *locus*: el *exap2775* que forma parte del elemento *2xHAR142* (figura 26, A y B).

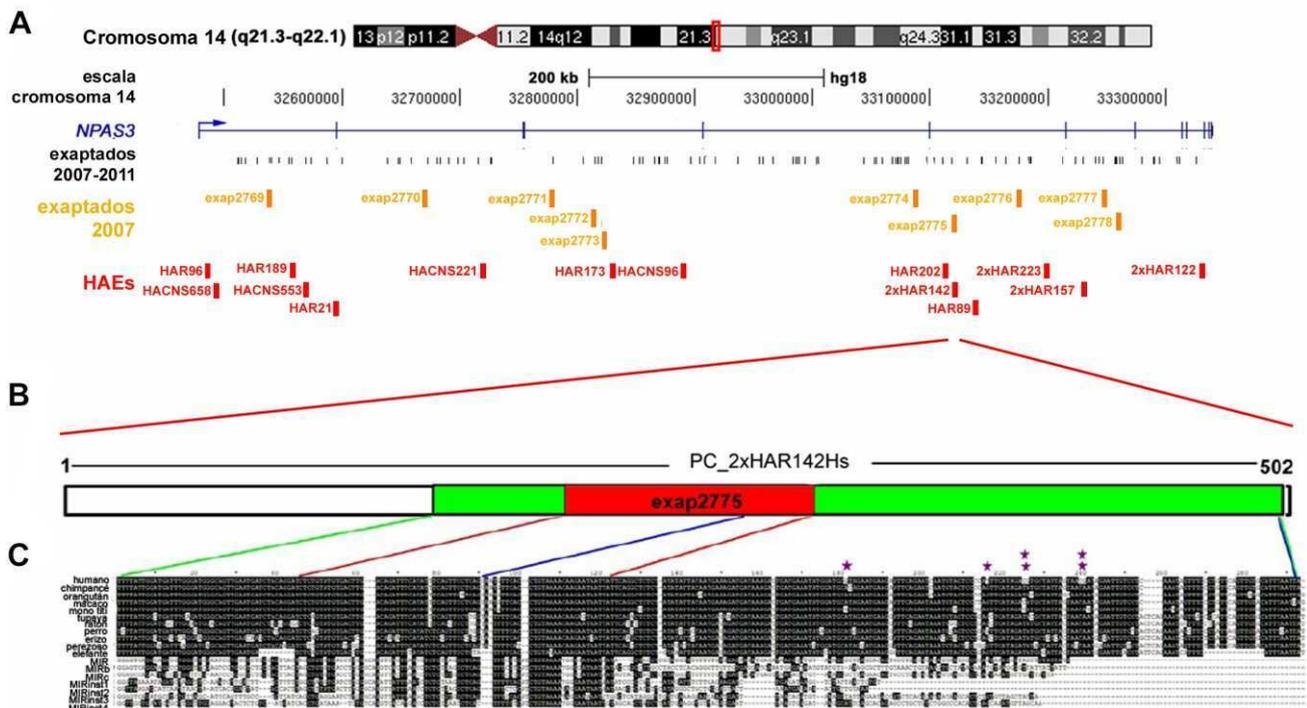


Figura 26: Localización de los elementos transponibles y HAEs en el *locus* de *NPAS3*. (A) Localización del gen *NPAS3* en el ensamblado del genoma humano versión del ensamblado NCBI36/hg18 mostrando los elementos exaptados como fueron definidos en (Lowe, Bejerano y Haussler 2007; Lindblad-Toh et al. 2011; Lowe y Haussler 2012) y HAEs en el contexto del cromosoma 14 de acuerdo al *USCS Genome Browser* (www.genome.ucsc.edu). (B) Diagrama del *enhancer* PC_2xHAR142Hs mostrando en blanco la secuencia completa de este elemento, en verde, la región altamente conservada definida por el programa PhansCons, y en rojo, la secuencia del elemento exapt2775 derivada de transposón MIR. (C) Alineamiento de las secuencias conservadas que contiene al elemento 2xHAR142Hs y regiones ortólogas en chimpancé, orangután, macaco, mono tití, tupaya, perro, erizo de tierra, perezoso, elefante y ratón. Adicionalmente, se muestran los alineamientos con las secuencias consenso de los retrotransposones MIR, MIRb, MIRc. Las regiones PhastCons, exapt2775 y 2xHAR142 (tal como fueron originalmente definidas) se muestran en el alineamiento por medio de barras verdes, rojas y azules respectivamente. Con estrellas violetas se indican las sustituciones nucleotídicas chimpancé-humano.

Recientemente, se publicó una segunda base de datos de secuencias provenientes de elementos móviles en el genoma humano con evidencia de exaptación (Lindblad-Toh et al. 2011; Lowe y Haussler 2012). Esta base de datos contiene un total de 284.687 regiones con evidencia de derivar de ET que fueron subsecuentemente sujetos a un proceso de evolución purificadora, y aquí se llamarán *exapted repeats 2011*. Usando esta segunda base de datos de ET con evidencia de exaptación, se encontró que a nivel genómico existen 133 *exapted repeats 2011* que solapan con algún HAE. Este hallazgo sugiere que, además de los 31 HAEs antes mencionados, existen HAEs

adicionales para los cuales hay evidencia de provenir de la inserción de un elemento móvil que subsecuentemente sufrió un proceso de selección purificadora (hipotéticamente por coopción para una determinada función). Por otro lado, se encontró que el *locus* de *NPAS3* contiene 138 elementos con evidencia de derivar de ET exaptados. Estos 138 elementos tienen un largo que va de entre 5 a 238 pb. Tres de estos *exapted repeats 2011* se solapan con dos *exapted repeats 2007* (los elementos *exapt2772* y *exapt2778*). Sin embargo, en este segundo caso no se encontró intersección entre la base de datos *exapted repeats 2011* y la de HAEs en el *locus* de *NPAS3*. Esto convierte al elemento 2xHAR142 en el único HAE de *NPAS3* que contiene parte de su secuencia derivada de la inserción de un elemento transponible con evidencia de coopción para una función regulatoria de acuerdo con las base de datos *exapted repeats 2007* y *exapted repeats 2011* (figura 26, A y B).

Por otro lado, con el objetivo de estudiar si *NPAS3* tiene una acumulación excepcional de secuencias provenientes de elementos transponibles en el genoma humano con evidencia de exaptación, se analizó en detalle la distribución de estos elementos por gen humano. Con este objetivo, se obtuvo una única base de datos de elementos conservados derivados de ET y exaptados no redundantes uniendo las dos bases de datos públicas de elementos móviles *exapted repeats 2007* y *exapted repeats 2011* (Lowe, Bejerano y Haussler 2007; Lindblad-Toh et al. 2011; Lowe y Haussler 2012). Esta nueva base de datos, llamada *exapted repeats 2007-2011*, se cruzó con una base de datos de unidades transcripcionales humanas, definidas como los intervalos genómicos delimitados por la versión más larga reportada por la base de datos de genes altamente curados *RefSeq* (Kamm et al. 2013b). Con este nuevo análisis se encontró que *NPAS3* contiene un total de 145 elementos conservados derivados de ET intrónicos y no redundantes, de los cuales uno solo solapa un HAE (el elemento presente en la base de datos *exapted repeats 2007*, *exapt2775*, mencionado anteriormente) (figura 18, A; figura 26, A y B). Por otro lado, se observó que si bien algunos elementos derivados de ET conservados están reportados como eventos únicos de inserción, se encuentran muy próximos a otros elementos *exapted repeats 2007-2011*. Esto sugiere que en realidad este grupo de elementos provendrían de un evento de movilización de un único ET.

Adicionalmente, se encontró que de los 19.897 genes humanos *RefSeq* no redundantes (ensamblado del genoma humano 2006; NCBI36/hb18), 10.754 contienen en su unidad transcripcional al menos un elemento derivado de un ET con evidencia de exaptación. De entre ellos, 10.707 genes tienen entre 1 y 144 elementos derivados de ET y posiblemente exaptados y solamente 47 genes tienen 145 o más elementos de este tipo (incluyendo en este último grupo, con 145 elementos, a *NPAS3*). Sorprendentemente, existen varios genes humanos que tienen 300 o más elementos derivados de ET y posiblemente exaptados en su unidad transcripcional (tabla 6). Para independizarse del largo de

las unidades transcripcionales, se analizó el agrupamiento de elementos conservados derivados de ET por Mb de genoma humano. Se encontró que las Mb que contienen más número de elementos derivados de ET y posiblemente exaptados (con más de 400 cada una) se encuentran en los *loci* que contienen a los genes *PTPRT* y *CSMD2* entre otros (tabla 6). Será interesante en un futuro estudiar la razón de esta inusual acumulación de este tipo de elementos en estos *loci*.

Tabla 6: Genes humanos y regiones genómicas de 1 Mb de longitud conteniendo la mayor acumulación de elementos exaptados.

Número de elementos exaptados	Posición del gen (NCBI36/hg18)	Símbolo del gen
410	chr20: 40,134,805-41,251,971	<i>PTPRT</i>
393	chr17: 28,364,218-29,507,938	<i>ASIC2</i>
364	chr1: 48,771,113-50,262,213	<i>AGBL4/BEND5</i>
312	chr1:57,236,166-58,488,799	<i>DAB1</i>
294	chr9:118,227,324-119,217,138	<i>ASTN2</i>
287	chr14: 67,356,248-68,132,491	<i>RAD51B</i>
282	chr11: 82,843,703-85,015,962	<i>DLG2</i>
278	chr20: 13,924,145-15,981,841	<i>MACROD2/FLRT3</i>
256	chr1: 33,752,195-34,404,030	<i>CSMD2/HMGB4</i>
255	chr11: 78,041,975-78,829,343	<i>TENM4</i>
250	chr12: 97,563,208-98,902,563	<i>APAF1 /ANKSBI</i>
Número de elementos exaptados	Posición de la Mb (NCBI36/hg18)	Genes en el intervalo
526	chr1:34,250,034-35,250,034	<i>GJB4, GJB3, GJA4, CSMD2, C1ORF94, GJB5, SMIM12, DLGAP3, LOC633160, ZMYM6NB, ZMYM6</i>
518	chr1: 34,000,034-35,000,034	<i>CSMD2, LOC402779, C1ORF94, HMGB4, GJB5, GJB4</i>
482	chr1:33,750,033-34,750,033	<i>CSMD2, HMGB4, LOC402779, C1ORF94</i>
410	chr10:99,500,099-100,500,099	<i>ZFYVE27, SFRP5, GOLGA7B, CRTAC1, R3HCC1L, LOXL4, PYROXD2, MIR1287, HPS1, MIR4685, HPSE2</i>
405	chr20:39,750,039-40,750,039	<i>PTPRT</i>
403	chr20:39,500,039-40,500,039	<i>PTPRT, CHD6, KIAA1335</i>
401	chr20:40,000,040-41,000,040	<i>PTPRT</i>
398	chr1:34,500,034-35,500,034	<i>GJB5, GJB4, GJB3, GJA4, SMIM12, DLGAP3, LOC633160, ZMYM6NB, ZMYM6, ZMYM1, SFPQ</i>
396	chr1:40,750,040-41,750,040	<i>EXO5, ZNF684, RIMS3, LOC100130557, NFYC, MIR30E, MIR30C1, KCNQ4, CITED4, CTPS1, SLFNLI-ASI, SLFNLI, SCMHI, MIR5095, EDN2, HIVEP3</i>
392	chr9:125,000,125-126,000,125	<i>STRPB, CRB2, DENND1A, MIR601, LHX2,</i>

A continuación, la Dra. Franchini analizó el tipo de elemento móvil del cual provienen los 145 elementos conservados derivados de ET hallados en el *locus* de *NPAS3* con el objetivo de caracterizarlos en mayor detalle y datar los eventos de inserción. Utilizando la base de datos de elementos transponibles *Repeat masker* (*Repeat Masker* v.3.2.7) (Smit 1996-2010), se lograron identificar 137 de los 145 elementos derivados de ET presentes en el *locus* de *NPAS3*. De acuerdo con esta base de datos, entre los 137 elementos móviles posiblemente exaptados identificados, 50 derivan de SINEs, 70 de LINEs y 10 de transposones de ADN (Kamm et al. 2013a). Estos resultados sugieren que los tipos de elementos móviles de los cuales derivan los ET con evidencia de coopción en el *locus* de *NPAS3* no difieren de la tendencia general reportada para el genoma humano completo (Lowe et al. 2003; Lowe y Haussler 2012). Los restantes elementos derivados de ET no anotados en la base *Repeat masker* fueron identificados utilizando la herramienta blast (Kamm et al. 2013a). Por otro lado, la datación de los eventos de inserción de los elementos derivados de ET y posiblemente exaptados presentes en el *locus* de *NPAS3* se realizó estudiando la conservación de sus secuencias en todos los genomas de vertebrados disponibles. Este estudio indicó que la gran mayoría de los elementos derivados de ET fueron probablemente cooptados en el linaje que dio origen a los mamíferos placentarios (85 elementos; 62,04%). Adicionalmente, 28,26% y 10,15% de los elementos probablemente hayan ganado una función en los linajes que dieron lugar a los mamíferos y a los vertebrados respectivamente (Kamm et al. 2013a).

El elemento acelerado no codificante 2xHAR142 contiene una secuencia derivada de un tranposón de la familia MIR

El próximo objetivo consistió en estudiar en más detalle el origen evolutivo del elemento con signos de evolución acelerada en el linaje humano 2xHAR142. Recurriendo a la base de datos *exapted repeats 2007* (Lowe, Bejerano y Haussler 2007), se encontró que 30 nucleótidos del elemento 2xHAR142 humano están anotados en esta base de datos como un elemento transponible posiblemente exaptado luego de su inserción (denominado exapt2775). El elemento exapt2775 descrito por Lowe *et. al.* (Lowe, Bejerano y Haussler 2007) contiene en su totalidad 73 pb que están incluidas completamente dentro del elemento conservado PhastCons de 280 pb, que a su vez, contiene la región 2xHAR142 (figura 18, A; figura 26, B).

Análisis bioinformáticos realizados por la Dra. Franchini demostraron que el elemento exapt2775 es parte remanente de un retrotransposón tipo SINE, probablemente de la familia de los *mammalian-wide interspread repeats* (MIRs, por las siglas en inglés). Adicionalmente, se encontró que el elemento exapt2775 está conservado entre los mamíferos placentarios estudiados pero no está

presente en los genomas de los mamíferos no placentarios *Monodelphis domestica* y *Ornitorhynchus anatinus* u otros vertebrados (figura 18, B) sugiriendo que la inserción de este elemento ocurrió en el linaje que dio lugar a los primeros (Kamm et al. 2013a).

Los retrotransposones de la familia MIR tienen una extensión de aproximadamente 250 pb (Smit y Riggs 1995). Sin embargo, solo una pequeña porción del MIR original (73pb) es reconocible como derivada de un retrotransposón en la secuencia conservada que contiene al elemento 2xHAR142 (figura 26, B). Estudios adicionales realizados por la Dra. Franchini alineando las secuencias consenso completas de los transposones MIR (familias MIR, MIRb y MIRc) con las secuencias ortólogas del elemento PC_2xHAR142, permitieron identificar como derivadas del retrotransposon original regiones adicionales de este elemento (figura 26, C). Estos hallazgos sugieren que un transposón completo de la familia MIR se insertó en esta región en el linaje que dio origen a los mamíferos placentarios y que posteriormente acumuló mutaciones antes de comenzar a evolucionar por selección purificadora (Kamm et al. 2013a). Si bien hacen falta estudios funcionales adicionales para evaluar el impacto de las regiones derivadas del transposon MIR en el *enhancer* codificado por el elemento PC_2xHAR142Hs, los resultados mostrados hasta aquí sugieren que secuencias derivadas de un transposon MIR pudieron haber sido exaptadas en una región regulatoria en *cis* (posiblemente como un *enhancer* de *NPAS3*) y mantenidas bajo selección purificadora todo a lo largo del linaje de los mamíferos placentarios.

Capítulo III

Estudio funcional del gen *NPAS3* en pez cebra

INTRODUCCIÓN ESPECÍFICA

Morfolinos

Una forma rápida de estudiar el rol de un gen particular durante el desarrollo embrionario consiste en reducir los niveles de expresión de su ARNm. Un método particularmente eficiente y sencillo utiliza morfolinos antisentido para inhibir o reducir los niveles de expresión de un determinado mensajero en peces cebra (Nasevicius y Ekker 2000). Los morfolinos son derivados sintéticos del ADN con dos modificaciones principales: 1) el anillo de desoxirribosa de 5 átomos de carbono es reemplazado por un anillo morfolino con 4 átomos de carbono, uno de nitrógeno y uno de oxígeno; 2) el enlace aniónico fosfodiéster es reemplazado por una unión no iónica fosforodiamidato (Bedell, Westcot y Ekker 2011) (figura 27). Los morfolinos consisten en oligonucléotidos de aproximadamente 25 bases que actúan vía complementariedad de bases con el ARNm blanco. En pez cebra se utilizan dos tipos principales de morfolinos (MOs): los morfolinos que bloquean la traducción y los morfolinos que bloquean el procesamiento del mensajero (particularmente el *splicing*). El primer tipo de morfolino (MOs traduccional) se une por complementariedad de bases con la región 5' no traducida (5'UTR por sus siglas en inglés) del mensajero blanco impidiendo el ensamblado del complejo ribosomal, y por lo tanto, previniendo la traducción. Por otro lado, el mecanismo de acción de los morfolinos de procesamiento (MOs de *splicing*) consiste en bloquear la maduración del pre-ARNm inhibiendo la formación de los complejos del spliceosoma (Bill et al. 2009).

Los diferentes tipos de MOs tienen usos complementarios a la hora de estudiar la función de los genes en pez cebra. Por un lado, los MOs de *splicing* solo inhiben los mensajeros sintetizados *de novo* por el embrión. Mientras que, los MOs de traducción son capaces de inhibir tanto los mensajeros cigóticos como los maternos, cubriendo de esta forma, fenotipos tal vez enmascarados por los mensajeros maternos (Bill et al. 2009). En general, los fenotipos producidos por los MOs se analizan durante los tres primeros días del desarrollo embrionario del pez cebra. Sin embargo, muchos estudios reportan que los efectos de los MOs pueden perdurar hasta los 5 días (Kimmel et al. 2003; Smart, De Rose y Farber 2004).

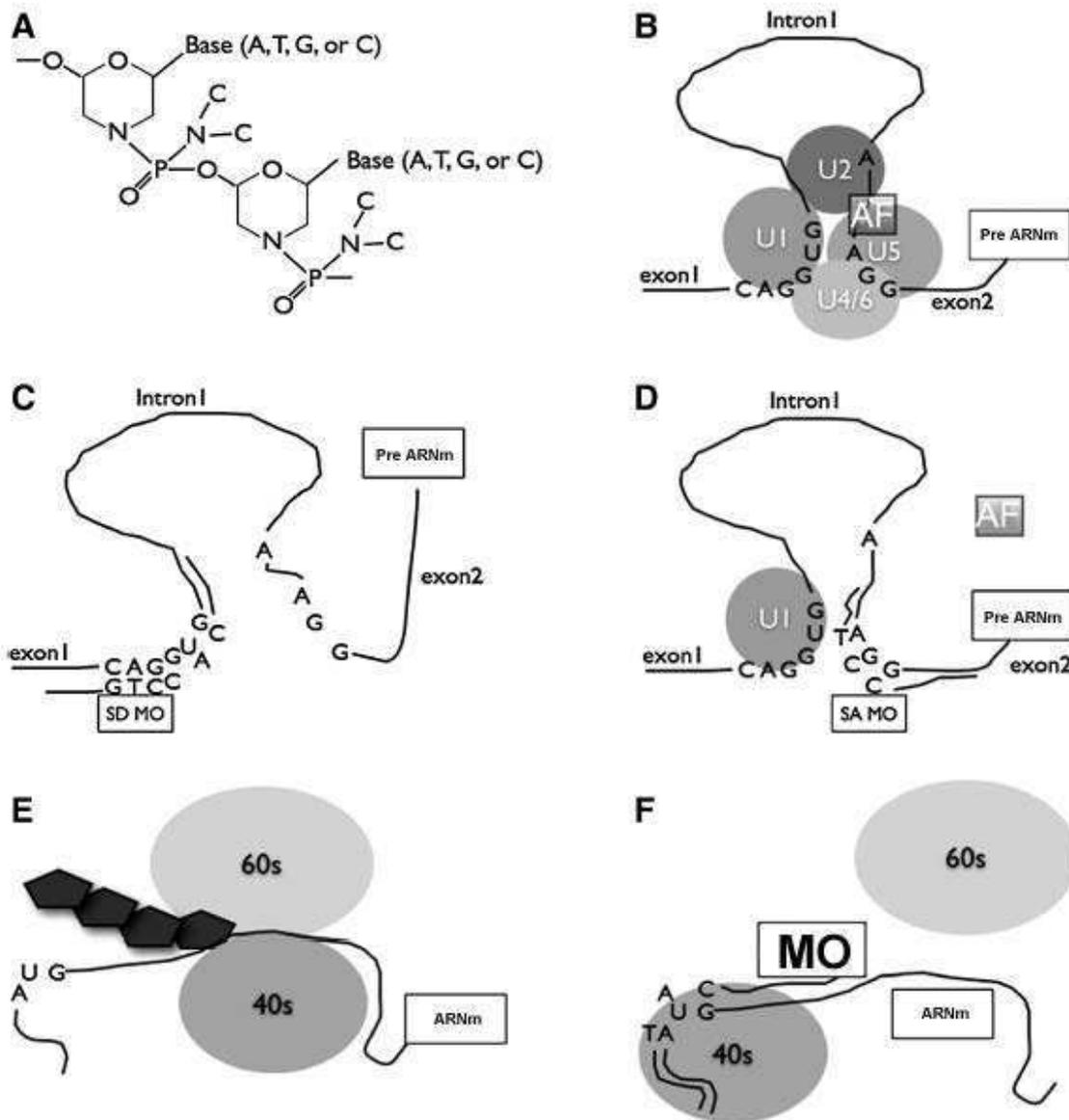


Figura 27: Oligonucleótidos antisentido de morfolinis y sus mecanismos de acción. (A) Los MOs están compuestos por unidades de anillos morfolino unidos por enlaces fosforodiamidato y las mismas bases que tiene el ADN. El mecanismo de acción de los MOs consiste en: (B) impedir el splicing normal o (E) bloquear la traducción. Los MOs diseñados contra el sitio donador (SD) de splicing inhiben la unión del complejo U1, impidiendo la formación del lazo y finalmente promoviendo la incorporación del intrón (C). La incorporación de un intrón frecuentemente produce la finalización prematura de la traducción y la degradación del mensajero por el sistema de *nonsense-mediated decay*. Los MOs diseñados contra el sitio aceptor (SA) funcionan teóricamente previniendo la unión de la proteína U2AF, necesaria para reclutar el complejo de splicing U2 (D). Los MOs traducciónales se unen al sitio ATG o a las secuencias 5'UTR de los mensajeros impidiendo de esta forma el escaneo de la subunidad ribosomal 40S y/o formación del complejo ribosomal completo (F). Figura modificada de (Bill et al. 2009).

HIPÓTESIS DE TRABAJO

Dado que el gen *NPAS3* juega un papel clave en el desarrollo y funcionamiento del sistema nervioso en mamíferos, la hipótesis de trabajo de este capítulo es que la expresión de éste gen también será determinante en el desarrollo del SNC del pez cebra.

OBJETIVO ESPECÍFICO

- ✓ Estudiar la función de *NPAS3* en pez cebra. Para ello se realizarán experimentos de genética inversa utilizando una serie de morfolidos dirigidos contra el mensajero del gen

RESULTADOS

Análisis del patrón de expresión de *NPAS3* en larvas de pez cebra

En este capítulo se estudiará la función de *NPAS3* en pez cebra utilizando morfolinós para reducir los niveles de esta proteína. Como algunos de los fenotipos podrían aparecer a edades postembrionales del desarrollo del pez cebra, más allá de las 72 hpf (Parichy et al. 2009), se decidió extender el análisis de expresión de *NPAS3* realizado en el capítulo II. Con este objetivo, se utilizaron dos sondas diseñadas para detectar el ARNm de *NPAS3* y se realizaron ensayos de ISH en larvas completas y cortes de crióstato a 3 y 5 días post fecundación (dpf). Estos ensayos mostraron que a 3 dpf *NPAS3* se expresa abundantemente en techo de la boca y los arcos faríngeos. También se observó expresión en el corazón y las aletas pectorales. A nivel del sistema nervioso se observaron altos niveles de expresión en el cerebro y la médula espinal. Se detectó expresión en el telencéfalo, diferentes estructuras diencefálicas incluidas el tálamo y el hipotálamo y también en el techo óptico. En el rombencéfalo, se detectó expresión en la médula oblonga (figura 28, A).

Dos días después, a 5 dpf, el patrón de expresión de *NPAS3* se mantiene estable. Se detectaron altos niveles de expresión en el cerebro y la médula espinal. Por fuera del sistema nervioso, la expresión es particularmente fuerte a nivel de los arcos faríngeos. También se mantiene la expresión de *NPAS3* en las aletas pectorales y el corazón. Además, se detectó expresión en la vejiga natatoria (figura 28, B). Las dos sondas utilizadas revelaron patrones de expresión similares, mientras que una sonda inespecífica fue incapaz de producir un precipitado coloreado (datos no mostrados).

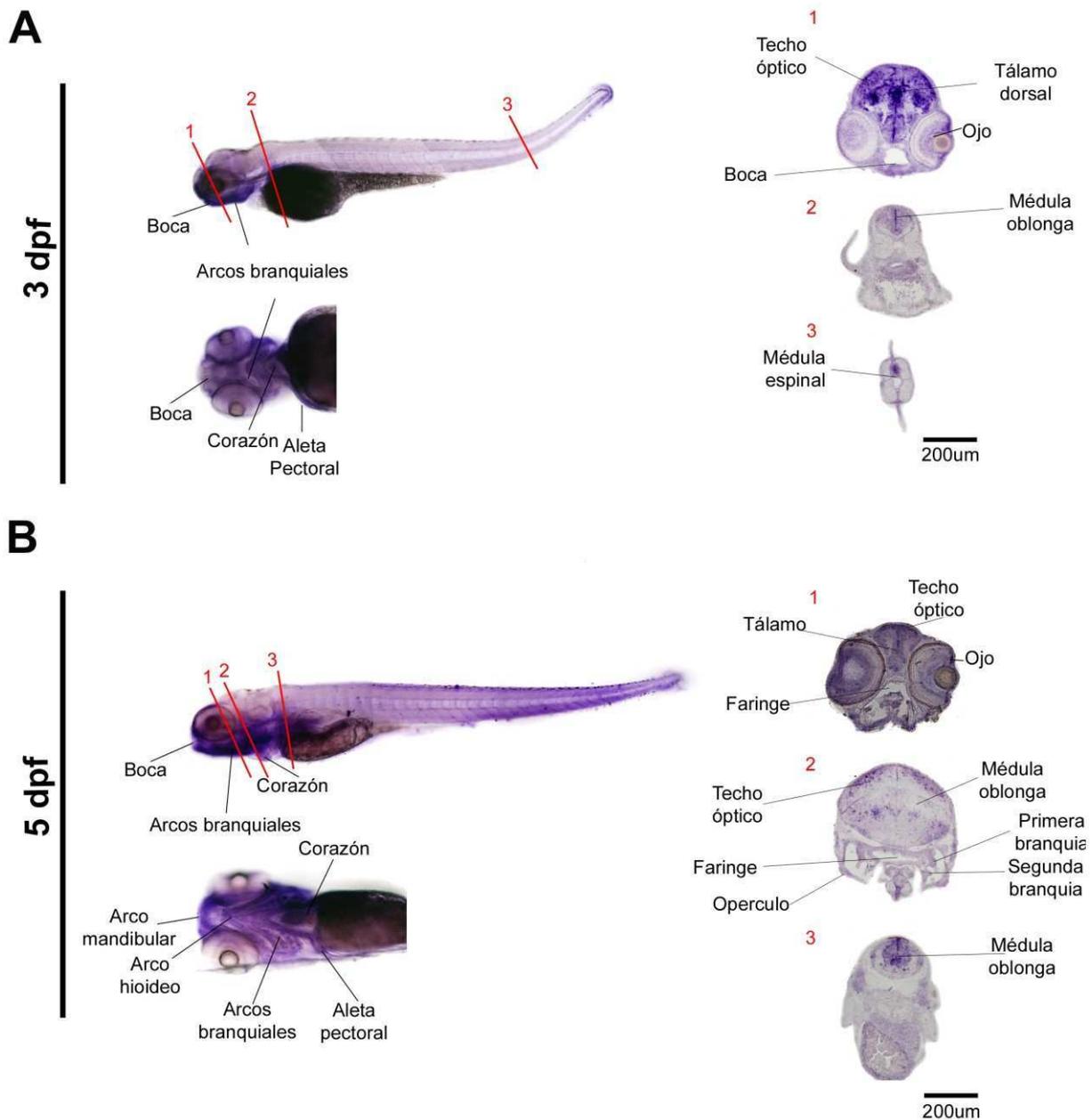


Figura 28: Expresión del gen *NPAS3* durante el desarrollo del pez cebra a 3 y 5 dpf. (A) Patrón de expresión de *NPAS3* a 3 dpf revelado por hibridación *in situ* sobre (izquierda) larvas completas y (derecha) cortes transversales de crióstato. (B) Patrón de expresión de *NPAS3* a 5 dpf revelado por hibridación *in situ* sobre (izquierda) larvas completas y (derecha) cortes transversales de crióstato.

Fenotipo de larvas de pez cebra tratadas con morfolinós contra *NPAS3*

Se decidió estudiar el efecto de los morfolinós diseñado contra el ARNm de *NPAS3* a las edades postembrionales 3 y 5 dpf (luego de ocurrida la eclosión del animal del corion y momento donde comienza la protrusión de la boca anteriormente a los ojos)(Kimmel et al. 1995; Parichy et al. 2009). Esta etapa del desarrollo es muy importante ya que en ella tienen lugar cambios claves en una gran variedad de estructuras. Por ejemplo, en el pez cebra, el desarrollo postembrional

comienza aproximadamente a las 72 hpf y abarca la aparición de la aleta pélvica y la estratificación de la piel. Además, en esta etapa continúa el desarrollo del esqueleto, tiene lugar cambios marcados en el sistema digestivo, los riñones y las gónadas, así como también en el sistema nervioso central y periférico (Parichy et al. 2009). Además, *NPAS3* se expresa abundantemente en múltiples tejidos en las larvas de 3 y 5 dpf (figura 28).

Para estudiar si este gen podría tener un rol conservado evolutivamente en los vertebrados, se realizaron ensayos de reducción de los niveles de expresión de *NPAS3* por medio de oligonucleótidos antisentido específicos para el ARNm de este gen en pez cebra. Se utilizaron dos morfolinós específicamente diseñados para este estudio. El primero es un morfolino traduccional (MO1) diseñado para unirse al codón de iniciación de la traducción del ARNm de *NPAS3* (figura 29, A). Un segundo morfolino de bloqueo del procesamiento del splicing (MO2) fue dirigido contra el sitio de splicing entre el exón 3 y el intrón 4 (figura 29, A). Por otro lado, no se encontraron diferencias significativas entre el porcentaje de supervivencia a 24 hpf de los embriones tratados con 0,75 μ M de MO1, 0,75 μ M de MO2, inyectados en las mismas condiciones pero donde se reemplazó la solución con morfolino por buffer (S.MO) o embriones sin inyectar (S.I.) (fracción promedio de supervivencia de los embriones a 24 hpf \pm desvío estándar: MO1=0,880 \pm 0,073, n=3; MO2=0,792 \pm 0,183, n=3; S.MO=0,942 \pm 0,038, n=3; S.I.=0,923 \pm 0,0245, n=3; P>0,05 para todas las comparaciones; Kruskal-Wallis seguido de contrastes de Dunn's).

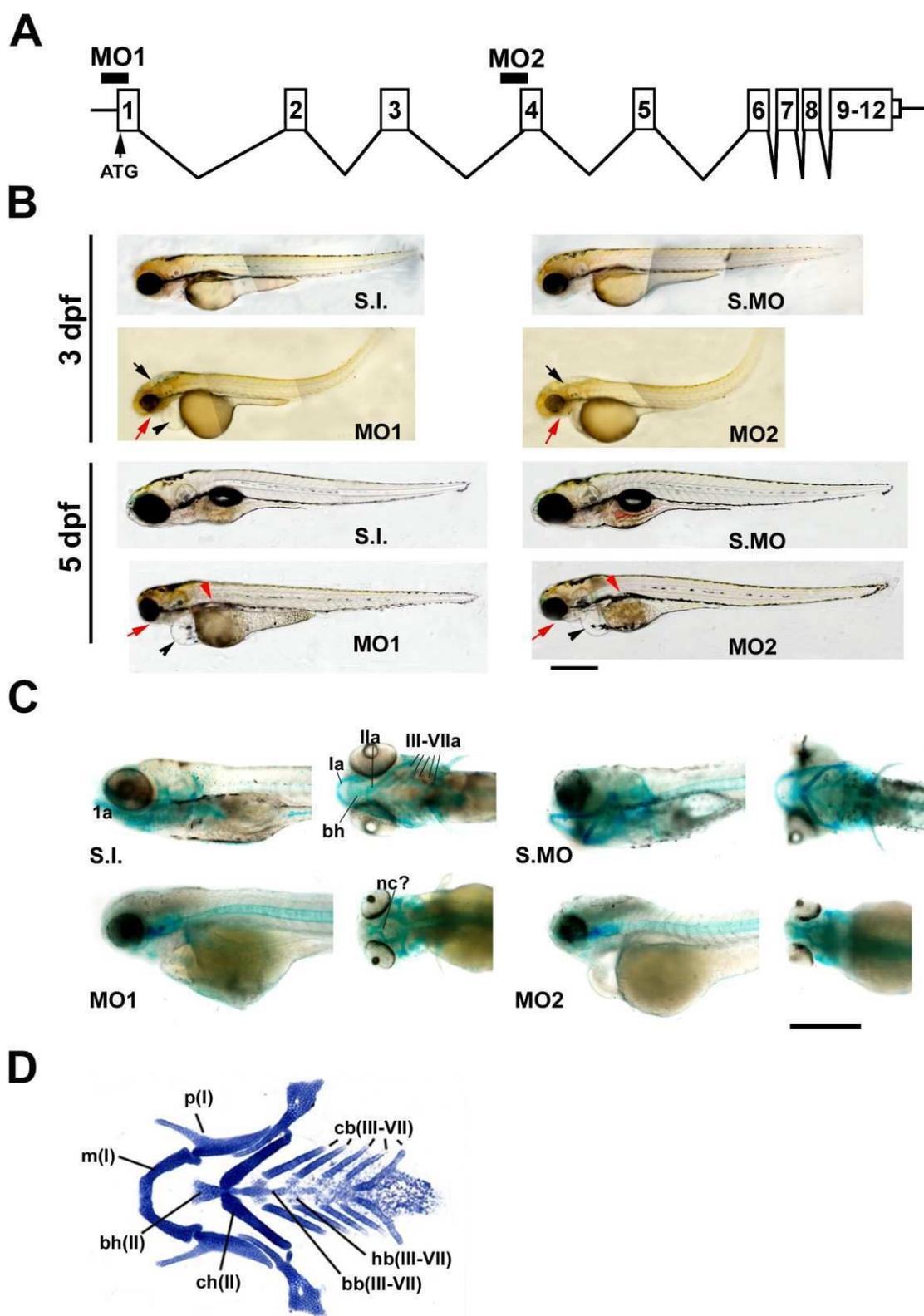


Figura 29: Defectos en las larvas de pez cebra tratadas con morfolinios contra *NPAS3*. (A) Esquema del ARNm de *NPAS3* en pez cebra de acuerdo con la base de datos de Ensembl (www.ensembl.org). En el esquema se indican los sitios predichos de reconocimiento de los morfolinios contra *NPAS3*: MO1 (morfolino traduccional) y MO2 (morfolino de bloqueo del *splicing* del intrón 3). (B) Peces cebra

morfantes tratados con 0,75 μ M de morfolino MO1 o MO2 y controles a (arriba) 3 dpf y (abajo) 5 dpf. Las flechas negras indican la presencia de hidrocefalia a nivel del ventrículo cerebral. Las flechas rojas indican la presencia de defectos y atrofas a nivel de las estructuras craneofaciales. Con cabezas de flecha negras se muestra el desarrollo de edema de pericardio (C) Peces cebra morfantes y controles tratados con el colorante alcian blue a 5 dpf en vista lateral y vista ventral. (D) Esquema en vista ventral de los elementos cartilagosos que forman los arcos branquiales en peces cebra de 5 dpf. Los números entre paréntesis indican los arcos branquiales de donde deriva cada elemento. Esquema sección (D) modificado de (Laue et al. 2008). Referencias: S.I., larvas no inyectadas; S.MO., larvas inyectadas con buffer pero sin morfolino; MO1, larvas inyectadas con MO1; MO2, larvas tratadas con el MO2; Ia, primer arco faríngeo (arco mandibular); IIa, segundo arco faríngeo (arco hioideo); IIIa-VIIa, tercer a séptimo arco faríngeo (primer a quinto arco branquial); nc, neurocráneo; ap, aleta pectoral; cb, cartílago ceratobranquial; hb, cartílago hipobranquial; m, cartílago de Merkel; bb, cartílago basibranquial; p, palatoc cuadrado; ch, cartílago ceratohial; bh, cartílago basihial.. Escala 500 μ m.

Los resultados obtenidos de tres días de inyección independientes mostraron que, las larvas morfantes de 3 dpf presentaron edema del pericardio (50% y 33% de los animales inyectados con los morfolinios MO1 y MO2, respectivamente) (figura 29, B; figura 30, A). Por otro lado, un 91% y 63% de los peces inyectados con los morfolinios MO1 y MO2 respectivamente presentaron anomalías en el desarrollo de estructuras craneofaciales incluyendo deformidades severas o ausencia de elementos de la mandíbula (figura 29 B; figura 30, B). Además, el 75% y 60% de los peces inyectados con los morfolinios MO1 y MO2 respectivamente desarrolló hidrocefalia a esta edad (figura 29, B; figura 30, C). Salvo en menos de un 3% de animales tratados con buffer que desarrolló edema de pericardio, en las larvas que fueron inyectadas con buffer de inyección (sin morfolino: S.MO.) y en los animales no inyectados (sin inyectar: S.I.) mantenidos en las mismas condiciones de cría, no se encontraron ninguna de las alteraciones morfológicas descritas anteriormente (figura 29, B; figura 30, A, B, C). A 5 dpf se encontró que un 50% de las larvas presentó edema de pericardio con ambos morfolinios, mientras que el 100% de las larvas control no desarrolló esta condición (figura 29, B; figura 30, A). Además, 75% y 76% de los animales tratados con los MO1 y MO2, respectivamente, desarrolló alguna deformidad craneofacial. Solo se encontró deformidad en elementos mandibulares en un 2,7% de las larvas no inyectadas y estuvo ausente en el 100% de los animales inyectados con buffer (figura 29, B; figura 30, B). Por otro lado, se observó un marcado déficit en el inflado de la vejiga natatoria en las larvas inyectadas con morfolino. Mientras que un 100% de los animales inyectados son buffer o no inyectados presentaban la vejiga natatoria inflada a 5 dpf (se observa como una burbuja con brillo metálico en las microscopías de

campo claro), solo un 6,25% y 12% de los animales tratados con los MO1 y MO2 respectivamente presentaba la vejiga natatoria inflada a dicha edad (figura 29, B; figura 20, D).

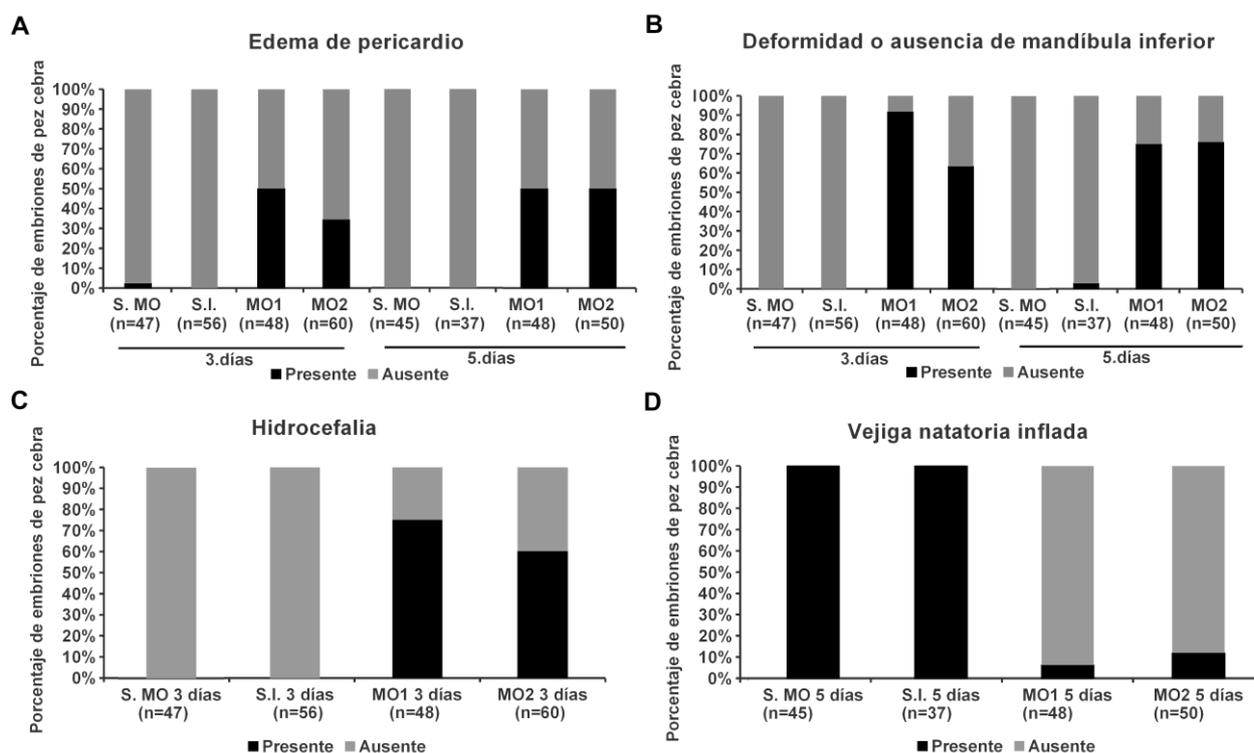


Figura 30: Gráfico del porcentaje de larvas estudiadas presentando diferentes defectos (A-C) y características fenotípicas (D). (A) Porcentaje de larvas tratadas con morfolino o larvas control que presentaron edema de pericardio a 3 y 5 dpf. (B) Porcentaje de larvas tratadas con morfolino o larvas control que presentaron ausencia/deformidad de elementos craneofaciales a 3 y 5 dpf. (C) Porcentaje de larvas tratadas con morfolino o larvas control que presentaron hidrocefalia a nivel del ventrículo cerebral a 3 dpf. (D) Porcentaje de larvas tratadas con morfolino o larvas control que mostraron la vejiga natatoria inflada a 5 dpf. Referencias: S.I., larvas no inyectadas; S.MO., larvas inyectadas con buffer pero sin morfolino; MO1, larvas inyectadas con MO1; MO2, larvas tratadas con el MO2; n, número de animales analizados provenientes de tres días independientes de inyección.

Con el objetivo de estudiar con más detalle las anomalías en el desarrollo craneofacial en los animales tratados con los morfolidos contra *NPAS3*, se realizó una tinción con el colorante alcian blue. Este colorante tiene afinidad por la matriz extracelular asociada con los condrocitos y es muy útil para visualizar la mayor parte de los elementos cartilaginosos de la mandíbula y arcos faríngeos en pez cebra (Bird y Mabee 2003; Solomon et al. 2003). A 5 dpf, se observó que las larvas inyectadas con ambos morfolidos presentaban diversas ausencias y atrofas en los cartílagos craneofaciales. Específicamente, en estos animales no se forman los cartílagos derivados del primer (arco mandibular) ni del segundo (arco hioideo) arco faríngeo (figura 29, C). Además, hay una

ausencia completa de los cartílagos asociados con el resto de los arcos faríngeos (del III al VII) (figura 29, C). Las estructuras cartilaginosas del neurocráneo resultan menos afectadas por el tratamiento con morfolidos contra *NPAS3*. Estos cartílagos llegan a formarse, aunque con alteraciones en su morfología (figura 29, C).

Ha sido reportado que ratones deficientes en la proteína *Npas3*, son más pequeños (pesando un 20% menos) que los animales de fenotipo salvaje (Brunskill et al. 2005). Para estudiar si los peces cebras tratados con morfolidos diseñados contra el ARNm de *NPAS3* presentan un fenotipo similar, se tomaron diferentes mediciones del cuerpo de estos animales. Por un lado, a 3 dpf las larvas inyectadas con morfolidos son aproximadamente un 10% más pequeñas en largo que los animales inyectados con buffer y esta diferencia es significativa (largo promedio del cuerpo medido desde la cabeza a la base de la aleta caudal en $\text{mm} \pm$ desvío estándar: S.MO= $3,30 \pm 0,13$, $n = 37$; MO1= $2,94 \pm 0,24$, $n=33$, $P < 0,05$; MO2= $2,89 \pm 0,22$, $n=44$, $P < 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, A). No se encontraron diferencias significativas en el largo del cuerpo a esta edad entre los animales inyectados con buffer y los animales sin inyectar (largo promedio del cuerpo medido desde la cabeza a la base de la aleta caudal en $\text{mm} \pm$ desvío estándar: S.MO= $3,30 \pm 0,13$, $n = 37$; S.I.= $3,17 \pm 0,17$, $n=50$, $P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, A). Tampoco difirieron significativamente los largos del cuerpo entre el grupo de larvas tratadas con el MO1 o con el MO2 (largo promedio del cuerpo medido desde la cabeza a la base de la aleta caudal en $\text{mm} \pm$ desvío estándar: MO1= $2,94 \pm 0,24$, $n=33$; MO2= $2,89 \pm 0,22$, $n=44$, $P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's). Por otro lado, a 5 dpf sólo se encontraron diferencias significativas en el largo del cuerpo entre los animales inyectados con buffer y los animales tratados con el MO1 (largo promedio del cuerpo medido desde la cabeza a la base de la aleta caudal en $\text{mm} \pm$ desvío estándar: S.MO= $3,67 \pm 0,13$, $n = 50$; S.I.= $3,66 \pm 0,12$, $n=44$, $P > 0,05$; MO1= $3,51 \pm 0,27$, $n=40$, $P < 0,05$; MO2= $3,62 \pm 0,24$, $n=56$, $P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, A).

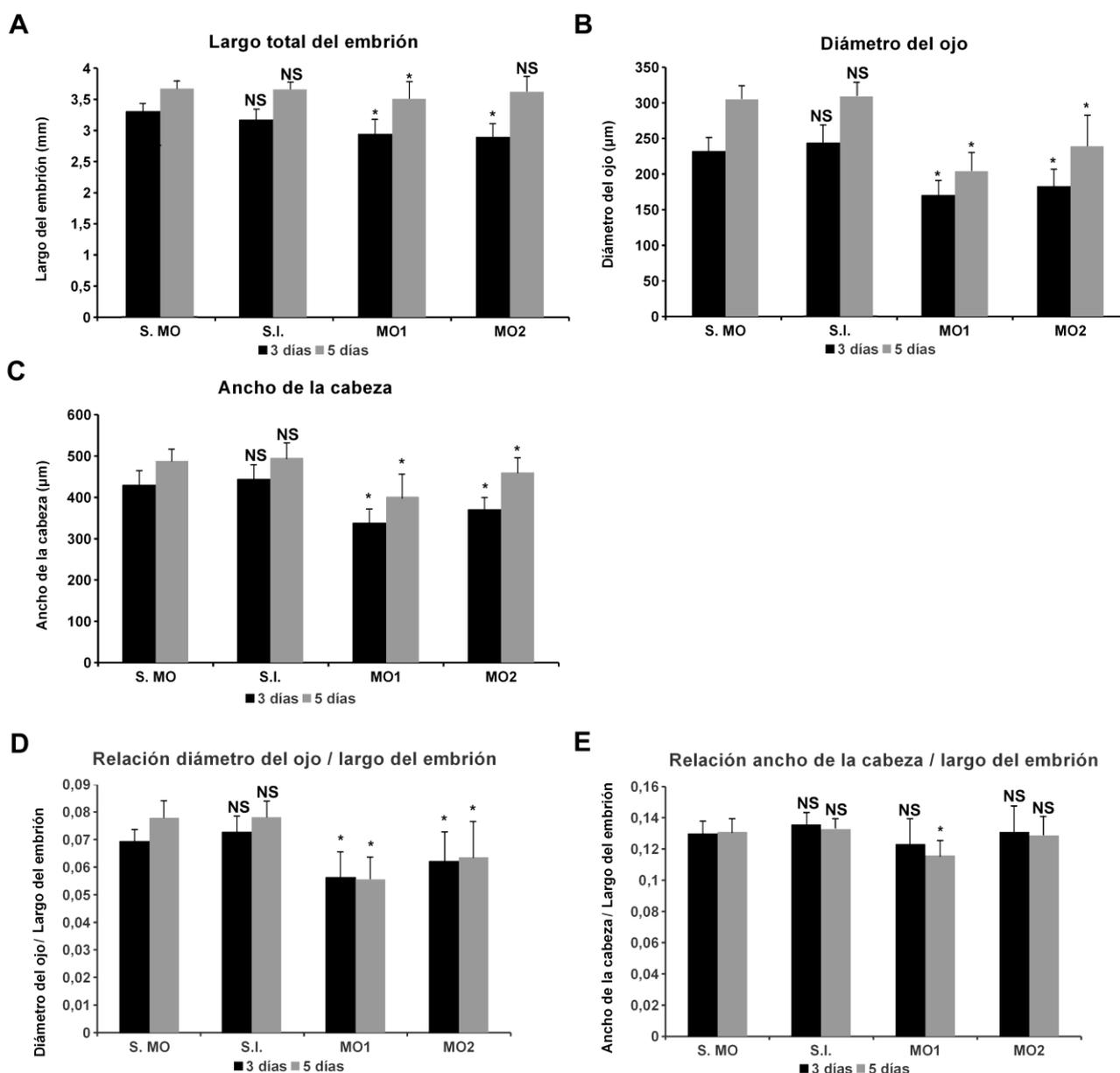


Figura 31: (A) Gráfico del largo del cuerpo promedio en mm de larvas inyectadas con morfolino o controles a 3 y 5 dpf. (B) Gráfico del diámetro del ojo promedio en µm de larvas inyectadas con morfolino o controles a 3 y 5 dpf. (C) Gráfico de la longitud promedio de la cabeza µm de larvas inyectadas con morfolino o controles a 3 y 5 dpf. (D) Gráfico del diámetro del ojo estandarizado por el largo del cuerpo de larvas inyectadas con morfolino o controles a 3 y 5 dpf. (E) Gráfico de la longitud de la cabeza estandarizada por el largo del cuerpo de larvas inyectadas con morfolino o controles a 3 y 5 dpf. Los valores expresan las medias ± desviación estándar. Kruskal-Wallis seguido de contrastes de Dunn's. Referencias: NS, no significativo respecto del tratamiento S.MO., $P > 0,05$; Con asterisco se indican las comparaciones con el tratamiento S.MO. con $P < 0,05$; S.I., larvas no inyectadas; S.MO., larvas inyectadas con buffer pero sin morfolino; MO1, larvas inyectadas con MO1; MO2, larvas tratadas con el MO2.

A 3 y 5 dpf el diámetro del ojo de las larvas (medido perpendicularmente a la superficie de la boca) inyectadas con morfolino fue aproximadamente un 20% menor que el diámetro del ojo en los animales inyectados con buffer y esta diferencia fue significativa para ambas edades (diámetro promedio del ojo medido perpendicularmente a la superficie de la boca en $\mu\text{m} \pm$ desvío estándar; a 3 dpf: S.MO.= $231,7 \pm 19,78$, $n=44$; MO1= $169,94 \pm 20,34$, $n=40$ $P < 0,05$; MO2= $182,56 \pm 23,5$, $n=58$, $P < 0,05$; a 5 dpf: S.MO.= $304,92 \pm 19,31$, $n=34$; MO1= $204,13 \pm 19,31$, $n=34$, $P < 0,05$; MO2= $239,00 \pm 43,45$, $n=53$, $P < 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, B). No se encontraron diferencias significativas en el diámetro del ojo entre los animales inyectados con buffer y los animales no inyectados a 3 y 5 dpf (diámetro promedio del ojo medido perpendicularmente a la superficie de la boca en $\mu\text{m} \pm$ desvío estándar; a 3 dpf: S.MO.= $231,7 \pm 19,78$, $n=44$; S.I.= $243,18 \pm 24,94$, $n=44$ $P > 0,05$; a 5 dpf: S.MO.= $304,92 \pm 19,31$, $n=34$; S.I.= $308,93 \pm 19,04$, $n=29$, $P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's). Tampoco difirieron significativamente los grupos de larvas inyectadas con MO1 o con MO2 respecto del diámetro del ojo ni a 3 ni a 5 dpf (diámetro promedio del ojo medido perpendicularmente a la superficie de la boca en $\mu\text{m} \pm$ desvío estándar; a 3 dpf: MO1= $169,94 \pm 20,34$, $n=40$; MO2= $182,56 \pm 23,5$, $n=58$, $P > 0,05$; a 5 dpf: MO1= $204,13 \pm 19,31$, $n=34$; MO2= $239,00 \pm 43,45$, $n=53$, $P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's).

Por otro lado, también se encontró que a 3 y 5 dpf, la longitud de la cabeza (medida como la longitud máxima desde la base de la boca al techo de la cabeza entre el ojo y la vesícula ótica) fue aproximadamente un 20% menor en las larvas inyectadas con morfolino respecto de las larvas tratadas con buffer (longitud promedio de la cabeza medida como distancia máxima desde la base de la boca al techo de la cabeza entre la vesícula ótica y el ojo en $\mu\text{m} \pm$ desvío estándar; a 3 dpf: S.MO.= $429,36 \pm 34,32$, $n=44$; MO1= $337,81 \pm 34,63$, $n=40$, $P < 0,05$; MO2= $370,25 \pm 28,27$, $n=58$, $P < 0,05$; a 5 dpf: S.MO.= $488,42 \pm 27,45$, $n=37$, MO1= $401,86 \pm 54,89$, $n=48$, $P < 0,05$; MO2= $460,60 \pm 34,16$, $n=52$, $P < 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, C). No se encontraron diferencias significativas en la longitud de la cabeza entre los animales inyectados con buffer y los animales no inyectados a 3 y 5 dpf (longitud promedio de la cabeza medida como distancia máxima desde la base de la boca al techo de la cabeza entre la vesícula ótica y el ojo en $\mu\text{m} \pm$ desvío estándar: S.MO.= $429,36 \pm 34,32$, $n=44$; S.I.= $443,43 \pm 36,44$, $n=45$, $P > 0,05$; a 5 dpf: S.MO.= $488,42 \pm 27,45$, $n=37$, S.I.= $495,43 \pm 37,25$, $n=29$, $P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, C). Tampoco difirieron significativamente los grupos de larvas inyectadas con MO1 o con MO2 respecto del diámetro del ojo a 3 dpf (longitud promedio de la cabeza medida como distancia máxima desde la base de la boca al techo de la cabeza entre la vesícula ótica y el ojo en $\mu\text{m} \pm$ desvío estándar; a 3 dpf: MO1= $337,81 \pm 34,63$, $n=40$;

MO2=370,25±28,27, n=58, P>0,05; Kruskal-Wallis seguido de contrastes de Dunn's). Sin embargo, si se encontraron diferencias significativas entre morfolinios a 5 dpf (longitud promedio de la cabeza medida como distancia máxima desde la base de la boca al techo de la cabeza entre la vesícula ótica y el ojo en $\mu\text{m} \pm$ desvío estándar; a 5 dpf: MO1=401,86±54,89, n=48; MO2=460,60±34,16, n=52, P<0,05; Kruskal-Wallis seguido de contrastes de Dunn's).

Dado se que encontraron diferencias significativas en el largo total del cuerpo de las larvas inyectadas con morfolino, especialmente a 3 dpf (figura 31, A), se relativizaron las medidas del diámetro del ojo y longitud de la cabeza al largo total del animal. Se encontró que la relación diámetro del ojo/largo total de embrión es menor en los animales inyectados con morfolino respecto de los animales tratados con buffer tanto 3 como a 5 dpf (relación diámetro del ojo medido perpendicularmente a la superficie de la boca frente a largo del cuerpo medido desde la cabeza a la base de la aleta caudal promedio \pm desvío estándar; a 3 dpf: S.MO= 0,06933674±0,00411859, n=38; MO1= 0,05626571±0,00945678, n=28; P<0,05; MO2= 0,06208432± 0,01080937, n=35, P<0,05; a 5 dpf: S.MO= 0,07787234± 0,00633251, n=31; MO1= 0,05555062±0,00814516, n=28; P<0,05; MO2= 0,06355353±0,01296877, n=35, P<0,05; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, D). Estos resultados indicarían que la reducción en el diámetro de los ojos no es proporcional a la reducción sufrida por el cuerpo completo de la larva. No se encontraron diferencias significativas la relación diámetro del ojo/largo total de embrión entre los animales inyectados con buffer y los animales no inyectados a 3 y 5 dpf (relación diámetro del ojo medido perpendicularmente a la superficie de la boca frente a largo del cuerpo medido desde la cabeza a la base de la aleta caudal promedio \pm desvío estándar: a 3 dpf, S.MO= 0,06933674±0,00411859, n=38; S.I.= 0,07270926± 0,005877, n=40, P>0,05; a 5 dpf S.MO= 0,07787234± 0,00633251, n=31; S.I.= 0,07809707±0,00594651, n=35, P>0,05; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, D).

Por otro lado, cuando la longitud de la cabeza se relativizó con el largo del cuerpo de la larva, se encontró que esta relación sólo difiere significativamente a 5 dpf cuando se compara los animales tratados con MO1 frente a los animales inyectados con buffer (relación longitud de la cabeza medida como distancia máxima desde la base de la boca al techo de la cabeza entre la vesícula ótica y el ojo frente a largo del cuerpo medido desde la cabeza a la base de la aleta caudal promedio \pm desvío estándar: a 3 dpf S.MO= 0,12994257± 0,0079290, n=38; S.I.=0,13556937±0,0073194, n=40, P>0,05; MO1=0,12307005±0,0162792, n=28; P>0,05; MO2=0,13092635±0,0163260, n=35, P>0,05; a 5 dpf: S.MO= 0,1308406± 0,0082897, n=31; S.I.= 0,13274366±0,0066470, n=30, P>0,05; MO1=0,11588376± 0,0093361, n=28; P<0,05; MO2= 0,1287245± 0,0119848, n=35,

$P > 0,05$; Kruskal-Wallis seguido de contrastes de Dunn's) (figura 31, E). Estos resultados indicarían que la reducción en la longitud de la cabeza de los animales tratados con morfolino sería proporcional a la reducción sufrida por el cuerpo de estos animales (con excepción de las larvas tratadas con MO1 a 5 dpf, donde se observa que la reducción no fue proporcional a la longitud del cuerpo). Todos estos resultados indicarían que la reducción de los niveles de expresión de *NPAS3* en pez cebra podría causar defectos en el crecimiento con ciertos puntos de similitud con los encontrados en los ratones *knock out* para *Npas3* (Erbel-Sieler et al. 2004; Brunskill et al. 2005; Pieper et al. 2005; Zhou et al. 2009)

Por otro lado, los peces cebra tratados con morfolino mostraron diversos defectos en el nado a 5 dpf. Las larvas inyectadas con el MO1 permanecieron de lado e inmóviles. Además, fueron incapaces de huir frente a un estímulo aversivo (estimulación con un pincel). En estos animales se observó la ausencia de las aletas pectorales en movimiento (que se observan claramente en los animales tratados con buffer) (videos adjuntos con el material multimedia).

Finalmente, los fenotipos producidos por los morfolinios MO1 y MO2 a 3 y 5 dpf fueron similares, siendo esto evidencia de especificidad de estos oligonucleótidos por *NPAS3* (figura 29, B y C; figura 30, figura 31). Ambos morfolinios produjeron efecto en las variables fenotípicas analizadas a la menor concentración utilizada ($0,75 \mu\text{M}$ en el embrión). En el futuro se espera poder complementar estos resultados con controles adicionales que apunten a corroborar la especificidad de estos morfolinios.

DISCUSIÓN DE RESULTADOS

Agrupamiento de HAEs en el genoma humano

Gracias al análisis de distribución de HAEs en el genoma humano, fue posible identificar los genes y regiones genómicas con la mayor densidad de elementos con signos de evolución acelerada en el linaje humano. En general, se encontró que los genes con unidades transcripcionales gigantes (con más de 800 kpb) tienen mayor probabilidad de contener un HAE. Este resultado es consistente con el hecho que estos genes tienen secuencias no codificantes (fundamentalmente intrones) más largas y acumulan más pares de bases conservadas por par de bases de gen. Dentro de este grupo de genes gigantes, se encuentra *NPAS3*, el gen que contiene el mayor número de HAEs en todo el genoma humano, con 14 elementos no codificantes ubicados en sus intrones. El análisis de agrupamiento de HAEs por Mb de genoma humano también mostró que la unidad transcripcional de *NPAS3* se encuentra ubicada dentro de las dos Mb más densamente pobladas de HAEs de todo nuestro genoma, con 14 y 12 HAEs respectivamente. Estos dos análisis indicarían que este *locus* pudo haber sido un sitio genómico donde se acumuló mucho cambio molecular a nivel de secuencias no codificantes conservadas desde la divergencia con el linaje de los chimpancés.

Análisis del patrón evolutivo en regiones codificantes y no codificantes en el *locus* de *NPAS3*

Por otro lado, el incremento dramático de la tasa de evolución molecular experimentada por estas 14 regiones no codificantes contrasta notablemente con la conservación a nivel de la secuencia codificante de *NPAS3*. En particular, los análisis evolutivos de rama por sitio sugieren que la región codificante de *NPAS3* evolucionó por selección negativa en el linaje humano, ocurriendo en la proteína únicamente dos sustituciones aminoacídicas en 940 residuos comparado con la región ortóloga de chimpancé. Adicionalmente, no se encontraron evidencias de evolución acelerada en la secuencia codificante de *NPAS3*, ni en las regiones no codificantes conservadas que no contienen HAEs presentes en este *locus* y/o regiones adyacentes, ni en el linaje humano ni en de los primates no humanos estudiados. Las únicas regiones con signos de aceleración en su tasa de evolución molecular son los HAEs y solamente para la rama terminal de los humanos (observándose en estas mismas regiones un patrón de evolución purificadora para el resto de los primates no humanos). Todos estos datos sugieren que el proceso de aceleración ocurrió específicamente en ciertas regiones no codificantes conservadas presentes en los intrones de *NPAS3* y no como consecuencia de fenómenos evolutivos a gran escala que involucraron grandes porciones de este *locus*.

Una de las desventajas cuando se estudian cambios en la tasa de evolución molecular es que resulta muy difícil identificar los procesos evolutivos responsables de la aceleración (Kostka et al. 2012). Entonces se requieren pruebas adicionales para distinguir los procesos evolutivos subyacentes, sean estos selección positiva, relajación de la selección negativa o eventos poblacionales (Pollard et al. 2006a; Kostka et al. 2012). Adicionalmente, la aceleración puede ser producto de procesos que no involucran necesariamente cambios en el modo de selección o eventos poblacionales como por ejemplo, el proceso evolutivo selectivamente neutro conocido como *GC-biased gene conversion* (gBGC) (Duret et al. 2002; Pollard et al. 2006; Duret y Arndt 2008). Conocer cuales fueron las causas últimas de la aceleración puede brindar indicios importantes sobre el impacto funcional de estos cambios en la biología de las especies. Dentro de los HAEs, los HARs muestran un sesgo en sustituciones del tipo AT→CG y una tendencia a ubicarse cerca de los extremos de los cromosomas (Kostka et al. 2012) (Pollard et al. 2006), donde las tasas de sustitución tienden a ser mayores (Lander et al. 2001). Una hipótesis sugiere que zonas con tasas altas de sustitución y sesgo fuerte en el tipo de sustituciones de bases débiles a bases fuertes puede producirse como consecuencia del proceso de gBGC (Duret et al. 2002; Duret y Arndt 2008; Duret y Galtier 2009). Se observó que 8 de los 14 HAEs presentes en el *locus* de *NPAS3* muestran un sesgo importante (mayor al 66%) de sustituciones del tipo débil (AT) a fuerte (CG), mientras que solo el HAR21 muestra la tendencia inversa. Aunque *NPAS3* no se encuentra en el extremo del cromosoma 14, esta fuerte tendencia en el tipo de sustituciones ocurrida en estos elementos acelerados podría indicar que el proceso de gBGC pudo ser una de las fuerzas evolutivas responsables del cambio en la tasa de evolución molecular de estos elementos. Por otro lado, un estudio reciente analizó *a posteriori* las posibles fuerzas evolutivas responsables de la aceleración de la tasa de sustitución en los HARs. Se encontró que más de la mitad de los HARs (55%) tienen tasas de sustitución que exceden significativamente la tasa neutra de evolución estimada localmente, sugiriendo que estas regiones pudieron estar sujetas a un proceso de evolución adaptativa (Kostka et al. 2012). Respecto de los HARs presentes en los intrones de *NPAS3*, este estudio encontró que el elemento HAR21 muestra signos de evolución positiva, mientras que el patrón de evolución de los elementos acelerados HAR89, HAR96 y HAR202 es mejor explicado por un proceso de relajación de la selección purificadora. Además, los elementos HAR173 y HAR189 tienen signos de gBGC (Kostka et al. 2012). Posteriormente, un estudio a escala genómica identificó regiones con evidencias de gBGC en humanos y chimpancés. Este análisis sugirió que los elementos HAR173, HAR189, HACNS96 y 2xHAR122 tienen patrones de sustitución nucleotídica compatibles con este proceso de evolución estos linajes (Capra et al. 2013b).

Entonces, es posible que el cambio en la tasa de evolución molecular experimentada por parte de los HAEs presentes en el *locus* de *NPAS3* (al menos 4 de los 14 elementos) haya sido consecuencia del proceso de gBGC. Sin embargo, serán necesarios estudios adicionales para poner en evidencia los procesos evolutivos responsables de la evolución acelerada experimentada por el resto de los HAEs de *NPAS3* aún no analizados. Es por ello que, hasta el momento, no es posible derivar conclusiones definitivas sobre la fuerza evolutiva subyacente predominante en la evolución acelerada experimentada por los HAEs de *NPAS3*.

Análisis de las sustituciones exclusivas de homínidos en los HAEs de *NPAS3*

Por otro lado, si bien en la actualidad *Homo sapiens* es el único representante vivo de nuestro linaje directo, durante los últimos cinco millones de años, diferentes grupos de homínidos estuvieron surgiendo sobre la Tierra. Estos grupos, con sus características variadas, compitieron por los recursos, coexistieron con otros grupos de homínidos y colonizaron nuevos ambientes con mayor o menor éxito (Tattersall 2000). Entonces, además de conocer los procesos evolutivos responsables del cambio en la tasa de evolución de los HAEs, develar el momento histórico en el cual estos cambios tuvieron lugar podría brindar una información muy valiosa. Por ejemplo, la datación de los sucesos de aceleración en los HAEs permitiría asociar estos eventos con el registro fósil de nuestro linaje, las condiciones ambientales existentes en ese momento y las características morfológicas que diferencian a los humanos modernos de aquellas otras poblaciones de homínidos. De esta forma, se podrían generar mejores hipótesis sobre el rol que los HAEs pudieron haber tenido en la evolución humana (Capra et al. 2013a). En la actualidad contamos con dos genomas de homínidos arcaicos: los Neandertales y los Denisovanos, un grupo de humanos arcaicos relacionados con los primeros (Noonan et al. 2006; Abecasis et al. 2010; Reich et al. 2010; Meyer et al. 2012). Recientemente, aplicando modernos métodos de secuenciación fue posible generar secuencias de alta cobertura del genoma del Neandertal, específicamente en las regiones de los HAEs. De esta forma se pudo estimar cuales sustituciones (de las específicas de los humanos) presentes en los HAEs estaban compartidas con los humanos arcaicos e identificar cuales se adquirieron más recientemente y son específicas de los humanos modernos (llamadas sustituciones modernas) (Burbano et al. 2012). Las sustituciones compartidas con los homínidos arcaicos se consideran antiguas y muy probablemente hayan surgido antes de la divergencia entre las poblaciones de los ancestros de los humanos modernos y los ancestros de los Neandertales y Denisovanos, divergencia que se estima tuvo lugar 553.000 a 589.000 años atrás (Prüfer et al. 2014). Excluyendo a los elementos 2xHARs, que no fueron incluidos en este análisis, todas las sustituciones humano específicas presentes en los HAEs de *NPAS3* están compartidas con el genoma de referencia de los Neandertales y por lo tanto serían antiguas. Sin embargo, se detectaron otros 69 HAEs con al menos una sustitución moderna, no

presente en el genoma de los Neandertales (Burbano et al. 2012). Estos resultados sugieren que la evolución acelerada sufrida por el grupo de HAEs de *NPAS3* tuvo lugar antes de la divergencia de los humanos modernos de los Neandertales/Denisovanos, hace más de 600.000 años. Sin embargo, posteriormente, la primera secuenciación de alta calidad del genoma de un Neandertal de Siberia (Prufer et al. 2014) permitió detectar la presencia de algunas sustituciones modernas en los sitios específicos de humanos en algunos de los HAEs de *NPAS3*. Tres pares de bases presentes en HAEs de *NPAS3* fueron anotadas como derivadas en humanos (diferentes a las bases ortólogas en Neandertales, Denisovanos y chimpancés). Dos de estas resultan ser polimórficas en las poblaciones humanas actuales (rs10149308 en HACNS658 y rs79570848 en HAR202) y una tercera no polimórfica (chr14:34160070 en 2xHAR157) (Prufer et al. 2014).

Por otro lado, el proyecto 1000 genomas secuenció, en su primera fase, 1.092 genomas humanos provenientes de 14 poblaciones de África, Europa, Este de Asia y América (Abecasis et al. 2010). Este análisis detectó sitios polimórficos en todos los HAEs de *NPAS3*, presentes en al menos una población humana. De entre estos, al menos cinco sitios polimórficos presentes en los elementos HAR189, HAR202, HACNS658 y 2xHAR142 se localizan en pares de bases que cambiaron exclusivamente en el linaje humano. En tres casos (rs79570848 en HAR202, rs10149308 en HACNS658 y rs183390399 en 2xHAR142), una de las variantes alélicas es igual a la base ancestral inferida. Adicionalmente, dos polimorfismos (rs145548775 en HAR189 y rs114570257 en 2xHAR142) tienen cada uno dos variantes alélicas y ambas en condición derivada respecto del estado ancestral inferido para esa posición. Todos estos hallazgos sugieren que existe un cierto número de sustituciones en los HAEs de *NPAS3* que pudieron haber surgido antes de la divergencia con el linaje de los Neandertales/Denisovanos, pero aún no se fijaron (como por ejemplo los polimorfismos rs10149308 en HACNS658, rs79570848 en HAR202 y rs183390399 en 2xHAR142) y otras que surgieron luego de esta divergencia y serían, posiblemente, exclusivas de los humanos modernos (rs114570257 en 2xHAR142 y base chr14:34160070 en 2xHAR157). Sin embargo, la gran mayoría de sustituciones humano específicas en los HAEs de *NPAS3* seguramente surgieron antes de la divergencia entre nuestros ancestros directos y los ancestros de los Neandertales y Denisovanos y no serían polimórficas en las poblaciones de humanos modernos. La disponibilidad de más secuencias genómicas de humanos modernos así como también de arcaicos podrá modificar estos números y permitir una mejor datación de los eventos de evolución acelerada en los intrones de *NPAS3*.

Ensayos funcionales en pez cebra transgénicos

La hipótesis general de este trabajo, basada en la idea original de King y Wilson, plantea que las diferencias fenotípicas entre chimpancés y humanos se deben principalmente a cambios en regiones regulatorias más que a cambios en las secuencias de las proteínas (King 1975). En este sentido, la excepcional acumulación de regiones no codificantes, conservadas y con evidencia de haber sufrido un proceso de evolución acelerada exclusivamente en el linaje humano presentes en los intrones de *NPAS3* sugiere que este gen pudo haber sufrido un cambio en su regulación de la expresión en nuestro linaje. Adicionalmente, *NPAS3* codifica para un factor de transcripción que tiene un rol clave en el desarrollo del sistema nervioso de los mamíferos (Erbel-Sieler et al. 2004; Brunskill et al. 2005; Pieper et al. 2005) y su disfunción ha sido asociada con el desarrollo de diferentes enfermedades psiquiátricas en humanos, entre ellas esquizofrenia (Kamnasaran et al. 2003; Yu et al. 2013). Toda esta evidencia apuntó a un posible, y hasta el momento no explorado, rol de *NPAS3* en la evolución del cerebro humano.

El primer objetivo para comenzar a poner a prueba la hipótesis antes mencionada fue estudiar si los 14 HAEs presentes en el *locus* de *NPAS3* tienen la capacidad de activar la expresión de EGFP en un ensayo de gen reportero en peces transgénicos. Este sistema presenta varias características que alentaron su uso como modelo para el estudio de regiones regulatorias potencialmente involucradas en la expresión de *NPAS3*. Por un lado, se trata de un modelo vertebrado en el cual se puede analizar el patrón de expresión de un transgén en un individuo completo y durante múltiples etapas del desarrollo (Fisher et al. 2006b). Adicionalmente, ensayos de hibridación *in situ* mostraron que la única copia presente del gen *NPAS3* en el genoma del pez cebra se expresa abundantemente en el sistema nervioso central en desarrollo, sugiriendo conservación en el patrón de expresión general entre los vertebrados dentro de este dominio.

Los ensayos de expresión transiente en peces transgénicos mostraron que al menos 11 de los 14 HAEs presentes en intrones de *NPAS3* tienen la capacidad de activar la expresión del gen reportero a 24 y 48 hpf de manera reproducible, dirigiendo la expresión de EGFP principalmente a diferentes dominios de sistema nervioso en desarrollo. Estos resultados son consistentes con la hipótesis que cambios en el patrón de expresión de este gen en el linaje humano (mediados a través de cambios en la actividad de algunos de los *enhancers* transcripcionales codificados en los HAEs) pudieron haber tenido algún rol en la evolución de nuestro cerebro. Adicionalmente, en concordancia con lo demostrado para regiones no codificantes conservadas humanas (Fisher et al. 2006a), estos resultados muestran que el pez cebra es un modelo muy útil para estudiar de forma rápida la actividad *enhancer* de regiones con signos de evolución acelerada en el linaje humano.

Por otro lado, se generaron líneas transgénicas de peces cebra llevando los elementos HAR21Hs, HAR21Pt, HAR89Hs, HAR89Pt, HAR96Hs, HAR202Hs, HAR202Pt, HACNS96Hs y HACNS221Hs. Los resultados de expresión de EGFP hallados en estas líneas estables a 48 hpf reprodujeron mayoritariamente los resultados obtenidos en los ensayos de expresión transiente. Esta observación está en línea con observaciones previas que sugieren que la reconstrucción del patrón de expresión obtenida por observación de un gran número de embriones de peces cebra transgénicos mosaicos, es representativa del patrón observable en líneas estables (Fisher et al. 2006b). Todos estos elementos (con excepción del HAR202Hs) fueron capaces de activar la expresión de EGFP en el cerebro y médula espinal a 48 hpf, siendo ambas estructuras parte de los dominios de expresión endógenos del gen *NPAS3* en pez cebra a esta edad. Adicionalmente, por fuera del SNC se observó abundante expresión en corazón, arcos faríngeos y aleta pectoral, estructuras que también expresan *NPAS3* a esta edad.

Con el objetivo de estudiar en mayor profundidad el impacto del proceso de evolución acelerada experimentada por los HAEs presentes en el *locus* de *NPAS3*, se comparó el patrón de expresión transiente producido por regiones ortólogas humanas y de chimpancé de los elementos HAR21, HAR89, HAR96 y HAR202. En los tres primeros casos, no se detectaron diferencias en la capacidad de activar la expresión de EGFP a 24 y 48 hpf entre elementos ortólogos; ambos pares de secuencias ortólogas funcionaron como potentes *enhancers*. Sin embargo, mientras la región HAR202 humana fue incapaz de producir un patrón de expresión consistente de EGFP a 24 y 48 hpf, el elemento ortólogo de chimpancé activó fuertemente la expresión de la proteína reportera en las mismas edades, dirigiendo la expresión de EGFP principalmente al sistema nervioso en desarrollo. Resultados similares se obtuvieron al comparar líneas transgénicas llevando estos elementos. Estos datos son consistentes con resultados previos que sugieren que el patrón de evolución del elemento HAR202 en el linaje humano se ajusta a un modelo de relajación de restricciones funcionales (Kostka et al. 2012). Análisis *in silico* de sitios probables para la unión de factores de transcripción (*transcription factor binding site*; TFBS, por sus siglas en inglés) en los HAEs de *NPAS3* y elementos ortólogos de chimpancé indican que la evolución acelerada experimentada por el elemento HAR202 en el linaje humano pudo haber llevado a la pérdida de dos TFBS: uno para la proteína *STAT5* y otro para el factor *BCL6*. Llamativamente, ambas proteínas se expresan en el cerebro mamífero en desarrollo (Markham, Schuurmans y Weiss 2007; Leamey et al. 2008) y tienen un rol importante en la neurogénesis a nivel de la corteza (Markham, Schuurmans y Weiss 2007; Tiberi et al. 2012). Se requerirán estudios adicionales para confirmar si la pérdida de

estos dos FTBS puede ser la causa molecular de la falta de actividad del elemento HAR202 humano en comparación con su elemento ortólogo en chimpancé.

La diferencia en el patrón de expresión del elementos HAR202 producida por los elementos humanos y chimpancés probablemente constituya el primer registro de un cambio funcional en un *enhancer* transcripcional en el cerebro entre estas dos especies cercanamente emparentadas. Adicionalmente, la alta tasa de sustitución sufrida por el elemento HAR202 durante la evolución de nuestro linaje, en comparación con la menor tasa de evolución experimentada por el elemento ortólogo en el linaje de los chimpancés así como de otros mamíferos, sugiere que la actividad revelada por el elemento de chimpancé podría ser representativa del estado ancestral. Sin embargo, sin estudios funcionales realizados sobre secuencias ortólogas provenientes de otras especies relacionadas con los humanos y chimpancés, no se puede excluir la posibilidad que el patrón de expresión observado para el elemento chimpancé represente, en realidad, el estado derivado.

Por otro lado, no se encontraron diferencias sustanciales en el patrón de expresión de EGFP entre líneas transgénicas de peces zebra llevando secuencias ortólogas de humano y chimpancé de los elementos HAR21 y HAR89. Sin embargo, estos resultados no excluyen la posibilidad que estas secuencias dirijan diferencialmente la expresión de EGFP a otras edades no estudiadas o presenten cambios discretos en los dominios de expresión que no fueron aún analizados con el detalle necesario (por ejemplo a nivel de subdominios y/o tipos celulares dentro de estos dominios).

Por otra parte, es interesante destacar que todos los HAEs estudiados en líneas de peces transgénicos (HAR21, HAR96, HAR202, HAR89, HACNS96 y HACNS221), muestran a 48 hpf patrones de expresión amplios que abarcan varios dominios anatómicos, entre los que se incluyen el cerebro, la médula espinal, corazón, arcos faríngeos y aleta pectoral (lugares endógenos de expresión del gen *NPAS3* en pez cebra a esa edad). Llamativamente, los dominios de expresión de estos elementos son altamente redundantes. Por ejemplo, cada uno de los dominios de expresión mencionados anteriormente está representado por varios HAEs. Existen múltiples ejemplos de genes, especialmente de aquellos activos durante el desarrollo, que tienen múltiples *enhancers* con actividades similares (Jeong et al. 2006; Frankel et al. 2010; L'Honore et al. 2010; Ghiasvand et al. 2011). Sin embargo, la razón de esta aparente redundancia es aún materia 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 (Hong, Hendrix y Levine

2008; Swami 2010; Frankel 2013). Se requerirán estudios adicionales para comprobar si este puede ser el rol que estén cumpliendo estos múltiples HAEs presentes en el *locus* de *NPAS3*.

Finalmente, se debe mantener cierta cautela a la hora de extrapolar los resultados funcionales de elementos humanos obtenidos en pez cebra, principalmente a nivel de los dominios de expresión de los *enhancers* (Ariza-Cosano et al. 2012). Estudios realizados comparando *enhancers* humanos en ratones y peces transgénicos frente a secuencias ortólogas en pez cebra en peces transgénicos (experimentos de tres vías) mostraron que en aproximadamente 38% de los casos estudiados (5/13), los resultados obtenidos (a nivel de dominios de expresión) al estudiar el *enhancer* humano en peces cebra mosaicos no pudieron ser reproducidos en el modelo mamífero debido a cambios en *trans* (es decir, a nivel de las proteínas encargadas de decodificar la información presente en los *enhancers*) (Ritter et al. 2010). Más recientemente, estudios realizados comparando el patrón de expresión de *enhancers* humanos en líneas de ratones y peces transgénicos también reportaron discrepancias entre modelos (en al menos un dominio de expresión) en la gran mayoría de los casos analizados (39/47, 83%) (Ariza-Cosano et al. 2012).

A nivel de los elementos HAE presentes en *NPAS3*, se encontró que transgenes portadores de la secuencia humana o de chimpancé del elemento 2xHAR122 son incapaces de activar la expresión del gen reportero en líneas de ratones transgénicos a 11.5 dpc (Capra et al. 2013a). Estos resultados no son consistentes con los hallados en los ensayos de expresión transientes en peces cebra, donde este elemento es capaz de activar la expresión de EGFP de forma robusta y consistente a 24 y 48 hpf (Kamm et al. 2013b), estadios del desarrollo del pez cebra que incluirían al estadio homólogo en ratón de 11.5 dpc (Irie y Kuratani 2010). Sin embargo, es importante destacar que la región genómica utilizada para generar los transgenes estudiados en ratón tiene el doble de longitud que la utilizada en los ensayos de expresión transientes en pez cebra (aprox. 1kpb frente a 469 pb respectivamente) y además, la primera contienen elementos conservados adicionales en su secuencia (Capra et al. 2013a). Entre otras posibles razones de esta falta de reproducibilidad de los resultados entre estos dos modelos experimentales se encuentran: las técnicas de transgénesis utilizados (mediada por transposasa vs. inyección pronuclear), el uso de diferentes promotores mínimos (*cFos* vs. *HSP68*) y genes reporteros (por ejemplo la proteína fluorescente EGFP frente a la proteína *LacZ* cuya presencia se revela mediante un ensayo enzimático), diferentes características de los embriones de estas especies que permiten un acceso diferencial a la visualización de la expresión del gen reportero (embriones transparentes de pez cebra frente a embriones translúcidos de ratón), entre otros. Además, como se mencionó anteriormente, no se puede descartar cambios en *trans* entre estos dos modelos (Ritter et al. 2010; Ariza-Cosano et al. 2012). Todas estas variables

podrían dar cuenta de esta diferencia en la actividad del elemento 2xHAR122 cuando estudiado en ratones o peces transgénicos.

Ensayos funcionales en ratones transgénicos

Por otro lado, se demostró que los elementos PC_2xHAR142Hs, PC_2xHAR142Pt y PC_2xHAR142Mm son *enhancers* capaces de dirigir la expresión del gen reportero *LacZ* a diferentes dominios del sistema nervioso en desarrollo en ratones transgénicos. Además, las secuencias PC_2xHAR142Pt y PC_2xHAR142Mm son capaces de activar la expresión del gen reportero en un subdominio del patrón de expresión del gen *Npas3* endógeno a 12.5 dpc, particularmente en aquellas regiones donde se codetecta expresión de *Npas3* con los marcadores *DCX* y *β TubIII* en el rombencéfalo y médula espinal. Este hallazgo sugiere fuertemente que estos *enhancers* podrían estar regulando la transcripción del gen *NPAS3*. Al comparar los dominios de expresión producidos por los *enhancers* de humano, chimpancé y ratón, se encontró evidencia que el proceso de evolución acelerada sufrida por la secuencia 2xHAR142 en el linaje humano pudo haber conducido a una expansión en el patrón de expresión del *enhancer* PC_2xHAR142Hs a 12.5 dpc, particularmente dentro de la corteza en desarrollo. Los análisis *in silico* de sitios probables para la unión de factores de transcripción realizados sobre los HAEs de *NPAS3* indican que la evolución acelerada experimentada por el elemento 2xHAR142 en el linaje humano pudo haber llevado a la ganancia y pérdida de múltiples TFBS. Se requerirán estudios adicionales para confirmar cuales de estos cambios, u otros aún no descritos, pueden ser la causa molecular de la heterotopía observada para el *enhancer* 2xHAR142Hs respecto de sus ortólogos en chimpancé y humano. Además, un desafío mayor será develar el impacto que este cambio pudo tener en la evolución del cerebro humano. Al respecto, es probable que la expansión en el dominio de expresión del *enhancer* PC_2xHAR142Hs en el linaje humano no esté relacionada con la ganancia de expresión de *NPAS3* en la corteza en desarrollo como un todo. Ensayos de ISH contra el ARNm de *Npas3* indican que este mensajero se expresa abundantemente en la corteza en desarrollo del ratón. Sin embargo, se requerirán estudios más detallados a nivel de subdominios y tipos celulares específicos para detectar cambios sutiles entre los dominios de expresión de *NPAS3* entre ratón y humano. Otra de las hipótesis posibles que se desprende de este hallazgo es que la corteza humana en desarrollo podría requerir niveles más altos de expresión de *NPAS3* respecto de la de otros mamíferos.

Además, se encontró una muy buena correspondencia entre los lugares de expresión del elemento 2xHAR142Hs obtenidos en ensayos de expresión transientes en pez cebra y líneas de ratones transgénicos. Los dominios de expresión hallados en peces transgénicos llevando la región conservada que contiene al elemento 2xHAR142Hs a 24 y 48 hpf incluyen al prosencéfalo,

mesencéfalo, rombencéfalo, ojos, médula espinal, arcos faríngeos y músculo somítico. Con excepción de este último dominio, que además coincide con una región de alta expresión ectópica en ensayos de expresión en pez cebra (Oksenberg et al. 2013), las restantes 6 estructuras mostraron expresión consistente de *LacZ* en ratones transgénicos llevando el *enhancer* PC_2xHAR142Hs a estadios del desarrollo (entre 10.5. y 14.5 dpc) comparables (Irie y Kuratani 2010). Sin embargo, el *enhancers* PC_2xHAR142Hs estudiado en ratones transgénicos mostró algunos dominios de expresión adicionales, no revelados por el ensayo con peces transgénicos como ser: los primordios de miembros, oído interno en desarrollo y los primordios de las vértebras. Es importante destacar que la región genómica utilizada para generar los ratones transgénicos tiene aproximadamente el doble de longitud que la utilizada en los ensayos de expresión transientes en pez cebra (502 pb y 263 pb respectivamente). Si bien ambas construcciones comparten la región conservada entre especies de vertebrados que contiene al elemento 2xHAR142, la secuencia estudiada en ratón lleva regiones no conservadas filogenéticamente entre mamíferos adicionales. Esta diferencia en las secuencias PC_2xHAR142Hs (estudiado en ratón) y 2xHAR142Hs (estudiado en pez cebra) podrían dar cuenta de estas diferencias en los dominios de expresión. Entre otras posibles razones de esta discrepancia entre modelos experimentales están los detallados anteriormente (cambios en *trans*, diferentes técnicas de transgénesis, diferentes promotores mínimos y genes reporteros, etc.). Nuevamente, y como ya fuera discutido para el elemento 2xHAR122, es importante ser cauteloso con la extrapolación directa de los resultado obtenidos en pez cebra a la actividad intrínseca de un *enhancer* en un contexto mamífero, particularmente en lo que respecta a los dominios de expresión hallados en el pez cebra. Finalmente, no es posible conocer la actividad endógena de *enhancers* humanos y de chimpancé (exceptuando, tal vez, los casos de mutaciones naturales) y por lo tanto, no existen estudios *in vivo* que indiquen cuan significativa puede ser la influencia del entorno (cambios en *trans*) en los ensayos de expresión utilizando ratones transgénicos. Entonces, los resultados de expresión de los *enhancers* PC_2xHAR142Hs y PC_2xHAR142Pt obtenidos en el modelo de ratón deben considerarse sólo una aproximación de la actividad que estos desarrollen en sus respectivos contextos nativos.

Origen evolutivo del 2xHAR142 y regiones con signos de exaptación en el genoma humano

Diferentes análisis bioinformáticos permitieron demostrar que el elemento conservado que contiene la secuencia 2xHAR142 es una novedad evolutiva de los mamíferos. Este elemento probablemente se haya originado, al menos parcialmente, por la exaptación de un retrotransposón tipo SINE de la familia MIR en el linaje que dio origen a los mamíferos placentarios. Por otro lado, el elemento 2xHAR142 no es la única secuencia no codificante conservada en los intrones de *NPAS3* que probablemente se haya originado por la exaptación de un elemento móvil. Otras 145 regiones

conservadas no codificantes adicionales fueron originalmente descritas como eventos de exaptación de elementos transponibles en este *locus* (Lowe, Bejerano y Haussler 2007; Lindblad-Toh et al. 2011; Lowe y Haussler 2012). Adicionalmente, se encontró que de las 19.897 unidades transcripcionales no redundantes anotadas en el genoma humano, 10.754 llevan al menos un elemento móvil posiblemente exaptado. De entre estas unidades transcripcionales, 10.707 llevan entre 1 y 144 elementos posiblemente exaptados no codificantes, mientras que solo 47 genes tienen 145 o más de estos elementos en sus unidades transcripcionales. Estos resultados sugieren que el *locus* de *NPAS3* pudo estar sujeto a una extensa coopción de elementos móviles para funciones regulatorias durante la evolución de los vertebrados.

Comparación del patrón de expresión de *NPAS3* entre vertebrados

Si bien el estudio inicial del patrón de expresión de *NPAS3* en pez cebra a 24 y 48 hpf buscó darle sustento al uso de pez cebra como un modelo experimental en el cual estudiar posibles regiones regulatorias asociadas a este gen, un análisis más detallado y extensivo (incluyendo estadios del desarrollo adicionales), revela indicios de una profunda conservación en el patrón de expresión espacial de este gen entre vertebrados. Fundamentalmente, este análisis se restringe al estudio de dominios de expresión generales y fácilmente identificables entre diferentes especies de vertebrados (en este caso ratón y pez cebra) y no excluye posibles particularidades a nivel de subdominios o tipos celulares específicos. Además, si bien esta comparación trata de restringirse a estadios homólogos del desarrollo de los vertebrados, estos estadios no son morfológicamente idénticos y están sujetos a cambios heterocrónicos en los tiempos del desarrollo de las diferentes estructuras (Gould 2010a; Irie y Kuratani 2010). Por lo tanto, también se pueden estar pasando por alto cambios sutiles en el momento de activación (o desactivación) de la expresión de este gen entre estas especies.

En pez cebra, y como fuera demostrado previamente en ratón, humano y pollo (Brunskill et al. 1999; Shin et al. 2010; Gould y Kamnasaran 2011), la expresión de *NPAS3* es particularmente abundante en el sistema nervioso central en desarrollo. Fue esta característica la que le valió el nombre de *Neuronal PAS domain protein 3* a este gen (Brunskill et al. 1999). Sin embargo, *NPAS3* también se expresa abundantemente por fuera del SNC (Brunskill et al. 1999) y existe evidencia que varios de estos dominios de expresión además estarían conservados entre vertebrados. Por un lado, a 3 y 5 dpf, *NPAS3* se expresa abundantemente en el corazón en desarrollo en peces cebra. Ensayos de ISH en ratón mostraron que *Npas3* se expresa en el pericardio y en las válvulas aórtica y pulmonar del corazón entre los estadios 12.5 y 16.5 dpc (Brunskill et al. 1999). Estos hallazgos

sugieren que la expresión de *NPAS3* en algunos de los tejidos del corazón podría ser una característica compartida entre los vertebrados.

Además, podría también estar conservada, al menos en parte, la expresión a nivel de los arcos faríngeos entre peces y mamíferos. Se observó que *NPAS3* comienza a expresarse en el techo de la boca a las 48 hpf y que a 3 y 5 dpf se expresa claramente en todos los arcos faríngeos. En ratones se encontró que *Npas3* se expresa también en los arcos faríngeos a 10.5 dpf (particularmente en el arco mandibular) (Brunskill et al. 1999). La expresión en este dominio es particularmente interesante. La estructura cartilaginosa que forma parte de los arcos faríngeos se forma a partir de células de la cresta neural (que migran y se introducen en las paredes de la faringe, entre las sucesivas hendiduras branquiales). Solamente los vertebrados presentan células de la cresta neural durante el desarrollo embrionario (Kardong 2007a). En el futuro, será interesante estudiar el rol de este gen en la evolución y función de estas estructuras dentro de los vertebrados.

Llamativamente, también se encontró conservación en la expresión de este gen a nivel de los miembros pares. En pez cebra, se encontró que *NPAS3* comienza a expresarse en los primordios de las aletas pectorales a 48 hpf y esta expresión se mantiene durante los estadios larvales estudiados (3 y 5 dpf). Por otro lado, en ratón, se encontró expresión de *Npas3* en el mesénquima de los primordios de los miembros a 12.5 y 14.5 dpf (Brunskill et al. 1999). Más allá de las diferencias estructurales entre las aletas pares de los peces teleósteos y los miembros de los tetrápodos, existe abundante evidencia que sostiene su homología (como ser su posición respecto del cuerpo como apéndices pares, la homología entre los elementos cartilaginosos proximales que los forman, entre otros) (Hall 2007). Llamativamente, también serían homólogas las redes de genes y cascadas de señales que subyacen el desarrollo temprano de estas dos estructuras, tales como la expresión de los genes *HoxA*, *HoxD* y *Shh* (Hinchliffe 2002; Hall 2007). Nuevamente, estas observaciones sugieren que la expresión de *NPAS3* en los miembros en desarrollo podría ser una característica compartida entre los vertebrados.

A 5 dpf se encontró expresión de *NPAS3* en la vejiga natatoria del pez cebra. Si bien se encuentran muchas diferencias entre las vejigas natatorias gaseosas de los peces actinopterigios (como la del pez cebra) y los pulmones de los tetrápodos, estas dos estructuras presentan similitudes en su desarrollo (ambas se originan de la pared del tubo digestivo o de la faringe) y anatomía (tienen aproximadamente los mismos músculos y nervios y comparten semejanzas en su vasculatura) (Kardong 2007b) (Longo, Riccio y McCune 2013). Estas características compartidas han sido, en algunos casos, interpretadas como evidencia de homología entre vejigas gaseosas y pulmones

(Kardong 2007a; Longo, Riccio y McCune 2013). Llamativamente, se encontró que en ratones, *Npas3* comienza a expresarse en pulmones en desarrollo a los 10.5 dpc. La expresión de *NPAS3* (seguida por ISH y IHQ) en pulmones persiste hasta la adultez de los animales (Zhou et al. 2009).

Todos estos datos indicarían que el patrón de expresión de *NPAS3* estaría altamente conservado entre vertebrados, incluyendo el cerebro, médula espinal, corazón, arcos faríngeos y, los órganos posiblemente homólogos vejiga gaseosa y pulmones. Sería interesante estudiar el patrón de expresión de *NPAS3* en otros grupos de vertebrados más cercanamente emparentados para corroborar estas observaciones.

Por otro lado, ciertos aspectos de la expresión y/o función de *NPAS3* podrían haberse establecido mucho antes del origen de los vertebrados, divergencia que tuvo lugar hace aproximadamente 500 millones de años (Erbel-Sieler et al. 2004; Kardong 2007a). Por ejemplo, la expresión de *NPAS3* en sistema nervioso central podría haber surgido antes de la divergencia entre protostomados y deuterostomados, hace aproximadamente 600 millones de años (Aris-Brosou y Yang 2003). Se encontró que el gen ortólogo a *NPAS3* en mosca, *Trh*, se expresa en cordón nervioso ventral y el cerebro en desarrollo entre los estadios embrionarios 13-16 de *Drosophila melanogaster* (Tomancak et al. 2002; Tomancak et al. 2007). También existe evidencia que el gen ortólogo a *NPAS3* en el nematodo *Caenorhabditis elegans* se expresaría en sistema nervioso (Harris et al. 2010; Frederic et al. 2013). No obstante, todavía no está claro si el SNC de los vertebrados e invertebrados tiene un origen evolutivo común (Arendt y Nubler-Jung 1994; Benito-Gutierrez y Arendt 2009) o si evolucionaron de forma independiente (Holland 2003; Lowe et al. 2003). Además, tampoco se conoce la función de *Trh* a nivel del sistema nervioso en desarrollo de los invertebrados. Será interesante estudiar si *NPAS3/Trh* participa del desarrollo de sistema nervioso en los invertebrados (tanto protostomados como deuterostomados), como ya fuera demostrado para los vertebrados (Brunskill et al. 2005).

Ensayos funcionales con morfolinós contra el ARNm de *NPAS3*

Los ensayos con morfolinós contra el ARNm de *NPAS3* en pez cebra, si bien incompletos, son promisorios. Por un lado, los morfolinós utilizados, si bien actuarían por medio de dos mecanismos distintos (bloqueo de la traducción para el MO1 y bloqueo del *splicing* a nivel del intrón 3 para el MO2), produjeron fenotipos similares. Esta convergencia en los fenotipos de ambos oligonucleótidos apoyaría la idea de que son específicos para este mensajero. Sin embargo, quedan pendientes controles adicionales que apunten a demostrar específicamente este punto (por ejemplo,

RT-PCR contra el intrón 3 de *NPAS3* para demostrar su inclusión en el mensajero maduro como consecuencia del bloqueo producido por el MO2).

Por otro lado, si se compara el patrón de expresión de *NPAS3* a 3 y 5 dpf y los lugares donde se observaron defectos en los animales morfantes, se encuentra nuevamente una muy buena correspondencia. Por ejemplo, uno de los defectos más severos encontrados en los morfantes estuvo a nivel de los arcos faríngeos, donde se halló una completa ausencia de los elementos cartilagosos asociados a estas estructuras. Paralelamente, los ensayos de ISH demostraron que 3 y 5 dpf *NPAS3* se expresa abundantemente en estos elementos en el pez cebra. Este hallazgo sugeriría que *NPAS3* podría tener un rol clave en el desarrollo de los arcos faríngeos y estructuras derivadas. Por ejemplo, las branquias en los peces teleósteos derivan de los arcos faríngeos III, IV, V, VI y VII (Kardong 2007b). De estos arcos, los primordios de los filamentos de las branquias, si bien muy difíciles de ver con microscopía de campo claro, comienzan a formarse a los 2-3 dpf y son funcionales para el intercambio gaseoso a los 14 dpf (Kimmel et al. 1995; Rombough 2002). Además de participar en el intercambio gaseoso, las branquias de los peces son un órgano donde tienen lugar varias funciones homeostáticas como ser osmoregulación y control del pH, entre otras (Rombough 2002). Por lo tanto, la disrupción de la formación de branquias que muestran las larvas de pez cebra tratadas con morfolinolinos contra *NPAS3*, sería una condición letal. Por otro lado, si bien se encontró evidencia que *Npas3* se expresa en, al menos algunos, de los arcos faríngeos en embriones de ratón (Brunskill et al. 1999), estudios realizados en ratones donde se disrumpió el gen *Npas3* no reportaron la presencia de defectos morfológicos a nivel de los derivados de estas estructuras (Erbel-Sieler et al. 2004; Brunskill et al. 2005; Zhou et al. 2009). Estos resultados podrían sugerir que existe una diferente sensibilidad a la deficiencia en los niveles de expresión de *NPAS3* a nivel de los arcos faríngeos entre teleósteos y mamíferos o que la actividad de este gen a este nivel no estaría conservada entre estos grupos. Además, no se puede descartar que la hipoplasia encontrada a nivel del desarrollo de los arcos faríngeos en los animales tratados con morfolino sea consecuencia de los oligonucleótidos actuando sobre otro/s ARNm. Se requerirán ensayos adicionales para descartar esta última posibilidad.

Por otro lado, existen ciertas similitudes entre los fenotipos descritos en ratones *Npas3*^{-/-} y las larvas de pez cebra tratadas con morfolinolinos contra *NPAS3*. Por un lado, los ratones *knock out* para la expresión del gen *Npas3* son más chicos y pesan un 20% menos que los ratones de genotipo salvajes (Brunskill et al. 2005). Un fenotipo similar se encontró en las larvas de pez cebra morfantes. Se halló que a 3 dpf las larvas inyectadas con morfolinolinos son aproximadamente un 10% más pequeñas (considerando el largo del cuerpo medido desde la cabeza a la base de la aleta caudal)

que las larvas control, y esta diferencia es significativa. Un fenotipo similar se encontró en larvas de 5 dpf pero únicamente para el morfolino MO1. Además, también se encontraron diferencias significativas en el diámetro del ojo (medido perpendicularmente a la superficie de la boca) y la longitud de la cabeza (medida como distancia máxima desde la base de la boca al techo de la cabeza entre la vesícula ótica y el ojo) en los animales tratados con morfolino frente a los animales control a 3 y 5 dpf. Mientras que la reducción en la longitud de la cabeza en los animales tratados con morfolino fue proporcional a la reducción del tamaño corporal (con excepción del MO1 a 5 dpf), la reducción del diámetro del ojo en los animales morfantes no fue proporcional a la reducción del cuerpo, sino más severa. En su conjunto, estos resultados serían consistentes con los hallados en ratones deficientes para la expresión de *Npas3* (Brunskill et al. 2005), y podrían sugerir un rol conservado y aun no estudiado de *NPAS3* relacionado con el control de tamaño corporal y desarrollo del ojo.

Adicionalmente, se identificó un segundo defecto fenotípico semejante entre larvas de pez cebra morfantes de *NPAS3* y ratones *Npas3*^{-/-}. A 3 dpf se encontró que el 75% y 60% de los peces inyectados con los morfolinis MO1 y MO2 respectivamente desarrollaron hidrocefalia con agrandamiento del ventrículo cerebral. De forma similar, los ratones adultos *Npas3*^{-/-}, entre otras anomalías anatómicas del cerebro, muestran un agrandamiento de los ventrículos cerebrales (Brunskill et al. 2005). Estudios adicionales sugieren que *NPAS3* es un factor necesario para la correcta producción de nuevas neuronas durante el desarrollo embrionario de ratón y que, consecuentemente, la disminución de los niveles de expresión de *Npas3* podrían ocasionar la reducción del grosor de ciertas capas neuronales del cerebro (Pieper et al. 2005). En estudios futuros se analizará en más detalle el cerebro de los animales morfantes de *NPAS3* en busca de indicios de un rol pro neurogénico de esta proteína en el cerebro de pez cebra.

También, las larvas morfantes de *NPAS3* mostraron un marcado déficit en el inflado de la vejiga natatoria. Mientras que un 100% de los animales control presentaban la vejiga natatoria inflada a 5 dpf, esto ocurrió solo en un 6,25% y 12% de los animales tratados con los MO1 y MO2 respectivamente. Nuevamente, ensayos de ISH indican que *NPAS3* se expresa en la vejiga natatoria de estos animales a esta edad. Como se discutió anteriormente, existen indicios que sugieren que las vejigas natatorias gaseosas de los peces actinopterigios y los pulmones de los tetrápodos podrían ser estructuras homólogas (siendo la vejiga gaseosa un derivado de un pulmón ancestral) (Kardong 2007a; Longo, Riccio y McCune 2013). Llamativamente, los ratones *Npas3*^{-/-} muestran defectos severos en la estructura de los alvéolos con un posterior desarrollo de bronquiectasia. Adicionalmente, el 95% de los ratones *Npas3*^{-/-} muere dentro de las primeras 48 horas luego del

nacimiento debido fundamentalmente a problemas respiratorios como consecuencia de las anomalías a nivel de los pulmones (Zhou et al. 2009). Estudios futuros buscarán develar los mecanismos moleculares que subyacen este posible rol de *NPAS3* en el desarrollo de la vejiga natatoria (así como también otras estructuras), y conocer si éstos son similares a los descritos para el desarrollo de los pulmones de ratón.

Otra de las anomalías halladas en los animales tratados con morfolidos contra *NPAS3* a 3 y 5 dpf estuvo a nivel del corazón. Un gran porcentaje de las larvas tratadas desarrolló edema de pericardio. Nuevamente, el corazón es un dominio de alta expresión de *NPAS3* durante el desarrollo del pez cebra (y otros vertebrados). Sin embargo, estudios en ratones *knock out* para la expresión del gen *Npas3* no reportaron la presencia de defectos morfológicos a nivel del desarrollo y/o función del corazón (Erbel-Sieler et al. 2004; Brunskill et al. 2005; Zhou et al. 2009). Nuevamente, estos resultados podrían sugerir diferencias de sensibilidad a la deficiencia de *NPAS3* a nivel del corazón entre teleosteos y mamíferos o una función de *NPAS3* no conservada entre estos grupos. Además, no se puede descartar la posibilidad que el fenotipo observado en los animales tratados con morfolino sea consecuencia de inespecificidad en el blanco de acción de los morfolidos utilizados.

***NPAS3* posiblemente forme parte de un módulo genético muy estable y antiguo asociado a la morfogénesis por ramificación**

Hasta ahora se hizo hincapié principalmente en el rol que *NPAS3* dentro de los vertebrados. Sin embargo, existe evidencia que este gen pudo tener un origen mucho más antiguo, que precede incluso al origen de animales con simetría bilateral (Yan et al. 2014). Por otro lado, recientemente se sugirió que el gen *NPAS3* sería una innovación genética de los metazoos. En este sentido, se encontró genes ortólogos a *NPAS3* en los genomas de 24 especies de metazoos (incluyendo al cordado *Homo sapiens*, el urocordado *Ciona intestinalis*, el artrópodo *Drosophila melanogaster*, el nematodo *Caenorhabditis elegans* y el cnidario *Nematostella vectensis*), pero no se encontraron evidencias de su presencia en 112 genomas de no metazoos (incluyendo tanto eucariotas unicelulares y procariotas) (Frederic et al. 2013). Incluso se detectó un ortólogo a *NPAS3* en el genoma del placozoo *Trichoplax adhaerens* (Frederic et al. 2013), un grupo aparentemente intermedio entre las esponjas y los cnidarios (Srivastava et al. 2008). Aunque la reciente secuenciación del genoma del ctenóforo *Mnemiopsis leidyi* puso en duda la ubicación filogenética de los placozoos, sugiriendo que estos podrían constituir el grupo hermano de bilateria (Ryan et al. 2013). De todas formas, estos hallazgos permitirían suponer que *NPAS3* es un gen muy antiguo que surgió en el linaje de los metazoos y antes de la divergencia del grupo de los eumetazoos (cnidarios + bilateria) y cuya función ancestral pudo estar vinculada con el desarrollo de la multicelularidad en

los animales (Frederic et al. 2013). Además, la presencia de *NPAS3* en *T. adhaerens*, un organismo morfológicamente simple que tiene solo cuatro tipos celulares distintos y que carece de células nerviosas y musculares diferenciadas (como sí ocurre en los eumetazoos) (Srivastava et al. 2008), sugiere que la función ancestral de *NPAS3* no estaría relacionada con el sistema nervioso. Sin embargo, no se puede descartar que la simplicidad morfológica de este placo-zoo sea una condición derivada (Bridge et al. 1995; Ryan et al. 2013).

Por otro lado, además del origen profundo de *NPAS3* en el linaje de los metazoos, hay otros aspectos vinculados a este gen que parecen también estar compartidos entre diferentes grupos de animales. Llamativamente, las redes de genes que intervienen en el proceso de morfogénesis por ramificación de estructuras tales como las tráqueas en la mosca, el oído de los vertebrados y los pulmones de ratón son notoriamente similares (Fritzscht et al. 2007b). La morfogénesis por ramificación es un tipo de crecimiento que tiene lugar cuando se debe maximizar el área total en contacto entre una determinada estructura y el entorno que lo rodea, especialmente cuando el volumen de esta estructura debe permanecer en un volumen reducido (Davies 2005). La alta similitud entre las redes de genes ortólogos detrás del desarrollo de estas estructuras tan diversas sugiere que ciertos genes de esta red, como ser *NPAS1/3/Trh*, *ARNT/Tgo*, *FGF/Branchless* y *FGFR/Breathless*, formarían parte de una red regulatoria altamente conservada dentro de bilateria (Fritzscht et al. 2007b). Esta red estaría vinculada funcionalmente con varios aspectos de la morfogénesis por ramificación, incluyendo la intrusión y extrusión de órganos tales como pulmones, glándulas, tráqueas y miembros (Fritzscht et al. 2007b; Zhou et al. 2009; Chung, Chavez y Andrew 2011). La impresionante conservación de la red molecular que subyace el desarrollo de estos órganos no homólogos dentro de bilateria sugiere que la misma pudo haber sido cooptada múltiples veces como un módulo funcional muy estable e incorporada a los programas de desarrollo de diferentes estructuras (Fritzscht et al. 2007a). Adicionalmente, existe evidencia de que parte de los elementos de la vía de *NPAS3/Trh*, (como ser los genes *ARNT/Tango* y *FGFR/Breathless*) (Fritzscht et al. 2007a), también serían innovaciones genéticas de los metazoos (Frederic et al. 2013). Esto permite pensar que el surgimiento de este circuito fue muy antiguo y pudo haber sido crítico para la biología de los metazoos. Una pregunta clave para entender las bases genéticas de la evolución morfológica de estos órganos, será entender cómo este módulo se incorporó a los programas de desarrollo de diferentes órganos, de forma tal de conservar su arquitectura genética pero participando en el surgimiento de estructuras variadas (Fritzscht et al. 2007b).

Particularmente, sabiendo que *NPAS3* tiene una función clave en el desarrollo del SNC en mamíferos (Brunskill et al. 2005; Pieper et al. 2005), y el crecimiento de las neuritas (Yu et al.

2013) y que además la morfogénesis por ramificación participa de eventos tales como la arborización de las neuronas (Davies 2005), será muy interesante estudiar no solo el rol de este gen sino la actividad de esta red génica completa (que posiblemente incluya a *NPAS3*, *NPASI*, genes de las familias de factores de crecimiento de fibroblastos y sus receptores; *FGFRs* y *FGFs* por sus siglas en inglés, entre otros) (Zhou et al. 2009) en el desarrollo del cerebro mamífero. Los resultados provenientes de este estudio podrían ser clave para dilucidar el impacto que posibles cambios en la regulación de *NPAS3* exclusivamente en el linaje humano pudieron haber tenido en la evolución de nuestro cerebro. Este enfoque, si bien experimentalmente complejo, podría resultar mucho más apropiado para encarar el estudio de la relación entre genotipo y fenotipo (Carroll 2008; Lamb 2013).

***NPAS3*, ¿Un gen con una compleja red regulatoria?**

El cambio linaje específico en la actividad de un elemento regulatorio de la expresión génica puede llegar a tener un impacto nulo sobre el fenotipo de una especie si esta actividad no se traduce (de alguna forma) en cambios en las propiedades del patrón expresión de su gen blanco (u otros genes vecinos). Existe evidencia creciente que las regiones regulatorias de los genes de los metazoos (y particularmente los involucrados en el desarrollo de los organismos) tienen regiones regulatorias particularmente complejas extendidas a lo largo de amplias regiones genómicas y que incluyen múltiples *enhancers*, muy frecuentemente con dominios de expresión llamativamente similares (Barolo 2012; Frankel 2013). Por un lado, la redundancia de función de estos *enhancers* podría estar vinculada a una mayor estabilidad fenotípica (Hong, Hendrix y Levine 2008; Frankel et al. 2010) (Hong, Hendrix y Levine 2008; Frankel et al. 2010), así como también podrían ser fuente de novedades evolutivas (Hong, Hendrix y Levine 2008). Adicionalmente se observó que genes con regiones regulatorias particularmente complejas pueden tener *enhancers* cuya actividad no se traduce en el patrón de expresión de su gen blanco, sino que resultan ser filtradas por algún mecanismo aún no descrito pero posiblemente vinculado con la estructura tridimensional que adopta ese *locus in vivo* (Marinic et al. 2013).

Dada toda esta evidencia, resulta importante conocer la dinámica regulatoria de *NPAS3* para comprender el impacto evolutivo, si alguno, de cambios linaje específicos en la actividad de un *enhancer* (como el codificado en el elemento 2xHAR142Hs). Por ejemplo, esto involucra conocer cuantos elementos regulatorios podrían estar participando en la regulación de *NPAS3*, el tamaño (en el genoma) de la región regulatoria, si hay otros genes que podrían estar siendo coregulados junto con *NPAS3*, entre otros. Conociendo este contexto, se podrán hacer mejores hipótesis sobre el rol de cambios regulatorios puntuales.

Por un lado, existe evidencia que algunos aspectos de la red regulatoria de *NPAS3* estarían muy conservados (como sugiere la alta conservación en el patrón de expresión de este gen entre vertebrados). Por un lado, *NPAS3* forma parte de un *locus* microsinténico con el gen *Egl Nine Homolog 3 (EGLN3)*, un gen no relacionado evolutivamente con *NPAS3* (es decir, que no proviene de eventos de duplicación) (Irimia et al. 2012). Particularmente, se encontró esta asociación está conservada en deuterostomados (en los vertebrados *Homo sapiens*, *Gallus gallus* y *Xenopus tropicalis*, el cefalocordado *Branchiostoma floridae* y el equinodermo *Strongylocentrotus purpuratus*, representantes de este grupo) y en el anélido *Capitella teleta* (Irimia et al. 2012). Esto sugiere que la asociación física entre los genes *NPAS3-EGLN3* puede haber surgido hace más de 600 millones de años, en el último ancestro común de estos dos grupos (posiblemente en el ancestro de toda bilateria).

La conservación filogenética de la asociación física de genes tiene varias interpretaciones. Por un lado, puede representar un *locus* para el cual no se fijó todavía, por azar, eventos de rearrreglos cromosómicos. Siguiendo esta interpretación, la asociación sería meramente azarosa y sin impacto en la biología de las especies (Nadeau y Taylor 1984). Alternativamente, la asociación física de los genes, evidenciada como regiones de sintenia evolutivamente conservada, puede deberse a restricciones funcionales. En este segundo caso, las restricciones pueden deberse a que los genes en la región sinténica comparten elementos regulatorios, es decir, son co-regulados (Tena et al. 2011; Andrey et al. 2013). Alternativamente, la conservación de la sintenia puede deberse a la formación de bloques de regulación génica. Estos bloques de regulación génica se definen como regiones que contienen una alta densidad de regiones no codificantes altamente conservadas y elementos regulatorios distribuidos sobre extensos dominios génicos dentro y alrededor de su gen blanco. Esta conformación espacial de regiones regulatorias además se caracteriza por extenderse más allá de los genes que flanquean el gen blanco. Sin embargo, y a diferencia del caso de co-regulación, las regiones regulatorias del gen blanco no actúan sobre los genes flanqueantes (Maeso et al. 2012; Marinic et al. 2013). En este contexto, la sintenia se conserva como un subproducto de la selección actuando de forma tal que no se pierda la asociación entre el gen blanco y sus regiones regulatorias (Irimia et al. 2013).

El significado de la profunda conservación filogenética en la sintenia del *locus NPAS3-EGLN3* es incierto. Pero muy seguramente esta asociación refleje las restricciones funcionales asociadas con mantener los genes blanco y sus regiones regulatorias ligadas (Irimia et al. 2012). En línea con lo anterior, la conservación de la sintenia en este *locus* es particularmente sorprendente si se

contempla el linaje de los vertebrados. *NPAS3* se encuentra en el *locus* sinténico multigénico más extenso conservado entre los genomas de humanos y pez cebra (Kikuta et al. 2007). Abarcando más de 3 Mb en el genoma humano (chr14: 31972332-35409930, NCBI36/hg18) y conteniendo 18 genes, la organización de esta región genómica (tanto en regiones codificantes, como entre regiones conservadas intrónicas e intergénicas) se ha conservado por más de 450 millones de años de evolución independiente de los linajes de los teleosteos y mamíferos (Kumar y Hedges 1998; Kikuta et al. 2007). Nuevamente, esta fuerte conservación estructural muy probablemente esté generada por la existencia de elementos regulatorios capaces de actuar a larga distancia, que deben permanecer asociados a su/s gen/es blanco para mantener su función. Adicionalmente, la profundidad de la conservación y la extensión genómica del *locus* sinténico en el cual se halla *NPAS3* podrían estar dando cuenta de la complejidad y antigüedad de la red regulatoria involucrada.

Adicionalmente, no se puede descartar que la profundidad de la conservación de la sintenia en este *locus* responda a restricciones funcionales asociadas con mantener un dominio topológico en la cromatina (regiones de cromatina típicamente de cientos de kpb dentro de las cuales las secuencias interactúan más frecuentemente entre sí que con secuencias por fuera de este dominio) (Dekker, Marti-Renom y Mirny 2013). Se postula que estos dominios topológicos de asociación (o *transcriptional associates domains*; TADs por sus siglas en inglés), tienen por función asegurar y/o facilitar la interacción entre regiones regulatorias y sus genes blanco, limitando interacciones espurias (Dekker, Marti-Renom y Mirny 2013). A diferencia de la expresión de los genes en ellos, la estructura de los TADs es altamente estable y se mantiene entre los diferentes tipos celulares y estadios del desarrollo (Dixon et al. 2012; Nora et al. 2012). Adicionalmente, TADs adyacentes se pueden asociar en estructuras de orden superior y formar dominios estructurales de varias Mb de longitud donde el estado transcripcional general de la región puede estar activo o inactivo (Gibcus y Dekker 2013). Llamativamente, un estudio reciente, donde se comparó estructuras de la cromatina de alto orden (incluidos TADs) entre ratones y humanos, encontró que existe un cambio en el estado estructural de la cromatina en una región a menos de 200 kpb del *locus* sinténico en el cual se encuentra *NPAS3* (chr14: 31900000- 32899999, NCBI37/hg19) (Chambers, Bickmore y Semple 2013). Si bien este cambio no involucraría directamente el *locus* sinténico en el cual se halla *NPAS3*, es interesante especular sobre el impacto de este cambio evolutivo tuvo sobre la estructura tridimensional de la cromatina en esta región. Estudios más detallados sobre la interacción entre regiones genómicas dentro de este *locus* podrían brindar más indicios al respecto. También sería interesante estudiar si alguno de los otros genes humanos con alta densidad de HAEs se encuentra en regiones con alta divergencia en la estructura de la cromatina de alto nivel entre humanos y ratones.

Por otro lado, existe evidencia adicional de la compleja red de regulación que podría estar controlando la expresión génica de *NPAS3*. Se encontró que este gen se encuentra en la 9^{na} región genómica (chr14:33068679-34779105, NCBI37/hg19) con mayor número de elementos no codificantes ultraconservados (o *ultraconserved non-coding elements*; UCNEs por sus siglas en inglés) del genoma humano. Esta región contiene 71 UCNEs que se encuentran distribuidos en un *locus* genómico que abarca a los genes *AKAP6*, *NPAS3*, *EGLN3* y *SPTSSA* (Dimitrieva y Bucher 2012). Si bien existen diferentes definiciones de elementos ultraconservados (Dermitzakis et al. 2002; Bejerano et al. 2004; Woolfe et al. 2005), en este caso el término hace referencia a regiones no codificantes de más de 200 pares de bases de largo, que están conservadas más de un 95% entre los genomas de humano y pollo (Dimitrieva y Bucher 2012). Existe abundante evidencia funcional demostrando que gran parte de las regiones no codificantes altamente conservadas en vertebrados funcionan, en general, como elementos regulatorios de la expresión génica y este tipo de elementos tienden a encontrarse en altas densidades en torno de genes con funciones claves durante desarrollo embrionario (Sandelin et al. 2004; de la Calle-Mustienes et al. 2005; Pennacchio et al. 2006; Elgar y Vavouri 2008).

Toda esta evidencia sugiere que la región regulatoria de *NPAS3* posiblemente sea muy extensa y contenga múltiples *enhancers* y/u otros elementos regulatorios (muchos más de los estudiados funcionalmente en este trabajo). Incluso, estos resultados apuntan a que este gen podría ser uno de los genes con más regiones regulatorias de los vertebrados y/o incluso estar alojado en un TAD particularmente grande. Además, no se puede descartar que algunos de estos elementos regulatorios participen del control de la expresión de varios genes simultáneamente dentro de este *locus* (como *AKAP6*, *EGLN3* o *SPTSSA*) o, alternativamente, que haya un bloque de regulación génica, donde elementos distales (ubicados a más de 1 Mb del sitio de inicio de la transcripción) interaccionen, casi con exclusividad, con el promotor de *NPAS3*.

Dinámica de los elementos no codificantes conservados en el *locus* de *NPAS3*

En este trabajo se demostró que el *locus* de *NPAS3* contiene la mayor densidad de elementos de evolución acelerada del genoma humano. Además, los resultados obtenidos en peces cebra transgénicos indican que 11 de los 14 HAEs presentes en el *locus* de *NPAS3* actúan como *enhancers* transcripcionales dirigiendo la expresión de la proteína reportera en un patrón de expresión que es coincidente con el patrón de expresión de éste gen. Estos hallazgos sugieren fuertemente que estos elementos regulan la expresión del gen *NPAS3*.

Por otro lado, dada la gran acumulación de elementos no codificantes altamente conservados entre vertebrados (Dimitrieva y Bucher 2012) y la particular estructura genómica de este *locus* dentro los vertebrados y de bilateria (Kikuta et al. 2007; Irimia et al. 2012), es probable que este gen posea una gran cantidad de elementos regulatorios adicionales aún no estudiados. Además de los elementos no codificantes altamente conservados entre vertebrados, *NPAS3* es un gen que posee en sus intrones al menos, tres elementos no codificantes conservados específicos de primates y un elemento no codificante conservado específico de roedores (es decir, regiones no codificantes que surgieron exclusivamente en estos linajes y que luego de su surgimiento probablemente hayan estado sujetas a evolución purificadora) (Takahashi y Saitou 2012). Estos elementos se suman a las 85, 38 y 14 regiones derivadas de elementos móviles presentes en los intrones de *NPAS3* que probablemente hayan sido cooptadas para funciones regulatorias en el linaje que dio origen a los euterios, mamíferos y vertebrados respectivamente (Kamm et al. 2013a).

Como fuera sugerido para genes con regiones conservadas con diferentes niveles de profundidad (Takahashi y Saitou 2012), una posibilidad es que cada una de las regiones no codificantes y con distintos niveles de conservación dentro de los intrones de *NPAS3* constituyan verdaderos elementos regulatorios que actúan sobre este gen y que hayan contribuido (mediante la adición de actividades regulatorias novedosas) con la evolución de diferentes grupos de organismos (por ejemplo, vertebrados, mamíferos, roedores, primates, etc.). Otra posibilidad es que los elementos regulatorios que fueron surgiendo durante la evolución de los diversos linajes de vertebrados en los intrones de este gen tengan todas funciones similares (aun cuando no sean secuencias homólogas) (Takahashi y Saitou 2012). En este caso, si se demuestra la coopción independiente de estos elementos y sus funciones análogas, se podría estar frente a eventos de evolución convergente sobre los elementos regulatorios de este gen (de forma similar a lo reportado para los *enhancers* neuronales del gen de proopiomelanocortina) (Franchini et al. 2011).

Si bien harán falta experimentos adicionales en modelos mamíferos, los ensayos de expresión sobre las secuencias ortólogas de chimpancé de los elementos HAR21, HAR89, HAR96 y HAR202 (secuencias que probablemente representen mejor la función ancestral debido a que su evolución es mejor explicada por selección purificadora en el linaje primate no humano) (Pollard et al. 2006) en peces transgénicos, indican una notable similitud en sus dominios de expresión, independientemente de la profundidad de conservación de estas secuencias, al menos en los estadios del desarrollo analizados. Esta observación es consistente con la gran conservación del patrón general de expresión de *NPAS3* observada entre vertebrados. Sin embargo, el estudio comparativo realizado sobre el *enhancer* 2xHAR142 mostró que este elemento pudo haber cambiado su

actividad exclusivamente en el linaje humano, pudiendo de esta forma, haber contribuido con actividades regulatorias novedosas. La región HAR202 podría constituir un segundo ejemplo de cambio funcional en regiones regulatorias posiblemente asociadas a *NPAS3* entre humanos y chimpancés.

Una hipótesis que surge al analizar toda la evidencia actualmente disponible es que los elementos regulatorios de este gen podrían estar sujetos a un alto recambio dentro de los vertebrados. Este alto recambio involucraría ganancias de múltiples elementos linaje específicos, modificación de elementos preexistentes y, aunque no analizado en este trabajo, posiblemente también pérdida de elementos. Las razones detrás este patrón, sin embargo, son discutibles. Una posible explicación podría estar vinculada con la participación de *NPAS3* dentro de un módulo genético muy estable que participar en los programas del desarrollo de diversas estructuras (Fritzsche et al. 2007b), muchas de las cuales, cambiaron drásticamente dentro de los vertebrados. Si este fuera el caso, es esperable un alto cambio (con ganancias, pérdidas y modificaciones) en las regiones regulatorias de estos genes ya que la reconexión de los genes que participan del desarrollo sería la base genética principal que sustenta la evolución morfológica (Carroll 2008; Vavouri y Lehner 2009; Peter y Davidson 2011). Esta hipótesis sugiere además, que las diferentes regiones regulatorias que controlan la expresión de *NPAS3* estarían vinculando a este gen con procesos que tienen lugar a niveles diversos de la jerarquía de mecanismos que ocurren durante el desarrollo embrionario de los organismos (Erwin y Davidson 2009). Por ejemplo, algunos elementos regulatorios altamente conservados entre los vertebrados podrían ser importantes en procesos compartidos por los integrantes de este grupo. Por otro lado, elementos exclusivos de los roedores o primates (Takahashi y Saitou 2012), podrían ser claves para variaciones de los programas del desarrollo que evolucionaron específicamente en estos linajes. Si bien gran parte de los *enhancers* estudiados en este trabajo tienen dominios de expresión llamativamente similares, no se puede descartar que, en su contexto endógeno, estos elementos desarrollen actividades no equivalentes (como fuera sugerido para los elementos HAR202 y 2xHAR142). Alternativamente, la convergencia en la función de estos *enhancers* puede estar relacionada con una fuerte presión de selección a favor de estabilizar la actividad transcripcional de este gen frente a contextos celulares de alto recambio de las redes regulatorias (como los asociados con procesos macroevolutivos). Ambas hipótesis no son excluyentes.

Finalmente, será un gran desafío comprender el rol de la evolución acelerada sufrida por los HAEs presentes en los intrones de *NPAS3*. En este sentido, entender cuáles son las fuerzas evolutivas que están detrás del comportamiento dinámico de los elementos regulatorios posiblemente asociados a

NPAS3 a lo largo de los diversos linajes de vertebrados será una pieza de conocimiento clave para avanzar en este sentido. Existe evidencia que una variedad de fuerzas pudieron estar involucradas en el cambio molecular rápido de los HAEs presentes en este *locus* a lo largo de nuestro linaje. En algunos casos el cambio en la tasa de evolución molecular probablemente sea consecuencia de procesos de selección natural, en otros, este patrón de evolución es mejor explicado por procesos de evolución neutra (Kostka et al. 2012). Un punto crítico para entender la función que este gen pudo tener en la evolución de algunas de las características distintivas de nuestro cerebro estará, posiblemente, en comprender como los cambios evolutivos que ocurrieron en estos elementos durante la evolución de nuestro linaje se integraron al dinámico y complejo repertorio de secuencias probablemente regulatorias que caracterizan al *locus* de *NPAS3* y fundamentalmente, si esta integración le imprimió propiedades novedosas a la regulación de la expresión de este gen. La abundante literatura actual que asocia diferentes enfermedades psiquiátricas humanas (principalmente esquizofrenia) con variantes genéticas en el *locus* de *NPAS3* es, además, un fuerte indicio que este gen, al menos en los humanos, es importante para el correcto despliegue de capacidades cognitivas complejas (Pickard et al. 2006). Cuanto de ese rol es una novedad evolutiva de los humanos y si tiene un sustento molecular en cambios regulatorios, constituyen preguntas que quedaran aún abiertas.

Finalmente, mediante este trabajo se espera haber agregado un escalón más de conocimiento al entendimiento la evolución de nuestro linaje. Sin embargo la pregunta *¿Qué es lo que nos hace humanos?* seguirá sin resolverse por un largo tiempo y continuará como guía para cuestionar quiénes (o qué) somos como especie y particularmente qué nos distingue del resto de los animales.

MATERIALES Y MÉTODOS

Análisis de agrupamiento de regiones aceleradas en linaje humano

Generación de la base de datos de HAEs

Para el análisis de agrupamiento de regiones aceleradas en el genoma humano se utilizaron un total de 1820 secuencias de evolución rápida en el linaje humano provenientes de cuatro bases de datos públicas: 202 *human accelerated regions* (HARs por sus siglas en inglés), 992 *human accelerated non-coding sequences* (HACNS por sus siglas en inglés), 63 *human terminal branch elements* (HTBE; por sus siglas en inglés) y 563 *29-mammals human accelerated regions* (denominadas 2xHAR). Los intervalos genómicos originalmente anotados como intervalos en el ensamblado del genoma humano versión NCBI35/hg17 (HARs y HACNS) fueron pasados manualmente a intervalos en NCBI36/hg18 utilizando la herramienta BLAT (disponible en www.genome.ucsc.edu). Las 1820 regiones fueron reunidas en una única base de datos donde posteriormente, dos o más elementos solapados fueron convertidos en un único elemento utilizando la herramienta *merge* disponible en el sitio Galaxy (<http://main.g2.bx.psu.edu/>). De esta forma, se eliminaron redundancias y se obtuvieron un total de 1629 elementos de evolución acelerada en humanos (los cuales fueron denominados HAEs por sus siglas en inglés; Human Accelerated Elements) no solapados.

Generación de las bases de datos de intervalos genómicos

Para realizar el análisis de agrupamiento, se generaron cuatro bases de datos de regiones genómicas humanas de 1 Mb (Mega base; 1.000.000) de longitud utilizando una planilla de cálculo. Cada base de datos incluye elementos de 1 Mb que se hallan desaplazados 250 kpb entre sí. En algunos casos, los intervalos iniciales y terminales de cada cromosoma humano fueron dejados con una longitud inferior a la de 1 Mb.

Por otro lado, se obtuvieron los intervalos genómicos de los genes humanos. Para este estudio, se utilizó la base de datos de genes altamente curados RefSeq (base de datos de referencia disponible en www.genome.ucsc.edu). Como esta base de datos posee redundancias (es decir, varios intervalos para un mismo gen), se procedió a solapar los intervalos utilizando la herramienta *merge*, previamente mencionada. De esta forma, dos o más genes o versiones de un gen que comparten un mismo intervalo (o parte de este) se convierten en un único *locus*.

Agrupamiento de HAEs en el genoma humano

Para analizar la distribución de HAEs a lo largo del genoma humano, se utilizaron diversos procedimientos. Por un lado, se encolumnaron uno a uno los intervalos genómicos de 1 Mb de longitud con los HAEs contenidos en dichos intervalos utilizando la herramienta *join* (disponible en <http://main.g2.bx.psu.edu/>). Luego, se contó el número de veces que aparecía una Mb en la tabla (este número representa el número de HAEs que contenía cada región genómica) con la herramienta *count on column* (disponible en <http://main.g2.bx.psu.edu/>) y se construyó un histograma de número de HAEs por Mb. Lo mismo se repitió para cada una de las cuatro ventanas de intervalos genómicos previamente generadas. El mismo procedimiento se realizó para analizar el agrupamiento de HAEs en genes humanos. En este caso, se encolumnaron la base de datos de RefSeq antes mencionada con la base de datos de HAEs.

Finalmente, se verificó la identidad de los 14 HAEs más próximos entre sí de todo el genoma humano utilizando una planilla de cálculo. Se colocaron las coordenadas de inicio y fin de cada HAE respetando su orden espacial en cada uno de los cromosomas. Luego se restó la posición de inicio del HAE 14vo a la posición de finalización del primer HAE. Este número se colocó en una tercera columna. Este procedimiento se repitió para todos los pares de HAEs separados 12 elementos entre sí. Estas distancias se calcularon para los HAEs presentes en cada uno de los 22 cromosomas humanos con HAE y se ordenaron de menor a mayor. De esta forma, se identificó el grupo de 14 HAEs más próximos entre sí de todo el genoma humano.

Análisis de agrupamiento de HAEs en genes gigantes

Se generó una base de datos de genes humanos con una longitud ≥ 800 kpb. Para ello, la base de datos de genes humanos RefSeq no redundantes previamente descrita se convirtió en formato fasta usando las herramientas presentes en el servidor Galaxy. De la lista original únicamente se recuperaron las secuencias de largos iguales o mayores a 800 kpb. Las coordenadas en NCBI36/hg18 de las secuencias obtenidas se obtuvieron usando BLAT, de acuerdo con el servidor genómico de UCSC. De esta forma, fue posible identificar 92 intervalos que contienen genes humanos de más de 800 kpb (denominándose a estos “genes gigantes”). El intervalo más pequeño fue de 800.503 pb, el más largo de 2,304,636 pb y el largo promedio fue de 1,170,256 pb. El agrupamiento de HAEs en genes gigantes se realizó siguiendo la metodología ya descrita para los intervalos genómicos de 1 Mb.

Análisis de elementos conservados *PhastCons* en el genoma humano

Se obtuvieron todos los elementos conservados entre vertebrados del genoma humano de la base de datos de *17-way most conserved elements* (Siepel et al. 2005) disponible en el sitio www.genome.ucsc.edu (NCBI36/hg18). Esta base de datos se construyó utilizando el programa PhastConst como parte del paquete de programas PHAST (*PHylogenetic Analysis with Space/Time models*). Esta base de datos se filtró por el valor de conservación asignado, solamente recuperando aquellos que tuvieran un puntaje mayor a 400. Se seleccionó este puntaje ya que fue el utilizado para la identificación de regiones conservadas en la identificación de elementos acelerados en el genoma humano (Prabhakar et al. 2006). Utilizando las herramientas disponibles en el sitio www.genome.ucsc.edu (*Table Browser*). Adicionalmente, se obtuvo una base de datos de *spliced EST* (Benson et al. 2004), desde el sitio www.genome.ucsc.edu. Con esta base de datos se filtraron todos aquellos intervalos *PhastCons* que solaparan con intervalos asignados a *spliced EST*. Esto se realizó para eliminar los intervalos genómicos conservados del genoma humano asignados a exones. Para hacer esto se usó la herramienta *join* de Galaxy, recuperándose exclusivamente los intervalos de *PhastCons* que no solapa con *spliced EST*. De esta forma se construyeron dos bases de datos: una conteniendo todas las secuencias conservadas entre vertebrados del genoma humano (*PhastCons-todas*) y otra conteniendo las secuencias conservadas no exónicas (*PhastCons-noEST*).

Paralelamente, se generó una base de intervalos genómicos intergénicos sustrayendo al intervalo de los 24 cromosomas humanos la base de datos de RefSeq no redundantes antes mencionada utilizando las herramientas disponibles en Galaxy. Esta base de datos se llamó no RefSeq y contiene 19,921 regiones intergénicas (24 regiones más que la base datos de RefSeq, una por cromosoma). Finalmente, y utilizando la misma metodología aplicada para los agrupamientos de los HAEs, se cruzaron las bases de datos de RefSeq, no RefSeq y RefSeq gigantes con las bases de datos de *PhastConst-todas* y *PhastCons-noEST*. Se recuperaron cinco bases de datos: *PhastConst* en regiones intergénicas, *PhastCons-todos* en regiones RefSeq, *PhastCons-todos* en regiones RefSeq gigantes, *PhastConst-noEST* en regiones RefSeq y *PhastConst-noEST* en regiones RefSeq gigantes. El número de pares de bases nucleotídicas y el número de elementos en cada base de datos se obtuvo desde el sitio www.genome.ucsc.edu. Para ello, cada una de las cinco bases de datos se subieron al sitio como *custom tracks* y se obtuvo la información desde la sección *table browser* (herramienta *statistics*). Siguiendo estos procedimientos, también se obtuvo el número de HAEs y el número de intervalos conteniendo HAEs para cada una de las bases de datos (RefSeq, no RefSeq y RefSeq gigantes).

Análisis de la evolución de la regiones no codificantes conservadas del gen *NPAS3*

Las regiones no codificantes y conservadas del gen *NPAS3* fueron identificadas mediante el alineamiento de múltiples especies utilizando el programa PhastCons. Para ello, se usó la aplicación *17 way_most_conserved_multi-species_genomic_alignments* disponible en el sitio www.genome.ucsc.edu. Únicamente se utilizaron aquellas regiones incluidas dentro del intervalo del gen *NPAS3* y con un puntaje mayor a 400. Posteriormente, se filtraron los elementos que solapaban con EST procesados, regiones repetitivas y bloques duplicados aplicando los siguientes filtros: *spliced ESTs*, *repeat masker* y *self-chain*. Finalmente, se calculó el número de pares de bases presentes en las regiones conservadas recuperadas, dando un total de 101.140 bp.

Análisis de solapamiento de las cuatro bases de datos de HAEs

El grado de coincidencia entre bases de datos de HAEs se realizó encolumnando de a pares, tríos o cuartetos, los intervalos genómicos de HARs, HACNS, HTBE y 2xHARs en todas sus combinaciones posibles utilizando la herramienta *join* previamente introducida. De esta forma, los intervalos genómicos compartidos (al menos en parte) por dos o más elementos quedaron agrupados en filas. Luego, manualmente se procedió a contar el número de elementos compartidos entre dos o más bases de datos. Los resultados obtenidos se volcaron en un diagrama de Venn.

Producción de peces cebra transgénicos

Construcción de los transgenes para pez cebra

Se clonaron a partir de ADN genómico humano y de genómico de chimpancé las regiones conservadas *Phast Cons* que contienen los HAEs de *NPAS3* (y regiones ortólogas) en el vector de entrada pENTR/D (Invitrogen). Las secuencias de los primers usados fueron las que se detallan a continuación. Al extremo 5' de los *primers forward* se les agregó la secuencia CACC que permite el clonado direccionado de los insertos en el vector pENTR/D

TgHAR21Hs_primer_F	5'- <u>CACCT</u> CATCTCAAGGGCTTGCTTT-3'
TgHAR21Hs_primer_R	5'-TGCTTGTGTACCACATGCTTT-3'
TgHAR21Pt_primer_F	5'- <u>CACCA</u> ATCATCTCAAGGGCTTGCTTT-3'
TgHAR21Pt_primer_R	5'-TGTGCTTGTGTACCACATGCTTT-3'
TgHAR89Hs_primer_F	5'- <u>CACC</u> ACTGTGAGACCCACCAAAGG-3'

TgHAR89Hs_primer_R	5'-GGTTTGGTTTGATCGAATGG-3'
TgHAR89Pt_primer_F	5'- <u>CACC</u> CTACTGTGAGAACCACCAAAGG-3'
TgHAR89Pt_primer_R	5'-GGTTTGGTTTGATCAAATGG-3'
TgHAR96Hs_608_primer_F	5'- <u>CACC</u> TTGTGTATGTCTATCTAAAATGGAAAA-3'
TgHAR96Hs_608_primer_R	5'-AAGAAGAAACAGCATTCAAGAAA-3'
TgHAR96Pt_608_primer_F	5'- <u>CACC</u> TTGTGTATGTCTATCTAAAATGGAAAA-3'
TgHAR96Pt_608_primer_R	5'-AAGAAGAAACAGCATTCAAGAAA-3'
TgHAR173Hs_primer_F	5'- <u>CACC</u> GCTGGCTTTATATTTTTGAGGTATTC-3'
TgHAR173Hs_primer_R	5'-TTCCGTAATACCTTCTGAGTGTT-3'
TgHAR202Hs_primer_F:	5'- <u>CACC</u> CACCAGATTCCCTTTGCATT-3'
TgHAR202Hs_primer_R:	5'-CAAACAGCAGGAGACGCTAA-3'
TgHAR202Pt_primer_F:	5'- <u>CACC</u> ACTACTGACTGTAATTTACTTT-3'
TgHAR202Pt_primer_R:	5'-CAAACAGCAGGAGACGCTAA-3'
TgHACNS96Hs_415_primer_R:	5'-ATTTCCCTCTGCCTGGTCTT-3'
TgHACNS96HPt_415_primer_F	5'- <u>CACC</u> CGTTTCCCCTTTGGTACA-3'
TgHACNS96Pt_415_primer_R:	5'-ATTTCCCTCTGCCTGGTCTT-3'
TgHACNS221Hs_primer_F	5'- <u>CACC</u> GCACATGAATCAGTCTTTGTCA-3'
TgHACNS221Hs_primer_R	5'-GCAGCTATTTCTGATGCTGAG-3'
TgHACNS658Hs_primer_F	5'- <u>CACC</u> AATCAGCAGCCATCAGTTTA-3'
TgHACNS658Hs_primer_R	5'-TCCCCGCTCCCTTTAGATG-3'
TgHACNS489Hs_primer_F	5'- <u>CACC</u> GGAAGTGTGTTTGATTCGTACC-3'
TgHACNS489Hs_primer_R	5'-GGCAAATCAAAAAGTTAATTTCA-3'
TgHACNS553Hs_primer_F	5'- <u>CACC</u> TAGCCTCTAGCCTGGACCAC-3'
TgHACNS553Hs_primer_R	5'-GAAAAGGCTGGAAAATCAGG-3'
TgHACNS658Hs_primer_F	5'- <u>CACC</u> CGTCTCCCCTTCGGTACA-3'
TgHACNS658Hs_primer_R	5'-ATTTCCCTCTGCCTGGTCTT-3'

Tg2xHAR122Hs_primer_F	5'- <u>CACCT</u> TGGATGTGGCACTTTCAGAG-3'
Tg2xHAR122Hs_primer_R	5'-TATTACTTGGCTGCGGATCG-3'
Tg2xHAR142Hs_primer_F	5'- <u>CACCT</u> TGCAGAAATCAATCTCTACTGG-3'
Tg2xHAR142Hs_primer_R	5'-CCCCTAATTAAATGCACAACCTGA-3'
Tg2xHAR157Hs_primer_F	5'- <u>CACCGGC</u> CAGAAACCCTAGTTTCTCAA-3'
Tg2xHAR157Hs_primer_R	5'-CCTCCACATCCCACAATTTC-3'
Tg2xHAR223Hs_primer_F	5'- <u>CACCAGCCC</u> AGGACTTCCAAGAGT-3'
Tg2xHAR223Hs_primer_R	5'-AACCCCTTCAGTTTTGCCTTA-3'

Los fragmentos fueron amplificados en un termociclador (Biorad) usando la polimerasa con actividad correctora de errores Pfx (Invitrogen) junto con la polimerasa altamente procesiva Taq (producida en el laboratorio) en una relación 1:1 y 0,2 mM de MgSO₄. El programa usado fue el siguiente:

- 1) Desnaturalización inicial: 95 °C por 5 minutos.
- 2) Desnaturalización: 95 °C por 30 segundos.
- 3) *Annealing*: 60 °C por 30 segundos.
- 4) Extensión: 68 °C por 30 segundos a 1 minuto 30 segundos según el tamaño del fragmento (considerando una actividad polimerasa de 1000 bases por minuto).
- 5) 29 ciclos de los pasos 2 a 4.
- 6) Extensión final: 10 minutos a 68 °C.

Los fragmentos deseados (de largos comprendidos entre ~ 200 a 1100 pares de bases) fueron aislados en gel de agarosa. Se recuperó el ADN purificando las bandas con el kit para extracción de gel QIAquick Gel Extraction Kit (Qiagen). Los insertos resultantes se clonaron direccionadamente por medio del sistema TOPO (Invitrogen) en el vector pENTR/D, que a su vez funciona como vector de entrada para el sistema Gateway (Invitrogen). El sistema TOPO utiliza las enzimas topoisomerasas de *Vaccinia* para recombinar el ADN doble cadena del producto de PCR a los extremos libres del vector pENTR/D. La direccionalidad del clonado está garantizada por la complementariedad de bases del extremo 3' GTGG del vector con la secuencia CACC presente en el inserto (figura 32).

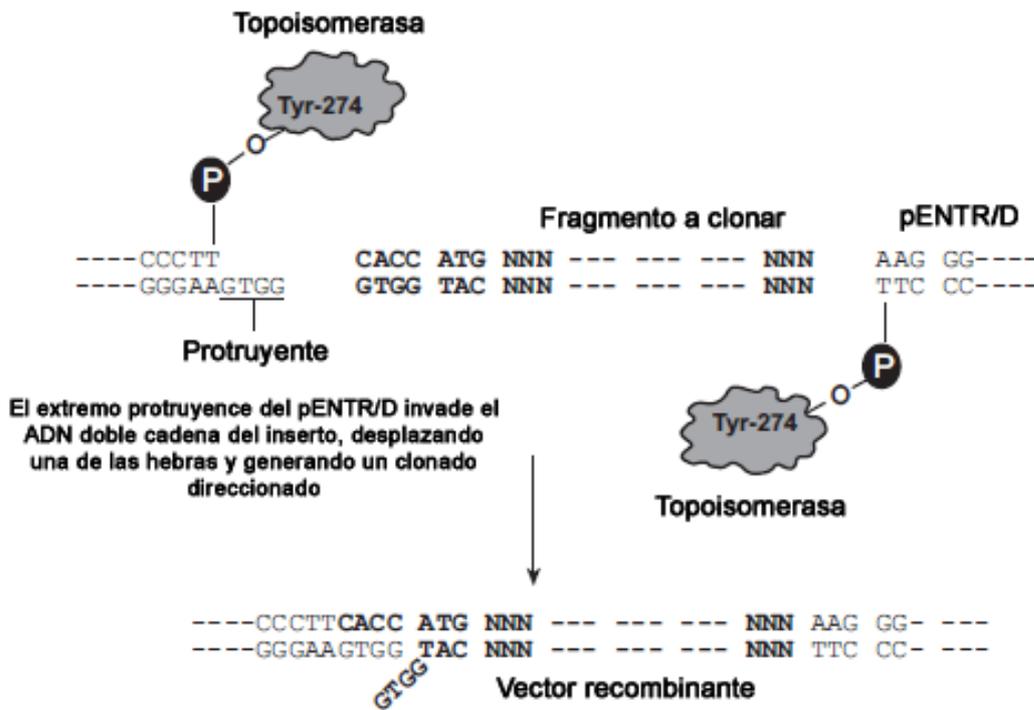


Figura 32: Esquema clonado en el vector pENTR/D por medio del sistema TOPO (Invitrogen). Modificado del manual del uso del vector pENTR/D.

Una vez secuenciados e identificados los clones que contenían los insertos deseados, se realizó el subclonado hacia el vector destino pXIG_cFosGW por medio del sistema Gateway (Invitrogen). Brevemente, el sistema Gateway utiliza las enzimas recombinantes integrasa y excisionasa proveniente del bacteriófago lambda y la proteína factor de integración de hospedador de *E. coli*. Esta mezcla de enzimas, conocidas como clonatas, catalizan *in vitro* la recombinación entre el inserto presente en el vector de entrada (pENTR/D en nuestro caso) y el vector destino. La recombinación ocurre de forma direccionada entre los sitios attL presentes en el vector de entrada (flanqueando el inserto) y las secuencias attR del vector destino.

En este sistema, las bacterias que incorporan vectores destino no recombinantes no crecen debido a que en estas, el vector destino no tiene interrumpida la expresión del gen letal *ccdB*. Así mismo, tampoco crecen las bacterias que incorporan el vector entrada debido a que la resistencia de este vector es kanamicina y la selección se realiza en LB sólido con ampicilina (figura 33). Todos los clonados se realizaron utilizando las cepas de *E. coli* TOP10 (Invitrogen). Esta cepa es susceptible a la proteína letal que interfiere con la girasa de *E. coli* codificada por el locus *ccdB*. Por el contrario, la amplificación del vector pXIG_cFosGW se realizó en la cepa DB3.1 que es resistente al producto del gen *ccdB*.

El vector destino pXIG_cFosGW contiene el promotor mínimo del gen *cFos* de ratón (sin actividad en ausencia de región activadora de la transcripción) río arriba de la secuencia codificante de la proteína reportera fluorescente *Enhanced Green Fluorescent Protein* (EGFP por sus siglas en inglés) (Fisher et al. 2006b). Además, flanqueando el sitio de clonado Gateway, el promotor mínimo y el gen reportero, este vector contiene las secuencias de reconocimiento del transposon *Tol2* de medaka (*Orzys latipes*). Estas secuencias son necesarias para el movimiento del transgén y la generación de peces transgénicos (Kawakami et al. 2004).

Se obtuvieron los transgenes para inyección por medio de mini preparaciones de ADN plasmídico de los clones HAE_pXIG_cFosGW deseados. Para ello se utilizó el sistema de purificación de ADN por columnas *Wizard Plus SV Minipreps DNA Purification System* (Promega).

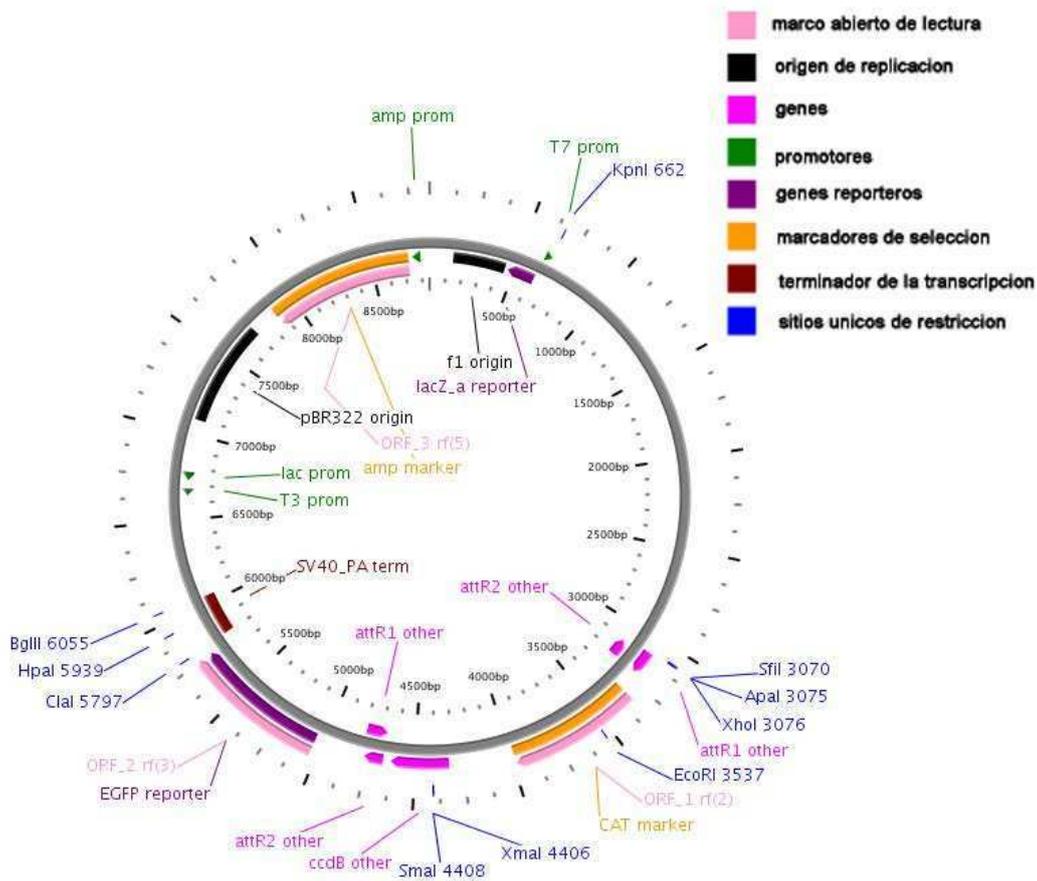


Figura 33: Mapa del vector pXIG_cFosGW (Fisher et al. 2006b).

Transcripción *in vitro* del ARN que codifica para la transposasa *Tol2*

Se preparó una mini preparación de ADN plasmídico del vector pCS-Tp con el *HiSpeed Plasmid Purification kit* (Qiagen). Este vector contiene la secuencia codificante para la transposasa de medaka *Tol2*. Se linealizaron 20 µg de plásmido con la enzima de restricción *NotI* (Invitrogen) en

un volumen final de 200 μl . Una vez verificada la linealización completa del vector, se degradó la enzima de restricción con proteinasa K (100 $\mu\text{g ml}^{-1}$.) Se purificó la preparación de ADN por dos extracciones orgánicas; la primera con 1 volumen de fenol:cloroformo:alcohol isoamílico (25:24:1), la segunda con 1 volumen cloroformo. La fase acuosa resultante se precipitó con acetato de sodio (0.3M) e isopropanol (1 volumen.) El precipitado se lavó con etanol 70% y finalmente se resuspendió en agua libre de nucleasas.

La transcripción *in vitro* se realizó con la ARN polimerasa de fago *sp6* (Roche). Una vez finalizada la reacción de síntesis, el ADN molde se degradó con *DNAseI* y se precipitó el ARN con acetato de litio (0,4 M) y 2 volúmenes de etanol frío. Se resuspendió el ARN en agua libre de nucleasas, se cuantificó y se llevó a una concentración de 175 $\text{ng } \mu\text{l}^{-1}$.

Animales y acuario

Los peces usados en este estudio pertenecen a cepa AB (Zebrafish International Resource Center, Universidad de Oregon, Oregon, USA) y a la cepa exocriada obtenida del laboratorio de la Dra. Nora Calcaterra (IBR, Rosario). Los animales se criaron en condiciones estándar en nuestro 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 diferentes tratamientos: aireación, desgaseado, filtración mecánica (*filter pad*), filtración biológica de amoníaco y nitritos (*Siporax* y *Kaldnes*), control de temperatura, segunda filtración mecánica (50 μm clartridge), filtración química por medio de carbón activado y esterilización del agua por UV. Sintéticamente, los animales se mantuvieron en peceras de policarbonato de 3 litros de capacidad, a una densidad máxima de 6 peces adultos por tanque con un flujo de agua de 9 litros por hora (6 recambios del agua de los tanques por hora). Entre otras condiciones, los animales se criaron en un ciclo regular de 14 horas de luz y 10 horas de obscuridad, a $28\pm 1^\circ\text{C}$, $7\pm 1,5$ mg/L de densidad de oxígeno, 600 ± 100 μS de conductividad del agua y $7\pm 1,5$ de pH. Los peces adultos recibieron una dieta de alimento balanceado comercial (Zeigler) dos veces al día suplementada con artemias (INVE).

Obtención de embriones para inyección

El día previo a las inyecciones, se armaron los apareos. Para ello, se usaron unas peceras más pequeñas que las usadas para cría donde hay dos compartimientos separados por una placa divisora de plástico transparente. La base de estas peceras tiene una superficie con ranuras que permite separar los huevos fertilizados de los peces progenitores. Se colocó un macho y una hembra en cada una de las peceras de apareo, dejándolos separados toda la noche por medio de la placa divisoria.

El día de la inyección, poco después del comienzo del ciclo de luz, se removió la división que separa machos y hembras en las peceras de apareo; esto pone en contacto a los animales. En general dentro de los primeros 30 minutos de interacción, comenzaron las primeras puestas de huevos. Las mismas estuvieron compuestas por aproximadamente 150 a 200 embriones cada una.

Inyección de los embriones con la transposasa

Los embriones obtenidos fueron colectados en una placa de Petri de 10 mm y se mantuvieron a 22 °C en medio de embriones (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄, 0.00001% peso/volumen de azul de metileno) hasta el momento de inyección.

Se preparó una mezcla fresca de inyección conteniendo:

ADN plasmídico del transgén HAE_ <i>cFos</i> EGFP	125 ng
ARN transposón Tol2 (175 ng μl^{-1})	1 μl
KCl 1 M	0,5 μl
Rojo fenol (2% en agua)	0,5 μl
Agua destilada c.s.p.	5 μl

Los peces transgénicos se generaron microinyectando 1 nl de la mezcla de inyección en el vitelo de embriones en estadio de 1-2 células. Se utilizaron pipetas preparadas a partir de capilares de borosilicato y estiradas con un estirador vertical de micropipetas (Vertical Pipette Puller, Model 720, KOPF, David Kopf Instruments, Tujunga, California, USA). Para esto, 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. Se absorbió el líquido excedente para dejar adheridos los embriones al vidrio del portaobjetos. Las inyecciones se generaron con un microinyector Narishige-IM300 Microinjector (Narishige) conectado a un tanque de N₂ y utilizando un micromanipulador para manejar la pipeta de inyección bajo una lupa Leica wild M3C (Leica).

Luego de la inyección, se identifican los embriones inyectados por la marca roja dejada por el rojo fenol en el vitelo y se retiraron los embriones dañados y/o no inyectados. Se colocaron 100 embriones inyectados por placa de Petri de 10 mm y se los dejó crecer a 28 °C en medio de embriones hasta la edad deseada en una incubadora.

Se realizaron al menos dos inyecciones independientes por transgén y por edad. En cada oportunidad, se inyectaron al menos 150-200 embriones. También se dejaron embriones sin inyectar para controlar la viabilidad de los embriones. En aquellos casos donde se analizaron embriones a

48 hpf, a las 24 hpf se agregó al medio de embriones 0.1mM de 1-fenil-2-tiourea (Sigma) para prevenir la pigmentación de los animales. Más del 80% de los embriones inyectados sobrevivió y se desarrolló correctamente.

Análisis del patrón de expresión de la proteína reportera EGFP

Se utilizó la metodología propuesta por el proyecto CONDOR (Woolfe et al. 2007) para estudiar *in vivo* la actividad *enhancer* de los HAEs. Brevemente, se analizó la expresión de EGFP a 24-30 y 48-54 hpf en embriones G0 decorionizados. Para ello, previamente se anestesiaron los embriones con 10 gotas de una solución 0,4% de tricaina (Sigma) por placa de 10 mm. Se montaron los animales en metil celulosa 3% p/v en agua (Sigma) y se registró la expresión de la proteína reportera EGFP usando un microscopio OLYMPUS BX51 acoplado a una cámara digital OLYMPUS DP72 y los filtros adecuados.

Se generaron histogramas de expresión por edad registrando el porcentaje de embriones positivos para la expresión de EGFP en cada una de las siguientes regiones anatómicas (como fueran propuestas en CONDOR): telencéfalo, mesencéfalo, rombencéfalo, médula espinal, ojo, notocorda, corazón, arcos faríngeos/branquiales-faringe, músculo somítico y aletas. La expresión de EGFP por fuera de estos dominios se registró como “otra”. Embriones con expresión de EGFP en 6 o más dominios simultáneamente fueron anotados como con “expresión ubicua”. Se consideró *enhancer* a todo HAE que produjera un patrón de expresión consistente en, al menos, 20% de los embriones analizados (Fisher et al. 2006b). Para hallar el patrón de expresión consistente primero, se consideraron todos aquellos dominios de expresión presentes, en al menos, un 30% de los embriones analizados. Se seleccionó este porcentaje de manera conservadora porque se encontró que el vector vacío (control) produce un 30% de embriones positivos (aunque no se encontró ningún dominio anatómico que estuviera presente en más del 20% de los animales registrados). El dominio de expresión consistente se consideró como aquel conjunto de dominios anatómicos más inclusivo que estuviera simultáneamente presente en, al menos, el 20% de los embriones totales analizados. Este patrón de expresión se consideró representativo de la expresión no mosaico.

Para aquellos transgenes que no mostraron expresión de EGFP apreciable (o niveles comparable a la expresión basal observada para el vector pXIG_cFosGW vacío), se incrementó la concentración de ADN en la mezcla de inyección hasta observar niveles elevados de toxicidad en los embriones (nunca superándose los 40 ng μl^{-1} de concentración final en la solución de microinyección) y se repitió el análisis detallado anteriormente a 24-30 hpf.

Generación de líneas de peces cebra transgénicos

Para la generación de líneas de peces cebra transgénicos, se inyectaron entre 200 a 600 embriones con el transgén de interés como se explicó anteriormente. Los embriones se mantuvieron en placas de petri con medio de embriones en una incubadora a 28 °C. Al día 5, los embriones fueron trasladados a peceras de cría y mantenidos sin flujo de agua. Se los alimentó dos a tres veces por día con una dieta de larvas especialmente formulada para estadios larvales y juveniles (Larval AP-100, Ziegler, USA). Al décimo día se conectó la pecera de cría al sistema de recirculado de agua bajo un régimen de flujo de agua lento y únicamente durante la noche. Tres días después se suplementó la alimentación de las crías de pez cebra con artemias frescas (INVE). Pasados los quince días, 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 dos meses, los animales entraron en régimen de alimentación de adultos.

Una vez alcanzada la madurez sexual (3 a 4 meses de edad), los animales fueron separados por sexo. Para analizar la expresión de proteína reportera EGFP, se cruzaron los animales adultos inyectados con animales de la línea AB (*wild type*). Los embriones obtenidos fueron analizados a diferentes edades como se explicó anteriormente. Se generaron al menos dos líneas de expresión consistente por transgén.

Producción de ratones transgénicos

Construcción de transgenes para ratón

Se clonaron a partir de ADN genómico humano, de ratón o de chimpancé las regiones ortólogas que contienen el HAE de *NPAS3* 2xHAR142. Las secuencias de los primers usados fueron las siguientes:

PC_2xHAR142Hs

Tg2xHAR142_Hs_F: 5' **CACC**CAACCAACACCAGCTAGAAAG-3'
 Tg2xHAR142_Hs_R: 5'-CAA**GAATT**CAATTAAATGCACAACACTGAACAGG

PC_2xHAR142Pt

Tg2xHAR142_Hs_F: 5' **CACC**CAACCAACACCAGCTAGAAAG-3'
 Tg2xHAR142_Hs_R: 5'-CAA**GAATT**CAATTAAATGCACAACACTGAACAGG

PC_2xHAR142Mm

Tg2xHAR142_Mm_F 5'-AC**GTCGAC**CTTTTGAGCCAGGTCCACTC
 Tg2xHAR142_Mm_R 5'-AC**CCCGGG**ACGAGGGGGAAAAGATCACT

En azul se muestran las secuencias de corte para la enzima de restricción EcoRV, en verde para la enzima Sal I, en rojo para la enzima Xma I y en amarillo las secuencias para clonado direccionado en pENTR/D.

Los fragmentos fueron amplificados en una termocicladora (Biorad) usando la polimerasa con actividad correctora de errores pfx (Invitrogen) junto con la polimerasa altamente procesiva taq (producida en el laboratorio) en una relación 1:1. El programa usado fue el siguiente:

- 1) Desnaturalización inicial: 95 °C por 5 minutos.
- 2) Desnaturalización: 95 °C por 30 segundos.
- 3) *Annealing*: 60 °C por 30 segundos.
- 4) Extensión: 68 °C por 30 segundos a 2 minutos según el tamaño del fragmento (considerando una actividad polimerasa de 1000 bases por minuto).
- 5) 29 ciclos de los pasos 2 a 4.
- 6) Extensión final: 10 minutos a 68 °C.

Los fragmentos deseados (de largos comprendidos entre ~ 500 a 2000 pares de bases) fueron aislados en gel de agarosa. Se recuperó el ADN purificando las bandas con el kit para extracción de gel QIAquick Gel Extraction Kit (Qiagen). Los insertos de ratón se clonaron con el sistema Zero BackgroundTM/Kan Cloning Kit (Invitrogen) que disminuye el número de colonias no recombinantes (Bernard et al. 1994). El vector pZErOTM-2 contiene el gen letal ccdB fusionado al extremo C-terminal de LacZ α -ccdB. La inserción de un fragmento de ADN en el sitio de clonado de este vector interrumpe la expresión de la proteína letal ccdB permitiendo a las bacterias portadoras de vectores recombinantes sobrevivir. Por el contrario, las células que contienen vectores no recombinantes, no pueden crecer. En este sistema es importante usar la cepa de *E. coli* TOP10 ya que esta no contiene el gen que codifica para el factor lacI^q y que funciona como inhibidor en *trans* del gen ccdB. Finalmente, las secuencias de humano y chimpancé fueron clonadas en el vector de entrada pENTR/D (Invitrogen) siguiendo la metodología detallada para el clonado de los transgenes de pez cebra. Se confirmó por secuenciación la identidad y calidad de todos los insertos.

Paso siguiente, los plásmidos pZeRO-2 conteniendo las secuencias de interés fueron digeridos con las enzimas Xma I y Sal I (sobre las secuencias de los primers se indican los sitios de corte para enzimas de restricción utilizadas). Se realizaron subclonados al vector final HSP68/LacZ. Este vector deriva del plásmido pBluescript y contiene el sitio múltiple de clonado río arriba del

promotor mínimo del gen HSP68 de ratón regulando la expresión de la proteína reportera LacZ. Se confirmó por restricción y secuenciación la orientación de los clones elegidos. En el caso de las secuencias de humano y chimpancé, estas fueron subclonadas direccionadamente al vector pHSP68/LacZGW (Poulin et al. 2005) por medio del sistema Gateway como se indicó anteriormente. Tanto la secuencias del elemento 2xHAR142 de chimpancé como la humana quedaron clonadas en el sitio EcoRV del vector HSP68/LacZ.

Preparación de los transgenes para microinyección pronuclear

Los transgenes para microinyecciones se obtuvieron digiriendo 20 µg de ADN plasmídico con la enzima Sal I. Esta enzima tiene dos sitios de corte en el plásmido pBluescript_HSP68 que aíslan el transgen del vector pBluescript. Se aisló la secuencia del transgén en gel de agarosa y se purificó el ADN por columnas Wizard Plus SV Minipreps DNA Purification System (Promega). Se realizó una segunda purificación usando columnas de intercambio iónico Elutip-D (Schleicher & Schuell) Luego, se recuperó el ADN precipitando con 10% v/v de NaAcO 3M, pH 5,2 y 2,5 volúmenes de etanol 100%. Se realizaron dos lavado con etanol 70% y se resuspendió en buffer de microinyección (Tris-HCl 5mM, pH 7,4; EDTA 0,1mM). La concentración final de ADN de transgén en la solución de microinyección fue de 500 moléculas/pl.

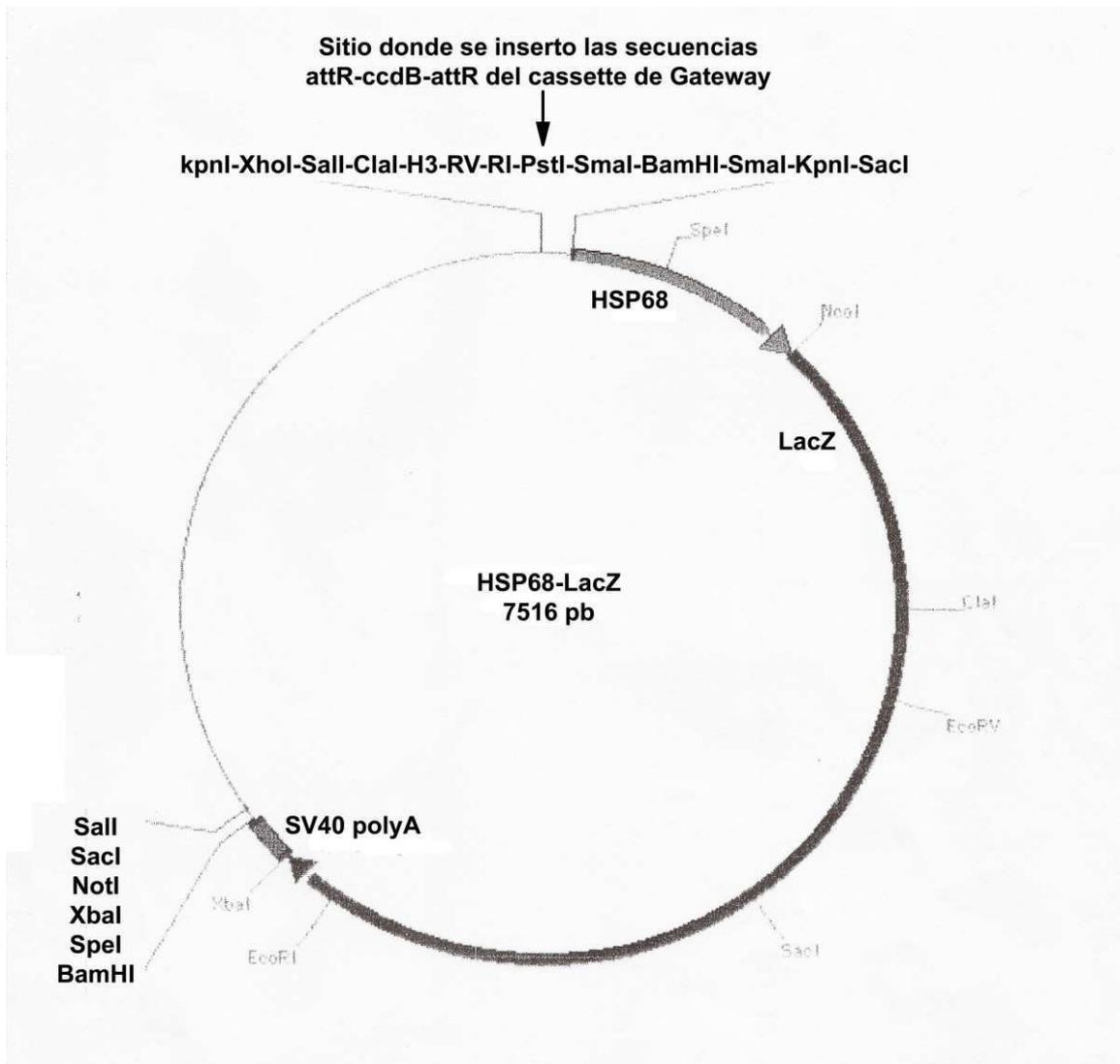


Figura 34: Mapa del vector pHSP68/LacZGW (Poulin et al. 2005).

Animales y bioterio

Los ratones se alojan en el bioterio del Instituto de Biología y Medicina Experimental (IBYME-CONICET.) Estos se mantienen en jaulas independientes que cuentan 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 de bioterio. Los animales están bajo un régimen de luz/oscuridad de 12 horas, cuentan con una dieta balanceada comercial y la temperatura de la sala donde se encuentran esta regulada (20-22 °C).

Producción de embriones para microinyección

Para obtener embriones aptos para microinyección se utilizan animales de la primera generación de la cepa híbrida B6CB. Esta cepa se obtiene por cruce de hembras C57BI/ con machos CBA/J. Hembras B6CBF1 prepúberes de 4 y 5 semanas de edad son súper ovuladas mediante un tratamiento hormonal. Este tratamiento consiste en una primera inyección de 5UI i.p. de PMS

(*pregnant mare serum*, 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 hGC (*human corionic gonadotrophin*, Sigma.)

Inmediatamente luego de la última inyección, las hembras súper ovuladas se ponen en jaulas de apareo con machos B6CBF1. A la mañana siguiente, las hembras copuladas por el macho (se verifica mediante la presencia de un tapón vaginal de semen) 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 (Sigma, 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 UI/ml, sulfato de estreptomycin 50 ug/ml, rojo fenol 0,001%, HEPES 25 mM y pH 7,4). Mediante un par de agujas de 30 G y bajo una lupa, se desgarraron los oviductos para liberar los ovocitos fecundados. Los mismos fueron aspirados con una pipeta Pasteur previamente estirada sobre llama. Los ovocitos junto con las células de granulosa que los rodean fueron colocados en una gota de 90 ul de medio M2 con 1% de hialuronidasa (Sigma). Una vez disgregadas las células de granulosa, se lavaron los embriones mediante 5 - 6 pasajes en medio M2. Finalmente, los embriones listos para la microinyección se mantuvieron en grupos de 15 a 20 en medio M16 (Sigma, 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 UI/ml, sulfato de estreptomycin 50 µg/ml, rojo fenol 0,001%, HEPES 25 mM y pH 7,4) en un incubador a 37 °C y 5% de CO₂.

Microinyección pronuclear

Los ratones transgénicos fueron generados por microinyección pronuclear de cigotas B6CBF2 utilizando micropipetas sujetadoras y de microinyección previamente preparadas a partir de capilares de borosilicato y estiradas con un estirador vertical de micropipetas (Vertical Pipette Puller, Model 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 el ADN del transgén de interés en una concentración de 500 moléculas de ADN por pl. Luego de la microinyección, los embriones se mantuvieron en estufa a 37 °C y 5% de CO₂ hasta el momento de la transferencia.

Transferencia de los embriones inyectados a hembras pseudo preñadas

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; que fueron aquellas copuladas por los machos estériles (se verifica nuevamente mediante la presencia de un tapón vaginal de semen). 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 fueron anestesiadas con avertina (300 mg/kg i.p.). Por medio de una incisión lateral, se expuso el oviducto. Luego de desgarrar la membrana que recubre el ovario y el oviducto, se transfirieron 20 a 30 embriones microinyectados. Para ello se introdujo en el ampulla de la hembra una pipeta Pasteur de vidrio adecuadamente estirada a llama cargada con los embriones en medio M2. Finalizada la maniobra se cerró la incisión con sutura.

Genotipificación de los ratones transgénicos

Extracción de ADN genómico

Los ratones nacidos de las hembras transferidas fueron apartados de las madres sustitutas a los 20-21 días y separados por sexo. El ADN genómico se extrajo de una muestra de tejido de la oreja. Este material se obtuvo mediante un sacabocado y fue el tejido de descarte que quedó luego del proceso de marcación numérica de los animales para su identificación en el bioterio.

Las muestras de oreja fueron incubadas 12-14 hs en agitación a 55 °C en 127 µl de buffer de digestión (Tris-HCl 50 mM; EDTA 100 mM; SDS 0,5%; 0,5 mg/ml de proteinasa K, pH 8). Luego de la digestión se agregó 40 µl de buffer de precipitación de proteínas (acetato de potasio 3M; pH 5,5). Se centrifugó a máxima velocidad por 10 minutos para eliminar restos de pelos, proteínas y SDS. Se precipitó el ADN con 1 volumen de 2-propanol dejando 2 horas a -20 °C. Luego de centrifugar 10 minutos a máxima velocidad, se lavó el precipitado con 200 µl de etanol 70%. Una vez evaporado el alcohol, se resuspendió el ADN en 20 µl de agua destilada. La cuantificación del ADN se realizó con NanoDrop. La presencia del transgén de interés fue detectada por medio de una PCR con oligonucleótidos específicos para la secuencia del transgén. Las secuencias de los primers fueron las siguientes:

TG PC_2xHAR142Mm_F: 5'-TTCGGAATAAATGGGCTCCT-3'

TG PC_2xHAR142Hs/Pt_HSP_GW_F: 5'- TCCCTGTTCAGTTGTGCATT -3'

LacZ_R1: 5'-AAGGGGGATGTGCTGCAAGGCG-3'

Para detectar la presencia del transgén PC_2xHAR142Mm se utilizaron los primers PC_2xHAR142Mm_F y LacZ_R1. Para detectar los transgenes PC_2xHAR142Pt PC_2xHAR142Hs y PC_2xHAR142Pt se utilizaron los primers PC_2xHAR142Hs/PtGW_F y LacZ_R1. En todos los casos, se utilizó una polimerasa de ADN taq de producción casera. El programa de amplificación se detalla a continuación:

- 1) Desnaturalización inicial: 95 °C por 5 minutos.
- 2) Desnaturalización: 95 °C por 30 segundos.
- 3) *Annealing*: 60 °C por 30 segundos.
- 4) Extensión: 72 °C por 1 minuto 30 segundos.
- 5) 35 ciclos de los pasos 2 a 4.
- 6) Extensión final: 10 minutos a 72°C.

Productos de amplificación tuvieron un largo de aprox. 1000 pb y se corrieron en un gel de agarosa 1,5% con bromuro de etidio (0,1 ug/ul de agarosa) y se visualizaron usando luz UV.

El numero de líneas independientes de ratones transgénicos por transgén fue la siguiente:

Transgén	# de líneas
PC_2xHAR142Mm	3
PC_2xHAR142Pt	3
PC_2xHAR142Hs	6

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).

Análisis de la expresión de LacZ en ratones transgénicos

Se analizaron embriones de ratón a diferentes edades del desarrollo embrionario (10.5 a 14.5 dpc). Para ello, se armaron apareos entre un macho adulto positivo para el transgén y dos hembras C57. A la mañana siguiente, las hembras copuladas por el macho (se verifica mediante la presencia de un tapón vaginal de semen) se separaron del macho y se reservaron en una jaula individual hasta el día

deseado de sacrificio (dependiendo de la edad de los embriones a analizar). El día de la recolección de los embriones, las hembras adultas fueron sacrificadas por dislocación cervical. Por medio de una incisión ventral, se aislaron los úteros y se colectaron los embriones. Los embriones fueron lavados en PBS y fijados en una solución de paraformaldehído 4% en PBS a TA y agitación suave. Los tiempos de fijación dependieron de la edad de los embriones: 30 minutos para los embriones 10.5-11.5 dpc, 1 hora para los embriones 12,5 dpc y 1 hora 30 minutos para los embriones 14.5 dpc. Luego de la fijación, se realizaron tres lavados de 30 minutos cada uno en *rinse buffer* (fosfato de sodio 100 mM, MgCl₂ 2 mM, deoxicolato de sodio 0.01%, NP-40 0.02%, pH 7.3). Finalmente, los embriones se dejaron 12-14 horas a 37°C en solución de coloración (*rinse buffer* + 1 mg/ml de 5-bromo-4-cloro-3-indolil-β-D-acido glucuronico (*X-Gal*), 2.12 mg/ml de ferrocianida de potasio y 1.64 mg/ml de ferricianida de potasio). Al día siguiente, se lavaron los embriones en PBS por 10 minutos. Los embriones fueron clarificados en glicerol 20% en PBS por 12-14 horas a 37 °C y fotografiados en una lupa Leica DFC320. De ser necesario, se realizó un segundo paso de clarificación en glicerol 50% en agua por 12-14 horas a 37 °C.

Para analizar la expresión de *LacZ* en cortes histológicos, los embriones previamente revelados fueron lavados en PBS por 12-14 horas y crioprottegidos en sacarosa 10% en PBS por 12-14 horas. El tejido, incluido en un medio a base de gelatina (gelatina 10%, sacarosa 10% en PBS), se congeló en 2-metilbutano (Sigma-Adrich) a -50 °C y se guardó a -70 °C hasta el momento de ser cortado. Se realizaron cortes en un crióstato Leica CM1850 dependiendo de la edad del embrión a 18-25 μm de grosor. Teniendo el tejido sobre portaobjetos Superfrost Plus (Fisherbrand), se realizó una contratinción en eosina Y 0,1% en agua (Sigma-Ardrich) por 2 horas. El tejido fue deshidratado pasándolo por un gradiente creciente de etanol en agua (70%, 80%, 95% y 100%) y clarificado con xileno. Los cortes fueron montados en DPX Mountant (Fluka BioChemika). Las imágenes fueron adquiridas usando una lupa Leica DFC320 y un microscopio Olympus BX51 acoplado a una cámara digital Olympus DP72.

Hibridación *In Situ* en cortes de crióstato

Extracción de ARN de embriones de pez cebra o cerebro de ratón

Se eutanasiaron 50 embriones de pez zebra de 24 o 48 hpf en agua helada. Los embriones fueron homogeneizados en 0,5 ml de Trizol (Invitrogen) usando un homogeneizador eléctrico. Así mismo, se homogeneizó un cerebro de ratón adulto en igual volumen de Trizol. Se extrajo el ARN con 0,1 ml de cloroformo, agitando por 5 minutos y centrifugando a 4 °C por 15 minutos a máxima velocidad. La fase acuosa se transfirió a un tubo limpio y el ARN se precipitó con 0,25 ml de

isopropanol, reservándolo 1 hora a -80 °C. Se realizó una segunda centrifugación por 15 minutos a máxima velocidad en frío y se lavó el precipitado con 0,2 ml de etanol 70% en agua. El ARN se resuspendió en 20 µl de agua libre de nucleasas. Se corrió 1 µl de la solución de ARN en un gel de agarosa 1% y se cuantificó la concentración de ARN en nanodrop.

Síntesis de ADN copia

Antes de comenzar con la síntesis del ADN copia (ADNc), los restos de ADN presente en las muestras se eliminaron tratando 20 µg de ARN con 1 µl de DNAsa I (Invitrogen) por 30 minutos a 37 °C. La enzima se eliminó utilizando las columnas para eliminación de enzimas Micropure-EZ (Millipore). Las muestras de ARN tratadas con DNAsa I se utilizaron como molde para la síntesis de ADNc. Para ello, se utilizaron 5 µg de ARN por reacción, 1 µl de solución de oligo dT (500 µg/ml), 1 µl de dNTPs (10 mM) y la enzima transcriptasa reversa (RT Superscript II, Invitrogen) en el buffer recomendado por los proveedores de la enzima. La reacción se incubó por 50 minutos a 42 °C. Finalmente, la enzima se inactivó tratando con calor 15 minutos a 70 °C.

Amplificación y clonado de las secuencias molde para las sondas de *In Situ*

Las secuencias de las sondas para detección del mensajero de *NPAS3* fueron amplificadas por PCR utilizando como molde una solución 1:50 de ADN copia de cerebro de ratón o de ADN copia de embriones de pez cebra de 24 y 48 hpf. Las secuencias de los primers utilizados fueron las siguientes:

Para clonar la sonda contra el ARN mensajero de *Npas3* de ratón (Brunskill et al. 1999):

Npas3probe1_F: 5'-CGAATAACTGCCAGCATC-3'

Npas3probe1_R: 5'-TTTTCTTCTTGATTTAGTGCAA-3'

Para clonar las sondas contra el ARN mensajero de *NPAS3* de pez cebra:

Npas3_Dr_E2_241_F: 5'-CACCCAGGCTCTGAGGAAGGAGAA-3'

Npas3_Dr_E2_241_R: 5'-CTTTGACGGAGGTGTTGG-3'

Npas3_Dr_EGF_451_F: 5'-CACCTCCGGTTCCTGCTCTAC-3'

Npas3_Dr_EGF_451_R: 5'-TGCGATCCAGAGTCTGTGTG-3'

La amplificación se realizó utilizando la polimerasa de alta fidelidad Pfx (Invitrogen). El programa de amplificación se detalla a continuación:

- 1) Desnaturalización inicial: 95 °C por 5 minutos.
- 2) Desnaturalización: 95 °C por 30 segundos.
- 3) *Annealing*: 64 °C por 30 segundos.
- 4) Extensión: 68 °C por 30 segundos.
- 5) 29 ciclos de los pasos 2 a 4.
- 6) Extensión final: 10 minutos a 72 °C.

Los productos de PCR fueron analizados en gel de agarosa 1,5% y las bandas de interés se aislaron directamente desde el gel. El ADN se purificó utilizando un kit para purificación de ADN desde gel de agarosa (QIAquick Gel Extraction Kit, Qiagen). Luego, los fragmentos de PCR conteniendo la secuencia de la sonda de ratón fueron clonados en el vector pGEM_Teasy (Promega) o en el vector pENTRD (Invitrogen) los fragmentos necesarios para obtener las sondas de pez cebra. Se confirmó la identidad de las secuencias clonadas por secuenciación de los clones aislados con los primers M13F.

Síntesis de la sonda

Se realizaron midipreparaciones de ADN plasmídico de los clones seleccionados conteniendo las secuencias de interés utilizando el kit HiSpeed Plasmid Purification (Qiagen). Se digirieron 30 µg de ADN plasmídico con las enzimas SpeI (sonda de ratón) o NotI (sonda de pez cebra) por 16 horas a 37 °C. Se confirmó la digestión total de los plásmidos corriendo 1 µl de la solución de digestión en gel de agarosa. Las enzimas fueron eliminadas utilizando las columnas de purificación de enzimas Micropure-EZ (Millipore) y el ADN se precipitó utilizando 2 volúmenes de etanol y 0,2 volúmenes de acetato de sodio 3 M. Luego de centrifugar, el precipitado se lavó con etanol 70% en agua y el ADN se resuspendió en 30 µl de agua libre de nucleasas. La concentración final de ADN se cuantificó en NanoDrop (Thermo Scientific).

Para la síntesis de las sondas, se utilizó como molde 1 µg de ADN plasmídico linealizado por reacción. Las síntesis y marcación del ARN de las sondas de secuencia antisentido se realizó utilizando el kit DIG RNA Labeling Mix (Roche) y la enzima ARN polimerasa T7 (Roche). El ARN resultante se purificó utilizando el *RNeasy mini kit* (Qiagen). Para obtener la sonda sentido contra el mensajero *Npas3* de ratón, el plásmido se linealizó con la enzima BamHI y la síntesis del ARN se realizó utilizando la enzima Sp6 (Promega). Los productos de síntesis resultantes se analizaron en un gel de agarosa 1% y se cuantificaron en NanoDrop (Thermo scientific). Los largos resultantes de las sondas fueron de aproximadamente de 460 pb para la sonda contra el mensajero

de *Npas3* de ratón y de 240 pb (sonda E2) y 450 pb (sonda EG) para las sondas contra el mensajero de *NPAS3* de pez cebra.

Congelado y corte del tejido

Los embriones de ratón o pez cebra de las edades deseadas se fijaron en una solución de paraformaldehído 4% en PBS (buffer fosfato de sodio 20 mM, NaCl 150 mM pH 7.4) por 14-16 horas a 4 °C. Al día siguiente, se lavó el tejido con PBS durante todo el día. Posteriormente, las muestras se equilibraron en una solución de sacarosa al 10% en PBS por 14-16 horas a 4 °C. Los embriones se embebieron en una solución de sacarosa 10%, gelatina 10% en PBS y se congelaron en isopentano enfriado a -55 °C con hielo seco. El material se guardó a -70 °C hasta el momento del corte.

El tejido se cortó en un crióstato (Leica) a 18-25 µm los embriones de ratón y a 14-16 µm los embriones de pez cebra, dependiendo de la edad. Los cortes obtenidos se adhirieron a portaobjetos de vidrio Superfrost Plus (Fisher Brand).

Hibridación *In situ* en cortes histológicos

Los vidrios se dejaron secar a temperatura ambiente por 30 minutos. Posteriormente el tejido se fijó en una solución de paraformaldehído 4% en PBS por 15 minutos y las muestras se lavaron en PBS 3 veces por 15 minutos. Se realizó un paso de acetilación en una solución con 3,2 ml de trietanolamina, 420 µl ácido clorhídrico 36%, 600 µl de anhídrido acético y 234 ml de agua libre de nucleasas por 10 minutos en agitación suave. Posteriormente, el tejido se permeabilizó tratándolo con una solución de tritón-X100 1% en PBS por 30 minutos. Paso seguido los cortes se lavaron en PBS y se realizó un paso de pre-hibridación con 400 µl de solución de hibridación (formamida 50%, dextrán sulfato 10%, ARNt de levaduras 1 mg/ml, solución de Denhart's 1 mg/ml en PBS preparado en agua libre de nucleasas) por corte a temperatura ambiente por 4 horas en cámara humidificada. Luego, la solución de prehibridación se reemplazó por 200 µl de solución de hibridación por vidrio conteniendo 1 µl de solución de sonda de ARN marcado con digoxigenina. Los vidrios se cubrieron con un cubreobjeto y se colocaron en una cámara humidificada por 14-16 horas a 72 °C.

Al día siguiente, se realizó un primer lavado en buffer citrato de sodio 0,2X (NaCl 0,3 M y 30 mM de citrato de sodio, pH 7) a 72 °C por 1 hora y un segundo lavado en el mismo buffer por 5 minutos 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) 3 veces 5 minutos. El bloqueo se realizó en 500 µl por vidrio de

solución B1 con 10% de suero fetal bovino por 4 horas a temperatura ambiente en cámara humidificada. Luego, los vidrios se cubrieron con 500 µl de solución de B1 con 10% de suero fetal bovino y anticuerpo anti-digoxigenina acoplado a fosfatasa alcalina 1/3500 (Roche) por 14-16 horas a temperatura ambiente.

Al día siguiente, se realizaron tres lavados en B1 por 10 minutos 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 300 µl por vidrio de solución B2 con los sustratos de la fosfatasa alcalina nitroblue tetrazodium (NBT, Roche) y 5-bromo-4cloro-3-indolil fosfato (BCIP, Roche). La reacción se finalizó sumergiendo los vidrios en PBS. 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* fluorescente en cortes histológicos

Los vidrios se dejaron secar a temperatura ambiente por 30 minutos. Posteriormente el tejido se fijó en una solución de paraformaldehído 4% en PBS por 15 minutos y las muestras se lavaron en PBS 3 veces por 15 minutos. Posteriormente, el tejido se permeabilizó tratándolo con una solución de tritón-X100 1% en PBS por 30 minutos. Paso seguido los cortes se lavaron en PBS y se realizó un paso de pre-hibridación con 400 µl de solución de hibridación (formamida 50%, dextrán sulfato 10%, ARNt de levaduras 1 mg/ml, solución de Denhart's 1 mg/ml en PBS preparado en agua libre de nucleasas) por corte a temperatura ambiente por 4 horas en cámara humidificada.

Luego, la solución de prehibridación se reemplazó por 200 µl de solución de hibridación por vidrio conteniendo 0,25 µl de solución de sonda de ARN marcado con digoxigenina. Los vidrios se cubrieron con un cubreobjeto y se colocaron en una cámara humidificada por 14-16 horas a 72 °C. Al día siguiente, se realizó un primer lavado en buffer citrato de sodio 0,2X (NaCl 0,3 M y 30 mM de citrato de sodio, pH 7) a 72 °C por 1 hora y un segundo lavado en el mismo buffer por 15 minutos con 2% de agua oxigenada a temperatura ambiente y en oscuridad. Los cortes se equilibraron en solución B1 (0,1 M Tris, 0,15 M NaCl, Tritón-X100 0,1%, pH 7,5) 3 veces 5 minutos. El bloqueo se realizó en 500 µl por vidrio de solución B1 con 10% de suero de cabra normal por 4 horas a temperatura ambiente en cámara humidificada. Luego, los vidrios se cubrieron con 500 µl de solución de B1 con 1% de suero de cabra normal y anticuerpo anti-digoxigenina acoplado a peroxidasa 1/500 (Roche) por 14-16 horas a 4 °C.

Al día siguiente, se realizaron tres lavados en B1 por 10 minutos. La reacción cromogénica de tiramida se produjo reemplazando la solución B1 por 200 µl/ vidrio de *Fluorophore Amplification Reagent* (1/200 de *Cyanine 5 Plus Amplification Reagen* en 1 X *Plus Amplification Diluen*, Perkin Elmer) por 30 minutos a temperatura ambiente. La reacción se finalizó sumergiendo los vidrios en 1 X PBS. Los cortes se lavaron 3 veces en 1 X PBS y se montaron en Vectashield. Las imágenes fueron adquiridas con un microscopio OLYMPUS BX51 equipado con una cámara OLYMPUS DP72.

Hibridación *In situ* fluorescente e inmunohistoquímica en corte de embriones de ratón

Una vez finalizada la ISH, los cortes se lavaron 7 veces X 1 hora en 1 X PBS y dejaron 16 hora en 1 X PBS para eliminar los restos de citrato de los pasos anteriores. Se realizó el bloqueo en una solución de 10 % suero normal de burro (Jackson Immuno Research), 0,1 % de Triton X-100 (Sigma) en 1 X PBS por 2 horas a temperatura ambiente. El anticuerpos primario se incubó en una solución conteniendo 0,1 % Triton X-100 en PBS 1 X durante toda la noche a 4 °C. Los anticuerpos usados fueron los siguientes: anticuerpo policlonal anti-SOX2 hecho en cabra (1/100, Santa Cruz), anticuerpo policlonal anti-Doublecortin hecho en cabra (1/100, Santa Cruz), anticuerpo monoclonal anti-β-Tubulina Neuronal Clase III (TUJ1) hecho en ratón (1/100, Covance) y anticuerpo policlonal anti-β-Galactosidasa hecho en conejo (1/100, Cappel). Al día siguiente los cortes se lavaron 3 veces en 1 X PBS y se incubaron con el anticuerpo secundario en una solución 0,1 % Triton X-100 en 1 X PBS por 2 horas a temperatura ambiente. Los anticuerpos secundarios utilizados fueron: anticuerpo anti-IgG de cabra hecho en burro acoplado a Alexa Fluor 488 (1/1000, Molecular probes), anticuerpo anti-IgG de ratón hecho en burro acoplado a Cyt3 (1/1000, Jackson Immuno Research) y anticuerpo policlonal anti-IgG de conejo acoplado a Alexa Fluor 488 (1/1000, Molecular probes). Los cortes se lavaron 3 veces en 1 X PBS y se montaron en Vectashield (Vector Laboratorios). Las imágenes fueron adquiridas en un microscopio OLYMPUS BX51 equipado con una cámara OLYMPUS DP72.

Doble Inmunohistoquímica en cortes de pez cebra

Se obtuvieron embriones de pez cebra de las líneas deseadas. Los peces se sacrificaron por inmersión en agua helada por 30 minutos y el tejido se fijó 2 horas en una solución de paraformaldehído 4% en 1 X PBS. Los embriones se lavaron 3 veces en 1 X PBS y se crioprotegieron y congelaron en una solución de gelatina 10%, sacarosa 10% en 1 X PBS siguiendo el protocolo utilizado para la ISH.

Se obtuvieron cortes transversales (14 μm) en crióstato y las muestras se adhirieron en vidrios portaobjetos Superfrost Plus (Fisher Scientific). Los cortes se secaron por 30 minutos a temperatura ambiente y se lavaron 3 veces en 1 X PBS para eliminar los restos de gelatina. A continuación, se realizó el bloqueo con una solución de 10 % suero normal de burro (Jackson Immuno Research), 0,1 % Triton X-100 (Sigma) en 1 X PBS. Los anticuerpos primarios se incubaron juntos en una solución conteniendo 0,1 % Triton X-100 en PBS 1 X durante toda la noche a 4 °C. Los anticuerpos usados fueron los siguientes: anticuerpo policlonal anti-GFP hecho en conejo (1/500, Abcam) y anticuerpo monoclonal anti-HuC/D hecho en ratón (1/20, Molecular Probes). Al día siguiente los cortes se lavaron 3 veces en 1 X PBS y se incubaron secuencialmente con los anticuerpos secundarios en una solución 0,1 % Triton X-100 en 1 X PBS por 1 hora a temperatura ambiente. Los anticuerpos secundarios utilizados fueron: anticuerpo anti-IgG de conejo hecho en cabra acoplado a Alexa Fluor 488 (1/1000, Molecular probes) y anticuerpo anti-IgG de ratón hecho en burro acoplado a Alexa Fluor 555 (1/1000, Jackson Immuno Research). Los cortes se lavaron 3 veces en 1 X PBS y se montaron en Vectashield (Vector Laboratorios). Las imágenes fueron adquiridas en un microscopio Leica MZFLIII acoplado a una cámara DFC320.

Hibridación *In Situ* en embriones completos de pez cebra

Preparación del tejido

Se sacrificaron aproximadamente 50 embriones de pez cebra de 24, 48 hpf o 5 días en agua helada. El tejido se fijó 14-16 horas en una solución de paraformaldehído 4% en PBS a 4 °C. Luego, los embriones se lavaron en PBT (PBS 1X con 0,1% de Tween-20) y se deshidrataron equilibrando los tejidos en soluciones de concentración creciente de metanol (25%, 50% y 75%). Finalmente, los embriones se pasaron a metanol 100% y se almacenaron hasta el momento de ser usados a -20 °C.

Hibridación *In Situ* en embriones completos

Los embriones se coloraron en placas multi-well y se rehidrataron pasándolos por soluciones de concentración creciente de PBT (25%, 50% y 75%). Se realizaron dos lavados de 5 minutos con PBT a temperatura ambiente. Los tejidos se permeabilizaron en 500 μl de PBT con 10 $\mu\text{g/ml}$ de proteinasa K por 5 minutos para embriones de 24-48 hpf, 10 minutos para embriones de 72 hpf y 30 minutos para embriones de 5 días o más. Se realizó una post fijación en paraformaldehído 4% en PBS por 10 minutos para embriones de 24 hps, 20 minutos para embriones de 48 y 72 hpf y 30 minutos para embriones de 5 días o mas. Posteriormente, se realizaron 5 lavados de 5 minutos cada uno en PBT. La prehibridación se realizó por 1 hora a 65 °C en 500 μl de solución de hibridación (50% formamida, buffer citrato 5X, heparina 50 $\mu\text{g/ml}$, 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 1 µl/ml de sonda de ARN marcada con digoxigenina.

Al día siguiente, se realizaron una serie de lavados en solución de hibridación con concentraciones crecientes de buffer citrato a 65 °C por 5 minutos (0,6X, 1,2X, 2X). Luego, se hicieron 3 lavados de 20 minutos a 65 °C en buffer citrato con Tween-20 0,1%, uno con buffer citrato 0,2X y dos con buffer citrato 0,1X. Finalmente, se hicieron dos lavados de 5 minutos cada uno a temperatura ambiente con buffer citrato 0,12X y 0,06 en PBT. El bloqueo se realizó por 1 hora a temperatura ambiente en 500 µl de solución de bloqueo (2% de suero normal de pollo, 2 mg/ml de BSA en PBT). Se reemplazó la solución anterior por 300 µl de solución de bloqueo fresca con 1/5000 de anticuerpo anti-digoxigenina acoplado a fosfatasa alcalina (Roche) y se incubó 14-16 horas a 4 °C en agitación suave.

Luego, se hicieron 5 lavados de 15 minutos en PBT. El tejido se equilibró en solución B2 y se reveló siguiendo el protocolo mencionado anteriormente para las *in situ* en corte de crióstato. La reacción se frenó sumergiendo los embriones en agua estéril y se montaron en una solución de glicerol 70% en agua. Los peces cebras fueron fotografiados en un microscopio OLYMPUS BX51 equipado con una cámara OLYMPUS DP72.

Cultivo primario de fibroblastos y preservación del cultivo celular

Fibroblastos provenientes de biopsias de piel de chimpancé (*Pan troglodytes*) (donación de Adrian Sestelo) se cultivaron en monocapa en placas de 10 cm de diámetro conteniendo 10 ml de medio mínimo esencial (DMEM, Gibco) suplementado con 10 % de suero fetal bovino inactivado por calor, 0,007% de glucosa 50%, NaHCO₃ 1,5 gr/l, L-glutamina 2 mM, amfotericina 2,5 µg/ml, penicilina 100 U/ml y estreptomina 100 µg/ml (medio DMEM completo), pH 7,4. Los cultivos se crecieron en estufa a 37 °C y atmósfera de 5% de CO₂. Antes de alcanzar la confluencia, los fibroblastos se pasaron a nuevas placas desadheriendo las células por tratamiento con tripsina. Para ello se aspiró el medio, se lavaron las células con *phosphate buffered saline* (PBS) estéril y se incubó cada caja con 1 ml de tripsina a 37 °C hasta detectar al microscopio (400X) que las células estuvieran despegadas de su sustrato. La inactivación de la enzima se logró por agregado de medio de cultivo suplementado con suero.

Para la determinación del número de células/ml se utilizó el método de recuento en cámara. Se colocó una gota de la suspensión celular en una cámara de Neubauer adecuadamente preparada y el recuento se efectuó al microscopio óptico (400X). El conteo de células (N) se realizó en los cuatro

cuadrantes principales de la cámara y el número de células totales por ml se calculó aplicando la siguiente fórmula:

$$\text{Células / ml} = N \times \frac{1}{4} \times 1.000 \times 25 \times 10$$

Los fibroblastos se plaquearon a una densidad de $2,5-3 \times 10^6$ células por caja de 10 cm de diámetro en 10 ml de medio.

Periódicamente, se congelaron suspensiones celulares de fibroblastos para conservar un stock de las células. Para ello, una vez tripsinizadas, las células se centrifugaron durante 3 minutos a 1000 rpm. Se aspiró el medio, las células se resuspendieron en 1 ml de medio de congelamiento preenfriado (DMEM 1X 65%, suero fetal bovino 25%, dimetilsulfóxido (DMSO) 10%) y se transfirieron a un criotubo. El criotubo envuelto en algodón se mantuvo por 24 horas a $-80\text{ }^{\circ}\text{C}$. Pasado ese lapso, se transfirió a nitrógeno líquido para su conservación (Freshney, 1987).

Para amplificar las células provenientes del nitrógeno líquido, en una placa de 10 cm de diámetro se mezcló el contenido de un criotubo parcialmente descongelado con 10 ml de medio fresco para diluir el DMSO. 24 horas después del descongelamiento, se aspiró el medio de cultivo y se reemplazó por medio DMEM completo fresco.

Extracción de ADN genómico de cultivos celulares

5 ml de suspensión de fibroblastos provenientes de 10 placas de cultivo de 10 cm se lisó con 5 ml de buffer de extracción (10mM Tris-Cl pH 8, 0,1 mM EDTA, 20 ug/ml RnaseA, 0,5% SDS) frío suplementado con proteinasa K (100 ug/ml). Se incubó durante toda la noche a $37\text{ }^{\circ}\text{C}$ en agitación suave (300 rpm). Se realizaron dos extracciones consecutivas con 5 ml de fenol:cloroformo:alcohol isoamilico (25:24:1). Se mezcló por inmersión por 10 minutos, se centrifugó por 10 minutos a 5000 G y se recuperó la fase acuosa. Posteriormente se realizaron dos extracciones con 5 ml de cloroformo: alcohol isoamilico (24:1), mezclando por inversión por 10 minutos. Se centrifugó nuevamente 10 minutos a 5000 G y se recuperó la fase acuosa. El ADN se precipitó agregando 0,2 volúmenes de acetato de amonio 10 M y 2 volúmenes de etanol 100%. La muestra se centrifugó 10 minutos a máxima velocidad y el precipitado se lavó con 200 ul de etanol 70%. Una vez seco el precipitado, el ADN se resuspendió en 500 ul de agua estéril. 5ul de la muestra se analizó en un gel de agarosa 1% para verificar la calidad del ADN y se cuantificó utilizando un NanoDrop.

Inyección de Morfolinos

Se utilizaron dos morfolinos, uno que bloquea la traducción (MO1) y uno que bloquea el *splicing* (MO2) de los ARNm de *NPAS3* de pez cebra (Gene Tools). Las secuencias de los morfolinos utilizados en este estudio fueron:

1) Morfolino de traducción contra el ARNm de *NPAS3* de pez cebra:

MO1: GGAAGATCCTGATCATCCCCGAGC

2) Morfolinos de *Splicing* contra ARNm de *NPAS3* de pez cebra:

MO2: CCCGATTGCTGTGTAGAAAAACACA

Los morfolinos se reconstituyeron en agua destilada libre de ARNsas y se dejaron a una concentración de stock de 1 mM a -20 °C. El día de la inyección se preparó una mezcla fresca conteniendo:

MO 1 mM	4 μ l
KCl 1 M	0,5 μ l
Rojo fenol (2% en agua)	0,5 μ l

Para la puesta a punto de los morfolinos, se inyectaron 1, 2, 2,5 o 5 nl de la mezcla anterior por embrión de en el estadio de 1-2 células, para alcanzar las concentraciones finales de morfolino en los embriones de 0,75, 1,5, 1,875 y 3,75 μ M respectivamente. Se inyectaron 100 embriones por concentración y tipo de morfolino. Como control, se inyectaron idénticos volúmenes de solución de inyección pero reemplazando el morfolino por el buffer en el cual se reconstituyó el morfolino (en este caso, agua libre de ARNsas). Se cuantificó el número total de embriones muertos o con signos de necrosis a 24 hpf para todas las condiciones y se estudió la penetrancia del fenotipo del morfolino para aquellas concentraciones que no elevaban la letalidad/necrosis por sobre los valores observados en los embriones no inyectados.

La concentración final utilizada en todos los ensayos fue de 0.75 μ M para los dos morfolinos. Esta concentración no produce una reducción significativa en la supervivencia de los embriones a 24 hpf respecto de los embriones no inyectados tratados en iguales condiciones que los inyectados y tiene niveles altos de penetrancia en el fenotipo (por ejemplos, superiores al 90% en el caso de defectos en el inflado de la vejiga natatoria). En todos los casos, los dos morfolinos se inyectaron en 100-150 embriones de 1-2 células por réplica durante la misma mañana de inyección y utilizando el mismo lote de reproductores (hembras y machos exocriados). Adicionalmente, se inyectaron 100-150

embriones en iguales condiciones (1 nl) con una solución donde el volumen correspondiente al morfolino fue reemplazado por el buffer de disolución. Adicionalmente, se mantuvo y se analizaron 100-150 embriones que no fueron inyectados como control adicional de la viabilidad. Se analizó el fenotipo de cada morfolino a 3 y 5 dpf.

A la edad deseada, los embriones fueron anestesiados como se detalló previamente y montados en metilcelulosa 2%. Los peces cebrá fueron fotografiados en un microscopio OLYMPUS BX51 equipado con una cámara OLYMPUS DP72. Las mediciones de largo del cuerpo, diámetro del ojo y longitud de la cabeza fueron realizadas utilizando el programa Image J. Con los resultados de las tres réplicas se estimó un valor medio y desvío estándar. Se evaluó el ajuste de la series de datos a la distribución Gaussiana mediante tests de Kolmogorov-Smirnov, D'Agostino-Pearson y Shapiro-Wilk. También se evaluó si estas series de datos presentaban homogeneidad de varianzas mediante un test de Bartlett. Sin embargo, en todos los casos, alguna de las series de datos no se ajustaron a una distribución normal con un valor $p \leq 0,05$ y/o no se pudo comprobar homocedacea. Entonces, se realizaron pruebas estadísticas no paramétricas. Para ello, se realizaron contrastes múltiples utilizando la prueba de Kruskal-Wallis seguida del test de contrastes de Dunn's. En todos los casos los valores de $p < 0,05$ fueron considerados significativos. El mismo procedimiento se aplicó para comparar los valores promedio de supervivencia entre tratamientos. Todos los análisis estadísticos se realizaron usando el programa Graph Pad Prism 4.00 para Windows (GraphPad Software).

Tinción de alcian blue

La tinción de cartílago con alcian blue se realizó sobre embriones de pez cebrá de 5dpf. Los embriones fueron sacrificados por inmersión en agua helada por 30 minutos. Posteriormente, se fijaron durante toda la noche en una solución de paraformaldeído al 4% en 1 X PBS. Luego, se lavaron tres veces por 10 minutos en una solución de tween20 al 0,1% en 1 X PBS. Se tiñeron por 90 minutos en una solución 0,37% HCl, 70% etanol y 0,1% p/v de alcian blue (Sigma-Aldrich, catálogo número 05500-10G). Se lavó tres veces por un minuto con una solución 1% HCl y 70% etanol para eliminar el exceso de colorante. Finalmente, los embriones se clarificaron en una solución 50% glicerol y 0,5% KOH durante toda la noche a 37 °C.

Tinción de Hematoxilina-Eosina

La tinción de hematoxilina-eosina se realizó sobre cortes de crióstato de embriones de pez cebrá de 5 dpf. Una vez cortados, los portaobjetos se dejaron secar por 30 minutos a temperatura ambiente antes de realizar la tinción. Luego se lavaron con agua corriente por un minuto. La coloración con hematoxilina se realizó sumergiendo los vidrios por 10 minutos en una solución de 0,73% agua

destilada, 0,25 etilen glicol, 0,02% p/v hematoxilina, 0,002 p/v iodato de sodio, 0,176% p/v de sulfato de aluminio y 0,02% ácido acético glacial. Luego se lavó el exceso de colorante sumergiendo los portaobjetos en una solución de 1% HCl y 70% etanol por 1 minuto. Paso seguido, los vidrios se lavaron en agua corriente por 10 minutos. La tinción de eosina se realizó sumergiendo los vidrios por 2 minutos en una solución de eosina alcohólica (Sigma-Aldrich, cat#HT110316-500ML). Se realizaron cuatro pasos de deshidratación de 1 minuto cada uno en soluciones de concentración creciente de etanol en agua (70, 85, 90 y 95% respectivamente). Finalmente, el tejido se clarificó sumergiéndolo en xilol por 1 minuto y se montó en DPX (Fluka, cat#44581).

APÉNDICE

Análisis evolutivo en la región codificante del gen *NPAS3*

Primero, se obtuvieron las secuencias codificantes del gen *NPAS3* de las siguientes especies: humano, chimpancé, orangután, gibón, mono rhesus, mono tití, tupaya, ratón, rata, cobayo, conejo, caballo, vaca, delfín, cerdo, murciélago, elefante y zarigüeya. Luego, las secuencias se alinearon usando el programa ClustalW y se calculó el porcentaje de identidad de secuencias usando MEGA5 (Tamura et al. 2011).

Tabla suplementaria s1: tabla indicando el porcentaje de identidad de secuencia de aminoácidos de *NPAS3* entre diferentes especies de mamíferos con/sin tracto de poliG.

	Chimpancé	Orangután	Gibón	Mono rhesus	Mono tití	Tupaya	Ratón	Rata	Cobayo	Conejo	Caballo	Vaca	Delfín	Cerdo	Murciélago	Elefante	Zarigüeya
Humano	99 / 99	99 / 99	98 / 98	99 / 99	98 / 99	95 / 96	95 / 96	94 / 96	97 / 98	97 / 98	97 / 99	95 / 97	96 / 98	95 / 97	95 / 97	96 / 98	89 / 91
Chimpancé	-	99	97 / 98	99 / 99	99 / 99	95 / 96	95 / 96	94 / 96	97 / 98	97 / 98	97 / 99	96 / 97	97 / 98	95 / 97	96 / 97	96 / 98	89 / 91
Orangután		-	98	99 / 97	99 / 97	96 / 93	96 / 93	96 / 93	98 / 95	98 / 96	99 / 96	97 / 94	98 / 95	97 / 93	97 / 94	98 / 95	91 / 88
Gibón			-	98	98	95	95	94	97	97	97	96	97	96	96	96	90
Mono rhesus				-	98 / 99	95 / 96	95 / 96	94 / 96	97 / 98	97 / 98	97 / 99	96 / 97	97 / 98	95 / 98	96 / 97	97 / 98	89 / 91
Mono tití					-	96	96	95	98	98	98	96	97	97	97	97	91
Tupaya						-	93	93	95	95	96	94	95	95	94	95	94
Ratón							-	99	96	96	96	95	96	95	95	96	90
Rata								-	96	95	96	94	95	95	94	95	90
Cobayo									-	98	98	97	98	97	97	97	91
Conejo										-	98	96	98	97	96	97	90
Caballo											-	95 / 97	96 / 98	95 / 97	96 / 97	95 / 97	88 / 91
Vaca												-	97 / 98	95 / 97	94 / 95	95 / 95	90
Delfín													-	96 / 98	94 / 96	96 / 96	91
Cerdo														-	93 / 95	94 / 96	88 / 91
Murciélago															-	96 / 95	89 / 90
Elefante																-	91

Con el fin de hacer un análisis de evolución adaptativa sobre el gen *NPAS3*, se utilizó un modelo de ramas por sitio (Zhang, Nielsen y Yang 2005) en las secuencias de mamífero usando PAML (Yang 2007). Las ramas focales estudiadas fueron la humana y la primate y se testearon dos hipótesis anidadas para cada rama focal. En la hipótesis alternativa, se calcularon tres valores de ω :

ω_0 , ω_1 y ω_2 representando los codones bajo selección negativa, neutros y selección positiva respectivamente. Sin embargo, en la hipótesis nula, solo se calcularon dos valores (ω_0 y ω_1) y no se permitieron sitios bajo selección positiva.

Adicionalmente, se estudió todo el *locus* en el cual se encuentra el gen *NPAS3*. Para ello se obtuvo la secuencia de la región genómica que abarca desde la última base del gen *AKAP6* hasta la primera base del gen *EGLN3* (comprendiendo un total de 1.091.152 pares de bases que cubren desde la base 32.372.020 hasta la base 33.463.171 del cromosoma 14 humano; NCBI36/ hg18). Esta secuencia se dividió en cuatro fragmentos representado regiones con diferentes características: regiones codificantes, regiones conservadas no codificantes, regiones HAE y regiones no codificantes no conservadas. Se obtuvieron las secuencias ortólogas de chimpancé, orangután y mono rhesus para los elementos presentes en cada una de las cuatro categorías antes descritas y se alinearon. Utilizando el programa phyloP presente en el paquete PHAST, se calcularon los valores P de conservación y aceleración en los alineamientos con el método SPH (Siepel A 2006). La conservación y aceleración se calcularon en cada caso para un sub-árbol conteniendo solamente a la rama humana y un super-árbol conteniendo las especies chimpancé, orangután y mono rhesus. Todos los alineamientos se compararon contra el modelo de evolución neutro *primates.mod* usado en el la herramienta de conservación del sitio UCSC.

Tabla suplementaria s2: Análisis de la región codificante de *NPAS3*

NPAS3	Log Likelihood					
M0 Model	-14,902.86123					
Branch-Site Models	LRT	P	Estimación parámetros			
Primates						
Hipótesis alternativa	-10,044.620481	0.000000	0.0000	$\rho_0 = 0.97012$	$\rho_1 = 0.02988$	$\rho_2 = 0.00000$
Hipótesis nula	-10,044.620481			$\omega_0 = 0.01766$	$\omega_1 = 1.00000$	$\omega_2 = 1.00000$
Humano						
Hipótesis alternativa	-10,044.620473	0.000000	0.0000	$\rho_0 = 0.97012$	$\rho_1 = 0.02988$	$\rho_2 = 0.00000$
Hipótesis nula	-10,044.620473			$\omega_0 = 0.01766$	$\omega_1 = 1.00000$	$\omega_2 = 1.00000$

NOTA: LRT (likelihood ratio test) = $2 \cdot [\text{LnLHA (hipótesis alternativa)} - \text{LnLH0 (hipótesis nula)}]$; P, P valor del test de chi-cuadrado.

Identificación de sitios probables de pegado de factores de transcripción y estructuras secundarias de ARN

Utilizando el programa MatInspector (www.genomatix.de), se realizó sobre las secuencias de los HAEs presentes en el gen *NPAS3* en humanos y chimpancé un análisis comparativo de ganancia y pérdida de sitios de unión para factores de transcripción. En todos los casos, se encontró que las

diferencias de secuencia entre los HAEs humanos y sus secuencias ortólogas en chimpancé predicen cambios en los sitios de reconocimiento para factores de transcripción.

Paralelamente, se analizó la capacidad predicha de los HAEs para formas estructuras secundarias estables usando el programa EvoFold (7), encontrándose que 7 de los 14 HAEs de NPAS3 podrían tener estructuras secundarias de ARN estables.

Tabla suplementaria s3: Análisis *in silico* de sitios probables de unión para factores de transcripción (TFBS) ganados y perdidos en humanos, conservación y análisis del sesgo en las sustituciones por sitios débiles a sitios fuertes.

HAE	Localización genómica (NCBI36/hg18)	Largo (bp)	Sust_Hs	SP	P Hs	Conservación con pez cebra	TFBS ganancia (↑) Y pérdida (↓)	Sesgo AT→CG
HAR96	chr14: 32486645-32487168	524	2	7	-	NO (Rana)	↑YTBP; ↓EVII ↑CREB; ↓RORA ↑ETSF; ↓AP4R ↓PERO	A→G; A→T
HACNS658	chr14: 32487503-32487837	335	6	6	1	NO (Lagarto)	↑LEFF; ↓HOMF ↑NR2F; ↓HOXF ↑XBBF; ↓APIR ↑PERO; ↓STEM	A→G; T→C T→A; C→T T→C; C→T
HAR189/ HACNS490	chr14: 32558123-32558864	742	6	10	1*	NO (Lagarto)	↑FKHD; ↓PLAG ↑E2FF; ↓GCMF ↑HOMF; ↓NR2F ↑GRHL; ↓TALE ↑INRE; ↓TEAF ↑CLOX; ↓RREB ↑HZIP; ↑BRNF ↑BNCF	C→G; T→C T→C; T→G A→G; G→A
HACNS553	chr14: 32566093-32566346	254	4	5	-	NO (Pollo)	↑EGRF; ↓GTBX ↑INSM; ↓CAAT ↓GATA ↓CLOX	A→G; T→G A→T; A→G
HAR21	chr14: 32596637-32596862	226	3	2	-	NO (Lagarto)	↑NKXH; ↓BRAC ↑RU49; ↓DMRT ↓EVII; ↓ABDB ↓TALE; ↓GATA	G→T; A→G C→T
HACNS221	chr14: 32718813-32719042	230	5	3	-	NO (Ornitorrinco)	↑SORY; ↓FKHD ↑BRN5; ↓EVII ↑EVII; ↓APIR ↓CREB; ↓OVOL ↓PARE; ↓GREF	C→T; T→G T→G; C→T T→C
HAR173	chr14: 32830590-32830773	184	2	2	-	NO (Lagarto)	↑PRDF; ↓HBOX ↑GFII; ↓SORY ↑CLOX; ↓HOXF ↓PLZF	A→C; A→G
HACNS96	chr14: 32888897-32889211	315	6	6	-	SI	↑WHNF; ↓CREB ↑MOKF; ↓HIF1 ↑HNF1; ↓HASF ↑GATA; ↓BRAC ↑MYBL; ↓ABDB ↑HOMF; ↓FKHD ↑MEF2; ↓MEF2 ↑MYT1; ↓CDXF ↑RORA; ↓VTBP ↓MYT1; ↓ARID ↓PAX5; ↓NKXH	A→G; A→C T→C; A→G T→C; T→G
HAR202	chr14: 33115156-33115595	440	5	8	1	SI	↓BCL6 ↓STAT	T→C; T→A A→G; C→T T→G

2xHAR142	chr14: 33118315-33118577	263	6	4	1/1*	NO (Mamíferos placentarios)	↑CLOX; ↓STEM ↑CREB; ↓ZF02 ↑HIFF; ↓NR2F ↑HNF1; ↓BCL6 ↑HZIP; ↓EVII ↑MYBL; ↓OCT1 ↓HOXF; ↓CART ↓MEF2; ↓SATB ↓PDX1	A→G; T→C G→T; A→G A→G; T→C
HAR89	chr14: 33135162-33135937	776	3	7	-	SI	↑PAX5; ↓SATB ↑HOMF; ↓OCT1 ↑MYBL; ↓ABDB ↑PERO; ↓HOXF ↑NR2F; ↓CART ↓NF1F; ↓RUSH ↓YY1F	A→G; A→G G→A
2xHAR223	chr14: 33199815-33200914	1100	8	13	-	SI	↑CART; ↓CHOP ↑ZF03; ↓EVII ↑NKXH; ↓SORY ↑SORY; ↓OCT1 ↑HOXC; ↓STEM	C→T; T→C T→C; A→G T→C; G→C T→C; G→T
2xHAR157	chr14: 33229602-33230232	631	10	5	-	SI	↑OAZF; ↓PARF ↑VTBP; ↓OCT1 ↑HOMF; ↓SORY ↑NKHX; ↓AIRE ↑GFII; ↓RREB ↑HOXC; ↑ZF06 ↑LEFF; ↑SMAD ↑FKHD	T→C; T→C T→C; A→G C→A; C→G A→C; C→A C→T; C→G
2xHAR122	chr14: 33331419-33331887	469	4	9	-	SI	↑MYBL; ↓SORY ↑CEBP; ↓FKHD ↑CARE; ↓IRFF ↓SNAP	T→C; T→C A→G; A→G

Nota: Análisis de polimorfismos humanos y cambios en los sitios de unión para factores de transcripción (TFBS por sus siglas en inglés) entre las secuencias de los HAEs y regiones ortólogas en chimpancé. Las coordenadas genómicas, el largo y el número de sustituciones humanas específicas corresponden a la región conservada completa que contiene al HAE y la que fue usada para generar los transgenes. Adicionalmente, en sesgo AT→CG se muestra el tipo de sustituciones ocurridas en los HAES en el linaje humano. En la última columna y en celdas amarillas se muestran aquellos elementos que presentan un sesgo fuerte (mayor al 60% de las sustituciones ocurridas) AT→CG (indicio del proceso evolutivo neutral BGC), mientras que en celeste se indica un sesgo CG→AT.

BIBLIOGRAFÍA

- Abecasis, G. R., D. Altshuler, A. Auton, L. D. Brooks, R. M. Durbin, R. A. Gibbs, M. E. Hurles, and G. A. McVean. 2010. A map of human genome variation from population-scale sequencing. *Nature* **467**:1061-1073.
- Abecasis, G. R., A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, and G. A. McVean. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56-65.
- Abzhanov, A., W. P. Kuo, C. Hartmann, B. R. Grant, P. R. Grant, and C. J. Tabin. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature* **442**:563-567.
- Ahituv, N. 2012. *Gene Regulatory Sequences and Human Disease*. Springer, New York.
- Ahituv, N., S. Prabhakar, F. Poulin, E. M. Rubin, and O. Couronne. 2005. Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. *Hum Mol Genet* **14**:3057-3063.
- Arendt, D., and K. Nubler-Jung. 1994. Inversion of dorsoventral axis? *Nature* **371**:26.
- Aris-Brosou, S., and Z. Yang. 2003. Bayesian models of episodic evolution support a late precambrian explosive diversification of the Metazoa. *Mol Biol Evol* **20**:1947-1954.
- Ariza-Cosano, A., A. Visel, L. A. Pennacchio, H. B. Fraser, J. L. Gomez-Skarmeta, M. Irimia, and J. Bessa. 2012. Differences in enhancer activity in mouse and zebrafish reporter assays are often associated with changes in gene expression. *BMC Genomics* **13**:713.
- Arnone, M. I., Dmochowski, I.J., Gache, C. 2004. Using reporter genes to study cis-regulatory elements. *Methods Cell Biol.* **74**:621-52.
- Atchison, M. L. 1988. Enhancers: mechanisms of action and cell specificity. *Annu Rev Cell Biol* **4**:127-153.
- Azevedo, F. A., L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. Jacob Filho, R. Lent, and S. Herculano-Houzel. 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* **513**:532-541.
- Bacsi, S. G., & Hankinson, O. . 1996. Functional characterization of DNA-binding domains of the subunits of the heterodimeric aryl hydrocarbon receptor complex imputing novel and canonical basic helix-loop-helix protein-DNA interactions. . *The Journal of Biological Chemistry* **271**, **8843-8850**.
- Ballard, W. W. 1981. Morphogenetic movements and fate maps of vertebrates. . *Am. Zool.* **21**:391-399.
- Barolo, S. 2012. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* **34**:135-141.
- Baron, G., H. D. Frahm, and H. Stephan. 1988. Comparison of brain structure volumes in insectivora and primates. VIII. Vestibular complex. *J Hirnforsch* **29**:509-523.
- Bedell, V. M., S. E. Westcot, and S. C. Ekker. 2011. Lessons from morpholino-based screening in zebrafish. *Brief Funct Genomics* **10**:181-188.
- Beermann, F., K. Kaloulis, D. Hofmann, F. Murisier, P. Bucher, and A. Trumpp. 2006. Identification of evolutionarily conserved regulatory elements in the mouse *Fgf8* locus. *Genesis* **44**:1-6.
- Begun, D. R. 2003. Planet of the apes. *Sci Am* **289**:74-83.
- Benito-Gutierrez, E., and D. Arendt. 2009. CNS evolution: new insight from the mud. *Curr Biol* **19**:R640-642.
- Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. 2004. GenBank: update. *Nucleic Acids Res* **32**:D23-26.
- Bernard, P., P. Gabant, E. M. Bahassi, and M. Couturier. 1994. Positive-selection vectors using the F plasmid *ccdB* killer gene. *Gene* **148**:71-74.

- Bill, B. R., A. M. Petzold, K. J. Clark, L. A. Schimmenti, and S. C. Ekker. 2009. A primer for morpholino use in zebrafish. *Zebrafish* **6**:69-77.
- Bird, C. P., B. E. Stranger, M. Liu, D. J. Thomas, C. E. Ingle, C. Beazley, W. Miller, M. E. Hurles, and E. T. Dermitzakis. 2007. Fast-evolving noncoding sequences in the human genome. *Genome Biol* **8**:R118.
- Bird, N. C., and P. M. Mabee. 2003. Developmental morphology of the axial skeleton of the zebrafish, *Danio rerio* (Ostariophysi: Cyprinidae). *Dev Dyn* **228**:337-357.
- Blackwood, E. M., and J. T. Kadonaga. 1998. Going the distance: a current view of enhancer action. *Science* **281**:60-63.
- Boffelli, D., M. A. Nobrega, and E. M. Rubin. 2004. Comparative genomics at the vertebrate extremes. *Nat Rev Genet* **5**:456-465.
- Bohne, A., F. Brunet, D. Galiana-Arnoux, C. Schultheis, and J. N. Volff. 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res* **16**:203-215.
- Bourque, G., B. Leong, V. B. Vega, X. Chen, Y. L. Lee, K. G. Srinivasan, J. L. Chew, Y. Ruan, C. L. Wei, H. H. Ng, and E. T. Liu. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**:1752-1762.
- Britten, R. J., and E. H. Davidson. 1969. Gene regulation for higher cells: a theory. *Science* **165**:349-357.
- Brosius, J., and S. J. Gould. 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc Natl Acad Sci U S A* **89**:10706-10710.
- Brunskill, E. W., L. A. Ehrman, M. T. Williams, J. Klanke, D. Hammer, T. L. Schaefer, R. Sah, G. W. Dorn, 2nd, S. S. Potter, and C. V. Vorhees. 2005. Abnormal neurodevelopment, neurosignaling and behaviour in *Npas3*-deficient mice. *Eur J Neurosci* **22**:1265-1276.
- Brunskill, E. W., D. P. Witte, A. B. Shreiner, and S. S. Potter. 1999. Characterization of *npas3*, a novel basic helix-loop-helix PAS gene expressed in the developing mouse nervous system. *Mech Dev* **88**:237-241.
- Bulger, M., and M. Groudine. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**:327-339.
- Burbano, H. A., R. E. Green, T. Maricic, C. Lalueza-Fox, M. de la Rasilla, A. Rosas, J. Kelso, K. S. Pollard, M. Lachmann, and S. Paabo. 2012. Analysis of human accelerated DNA regions using archaic hominin genomes. *PLoS One* **7**:e32877.
- Bush, E. C., and B. T. Lahn. 2008. A genome-wide screen for noncoding elements important in primate evolution. *BMC Evol Biol* **8**:17.
- Butler, J. E., and J. T. Kadonaga. 2002. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* **16**:2583-2592.
- C. Armour, E. B., J. McGowan-Jordan. 2013. Intragenic deletion of NPAS3 in a child with developmental delay. Presented at the 63rd Annual Meeting of The American Society of Human Genetic, Boston, MA.
- Cann, R. L., M. Stoneking, and A. C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* **325**:31-36.
- Capra, J. A., G. D. Erwin, G. McKinsey, J. L. Rubenstein, and K. S. Pollard. 2013a. Many human accelerated regions are developmental enhancers. *Philos Trans R Soc Lond B Biol Sci* **368**:20130025.
- Capra, J. A., M. J. Hubisz, D. Kostka, K. S. Pollard, and A. Siepel. 2013b. A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS Genet* **9**:e1003684.
- Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**:25-36.
- 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.

- Consortium, T. C. S. a. A. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**:69-87.
- Crespi, S. a. D. 2007. Adaptive evolution of genes underlying schizophrenia. *Proc. R. Soc. B* **274**, **2801-2810**.
- Cretekos, C. J., Y. Wang, E. D. Green, J. F. Martin, J. J. t. Rasweiler, and R. R. Behringer. 2008. Regulatory divergence modifies limb length between mammals. *Genes Dev* **22**:141-151.
- Crossley, P. H., and G. R. Martin. 1995. The mouse *Fgf8* gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**:439-451.
- Chambers, E. V., W. A. Bickmore, and C. A. Semple. 2013. Divergence of mammalian higher order chromatin structure is associated with developmental loci. *PLoS Comput Biol* **9**:e1003017.
- Chung, S., C. Chavez, and D. J. Andrew. 2011. Trachealess (*Trh*) regulates all tracheal genes during *Drosophila* embryogenesis. *Dev Biol* **360**:160-172.
- Darwin, C. 2007. El origen de las especies. Centro Editor de Cultura.
- Davies, J. A. 2005. Branching Morphogenesis. Springer Science+Business Media. New York.
- de la Calle-Mustienes, E., C. G. Feijoo, M. Manzanares, J. J. Tena, E. Rodriguez-Seguel, A. Letizia, M. L. Allende, and J. L. Gomez-Skarmeta. 2005. A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. *Genome Res* **15**:1061-1072.
- de Souza, F. S., L. F. Franchini, and M. Rubinstein. 2013. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* **30**:1239-1251.
- Deaner, R. O., K. Isler, J. Burkart, and C. van Schaik. 2007. Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain Behav Evol* **70**:115-124.
- Dekker, J., M. A. Marti-Renom, and L. A. Mirny. 2013. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* **14**:390-403.
- Dennis, M. Y., X. Nuttle, P. H. Sudmant, F. Antonacci, T. A. Graves, M. Nefedov, J. A. Rosenfeld, S. Sajadian, M. Malig, H. Kotkiewicz, C. J. Curry, S. Shafer, L. G. Shaffer, P. J. de Jong, R. K. Wilson, and E. E. Eichler. 2012. Evolution of human-specific neural *SRGAP2* genes by incomplete segmental duplication. *Cell* **149**:912-922.
- Dimitrieva, S., and P. Bucher. 2012. UCNEbase--a database of ultraconserved non-coding elements and genomic regulatory blocks. *Nucleic Acids Res* **41**:D101-109.
- Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**:376-380.
- Doolittle, W. F., and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**:601-603.
- Dumas, L., and J. M. Sikela. 2009. DUF1220 domains, cognitive disease, and human brain evolution. *Cold Spring Harb Symp Quant Biol* **74**:375-382.
- Duret, L., and P. F. Arndt. 2008. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet* **4**:e1000071.
- Duret, L., and N. Galtier. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet* **10**:285-311.
- Duret, L., M. Semon, G. Piganeau, D. Mouchiroud, and N. Galtier. 2002. Vanishing GC-rich isochores in mammalian genomes. *Genetics* **162**:1837-1847.
- Elgar, G., and T. Vavouri. 2008. Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet* **24**:344-352.
- Ema, M., M. Morita, S. Ikawa, M. Tanaka, Y. Matsuda, O. Gotoh, Y. Saijoh, H. Fujii, H. Hamada, Y. Kikuchi, and Y. Fujii-Kuriyama. 1996. Two new members of the murine *Sim* gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis. *Mol Cell Biol* **16**:5865-5875.

- Enard, W., M. Przeworski, S. E. Fisher, C. S. Lai, V. Wiebe, T. Kitano, A. P. Monaco, and S. Paabo. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**:869-872.
- Erbel-Sieler, C., C. Dudley, Y. Zhou, X. Wu, S. J. Estill, T. Han, R. Diaz-Arrastia, E. W. Brunskill, S. S. Potter, and S. L. McKnight. 2004. Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. *Proc Natl Acad Sci U S A* **101**:13648-13653.
- Ertzer, R., F. Muller, Y. Hadzhiev, S. Rathnam, N. Fischer, S. Rastegar, and U. Strahle. 2007. Cooperation of sonic hedgehog enhancers in midline expression. *Dev Biol* **301**:578-589.
- Erwin, D. H., and E. H. Davidson. 2009. The evolution of hierarchical gene regulatory networks. *Nat Rev Genet* **10**:141-148.
- Evans, P. D., J. R. Anderson, E. J. Vallender, S. S. Choi, and B. T. Lahn. 2004a. Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. *Hum Mol Genet* **13**:1139-1145.
- Evans, P. D., J. R. Anderson, E. J. Vallender, S. L. Gilbert, C. M. Malcom, S. Dorus, and B. T. Lahn. 2004b. Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum Mol Genet* **13**:489-494.
- Evans, P. D., S. L. Gilbert, N. Mekel-Bobrov, E. J. Vallender, J. R. Anderson, L. M. Vaez-Azizi, S. A. Tishkoff, R. R. Hudson, and B. T. Lahn. 2005. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science* **309**:1717-1720.
- Evans, P. D., E. J. Vallender, and B. T. Lahn. 2006. Molecular evolution of the brain size regulator genes CDK5RAP2 and CENPJ. *Gene* **375**:75-79.
- Eyre-Walker, A., and L. D. Hurst. 2001. The evolution of isochores. *Nat Rev Genet* **2**:549-555.
- Fisher, S., E. A. Grice, R. M. Vinton, S. L. Bessling, and A. S. McCallion. 2006a. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* **312**:276-279.
- Fisher, S., E. A. Grice, R. M. Vinton, S. L. Bessling, A. Urasaki, K. Kawakami, and A. S. McCallion. 2006b. Evaluating the biological relevance of putative enhancers using Tol2 transposon-mediated transgenesis in zebrafish. *Nat Protoc* **1**:1297-1305.
- Fleagle, J. 2013. *Primate Adaptation and Evolution*. Academic Press, Waltham.
- Foley, R. L. a. R. A. 2004. *Principles of Human Evolution*. Blackwell Publishing Company, Malden.
- Frahm, H. D., H. Stephan, and M. Stephan. 1982. Comparison of brain structure volumes in Insectivora and Primates. I. Neocortex. *J Hirnforsch* **23**:375-389.
- Franchini, L. F., R. Lopez-Leal, S. Nasif, P. Beati, D. M. Gelman, M. J. Low, F. J. de Souza, and M. Rubinstein. 2011. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. *Proc Natl Acad Sci U S A* **108**:15270-15275.
- Frankel, N. 2013. Multiple layers of complexity in cis-regulatory regions of developmental genes. *Dev Dyn* **241**:1857-1866.
- Frankel, N., G. K. Davis, D. Vargas, S. Wang, F. Payre, and D. L. Stern. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**:490-493.
- Frederic, M. Y., V. F. Lundin, M. D. Whiteside, J. G. Cueva, D. K. Tu, S. Y. Kang, H. Singh, D. L. Baillie, H. Hutter, M. B. Goodman, F. S. Brinkman, and M. R. Leroux. 2013. Identification of 526 Conserved Metazoan Genetic Innovations Exposes a New Role for Cofactor E-like in Neuronal Microtubule Homeostasis. *PLoS Genet* **9**:e1003804.
- Fritsch, B., K. W. Beisel, S. Pauley, and G. Soukup. 2007a/b. Molecular evolution of the vertebrate mechanosensory cell year. *Int J Dev Biol* **51**:663-678.
- Gibcus, J. H., and J. Dekker. 2013. The hierarchy of the 3D genome. *Mol Cell* **49**:773-782.
- Gilbert, S. F. 2005. *Biología del desarrollo*. Ed. Médica Panamericana, Buenos Aires.
- Glass, C. K., and M. G. Rosenfeld. 2000. The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* **14**:121-141.

- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**:725-736.
- Gould, P., and D. Kamnasaran. 2011. Immunohistochemical analyses of NPAS3 expression in the developing human fetal brain. *Anat Histol Embryol* **40**:196-203.
- Gould, S. J. 2010a. *Ontogenia y Filogenia: La ley fundamental biogenética*. Crítica.
- Gould, S. J. 2006. *La Vida Maravillosa*. Crítica.
- Gould, S. J. 2004. *Dientes de gallina y dedos de caballo*. Crítica, Barcelona.
- Gould, S. J. 2010b. *Dientes de gallina y dedos de caballo. Reflexiones sobre historia natural*. Drakontos Bolsillo.
- Graur, D., Y. Zheng, N. Price, R. B. Azevedo, R. A. Zufall, and E. Elhaik. 2013. On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol* **5**:578-590.
- Hall, B. K. 2007. *Fins into Limbs: Evolution, Development, and Transformation*. University of Chicago Press.
- Hall, B. K. 2003. Evo-Devo: evolutionary developmental mechanisms. *Int J Dev Biol* **47**:491-495.
- Hatzis, P., and I. Talianidis. 2002. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol Cell* **10**:1467-1477.
- Hawks, J., K. Hunley, S. H. Lee, and M. Wolpoff. 2000. Population bottlenecks and Pleistocene human evolution. *Mol Biol Evol* **17**:2-22.
- Heintzman, N. D., and B. Ren. 2009. Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev* **19**:541-549.
- Heintzman, N. D., R. K. Stuart, G. Hon, Y. Fu, C. W. Ching, R. D. Hawkins, L. O. Barrera, S. Van Calcar, C. Qu, K. A. Ching, W. Wang, Z. Weng, R. D. Green, G. E. Crawford, and B. Ren. 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.
- Herculano-Houzel, S. 2011. Not all brains are made the same: new views on brain scaling in evolution. *Brain Behav Evol* **78**:22-36.
- Herculano-Houzel, S., C. E. Collins, P. Wong, and J. H. Kaas. 2007. Cellular scaling rules for primate brains. *Proc Natl Acad Sci U S A* **104**:3562-3567.
- Herculano-Houzel, S., and J. H. Kaas. 2011. Gorilla and orangutan brains conform to the primate cellular scaling rules: implications for human evolution. *Brain Behav Evol* **77**:33-44.
- Herculano-Houzel, S., B. Mota, and R. Lent. 2006. Cellular scaling rules for rodent brains. *Proc Natl Acad Sci U S A* **103**:12138-12143.
- Hinchliffe, J. R. 2002. Developmental basis of limb evolution. *Int J Dev Biol* **46**:835-845.
- Hofman, M. A. 1983. Encephalization in hominids: evidence for the model of punctuationalism. *Brain Behav Evol* **22**:102-117.
- Hofman, M. A. 1988. Size and shape of the cerebral cortex in mammals. II. The cortical volume. *Brain Behav Evol* **32**:17-26.
- Hofman, M. A. 1985. Size and shape of the cerebral cortex in mammals. I. The cortical surface. *Brain Behav Evol* **27**:28-40.
- Hong, J. W., D. A. Hendrix, and M. S. Levine. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* **321**:1314.
- Hopkins, R., and M. D. Rausher. 2011. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature* **469**:411-414.
- Huang, J., R. H. Perlis, P. H. Lee, A. J. Rush, M. Fava, G. S. Sachs, J. Lieberman, S. P. Hamilton, P. Sullivan, P. Sklar, S. Purcell, and J. W. Smoller. 2010. Cross-disorder genomewide analysis of schizophrenia, bipolar disorder, and depression. *Am J Psychiatry* **167**:1254-1263.
- Irie, N., and S. Kuratani. 2010. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat Commun* **2**:248.

- Irimia, M., I. Maeso, S. W. Roy, and H. B. Fraser. 2013. Ancient cis-regulatory constraints and the evolution of genome architecture. *Trends Genet* **29**:521-528.
- Irimia, M., J. J. Tena, M. S. Alexis, A. Fernandez-Minan, I. Maeso, O. Bogdanovic, E. de la Calle-Mustienes, S. W. Roy, J. L. Gomez-Skarmeta, and H. B. Fraser. 2012. Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res* **22**:2356-2367.
- Jacob, F., and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**:318-356.
- Johanson, D. C., and T. D. White. 1979. A systematic assessment of early African hominids. *Science* **203**:321-330.
- Jolly, A. 1985. *The Evolution of Primate Behavior*. Macmillan Series in Physical Anthropology.
- Kadonaga, J. T. 1998. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**:307-313.
- Kaindl, A. M., S. Passemard, P. Kumar, N. Kraemer, L. Issa, A. Zwirner, B. Gerard, A. Verloes, S. Mani, and P. Gressens. 2010. Many roads lead to primary autosomal recessive microcephaly. *Prog Neurobiol* **90**:363-383.
- Kamm, G. B., R. Lopez-Leal, J. R. Lorenzo, and L. F. Franchini. 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., F. Pisciotto, R. Kliger, and L. F. Franchini. 2013b. The Developmental Brain Gene NPAS3 Contains the Largest Number of Accelerated Regulatory Sequences in the Human Genome. *Mol Biol Evol*.
- Kamnasaran, D., W. J. Muir, M. A. Ferguson-Smith, and D. W. Cox. 2003. Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. *J Med Genet* **40**:325-332.
- Kardong, K. V. 2007a. *Vertebrados: anatomía comparada, función and evolución*. Mc Graw-Hill-Interamericana de eEspaña, Madrid.
- Kardong, K. V. 2007b. *Vertebrados: anatomía comparada, función y evolución*. . Mc Graw-Hill-Interamericana de España.Madrid.
- Karlsson, J., J. von Hofsten, and P. E. Olsson. 2001. Generating transparent zebrafish: a refined method to improve detection of gene expression during embryonic development. *Mar Biotechnol (NY)* **3**:522-527.
- Kawakami, K., and A. Shima. 1999. Identification of the Tol2 transposase of the medaka fish *Oryzias latipes* that catalyzes excision of a nonautonomous Tol2 element in zebrafish *Danio rerio*. *Gene* **240**:239-244.
- Kawakami, K., A. Shima, and N. Kawakami. 2000. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc Natl Acad Sci U S A* **97**:11403-11408.
- Kewley, R. J., M. L. Whitelaw, and A. Chapman-Smith. 2004. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int J Biochem Cell Biol* **36**:189-204.
- Kikuta, H., M. Laplante, P. Navratilova, A. Z. Komisarczuk, P. G. Engstrom, D. Fredman, A. Akalin, M. Caccamo, I. Sealy, K. Howe, J. Ghislain, G. Pezeron, P. Mourrain, S. Ellingsen, A. C. Oates, C. Thisse, B. Thisse, I. Foucher, B. Adolf, A. Geling, B. Lenhard, and T. S. Becker. 2007. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res* **17**:545-555.
- Kimmel, C. B., W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling. 1995. Stages of embryonic development of the zebrafish. *Dev Dyn* **203**:253-310.
- Kimmel, C. B., B. Ullmann, M. Walker, C. T. Miller, and J. G. Crump. 2003. Endothelin 1-mediated regulation of pharyngeal bone development in zebrafish. *Development* **130**:1339-1351.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge.
- Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* **217**:624-26.

- King, M. C. a. W. A. C. 1975. Evolution at two levels in human and chimpanzees. Pp. 107-116. Science.
- Kingston, R. E., and G. J. Narlikar. 1999. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev* **13**:2339-2352.
- Komisarczuk, A. Z., K. Kawakami, and T. S. Becker. 2009. Cis-regulation and chromosomal rearrangement of the *fgf8* locus after the teleost/tetrapod split. *Dev Biol* **336**:301-312.
- Kosiol, C., T. Vinar, R. R. da Fonseca, M. J. Hubisz, C. D. Bustamante, R. Nielsen, and A. Siepel. 2008. Patterns of positive selection in six Mammalian genomes. *PLoS Genet* **4**:e1000144.
- Kostka, D., M. J. Hubisz, A. Siepel, and K. S. Pollard. 2012. The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. *Mol Biol Evol* **29**:1047-1057.
- Kothary, R., S. Clapoff, S. Darling, M. D. Perry, L. A. Moran, and J. Rossant. 1989. Inducible expression of an *hsp68-lacZ* hybrid gene in transgenic mice. *Development* **105**:707-714.
- Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. *Annu Rev Genomics Hum Genet* **1**:539-559.
- Krivega, I., and A. Dean. 2012. Enhancer and promoter interactions-long distance calls. *Curr Opin Genet Dev* **22**:79-85.
- Kudla, G., L. Lipinski, F. Caffin, A. Helwak, and M. Zylicz. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol* **4**:e180.
- Kumar, A., S. C. Girimaji, M. R. Duvvari, and S. H. Blanton. 2009. Mutations in STIL, encoding a pericentriolar and centrosomal protein, cause primary microcephaly. *Am J Hum Genet* **84**:286-290.
- Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917-920.
- Lahr, M. M., and R. A. Foley. 1998. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am J Phys Anthropol* **Suppl 27**:137-176.
- Lamb, E. J. y. M. J. 2013. Evolución en Cuatro Dimensiones: genética, epigenética, comportamientos y variación simbólica en la historia de la vida. Capital Intelectual, Buenos Aires.
- Lander, E. S.L. M. LintonB. BirrenC. NusbaumM. C. ZodyJ. BaldwinK. DevonK. DewarM. DoyleW. FitzHughR. FunkeD. GageK. HarrisA. HeafordJ. HowlandL. KannJ. LehoczkyR. LeVineP. McEwanK. McKernanJ. MeldrimJ. P. MesirovC. MirandaW. MorrisJ. NaylorC. RaymondM. RosettiR. SantosA. SheridanC. SougnezN. Stange-ThomannN. StojanovicA. SubramanianD. WymanJ. RogersJ. SulstonR. AinscoughS. BeckD. BentleyJ. BurtonC. CleeN. CarterA. CoulsonR. DeadmanP. DeloukasA. DunhamI. DunhamR. DurbinL. FrenchD. GrafhamS. GregoryT. HubbardS. HumphrayA. HuntM. JonesC. LloydA. McMurrayL. MatthewsS. MercerS. MilneJ. C. MullikinA. MungallR. PlumbM. RossR. ShownkeenS. SimsR. H. WaterstonR. K. WilsonL. W. HillierJ. D. McPhersonM. A. MarraE. R. MardisL. A. FultonA. T. ChinwallaK. H. PepinW. R. GishS. L. ChissoeM. C. WendlK. D. DelehauntyT. L. MinerA. DelehauntyJ. B. KramerL. L. CookR. S. FultonD. L. JohnsonP. J. MinxS. W. CliftonT. HawkinsE. BranscombP. PredkiP. RichardsonS. WenningT. SlezakN. DoggettJ. F. ChengA. OlsenS. LucasC. ElkinE. UberbacherM. FrazierR. A. GibbsD. M. MuznyS. E. SchererJ. B. BouckE. J. SodergrenK. C. WorleyC. M. RivesJ. H. GorrellM. L. MetzkerS. L. NaylorR. S. KucherlapatiD. L. NelsonG. M. WeinstockY. SakakiA. FujiyamaM. HattoriT. YadaA. ToyodaT. ItohC. KawagoeH. WatanabeY. TotokiT. TaylorJ. WeissenbachR. HeiligW. SaurinF. ArtiguenaveP. BrottierT. BrulsE. PelletierC. RobertP. WinckerD. R. SmithL. Doucette-StammM. RubenfieldK. WeinstockH. M. LeeJ. DuboisA. RosenthalM. PlatzerG. NyakaturaS. TaudienA. RumpH. YangJ. YuJ. WangG. HuangJ. GuL. HoodL. RowenA. MadanS. QinR. W. DavisN. A. FederspielA. P. AbolaM. J. ProctorR. M. MyersJ. SchmutzM. DicksonJ. GrimwoodD. R. CoxM. V. OlsonR. KaulN. ShimizuK. KawasakiS. MinoshimaG. A. EvansM. AthanasiouR.

- Schultz B. A. Roe F. Chen H. Pan J. Ramser H. Lehrach R. Reinhardt W. R. McCombie M. de la Bastide N. Dedhia H. Blocker K. Hornischer G. Nordsiek R. Agarwala L. Aravind J. A. Bailey A. Bateman S. Batzoglou E. Birney P. Bork D. G. Brown C. B. Burge L. Cerutti H. C. Chen D. Church M. Clamp R. R. Copley T. Doerks S. R. Eddy E. E. Eichler T. S. Furey J. Galagan J. G. Gilbert C. Harmon Y. Hayashizaki D. Haussler H. Hermjakob K. Hokamp W. Jang L. S. Johnson T. A. Jones S. Kasif A. Kasprzyk S. Kennedy W. J. Kent P. Kitts E. V. Koonin I. Korf D. Kulp D. Lancet T. M. Lowe A. McLysaght T. Mikkelsen J. V. Moran N. Mulder V. J. Pollara C. P. Ponting G. Schuler J. Schultz G. Slater A. F. Smit E. Stupka J. Szustakowski D. Thierry-Mieg J. Thierry-Mieg L. Wagner J. Wallis R. Wheeler A. Williams Y. I. Wolf K. H. Wolfe S. P. Yang R. F. Yeh F. Collins M. S. Guyer J. Peterson A. Felsenfeld K. A. Wetterstrand A. Patrinos M. J. Morgan P. de Jong J. J. Catanese K. Osoegawa H. Shizuya S. Choi, and Y. J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921.
- Laue, K., S. Daujat, J. G. Crump, N. Plaster, H. H. Roehl, C. B. Kimmel, R. Schneider, and M. Hammerschmidt. 2008. The multidomain protein Brpf1 binds histones and is required for Hox gene expression and segmental identity. *Development* **135**:1935-1946.
- Leakey, L. S., P. V. Tobias, and J. R. Napier. 1964. A New Species of the Genus Homo from Olduvai Gorge. *Nature* **202**:7-9.
- Leamey, C. A., K. A. Glendining, G. Kreiman, N. D. Kang, K. H. Wang, R. Fassler, A. Sawatari, S. Tonegawa, and M. Sur. 2008. Differential gene expression between sensory neocortical areas: potential roles for Ten_m3 and Bcl6 in patterning visual and somatosensory pathways. *Cereb Cortex* **18**:53-66.
- Lee, T. I., and R. A. Young. 2000. Transcription of eukaryotic protein-coding genes. *Annu Rev Genet* **34**:77-137.
- Lemon, B., and R. Tjian. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev* **14**:2551-2569.
- Lettice, L. A., S. J. Heaney, L. A. Purdie, L. Li, P. de Beer, B. A. Oostra, D. Goode, G. Elgar, R. E. Hill, and E. de Graaff. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**:1725-1735.
- Lewin, R. 1997. *Bones of Contention*. The University of Chicago Press, London.
- Lin, Y., B. L. Bloodgood, J. L. Hauser, A. D. Lapan, A. C. Koon, T. K. Kim, L. S. Hu, A. N. Malik, and M. E. Greenberg. 2008. Activity-dependent regulation of inhibitory synapse development by Npas4. *Nature* **455**:1198-1204.
- Lindblad-Toh, K., M. Garber, O. Zuk, M. F. Lin, B. J. Parker, S. Washietl, P. Kheradpour, J. Ernst, G. Jordan, E. Mauceli, L. D. Ward, C. B. Lowe, A. K. Holloway, M. Clamp, S. Gnerre, J. Alföldi, K. Beal, J. Chang, H. Clawson, J. Cuff, F. Di Palma, S. Fitzgerald, P. Flicek, M. Guttman, M. J. Hubisz, D. B. Jaffe, I. Jungreis, W. J. Kent, D. Kostka, M. Lara, A. L. Martins, T. Massingham, I. Moltke, B. J. Raney, M. D. Rasmussen, J. Robinson, A. Stark, A. J. Vilella, J. Wen, X. Xie, M. C. Zody, J. Baldwin, T. Bloom, C. W. Chin, D. Heiman, R. Nicol, C. Nusbaum, S. Young, J. Wilkinson, K. C. Worley, C. L. Kovar, D. M. Muzny, R. A. Gibbs, A. Cree, H. H. Dihn, G. Fowler, S. Jhangiani, V. Joshi, S. Lee, L. R. Lewis, L. V. Nazareth, G. Okwuonu, J. Santibanez, W. C. Warren, E. R. Mardis, G. M. Weinstock, R. K. Wilson, K. Delehaunty, D. Dooling, C. Fronik, L. Fulton, B. Fulton, T. Graves, P. Minx, E. Sodergren, E. Birney, E. H. Margulies, J. Herrero, E. D. Green, D. Haussler, A. Siepel, N. Goldman, K. S. Pollard, J. S. Pedersen, E. S. Lander, and M. Kellis. 2011. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**:476-482.
- Lindebro, M., Poellinger, L., & Whitelaw, M. . 1995. Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. . *EMBO Journal* **14**,3528-3539.
- Lomvardas, S., G. Barnea, D. J. Pisapia, M. Mendelsohn, J. Kirkland, and R. Axel. 2006. Interchromosomal interactions and olfactory receptor choice. *Cell* **126**:403-413.

- Longo, S., M. Riccio, and A. R. McCune. 2013. Homology of lungs and gas bladders: insights from arterial vasculature. *J Morphol* **274**:687-703.
- Lordkipanidze, D., M. S. Ponce de Leon, A. Margvelashvili, Y. Rak, G. P. Rightmire, A. Vekua, and C. P. Zollikofer. 2013. A complete skull from Dmanisi, Georgia, and the evolutionary biology of early Homo. *Science* **342**:326-331.
- Lowe, C. B., G. Bejerano, and D. Haussler. 2007. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A* **104**:8005-8010.
- Lowe, C. B., and D. Haussler. 2012. 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. *PLoS One* **7**:e43128.
- Lowe, C. J., M. Wu, A. Salic, L. Evans, E. Lander, N. Stange-Thomann, C. E. Gruber, J. Gerhart, and M. Kirschner. 2003. Anteroposterior patterning in hemichordates and the origins of the chordate nervous system. *Cell* **113**:853-865.
- Macintyre, G., T. Alford, L. Xiong, G. A. Rouleau, P. G. Tibbo, and D. W. Cox. 2010. Association of NPAS3 exonic variation with schizophrenia. *Schizophr Res* **120**:143-149.
- Marinic, M., T. Aktas, S. Ruf, and F. Spitz. 2013. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* **24**:530-542.
- Marino, L. 1998. A comparison of encephalization between odontocete cetaceans and anthropoid primates. *Brain Behav Evol* **51**:230-238.
- Markham, K., C. Schuurmans, and S. Weiss. 2007. STAT5A/B activity is required in the developing forebrain and spinal cord. *Mol Cell Neurosci* **35**:272-282.
- Martin, R. D. 1990. *Primate Origins and Evolution: A Phylogenetic Reconstruction*. Princeton University Press.
- Martin, R. D. 1986. *Primates: a definition*. Cambridge University Press.
- Marygold, S. J., P. C. Leyland, R. L. Seal, J. L. Goodman, J. Thurmond, V. B. Strelets, and R. J. Wilson. 2013. FlyBase: improvements to the bibliography. *Nucleic Acids Res* **41**:D751-757.
- Massari, M. E., and C. Murre. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20**:429-440.
- Maston, G. A., S. K. Evans, and M. R. Green. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**:29-59.
- Matys, V., E. Fricke, R. Geffers, E. Gossling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, D. U. Kloos, S. Land, B. Lewicki-Potapov, H. Michael, R. Munch, I. Reuter, S. Rotert, H. Saxel, M. Scheer, S. Thiele, and E. Wingender. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31**:374-378.
- Mayr, E. 2001. *What is Evolution*. Phoenix.
- McEwen, G. K., A. Woolfe, D. Goode, T. Vavouri, H. Callaway, and G. Elgar. 2006. Ancient duplicated conserved noncoding elements in vertebrates: a genomic and functional analysis. *Genome Res* **16**:451-465.
- McLean, C. Y., P. L. Reno, A. A. Pollen, A. I. Bassan, T. D. Capellini, C. Guenther, V. B. Indjeian, X. Lim, D. B. Menke, B. T. Schaar, A. M. Wenger, G. Bejerano, and D. M. Kingsley. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**:216-219.
- McHenry, H. M. 1982. The Pattern of Human Evolution: Studies on Bipedalism, Mastication, and Encephalization. *Annual Review of Anthropology* **Vol. 11**: 151-173.
- McHenry, H. M. 1975. Fossils and the mosaic nature of human evolution. *Science* **190**:425-431.
- Mekel-Bobrov, N., S. L. Gilbert, P. D. Evans, E. J. Vallender, J. R. Anderson, R. R. Hudson, S. A. Tishkoff, and B. T. Lahn. 2005. Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens. *Science* **309**:1720-1722.
- Meyer, M., M. Kircher, M. T. Gansauge, H. Li, F. Racimo, S. Mallick, J. G. Schraiber, F. Jay, K. Prufer, C. de Filippo, P. H. Sudmant, C. Alkan, Q. Fu, R. Do, N. Rohland, A. Tandon, M. Siebauer, R. E. Green, K. Bryc, A. W. Briggs, U. Stenzel, J. Dabney, J. Shendure, J.

- Kitzman, M. F. Hammer, M. V. Shunkov, A. P. Derevianko, N. Patterson, A. M. Andres, E. E. Eichler, M. Slatkin, D. Reich, J. Kelso, and S. Paabo. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**:222-226.
- Miller, K. a. 2006. Resolving the paradox of common, harmful, heritable mental disorders: Which evolutionary genetic models work best? *BEHAVIORAL AND BRAIN SCIENCES* **29**, 385–452.
- Moalic, L. S., Lepagnol-Bestel, Ramoz, Loe-Mie, Maussion, Gorwood and Simonneau. 2010. Primate-Accelerated Evolutionary Genes: Novel Routes to Drug Discovery in Psychiatric Disorders. *Current Medicinal Chemistry* **17**, 1300-1316.
- Monod, J., and F. Jacob. 1961. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harb Symp Quant Biol* **26**:389-401.
- Montgomery, S. H., I. Capellini, R. A. Barton, and N. I. Mundy. 2010. Reconstructing the ups and downs of primate brain evolution: implications for adaptive hypotheses and *Homo floresiensis*. *BMC Biol* **8**:9.
- Muchmore, E. A., S. Diaz, and A. Varki. 1998. A structural difference between the cell surfaces of humans and the great apes. *Am J Phys Anthropol* **107**:187-198.
- Nadeau, J. H., and B. A. Taylor. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci U S A* **81**:814-818.
- Nasevicius, A., and S. C. Ekker. 2000. Effective targeted gene 'knockdown' in zebrafish. *Nat Genet* **26**:216-220.
- Nei, M., and A. P. Rooney. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* **39**:121-152.
- Nielsen, R., and J. P. Huelsenbeck. 2002. Detecting positively selected amino acid sites using posterior predictive P-values. *Pac Symp Biocomput*:576-588.
- Noonan, J. P. 2009. Regulatory DNAs and the evolution of human development. *Curr Opin Genet Dev* **19**:557-564.
- Noonan, J. P., and A. S. McCallion. 2010. Genomics of long-range regulatory elements. *Annu Rev Genomics Hum Genet* **11**:1-23.
- Nora, E. P., B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, and E. Heard. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**:381-385.
- Nusslein-Volhard, C. a. D., R. 2002. *Zebrafish: A Practical Approach*. Oxford University Press, Oxford.
- O'Bleness, M., V. B. Searles, A. Varki, P. Gagneux, and J. M. Sikela. 2012. Evolution of genetic and genomic features unique to the human lineage. *Nat Rev Genet* **13**:853-866.
- Ohno, S. 1970. *Evolution by gene duplication*. Springer, Berlin.
- Oksenberg, N., L. Stevison, J. D. Wall, and N. Ahituv. 2013. Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genet* **9**:e1003221.
- Ong, C. T., and V. G. Corces. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* **12**:283-293.
- Orgel, L. E., and F. H. Crick. 1980. Selfish DNA: the ultimate parasite. *Nature* **284**:604-607.
- Ostrander, G. K. 2000. *The Laboratory Fish*. Academic Press Inc
- Ovcharenko, I., G. G. Loots, M. A. Nobrega, R. C. Hardison, W. Miller, and L. Stubbs. 2005. Evolution and functional classification of vertebrate gene deserts. *Genome Res* **15**:137-145.
- Parichy, D. M., M. R. Elizondo, M. G. Mills, T. N. Gordon, and R. E. Engeszer. 2009. Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. *Dev Dyn* **238**:2975-3015.
- Park, H. C., S. K. Hong, H. S. Kim, S. H. Kim, E. J. Yoon, C. H. Kim, N. Miki, and T. L. Huh. 2000. Structural comparison of zebrafish Elav/Hu and their differential expressions during neurogenesis. *Neurosci Lett* **279**:81-84.

- Pennacchio, L. A., N. Ahituv, A. M. Moses, S. Prabhakar, M. A. Nobrega, M. Shoukry, S. Minovitsky, I. Dubchak, A. Holt, K. D. Lewis, I. Plajzer-Frick, J. Akiyama, S. De Val, V. Afzal, B. L. Black, O. Couronne, M. B. Eisen, A. Visel, and E. M. Rubin. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**:499-502.
- Perry, M. W., A. N. Boettiger, J. P. Bothma, and M. Levine. 2010. Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr Biol* **20**:1562-1567.
- Peter, I. S., and E. H. Davidson. 2011. Evolution of gene regulatory networks controlling body plan development. *Cell* **144**:970-985.
- Pickard, B. S., A. Christoforou, P. A. Thomson, A. Fawkes, K. L. Evans, S. W. Morris, D. J. Porteous, D. H. Blackwood, and W. J. Muir. 2009. Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. *Mol Psychiatry* **14**:874-884.
- Pickard, B. S., M. P. Malloy, D. J. Porteous, D. H. Blackwood, and W. J. Muir. 2005. Disruption of a brain transcription factor, NPAS3, is associated with schizophrenia and learning disability. *Am J Med Genet B Neuropsychiatr Genet* **136B**:26-32.
- Pickard, B. S., A. A. Pieper, D. J. Porteous, D. H. Blackwood, and W. J. Muir. 2006. The NPAS3 gene--emerging evidence for a role in psychiatric illness. *Ann Med* **38**:439-448.
- Pieper, A. A., X. Wu, T. W. Han, S. J. Estill, Q. Dang, L. C. Wu, S. Reece-Fincannon, C. A. Dudley, J. A. Richardson, D. J. Brat, and S. L. McKnight. 2005. The neuronal PAS domain protein 3 transcription factor controls FGF-mediated adult hippocampal neurogenesis in mice. *Proc Natl Acad Sci U S A* **102**:14052-14057.
- Pollard, K. S., S. R. Salama, B. King, A. D. Kern, T. Dreszer, S. Katzman, A. Siepel, J. S. Pedersen, G. Bejerano, R. Baertsch, K. R. Rosenbloom, J. Kent, and D. Haussler. 2006. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* **2**:e168.
- Poulin, F., M. A. Nobrega, I. Plajzer-Frick, A. Holt, V. Afzal, E. M. Rubin, and L. A. Pennacchio. 2005. In vivo characterization of a vertebrate ultraconserved enhancer. *Genomics* **85**:774-781.
- Prabhakar, S., J. P. Noonan, S. Paabo, and E. M. Rubin. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**:786.
- Prud'homme, B., N. Gompel, and S. B. Carroll. 2007. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A* **104 Suppl 1**:8605-8612.
- Prufer, K., F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E. Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, and S. Paabo. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**:43-49.
- Pruitt, K. D., T. Tatusova, and D. R. Maglott. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**:D501-504.
- Purvis, A. 1995. A composite estimate of primate phylogeny. *Philos Trans R Soc Lond B Biol Sci* **348**:405-421.
- Rakic, P. 2000. Radial unit hypothesis of neocortical expansion. *Novartis Found Symp* **228**:30-42; discussion 42-52.
- Reich, D., R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo, M. V. Shunkov, A. P. Derevianko, J. J. Hublin, J. Kelso, M. Slatkin, and S. Paabo. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**:1053-1060.
- Rilling, J. K., and T. R. Insel. 1999. The primate neocortex in comparative perspective using magnetic resonance imaging. *J Hum Evol* **37**:191-223.

- Ritter, D. I., Q. Li, D. Kostka, K. S. Pollard, S. Guo, and J. H. Chuang. 2010. The importance of being cis: evolution of orthologous fish and mammalian enhancer activity. *Mol Biol Evol* **27**:2322-2332.
- Rombough, P. 2002. Gills are needed for ionoregulation before they are needed for O₂ uptake in developing zebrafish, *Danio rerio*. *J Exp Biol* **205**:1787-1794.
- Romero, I. G., I. Ruvinsky, and Y. Gilad. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet* **13**:505-516.
- Rosenfeld, M. G., V. V. Lunnyak, and C. K. Glass. 2006. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev* **20**:1405-1428.
- Rouleux-Bonnin, S. B. a. F. 2011. Transposable Elements as Tools for Reshaping the Genome: It Is a Huge World After All! *Mobile Genetic Elements: Protocols and Genomic Applications, Methods in Molecular Biology*, vol. **859**,
- Rubinstein, M., and F. S. de Souza. 2013. Evolution of transcriptional enhancers and animal diversity. *Philos Trans R Soc Lond B Biol Sci* **368**:20130017.
- Ryan, J. F., K. Pang, C. E. Schnitzler, A. D. Nguyen, R. T. Moreland, D. K. Simmons, B. J. Koch, W. R. Francis, P. Havlak, S. A. Smith, N. H. Putnam, S. H. Haddock, C. W. Dunn, T. G. Wolfsberg, J. C. Mullikin, M. Q. Martindale, and A. D. Baxevanis. 2013. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**:1242592.
- Sagai, T., M. Hosoya, Y. Mizushina, M. Tamura, and T. Shiroishi. 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.
- Samonte, R. V., and E. E. Eichler. 2002. Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* **3**:65-72.
- Sandelin, A., P. Bailey, S. Bruce, P. G. Engstrom, J. M. Klos, W. W. Wasserman, J. Ericson, and B. Lenhard. 2004. Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **5**:99.
- Shin, J., H. Y. Jeong, K. E. Lee, and J. Kim. 2010. Isolation and Characterization of Chicken NPAS3. *Exp Neurobiol* **19**:71-74.
- Siepel, A., G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, and D. Haussler. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**:1034-1050.
- Siepel A, P. K., Haussler D. . 2006. New methods for detecting lineage-specific selection. . Pp. P. 190-205. *Proceedings of the 10th International Conference on Research in Computational Molecular Biology (RECOMB 2006)*. Venice, Italy.
- Smale, S. T., and J. T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* **72**:449-479.
- Smart, E. J., R. A. De Rose, and S. A. Farber. 2004. Annexin 2-caveolin 1 complex is a target of ezetimibe and regulates intestinal cholesterol transport. *Proc Natl Acad Sci U S A* **101**:3450-3455.
- Smit, and Riggs. 1995. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.* **23**, **98-102**.
- Smit, A., Hubley, R & Green, . 1996-2010 RepeatMasker Open-3.0.
- Smith, A. M., M. J. Sanchez, G. A. Follows, S. Kinston, I. J. Donaldson, A. R. Green, and B. Gottgens. 2008. A novel mode of enhancer evolution: the Tal1 stem cell enhancer recruited a MIR element to specifically boost its activity. *Genome Res* **18**:1422-1432.
- Solomon, K. S., T. Kudoh, I. B. Dawid, and A. Fritz. 2003. Zebrafish foxi1 mediates otic placode formation and jaw development. *Development* **130**:929-940.
- Stephan, H., H. Frahm, and G. Baron. 1981. New and revised data on volumes of brain structures in insectivores and primates. *Folia Primatol (Basel)* **35**:1-29.

- Strahl, B. D., and C. D. Allis. 2000. The language of covalent histone modifications. *Nature* **403**:41-45.
- Strathern, J. N., B. K. Shafer, and C. B. McGill. 1995. DNA synthesis errors associated with double-strand-break repair. *Genetics* **140**:965-972.
- Striedter, G. F. 2005. *Principles Of Brain Evolution*. Sinauer Associates, Incorporated.
- Stringer, C. B., and P. Andrews. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* **239**:1263-1268.
- Studer, A., Q. Zhao, J. Ross-Ibarra, and J. Doebley. 2011. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* **43**:1160-1163.
- Suzuki, Y., and T. Gojobori. 1999. A method for detecting positive selection at single amino acid sites. *Mol Biol Evol* **16**:1315-1328.
- Swami, M. 2010. Transcription: Shadow enhancers confer robustness. *Nat Rev Genet* **11**:454.
- Swanson, C. I., N. C. Evans, and S. Barolo. 2010. Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev Cell* **18**:359-370.
- Tagle, D. A., B. F. Koop, M. Goodman, J. L. Slightom, D. L. Hess, and R. T. Jones. 1988. Embryonic epsilon and gamma globin genes of a prosimian primate (*Galago crassicaudatus*). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. *J Mol Biol* **203**:439-455.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585-595.
- Takahashi, M., and N. Saitou. 2012. Identification and characterization of lineage-specific highly conserved noncoding sequences in Mammalian genomes. *Genome Biol Evol* **4**:641-657.
- Tattersall, I. 2009. Out of Africa: modern human origins special feature: human origins: out of Africa. *Proc Natl Acad Sci U S A* **106**:16018-16021.
- Tattersall, I. 2000. Once We Were Not Alone. *Scientific American* **56. January 2000**:56 - 62.
- Tavare, S., C. R. Marshall, O. Will, C. Soligo, and R. D. Martin. 2002. Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature* **416**:726-729.
- Templeton, A. 2002. Out of Africa again and again. *Nature* **416**:45-51.
- Thorne, A. G., and M. H. Wolpoff. 1992. The multiregional evolution of humans. *Sci Am* **266**:76-79, 82-73.
- Tiberi, L., J. van den Ameele, J. Dimidschstein, J. Piccirilli, D. Gall, A. Herpoel, A. Bilheu, J. Bonnefont, M. Iacovino, M. Kyba, T. Bouschet, and P. Vanderhaeghen. 2012. BCL6 controls neurogenesis through Sirt1-dependent epigenetic repression of selective Notch targets. *Nat Neurosci* **15**:1627-1635.
- Vallender, E. J. 2008. Exploring the origins of the human brain through molecular evolution. *Brain Behav Evol* **72**:168-177.
- Vallender, E. J., N. Mekel-Bobrov, and B. T. Lahn. 2008. Genetic basis of human brain evolution. *Trends Neurosci* **31**:637-644.
- van Heyningen, V., and W. Bickmore. 2013. Regulation from a distance: long-range control of gene expression in development and disease. *Philos Trans R Soc Lond B Biol Sci* **368**:20120372.
- Vaquerizas, J. M., S. K. Kummerfeld, S. A. Teichmann, and N. M. Luscombe. 2009. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**:252-263.
- Varki, O. a. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* **4**, 20-28.
- Vavouri, T., and B. Lehner. 2009. Conserved noncoding elements and the evolution of animal body plans. *Bioessays* **31**:727-735.
- Ventura, M., C. R. Catacchio, S. Sajjadian, L. Vives, P. H. Sudmant, T. Marques-Bonet, T. A. Graves, R. K. Wilson, and E. E. Eichler. 2012. The evolution of African great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome Res* **22**:1036-1049.

- Visel, A., M. J. Blow, Z. Li, T. Zhang, J. A. Akiyama, A. Holt, I. Plajzer-Frick, M. Shoukry, C. Wright, F. Chen, V. Afzal, B. Ren, E. M. Rubin, and L. A. Pennacchio. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**:854-858.
- Waddington, C. H. 1942. Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563-565.
- Wang, G. L., B. H. Jiang, E. A. Rue, and G. L. Semenza. 1995. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc Natl Acad Sci U S A* **92**:5510-5514.
- Wang, X., W. E. Grus, and J. Zhang. 2006. Gene losses during human origins. *PLoS Biol* **4**:e52.
- Weber, H., S. Kittel-Schneider, A. Gessner, K. Domschke, M. Neuner, C. P. Jacob, H. N. Buttenschon, A. Boreatti-Hummer, J. Volkert, S. Herterich, B. T. Baune, S. Gross-Lesch, J. Kopf, S. Kreiker, T. T. Nguyen, L. Weissflog, V. Arolt, O. Mors, J. Deckert, K. P. Lesch, and A. Reif. 2011. Cross-disorder analysis of bipolar risk genes: further evidence of DGKH as a risk gene for bipolar disorder, but also unipolar depression and adult ADHD. *Neuropsychopharmacology* **36**:2076-2085.
- West, A. G., M. Gaszner, and G. Felsenfeld. 2002. Insulators: many functions, many mechanisms. *Genes Dev* **16**:271-288.
- Wilk, R., I. Weizman, and B. Z. Shilo. 1996. tracheless encodes a bHLH-PAS protein that is an inducer of tracheal cell fates in *Drosophila*. *Genes Dev* **10**:93-102.
- Wilson, A. C., and R. L. Cann. 1992. The recent African genesis of humans. *Sci Am* **266**:68-73.
- Wittkopp, P. J., and G. Kalay. 2011. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* **13**:59-69.
- Wolpoff, M. H. 1999. The systematics of *Homo*. *Science* **284**:1774-1775.
- Wood, B. 1996. Human evolution. *Bioessays* **18**:945-954.
- Wood, B., and M. Collard. 1999. The human genus. *Science* **284**:65-71.
- Woods, C. G., J. Bond, and W. Enard. 2005. Autosomal recessive primary microcephaly (MCPH): a review of clinical, molecular, and evolutionary findings. *Am J Hum Genet* **76**:717-728.
- Woolfe, A., D. K. Goode, J. Cooke, H. Callaway, S. Smith, P. Snell, G. K. McEwen, and G. Elgar. 2007. CONDOR: a database resource of developmentally associated conserved non-coding elements. *BMC Dev Biol* **7**:100.
- Wray, G. A. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet* **8**:206-216.
- Yan, J., Z. Ma, X. Xu, and A. Y. Guo. 2014. Evolution, functional divergence and conserved exon-intron structure of bHLH/PAS gene family. *Mol Genet Genomics* **289**:25-36.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586-1591.
- Yang, Z., W. S. Wong, and R. Nielsen. 2005. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**:1107-1118.
- Yu, L., N. Arbez, L. G. Nucifora, G. L. Sell, L. E. Delisi, C. A. Ross, R. L. Margolis, and F. C. Nucifora, Jr. 2013. A mutation in NPAS3 segregates with mental illness in a small family. *Mol Psychiatry*.
- Yunis, J. J., and O. Prakash. 1982. The origin of man: a chromosomal pictorial legacy. *Science* **215**:1525-1530.
- Yunis, S. a. D. 1980. The Striking Resemblance of High-Resolution G-Banded Chromosomes of Man and Chimpanzee. *Science* **208**, 4448, 1145-1148.
- Zhang, J., R. Nielsen, and Z. Yang. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* **22**:2472-2479.
- Zhang, J., D. M. Webb, and O. Podlaha. 2002. Accelerated protein evolution and origins of human-specific features: *Foxp2* as an example. *Genetics* **162**:1825-1835.
- Zhou, S., S. Degan, E. N. Potts, W. M. Foster, and M. E. Sunday. 2009. NPAS3 is a tracheless homolog critical for lung development and homeostasis. *Proc Natl Acad Sci U S A* **106**:11691-11696.