


## RESEARCH

## Open Access



# Characterization of an eutherian gene cluster generated after transposon domestication identifies *Bex3* as relevant for advanced neurological functions

Enrique Navas-Pérez<sup>1†</sup>, Cristina Vicente-García<sup>2†</sup>, Serena Mirra<sup>1,3,4,5†</sup>, Demian Burguera<sup>1,6</sup>, Noèlia Fernández-Castillo<sup>1,4,7</sup>, José Luis Ferrán<sup>8</sup>, Macarena López-Mayorga<sup>2</sup>, Marta Alaiz-Noya<sup>9,10</sup>, Irene Suárez-Pereira<sup>9,11</sup>, Ester Antón-Galindo<sup>1</sup>, Fausto Ulloa<sup>3,5</sup>, Carlos Herrera-Úbeda<sup>1</sup>, Pol Cuscó<sup>12,13</sup>, Rafael Falcón-Moya<sup>9</sup>, Antonio Rodríguez-Moreno<sup>9</sup>, Salvatore D'Aniello<sup>14</sup>, Bru Cormand<sup>1,4,7</sup>, Gemma Marfany<sup>1,4,7</sup>, Eduardo Soriano<sup>3,5,15</sup>, Ángel M. Carrión<sup>9</sup>, Jaime J. Carvajal<sup>2\*</sup> and Jordi Garcia-Fernàndez<sup>1\*</sup> 

\* Correspondence: [j.carvajal@csic.es](mailto:j.carvajal@csic.es); [jordigarcia@ub.edu](mailto:jordigarcia@ub.edu)

<sup>†</sup>Enrique Navas-Pérez, Cristina Vicente-García and Serena Mirra contributed equally to this work.

<sup>2</sup>Centro Andaluz de Biología del Desarrollo, CSIC-UPO-JA, Universidad Pablo de Olavide, 41013 Sevilla, Spain

<sup>1</sup>Department of Genetics, Microbiology and Statistics, Faculty of Biology, and Institut de Biomedicina (IBUB), University of Barcelona, 08028 Barcelona, Spain Full list of author information is available at the end of the article

## Abstract

**Background:** One of the most unusual sources of phylogenetically restricted genes is the molecular domestication of transposable elements into a host genome as functional genes. Although these kinds of events are sometimes at the core of key macroevolutionary changes, their origin and organismal function are generally poorly understood.

**Results:** Here, we identify several previously unreported transposable element domestication events in the human and mouse genomes. Among them, we find a remarkable molecular domestication that gave rise to a multigenic family in placental mammals, the *Bex/Tceal* gene cluster. These genes, which act as hub proteins within diverse signaling pathways, have been associated with neurological features of human patients carrying genomic microdeletions in chromosome X. The *Bex/Tceal* genes display neural-enriched patterns and are differentially expressed in human neurological disorders, such as autism and schizophrenia. Two different murine alleles of the cluster member *Bex3* display morphological and physiopathological brain modifications, such as reduced interneuron number and hippocampal electrophysiological imbalance, alterations that translate into distinct behavioral phenotypes.

(Continued on next page)



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Conclusions:** We provide an in-depth understanding of the emergence of a gene cluster that originated by transposon domestication and gene duplication at the origin of placental mammals, an evolutionary process that transformed a non-functional transposon sequence into novel components of the eutherian genome. These genes were integrated into existing signaling pathways involved in the development, maintenance, and function of the CNS in eutherians. At least one of its members, *Bex3*, is relevant for higher brain functions in placental mammals and may be involved in human neurological disorders.

**Keywords:** Genetic novelty, Transposon domestication, *Bex3*, *Tceal*, Placental mammals, Gene cluster, Neurodevelopmental disorders, mTOR, Autism spectrum disorder

## Background

Newly evolved genes in a given lineage showing no homologs in other taxa are known as “orphan” or “taxonomically restricted” genes [1]. One of the most striking sources for the birth of lineage-restricted genes is the molecular domestication of transposable element (TE) proteins into novel coding genes [2], which are sometimes involved in the appearance of clade-specific traits and even true evolutionary novelties [3]. Despite their evolutionary relevance, a systematic search for domesticated transposons taking advantage of current and improved genomic annotations was lacking in human and mouse. We have identified now several domestication cases in these species. Among them, we highlight here a previously unreported event that took place at the origin of eutherian mammals, which gave rise to a multigenic family known as *Bex/Tceal* and shaped a cluster of 14 genes on the X chromosome of the placental ancestor. While most are scarcely studied, several reports have linked some of these genes to processes such as cancer proliferation [4–9], cellular reprogramming [10] and differentiation [11], or cell cycle modulation [12, 13]. To investigate the function of this gene family at the organism level, we generated and phenotyped mutant mice for one of its members, *Bex3*. Mutants showed molecular, cellular and anatomical alterations in the brain, as well as important neurological and behavioral alterations typical of neurodevelopmental defects. Altogether, we describe the evolutionary pathways of a TE-derived gene family that integrated into complex molecular routes, while underscoring its impact on neural development and its neuropsychiatric significance.

## Results

### Identification of genes derived from molecular domestication of TEs

In order to detect new transposon domestication events in the mammalian lineage, we looked for protein-coding genes made up by TE-derived sequences. By identifying genes with a coding sequence overlapping greater than 50% with annotated TEs and present in more than one species, we obtained a list of 28 and 9 candidates in the human and mouse genomes, respectively (Additional file 1: Table S1). We recovered well-known TE-derived genes, such as *syncytin-1* [14], *syncytin-2* [15], *SETMAR* [16], and the *ZBED* [17] and *PNMA* families [18]. More ancient domesticated transposons like the vertebrate-specific *RAG1* and *RAG2* genes [3] were not detected, probably due to higher sequence divergence in relation to the ancestral TE element. We also found new

putatively domesticated TEs, most with unknown function, confined to either primate or *Mus* species. However, our attention was drawn to the remarkable case of *Tceal7*, a gene present in all major groups of placental mammals. Due to its broad phylogenetic range, we decided to study this molecular domestication event in further detail.

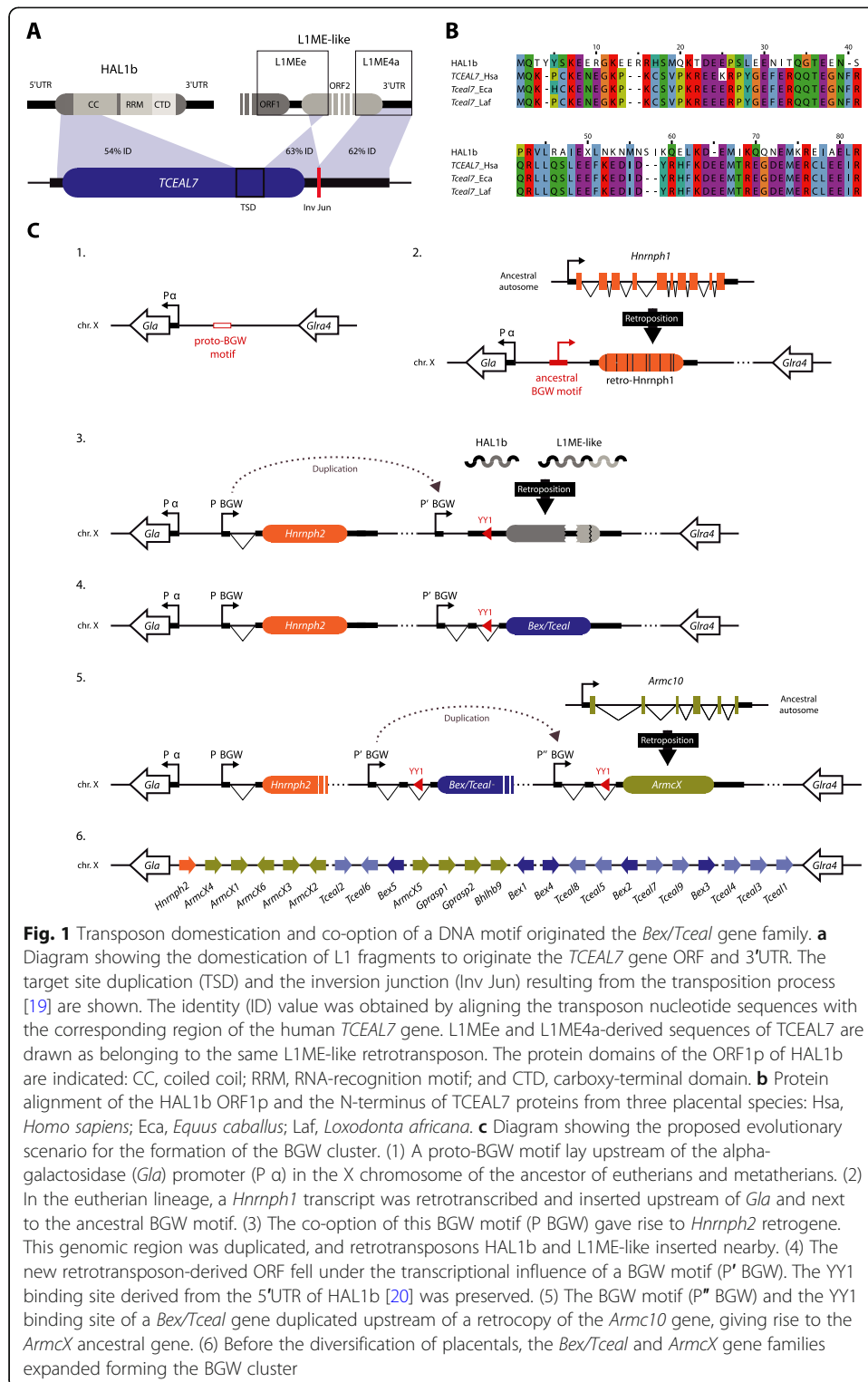
### Molecular composition of the *Tceal7* gene

*Tceal7* is a small gene consisting of two 5' non-coding exons and a third exon that includes the whole open reading frame (ORF) and the 3'UTR. We observed an overlap of 76% between the *Tceal7* ORF sequence and that of HAL1b, a non-LTR retrotransposon belonging to the long interspersed nuclear element-1 (LINE-1 or L1) superfamily (Fig. 1a). Moreover, the last 18 nucleotides of the *Tceal7* ORF and most of the 3'UTR originated from two other L1 subfamilies: the elements L1MEe and L1ME4a (Fig. 1a and Additional file 1: Fig. S1). We observed a partial retention of the original TE coding frame, as human TCEAL7 amino acid sequence shares 31.7% identity with the ORF1p of HAL1b (Fig. 1b). From these results, we determined that the *Tceal7* gene arose from a composite sequence derived from two L1 elements (Fig. 1).

Although *Tceal7* stands out as the most similar to the ancestral L1 retrotransposon sequences, it is but a representative member of a multigenic family called *Bex/Tceal*. This family forms a gene cluster on the X chromosome of placental mammals, with no detectable orthologs outside this clade [21]. In humans, the cluster consists of 5 *Bex* (brain-expressed X-linked) and 9 *Tceal* (transcription elongation factor A (SII)-like) genes, spanning ~1.5 Mb. In mouse, the number is reduced to 11 genes due to the lineage-specific loss of *Tceal2*, *Tceal4*, and *Bex5*. Interestingly, we detected the presence of clustered *Bex/Tceal* genes in each major eutherian lineage (Additional file 1: Fig. S2), confirming that the origin and expansion of the family took place after the divergence of the marsupial-placental clades, and before the radiation of the latter. In agreement, HAL1b, L1MEe, and L1ME4a elements have been suggested to be active retrotransposons during the appearance of early eutherians ~150 Mya [22].

### Evolutionary diversification of the *Bex/Tceal* family

BEX and TCEAL protein sequences are relatively divergent [21], likely due to low selective constraints after the retrotransposon domestication. However, both families share a conserved region toward their C-terminal end (Additional file 1: Fig. S3A). BEX proteins, with the exception of BEX4, are predicted to have a coiled coil within this region [23–25], and we found that TCEAL1, TCEAL7, and TCEAL8 are also predicted to harbor a C-terminal coiled coil domain (Additional file 1: Fig. S3A). Remarkably, most of the protein-protein interactions of BEX1, BEX2, and BEX3 have been mapped to this domain [24], highlighting its potential functional relevance. Furthermore, BEX proteins have been classified as intrinsically disordered proteins (IDPs) [24]. We found that TCEAL proteins, as previously described for BEX members [24], are predicted to have a disordered N-terminal region and C-terminal  $\alpha$ -helices (Additional file 1: Fig. S4). Interestingly, we also detected all these structural features in the ancestral transposon HAL1b (Additional file 1: Fig. S4), suggesting that these features were preserved along the domestication process and inherited by the *Bex/Tceal* genes.



Previous studies using a reduced number of mammalian species reported gene conversion events and positive selection signatures on some *Bex/Tceal* genes [21, 26]. We decided to perform an expanded analysis adding species from the major eutherian clades, and found homogenization of coding sequences among three

groups of *Bex/Tceal* paralogs (*Tceal2* with *Tceal4*; *Tceal3* with *Tceal5* and *Tceal6*; and *Bex1* with *Bex2*) across all studied lineages (Additional file 1: Fig. S3B). Moreover, after filtering out sequences experiencing gene conversion [27], we detected several sites with signatures of positive selection, mainly in the *Bex* subfamily tree, and an episode of positive selection after the branching of *Bex5* (Additional file 1: Fig. S5). This suggests that an ancestral gene within the branch leading to *Bex3* and *Bex4* went through an adaptive process in an eutherian ancestor before the diversification of placental mammals.

### Neighboring DNA sequence co-option as a central regulatory element

Within the same eutherian-specific chromosomal region containing *Bex* and *Tceal* genes, there are multiple retrocopy genes belonging to another gene family (the *Armcx* family, also known as *ALEX*) [28]. Despite no protein similarity between both families, they share a homologous DNA sequence motif in their promoter region known as the BGW motif [21], a ~60 base pair (bp)-long sequence containing an internal E-box, which has been shown to be essential for the regulation of mouse *Tceal7* [29] and human *ARMCX1* [30] expression. The promoter region of *Hnrnp2*, another eutherian-specific retrogene located at the centromeric end of the cluster, also harbors this unique motif [21]. This multiplicity of BGW motifs raises the question of how genes with three independent origins ended up with separate, but homologous, regulatory elements. The promoter of *Hnrnp2* is bidirectional and shared with the galactosidase-alpha (*GLA*) gene [31]. We found a BGW-like motif upstream of the *GLA* promoter in marsupials that lacks an eutherian-restricted 11 bp sequence required for the proper transcription of human *ARMCX1* gene [30] (Additional file 1: Fig. S6). Therefore, the origin of the BGW motif can be traced back to sequences already present in the *GLA* promoter of the last therian common ancestor. Although we cannot determine the precise order of the events leading to the assembly of the whole cluster, the inferred co-option of the BGW motif by the ancestors of the *Bex/Tceal* and *Armcx* families allows us to reconstruct the main steps of this evolutionary process (diagrams depicted in Fig. 1c, see legend for details).

### Brain expression of the *Bex/Tceal* genes and deregulation in neuropsychiatric disorders

Information about expression patterns for *Bex/Tceal* genes is sparse, with data restricted to human, mouse, or rat [12, 29, 32–37]. We gathered publicly available transcriptomic data from eight homologous adult organs of five species belonging to main placental lineages. We observed that most genes present a tissue-enriched pattern, with brain being the organ showing the highest expression levels for most paralogs across species (Fig. 2a). Although brain is an organ where many genes tend to be expressed [38], this result suggests that some of the neural functions reported in mouse and human for this gene family [11, 12, 39] might be conserved among the eutherian clade.

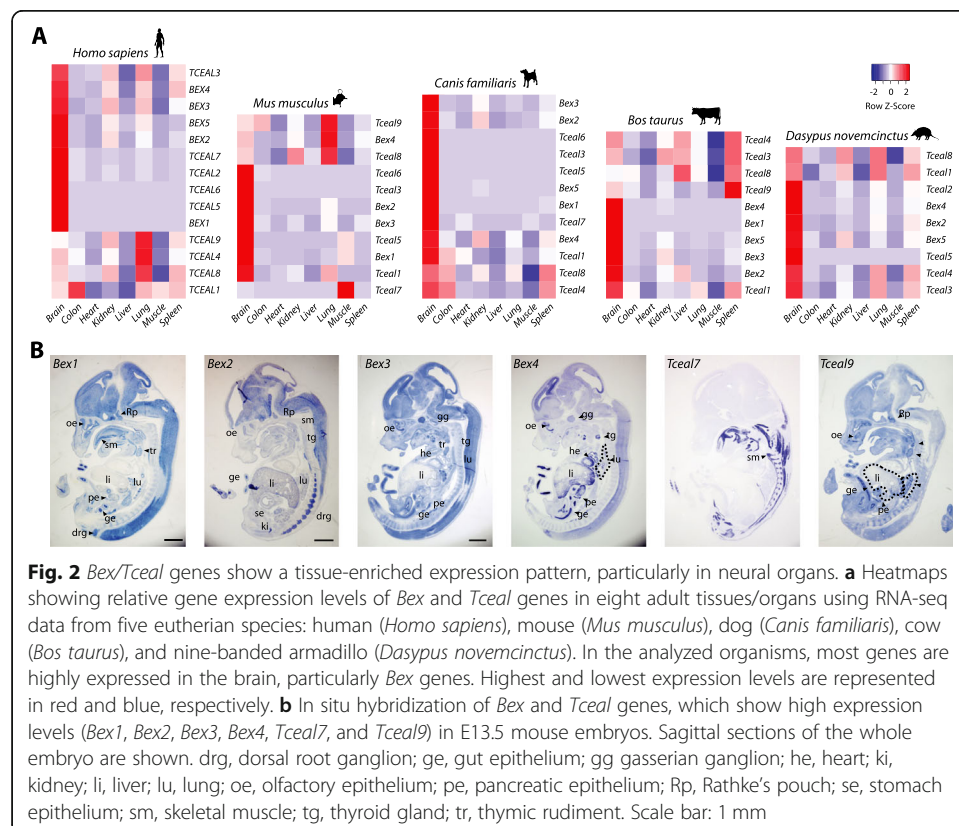
We also investigated the expression patterns of this group of genes during mouse development by *in situ* hybridization and observed a subset of *Bex/Tceal* genes being highly and widely expressed during mouse embryogenesis, especially *Bex* genes (Fig. 2b and Additional file 1: Fig. S7). *Bex3* expression was particularly strong in

the central nervous system during murine development compared to other members of the family (Fig. 2b).

By analyzing publicly available human transcriptomic data of autism spectrum disorder (ASD) and schizophrenia, two well-studied neuropsychiatric disorders, we found a significant decrease in the expression of *BEX/TCEAL* genes in patients compared to controls in different brain regions and datasets (Additional file 1: Table S2), being *BEX* but not *TCEAL* genes significantly enriched among the differentially expressed genes in most datasets (Additional file 1: Table S3). Notably, *BEX3* is located in the interval associated with the neurological features of patients diagnosed with early-onset neurological disease trait (EONDT), which harbor different genomic deletions encompassing *BEX/TCEAL* genes [40–42].

### Activation of neural tube progenitor proliferation by *BEX/TCEAL* proteins in a non-eutherian vertebrate

Next, we investigated the putative function of the inferred, original TE composite sequence that later evolved into the *Bex/Tceal* family. To understand the potential physiological response of this ancient element in the eutherian ancestor, we aimed to mimic the original scenario by using a non-eutherian organism. For this purpose, we synthetically reconstructed a version of the ancestral *Bex/Tceal* protogene, termed *HALEX* (HAL1b-derived on eutherian X chromosome), based on TE consensus sequences (see “Methods”). Moreover, we also studied two *Bex/Tceal* members: *Tceal7*, the gene with the highest sequence similarity to the original protogene; and *Bex3*,



**Fig. 2** *Bex/Tceal* genes show a tissue-enriched expression pattern, particularly in neural organs. **a** Heatmaps showing relative gene expression levels of *Bex* and *Tceal* genes in eight adult tissues/organs using RNA-seq data from five eutherian species: human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis familiaris*), cow (*Bos taurus*), and nine-banded armadillo (*Dasypus novemcinctus*). In the analyzed organisms, most genes are highly expressed in the brain, particularly *Bex* genes. Highest and lowest expression levels are represented in red and blue, respectively. **b** In situ hybridization of *Bex* and *Tceal* genes, which show high expression levels (*Bex1*, *Bex2*, *Bex3*, *Bex4*, *Tceal7*, and *Tceal9*) in E13.5 mouse embryos. Sagittal sections of the whole embryo are shown. drg, dorsal root ganglion; ge, gut epithelium; gg, gasserian ganglion; he, heart; ki, kidney; li, liver; lu, lung; oe, olfactory epithelium; pe, pancreatic epithelium; Rp, Rathke's pouch; se, stomach epithelium; sm, skeletal muscle; tg, thyroid gland; tr, thymic rudiment. Scale bar: 1 mm

which shows the strongest expression in the embryonic nervous system. We electroporated murine *Bex3*, *Tceal7*, and *HALEX* genes into the neural tube of chicken embryos at stage HH12 to investigate their capacity to elicit a cellular response in neuronal progenitors. Expression of *Bex3* and *Tceal7*, but not *HALEX*, generated a significant increase in cell proliferation in the chicken embryonic neural tube, similarly to previous reports in mammalian cultured cells [11, 43] (Additional file 1: Fig. S8). Although the heterologous approach we have used cannot reproduce the regulatory and signaling environments of the eutherian ancestor, the observation that only the murine constructs, but not the ancient element, produced a measurable response suggests that the current ability of *Bex/Tceal* genes to effectively modulate cellular physiology was not present in the ancestral protogene, being acquired during eutherian evolution.

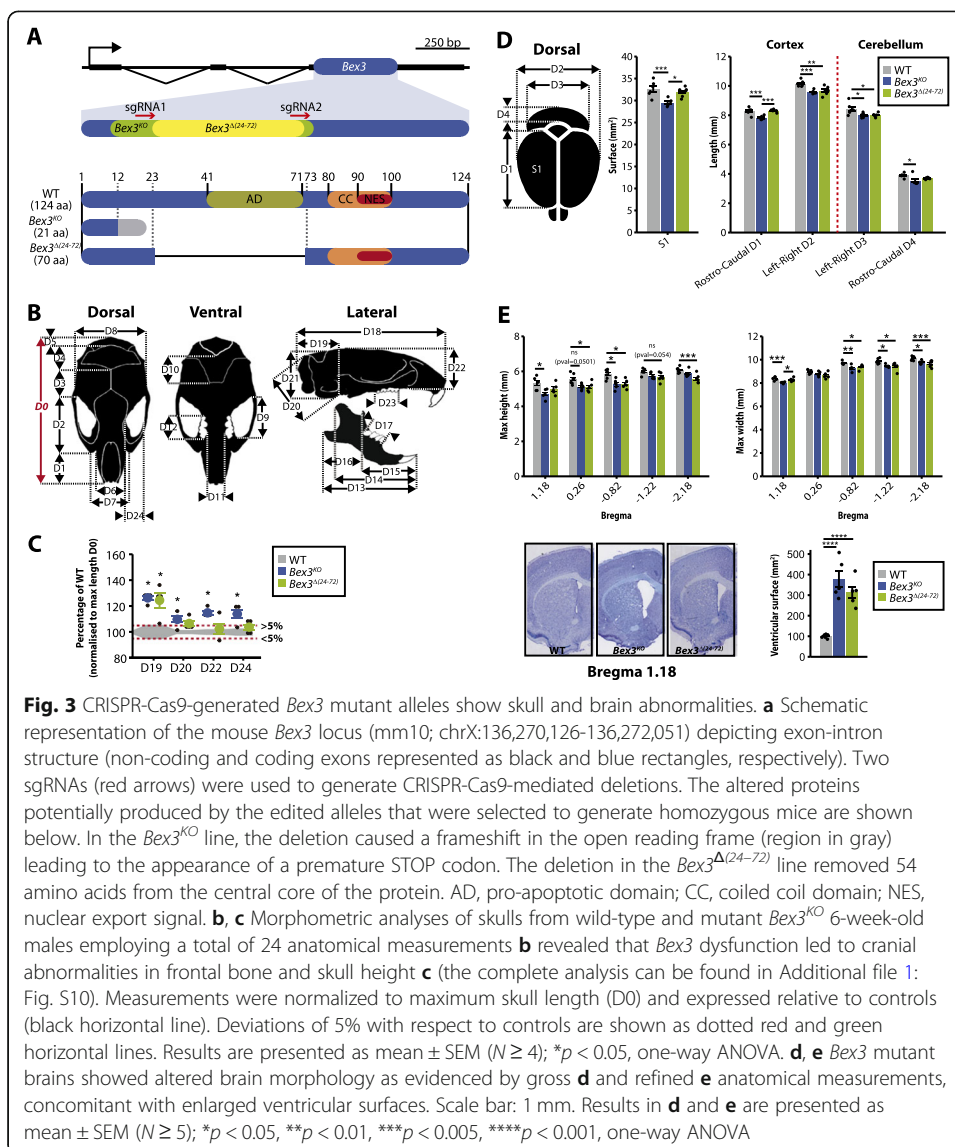
### Generation of two independent *Bex3* mutant lines

While functional studies for the *Bex/Tceal* genes have focused mainly on in vitro assays, little is known about their role at the organism level. Based on previous reports linking *Bex3* to neuronal physiology [11, 43, 44], the analyses on patients with neurological features harboring deletions encompassing *BEX/TCEAL* genes [40–42], and its neural-enriched expression in adult and embryonic tissues, we decided to generate mouse mutant lines for this gene. We used CRISPR-Cas9 technology to generate mutant alleles for *Bex3* in mice and selected two for in-depth characterization (Fig. 3a and Additional file 1: Fig. S9). One of them, namely *Bex3<sup>KO</sup>*, carried a 196-bp deletion that caused a frameshift mutation, introducing a premature stop codon that led to a truncated coding sequence. The second line, which we named *Bex3<sup>Δ(24–72)</sup>*, carried a 147-bp deletion that removed 49 amino acids of the central core of the protein, specifically the pro-apoptotic domain, and retained the C-terminal coiled coil domain required for BEX3 dimerization, nuclear import, ubiquitination [44], and interaction with its multiple partners [24]. Therefore, this mutation potentially expresses a protein lacking an essential functional domain, which is likely to act as a hypomorph or a dominant negative allele, as it has been shown to be the case in cell culture experiments [44].

### Anatomical alterations in skull and brain of *Bex3* mutant mice

Homozygous mutant mice for both lines could be distinguished from wild-type counterparts by external observation of subtle facial differences (Fig. 3b, c), hence skull measurements were taken out in order to identify the origin of these alterations. Snout-to-midbone length (D2 in Additional file 1: Fig. S10) tended to be smaller in the mutants, while the width of the eye sockets (D24 in Fig. 3c) was significantly larger in *Bex3<sup>KO</sup>* mice. Further, the ratio between these measurements was reduced by  $19.4\% \pm 3.4$  in *Bex3<sup>KO</sup>* ( $p = 0.001$ ) and  $10.2\% \pm 3.4$  in *Bex3<sup>Δ(24–72)</sup>* ( $p = 0.064$ ), suggesting abnormal frontal bone morphology that resulted in a rounder eye appearance. Additionally, lack of fully functional *Bex3* also led to increased size of posterior parameter measurements (D5, D19–D21 in Fig. 3c and Additional file 1: Fig. S10), indicating that the bone cavity harboring the cerebellum was larger in the mutants, particularly in *Bex3<sup>KO</sup>* mice.

The overall brain and cerebellum anatomical structures of *Bex3<sup>KO</sup>* and *Bex3<sup>Δ(24–72)</sup>* mice appeared close to normal, although they showed a reduction in cortical surface and size, as well as in cerebellum size (Fig. 3d, e) despite the layered



structures of the cortex, hippocampus, and cerebellum, being preserved in the mutants (data not shown). In addition, we observed a marked increase in brain ventricular surface in both mutant lines (Fig. 3e). It is worth noting that reduced brain size and enlargement of brain ventricles have been described extensively in neurodevelopmental diseases [45–47].

### Behavioral defects in *Bex3* mutant mice

To determine if the anatomical and physiological brain defects observed in the *Bex3* mutant mice have an effect on animal behavior, adult mutant and control mice were subjected to a comprehensive battery of behavioral tests (Fig. 4 and Additional file 1: Fig. S11). In non-stressful open field, *Bex3* mutant mice showed normal locomotor activity [F (10,1) = 0.291; p = 0.601 and F (10,1) = 1.628; p = 0.230] for *Bex3*<sup>KO</sup> and *Bex3*<sup>Δ(24–72)</sup> lines respectively (Fig. 4a). However, we observed that both *Bex3*<sup>KO</sup> and *Bex3*<sup>Δ(24–72)</sup> lines displayed significantly more repetitive behavior events [rearing: *Bex3*<sup>KO</sup>: F (10,1) = 5.078; p =

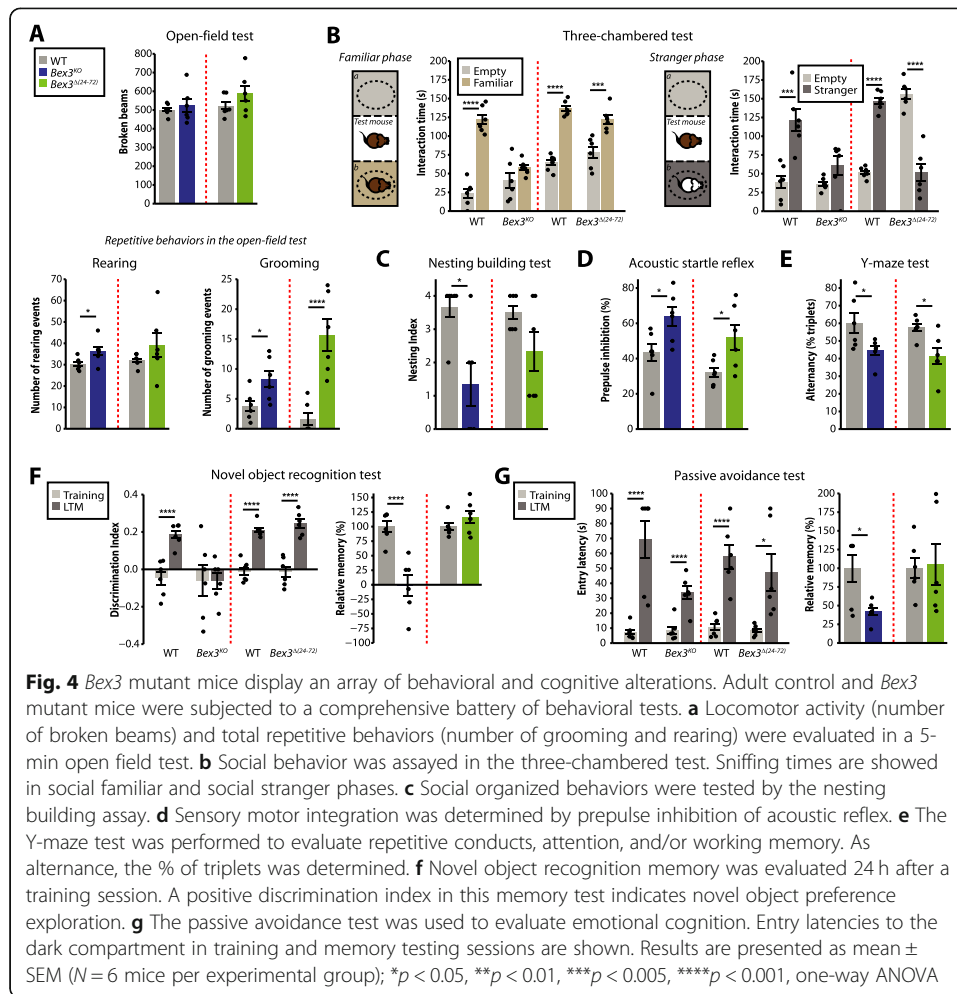


0.047 and  $Bex3^{\Delta(24-72)}$ :  $F(10,1) = 1.355$ ;  $p = 0.271$ ; grooming:  $Bex3^{KO}$ :  $F(10,1) = 6.188$ ;  $p = 0.032$  and  $Bex3^{\Delta(24-72)}$ :  $F(10,1) = 22.216$ ;  $p < 0.001$ ] when compared to their wild-type littermates (Fig. 4a). We evaluated social interaction using a three-chambered assay to study the interactions with familiar or stranger mice (Fig. 4b). While wild-type mice showed preferential interaction with familiar mice against the empty compartment [ $F(10,1) = 94.218$ ;  $p < 0.001$ ], we only found impairment in the interaction with the familiar mice in  $Bex3^{KO}$  line, with no significant differences in the sniffing time between empty and the familiar mice containing compartments [ $F(10,1) = 2.053$ ;  $p = 0.182$ ]. In contrast, in a social novelty paradigm, control animals spent significantly more time in close interaction with the novel animal, whereas  $Bex3^{KO}$  displayed no significant preference for social novelty [ $F(10,1) = 3.114$ ;  $p = 0.108$ ] or even avoided social novel interaction in  $Bex3^{\Delta(24-72)}$  [ $F(10,1) = 51.494$ ;  $p < 0.001$ ]. Furthermore, only  $Bex3^{KO}$  mice also showed impairment in nest building [ $F(10,1) = 9.800$ ;  $p = 0.010$ ; Fig. 4c], which has been correlated with abnormal social organized behavior [48, 49]. Sensorimotor gating, whose impairment is also strongly associated with some neurodevelopmental disorders [50], was assayed by prepulse inhibition (PPI) of acoustic startle reflex. Although  $Bex3$  mutants showed similar acoustic startle reflex than wild-type mice (Additional file 1: Fig. S11), both mutant lines exhibited a significant enhancement of PPI [ $Bex3^{KO}$ :  $F(10,1) = 6.588$ ;  $p = 0.028$  and  $Bex3^{\Delta(24-72)}$ :  $F(10,1) = 6.038$ ;  $p = 0.033$ ; Fig. 4d]. Moreover, cognitive behavior was assayed in the Y-maze test, object recognition memory, and passive avoidance tests. In the Y-maze test,  $Bex3$  mutant mice showed significantly lower spontaneous alternation indexes, indicative of attention deficits and/or working memory [ $Bex3^{KO}$ :  $F(10,1) = 5.186$ ;  $p = 0.045$  and  $Bex3^{\Delta(24-72)}$ :  $F(10,1) = 8.125$ ;  $p = 0.017$ ; Fig. 4e]. On the contrary, only  $Bex3^{KO}$  mice showed significant lower performance compared to their control littermates in object recognition memory [ $Bex3^{KO}$ :  $F(20,3) = 4.843$ ;  $p = 0.039$  and  $Bex3^{\Delta(24-72)}$ :  $F(20,3) = 0.549$ ;  $p = 0.467$ , analyzed by 2-way ANOVA for genotype X session interaction; Fig. 4f] and passive avoidance tests [ $Bex3^{KO}$ :  $F(20,3) = 6.548$ ;  $p = 0.018$  &  $Bex3^{\Delta(24-72)}$ :  $F(20,3) = 0.262$ ;  $p = 0.613$ , analyzed by 2-way ANOVA for genotype X session interaction; Fig. 4g]. In summary, the behavioral experiments indicated that both  $Bex3$  mutant lines display repetitive behaviors and abnormal social conducts and that  $Bex3^{KO}$  mice presented more severe phenotypes, especially in cognitive tests.

### Reduction in the number of cortical and subcortical interneurons in $Bex3$ -deficient mice

A link between the disruption of interneuron inhibitory circuits in the neocortex and some clinical aspects of neurodevelopment and psychiatric disorders has already been established [51]. Our data show that both  $Bex3$  mutant lines display gross brain morphological alterations and have abnormal behavioral traits and that  $Bex3^{KO}$  mice presented more severe phenotypes, especially in terms of memory and learning impairment. These results, together with transcriptomic and structural genomic data related to human neurological disorders (Additional file 1: Tables S2 and S3) [40–42], suggest that  $Bex3$  may function through the maintenance/renewal of specific neurons, and its absence can give rise to neurological conditions.

We measured parvalbumin-positive (PV+) interneuron density and found it was significantly decreased in motor, cingulate and sensory cortices, and caudate-putamen of  $Bex3^{KO}$  mice;  $Bex3^{\Delta(24-72)}$  mice showed a significant reduction of PV+ interneurons



only in cingulate and sensory cortices (Fig. 5a, b) and minor, non-significant, decrease in motor cortex and caudate-putamen. On the other hand, altered numbers of hippocampal neurons and interneurons have been associated to the onset of social and cognitive deficits in several animal models for neurological disorders (reviewed in [52]). Even though the hippocampus-to-hemisphere size ratio was not significantly altered in *Bex3* mutant mice (Fig. 5c), we observed a strong reduction in the density of total neurons and, particularly, calretinin-positive (CR+) inhibitory interneurons in the *stratum radiatum* of the CA1–2 hippocampal fields of *Bex3<sup>KO</sup>* mice (Fig. 5d, e).

We decided to analyze the status of the adult hippocampal neurogenic niche in *Bex3* mutant mice. Interestingly, the density of immature neurons was significantly decreased in the subgranular zone of *Bex3<sup>KO</sup>* mice, but not in *Bex3<sup>Δ(24–72)</sup>* (Fig. 5f, g), thus suggesting that adult neurogenesis could be partially compromised in these animals.

**Alteration in the excitation/inhibition balance in the hippocampal CA2 region of *Bex3<sup>KO</sup>* mice** Impaired social interaction and social learning is known to rely on the CA2 neuronal circuit within the hippocampus [53, 54], while altered interneuron numbers in the CA2 hippocampal field can disrupt the excitatory/inhibitory hippocampal balance. Thus, we

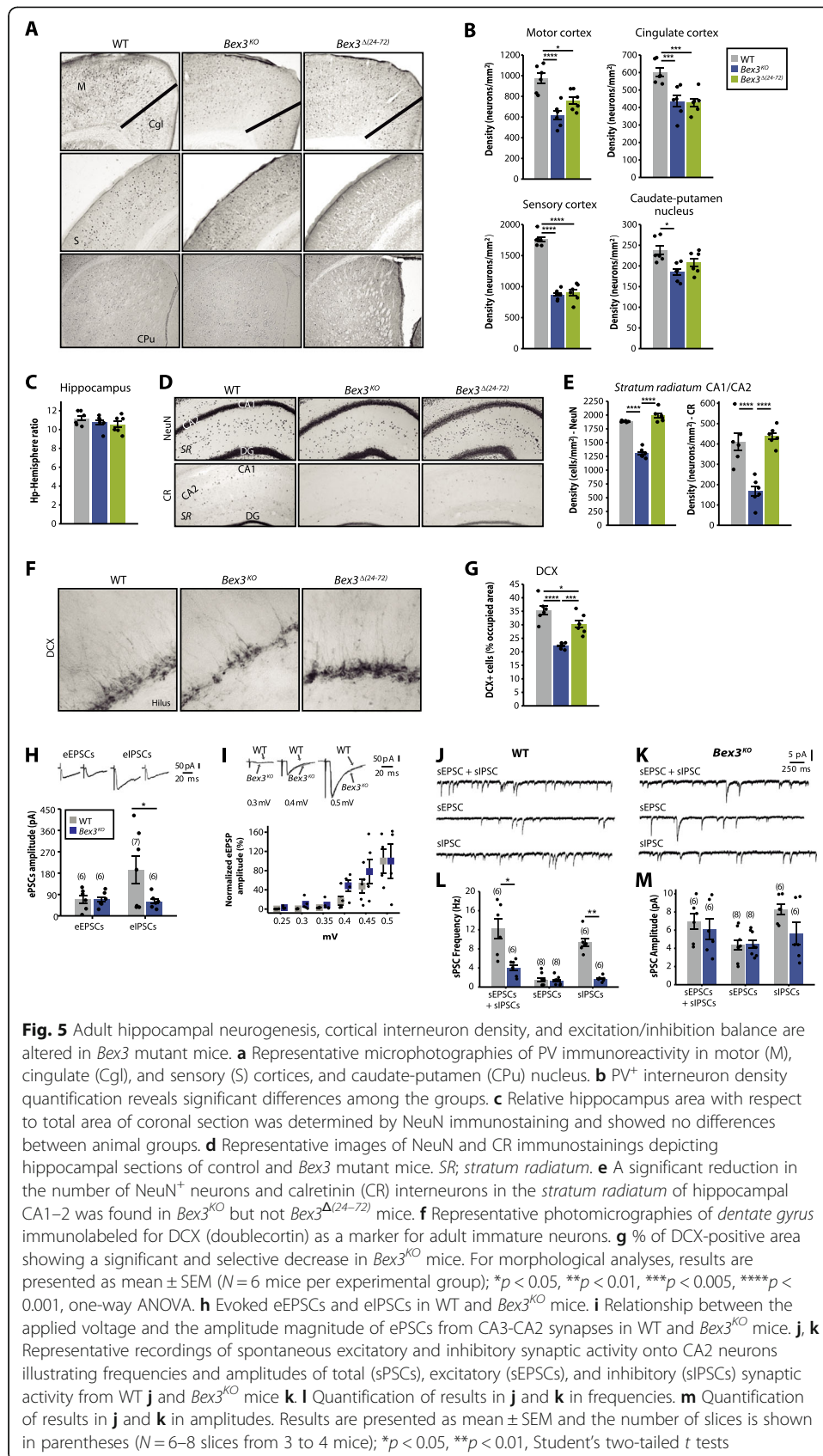
tested if CA2 synaptic transmission was altered in *Bex3*<sup>KO</sup> mice, which have reduced numbers of hippocampal interneurons in this area. To determine whether the evoked synaptic activity was altered in hippocampal pyramidal CA2 neurons, we monitored evoked excitatory (eEPSCs) and inhibitory (eIPSCs) postsynaptic currents. We found that the amplitude of eIPSCs—but not eEPSCs—was clearly decreased in mutant mice (Fig. 5h). However, no statistically significant differences were found between wild-type and *Bex3*<sup>KO</sup> mice in the input-output curve, indicating that the net basal synaptic activity was not affected in these animals (Fig. 5i). Additionally, we studied spontaneous synaptic transmission onto hippocampal CA2 pyramidal neurons (Fig. 5j-m). Total spontaneous activity frequency (sEPSC + sIPSC) was significantly decreased in *Bex3*<sup>KO</sup> mice respect to wild-type (Fig. 5l) due to a strong reduction in sIPSCs frequency. On the other hand, no differences were found between sEPSCs and sIPSCs amplitude when comparing wild-type and mutant mice (Fig. 5m). These results indicate an excitation/inhibition imbalance in the hippocampal CA2 region of *Bex3*<sup>KO</sup> mice due to a strong decrease in inhibitory synaptic transmission, which is in agreement with the reduced density of inhibitory interneurons observed in this region.

#### Altered mTOR signaling in the adult brain of *Bex3*-deficient mice

BEX3 protein physically interacts with the TSC1/TSC2 complex [55], which is essential for the proper regulation of the mTOR signaling cascade [56]. This pathway controls brain development and function at multiple levels, and its dysregulation has been implicated in several neurological diseases [57]. Intriguingly, we found a decrease in the phosphorylation of the mTORC1 readout S6K1 in brain lysates of both *Bex3* mutant lines, which would suggest mTORC1 hypoactivation; in contrast, no changes were detected in other targets like 4E-BP1 or S6. The mTORC2 readouts AKT and NDRG1 were hyperphosphorylated, whereas PKC $\alpha$  and PKC $\gamma$  protein levels were unaltered (Fig. 6 and Additional file 1: Fig. S12). All these data suggest a hyperactivation of mTORC2 signaling in the adult brain as one of the possible molecular consequences linked to *Bex3* deficiency.

#### Discussion

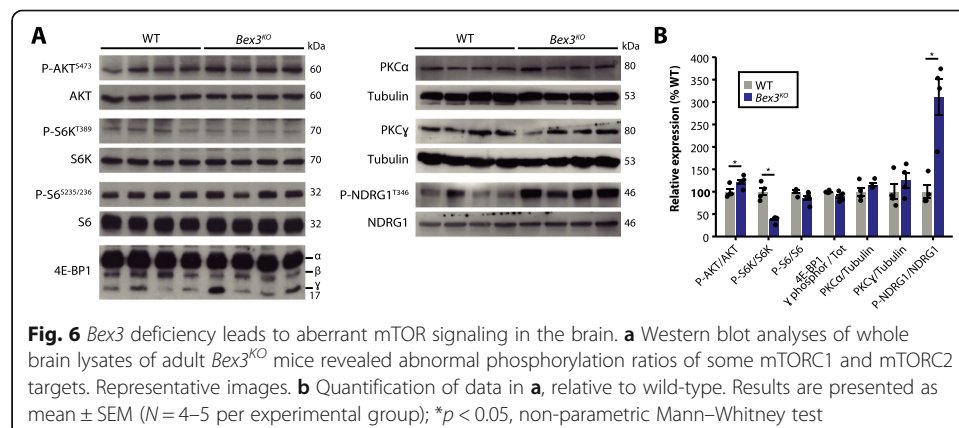
By screening the human and mouse genomes for domesticated transposons, we have identified the footprints of two L1 retrotransposons within the human *TCEAL7* gene, indicating that a domestication event restricted to placental mammals gave rise to a whole new gene family in the X chromosome. After the eutherian-specific gene L1TD1 [58], the *Bex/Tceal* family represents the second case of domestication of non-LTR transposons reported in metazoans, and the first to give rise to a multigenic family. This domestication generated an ancestral protogene that we termed *HALEX*, and our experiments in a non-eutherian model suggest an evolutionary scenario where this element was progressively integrated into anciently established gene networks through complete neofunctionalization before the diversification of eutherians. This is in contrast with most other known domestication cases, where new genes perform tasks at the cellular level similar to those of the ancestral element [59]. Several members of the *Bex/Tceal* gene family have been reported to code for hub proteins, due to their high number of interactions with multiple proteins belonging to several signaling pathways



[11, 12, 35, 60–64]. Remarkably, we found that both BEX/TCEAL proteins and the ancestral HAL1b sequence from which they derive are predicted to be predominantly disordered at their N-terminus and contain C-terminal  $\alpha$ -helices (Additional file 1: Fig. S4). Relatedly, disorder is a structural property that frequently translates into variable conformations and the ability to bind multiple partners [65]. Thus, we suggest that the evolutionary potential of *Bex/Tceal* genes for a multiplicity of protein-protein interactions might be a direct legacy of the ancestral HAL1b retrotransposon protein.

The mammalian target of rapamycin (mTOR) pathway stands out as one of the molecular networks in which *Bex/Tceal* genes integrated, as shown by previous work on *BEX2* and *BEX4* [62, 64]. The mTOR protein kinase is involved in a myriad of processes related to cell growth, proliferation, and survival. It can assemble into two molecularly and functionally different complexes that include some specific components such as mTORC1 and mTORC2 [66]. mTORC1 main substrates are S6K and 4E-BP kinases, while mTORC2 downstream targets include AKT and SGK kinases. Upon phosphorylation, AKT inhibits the TSC1/2 complex [67], which regulates both mTORC signaling cascades in opposite directions: it inhibits mTORC1 and activates mTORC2. Yet, it can interact physically with mTORC2 only, specifically through Tsc2 [67].

Our data show that *Bex3* deficiency leads to hyperphosphorylation of Akt on Ser-473 in brain extracts, suggesting mTORC2 hyperactivation. Akt hyperphosphorylation should also induce hyperactivation of the mTORC1 route but no changes in 4E-BP1 phosphorylation were detected, while S6K1 was found to be hypophosphorylated on Thr-389. Importantly, earlier phosphorylation steps on the autoinhibitory domain of S6K1 are needed to render residue Thr-389 accessible to mTORC1. Both mTORC1 and JNK are involved in the phosphorylation of this domain of S6K1 in muscle [68]. JNK is inhibited by mTORC2, and therefore, mTORC2 hyperactivation in our models could indirectly lead to the observed S6K1 hypophosphorylation on Thr-389. Surprisingly, phosphorylation levels of S6K1 main effector S6 on residues Ser235/236 were not affected, suggesting the possible existence of compensatory mechanisms by other kinases that can target these residues [69]. In summary, our biochemical data and the fact that BEX3 interacts with the TSC1/TSC2 complex [55] suggest the putative involvement of the mTOR pathway in some of the reported phenotypic alterations. However, the weak differences found in some mTOR effectors imply that additional signaling pathways may also be involved in the reported changes.



Given that BEX3 physically interacts with TSC1, an interaction shown to prevent proteasome degradation and thus necessary for NGF-p75NTR-BEX3-mediated apoptosis [55], we propose that BEX3 could prevent the TSC1/2 complex from interacting with mTORC2, eventually reducing the activity of the pathway. The absence of *Bex3* would then cause a hyperactivation of the mTORC2 pathway without affecting mTORC1. Thus, the function of BEX3 could be that of fine-tuning the regulation of these cascades. The fact that the *Bex3*<sup>KO</sup> and *Bex3*<sup>Δ(24–72)</sup> lines show analogous results suggests that the missing central core of the mutant *Bex3*<sup>Δ(24–72)</sup> protein is necessary for TSC1/2 recognition or, alternatively, that its coupling with the TSC1/2 complex does not interfere with mTORC2 binding, presumably due to the smaller size of the resultant protein. Furthermore, since mTOR signaling has been involved in adult neurogenesis and neuronal excitability and is dysregulated in several neurological diseases, the mTOR pathway is a compelling candidate to explain, at least in part, the molecular alterations behind the phenotype observed in *Bex3*-deficient mice.

The acquisition of regulatory sequences enabling transcription is also a crucial but scarcely studied step during the process of domestication and subsequent neofunctionalization of transposable elements [70]. We have described here a short, non-coding regulatory region called BGW motif that was co-opted during eutherian evolution by three unrelated gene ancestors generated from retrotransposition events.

Although transposons and mammalian-restricted genes predominantly show tissue-specific patterns [71, 72], we observed a relatively wide expression for most *Bex/Tceal* genes. In mouse embryos, expression profiles of *Bex/Tceal* genes are consistently associated with proliferative tissues within organs such as the stomach, lungs, pancreas, or central nervous system. Interestingly, some of these genes are used as markers for progenitor cells in developing tissues and are involved in cell proliferation, differentiation, and cell death [11, 29, 34, 73, 74]. In adult tissues, our results suggest that *Bex3* regulates adult hippocampal neurogenesis, while some *Bex/Tceal* genes are induced upon injury, having an impact on axonal, muscular, and hepatic regeneration [29, 37, 39].

*BEX1* and *BEX3* have been recently identified among the most significantly downregulated genes in excitatory neurons of the prefrontal cortex of Alzheimer's disease patients [75]. Furthermore, patients with severe intellectual disability, craniofacial dysmorphism, and autism have been reported to carry different genomic microdeletions in Xq22 encompassing *BEX/TCEAL* genes, with *BEX3* pinpointed as one of the main candidates to cause these neurological features [40–42]. Also, a small 252-kb duplication spanning *BEX3*, *TCEAL4*, *TCEAL9*, and *RAB40A* has been reported in a patient with autism (Decipher database, ID: 290829) [76]. Here we show that mutations of murine *Bex3* lead to subtle craniofacial changes and have a profound impact on repetitive and social behavioral performance, two important behavior alterations required to diagnose autism spectrum disorders (ASD) [77]. Consistent with ASD-like behaviors, some cerebral cortex areas, related with social behaviors [78, 79], and the striatum, related with repetitive behavior [80], showed alterations in the number of parvalbumin interneurons, another feature found in ASD mice models and patients [81–88]. Cognitive alterations are frequently found in ASD patients [89]. In this aspect, only *Bex3*<sup>KO</sup> mice show learning and memory defects in object recognition and passive avoidance tests, two hippocampus-dependent paradigms [90–92]. In parallel, alterations in calretinin interneurons in the CA1–2 hippocampal *stratum radiatum*, excitatory-inhibitory

imbalance in the CA2 field, and potential defects in adult *dentate gyrus* neurogenesis were also found only in *Bex3*<sup>KO</sup> mice. In this regard, the less severe phenotype of *Bex3*<sup>Δ(24–72)</sup> mice suggests that the resulting protein may act as a hypomorph version of the wild-type allele. In brief, the phenotypical features of *Bex3* mutant mice makes this gene an exciting candidate for future research into human neurological disorders that impact upon repetitive behavior, sociability, and intellectual disability.

## Conclusions

We characterized the evolutionary process by which an ancient retrotransposon was inserted into the genome of an eutherian ancestor and progressively integrated into host molecular networks while acquiring crucial functions for the organism. We suggest that this process of strict neofunctionalization was channeled by the inherited structural properties of the ancestral transposable element, facilitating the evolution of protein interactions. This domesticated gene subsequently duplicated, generating the *Bex/Tceal* cluster, whose expression is enriched in neural tissues. By generating new alleles for one of its members, *Bex3*, we show that this gene is involved in the development, maintenance, and function of specific areas of the central nervous system. The mutant alleles show subtle anatomical alterations in skull morphology, and important variations in cortical and subcortical neuron populations, as well as electrophysiological hippocampal imbalance, which can explain some of the behavioral changes identified. We also find a disruption in the mTOR pathway that might explain the molecular cause underlying some of the observed phenotypes. It remains to be elucidated if other genes in the cluster, most of which are highly expressed in the central nervous system, play similar or complementary roles in the establishment of higher neurological functions in placental mammals, or into which existing signaling pathways have been incorporated.

## Methods

### Study design

Our original hypothesis posited that recently improved genome and repetitive element annotations may help uncover new events of transposon domestication in the human and mouse genomes. Thus, the primary objective was to identify previously undetected TE-derived genes in the human and mouse genomes. A bioinformatic genome-wide screening provided a list of putative candidates, including previously uncharacterized cases. We decided to study in depth the event that gave rise to the *Bex/Tceal* gene family, which originated at the base of the eutherian lineage. We hypothesized that they could have provided an evolutionary novelty and sought to functionally characterize one of its members, *Bex3*, to explore this possibility. Using the CRISPR/Cas9 technology in mice, we generated two mutant lines for this gene, namely *Bex3*<sup>KO</sup> and *Bex3*<sup>Δ(24–72)</sup>, which were maintained in a CBA/C57Bl6 hybrid background. These mice, together with wild-type counterparts, were subjected to a battery of morphological, behavioral, physiological, histological, and molecular analyses. Additionally, chick embryos were employed as a non-eutherian animal model. All experiments were performed using 4 to 7 animals per experimental group, as presented in the manuscript. They were randomly chosen from the available animals in each case. For quantitative data, most

measurements were made automatically and, thus, not subject to operator bias. Additionally, the limited number of researchers working on each experiment prevented blinding procedures. No data were excluded from the analysis.

#### Identification of domesticated transposon candidates

Gene and repetitive element annotations for the hg38 human and mm10 mouse genome assemblies were downloaded from the table browser tool of the UCSC website, selecting UCSC genes and RepeatMasker tracks, respectively. In order to discard non-TE elements, we filtered the RepeatMasker output as in [93]. These gene and repetitive element annotations were directly compared, and the results were filtered according to the following criteria: (i) overlap above 50% across the candidate gene coding region; (ii) tx-CdsPredict score above 800, which is approximately 90% predictive of protein-coding genes (it considers the length of the ORF, the presence of a Kozak consensus sequence, nonsense mediated decay mechanisms, and upstream ORFs); and (iii) conservation of the ORF in more than one species. Mammalian species with genome available at Ensembl genome browser (<http://www.ensembl.org>) were considered when checking for ORF conservation.

#### Sequence retrieval

BLASTn and BLAT searches in NCBI and UCSC databases using human sequences were carried out in order to identify *Bex/Tceal* family genes in eutherian species. To clarify the orthology between genes undergoing gene conversion, surrounding intergenic sequences were added to the corresponding human query. The genomic assemblies used were as follows: GRCh38/hg38 for human (*Homo sapiens*); Broad CanFam3.1/canFam3 for dog (*Canis familiaris*); GRCm38/mm10 for mouse (*Mus musculus*); Bos\_taurus\_UMD\_3.1.1/bosTau8 for cow (*Bos taurus*); Broad/equCab2 for horse (*Equus caballus*); Baylor/dasNov3 for nine-banded armadillo (*Dasypus novemcinctus*); Broad/choHof1 for Hoffmann's two-toed sloth (*Choloepus hoffmanni*); Broad v1.0/triMan1 for Florida manatee (*Trichechus manatus latirostris*); Broad/monDom5 for gray short-tailed opossum (*Monodelphis domestica*); WTSI Devil\_ref v7.0/sarHar1 for Tasmanian devil (*Sarcophilus harrisii*), TWGS Meug\_1.1/macEug2 for tammar wallaby (*Macropus eugenii*). The consensus sequences of TE subfamilies were retrieved from the RepBase-derived RepeatMasker library update 20170127 (<http://www.girinst.org/server/RepBase/>).

#### Evolutionary analyses and secondary structure prediction

Nucleotide and protein sequences were aligned using the L-INS-i iterative refinement method of MAFFT [94], and the resulting alignments were edited with Jalview [95]. The phylogenetic reconstruction was performed using IQ-TREE [96] and built-in software ModelFinder [97]. Branch support was calculated running 1000 replicates of the SH-like approximate likelihood ratio test [98] and ultrafast bootstrap [99]. Phylogenetic trees were visualized and edited with FigTree v1.4.2 [100] and Dendroscope 3 [101]. PCOILS was used for coiled coil prediction [102]. PSI-PRED 4.0 [103, 104] and DISOPRED3 [105] were used for  $\alpha$ -helix and protein disorder prediction, respectively. Finally, the positive selection analysis was performed using HyPhy [106], specifically the



MEME [107] and aBSREL [108] methods. Taking into account the divergence observed between the *Bex* and *Tceal* subfamilies, the analysis was split in two parts to avoid saturation of substitutions.

#### Differential expression analysis of eutherian adult tissues

RNA-seq data from eight homologous adult organs of dog, cow, and nine-banded armadillo were downloaded from SRA database: SRP016501 (GSE41637, *Bos taurus*), SRP114662 (GSE20113, *Canis familiaris*), and SRP012922 (GSE106077, *Dasypus novemcinctus*). Protein-coding cDNA sequences were downloaded from Ensembl (<http://www.ensembl.org>) as a transcriptomic index to map against. When *Bex/Tceal* orthologs were found to be incomplete or absent in these files, we replaced the partial sequences or introduced the missing ones, respectively. RNA-seq data was trimmed to 50 base pairs and mapped using Bowtie (allowing no more than two mismatches and discarding reads that mapped more than once) against their respective libraries for each species, and expression levels were calculated correcting for the effective length of each transcript (read-long positions repeated in other transcripts were excluded) to obtain cRPKM metrics [109]. cRPKM values for human and mouse were obtained from VAST DB [110]. Heatmaps were produced with Heatmapper [111] using Pearson's distance measurement and average linkage as the clustering method.

#### Expression of *BEX/TCEAL* genes in autism spectrum disorder and schizophrenia

Differential expression of *BEX/TCEAL* genes was assessed using transcriptomic data from all publicly available human transcriptomic datasets for schizophrenia and autism spectrum disorder, either in GEO (<http://www.ncbi.nlm.nih.gov/geo>) or published articles (Additional file 1: Table S2). Statistical analysis of these data is specified in the corresponding "Statistical analysis" section.

#### In situ hybridization

Mouse embryos were collected at E13.5, cryoprotected, and cut in 20- $\mu$ m-thick sections. Primer pairs (Additional file 1: Table S4) were designed in order to amplify murine *Bex/Tceal* genes by PCR using cDNA. Next, digoxigenin-labeled RNA antisense probes were synthesized, and in situ hybridization was performed as described elsewhere [112]. Images were obtained with a Leica MZ16 F stereomicroscope.

#### Generation of transgenic mice using the CRISPR-Cas9 system

Two CRISPR guide RNAs (sgRNAs) targeting the mouse *Bex3* gene were designed on exon 1 using CRISPR DESIGN (<http://crispr.mit.edu>) and CRISPRSCAN [113], considering potential off-target effects and predicted functional activity. They were generated by in vitro transcription from sgRNA DNA templates as previously described [114]. Briefly, for each sgRNA, two complementary oligos containing the CRISPR-Cas9 target sequence were annealed leaving BbsI-compatible overhangs. Oligo sequences for each sgRNA can be found in Additional file 1: Table S4. Annealed products were then cloned into the pgrNAbasic plasmid, a kind gift from Dr. Moises Mallo (Instituto Gulbenkian de Ciencia, Portugal), at the *BbsI* site located downstream the T7 promoter and upstream the universal tracrRNA sequence necessary for sgRNA folding and

activity. Plasmids were linearized with *FspI* and the purified product used for T7 in vitro transcription according to the manufacturer's instructions (New England Biolabs). sgRNAs were purified by phenol:chloroform extraction after DNaseI treatment (Roche). Next, in vitro digestion assays were performed to evaluate sgRNA activity [115]. In brief, a DNA template containing the CRISPR-Cas9 target sites was amplified by nested PCR with the primers listed in Additional file 1: Table S4. Purified PCR products (200 ng) were incubated for 3 h at 37 °C with Cas9 protein (20 ng/μl, Addgene vector #47327) and sgRNA (2.5 ng/μl) in digestion buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA, 10 mM MgCl<sub>2</sub>). Cleavage efficiency was then analyzed by electrophoresis in a 1% agarose gel stained with ethidium bromide.

Cas9 mRNA (100 ng/μl, SBI) and sgRNA (10 ng/μl) were co-injected into the cytoplasm of CBA/C57Bl6 fertilized eggs using standard methods. Deletions in F<sub>0</sub> pups were detected by nested PCR (see Additional file 1: Table S4) of genomic DNA obtained from tail biopsies, and confirmed by sequencing after TA cloning into pCR2.1 (Invitrogen). Mutant F<sub>0</sub> carriers were crossed with wild-type CBA/C57Bl6 hybrids, and their descendants were likewise analyzed to identify the specific deletion allele transmitted. Mutant lines with deletions in mm10 chrX:136271359-136271505, for *Bex3*<sup>Δ(24-72)</sup>, and in chrX:136271327-136271522, for *Bex3*<sup>KO</sup>, were established and maintained in a hybrid background.

#### RT-PCR analyses

Brain RNA was extracted from three adult male wild-type and mutant mice using TriPure isolation reagent (Roche), and cDNA was then synthesized with the first-strand cDNA synthesis kit for RT-PCR (AMV) (Roche), according to the manufacturer's instructions. *Bex3* and *actin* expression were analyzed with the primers listed in Additional file 1: Table S4.

#### Morphometric analysis of mouse skulls

Control and mutant adult male mice at 6 to 8 weeks of age were sacrificed by cervical dislocation. Their heads were severed, skinned, defleshed, and incubated for 2–3 days in 2% NaOH/PBS in agitation to remove all remaining soft tissue. Undigested tissues were removed with forceps and skulls washed 3 times in PBS for 20 min. Samples were oriented and photographed using a high-resolution 3-CCD JVC camera (model KY-F55B) fitted onto a Nikon SMZ1500 stereomicroscope. Images were assembled, and 25 anatomical measurements were taken in Adobe Photoshop. Results were normalized to maximum skull length (D0 in Fig. 3) and relativized with respect to controls.

#### Histological staining

Adult control and *Bex3* mutant mice were anesthetized and transcardially perfused for 20 min with 4% PFA/PBS. Brains were removed, post-fixed 24–48 h in 4% PFA/PBS, cryoprotected with 30% sucrose/PBS overnight, frozen in isopentane, and stored at –80 °C until use. Brains were sectioned coronally (30 μm), sections were collected in cryoprotectant solution (85% glycerol, 100% ethylene glycol, 0.1 M PBS) and kept at –20 °C until analysis. In the case of Nissl staining, sections were dehydrated and

mounted (Eukitt). Slide images were captured using a Nanozoomer slide scanner (Hamamatsu) and analyzed by measuring the maximum width (mm) and height (mm) in 5 sections per animal, corresponding to different anterior-posterior levels based on Bregma. Measurements were performed by using NanoZoomer Digital Pathology software. Ventricular surface was measured in 2 histological sections per animal by using ImageJ software. Additionally, immunohistochemistry (IHC) labelling was performed on 50- $\mu\text{m}$  coronal brain cryosections with the following primary antibodies: mouse anti-NeuN (1:3000, Millipore), rabbit anti-parvalbumin (1:3000, Swant), mouse anti-calretinin (1:1000, Novocastra), rat anti-GFAP (1:3000, Calbiochem), and goat anti-doublecortin (1:500, Santa Cruz). Immunoreactivity was developed with DAB-peroxidase reaction. To minimize variability, at least 2–3 sections from each area were analyzed per animal on a bright-field DMRB RFY HC microscope (Leica). In each section, the total number of labeled cells per area of tissue was quantified using ImageJ software.

Chick embryos were fixed for 2–4 h at 4 °C in 4% (w/v) paraformaldehyde in PB, washed in PBS, and vibratome sectioned (45  $\mu\text{m}$ ). BrdU detection and immunostaining were performed following standard procedures [116, 117]. The following antibodies were used: rabbit anti-GFP (1:500, Invitrogen), rat anti-BrdU (1:500, AbDSerotec), rabbit anti-Sox2 (1:500, Invitrogen), Alexa488- and Alexa562-conjugated antibodies (Invitrogen). The sections were recorded using a Leica SPE confocal microscope. Cell counting was carried out on 10–17 pictures obtained from 5 to 7 chick embryos per experimental condition.

### **Behavioral tests**

Behavioral tests were performed in 3- to 5-month-old mice in a room with constant sound and light after 1 h of habituation. In all tests, during mice manipulation and behavioral evaluation, the researcher was blind to mice genotypes. The order in which the tests were run was always the same: open field, three-chambered test, nesting test, startle, Y-maze, object recognition memory, and step-down passive avoidance.

#### ***Open field test***

Motor activity was assessed in an open field for 5 min (38 × 21 × 15 cm: Cybertec S.A., Madrid, Spain) as previously described [92]. This apparatus, which is coupled with infrared (IR) emitters and sensors connected to a computer, allowed to quantify the number of times mice interrupted the IR beams/min. In parallel, sessions were video recorded in order to evaluate rearing and grooming.

#### ***Three-chambered sociability test***

Sociability was analyzed in an arena (60 × 60 cm) divided in three equal compartments. Each lateral compartment contained a clear Plexiglas cylinder (each 7 cm in diameter, 12 cm tall) with multiple holes (1 cm in diameter) to allow auditory, visual, and olfactory interaction between the stimulus mouse and the test mouse placed inside and outside of the cylinder, respectively. The paradigm, lasting 5 min per animal, consisted in a three-stage procedure: During the *habituation phase*, the test mouse was allowed to explore the

apparatus with the cylinders empty. In the *social familiar subject phase*, a conspecific from its same home cage was placed in one of the cylinders while the other remained empty. Finally, the *social stranger subject phase* was performed with an unfamiliar age- and sex-matched CD1 mouse placed in one of the cylinders maintaining the other empty. Social interaction was evaluated by the time that test mice spent sniffing each cylinder.

#### ***Nesting test***

The ability to build a nest was assessed following previously described procedures [118].

#### ***Acoustic startle reflex and prepulse inhibition***

During training, the mouse was placed in the startle chamber (Cibertec S.A., Madrid, Spain) for 3 min for acclimation; then, baseline startle responses were measured and averaged from 25 to 30 recordings after the presentation of 20 sounds (125 dB, 100 ms long). From this phase, the average response and peak latencies, as well as the peak response, were determined. During prepulse inhibition (PPI) trials, the same sound was preceded (250 ms) by a prepulse stimulus of 85 dB, 50 ms long. Trials including prepulse stimuli were randomly presented with normal startle stimuli, the final total being 25 of each, and the proportion of PPI was determined as  $[(1 - \text{prepulse/startle}) \times 100]$ . The ambient background noise was 70 dB.

#### ***Y-maze***

Each mouse was placed in the center of a Y-maze consisting of three equally sized arms ( $8 \times 40 \times 20$  cm), with white opaque walls at a  $120^\circ$  angle from each other, and allowed to freely explore the arms during 5 min. The number of arm entries and triads were used to calculate the alternation index. An entry was considered to occur when all four limbs were within the arm.

#### ***Object recognition memory***

The object recognition protocol was described extensively elsewhere [92]. Briefly, two equal objects were placed in a rectangular arena ( $55 \times 40 \times 40$  cm) during the training phase. The next day, one object was replaced by a novel one, and the animal's memory of the original object was assessed by comparing the amount of time spent actively exploring the novel object against that for the familiar one using a discrimination index  $[DI = (t_{\text{novel}} - t_{\text{familiar}})/(t_{\text{novel}} + t_{\text{familiar}})]$ . Exploration of an object was defined as directing the nose toward the object at a distance of  $\leq 1.5$  cm or touching the object with the nose or vibrissae. Circling or sitting on the object were not considered exploratory behaviors.

#### ***Step-through passive avoidance test***

During habituation, mice were allowed to explore for 1 min a chamber ( $47 \times 18 \times 26$  cm, Ugo Basile) symmetrically divided into one light and one dark compartment separated by a door. During the training phase, mice were confined to the light compartment for 30 s and then allowed to access the dark compartment. Once inside, the door was closed automatically and the mice received an electrical stimulation (0.3 mA, 5 s) through the metal floor. Retention tests were performed the next day. Here, the latency

to enter into the dark compartment (escape latency) was assessed as a measure of memory retention.

### Electrophysiological recordings

Hippocampal slices were prepared as previously described [119, 120]. Five-month-old mice were anesthetized with isoflurane (2%) and decapitated for slice preparation. Briefly, after decapitation, the whole brain containing the two hippocampi, was placed into ice-cold solution (I) consisting of the following (in mM): 126 NaCl, 3 KCl, 1.25  $\text{KH}_2\text{PO}_4$ , 2  $\text{MgSO}_4$ , 2  $\text{CaCl}_2$ , 26  $\text{NaHCO}_3$ , and 10 glucose (pH 7.2, 300 mOsm), positioned on the stage of a vibratome slicer and cut to obtain transverse hippocampal slices (350  $\mu\text{m}$ ). Slices were maintained continuously oxygenated for at least 1 h before use. All experiments were carried out at room temperature (22–25 °C). For experiments, slices were continuously perfused with the solution described above. Whole-cell patch clamp recording of pyramidal cells located in the CA2 field of the hippocampus was obtained under visual guidance by IR differential interference contrast (DIC) microscopy and were verified as pyramidal cells through their characteristic voltage response to a current step protocol. Neurons were recorded in the current-clamp configuration with a Multiclamp 700B patch clamp amplifier. Data were acquired using pCLAMP 10.2 software (Molecular Devices). To record evoked excitatory postsynaptic currents (eEPSCs), electrical pulses were delivered to Schaffer collateral axons. To evoke inhibitory postsynaptic currents (eIPSCs), electrical pulses were delivered to interneurons situated in the *stratum oriens*. Spontaneous synaptic activity (sPSCs), either excitatory (sEPSCs) or inhibitory (sIPSCs), was recorded from hippocampal CA2 neurons.

Patch electrodes were pulled from borosilicate glass and had a resistance of 4–7 M $\Omega$  when filled with the following (in mM): 120 CsCl, 8 NaCl, 1  $\text{MgCl}_2$ , 0.2  $\text{CaCl}_2$ , 10 HEPES, 2 EGTA, and 20 QX-314 (pH 7.2, 290 mOsm). Experiments were performed at –70 mV. Only cells with a stable resting membrane potential of –55 mV were assessed, and the cell recordings were discarded if the series resistance changed by more than 15%. All recordings were low-pass filtered at 2 kHz and acquired at 10 kHz. Excitatory postsynaptic currents (evoked, eEPSCs, and spontaneous, sEPSCs) were isolated by adding bicuculline (20  $\mu\text{M}$ ) to the perfusion solution to block GABAA receptors. Inhibitory postsynaptic currents (evoked, eIPSCs, and spontaneous, sIPSCs) were isolated adding D-AP5 (50  $\mu\text{M}$ ) and NBQX (10  $\mu\text{M}$ ) to the perfusion solution to block NMDA receptor and AMPA/Kainate receptor-mediated currents, respectively. Signals were averaged every 12 traces. Spontaneous recordings consisted of 60 sweeps/5 s long that were analyzed for amplitude and frequency of detected events.

### Western blotting

Total brain tissue was homogenized in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% (v/v) Triton X-100, 0.1% (w/v) SDS), containing protease inhibitor cocktail (Roche) and phosphatase inhibitors (2 mM sodium orthovanadate, 1 mM sodium pyrophosphate, 10 mM sodium fluoride). Protein concentration was measured using the bicinchoninic acid (BCA) protein assay as specified by the manufacturer (Pierce, Thermo Fisher Scientific). Samples were resolved by SDS-polyacrylamide

gels and transferred onto nitrocellulose membranes. Blocking and incubation with primary and secondary antibodies were performed following the manufacturer instructions. Membranes were developed with the ECL system (GE Healthcare), and acquired images were quantified using ImageJ software. Tubulin or GAPDH loading controls were used when needed. The primary antibodies used were the following: S6 1/1000 (#2217; Cell Signaling), p-S6 (Ser235/236) 1/1000 (#2217; Cell Signaling), S6K1 1/1000 (#9202; Cell Signaling), p-S6K1 (Thr389) 1/1000 (#9205; Cell Signaling), 4E-BP1 1/1000 (#9644; Cell Signaling), AKT 1/1000 (C-20, Santa Cruz Biotechnology), p-AKT (Ser473) 1/1000 (#9271; Cell Signaling), PKC $\alpha$  1/1000 (10/2018, Cell Signaling), PKC $\gamma$  1/1000 (C-4, SC-166385, Santa Cruz Biotechnology), NDRG1 1/1000 (B-5, SC-398291, Santa Cruz Biotechnology), p-NDRG1 (Thr346) 1/1000 (#3217; Cell Signaling), GAPDH (6C5, ab 8245, Abcam), and tubulin 1/1000 (#9026; Sigma-Aldrich). The secondary antibodies used were HRP-labeled anti-mouse 1/2000 (P447-01, Vector) and anti-rabbit 1/2000 (P217-02, Vector).

#### HALEX ancestral protogene reconstruction and in ovo electroporation

In order to obtain an approximate reconstruction of the ORF of the ancestral *HALEX* protogene, a segment of the consensus sequences of HAL1b and inverted LIMEe derived from RepeatMasker, together with the sequence of the *TCEAL7* human gene corresponding to the hypothetical target site duplication of an ancestral LIME element, were combined (Additional file 1: Table S4). Finally, the synthesis of an artificial gene cloned into a pcDNA<sup>™</sup>3.1/*myc*-His was ordered to GenScript.

Eggs from white-Leghorn chickens were incubated at 38.5 °C in an atmosphere with 70% humidity and staged according to the method of Hamburger and Hamilton [121]. In ovo electroporation was performed at stage HH11-12 (48 h of incubation) with DNA plasmids as described previously [117, 122]. Bromo-deoxyuridine (1 mM; Sigma) was injected into the lumen of the chick neural tube at 20 min before harvesting, to label dividing cells. The embryos were recovered at 24 h post-electroporation.

#### Statistical analysis

Enrichment analyses for *BEX* and *TCEAL* genes in gene expression datasets for autism spectrum disorder and schizophrenia were performed for each dataset using a hypergeometric test;  $p < 0.05$  was considered as a threshold for statistical significance. Other statistical analyses were performed using Student's two-tailed  $t$  tests or one-way ANOVA with Tukey's multiple comparison tests, where appropriate. When data were not normally distributed, non-parametric Mann–Whitney tests were used to determine the statistical significance. Calculations were performed with GraphPad Prism version 6 or in R; \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.005$ , \*\*\*\* $p < 0.001$ .

#### Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13059-020-02172-3>.

**Additional file 1:** Fig. S1 The *Tceal7* gene is derived from the domestication of L1 retrotransposon fragments. Fig. S2 The *Bex/Tceal* gene cluster was established before the diversification of extant eutherians. Fig. S3 Highly diverged *BEX/TCEAL* proteins share a coiled coil domain. Fig. S4 *BEX/TCEAL* proteins might have inherited some of their structural properties from the ancestral transposon. Fig. S5 Selection pressure analyses reveal signatures of positive selection in the *Bex/Tceal* genes. Fig. S6 A BGW-like sequence was already present in the *GLA* promoter of the last therian common ancestor. Fig. S7 *Bex/Tceal* genes show tissue-enriched expression patterns during

development. Fig. S8 *Bex3* and *Tceal7* genes, but not the ancestral HALEX element, induce cell proliferation in chicken neural tube. Fig. S9 The deletions introduced using CRISPR-Cas9 technology can be observed in the mRNA expressed from the *Bex3* mutant alleles. Fig. S10 CRISPR-Cas9-generated *Bex3* mutant alleles show subtle skull abnormalities. Fig. S11 *Bex3* mutant mice show normal acoustic startle reflex. Fig. S12 *Bex3* deficiency leads to aberrant mTOR signaling in the brain. Table S1 Coding genes putatively derived from transposable elements in the human and mouse genomes. Table S2 Altered expression of *BEX* and *TCEAL* genes in subjects with autism spectrum disorder or schizophrenia. Table S3 Enrichment of differential gene expression in *BEX* and *TCEAL* gene families in subjects with autism spectrum disorder or schizophrenia. Table S4 Primers and reconstructed gene sequences used in this work. Supplementary references.

**Additional file 2.** Review history.

### Acknowledgments

We wish to thank Claudia Pérez and Alba del Valle for technical support, Jose Ramón Bayascas for help with western blots, Alejandro Sanchez-Gracia and Julio Rozas for advice on selection analyses, and Manuel Irimia and Ignacio Maeso for helpful discussions.

### Peer review information

Kevin Pang was the primary editor on this article and managed its peer review and editorial process in collaboration with the rest of the editorial team.

### Review history

The review history is available as Additional file 2.

### Authors' contributions

E.N.P. designed the study, performed molecular biology and in situ hybridization experiments, did bioinformatics and histologic analyses, and wrote the manuscript. C.V.G. designed the study, performed molecular biology experiments, and wrote the manuscript. S.M. did molecular biology experiments, performed histological analyses, and wrote the manuscript. D.B. designed the study, performed bioinformatics analyses and molecular biology experiments, and wrote the manuscript. N.F.C. did bioinformatics analyses. J.L.F. supervised in situ hybridization experiments in mice. M.L.M. did all skull measurements and preliminary behavioral tests. M.A.N. performed behavioral tests. I.S.-P. performed behavioral tests. E.A.G. did bioinformatics analyses. F.U. supervised molecular biology analyses. C.H.U. conducted selection pressure analysis. P.C. performed bioinformatics analyses. R.F.M. performed excitation/inhibition balance experiments. A.R.M. supervised excitation/inhibition balance experiments. S.D'A. helped in writing the manuscript. B.C. supervised bioinformatics analyses and provided resources. G.M. supervised molecular biology experiments and provided resources. E.S. supervised molecular biology and histology experiments and provided resources. A.M.C. performed and supervised behavioral and histology experiments and provided resources. J.J.C. designed the study, generated the mouse mutant lines, performed and supervised molecular biology experiments, and provided resources. J.G.F. designed the study and provided resources and supervision. The author(s) read and approved the final manuscript.

### Authors' information

Twitter handles: @cvcgar (Cristina Vicente-García); @MirraSerena (Serena Mirra); @Noeliafdez6 (Noèlia Fernàndez-Castillo); @jlferran (José Luis Ferrán); @maclopmay (Macarena López-Mayorga); @yomellamocar (Carlos Herrera-Úbeda); @Salv\_DANIELLO (Salvatore D'Aniello); @bcormand (Bru Cormand); @GMarfanyN (Gemma Marfany); @jcarvajal (Jaime J. Carvajal); @jordigarciafdez (Jordi Garcia-Fernàndez).

### Funding

Major financial support for this research was received from Spanish "Ministerio de Ciencia, Innovación y Universidades." Grants BFU2015-68655-P and BFU2017-861152-P to J.G.F., RTI2018-100968-B-I00, 2017-SGR-738, H2020/2014-2020 under grant agreements n°667302, n°643051, and n°728018 to B.C., PGC2018-098229-B-I00 to J.L.F., BES-2016-077374 to E.A.-G., CVI-7290 Junta de Andalucía to A.R.M., SAF2016-80937-R (Ministerio de Economía y Competitividad/FEDER) to G.M., Institutional Grant MDM-2016-0687 (María de Maeztu Excellence Unit, Department of Gene Regulation and Morphogenesis at CABD) and BFU2017-83150-P to J.J.C., BFU2017-89780-R and P12-CTS-2257 to A.M.C. and SAF2016-76340-R and María de Maeztu Excellence Unit, Institute of Neurosciences to E.S.. E.N.P. held an FPI pre-doctoral fellowship (Spanish "Ministerio de Ciencia, Innovación y Universidades"). S.M. was first supported by a contract with the "Centro de Investigación Biomédica en Enfermedades Neurodegenerativas," and later by "Centro de Investigación Biomédica en Red de Enfermedades Raras" (CIBERER). N.F.C. is also under contract by CIBERER. This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available at <http://decipher.sanger.ac.uk> and via email from [decipher@sanger.ac.uk](mailto:decipher@sanger.ac.uk). Funding for the project was provided by the Wellcome Trust.

### Availability of data and materials

All data generated during this study are included in this published article and its supplementary information files. All data reanalyzed in this study is publicly available from the sources specified in the methods. Specifically, RNA-seq datasets SRP016501 (GSE41637, *Bos taurus*), SRP114662 (GSE20113, *Canis familiaris*), and SRP012922 (GSE106077, *Dasypus novemcinctus*) were downloaded from SRA database, while protein-coding cDNA sequences were obtained from Ensembl (<http://www.ensembl.org>), and cRPKM values for human and mouse tissues, from VASTDB [110]. Further, human transcriptomic datasets for schizophrenia and autism spectrum disorder were obtained from GEO database (GSE38322, GSE35978, GSE53987, GSE87610) or published articles (Additional file 1: Supplementary references [16–18, 21–26]).

**Ethics approval and consent to participate**

Experiments using animals were performed under protocols approved by the Universidad Pablo de Olavide Ethical Committee (Sevilla, Spain; protocols 24/04/2018/056 and 24/04/2018/043) and the Ethical Committee for Animal Experimentation (AEC) of the Generalitat of Catalonia (protocol 9431) in accordance with Spanish Royal Decree 53/2013, European Directive 2010/63/EU, and other relevant guidelines.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare no competing interests.

**Author details**

<sup>1</sup>Department of Genetics, Microbiology and Statistics, Faculty of Biology, and Institut de Biomedicina (IBUB), University of Barcelona, 08028 Barcelona, Spain. <sup>2</sup>Centro Andaluz de Biología del Desarrollo, CSIC-UPO-JA, Universidad Pablo de Olavide, 41013 Sevilla, Spain. <sup>3</sup>Department of Cell Biology, Physiology and Immunology, and Institute of Neurosciences, University of Barcelona, 08028 Barcelona, Spain. <sup>4</sup>Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III (ISCIII), Madrid, Spain. <sup>5</sup>Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain. <sup>6</sup>Department of Zoology, Charles University, Vinicna 7, 12844 Prague, Czech Republic. <sup>7</sup>Institut de Recerca Sant Joan de Déu (IR-SJD), Esplugues de Llobregat, 08950 Barcelona, Spain. <sup>8</sup>Department of Human Anatomy, School of Medicine, University of Murcia and IMIB-Arrixaca Institute, 30120 Murcia, Spain. <sup>9</sup>Department of Physiology, Anatomy and Cell Biology, Universidad Pablo de Olavide, 41013 Sevilla, Spain. <sup>10</sup>Present Address: Instituto de Neurociencias de Alicante (Universidad Miguel Hernández - Consejo Superior de Investigaciones Científicas), Alicante, Spain. <sup>11</sup>Present Address: Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Neuropsychopharmacology and psychobiology research group, UCA, INiBICA, Cádiz, Spain. <sup>12</sup>Genome Architecture, Gene Regulation, Stem Cells and Cancer Programme, Centre for Genomic Regulation (CRG), the Barcelona Institute of Science and Technology, 08003 Barcelona, Spain. <sup>13</sup>Università Pompeu Fabra (UPF), 08003 Barcelona, Spain. <sup>14</sup>Department of Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn, 80121 Naples, Italy. <sup>15</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

Received: 25 April 2020 Accepted: 25 September 2020

Published online: 26 October 2020

**References**

1. Tautz D, Domazet-Lošo T. The evolutionary origin of orphan genes. *Nat Rev Genet.* 2011;12:692–702.
2. Miller WJ, McDonald JF, Pinsker W. Molecular domestication of mobile elements. *Genetica.* 1997;100:261–70.
3. Fugmann SD. The origins of the Rag genes—from transposition to V(D) J recombination. *Semin Immunol.* 2010;22:10–6.
4. Kim AJ, Lee CS, Schlessinger D. Bex3 associates with replicating mitochondria and is involved in possible growth control of F9 teratocarcinoma cells. *Gene.* 2004;343:79–89.
5. Chien J, Narita K, Rattan R, Giri S, Shridhar R, Staub J, et al. A role for candidate tumor-suppressor gene TCEAL7 in the regulation of c-Myc activity, cyclin D1 levels and cellular transformation. *Oncogene.* 2008;27(58):7223–34.
6. Rattan R, Narita K, Chien J, Maguire JL, Shridhar S, Giri S, et al. TCEAL7, a putative tumor suppressor gene, negatively regulates NF- $\kappa$ B pathway. *Oncogene.* 2009;29:1362–73.
7. Kazi JU, Kabir NN, Ronnstrand L. Brain-expressed X-linked (BEX) proteins in human cancers. *Biochim Biophys Acta.* 2015;1856:226–33.
8. Gao W, Li JZ, Chen SQ, Chu CY, Chan JY, Wong TS. BEX3 contributes to cisplatin chemoresistance in nasopharyngeal carcinoma. *Cancer Med.* 2017;6(2):439–51.
9. Ward C, Cauchy P, Garcia P, Frampton J, Esteban MA, Volpe G. High WBP5 expression correlates with elevation of HOX genes levels and is associated with inferior survival in patients with acute myeloid leukaemia. *Sci Rep.* 2020;10(1):3505.
10. Schwarz BA, Cetinbas M, Clement K, Walsh RM, Cheloufi S, Gu H, et al. Prospective isolation of poised iPSC intermediates reveals principles of cellular reprogramming. *Cell Stem Cell.* 2018;23(2):289–305.
11. Calvo L, Anta B, López-Benito S, Martín-Rodríguez C, Lee FS, Pérez P, et al. Bex3 dimerization regulates NGF-dependent neuronal survival and differentiation by enhancing trkA gene transcription. *J Neurosci.* 2015;35:7190–202.
12. Vilar M, Murillo-Carretero M, Mira H, Magnusson K, Besset V, Ibáñez CF. Bex1, a novel interactor of the p75 neurotrophin receptor, links neurotrophin signaling to the cell cycle. *EMBO J.* 2006;25:1219–30.
13. Judd J, Lovas J, Huang GN. Defined factors to reactivate cell cycle activity in adult mouse cardiomyocytes. *Sci Rep.* 2019;9(1):18830.
14. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature.* 2000;403:785–9.
15. Blaise S, de Parseval N, Bénéit L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci U S A.* 2003;100:13013–8.
16. Robertson HM, Zumpano KL. Molecular evolution of an ancient *mariner* transposon, *Hsmarl*, in the human genome. *Gene.* 1997;205:203–17.
17. Hayward A, Ghazal A, Andersson G, Andersson L, Jern P. ZBED evolution: repeated utilization of DNA transposons as regulators of diverse host functions. *PLoS One.* 2013;8:e59940.
18. Campillos M, Doerks T, Shah PK, Bork P. Computational characterization of multiple gag-like human proteins. *Trends Genet.* 2006;22:585–9.
19. Ostertag EM, Kazazian J. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* 2001;11:2059–65.



20. Athanikar JN, Badge RM, Moran JV. A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* 2004;32:3846–55.
21. Winter EE, Ponting CP. Mammalian BEX, WEX and GASP genes: coding and non-coding chimaerism sustained by gene conversion events. *BMC Evol Biol.* 2005;5:54.
22. Giordano J, Ge Y, Gelfand Y, Abrusán G, Benson G, Warburton PE. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comput Biol.* 2007;3:1321–34.
23. Cabral KMS, Raymundo DP, Silva VS, Sampaio LAG, Johanson L, Hill LF, et al. Biophysical studies on BEX3, the p75NTR-associated cell death executor, Reveal a High-Order Oligomer with Partially Folded Regions. *PLoS One.* 2015;10:e0137916.
24. Fernandez EM, Díaz-Ceso MD, Vilar M. Brain expressed and X-linked (Bex) proteins are intrinsically disordered proteins (IDPs) and form new signaling hubs. *PLoS One.* 2015;10:e0117206.
25. do Amaral MJ, Araujo TS, Díaz NC, Accornero F, Polycarpo CR, Cordeiro Y, et al. Phase separation and disorder-to-order transition of human brain expressed X-Linked 3 (hBEX3) in the presence of small fragments of tRNA. *J Mol Biol.* 2020; 432(7):2319–48.
26. Zhang L. Adaptive evolution and frequent gene conversion in the brain expressed X-linked gene family in mammals. *Biochem Genet.* 2008;46(5–6):293–311.
27. Ratnakumar A, Mousset S, Glémin S, Berglund J, Galtier N, Duret L, et al. Detecting positive selection within genomes: the problem of biased gene conversion. *Philos Trans R Soc B Biol Sci.* 2010;365:2571–80.
28. López-Doménech G, Serrat R, Mirra S, D'Aniello S, Somorjai I, Abad A, et al. The Eutherian *Armcx* genes regulate mitochondrial trafficking in neurons and interact with Miro and Trak2. *Nat Commun.* 2012;3:814.
29. Shi X, Garry DJ. Myogenic regulatory factors transactivate the *Tceal7* gene and modulate muscle differentiation. *Biochem J.* 2010;428:213–21.
30. Iseki H, Takeda A, Andoh T, Takahashi N, Kurochkin IV, Yarmishyn A, et al. Human Arm protein lost in epithelial cancers, on chromosome X 1 (ALEX1) gene is transcriptionally regulated by CREB and Wnt/beta-catenin signaling. *Cancer Sci.* 2010;101:1361–6.
31. Andrianaki A, Siapati EK, Hirata RK, Russell DW, Vassilopoulos G. Dual transgene expression by foamy virus vectors carrying an endogenous bidirectional promoter. *Gene Ther.* 2010;17:380–8.
32. Alvarez E, Zhou W, Witt SE, Freed CR. Characterization of the Bex gene family in humans, mice, and rats. *Gene.* 2005; 357:18–28.
33. Koo JH, Saraswati M, Margolis FL. Immunolocalization of Bex protein in the mouse brain and olfactory system. *J Comp Neurol.* 2005;487:1–14.
34. Ito K, Yamazaki S, Yamamoto R, Tajima Y, Yanagida A, Kobayashi T, et al. Gene targeting study reveals unexpected expression of brain-expressed X-linked 2 in endocrine and tissue stem/progenitor cells in mice. *J Biol Chem.* 2014;289: 29892–911.
35. Accornero F, Schips TG, Petrosino JM, Gu S-Q, Kanisicak O, van Berlo JH, et al. BEX1 is an RNA-dependent mediator of cardiomyopathy. *Nat Commun.* 2017;8:1875.
36. Yu W, Yaping L, Mingjun W, Jie H, Xiaogang L, Gang L. BEX4 upregulation alters Sertoli cell growth properties and protein expression profiles: An explanation for cadmium-induced testicular Sertoli cell injury. *J Biochem Mol Toxicol.* 2017;31 <https://doi.org/10.1002/jbt.21908>.
37. Gu Y, Wei W, Cheng Y, Wan B, Ding X, Wang H, et al. A pivotal role of BEX1 in liver progenitor cell expansion in mice. *Stem Cell Res Ther.* 2018;9:164.
38. Myers A, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, et al. A survey of genetic human cortical gene expression. *Nat Genet.* 2007;39(12):1494–9.
39. Khazaei MR, Halfter H, Karimzadeh F, Koo JH, Margolis FL, Young P. Bex1 is involved in the regeneration of axons after injury. *J Neurochem.* 2010;115:910–20.
40. Yamamoto T, Wilsdon A, Joss S, Isidor B, Erlandsson A, Suri M, et al. An emerging phenotype of Xq22 microdeletions in females with severe intellectual disability, hypotonia and behavioral abnormalities. *J Hum Genet.* 2014;59:300–6.
41. Shirai K, Higashi Y, Shimojima K, Yamamoto Y. An Xq22.1q22.2 nullisomy in a male patient with severe neurological impairment. *Am J Med Genet Part A.* 2017;173:1124–7.
42. Hijazi H, Coelho FS, Gonzaga-Jauregui C, Bernardini L, Mar SS, Manning MA, et al. Xq22 deletions and correlation with distinct neurological disease traits in females: further evidence for a contiguous gene syndrome. *Hum Mutat.* 2020;41: 150–68.
43. Mukai J, Hachiya T, Shoji-Hoshino S, Kimura MT, Nadano D, Suvanto P, et al. NADE, a p75NTR-associated cell death executor, is involved in signal transduction mediated by the common neurotrophin receptor p75NTR. *J Biol Chem.* 2000;275:17566–70.
44. Mukai J, Shoji S, Kimura MT, Okubo S, Sano H, Suvanto P, et al. Structure-function analysis of NADE: identification of regions that mediate nerve growth factor-induced apoptosis. *J Biol Chem.* 2002;277:13973–82.
45. Fombonne E, Rogé B, Claverie J, Courty S, Frémolle J. Microcephaly and macrocephaly in autism. *J Autism Dev Disord.* 1999;29:113–9.
46. Wright IC, Rabe-Hesketh S, Woodruff PWR, David AS, Murray RM, Bullmore ET. Meta-analysis of regional brain volumes in schizophrenia. *Am J Psychiatry.* 2000;157:16–25.
47. Turner AH, Greenspan KS, van Erp TGM. Pallidum and lateral ventricle volume enlargement in autism spectrum disorder. *Psychiatry Res Neuroimaging.* 2016;252:40–5.
48. Lijam N, Paylor R, McDonald MP, Crawley JN, Deng CX, Herrup K, et al. Social interaction and sensorimotor gating abnormalities in mice lacking *Dvl1*. *Cell.* 1997;90:895–905.
49. El-Kordi A, Winkler D, Hammerschmidt K, Kästner A, Krueger D, Ronnenberg A, et al. Development of an autism severity score for mice using *Nlgn4* null mutants as a construct-valid model of heritable monogenic autism. *Behav Brain Res.* 2013;251:41–9.
50. Kohl S, Heekeren K, Klosterkötter J, Kuhn J. Prepulse inhibition in psychiatric disorders—apart from schizophrenia. *J Psychiatr Res.* 2013;47:445–52.
51. Marín O. Interneuron dysfunction in psychiatric disorders. *Nat Rev Neurosci.* 2012;13:107–20.

52. Pelkey KA, Chittajallu R, Craig MT, Tricoire L, Wester JC, McBain CJ. Hippocampal gabaergic inhibitory interneurons. *Physiol Rev*. 2017;97:1619–747.
53. Hitti FL, Siegelbaum SA. The hippocampal CA2 region is essential for social memory. *Nature*. 2014;508:88–92.
54. Dudek SM, Alexander GM, Farris S. Rediscovering area CA2: unique properties and functions. *Nat Rev Neurosci*. 2016;17:89–102.
55. Yasui S, Tsuzaki K, Ninomiya H, Floricel F, Asano Y, Maki H, et al. The TSC1 gene product hamartin interacts with NADE. *Mol Cell Neurosci*. 2007;35:100–8.
56. Huang J, Dibble CC, Matsuzaki M, Manning BD. The TSC1-TSC2 complex is required for proper activation of mTOR complex 2. *Mol Cell Biol*. 2008;28:4104–15.
57. Lipton JO, Sahin M. The neurology of mTOR. *Neuron*. 2014;84:275–91.
58. McLaughlin RN, Young JM, Yang L, Neme R, Wichman HA, Malik HS. Positive selection and multiple losses of the LINE-1-derived L1TD1 gene in mammals suggest a dual role in genome defense and pluripotency. *PLoS Genet*. 2014;10:e1004531.
59. Shaheen M, Williamson E, Nickoloff J, Lee S-H, Hromas R. Metnase/SETMAR: a domesticated primate transposase that enhances DNA repair, replication, and decatenation. *Genetica*. 2010;138:559–66.
60. Gewurz BE, Towfic F, Mar JC, Shinnars NP, Takasaki K, Zhao B, et al. Genome-wide siRNA screen for mediators of NF- $\kappa$ B activation. *Proc Natl Acad Sci U S A*. 2012;109:2467–72.
61. Arroyo R, Suñé G, Zanzoni A, Duran-Frigola M, Alcalde V, Stracker TH, et al. Systematic identification of molecular links between core and candidate genes in breast cancer. *J Mol Biol*. 2015;427:1436–50.
62. Hu Z, Wang Y, Huang F, Chen R, Li C, Wang F, et al. Brain-expressed X-linked 2 is pivotal for hyperactive mechanistic target of rapamycin (mTOR)-mediated tumorigenesis. *J Biol Chem*. 2015;290:25756–65.
63. Lee JK, Lee J, Go H, Lee CG, Kim S, Kim HS, et al. Oncogenic microtubule hyperacetylation through BEX4-mediated sirtuin 2 inhibition. *Cell Death Dis*. 2016;7:e2336.
64. Zhao Z, Li J, Tan F, Gao S, He J. mTOR up-regulation of BEX4 promotes lung adenocarcinoma cell proliferation by potentiating OCT4. *Biochem Biophys Res Commun*. 2018;500:302–9.
65. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, et al. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol*. 2006;2(8):e100.
66. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell*. 2012;149:274–93.
67. Huang J, Manning BD. A complex interplay between Akt, TSC2 and the two mTOR complexes. *Biochem Soc Trans*. 2009;37:217–22.
68. Martin TD, Dennis MD, Gordon BS, Kimball SR, Jefferson LS. mTORC1 and JNK coordinate phosphorylation of the p70S6K1 autoinhibitory domain in skeletal muscle following functional overloading. *Am J Physiol Endocrinol Metab*. 2014;306:E1397–405.
69. Biever A, Valjent E, Puighermanal E. Ribosomal protein S6 phosphorylation in the nervous system: from regulation to function. *Front Mol Neurosci*. 2015;8:75.
70. Kokošar J, Kordiš D. Genesis and regulatory wiring of retroelement-derived domesticated genes: a phylogenomic perspective. *Mol Biol Evol*. 2013;30:1015–31.
71. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet*. 2009;41(5):563–71.
72. Villanueva-Cañas JL, Ruiz-Orera J, Agea MI, Gallo M, Andreu D, Albà MM. New genes and functional innovation in mammals. *Genome Biol Evol*. 2017;9:1886–900.
73. De Angelis MT, Russo F, D'Angelo F, Federico A, Gemei M, Del Vecchio L, et al. Novel pancreas organogenesis markers refine the pancreatic differentiation roadmap of embryonic stem cells. *Stem Cell Rev Rep*. 2014;10:269–79.
74. Gergics P, Christian HC, Choo MS, Ajmal A, Camper SA. Gene expression in mouse thyrotrope adenoma: transcription elongation factor stimulates proliferation. *Endocrinology*. 2016;157:3631–46.
75. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*. 2019;570:332–7.
76. Firth HV, Richards SM, Bevan AP, Clayton S, Corvas M, Rajan D, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am J Hum Genet*. 2009;84:524–33.
77. Silverman JL, Yang M, Lord C, Crawley JN. Behavioural phenotyping assays for mouse models of autism. *Nat Rev Neurosci*. 2010;11:490–502.
78. Sandi C, Haller J. Stress and the social brain: behavioural effects and neurobiological mechanisms. *Nat Rev Neurosci*. 2015;16:290–304.
79. del Pino I, Rico B, Marín O. Neural circuit dysfunction in mouse models of neurodevelopmental disorders. *Curr Opin Neurobiol*. 2018;48:174–82.
80. Langen M, Durston S, Kas MJH, van Engeland H, Staal WG. The neurobiology of repetitive behavior: ... and men. *Neurosci Biobehav Rev*. 2011;35:356–65.
81. Selby L, Zhang C, Sun Q-Q. Major defects in neocortical GABAergic inhibitory circuits in mice lacking the fragile X mental retardation protein. *Neurosci Lett*. 2007;412:227–32.
82. Gant JC, Thibault O, Blalock EM, Yang J, Bachstetter A, Kotick J, et al. Decreased number of interneurons and increased seizures in neuropilin 2 deficient mice: implications for autism and epilepsy. *Epilepsia*. 2009;50:629–45.
83. Gogolla N, LeBlanc JJ, Quast KB, Südhof TC, Fagiolini M, Hensch TK. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J Neurodev Disord*. 2009;1:172–81.
84. Lawrence YA, Kemper TL, Bauman ML, Blatt GJ. Parvalbumin-, calbindin-, and calretinin-immunoreactive hippocampal interneuron density in autism. *Acta Neurol Scand*. 2010;121:99–108.
85. Zikopoulos B, Barbas H. Altered neural connectivity in excitatory and inhibitory cortical circuits in autism. *Front Hum Neurosci*. 2013;7:609.
86. Stoner R, Chow ML, Boyle MP, Sunkin SM, Mouton PR, Roy S, et al. Patches of disorganization in the neocortex of children with autism. *N Engl J Med*. 2014;370:1209–19.
87. Wöhr M, Orduz D, Gregory P, Moreno H, Khan U, Vörckel KJ, et al. Lack of parvalbumin in mice leads to behavioral deficits relevant to all human autism core symptoms and related neural morphofunctional abnormalities. *Transl Psychiatry*. 2015;5:e525.

88. Cai Y, Tang X, Chen X, Li X, Wang Y, Bao X, et al. Liver X receptor  $\beta$  regulates the development of the dentate gyrus and autistic-like behavior in the mouse. *Proc Natl Acad Sci U S A*. 2018;115:E2725–33.
89. Goh S, Peterson BS. Imaging evidence for disturbances in multiple learning and memory systems in persons with autism spectrum disorders. *Dev Med Child Neurol*. 2012;54:208–13.
90. Baarendse PJJ, Van Grootheest G, Jansen RF, Pieneman AW, Ögren SO, Verhage M, et al. Differential involvement of the dorsal hippocampus in passive avoidance in C57BL/6J and DBA/2J mice. *Hippocampus*. 2008;18:11–9.
91. Winters BD, Saksida LM, Bussey TJ. Object recognition memory: neurobiological mechanisms of encoding, consolidation and retrieval. *Neurosci Biobehav Rev*. 2008;32:1055–70.
92. Suárez-Pereira I, Canals S, Carrión AM. Adult newborn neurons are involved in learning acquisition and long-term memory formation: the distinct demands on temporal neurogenesis of different cognitive tasks. *Hippocampus*. 2015;25:51–61.
93. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay LA, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet*. 2013;9:e1003470.
94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
95. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25:1189–91.
96. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74.
97. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9.
98. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59:307–21.
99. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*. *Mol Biol Evol*. 2018;35:518–22.
100. Rambaut A. FigTree v1.4.2, a graphical viewer of phylogenetic trees. Available from <http://tree.bio.ed.ac.uk/software/figtree/>, 2014.
101. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol*. 2012;61(6):1061–7.
102. Gruber M, Söding J, Lupas AN. Comparative analysis of coiled-coil prediction methods. *J Struct Biol*. 2006;155(2):140–5.
103. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999;292:195–202.
104. Buchan DWA, Jones DT. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res*. 2019;47:W402–7.
105. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015;31:857–63.
106. Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*. 2005;21(5):676–9.
107. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 2012;8(7):e1002764.
108. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 2015;32(5):1342–53.
109. Labbé RM, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DDR, et al. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells*. 2012;30:1734–45.
110. Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, et al. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res*. 2017;27:1759–68.
111. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A, et al. Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res*. 2016;44:W147–53.
112. Ferran JL, Ayad A, Merchán P, Morales-Delgado N, Sánchez-Arrones L, Alonso A, et al. Exploring brain genoarchitecture by single and double chromogenic in situ hybridization (ISH) and immunohistochemistry (IHC) on cryostat, paraffin, or floating sections. In: Hauptmann G, editor. *In Situ Hybridization Methods, Neuromethods*, vol. 99. New York: Springer; 2015. p. 83–107.
113. Moreno-Mateos MA, Vejnar CE, Beaudoin JD, Fernandez JP, Mis EK, Khokha MK, et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods*. 2015;12:982–8.
114. Casaca A, Nóvoa A, Mallo M. Hoxb6 can interfere with somitogenesis in the posterior embryo through a mechanism independent of its rib-promoting activity. *Development*. 2016;143:437–48.
115. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–21.
116. Lobjois V, Bel-Vialar S, Trousse F, Pituello F. Forcing neural progenitor cells to cycle is insufficient to alter cell-fate decision and timing of neuronal differentiation in the spinal cord. *Neural Dev*. 2008;3:4.
117. Mirra S, Ulloa F, Gutierrez-Vallejo I, Marti E, Soriano E. Function of *Armcx3* and *Armc10/SVH* genes in the regulation of progenitor proliferation and neural differentiation in the chicken spinal cord. *Front Cell Neurosci*. 2016;10:47.
118. Deacon RMJ. Assessing nest building in mice. *Nat Protoc*. 2006;1:1117–9.
119. Andrade-Talavera Y, Duque-Feria P, Paulsen O, Rodríguez-Moreno A. Presynaptic spike timing-dependent long-term depression in the mouse hippocampus. *Cereb Cortex*. 2016;26:3637–54.
120. González-Rodríguez P, Ugidos IF, Pérez-Rodríguez D, Anuncibay-Soto B, Santos-Galdiano M, Font-Belmonte E, et al. Brain-derived neurotrophic factor alleviates the oxidative stress induced by oxygen and glucose deprivation in an ex vivo brain slice model. *J Cell Physiol*. 2019;234:9592–604.
121. Hamburger V, Hamilton HL. A series of normal stages in the development of the chick embryo. *Dev Dyn*. 1992;195:231–72.
122. Megason SG, McMahon AP. A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development*. 2002;129:2087–98.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.