Oct4 is a key regulator of vertebrate trunk length diversity

Rita Aires[1], Arnon D. Jurberg[1]†, Francisca Leal[2], Ana Nóvoa[1], Martin J. Cohn[2, 3] and Moisés Mallo[1*]

[1] Instituto Gulbenkian de Ciencia, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

[2] Department of Biology, University of Florida, P.O. Box 103610, Gainesville, FL 32610, USA.

[3] Howard Hughes Medical Institute, Department of Molecular Genetics and Microbiology, Genetics Institute, University of Florida. P.O. Box 103610, Gainesville, FL 32610, USA.

† Current address: Laboratory on Thymus Research, Oswaldo Cruz Institute, Oswaldo Cruz Foundation/Fiocruz, Av. Brasil, 4365. Pavilhão Leônidas Deane, Manguinhos, Rio de Janeiro-RJ, Brazil. 21040-360 and Graduate Program in Cell and Developmental Biology, Institute of Biomedical Sciences, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

* Author for correspondence: mallo@igc.gulbenkian.pt

SUMMARY

Vertebrates exhibit a remarkably broad variation in trunk and tail lengths. However, the evolutionary and developmental origins of this diversity remain largely unknown. Posterior *Hox* genes were proposed to be major players in trunk length diversification in vertebrates but functional studies have so far failed to support this view. Here we identify the pluripotency factor *Oct4* as a key regulator of trunk length in vertebrate embryos. Maintaining high *Oct4* levels in axial progenitors throughout development was sufficient to extend trunk length in mouse embryos. *Oct4* also shifted posterior *Hox* gene expression boundaries in the extended trunks, thus providing a link between activation of these genes and the transition into tail development. Furthermore, we show that the exceptionally long trunks of snakes are likely to result from heterochronic changes in *Oct4* activity during body axis extension, which may have derived from differential genomic rearrangements at the *Oct4* locus during vertebrate evolution.

INTRODUCTION

The Vertebrate clade encompasses remarkable diversity of body shapes and sizes, yet the causes of this variation are poorly understood. Research into this topic has been particularly focused in the wide differences in body length and in regional patterns along the main body axis of vertebrates (Burke et al., 1995; Cohn and Tickle, 1999; Di-Poï et al., 2010; Gomez et al., 2008; Head and Polly, 2015; Woltering, 2012; Woltering et al., 2009). These studies often combined the analysis of animals with extreme variants of specific features (like the long necks of birds or the large rib numbers of snakes) with functional approaches in tractable animal models, allowing for a direct experimental evaluation of the factors potentially involved in generating those patterns. *Hox* genes were identified this way as having a central role in the generation of the different vertebral patterns observed along the body's anterior-posterior axis and in its variations among species (Di-Poï et al., 2010; Head and Polly, 2015; Woltering et al., 2009). However, the mechanisms regulating other global features of the vertebrate body plan are much less understood. A particularly interesting case is regionalization of the body into head, trunk and tail regions, given that their relative proportions vary widely among vertebrates. Snake bodies provide an extreme example of an uneven distribution of these regions, since these animals are mostly composed of very long, organ-filled, trunks. Snake trunks are not just a simple consequence of their body length, since other vertebrates with remarkably long bodies, like some lizards, are mostly dominated by a long tail and have their organs confined to a relatively small trunk. Also, although long trunks are closely associated with rib-bearing vertebrae, functional studies indicate that *Hox* genes play no major role in the regulation of overall trunk length (Carapuço et al., 2005; Jurberg et al., 2013; Mallo et al., 2010; Vinagre et al., 2010; Wellik and Capecchi, 2003), even if they do determine the identity of its segments. This shows that Hox genes are downstream effectors of a still unknown mechanism controlling body region distribution.

Body region allocation and the transition between regions have important morphological, physiological and evolutionary consequences. The partitioning of the body into morphologically discrete regions is defined during embryonic development through the process of axial extension. During this process, the vertebrate body is built progressively from head to tail by the sequential addition of new tissue produced by dedicated axial progenitors located at its posterior end (Wilson et al., 2009; Wymeersch et al., 2016). Embryological studies indicate that, while

continuous, axial extension relies on different mechanisms to generate trunk or tail structures. Axial elongation through the trunk requires activity of various types of progenitors located within the epiblast, an epithelial sheet at the posterior embryonic end (Stern et al., 2006; Wilson et al., 2009; Wymeersch et al., 2016). These include neuro-mesodermal progenitors (NMPs) that elongate the neural tube and lay down the paraxial mesoderm that will form the musculoskeletal case of the trunk (Cambray and Wilson, 2007; Henrique et al., 2015; Stern et al., 2006; Tzouanacou et al., 2009), and progenitors for the intermediate and lateral mesoderm, which together with the endoderm will build the trunk organs involved in digestive, excretory and reproductive functions (Stern et al., 2006). In contrast, tail development derives almost exclusively from the activity of NMPs that at this stage are embedded within the tail bud mesenchyme (Wilson et al., 2009). Transition from trunk to tail development thus entails a number of morphological and functional changes, including the terminal differentiation of the progenitors for the lateral and intermediate mesoderm and relocation of NMPs to the tail bud (Jurberg et al., 2013).

Genetic data indicate that Gdf11 signaling is a key regulator of the trunk to tail transition. *Gdf11* mutant mice have longer trunks due to the delayed onset of this transition (Jurberg et al., 2013; McPherron et al., 1999). Conversely, premature activation of Gdf11 signaling produces embryos with smaller or even absent trunks as the result of an early transition into tail-producing mechanisms (Jurberg et al., 2013). The intensification of the trunk length phenotype of *Gdf11* mutants when *Gdf8* (*Myostatin*) is simultaneously inactivated indicates partial compensation of Gdf11 activity by *Gdf8* (McPherron et al., 2009). The extended trunks in *Gdf11* mutants suggest that Gdf11 counteracts the activity of hitherto non-identified trunk-promoting factors as part of the program that activates tail development. The identification of these factors might provide important insights into the mechanisms responsible for trunk length diversity among vertebrates.

Here we show that the pluripotency factor *Oct4* is a key regulator of vertebrate trunk length. Sustained *Oct4* expression in axial progenitors during embryonic elongation using a transgenic approach in mice resulted in extended trunks. This factor was also able to coordinate other aspects of anterior-posterior body patterning, most notably the activation of posterior *Hox* genes. Our results also indicate that persistent *Oct4* expression could be in the origin of the extreme length of snake trunks, since its expression in snake embryos seems to remain active for a longer developmental period when compared to mouse embryos. These heterochronic changes

in *Oct4* expression could have resulted from extensive genomic rearrangements during the evolution of vertebrates, generating entirely different genomic configurations in squamates and mammals 5' from the *Oct4* gene. *Oct4* would thereby serve as a link between the overall trunk length and patterning of the paraxial mesoderm, coordinating axial extension with the production of the appropriate skeletal structures for the trunk and tail regions through *Hox* genes.

RESULTS

Incomplete epiblast resolution in *Gdf11* mutant embryos

We have recently described that *Gdf11* mutant embryos have an ectopic *T* (*Brachyury*)-positive tissue in their tails (Jurberg et al., 2013) (Figure 1B). Considering the important role of Gdf11 signaling in setting the trunk to tail transition, we hypothesized that this ectopic mass could contain progenitor cells still displaying trunk-generating characteristics. A close examination of this tissue revealed that it was composed of two main elements: an epithelial sheet continuous with the tail-associated gut, projecting inward towards the gut's lumen, and a mass of mesenchymal cells mostly confined within the epithelial pocket (Figures 1 and S1A,B). Expression analyses indicated that the epithelium in the ectopic *T*-positive mass is not part of the tail endoderm, since the observed gene expression patterns in mutant tails did not abide by the molecular characteristics of the caudal gut. In particular, it contained transcripts for *T* (*Brachyury*), *Sox2, Fgf8*, *Wnt3a* and *Nkx1.2* (Figure 1B,D,K-O) that were either not observed in the gut endoderm or confined to its posterior end in wild type embryos at this stage of development (Albors et al., 2016; Cambray and Wilson, 2007) (Figure 1A,C,E-I). Similarly, expression of the gut endodermal marker *Foxa2* was weak and spotty in this epithelium, which contrasted with the strong signal observed in the epithelium surrounding the ectopic tissue representing the bona fide gut endoderm (Figure 1J,P). The ectopic mass epithelium also had areas expressing both Sox2 and T proteins (Figure 1T). Surprisingly, Sox2 immunoreactivity was also observed in the gut endoderm of both wild type and *Gdf11* mutant embryos (Figure 1Q,R,T,U), which contrasts with the absence of *Sox2* transcripts in this structure (Figure 1C,F). Together, these patterns resemble some typical features of the NMP-containing epiblast (Cambray and Wilson, 2007), suggesting that the epithelium of the ectopic *T*-positive mass in *Gdf11* mutants might represent incomplete resolution of the epiblast during the trunk to tail transition.

From all the known factors involved in the regulation of axial progenitor activity (Wilson et al., 2009), *Oct4* seems to be unique in being required during trunk formation, yet dispensable at tail bud stages (DeVeale et al., 2013). Interestingly, in about half of the analyzed cases, premature activation of Gdf11/Alk5 signaling in axial progenitors at the beginning of the trunk-formation stage produced embryos with strong morphological resemblance to those where *Oct4* was inactivated at late primitive streak stages (Jurberg et al., 2013). These observations suggest that the epiblast-like structure in the *Gdf11* mutant tails might be somehow associated with persistent Oct4 activity. Analysis of *Oct4* expression was consistent with this hypothesis. While *Oct4* expression could not be detected in the tail bud of wild type embryos at E10.5 (DeVeale et al., 2013; Downs, 2008; Osorno et al., 2012) (Figure 1R,S), it was observed in the tail tip, the ventral posterior neural tube and the ectopic *T*-positive tissue mass of similarly staged *Gdf11* mutant embryos (Figure 1U,V). In this latter structure, we found Oct4 expression in both its epithelial and mesenchymal components, partially overlapping with T and Sox2 (Figures 1T,U). This observation further supports the hypothesis that the epithelium in the ectopic *T*-positive mass is likely to be an epiblast remnant and is consistent with the role of *Gdf11* in promoting the switch in axial extension from epiblast to tailbud-dependent growth (Jurberg et al., 2013).

Sustained *Oct4* activity in axial progenitors extends the trunk

As conditional inactivation of *Oct4* suggested a role for this gene in the maintenance of the primitive streak (DeVeale et al., 2013), we hypothesized that the longer trunks of *Gdf11* mutants could result from persistent Oct4 activity, which would keep axial growth in a trunk-generating configuration. To test this, we produced transgenic embryos expressing *Oct4* under the control of the *Cdx2* enhancer (Benahmed et al., 2008) (*Cdx2-Oct4* transgenics), thus overcoming the normal progressive down-regulation of *Oct4* expression in the epiblast (Osorno et al., 2012). From the ten *Cdx2-Oct4* transgenic fetuses recovered at E18.5 nine had abnormal phenotypes that could be divided in two groups. The first group (4 embryos) showed a variable increase in the number of thoracic and lumbar segments in their axial skeletons. The most affected fetus of this group had 17 instead of 13 rib-containing vertebrae and 8 instead of 6 lumbar segments (Figure 2A,C). These transgenic skeletons phenocopy to a large extent the main axial features observed in *Gdf11* mutant fetuses (McPherron et al., 1999) (Figure 2B), showing

that the longer trunks of *Gdf11* mutants could indeed result from an extended period of Oct4 activity.

The second group of *Cdx2-Oct4* transgenic embryos (5 embryos) was readily identified by the presence of a sacrococcigeal teratoma of variable sizes (Figure S2A-C). The axial skeleton in this group of transgenics was characterized by an abnormally large number of ribs (up to about 30 in the most strongly affected specimen), covering most of the body length and associated with the absence of recognizable sacral or caudal structures (Figure 2D). Histological analyses of this type of transgenic revealed that their neural tubes also extended further posteriorly than in wild type fetuses (Figure 2SE-H), thus fitting with the posterior extension of thoracic characteristics observed in *Cdx2-Oct4* transgenic fetuses. Remarkably, even in the most strongly affected fetuses, the neck and anterior thoracic segments seemed mostly normal (Figure 2D). These observations indicate that while *Oct4* overexpression had little or no effect in its normal domain of activity, it interfered dramatically with the development of areas formed after the switch into tailbud-dependent extension. Overall, these anatomical patterns are consistent with a partial or total conversion of the posterior body into a trunk.

In addition to teratomas and extended rib cages, this group of *Cdx2-Oct4* transgenics also had variable malformations in their hindlimbs. In all analyzed specimens, hindlimbs were found on the ventral side of the body with no connection with the axial skeleton (Figures 2D and S2A). Skeletal hindlimb morphologies varied from almost normal to strongly malformed, most notably in the distal structures, where some of the bones were either missing or reduced in size (Figures 2D and S2D).

At mid gestation stages (E10.5), *Cdx2-Oct4* transgenic embryos had malformations that, consistent with the skeletal phenotypes, were restricted to their posterior region. In these embryos, the distance between the fore and hindlimb buds was increased to different extents, in accordance with the abnormally large numbers of thoracic vertebrae observed at E18.5 (Figure 2G-I). Some of these embryos resembled *Gdf11* mutants (Figures 2F and S3B,E), further supporting the inverse functional link between Gdf11 and Oct4 activities. Another distinctive feature of *Cdx2-Oct4* transgenic embryos was the abnormal morphology of their posterior embryonic end, a trait most clearly observed at stages when their littermates had reached tail bud stages. Strongly affected embryos, such as those containing exceptionally long interlimb regions, lacked a recognizable tail bud and failed to close at the posterior end altogether, leaving a

dorsally exposed epithelium that resembled the epiblast of younger wild type embryos (Figures 2I and 3A,B,C,F,G,H). The epiblast-like nature of this epithelium was further evidenced by analysis of molecular markers such as *Fgf8, T(Bra)* and *Nkx1.2* that at E10.5 were expressed in the open epithelium in patterns resembling those typical of the epiblast of E8.5-E9.0 wild type embryos (Figures 3A,A',B,B',C,C',F,F',G,G',H,H'). These embryos also produced lateral mesoderm up to the very posterior end of the main body axis, as revealed by *Hand2* expression. This contrasts with the absence of this mesodermal compartment in the tail region of wild type embryos (Figure 3D,I).

Taken together, skeletal, morphological and expression phenotypes of *Cdx2-Oct4* transgenics indicate that persistent Oct4 activity in posterior epiblast was sufficient to keep embryonic extension of the main body axis in a trunk-generating mode. Also, Gdf11 signaling seems to be required to counteract Oct4 activity as part of its transition-triggering program into a "tail mode" of development.

Oct4 expression delays activation of posterior *Hox* genes

The remarkable increase in rib number observed in *Cdx2-Oct4* skeletons suggested delayed activation of posterior *Hox* genes (Carapuço et al., 2005; Mallo et al., 2010; Wellik and Capecchi, 2003). Expression analyses confirmed that *Hox* genes of the paralog group 10 were activated at more posterior levels than in wild type embryos, following the posterior shift in hindlimb location (Figures 3K-P and S3A,B). A closer look at these expression patterns indicated that activation of *Hox* group 10 genes was particularly delayed in the paraxial mesoderm of *Cdx2-Oct4* transgenics. Indeed, in these transgenic embryos, transcripts for *Hox* group 10 genes were barely detectable in the somites adjacent to the hindlimb, whereas in wild type embryos, the anterior expression limit of these genes in the somitic mesoderm roughly coincided with the anterior hindlimb border (Figure S3A,A',B,B'). In strongly affected *Cdx2-Oct4* embryos, activation of *Hox* group 10 genes was barely detectable in the paraxial mesoderm, albeit some expression was apparent in the emerging hindlimb buds and neural tube (Figure 3K-P). Expression of *Hox* genes belonging to more posterior paralog groups, like *Hoxd11* or *Hoxd13,* was also affected, following patterns similar to those observed for *Hox* group 10 genes (Figures 3E,J and S3D,E,G,H). Conversely, activation of more anterior *Hox* genes, like *Hoxb6*, was essentially normal (Figure S3I, J). In general, these expression patterns resembled the

delayed activation of posterior *Hox* genes observed in *Gdf11* mutants (Jurberg et al., 2013; McPherron et al., 1999; Szumska et al., 2008) and were complementary to the early activation of the same *Hox* genes in *Cdx2-Alk5*^{CA} embryos that presented a premature induction of the trunk to tail transition (Jurberg et al., 2013) (Figure S3C,F). These data further support an inverse functional connection between Gdf11 signaling and Oct4.

Oct4 expression is maintained for longer developmental times in snake embryos

Long rib cages, extended production of lateral mesoderm and delayed activation of posterior *Hox* genes, as observed in *Cdx2-Oct4* transgenic embryos, are also the very same traits found in snakes (Di-Poï et al., 2010; Woltering, 2012; Woltering et al., 2009), suggesting that their elongated trunks might result from sustained Oct4 activity for longer developmental times. To test this hypothesis, we investigated *Oct4* expression by *in situ* hybridization in corn snake embryos and found that *Oct4* continued to be transcribed in the posterior part of the embryonic trunk region at a stage when the tailbud starts becoming evident (Figure 4A,B), in sharp contrast with equivalently staged mouse embryos where *Oct4* expression had already disappeared completely (Osorno et al., 2012). These results were confirmed by RT-PCR, using python trunk cDNA as template (Figure 4H, I). This indicates that *Oct4* expression persists for longer developmental times in snake embryos than in mammalian embryos and that heterochronic changes in gene expression could thus be involved in production of the extended, organ-filled trunks characteristic of snakes.

Genomic organization of the mouse and snake *Oct4* loci

To understand the possible origin of the differences in snake and mammalian *Oct4* expression, we compared the chromosomal environment of *Oct4* in the genomes of a basal snake, python (Castoe et al., 2013), and the mouse (the mouse genome will be used as reference, but other mammals have similar configurations in this area) (Figure 4C, D). The chromosomal organization downstream of the *Oct4* locus was fairly similar in python and mouse, with the presence of the *Tcf19* and *Cchcr1* genes. However, the python and mammalian genomes seemed to have lost synteny upstream of the *Oct4* locus. The only existing similarity between these two genomes in this area was the presence of the *Lsm2* and *Vars* genes, located about 500 kb upstream of *Oct4* in the mouse and about 60 kb in the python. Between these genes and *Oct4,* the

mouse genome has a 220 kb gene rich region next to *Lsm2*, followed by an almost gene-free 280 kb region containing just a few scattered histocompatibility complex genes. In python, the region between the *Vars* and *Oct4 loci* contains only exons coding for a protein with high homology with *Npdc1*, some 20 kb upstream of the exons homolog to *Oct4*. The presence of *Npdc1* next to *Oct4* might be a general characteristic of snakes, as it is also present in the king cobra genome (Vonk et al., 2013). In mammals, *Npdc1* and *Oct4* are located in different chromosomes (e.g. mouse chromosomes 2 and 17 for *Npdc1* and *Oct4*, respectively). This indicates that the area 5' from the *Oct4 locus* underwent divergent reorganization after the divergence of snakes and mammals, which might have influenced *Oct4* regulation or even activity in these two taxa.

The annotated *Npdc1* region of python (Castoe et al., 2013) suggested that this gene had lost the last coding exon, which was also not identified in the annotated king cobra genome (Vonk et al., 2013). Actually, in the king cobra *Npdc1* annotation, an additional exon was missing but we could identify it upon comparison with the python genome (Figure S4A,B). The absence of the last *Npdc1* exon could have promoted an exon-shuffling event, bringing the *Oct4*-coding exons into the *Npdc1* transcript and creating a fusion between the two genes/proteins that could have affected *Oct4* regulation and/or activity. Such a chimeric product was suggested in the python annotation (Castoe et al., 2013). We directly assessed this possibility by searching for the different transcripts potentially produced from this genomic region. By RT-PCR, we detected mRNAs derived from exons with homology to *Npdc1* or to *Oct4* (Figure 4F,H,I). However, we were unable to amplify PCR products compatible with the existence of the hypothetical chimeric transcript containing *Npdc1* and *Oct4* exons (Figure 4G). Yet, the free donor and acceptor splice sites in *Npdc1* and *Oct4* exons suggested in the python annotation indicated either incomplete annotation or the existence of additional exons that might have compromised our search for the chimeric transcript. Comparison of the python and king cobra genomes revealed the presence of a ~0.8 kb-long highly conserved region about 3 kb downstream of the last annotated *Npdc1* exon (Figure 5B). This sequence contained a potential splice acceptor that, when spliced to the last *Npdc1* exon, would extend the open reading frame by 21 amino acids with high homology to the corresponding area of the mouse protein (Figure 5B,D). We confirmed that this region is a bona fide *Npdc1* exon by RT-PCR (Figure 5C). Therefore, the structure of the python and king cobra *Npdc1* resembles that of its mammalian homolog.

Annotation of the king cobra *Oct4* differed from that of python in two main aspects: no splice acceptor was identified at the 5' end of the first exon and the Oct4 protein was given a start site about 250 bp downstream of the splice acceptor annotated in python (Figure S4C). However, the king cobra sequence upstream of the mapped start codon showed strong homology with the corresponding python sequence both in the nucleotide sequence and predicted translation product (Figure S4C), suggesting that it is part of the *Oct4*-coding region. Consistent with this, comparison of the two sequences after filling the gaps in the published sequences revealed the presence of an in frame ATG that marks the start of the conservation between the two snake sequences (Figure S4C). The protein encoded from this start codon bares almost no homology with the mouse protein other than the first five amino acids (Figure 5E). This contrasts with the high conservation exhibited by parts of the protein encoded by further downstream exons. We confirmed by RT-PCR that this region is a genuine exon of the *Oct4* transcript (Figure 4H). This indicates that the differential rearrangements at the 5' end of the *Oct4* gene led to a different selection of the first exon(s). Considering that the enhancers controlling *Oct4* expression in the mouse are relatively close to its transcription start site (Yeom et al., 1996), the differences in the genomic region 5' of *Oct4* might have also impacted its regulation. Indeed, we were unable to detect any regions in the python genome with homology to the first *Oct4* exon of mammals or to mammalian *Oct4* regulatory regions (Figure S5A). Similarly, a BLAST search with the sequence of the first python *Oct4* exon failed to identify significant homology in mouse databases.

Organization of the *Oct4* locus in lizards

To further explore the role that the genomic organization at the *Oct4* locus might have played in *Oct4* expression during snake embryonic development, we first analyzed the extent of structural and sequence conservation around the *Oct4* locus between lizards and snakes. Analysis of the *Anolis* genome was inconclusive; although in the available annotated sequence there was no reference to *Npdc1* next to *Oct4*, the sequence contains a 30 kb gap where *Npdc1* might be positioned according to the snake genome. However, a VISTA analysis using available genomic sequence data from gecko and glass lizard revealed the presence of *Npdc1*-coding exons in the area 5' from *Oct4*, indicating that the rearrangement placing *Npdc1* upstream of *Oct4* occurred before the snake/lizard divergence (Figure 6). Interestingly, addition of other snake and lizard

species to the analysis showed substantial differences in this area between snakes and lizards. In particular, while homology among representatives of both squamate groups was mostly restricted to coding regions, the genome of all snake species displayed extensive homology in non-coding regions as well. A similar VISTA analysis performed taking gecko genomic sequence as the reference, revealed that homology among lizards in non-coding regions was not substantially higher than between gecko and python (Figure S5B). These results indicate that while the general gene structure around the *Oct4* locus is similar in snakes and lizards, non-coding regions within this genomic area were subject to differential evolutionary constraints in these two squamate lineages. Also, the extensive sequence homology in non-coding regions around *Oct4* in snakes suggests the existence of highly conserved regulatory information.

Gdf11 in squamates

Our data from mice indicate functional interactions between Oct4 and Gdf11 signaling during axial extension and, therefore, we explored whether modifications affecting Gdf11 could also have contributed to the snake body plan. Comparison of the *Gdf11*-containing genomic region in mice and snakes revealed that this region underwent differential rearrangement involving an inversion that affected the genomic context around *Gdf11* (Figure S6A). Despite this rearrangement, *Gdf11* expression in snake embryos kept features compatible with the existence of a balance between Oct4 and Gdf11 signaling activities similar to that observed in mice, as it was restricted to the embryonic area posterior to the *Oct4* expression domain (Figure S6B).

As with the *Oct4* genomic region, lizards also shared the global genomic structure with snakes at the *Gdf11* locus. Interestingly, a VISTA analysis revealed a high degree of similarity among squamates in this genomic area, including both coding and non-coding regions (Figure 6SC). This clearly contrasted with the strong differential conservation in non-coding regions observed between snakes and lizards at the *Oct4* locus, which highlights the possible importance of the conserved regions at the *Oct4* locus for the production of the snake-like pattern of *Oct4* expression.

Regulation of *Oct4* expression in snakes

In an attempt to understand the regulatory potential of the region upstream of the snake *Oct4* locus we tested several highly conserved regions for their ability to activate reporter expression

in transgenic mouse embryos. Only two of the tested regions (see methods) seemed to be active in our assay. Interestingly, both regions included the only non-coding fragment (~250 bp) showing significant homology between snakes and lizards according to the VISTA analysis. On its own, this element was very active as an enhancer in mouse embryos, inducing reporter expression in neural tissues and neural crest along most of the AP axis, including head and body structures (Figures 7A-C and S7). This pattern was consistently observed in five different transgenic embryos, only with with slightly different levels of intensity, indicating that it likely reflects the element's regulatory potential in the context of a mouse embryo. Clear expression in the paraxial mesoderm was also observed in the most posterior part of one embryo, but this expression was much reduced in the other embryos (Figures 7B,C and S7B,C). The lateral plate mesoderm was negative in all cases.

In snakes, this 250 bp fragment is part of a larger segment of very high sequence conservation (about 1.2 kb). This larger fragment was also active in transgenic mice but in a much reduced spatial domain than the 250 bp element alone (Figure 7D-F). It also activated reporter expression in the neural tube and neural crest, but this was restricted to specific areas of the hindbrain and to the spinal cord corresponding to the posterior part of the embryo. This pattern was consistent in four different embryos, indicating that it represents the regulatory potential of this element in mouse embryos. The more restricted activity of the 1.2kb fragment suggests that the full potential of the 250bp sequence might be affected by surrounding conserved snake sequences. We therefore explored the expression patterns produced by the 250 bp element when embedded in a more complete snake-like genomic context by generating transgenic embryos with a Bacterial Artificial Chromosome (BAC) containing a corn snake genomic region spanning over the *Oct4* and *Npdc1* loci. None of the six transgenic embryos analyzed by in situ hybridization gave a signal for the corn snake *Oct4* gene above what was observed in non-transgenic mouse embryos (Figure 7G,H). Together, these results indicate that the region around *Oct4* might contain discrete enhancer elements able to activate transcription in the progenitors of axial structures but that their activity might be under the control of additional regulatory elements.

DISCUSSION

Oct4 has been extensively studied in the context of its function as a major regulator of pluripotency (Shi and Jin, 2010). We show here that this gene is also a key regulator of trunk length during vertebrate development. It has been previously shown that Gdf11 activity promotes the transition from trunk forming to tail forming mechanisms (Jurberg et al., 2013; Liu, 2006; McPherron et al., 1999, 2009). These observations, together with the data presented in this paper, suggest that the balance between Gdf11 and Oct4 activities is a determining factor that regulates regionalization of the vertebrate body into trunk and tail domains. This is consistent with the observation that experimental changes in Oct4 or Gdf11 activities produced complementary shifts in the trunk to tail transition. The mechanistic details of this interaction are currently unknown. However, it has been shown that in embryonic stem cells Smad3 is directed to its targets in the genome through interactions with Oct4 (Mullen et al., 2011). It is therefore possible that Gdf11 signaling might generate a Smad or Smad-like product that counterbalances Oct4 activity instead of promoting a positive functional interaction. Whether or not this is the case remains to be determined.

One of the most interesting consequences of our findings is the possibility that changes in *Oct4* regulation might have played a central role in the evolution of the vertebrate body plan. Our results indicate that long trunks could be a consequence of different *Oct4* regulatory mechanisms operating in snake taxa that would keep its expression at such levels as to maintain axial growth in a trunk-forming mode for extended developmental periods. Genomic analyses suggest that this resulted from genomic rearrangements involving an extensive region upstream of the *Oct4* locus. These changes might have occurred in sequential steps. A first phase, which seemed to have occurred at an early stage of squamate evolution, involved major rearrangements resulting in the general gene structure observed in this particular chromosomal region. This rearrangement might have had significant impact on the basic *Oct4* regulatory landscape, to the extent that it seemed to have resulted in a different selection of the first *Oct4* exon in squamates and mammals. After this initial major event, the *Oct4* locus diverged considerably in snakes and lizards as estimated by sequence comparisons involving this genomic region.

The observation that snakes exhibit remarkable sequence conservation in the non-coding regions around Oct4, which is not shared with lizards, is particularly relevant. Considering that conserved non-coding regions are often part of regulatory processes, it is possible that this area contains elements associated with a common snake trait or function, which could include

regulation of *Oct4* expression during axial extension. The finding that one of these conserved regions was indeed able to activate transcription in transgenic mouse embryos in a pattern compatible with activity in a subset of axial progenitors provides some support to this idea. Interestingly, the most active sequence in the mouse transgenic assay was the only non-coding region upstream of *Oct4* that seemed to have significant homology between snakes and lizards. This could indicate that the element is part of an ancestral regulatory network resulting from the initial recombination event, whose activity was later differentially modulated by additional regulatory elements. Our observation that the activity of this element was substantially different when tested alone or embedded within a larger snake-derived genomic context is consistent with this hypothesis, at least with respect to the snake locus. It is actually somewhat surprising that given the strong intrinsic activity of the 250bp element, the BAC containing the whole genomic region seemed to be unable to activate the snake *Oct4* in mouse embryos. This could indicate that proper control of this and/or other relevant elements requires factors not present in the appropriate combination or in the right spatial-temporal pattern in mouse embryos. This possibility would fit with the extreme molecular adaptations that seem to have occurred in snakes (Castoe et al., 2013; Vonk et al., 2013). If this is the case, the mouse transgenic approach will be rather limited to understand *Oct4* regulation in snakes; a direct analysis of chromatin interactions and histone modification profiles within the *Oct4* locus obtained from snake embryos might be a more appropriate strategy to identify key elements of the snake *Oct4* regulation landscape.

If the 250bp element is involved in the snake-like type of *Oct4* expression, its presence in lizards opens additional questions. One possibility is that this element plays no role in *Oct4* expression in snakes, but is required for another function shared by snakes and lizards. An alternative possibility is that the genomic environment brought this element into a different function in lizards. One such function could be associated with the tail regeneration capacity characteristic of many lizard species. Regeneration processes typically involve reactivation of progenitors and/or dedifferentiation of somatic cells into proliferative cells (Foglia and Poss, 2016; Poss, 2010) and a factor with the characteristics of Oct4 might have been co-opted for those types of processes. If this would indeed be the case, the 250bp element might fall into the recently described category of regeneration enhancer elements (Kang et al., 2016).

In addition to sustained *Oct4* expression, our experiments also suggest that successful trunk extension in snakes must have required substantial changes in its tail-promoting

mechanisms. In particular, the phenotypes of *Cdx2-Oct4* transgenics show that areas of the main embryonic axis that naturally form trunk structures tolerate increased levels of *Oct4* and develop without major problems. However, as soon as the embryo starts laying down the caudal-most areas, tail-promoting factors seem to start taking hold of morphogenesis of these areas. As a consequence, although *Oct4* is still able to promote recognizable trunk structures, these become progressively more disorganized and cells eventually reach a developmental dead-end in large sacroccocigeal teratomas. The presence of such tumors has also been described in other experimental settings upon ectopic *Oct4* expression (Economou et al., 2015). These observations highlight the need of a tight control of *Oct4* expression and of its interactions with other patterning factors to guide the differentiation potential of progenitor cells towards proper physiological routes.

A variety of studies have shown a close association between activation of posterior *Hox* genes and the position of the trunk to tail transition (Carapuço et al., 2005; Di-Poï et al., 2010; Jurberg et al., 2013; McPherron et al., 1999; Szumska et al., 2008; Woltering et al., 2009). Given the role that these genes play in anterior-posterior patterning processes (Pearson et al., 2005), these observations suggested a possible role for *Hox* genes in the evolution of vertebrate trunk length. However, functional assays in the mouse failed to support this hypothesis, as loss and gain of function experiments involving *Hox* genes consistently failed to produce a significant change in the hindlimb position (a proxy for the position of the trunk to tail transition) (Carapuço et al., 2005; Jurberg et al., 2013; Vinagre et al., 2010; Wellik and Capecchi, 2003). Nonetheless, what those experiments clearly showed was that *Hox* genes are crucial in the specification of skeletal features typically associated with the different body sections (Mallo et al., 2010), indicating that both processes must be coordinated. The finding that *Oct4* suppresses posterior *Hox* gene activation when promoting trunk formation, whereas *Gdf11* stimulates their expression when inducing tail development, suggests that the Oct4/Gdf11 system is the primary determinant of global trunk or tail growth modes and that *Hox* genes are then used to transmit patterning information into the mesodermal and neural derivatives of axial progenitors. This, in turn, ensures that all appropriate neural and skeletal elements for the trunk or tail bud-derived regions of the body are properly formed, positioned and coordinated during axial extension. The close association of growth modes with subsequent skeletal patterning increases not only the system's

robustness, but also creates developmental modules that could be readily acted upon by selection during the course of evolution.

The finding that both Oct4 and Gdf11 signaling are able to modulate expression of posterior *Hox* genes is also interesting from a regulatory perspective. It will be important to understand how these two activities fit within known regulatory landscapes of *Hox* genes (Darbellay and Duboule, 2016) and whether their regulatory capacity is implemented by direct interaction with discrete elements within or around the *Hox* clusters or indirectly through the control of additional factors that will then modulate *Hox* cluster activity.

EXPERIMENTAL PROCEDURES

Mouse and snake embryos

Gdf11 mutant embryos were produced from intercrosses between *Gdf11*$^{+/-}$ mice (McPherron et al., 1999). The *Cdx2-Oct4* transgenic construct contained the mouse *Oct4* cDNA (Osorno et al., 2012) cloned between the *Cdx2* enhancer (Benahmed et al., 2008) and the SV40 polyadenylation signal. The *Cdx2-Alk5*^{CA} transgenic construct has been previously described (Jurberg et al., 2013). For reporter analyses, six different highly conserved regions 5' from the snake *Oct4* gene were amplified by PCR from python genomic DNA (the primers are the "Tra" series of those specified on Table S1) and cloned upstream of a cassette containing the adenovirus2 minimal late promoter, the *β-galactosidase* cDNA and the polyadenylation signal from SV40 (Jurberg et al., 2013). The BAC containing the *Oct4* locus was obtained from a corn snake genomic library (kindly provided by Isabel Guerreiro and Denis Duboule). A BAC clone that included the *Npdc1* and *Oct4* genes was linearized with lambda terminase and used to produce transgenic mouse embryos. Transgenics were produced by pronuclear injection (Hogan et al., 1994). Mouse embryos were recovered by cesarean section at different developmental stages and processed for whole mount *in situ* hybridization, immunofluorescence, β-galactosidase staining, histology or skeletal analysis. *Python regius* eggs were obtained from the python breeding colony at University of Florida. Eggs from *Pantherophis guttatus* (corn snake) were obtained from local breeders. Eggs were collected immediately after (corn snake) or slightly before (python) oviposition and embryos were dissected from the eggs and processed for *in situ* hybridization or RNA/DNA extraction. All experiments conducted on animals followed

the Portuguese (Portaria 1005/92) and European (Directive 2010/63/EU) legislations, concerning housing, husbandry, and welfare. The project was reviewed and approved by the Ethics Committee of ‘‘Instituto Gulbenkian de Ciência’’ and by the Portuguese National Entity, ‘‘Direcção Geral de Alimentação Veterinária’’ (license reference: 014308).

Phenotypic analyses

Skeletal analyses were performed at E18.5 by alcian blue/alizarin red staining as previously described (Mallo and Brändlin, 1997). For histological analyses specimens were fixed in Bouin's fixative, embedded in paraffin and 10 μm sections were stained with hematoxylin and eosin according to standard procedures. Expression analyses were performed by whole mount *in situ* hybridization (Kanzler et al., 1998) using *in vitro* translated digoxigenin-labelled antisense RNA probes. The probe for the corn snake *Oct4* gene consisted of the three last exons and the probe for the corn snake Gdf11 consisted of the last two coding exons. In both cases exons were individually amplified from genomic DNA (primers are specified in Table S1) and linked together by regular cloning strategies. Stained embryos were included in gelatin/albumin and cut at 35 μm with a vibratome. Immunofluorescence analyses were performed as previously described in (Osorno et al., 2012), either on whole mount embryos or cryostat sections using antibodies against Oct4 (sc-8628 Santa Cruz Biotechnology, 1:250), Sox2 (ab92494 Abcam, 1:250) and Brachyury (AF2085 R&D Systems, 1:250). Confocal Z-series stacks of immunostained material were acquired on a Leica SP5 confocal microscope. β-galactosidase activity was identified by X-gal staining as previously described (Jurberg et al. 2013).
For RT-PCR analyses, total RNA was extracted from python embryonic tissues using Trizol (Sigma-Aldrich). cDNAs were produced by reverse transcription using oligo dT priming and the presence of specific transcripts was investigated by PCR. The position of the primers is shown on Fig. 4 and their sequences listed in Table S1. The identity of the PCR products was confirmed by sequencing.

Genomic analyses

Python, king cobra and mouse genomic sequences were obtained from public databases. The mouse genome corresponded to assembly GRCm38/mm10. Python sequences were obtained from contig NW_006534040 and the king cobra genome from contig AZIM01002363. Gaps in

the published genomic sequence of the area around the 5' end of the first *Oct4* coding exon of python were filled sequencing of PCR-amplified fragments using primers Py-Oct4-gen-F and Py-Oct4-gen-R (Table S1). Sequence comparisons were performed using ClustalW2.

We computed long global alignments of the *Oct4* regions (100Kb) from multiple species and represented sequence similarity by curve-based visualization using the software mVISTA (Frazer et al., 2004) (http://genome.lbl.gov/vista/index.shtml). We used the default parameters from the VISTA browser plot to calculate conserved regions and to display VISTA graphs. The genomic regions of interest were extracted from NCBI genome assemblies using standalone Blast. The *Boa constrictor* genome assembly was downloaded from GigaDB repository (http://gigadb.org/site/index). We identified the scaffolds containing the genomic region where *Oct4* was located (Table S2) using tblastn with the anole and mouse *Oct4* (anole, XP_008120168; mouse, NP_038661) and *Tcf19* (anole, XP_008120167; mouse, AAH04617) as queries. Then we used blastdbcmd in order to parse 100-200kb from the syntenic regions containing *Oct4* and neighboring genes from each species (*Npdc1*, *Tcf19* and *Cchrc1*). Gene annotations from the corresponding scaffolds were extracted from NCBI and used for the VISTA analysis in order to identify coding and non-coding regions. Before using Python and King cobra annotations in VISTA, *Npdc1* and *Oct4* exon annotations were manually curated as described in the results section.

Genomic regions of interest and their corresponding gene annotation for *Anolis carolinensis* and *Pogona vitticeps* were downloaded using the equivalent tools available on their genome browsers. For *Pogona*: https://genomics.canberra.edu.au/gbrowse/gbrowse/pogona_pvi1.1/. For *Anolis*: http://www.ensembl.org/Anolis_carolinensis/Info/Index.

AUTHOR CONTRIBUTIONS

R.A. and M.M. designed experiments; R.A., A.D.J., F.L., A.N. and M.M. performed experiments; R.A., F.L., M.J.C and M.M. analyzed data; and M.M. wrote the paper with help from all other authors.

ACKNOWLEDGMENTS

We would like to thank Val Wilson for providing *Oct4* cDNA, Jose Belo, Denis Duboule, Andreas Kispert, Andy McMahon, Erik Olson and Cliff Tabin for *in situ* probes, Denis Duboule and Isabel Guerreiro for the corn snake BAC library and Monica Dias, Florence Janody, Ana Stankovic, Miguel Manzanares and members of the Mallo lab for useful comments during the course of this project. This work has been supported by grants PTDC/BEX-BID/0899/2014 (FCT, Portugal) and SCML-MC-60-2014 (from Santa Casa da Misericordia de Lisboa, Portugal) to MM and by Howard Hughes Medical Institute funding to MJC. RA is supported by a PhD fellowship (SFRH/BD/51876/2012, from FCT, Portugal) and FL is supported by a HHMI international graduate student research fellowship.

REFERENCES

Albors, A.R., Halley, P.A., Storey, K.G., and Storey, K.G. (2016). Fate mapping caudal lateral epiblast reveals continuous contribution to neural and mesodermal lineages and the origin of secondary neural tube. 1–22.

Benahmed, F., Gross, I., Gaunt, S.J., Beck, F., Jehan, F., Domon–Dell, C., Martin, E., Kedinger, M., Freund, J., and Duluc, I. (2008). Multiple Regulatory Regions Control the Complex Expression Pattern of the Mouse Cdx2 Homeobox Gene. Gastroenterology *135*, 1238–1247.e3.

Burke, A.C., Nelson, C.E., Morgan, B.A., and Tabin, C.J. (1995). Hox genes and the evolution of vertebrate axial morphology. Development *121*, 333–346.

Cambray, N., and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. Development *134*, 2829–2840.

Carapuço, M., Nóvoa, A., Bobola, N., and Mallo, M. (2005). Hox genes specify vertebral types in the presomitic mesoderm. Genes Dev. *19*, 2116–2121.

Castoe, T.A., de Koning, A.P.J., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero, R.P., Degner, J.F., Daza, J.M., Gu, W., et al. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. *110*, 20645–20650.

Cohn, M.J., and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in

snakes. Nature *399*, 474–479.

Darbellay, F., and Duboule, D. (2016). Chapter Sixteen - Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci. In Essays on Developmental Biology, Part A, P.M.W.B.T.-C.T. in D. Biology, ed. (Academic Press), pp. 299–314.

DeVeale, B., Brokhman, I., Mohseni, P., Babak, T., Yoon, C., Lin, A., Onishi, K., Tomilin, A., Pevny, L., Zandstra, P.W., et al. (2013). Oct4 Is Required □E7.5 for Proliferation in the Primitive Streak. PLoS Genet. *9*, e1003957.

Di-Poï, N., Montoya-Burgos, J.I., Miller, H., Pourquié, O., Milinkovitch, M.C., and Duboule, D. (2010). Changes in Hox genes' structure and function during the evolution of the squamate body plan. Nature *464*, 99–103.

Downs, K.M. (2008). Systematic localization of Oct-3/4 to the gastrulating mouse conceptus suggests manifold roles in mammalian development. Dev. Dyn. *237*, 464–475.

Economou, C., Tsakiridis, A., Wymeersch, F.J., Gordon-Keylock, S., Dewhurst, R.E., Fisher, D., Medvinsky, A., Smith, A.J., and Wilson, V. (2015). Intrinsic factors and the embryonic environment influence the formation of extragonadal teratomas during gestation. BMC Dev. Biol. *15*, 35.

Foglia, M.J., and Poss, K.D. (2016). Building and re-building the heart by cardiomyocyte proliferation. Development *143*, 729–740.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: Computational tools for comparative genomics. Nucleic Acids Res. *32*, 273–279.

Gomez, C., Ozbudak, E.M., Wunderlich, J., Baumann, D., Lewis, J., and Pourquie, O. (2008). Control of segment number in vertebrate embryos. Nature *454*, 335–339.

Head, J.J., and Polly, P.D. (2015). Evolution of the snake body form reveals homoplasy in amniote Hox gene function. Nature *520*, 86–89.

Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors

and the making of the spinal cord. Development *142*, 2864–2875.

Hogan, B., Beddington, R., Constantini, F., and Lacy, E. (1994). Manipulating the mouse embryo: a laboratory manual (Cold Spring Harbour Laboratory Press).

Jurberg, A.D., Aires, R., Varela-Lasheras, I., Nóvoa, A., and Mallo, M. (2013). Switching Axial Progenitors from Producing Trunk to Tail Tissues in Vertebrate Embryos. Dev. Cell *25*, 451–462.

Kang, J., Karra, R., Dickson, A.L., Nachtrab, G., and Goldman, J.A. (2016). Modulation of tissue repair by regeneration enhancer elements.

Kanzler, B., Kuschert, S.J., Liu, Y.H., and Mallo, M. (1998). Hoxa-2 restricts the chondrogenic domain and inhibits bone formation during development of the branchial area. Development *125*, 2587–2597.

Liu, J.-P. (2006). The function of growth/differentiation factor 11 (Gdf11) in rostrocaudal patterning of the developing spinal cord. Development *133*, 2865–2874.

Mallo, M., and Brändlin, I. (1997). Segmental identity can change independently in the hindbrain and rhombencephalic neural crest. Dev. Dyn. *210*, 146–156.

Mallo, M., Wellik, D.M., and Deschamps, J. (2010). Hox genes and regional patterning of the vertebrate body plan. Dev. Biol. *344*, 7–15.

McPherron, A.C., Lawler, A.M., and Lee, S.-J. (1999). Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. Nat. Genet. *22*, 260–264.

McPherron, A.C., Huynh, T. V, and Lee, S.-J. (2009). Redundancy of myostatin and growth/differentiation factor 11 function. BMC Dev. Biol. *9*, 24.

Mullen, A.C., Orlando, D.A., Newman, J.J., Lovén, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. (2011). Master Transcription Factors Determine Cell-Type-Specific Responses to TGF-β Signaling. Cell *147*, 565–576.

Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting,

P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. Development *139*, 2288–2298.

Pearson, J.C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. Nat. Rev. Genet. *6*, 893–904.

Poss, K.D. (2010). Advances in understanding tissue regenerative capacity and mechanisms in animals. Nat. Rev. Genet. *11*, 710–722.

Shi, G., and Jin, Y. (2010). Role of Oct4 in maintaining and regaining stem cell pluripotency. Stem Cell Res. Ther. *1*, 39.

Stern, C.D., Charité, J., Deschamps, J., Duboule, D., Durston, A.J., Kmita, M., Nicolas, J.-F., Palmeirim, I., Smith, J.C., and Wolpert, L. (2006). Head-tail patterning of the vertebrate embryo: one, two or many unresolved problems? Int. J. Dev. Biol. *50*, 3–15.

Szumska, D., Pieles, G., Essalmani, R., Bilski, M., Mesnard, D., Kaur, K., Franklyn, A., El Omari, K., Jefferis, J., Bentham, J., et al. (2008). VACTERL/caudal regression/Currarino syndrome-like malformations in mice with mutation in the proprotein convertase Pcsk5. Genes Dev. *22*, 1465–1477.

Tzouanacou, E., Wegener, A., Wymeersch, F.J., Wilson, V., and Nicolas, J.-F. (2009). Redefining the progression of lineage segregations during mammalian embryogenesis by clonal analysis. Dev. Cell *17*, 365–376.

Vinagre, T., Moncaut, N., Carapuço, M., Nóvoa, A., Bom, J., and Mallo, M. (2010). Evidence for a myotomal Hox/Myf cascade governing nonautonomous control of rib specification within global vertebral domains. Dev. Cell *18*, 655–661.

Vonk, F.J., Casewell, N.R., Henkel, C. V, Heimberg, A.M., Jansen, H.J., McCleary, R.J.R., Kerkkamp, H.M.E., Vos, R.A., Guerreiro, I., Calvete, J.J., et al. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. U. S. A. *110*, 20651–20656.

Wellik, D.M., and Capecchi, M.R. (2003). Hox10 and Hox11 genes are required to globally

pattern the mammalian skeleton. Science *301*, 363–367.

Wilson, V., Olivera-Martinez, I., Storey, K.G., Olivera-Martínez, I., Storey, K.G., Olivera-Martinez, I., and Storey, K.G. (2009). Stem cells, signals and vertebrate body axis extension. Development *136*, 2133–2133.

Woltering, J.M. (2012). From Lizard to Snake; Behind the Evolution of an Extreme Body Plan. Curr. Genomics *13*, 289–299.

Woltering, J.M., Vonk, F.J., Müller, H., Bardine, N., Tuduce, I.L., de Bakker, M. a. G., Knöchel, W., Sirbu, I.O., Durston, A.J., and Richardson, M.K. (2009). Axial patterning in snakes and caecilians: Evidence for an alternative interpretation of the Hox code. Dev. Biol. *332*, 82–89.

Wymeersch, F.J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F.C., and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. Elife *5*.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development *122*, 881–894.

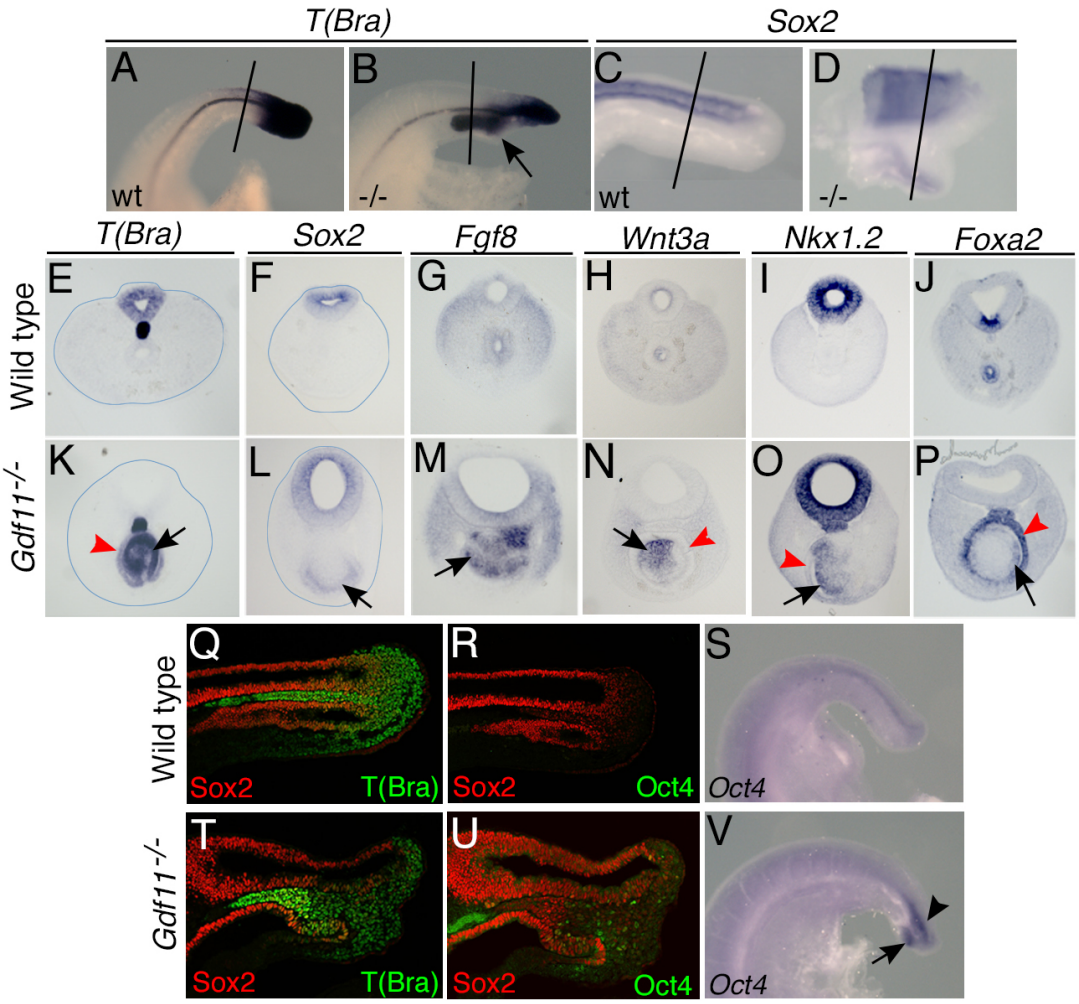

Figure 1. Characterization of the ventral ectopic mass in *Gdf11* mutant embryos. The tail region of E10.5 wild type (A, C, E-J, Q-S) or of *Gdf11*$^{-/-}$ (B, D, K-P, T-V) embryos was analyzed by whole mount *in situ* hybridization against *T(Brachyury)* (A, B, E, K), *Sox2* (C, D, F, L), *Fgf8* (G, M), *Wnt3a* (H, N), *Nkx1.2* (I, O), *Foxa2* (J, P) or *Oct4* (S, V) or by immunofluorescence with antibodies against Sox2 (Q, R, T, U), T(Brachyury) (Q, T) or Oct4 (R, U). E-P show transverse sections through the areas indicated in A-D or in supplementary Figure S1C-J, with the dorsal part oriented to the top. In E, F, K and L the outline of the embryo is indicated. Q, R, T, U show sagittal sections through the tail region. The red arrowheads in K-P indicate the gut epithelium and the arrows the ectopic mass-associated epithelium. The arrow in V indicates expression in the ectopic mass and the arrowhead shows expression in the posterior neural tube.

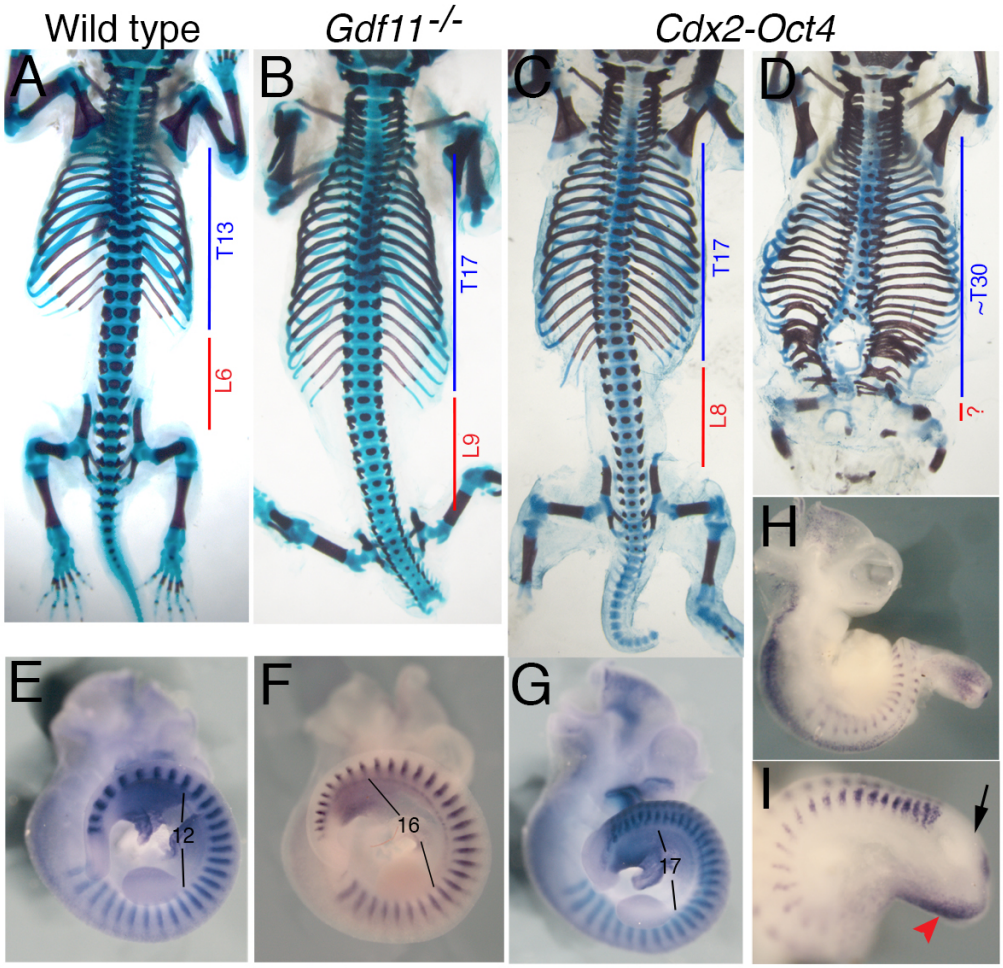

Figure 2. Consequences of sustained *Oct4* expression in mouse embryos. A-D. Skeletal analysis of a wild type (A), a *Gdf11*$^{-/-}$ (B) and two *Cdx2-Oct4* transgenic (C, D) embryos at E18.5. The number of thoracic (T) and lumbar (L) vertebrae is shown. E-I. Analysis of a wild type (E), *Gdf11*$^{-/-}$ (F) and two *Cdx2-Oct4* transgenic (G-I) embryos at E10.5 stained simultaneously for *Uncx4.1* and *Ptx1* (E-G) or with *Uncx4.1* and *Tbx4* (H, I). The number of interlimb somites is indicated. I shows a close up of the caudal end of the embryo in H. The embryo in G is expected to produce a skeleton similar to the one shown in C and the embryo in H and I is expected to produce a skeleton similar to that in D. The arrow in I indicates the posterior embryonic end showing the absence of a tailbud and the red arrowhead the position of the hindlimb/ventral mesoderm.

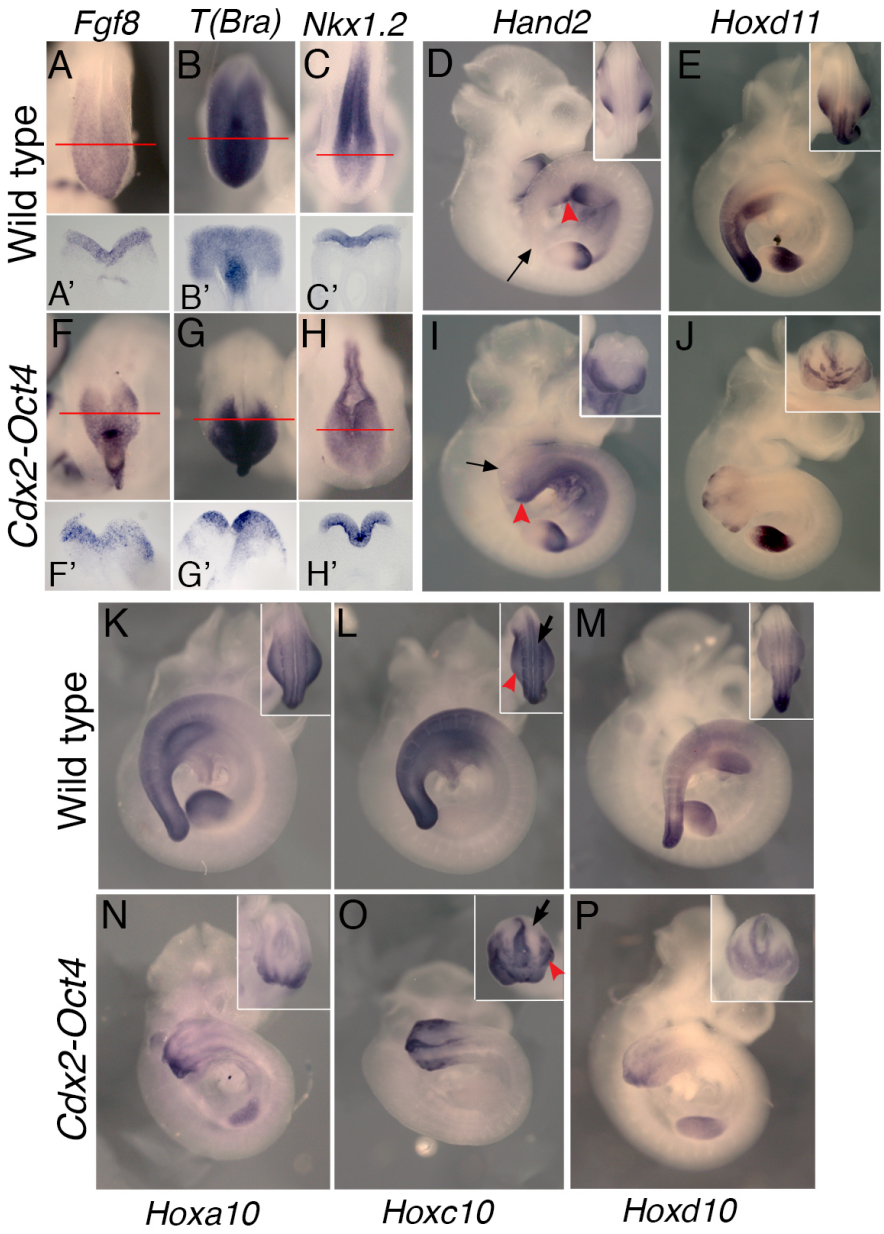

Figure 3. Molecular characterization of *Cdx2-Oct4* transgenics. E10.5 wild type (A-E and K-M) and *Cdx2-Oct4* transgenic (F-J and N-P) embryos were analyzed by *in situ* hybridization with probes for *Fgf8* (A, F), *T(Brachyury)* (B, G), *Hand2* (D, I), *Hoxd11* (E, J), *Hoxa10* (K, N), *Hoxc10* (L, O) and *Hoxd10* (M, P). A, B, C, F, G and H show dorsal close ups of the posterior end of the embryos. A', B', C', F'G' and H' show transverse sections through the areas indicated in the corresponding whole mount stained embryo. The black arrow in D, I indicates the posterior end of the embryos, whereas the posterior end of the lateral mesoderm is indicated with a red arrowhead. In all images involving *Hox* genes the inset shows a dorsal view of the posterior embryonic end to show that the paraxial mesoderm of the transgenics is mostly negative for posterior *Hox* gene expression. The arrows in L and O indicate the paraxial mesoderm and the red arrowhead the hindlimb.

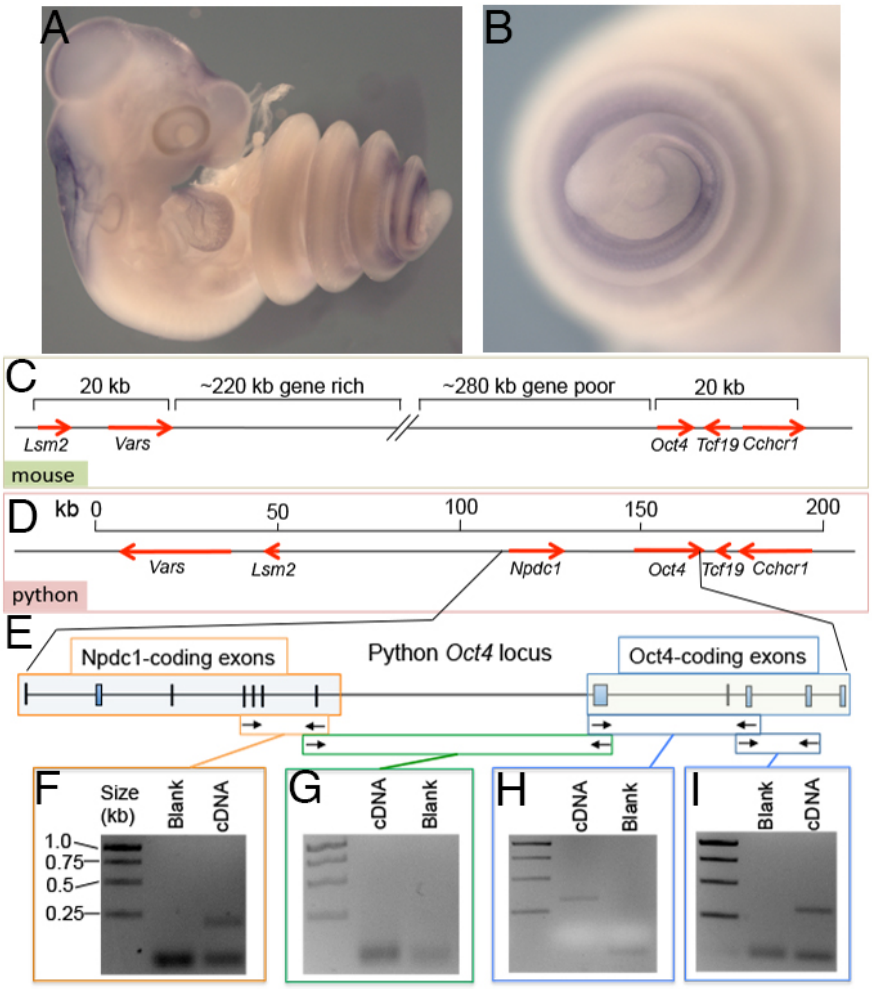

Figure 4. *Oct4* expression and genomic environment in snake embryos. A and B. *Oct4* expression in a corn snake embryo shortly after it underwent trunk to tail transition. B. shows a close up of the lower trunk/tail region. C. Structure of the genomic area surrounding the *Oct4* locus in mouse. D. Structure of the genomic area surrounding *Oct4* locus in python. In C and D the coding regions and their transcriptional orientation are indicated with red arrows. E. Close up of the region containing *Npdc1* and *Oct4*-coding exons in the python. Exons are represented as boxes. F-I. RT-PCR analysis of transcripts synthesized from the python *Npdc1* and *Oct4* transcription units using primer sets Py-Npdc-RT-F1 and Py-Npdc-RT-R1 (F), Py-Npdc-RT-F2 and Py-Oct4-RT-R1 (G), Py-Oct4-RT-F1 and Py-Oct4-RT-R2 (H) and Py-Oct4-RT-F2 and Py-Oct4-RT-R3 (I). The position of the primers is indicated in E.

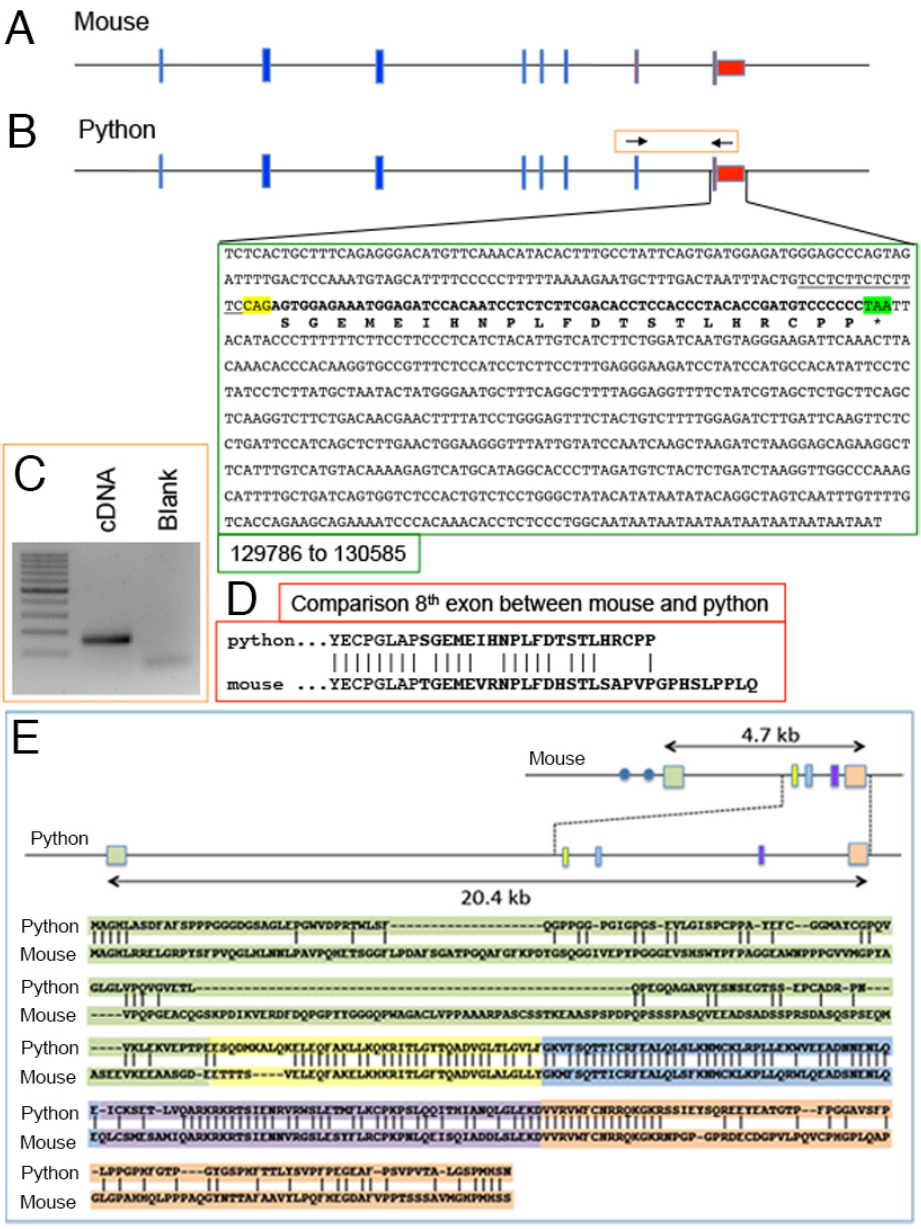

Figure 5. Comparison of last *Npdc1* exons and first *Oct4* exons between python and mouse. A. Schematic structure of the mouse *Npdc1* gene. B. Eighth python *Npdc1* exon. The sequence includes the end of the open reading frame (bold) and the 3' UTR. The king cobra has a similar sequence but is not shown in this figure. The splice acceptor site (in yellow), the termination codon (in green) and translation products are indicated, together with the coordinates of the exon in contig NW_006534040. C. RT-PCR analysis showing the existence of the eighth *Npdc1* exon in python using oligonucleotides form the regions indicated in the orange box in B. D. Comparison of the protein fragments encoded by the last exons of the mouse and python *Npdc1* genes. E. Comparison of the mouse and python Oct4 proteins. Regions encoded by the different exons are color-coded. Note the striking difference in exon1-encoded peptides, which contrasts with the high similarity in the rest of the molecules.

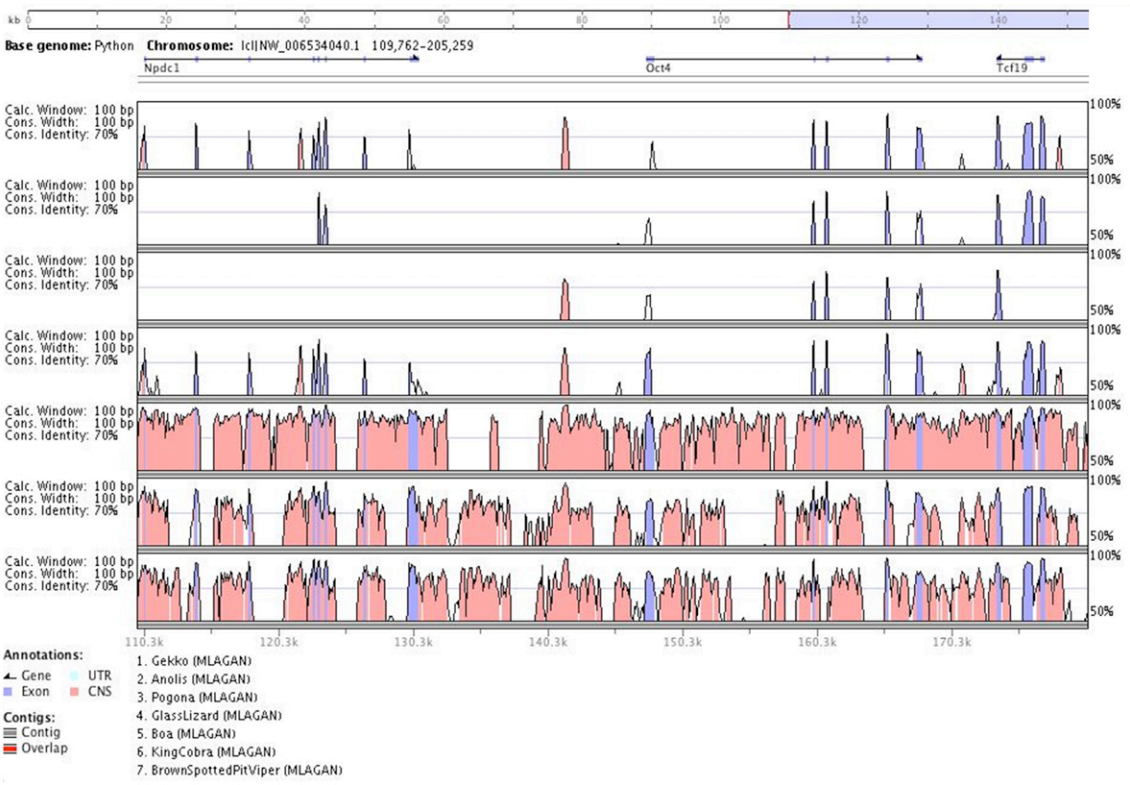

Figure 6. Sequence comparison in the vicinity of the *Oct4* locus in snakes and lizards. The genomic sequences from the area including *Oct4* from gecko, *Anolis*, *Pogona*, glasslizard, boa, king cobra and brown spotted pit viper were plotted against the corresponding area of python using a VISTA software. Represented in blue are homologies within coding exons and in red conservation in non-coding regions.

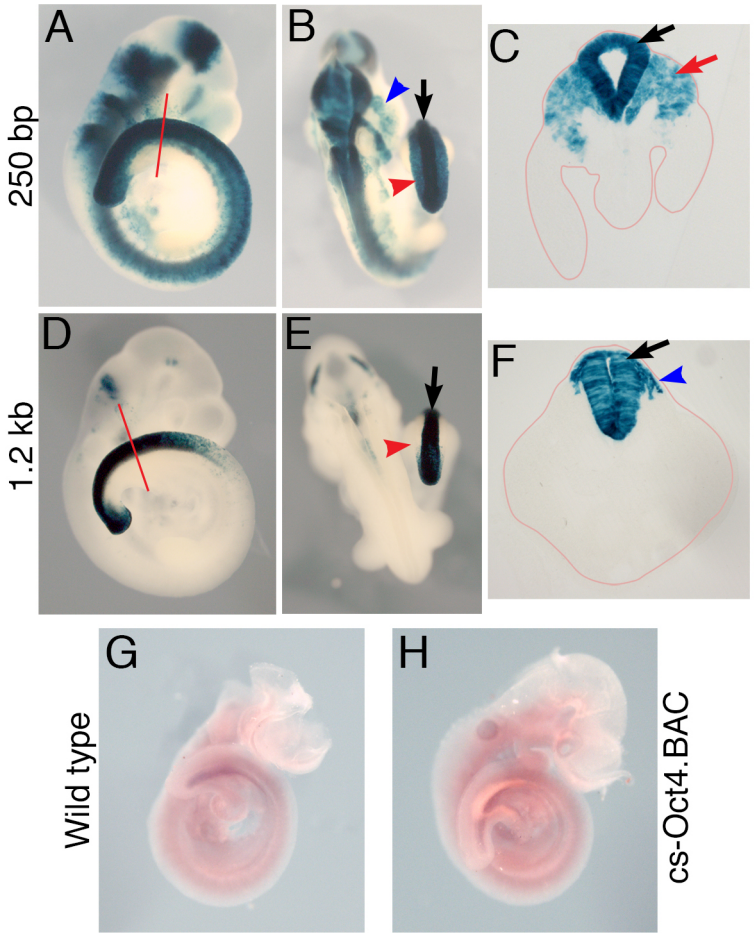

Figure 7. Regulatory capacity within the snake *Oct4* locus. A-F. β-galactosidase reporter assay to explore the capacity of the conserved 250 bp region (A-C) or the larger 1.2 kb fragment (D-F) from the genomic region upstream of the snake *Oct4* gene. The arrows indicate the neural tube, the blue arrowhead the neural crest and the red arrowhead the presomitic mesoderm. C and F show sections though the areas indicated in A and D, respectively. The black arrow indicate the neural tube and the red arrow the somitic mesoderm. The outline of the tissue is contoured in red. G,H. Whole mount *in situ* hybridization with a probe against the corn-snake *Oct4* mRNA, on wild type (G) or transgenic embryos for a BAC containing the *Oct4* locus (H).

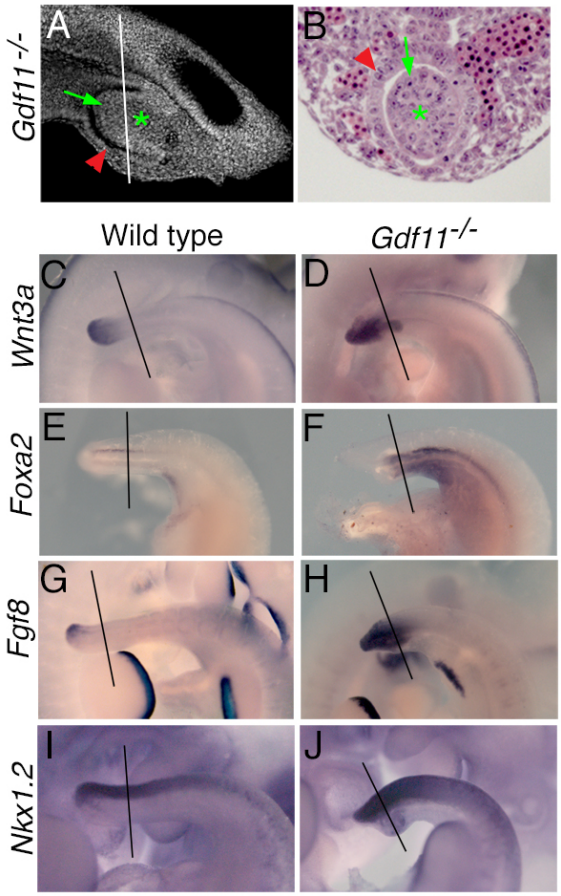

Figure S1 (related to Figure 1)**.** Ectopic ventral mass in *Gdf11* mutant tails. A. Sagittal section through the tail of an E10.5 *Gdf11*$^{-/-}$ embryo stained with DAPI. B. Transverse section through the area indicated in A stained with hematoxylin and eosin. The red arrowheads indicate the epithelium of the tail-associated gut. The epithelial (arrow) and the mesenchymal (asterisk) compartments of the ectopic mass are indicated in green. C-J. Close ups of E10.5 wild type (C, E, G, I) and *Gdf11*$^{-/-}$ (D, F, H, J) embryo tails stained by whole mount *in situ* hybridization with probes for *Wnt3a* (C, D), *Foxa2* (E, F), *Fgf8* (G, H) and *Nkx1.2* (I, J). The line indicates the position of the sections shown on Figure 1.

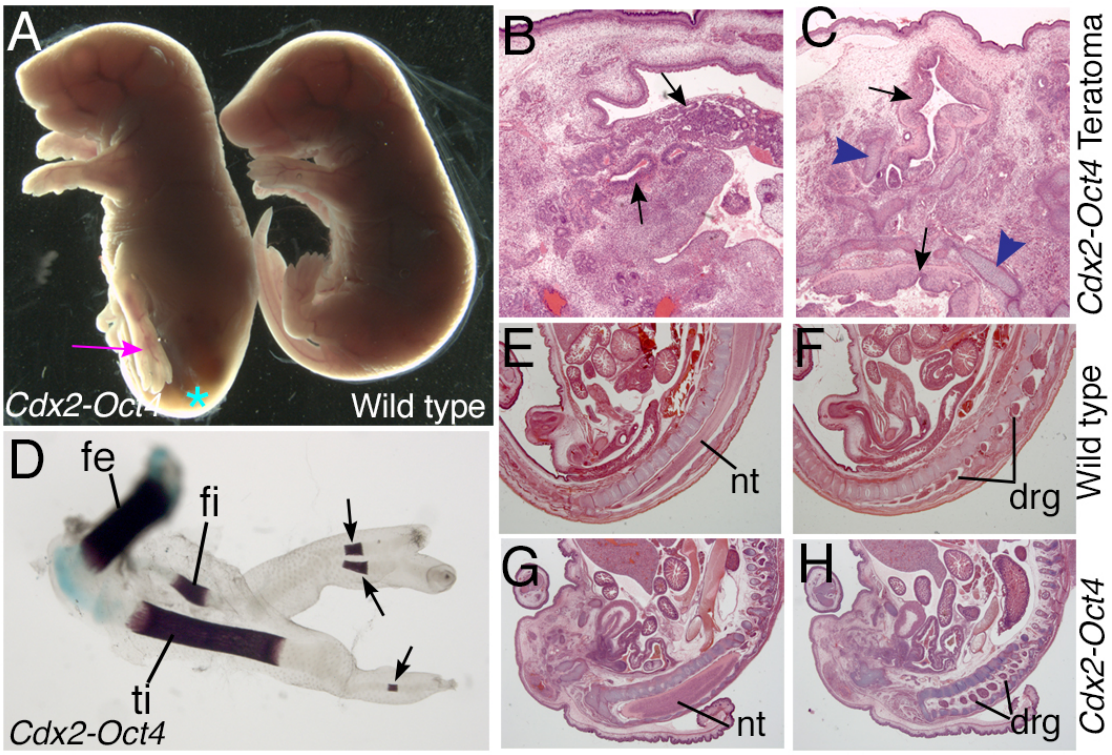

Figure S2 (related to Figure 2)**.** Additional phenotypes of *Cdx2-Oct4* transgenics. A. External morphology of a strongly affected transgenic fetus at E18.5 (left). A wild type littermate is shown for comparison. The asterisk indicates the position of the sacrococcigeal teratocarcinoma and the pink arrow indicates a hindlimb. B,C. Two histological sections of a teratocarcinoma found in a *Cdx2-Oct4* transgenic embryo, stained with hematoxylin and eosin. In B the arrows show glandular tissue. In C the arrows indicate neural tissues and the blue arrowheads cartilages. D. A hindlimb of a strongly affected *Cdx2-Oct4* transgenic fetus. The femur (fe) is fairly well formed. The tibia (ti) and fibula (fi) are more affected. Also, there are just a couple of digit-like structures (arrows) connecting directly to the tibia and fibula. E-H. Sagittal sections through the caudal part of a wild type (E,F) or a *Cdx2-Oct4* transgenic fetus (G,H). In the wild type embryo, the neural tube (nt) finishes around vertebra 34, whereas in the transgenic the neural tube fills the vertebral canal until the very posterior end of the axial skeleton (in this embryo around vertebra 38). In both wild type and *Cdx2-Oct4* transgenics the neural tube is associated with dorsal root ganglia (drg).

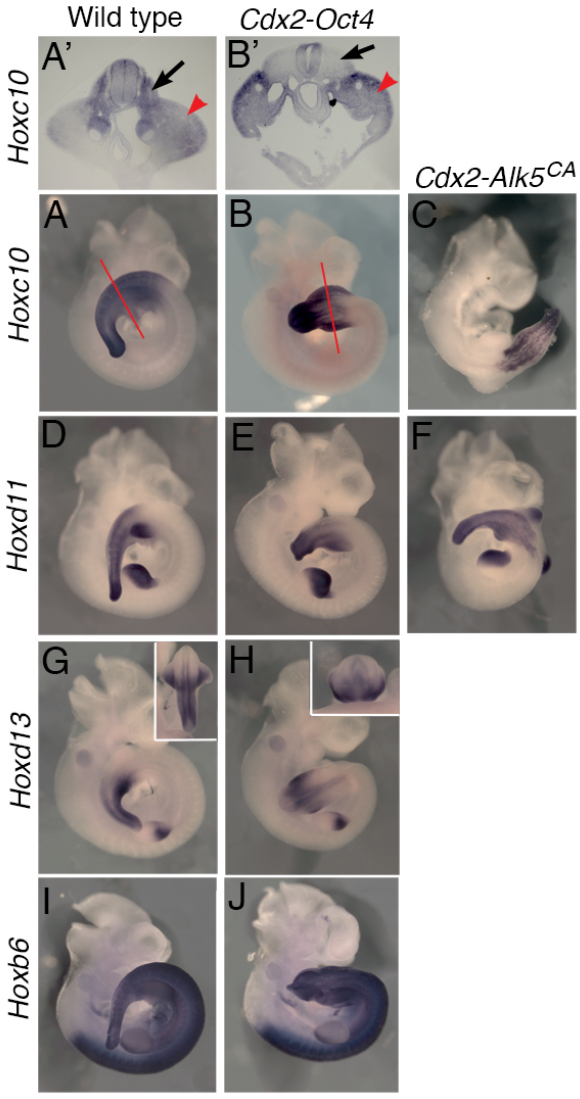

Figure S3 (related to Figure 3)**.** Hox gene expression in transgenic embryos. Expression of *Hoxc10* (A', B',A-C),*Hoxd11* (D-F), *Hoxd13* (G,H) and *Hoxb6* (I, J) was analyzed in E10.5 wild type (A, A', D, G, I), *Cdx2-Oct4* (B, B', E, H, J) or *Cdx2-Alk5*^{CA} (C, F) embryos by whole mount in situ hybridization. A-F show a comparison of two posterior Hox genes, *Hoxc10* and *Hoxd11*, in *Cdx2-Oct4* embryos with a milder phenotype (more similar to those of *Gdf11* mutant embryos, with a posterior displacement of the hindlimb position by about 4 somites) and in *Cdx2-Alk5*^{CA} embryos, in which these genes are activated more anteriorly following the premature trunk to tail transition. A' and B' show sections through the indicated region in A or B to indicate the presence of *Hoxa10* expression in somites (arrows) adjacent to the hindlimb (red arrowheads) in wild type embryos and its absence in *Cdx2-Oct4* transgenics. Expression of *Hoxd13* was shifted posteriorly following the position of the hindlimbs. In the paraxial mesoderm, expression was almost undetectable. The anterior expression border of *Hoxb6* expression is not altered in *Cdx2-Oct4* transgenic embryos.

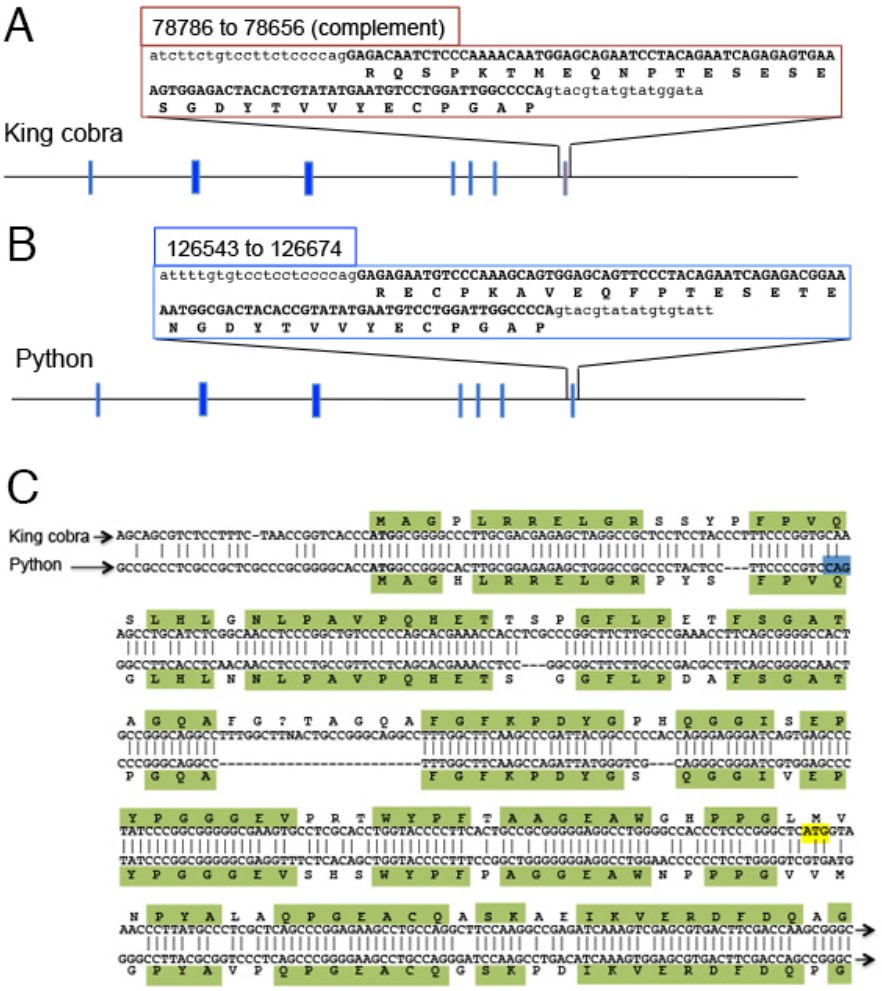

Figure S4 (related to Figure 5)**.** Comparison of king cobra and python Npdc1 and Oct4 sequences. A. Sequence of the seventh *Npdc1* exon from python (bold). The coordinates of the fragment in contig NW_006534040 and the sequence of the protein fragment encoded by this exon are also shown. B. Identification of the "missing" seventh *Npdc1* exon from king cobra through alignment with the python genome (coordinates in contig AZIM01002363). C. Identification of the N-terminal end of snake Oct4 proteins. This area of the python genome contained sequence gaps that were curated by amplification and sequencing. The resulting sequence provided an extension of the open reading frame with high homology with the corresponding area of the king cobra. For the alignment, a small gap (suggested 4 nucleotide-long) in the published king cobra sequence was disregarded to avoid a frame change. Nucleotide sequence similarity was lost upstream of the ATG. Indicated are the starting ATG of king cobra as described in the genome annotation (yellow). Also indicated is the position of the acceptor site (blue) according to the python annotation.

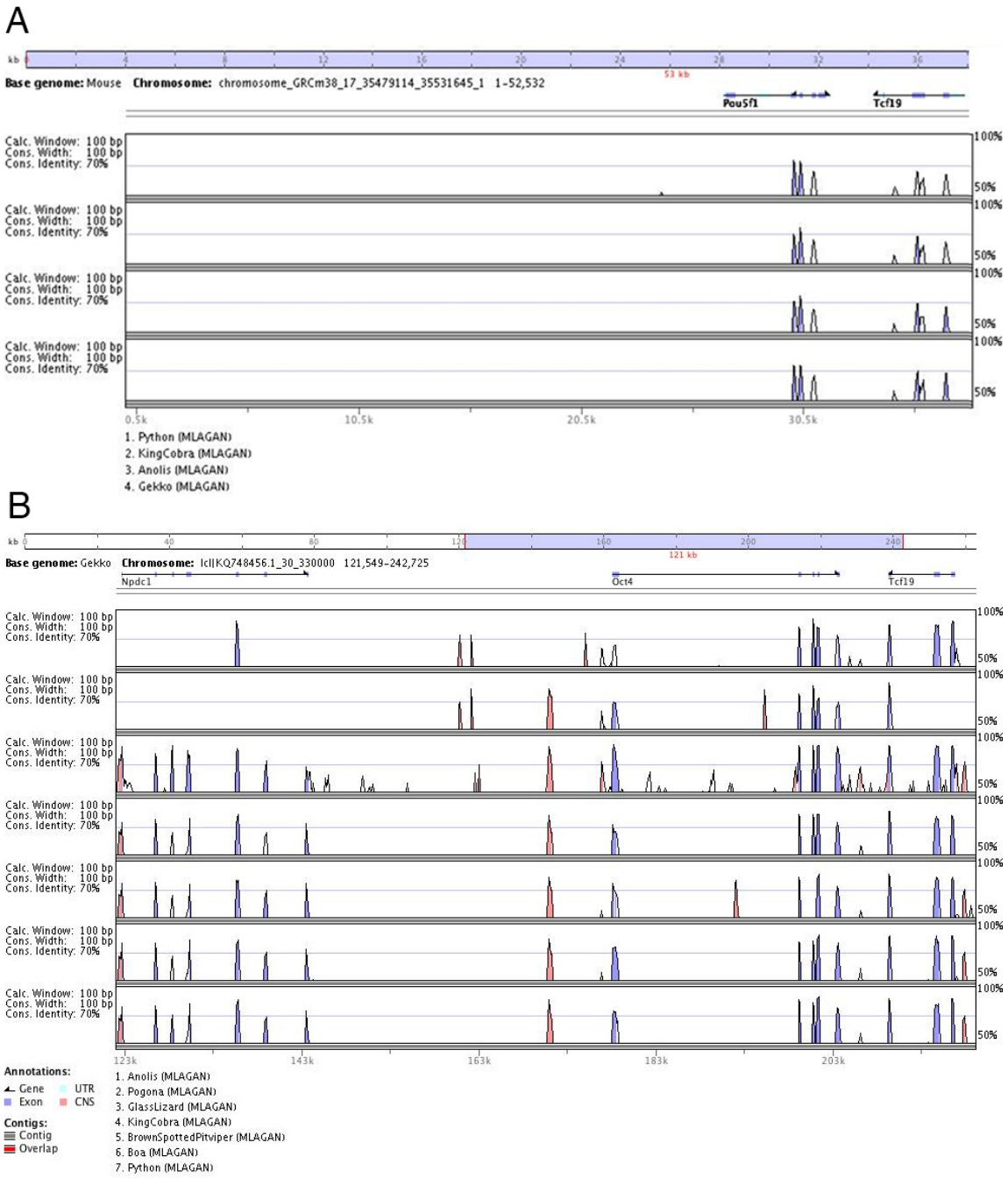

Figure S5 (related to Figure 6). Comparison of *Oct4*-containing genomic regions of different species. A. The mouse genomic region was compared with those of python, king cobra, Anolis and Gekko using VISTA. Homologies seem to be reduced to the coding exons, with the exception of the first *Oct4* exon. B. Comparison of the Gekko sequence with those of other lizards (Anolis, pogona, glasslizard) and snakes (king cobra, brown spotted pitviper, boa and python). Homology outside the coding regions was very reduced.

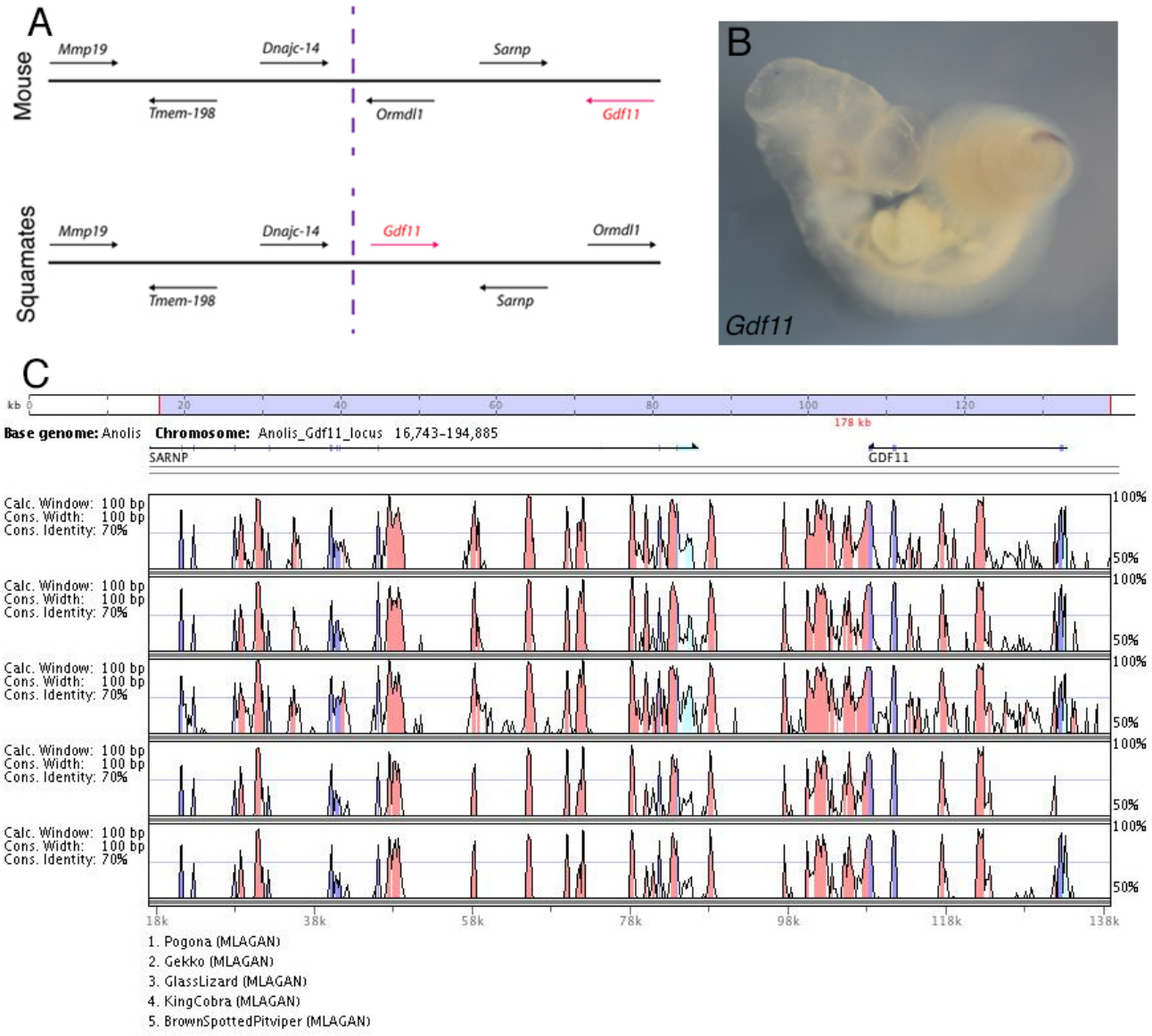

Figure S6 (related to Figure 6). *Gdf11* in squamates. A. Gene structure around the *Gdf11* locus in mammals and squamates, showing the inverted orientation of the region containing the *Gdf11*, *Sarnp* and *Ormdl1* genes. B. Whole mount in situ hybridization with a probe against *Gdf11* in corn snake embryos, showing expression in the tailbud. C. Sequence comparison in the vicinity of the *Gdf11* locus in snakes and lizards. Compared are sequences from *Anolis*, *Pogona*, gecko, glasslizard, king cobra and brown spotted pit viper using the VISTA software. Represented in blue are homologies within coding exons and in red conservation in non-coding regions.

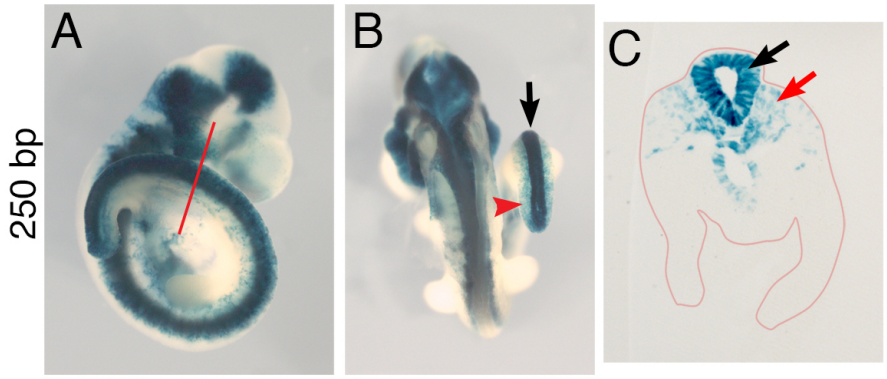

Figure S7. (related to Figure 7). Enhancer activity of the 250 bp homologous fragment. Lateral (A) and dorsal (B) views of a transgenic embryo stained for b-galactosidase activity. The arrow in B indicates the neural tube and the red arrowhead the presomitic mesoderm. C shows a section through the area indicated in A. Staining is strong in the neural tube (black arrow) but restricted to a few cells in the somitic mesoderm (red arrow). The contour of the tissue is outlined with the red line.

Table S1. Primers used in this work.

CS-Oct4-Ex3-F	CTGTCGACGGAAGATGTTTAGTCAAACC
CS-Oct4-Ex3-R	CTGAATTCTCCTGCAAATTCTCATTACT
CS-Oct4-Ex4-F	CTGAATTCAACTGTGCTCCATGGAGTC
CS-Oct4-Ex4-R	GTACTAGTATCCTTCTCCAGGCTGAGGTC
CS-Oct4-Ex5-F	GTACTAGTGTCTGGTTTTGCAACCG
CS-Oct4-Ex5-R	CTGCGGCCGCAGCTTGAATGCATCGGGTG
CS-Gdf11-Ex2-F	CAGTCGACGGACCCAGTGGTACAGATCG
CS-Gdf11-Ex2-R	CATCTAGACAGGCCTTCAGCACCAGGCCC
CS-Gdf11-Ex3-F	CATCTAGACACCCCTTCATGGAGCTGCGTG
CS-Gdf11-Ex3-R	GTGCGGCCGCTTAAGAGCATCCACACCTGTC
Py-Npdc-RT-F1	GATGGCATACACAGCTACTCG
Py-Npdc-RT-R1	ACACCGTATATGAATGTCCTG
Py-Npdc-RT-F2	CATTCATATACGGTGTAGTCG
Py-Npdc-RT-R2	CAGAAGATGACAATGTAGATGAG
Py-Oct4-RT-F1	GCCTGACATCAAAGTGGAGCGTG
Py-Oct4-RT-F2	GGCTTCACTCAGGCAGATGTG
Py-Oct4-RT-R1	CACGCTCCACTTTGATGTCAGGC
Py-Oct4-RT-R2	TGTTCCAACTCCACTGAGGTG
Py-Oct4-RT-R3	GGCCGACTCCATGGAGCACAG
Py-Oct4-gen-F	GACTCGAGCAGCACGAGCCTTCCGAGAGG
Py-Oct4-gen-R	CTGGATCCGGAAAGGGGTACCAGCTGTGAG
Tra-250-F	GTTTACAACTACAGTTGGATC
Tra-250-R	CTCCCTCCAGGACAATGTATC
Tra-1.2-F	TTGTGCAAATATGATTTGGATC
Tra-1.2-R	TTTCGCTTGGGGAAGGCAGAG
Tra-I-F	ATTATAGTTGCCTGGATGTAAGCC
Tra-I-R	ATACCTGCCTGTTGTGCAGCAGTG
Tra-II-F	TTTCAACTGCTACTTAAATTCTCCAGAG
Tra-II-R	CTGCTTGTGAAGGAAAGAATTAACTC
Tra-III-F	TTAACCTACTTGTTATTTTAATAACATTCCAC
Tra-III-R	ATCCTTTAAGAGACTACAATTGATTCAGAGTC
Tra-IV-F	AGCCTTTTTAGCATTGGAGCAG
Tra-IV-R	CTTTTCCCCATATGGCTTAGT

Table S2. Genome assemblies and scaffold accession numbers from the lizards and snake species analyzed in this study. Accession numbers from NCBI (http://www.ncbi.nlm.nih.gov/genome/) for all species but *Boa* and *Pogona*, which can be accessed at GigaDB (http://gigadb.org/site/index).

Species	Genome accession number	Genome assembly	Scaffold accession number
Gekko japonicus	GCA_001447785.1	Gekko_japonicus_V1.1	KQ748456.1
Python molurus	GCA_000186305.2	Python_molurus_bivittatus-5.0.2	NW_006534040.1
Ophiophagus hannah	GCA_000516915.1	OphHan1.0	AZIM01002363.1
Boa constrictor	ERS218597	snake_1C_scaffolds	SNAKE00002857
Protobothrops mucrosquamatus	GCA_001527695.2	P.Mucros_1.0	NW_015389406.1
Anolis carolinensis	GCA_000090745.2	AnoCar2.0	NW_003339170.1
Pogona vitticeps	ERA280782	pvi1.1.Jan.2013	scf001525