Cell Reports, Volume *32*

Supplemental Information

A Conserved Notochord Enhancer Controls Pancreas Development in Vertebrates

João Pedro Amorim, Ana Gali-Macedo, Hugo Marcelino, Renata Bordeira-Carriço, Silvia Naranjo, Solangel Rivero-Gil, Joana Teixeira, Mafalda Galhardo, Joana Marques, and José Bessa

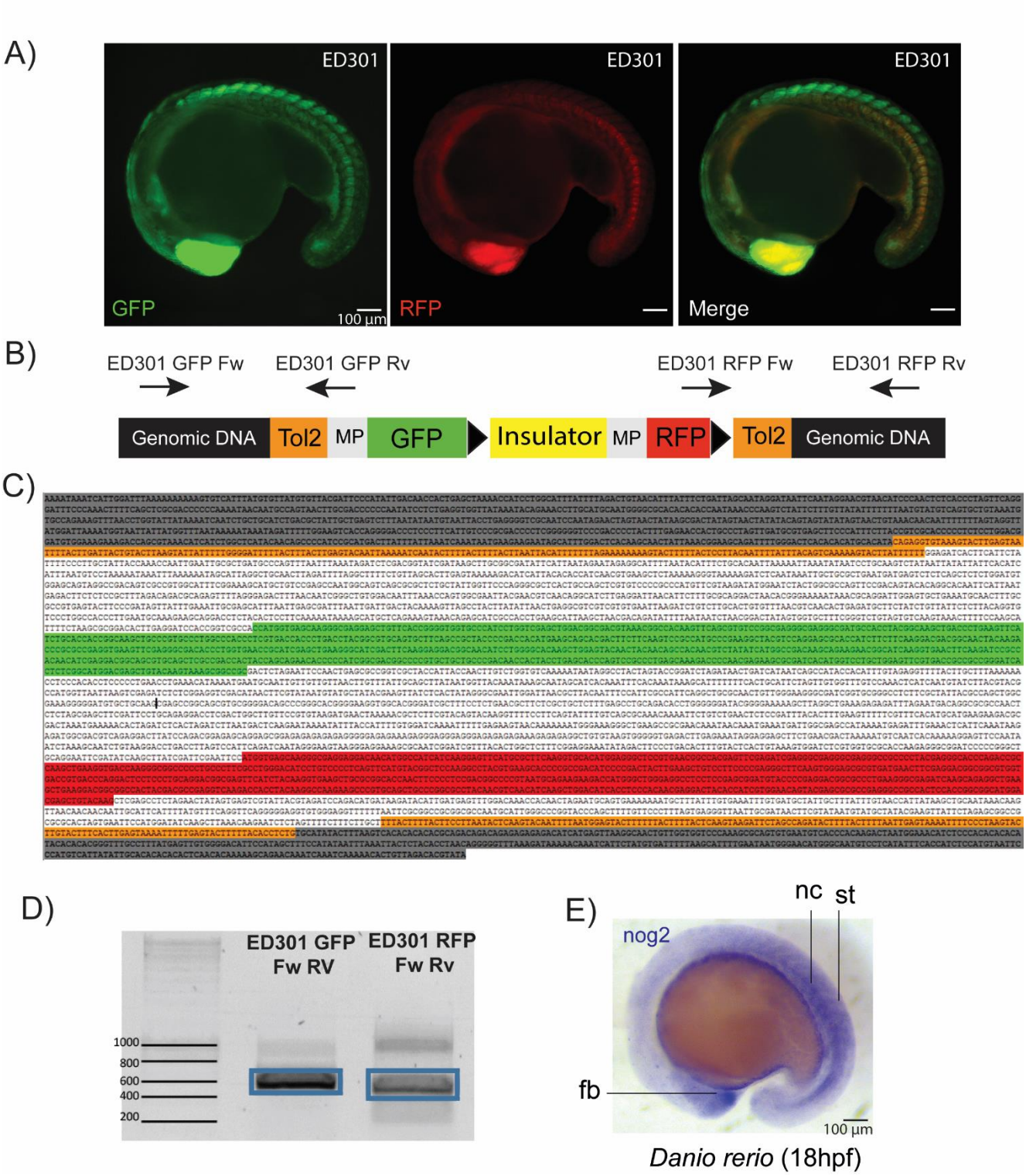


Figure S1. Related to Figures 1 and 2

A) Representative images of the ED301 line expression pattern at 18hpf. GFP (green) expression is seen in the forebrain, hindbrain, somites and notochord and RFP (red) expression is seen in forebrain and notochord. B) Diagram of the ED transposon located in the *nog2* locus in the ED301 transgenic line. Arrows represent primers used to confirm the ED transposon integration site. C) Sequence of the ED vector and adjacent genomic regions (dark grey) in the ED301 integration. D) Agarose gel showing PCR products obtained using the depicted primers (see Table S1), confirming the ED301 integration. E) In situ hybridization of *nog2*, in a 18hpf representative embryo. fb – forebrain, nc – notochord, st – somites. Scale bar represents 100 µm.

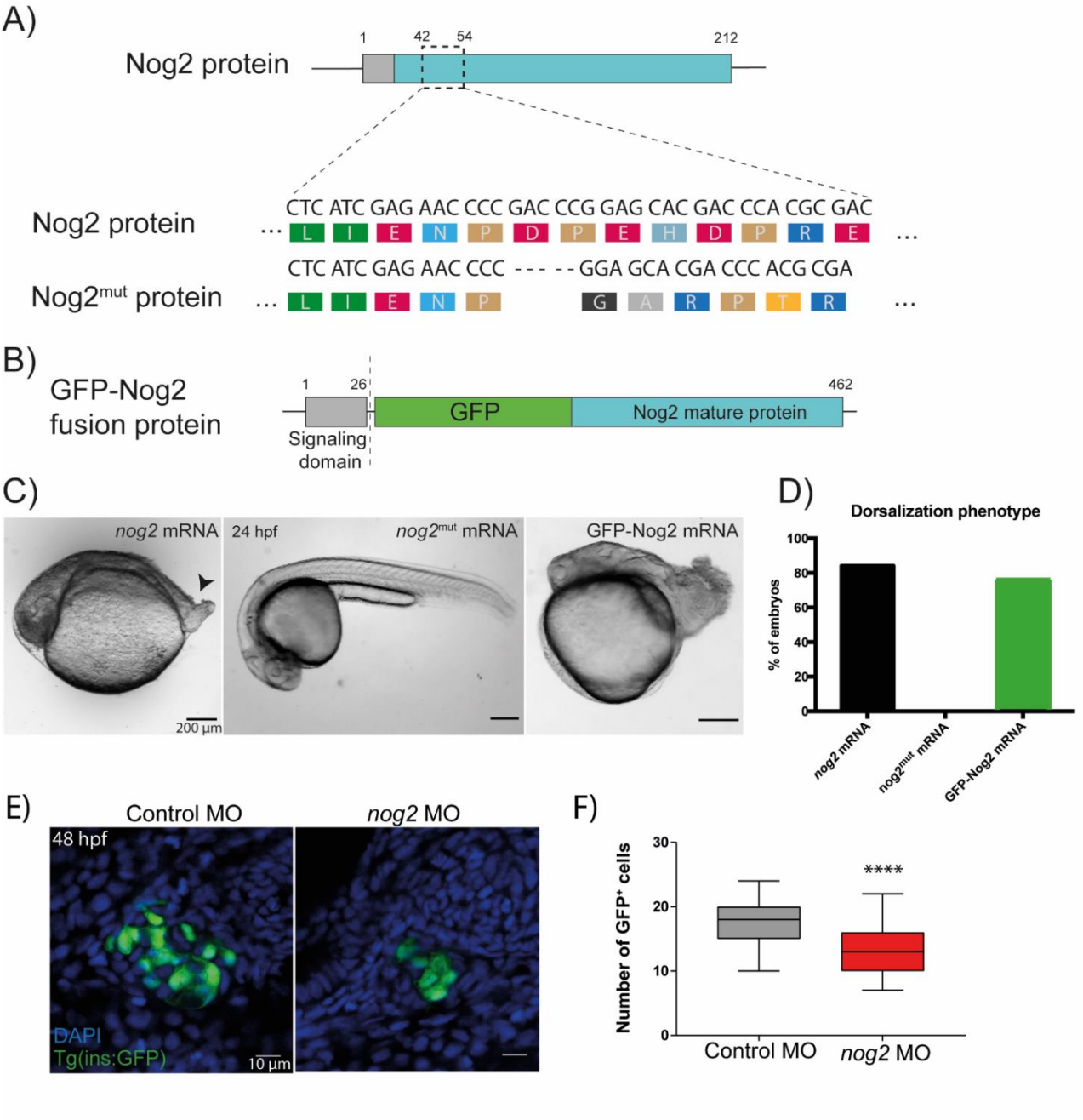
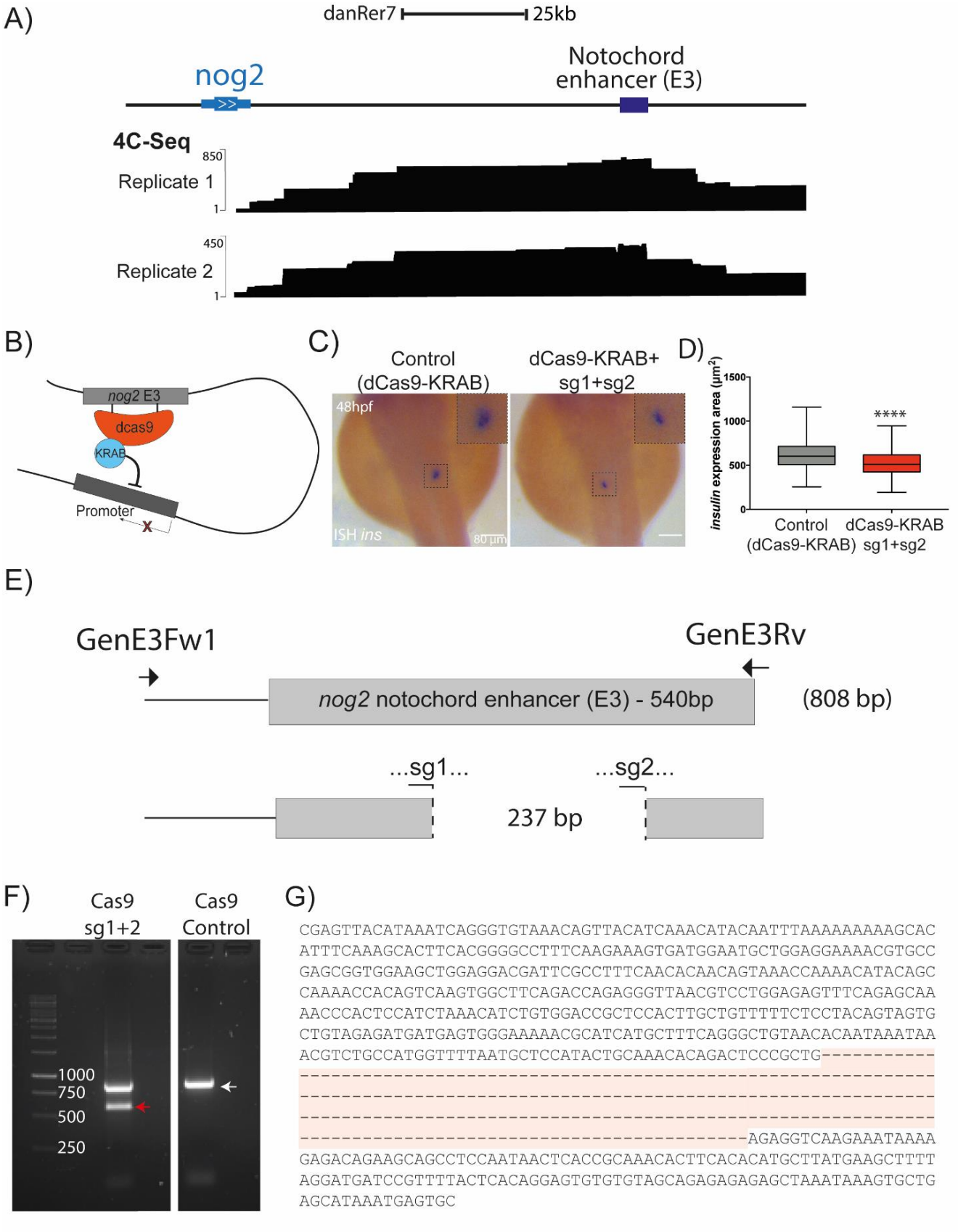


Figure S2. Related to Figures 1, 3 and 4

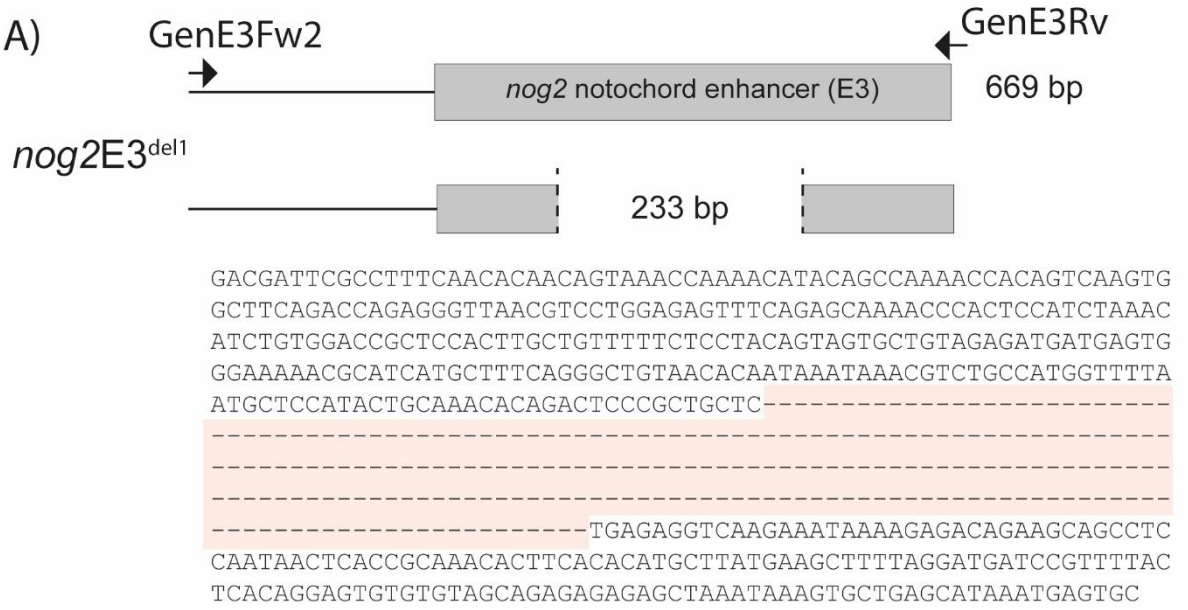
A) Representation of the nucleotide and corresponding translated amino acidsequences of Nog2 and Nog2^{mut} proteins, focusing on the region of the mutation. B) Diagram of the GFPNog2 fusion protein, depicting the signaling domain of Nog2, which is removed during the maturation of the protein. C) Representative images of 24hpf embryos injected with Nog2 mRNA, Nog2^{mut} mRNA and GFPNog2 mRNA. Only the mutant version of the protein fails to cause hyper-dorsalization phenotypes (embryos with ovoid shape and loss of ventral structures such as tail and trunk – arrowhead) Scale bars represent 200 µm. D) Quantification of thenumber of embryos showing dorsalization phenotypes. In all cases n≥61. E) Representative confocal images of an insulin GFP reporter line (Tg (ins:GFP); green) for 48hpf embryos injected with a control morpholino (Control MO) and a *nog2* targeting morpholino (*nog2*MO). Embryos were counterstained with DAPI nuclear marker (blue) Scale bars represent 10 µm. F) Quantification of the number of GFP positive cells from 48hpf Tg (ins:GFP) embryos injected with a control morpholino (Control) and a *nog2* targeting morpholino (*nog2*MO). Error bars represent SD and **** denotes p-value < 0.0001. In both cases n=52.


A)
danRer7
25kb
nog2
Notochord enhancer (E3)
4C-Seq
850
Replicate 1
1
450
Replicate 2
1
B)
nog2 E3
dcas9
KRAB
Promoter
C)
Control (dCas9-KRAB)
dCas9-KRAB+ sg1+sg2
48hpf
ISH ins
80 µm
D)
insulin expression area (µm²)
1500
1000
500
0
****
Control (dCas9-KRAB)
dCas9-KRAB sg1+sg2
E)
GenE3Fw1
GenE3Rv
nog2 notochord enhancer (E3) - 540bp
(808 bp)
...sg1...
...sg2...
237 bp
F)
Cas9 sg1+2
Cas9 Control
1000
750
500
250
G)
CGAGTTACATAAATCAGGGTGTAAACAGTTACATCAAACATACAATTTAAAAAAAAAGCAC
ATTTCAAAGCACTTCACGGGGCCTTTCAAGAAAGTGATGGAATGCTGGAGGAAAACGTGCC
GAGCGGTGGAAGCTGGAGGACGATTCGCCTTTCAACACAACAGTAAACCAAAACATACAGC
CAAAACCACAGTCAAGTGGCTTCAGACCAGAGGGTTAACGTCCTGGAGAGTTTCAGAGCAA
AACCCACTCCATCTAAACATCTGTGGACCGCTCCACTTGCTGTTTTTCTCCTACAGTAGTG
CTGTAGAGATGATGAGTGGGAAAAACGCATCATGCTTTCAGGGCTGTAACACAATAAATAA
ACGTCTGCCATGGTTTTAATGCTCCATACTGCAAACACAGACTCCCGCTG----------

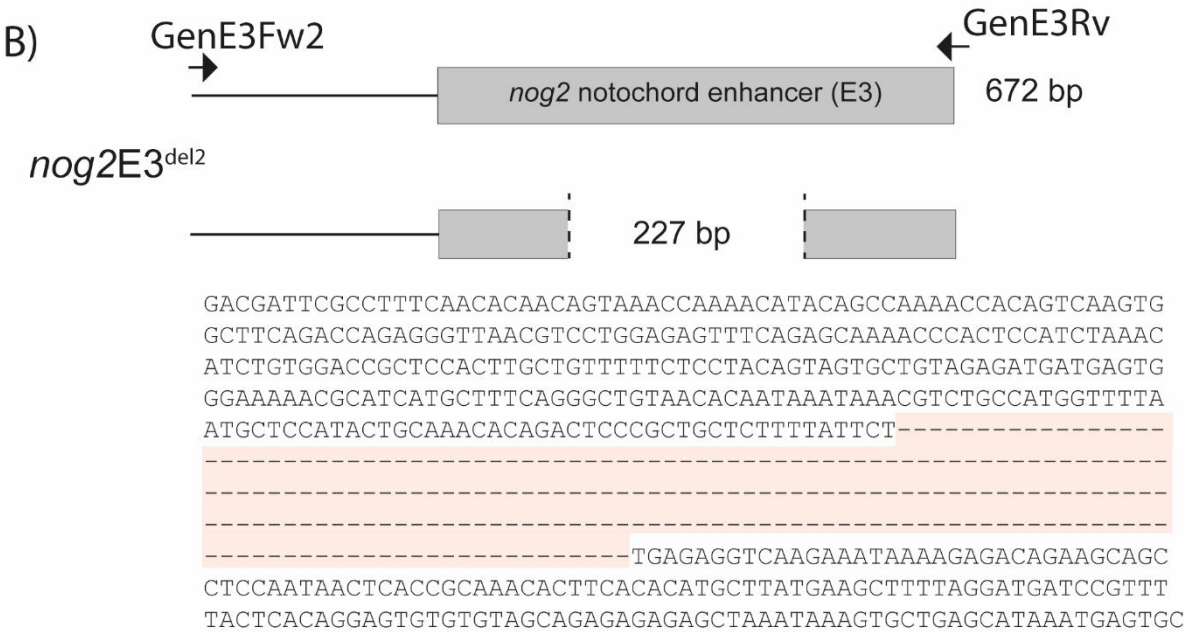
----------AGAGGTCAAGAAATAAAA
GAGACAGAAGCAGCCTCCAATAACTCACCGCAAACACTTCACACATGCTTATGAAGCTTTT
AGGATGATCCGTTTTACTCACAGGAGTGTGTGTAGCAGAGAGAGAGCTAAATAAAGTGCTG
AGCATAAATGAGTGC

Figure S3. Related to Figures 2 and 4

A) 4C-seq replicates (black tracks) showing a physical interaction between the *nog2* promoter and a region of approximately 100 kb downstream of the gene, containing the E3 notochord enhancer. Scale bar represents 25 kilobases. B) Representation of the repression of *nog2* expression, mediated by the KRAB domain, fused to a dCas9 protein, which targets the nog2E3 enhancer via sg1 and sg2. C) *In situ* hybridization of *insulin* in 48hpf representative embryos injected with dCas9-KRAB mRNA and two sgRNAs, comparing to control embryos, injected with only dCas9-KRAB. Scale bars represent 80 μm. D) Quantification of the *insulin* expression area, detected by *in situ* hybridization in injected embryos at 48hpf, comparing to controls. Error bars represent SD and **** denotes p-value < 0.0001. In all cases n≥99. E) Representation of the sg1 and sg2 location within the E3 enhancer and predicted genomic deletions with an expected size of 237bp. The full size of the E3 sequence is 540 bps and was selected based on ATAC-Seq and Chip-Seq data (see Figure 2 – pale blue box). Primers used to genotype the somatic deletions are also shown (Arrows; GenE3Fw1 and GenE3Rv; see Table S1). F) Upon injection of Cas9 protein, sg1 and sg2 RNAs (Cas9 sg1+2), a batch of 8 embryos was genotyped, using the GenE3Fw1 and GenE3Rv primers, amplifying a shorter band compatible with a 237bp genomic deletion (red arrow). This band was not present in control animals injected with Cas9 alone (Cas9 Control), where only the wild type band was observed (white arrow). Cas9 sg1+2 and Cas9 control are not consecutive lanes from the same gel. G) The short band observed in B) (red arrow) was sequenced, confirming a somatic deletion in the E3 enhancer.


A)
GenE3Fw2
GenE3Rv
nog2 notochord enhancer (E3)
669 bp
nog2E3del1
233 bp
GACGATTCGCCTTTCAACACAACAGTAAACCAAAACATACAGCCAAAACCACAGTCAAGTG
GCTTCAGACCAGAGGGTTAACGTCCTGGAGAGTTTCAGAGCAAAACCCACTCCATCTAAAC
ATCTGTGGACCGCTCCACTTGCTGTTTTTCTCCTACAGTAGTGCTGTAGAGATGATGAGTG
GGAAAAACGCATCATGCTTTCAGGGCTGTAACACAATAAATAAACGTCTGCCATGGTTTTA
ATGCTCCATACTGCAAACACAGACTCCCGCTGCTC----------

----------TGAGAGGTCAAGAAATAAAAGAGACAGAAGCAGCCTC
CAATAACTCACCGCAAACACTTCACACATGCTTATGAAGCTTTTAGGATGATCCGTTTTAC
TCACAGGAGTGTGTGTAGCAGAGAGAGAGCTAAATAAAGTGCTGAGCATAAATGAGTGC


B)
GenE3Fw2
GenE3Rv
nog2 notochord enhancer (E3)
672 bp
nog2E3del2
227 bp
GACGATTCGCCTTTCAACACAACAGTAAACCAAAACATACAGCCAAAACCACAGTCAAGTG
GCTTCAGACCAGAGGGTTAACGTCCTGGAGAGTTTCAGAGCAAAACCCACTCCATCTAAAC
ATCTGTGGACCGCTCCACTTGCTGTTTTTCTCCTACAGTAGTGCTGTAGAGATGATGAGTG
GGAAAAACGCATCATGCTTTCAGGGCTGTAACACAATAAATAAACGTCTGCCATGGTTTTA
ATGCTCCATACTGCAAACACAGACTCCCGCTGCTCTTTTATTCT----------

----------TGAGAGGTCAAGAAATAAAAGAGACAGAAGCAGC
CTCCAATAACTCACCGCAAACACTTCACACATGCTTATGAAGCTTTTAGGATGATCCGTTT
TACTCACAGGAGTGTGTGTAGCAGAGAGAGAGCTAAATAAAGTGCTGAGCATAAATGAGTGC

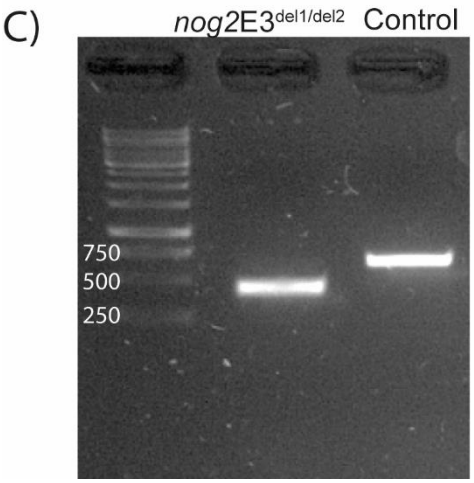

C)
nog2E3del1/del2
Control
750
500
250

Figure S4. Related to Figure 4

A and B) Representation of two generated genomic deletions of the E3 enhancer (del1 and del2), with sizes of 233 and 227bp, respectively. Adult heterozygous fish were genotyped by fin clipping followed by genomic DNA extraction and PCR amplification, using the GenE3Fw2 and GenE3Rv primers (see Table S1). The bands corresponding to alleles containing the deletion were extracted and sequenced. C) One heterozygous male and one female were crossed and their progeny was genotyped to select wt (control) and homozygous embryos for the deletion (*nog2*E3$^{del1/del2}$). Correspondent amplified bands are shown, in a 1.5% agarose gel.

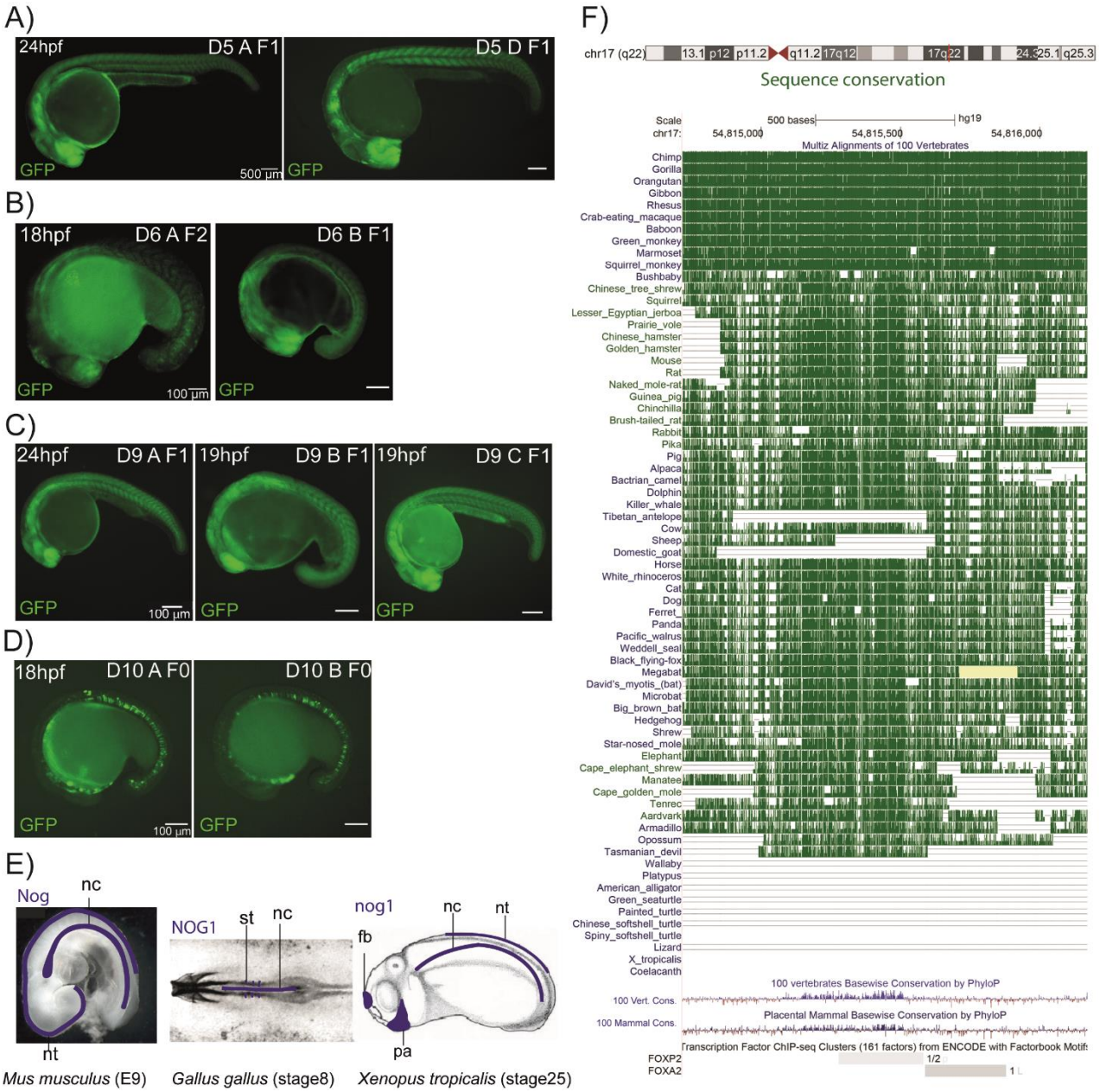
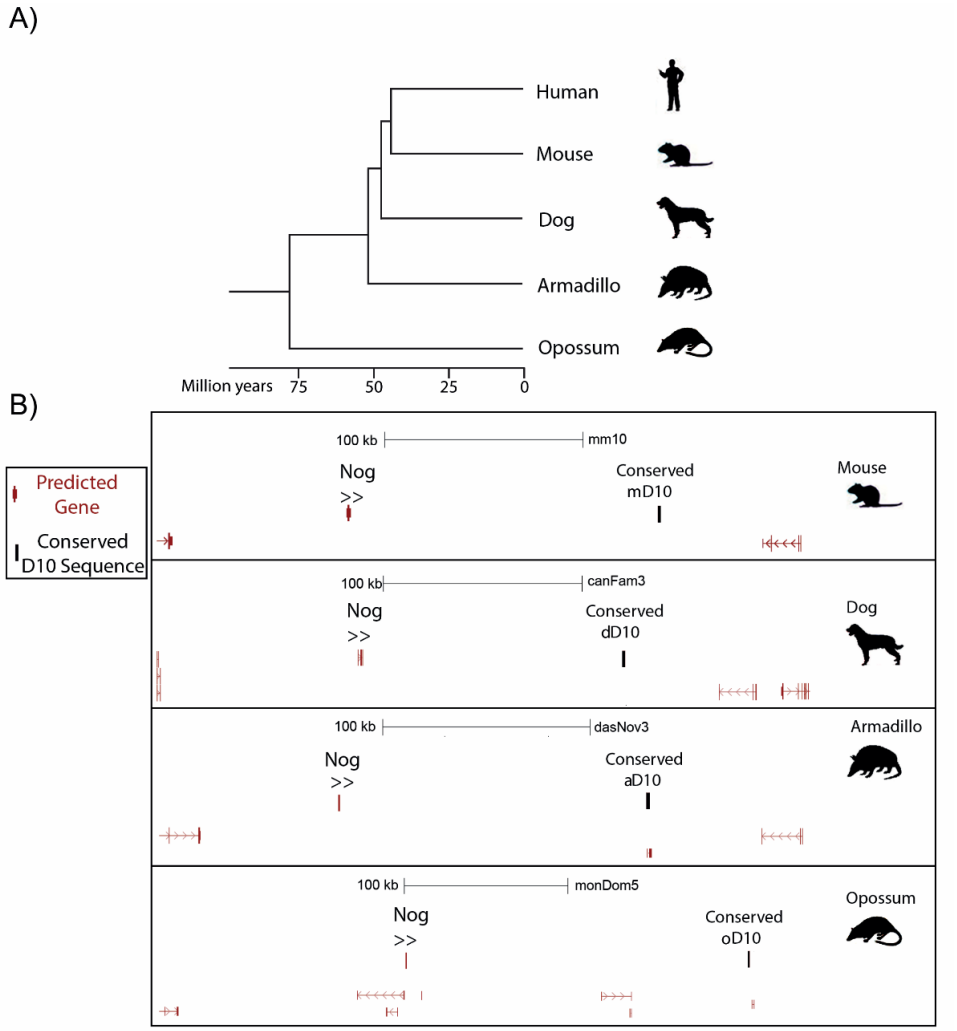


Figure S5. Related to Figure 5

Representative images of two or more independent transgenic lines showing a reproducible GFP (green) expression pattern for the A) D5 (somites, forebrain and hindbrain), B) D6 (somites and forebrain) and C) D9 (not restricted) enhancers. D) Representative images of two injected embryos showing reproducible expression pattern of GFP in the notochord, driven by the D10 enhancer. E) Representation of the expression patterns of *Nog* genes (dark blue) of *Mus musculus*, *Gallus gallus* and *Xenopus tropicalis* at early developmental stages. fb – forebrain, nc – notochord, nt – neural tube, pa – pharyngeal arches, st – somites. F) Tracks from USCS genome browser Multiz Alignments of 100 Vertebrates (Blanchette et al., 2004) showing conservation of the human D10 enhancer among mammal species. The most distantly related species that D10 is able to be aligned with, are two representatives of marsupials, Opossum (*Monodelphis domestica*) and Tasmanian Devil (*Sarcophilus harrisii*).


A)
Human
Mouse
Dog
Armadillo
Opossum
Million years 75 50 25 0
B)
Predicted Gene
Conserved D10 Sequence
100 kb
mm10
Nog
>>
Conserved mD10
Mouse
canFam3
Conserved dD10
Dog
dasNov3
Conserved aD10
Armadillo
monDom5
Conserved oD10
Opossum

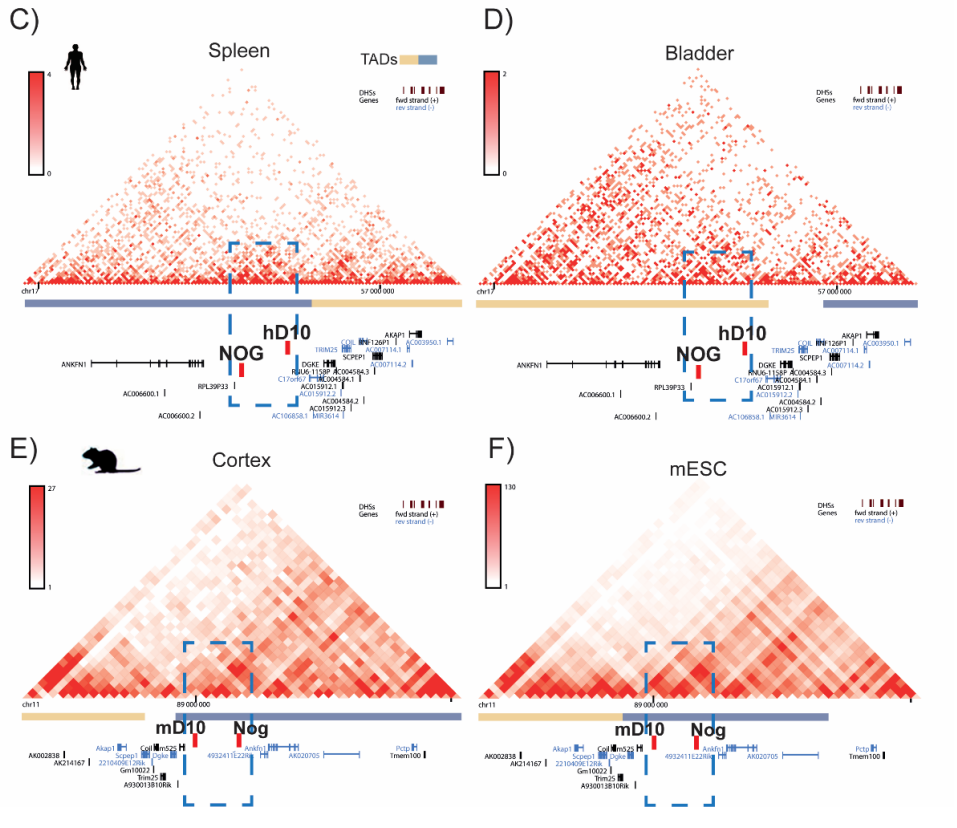

C)
Spleen
TADs
hD10
NOG
D)
Bladder
E)
Cortex
mD10
Nog
F)
mESC

Figure S6. Related to Figure 5

A) Phylogenetic tree representing the estimated time of divergence (million years) of selected mammals (humans, mouse, dog, armadillo and opossum) that share sequence conservation detected by alignment with the human D10 enhancer. B) Representation of the genomic landscape of Nog genes of selected mammals, showing synteny blocks containing Nog and ortholog sequences aligned with the human D10 enhancer. C and D) Representation of Hi-C data depicting the interaction between the human NOG gene and the hD10 enhancer (both in red) in spleen and bladder tissues (Schmitt et al., 2016), respectively. E and F) Representation of Hi--C data depicting the interaction between the mouse Nog gene and the mD10 enhancer (both in red) in cortex and mESCs (Dixon et al., 2012), respectively. Images were generated using the 3D Genome Browser (Wang et al., 2018).