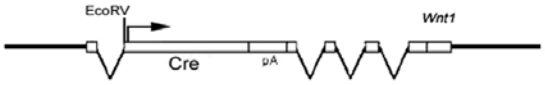
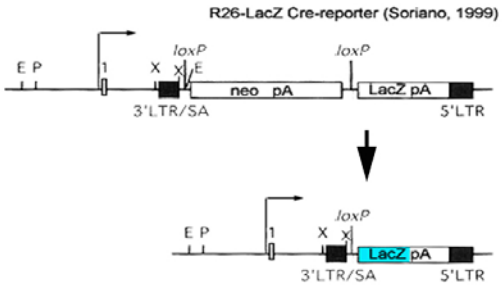
Supplemental information
S1: Materials, Methods and Analysis

Constructs

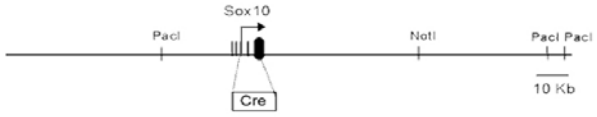
1. *Wnt1-Cre* transgenic (Danielian et al. 1998)



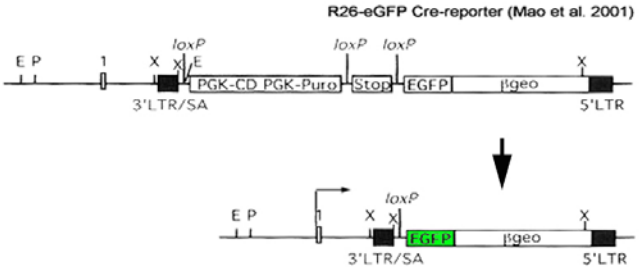
X



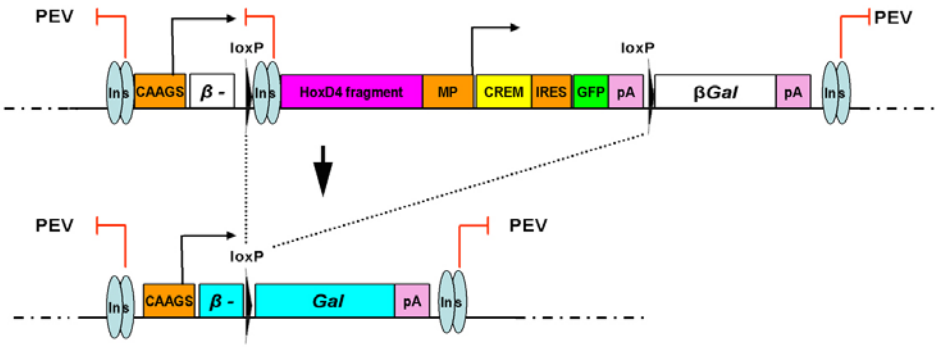
2. 170Kb *Sox10-Cre* PAC transgenic (this study)



X



3. *HoxD4*-controlled self-excising *Cre*-transgenic (this study)



A. Constructs, transgenesis and genotyping

1. *Wnt-1-Cre* construct

In order to permanently label the pre-migratory neural crest precursor population, we utilized the wellcharacterized
Wnt-1 dorsal neural tube enhancer region [51]. The *Wnt-1* enhancer region used is conferring transgene expression onto the entire dorsal neural tube neural crest precursor population with the exception of that in rhombomere 1 (Andrew McMahon, GK, unpublished). Thus, we cannot safely determine whether the unlabelled parietals (seen in Fig. 2b) are neural crest derivatives [4,8] of rhombomere 1 or mesodermal[9]. Only positive genetic labeling data would be admissible as valuable contribution to an old as yet unresolved controversy which has not been provided yet for head mesoderm. However, it is very interesting to note a razor-sharp boundary of blue staining in Wnt-1-Cre/reporter mice which does not coincide with the frontoparietal suture, but runs imperceptibly parallel to this suture (dotted line in Fig. 2b) inside the parietals. This corroborates the prevailing notion that developmental boundaries do not coincide with anatomical boundaries. From the *HoxD4* transgenics (see below) we can only safely infer that it is not derived from somites 5 or posterior. As our study only focuses on the post-otic neural crest (PONC) i.e. neural crest in the postparietal/occipital/ neck region this issue is not relevant in our context. Postmigratory neural crest does not express *Wnt1* nor does it activate the transgenic enhancer. *Wnt1-Cre* transgenic mice were constructed and genotyped as described previously [17,52].

2. *Sox10-Cre* PAC construct

To label the complementary entire post-migratory postotic neural crest (PONC) population we made Cre expression conditional upon expression of *Sox10*, the best *bona fide* PONC marker known to date [17, 53]. To generate the *Sox10*-Cre transgenic, a genomic PAC (RP21-529-I6 from RPCI mouse PAC library 21, HGMP Resource Centre, UK) spanning 170 Kb around the *Sox10* locus was used. The entire coding region of *Sox10* was replaced with that of a nuclear-targeted Cre recombinase (a gift from Sam Aparicio) by homologous recombination in bacteria [54]. Further details of construction and transgene expression will be published elsewhere (NK, PI and WDR). Several transgenic founders were obtained. Genomic DNA was prepared from the tip of the mouse tail and PCR with the primer sets of Cre5' and MSox10rev was performed for genotyping. The primer pair amplifies a 1400bp genomic DNA product in transgenic founders.

Cre-recombinase reporters

Several *Wnt-1-Cre* and 3 different *Sox10-Cre* founders were crossed with Cre-reporter females carrying transgenic floxed resistance-pA cassettes in front of LacZ [45] or eGFP [46] under the control of the Rosa26 locus. Notably, LacZ and GFP proteins donot carry a nuclear localization sequence (NLS), therefore fluorescence (and blue staining) is mainly cytoplasmic. After Cre-mediated excision of the floxed cassettes F_1 offspring expresses LacZ or GFP which was assayed in a detailed time course of daily stages between E9.5 and P7.

3. *HoxD4*-controlled self-excising Cre transgenic

a. A novel recombinase strategy

In order to genetically label the entire offspring of various mesenchymal stem cell populations we generated a novel vector construct (see Fig. above). Traditional crossing of Cre-drivers and reporters is time-consuming and costly and different transgenes (of driver and reporter) are subject to differential positional effect variegation (PEV) potentially leading to considerable variability. We therefore designed a vector which contains Cre- driver and reporter in one, thus unifying potential external (chromosomal) effects on both transgenes. Moreover, recent studies had shown that Cre recombinase in high concentration can recognize other sites than its cognate loxP sites in the mammalian genome, which can lead to deleterious double-strand break repair and fusion of non-sister-chromatids and ensuing apoptotic cascades [59]. This would create considerable noise in microarray-based expression profiling studies on the genetically labeled mesenchymal stem cells currently undertaken in our lab (GK). To obviate this serious problem of standard recombinase technology we designed the construct in a way that Cre-recombinase excises itself, after it has done the first (and essential) marker activation: indeed the marker activation itself was made conditional upon Cre-self excision by placing a floxed Cre-recombinase inside the LacZ coding region, thereby interrupting its LacZ reading frame. In order to accomplish this and carry Cre as well as its cognate loxP sites on the same vector we took advantage of CREM, an artificial intron-containing Cre coding

region whose protein cannot be produced by bacterial cloning hosts that are unable to splice [55]. The hrGFP was used to monitor the time window of CREM production prior to self- excision of the CREM-IREShrGFP cassette in the embryo. As the Rosa26 locus had shown downregulation in many skeletal tissues after P21 (Andrew McMahon, GK, unpublished observations) we used the ubiquitous CAAG/S promoter [56] to control the reporter transgene after Cre-excision. In order to ensure that Cre-activation is strictly controlled by an enhancer of our choice (in this case a HoxD4 region) we placed insulators between the strong CAAGS promoter region and a minimal promoter in front of the bicistronic CREM-IREShrGFP cassette (kind gift of Adam West and Gary Felsenfeld). The entire construct was flanked by insulators to block mis-regulation of the transgene by the chromatin region flanking its insertion (PEV: position effect variegation in figure 2a3.above, [57]).

b. Cloning strategy of basic Cre self-excision vector p6'tagExon

The *Eco*RI/*Sal*I fragment from p240 [55] containing the 3' portion of modified cre-recombinase (CREM) was removed and ligated into the *Eco*RI/*Xho*I sites of pIRES-hrGFP-1a (Stratagene) resulting in the generation of pCREM-IRES-hrGFP.

Minimal promoter (MP) and a 5' portion of CREM were amplified by recombinant PCR using a primer set of MPs (5' GAATTCTTAATTAAAGCTTTGGCCGGCCAGCTTTGTTTAAACGGCGCGCCGGGCTGGGCATAA AAGTCAGG-3') and MPas (5'-TTGGAATTCTGGTACCGTCAGTAAATTCGCCATGGCGCCGCGCTCTGCTTC-3') using the minimal promoter region of pBGZ40 as template [58]. The latter has been used extensively for *Hox* gene enhancer analysis. The product contains artificial *EcoRI, PacI,* and *PmeI* restriction sites at the 5' region and cloned into pCRII-TOPO (Invitrogen). The *Eco*RI fragment which contains MP was inserted into the *Eco*RI site of pCREM-IRES-hrGFP resulting in the generation of pMP-CREM-IRES-hrGFP.

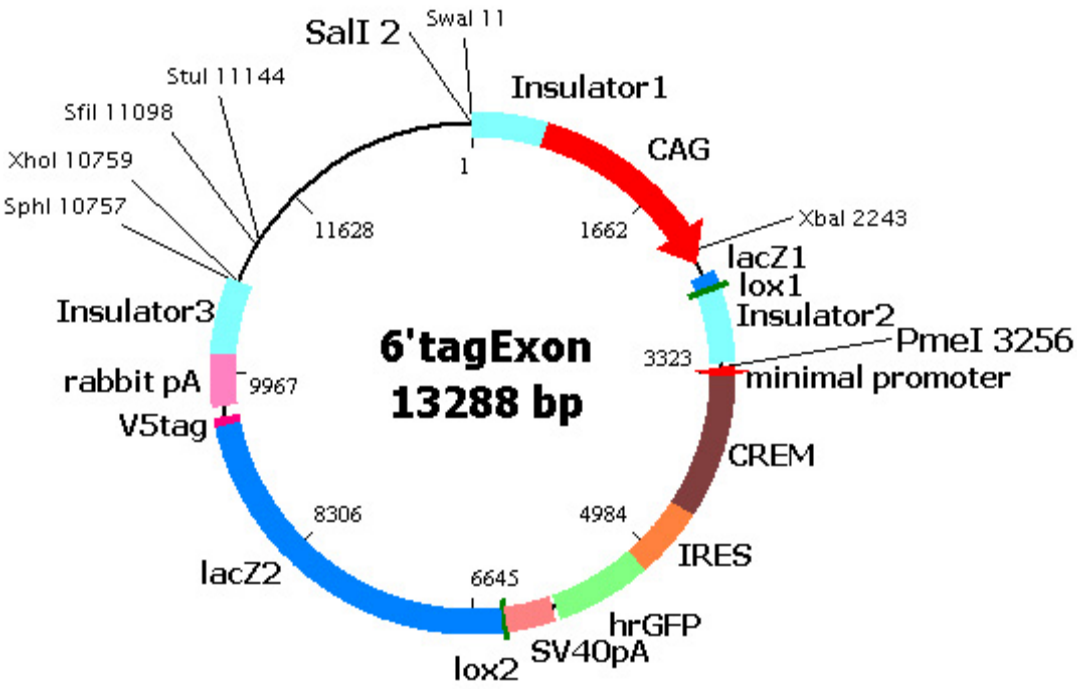
pCAG-EGFP/ß-gal (p169, [55].) was digested with *KpnI* to remove LOX and EGFP and synthetic oligonucleotides linker1s (5'-CGATAACTTCGTATAGCATACATTATACGAAGTTATTTAATTAAGGCGCGCCATAACTTCGTA TAGCATACATTATACGAAGTTATCGGTAC -3') and linker1as (5' CGATAACTTCGTATAATGTATGCTATACGAAGTTATGGCGCGCCTTAATTAAATAACTTCGTAT AATGTATGCTATACGAAGTTATCGGTAC -3'), designed to contain *Pac*I and *Asc*I site between the LOX sites, were cloned into *KpnI* site of p169 resulting in the generation of pCAG-PA/ß-gal.

PstI site of pCAG-PA/ß-gal was replaced by *XhoI* site by introducing the linker2s (5'-CCTCGAGGTGCA-3') and linker2as (5'-CCTCGAGGTGCA-3') into *PstI* site of this vector resulting in the generation of pCAG-PA/ß-galX.

The *PacI/MluI* fragment from pMP-CREM-IRES-hrGFP was removed and ligated into the *Pac*I/*Asc*I sites of pCAG-PA/ß-galX resulting in the generation of pCAG-MP-CREM-IRES-hrGFP/ß-galX.

pIRES-hrGFP-1a was digested with *SexAI* and synthetic oligonucleotides STOPs (5'- CCAGGA TAA CTGA TAG TTGA TAA TAG AT TAA TT TAG TGA ggttacc A-3') and STOPas (5'-CCTGGTggtaaccTCACTAAATTAATCTATTATCAACTATCAGTTAT -3') were inserted resulting in the generation of pIRES-hrGFP-STOP.

The 3' region of CREM was PCR amplified using CREMs (5'-CTGATTTCGACCAGGTTCGTTCACTCATGG-3') and CREMas (5'-acctggtggtaaccCTAATCGCCATCTTCCAGCAGGCGC-3') which contains *SexAI* restriction site at the 3' region and *BstEII* and *SexAI* site at the 5' region and cloned into pCRII-TOPO. At the *BstEII* site of this construct *BstEII* fragment from pIRES-hrGFP-STOP was cloned and resulted in the generation of pCREMIRES-hrGFP-STOP. The CREM-IRES-hrGFP region of pCAG-MP-CREM-IRES-hrGFP/ß-galX was replaced by the *SexAI* fragment from pCREM-IRES-hrGFP-STOP to remove the polyA from CREM and make sure stopping the translation before beta-galactosidase sequence, resulting in the generation of pCAG-MP-CREM-IRES-hrGFP-STOP/ß-galX. Then we added a V5-His tag epitope (Stratagene) to the end of the ß-galX cassette to allow alternative detection of the LacZ-V5-His fusion protein as well as the addition of other cistrons (such as IRES-RFP, -CFP etc in the future). Pairs of 250bp long HS4 insulator core elements [57,59] were introduced at the *SalI* site before CAAGS, into the *PacI* site before the MP and into the *XhoI* site after ß-gal-pA. This resulted in vector p6'tagExon which carries a PmeI site (3256) in its multiple cloning site for insertion of enhancer regions of choice.



c. Cloning of HoxD4 enhancer fragment

Landmark studies [18,47,60] on the correspondence between Hox gene expression and axial somitic identity had predicted Hox 4 paralogues to define the non-occipital/occipital boundary. In order to exclude neural-crest expression and confine genetic labelling to somite 5 and posterior (as the occipital plane exactly intersects the re-segmented somite 5 [14] we chose the HoxD4 gene, which is not expressed in neural crest, but in somite 5 and posterior[62,63]. This mesodermal and neural tube expression is regulated by a 12.5 Kb genomic region [48] encompassing a large 5' region with a RARE [61] and a 3' 3.6 kb genomic Region A [65]. We chose this entire fragment as it accurately reflects endogenous somitic gene expression from somite 5 backwards but conveniently excludes (endogenous) expression in lateral plate mesoderm and primitive streak: domains of expression that might have resulted in the entire transgenic embryo being labeled and that are under long-distance co-linear gene cluster control [66,67]. We TOPO-cloned (Invitrogen) this region by highfidelity
long-template PCR (Roche) using recombinant primers, sequenced, excised and blunted it and inserted it into the PmeI site of the multiple cloning site of p6'tagEXON in front of its minimal promoter. 4 transgenic founders were produced using standard pronuclear injection protocols. They were identified by genomic Southern hybridization with internal CREM-IRES-hrGFP radioactive probes after restriction digest (to determine numbers of independently inserted copies) and by PCR using primers in CREM and LacZ. 4 independant transgenic founders with multiple independent insertions were mated to CD1 wildtype females and F_1 offspring was analyzed.

d. Controls of CAAGS-CREM function and Cre self-excision

In order to test the CREM functionality of the p6'tagExon construct and insulator efficiency we removed insulator 2 from the construct and transfected both plasmids independently into the C_2C_{12} cell line using lipofectamine and assayed for hrGFP and LacZ expression in a time course over 3 days after transfection. As *HoxD4* is not expressed in this cell line, the enhancer fragments were not active to elicit CREMhrGFP expression or LacZ expression. No hrGFP or LacZ staining was discernable. However, when the insulator 2 was removed, the strong CAAGS promoter could cross-activate the minimal promoter-CREM-IRES-hrGFP cassette, leading to hrGFP production after 20 hrs (to monitor transcription and translation of the bicistronic transgene) and full LacZ expression after Cre-self excision (and disappearance of hrGFP) in 90% of cells after 1.5 days (data not shown). As LacZ expression is only possible after Cre-self excision from the transgene we could therefore determine the speed of CRE action. (TM, GK manuscript in preparation) This probably also explains why we could not see any postotic neural crest derivatives (such as glossopharyngeal, vagal ganglia, heart outflow tracts, superior cervical ganglion etc) labeled in the HoxD4

transgenics as neural crest would have parted from the r6/7 and posterior neural tube precursor population before HoxD4 (and the HoxD4 enhancer transgene driving Cre self-excision) and CREM becomes active.

e. Utility and Limitations of p6'tagExon construct

The 6'tagExon labellling construct is particularly useful in cases in which cell fates have to be determined after about P20 at which time expression from the traditional Rosa26 reporters becomes sketchy even in regions that were e strongly LacZ/GFP positive at P6-9 (Andy McMahon & GK pers. observations). The strong ubiquitous CAAGS promoter [56] obviates this problem. The dual insertion of Cre-driver and reporter and the fact that reporter expression is conditional upon Cre-excision solves the problem of potentially differential position effect variegation (PEV) effects as well as the deleterious effects of strong Cre expression which would be monitored if such transgenic lineages were used for sophisticated microarray expression profiling that would reveal extensive apoptotic cascades being activated (Stratagene). The insulator sequences we introduced block position effect variegation (PEV) that is caused by transgene insertion into silenced regions or cross-activation of constructs by concatemeric adjacent copies of CAAG/S. However, the regulatory region to drive the CREM-IRES-hrGFP transgene needs to be well known. Particularly important is the fact that this region shall not be expressed in the founders' germ-line, otherwise the entire transgenic embryonic offspring would be GFP/LacZ+. We retrospectively checked this fact by inspecting gonads of male and female second generation transgenic founders after sexual maturity for LacZ or V5 tag expression and found none. Due to multiple insertions of transgenic constructs the ratio of transgenic offspring with wild-type females is >50%, thus leading to considerable savings in time and resources. As different IRES cassettes can be placed behind lacZ in p6'tagExon, multiple lineage labeling vectors can be used simultaneously to 'triangulate' positions of cells in lineage trees based on multiple gene activations (TM, GK in preparation).

B. Methods

Alcian blue and Alizarin red stain

Wild-type mice skeletons were double-stained with Alcian blue 8 GX (Sigma A-5268) and Alizarin red S (Sigma A-5533) for cartilage and bone at P5 and subsequently cleared, and stored in glycerol as described previously [13].

Beta-Gal staining

Whole mount X-Gal staining was carried out on embryos from E10.5 to P5 as described [51,52,17]. Macro pictures were taken with a FUJIX digital camera HC-300Z (Fuji film) mounted on Leica MZFLIII microscope. Transgenic mice were embedded in OCT (Tissue-Tek, Mile, Inc.) over dry ice after equilibration in the following solutions until they settled at the bottom: PBS, 5% sucrose in PBS, 10% sucrose in PBS,15% sucrose in PBS, and a 1:1 mixture of OCT and 15% sucrose in PBS. Sections were cut at 8 μm, embeeded in Histoclear and photographed using Nomarski differential interference contrast on a Zeiss Axiophot.

Immunohistochemistry and confocal microscopy

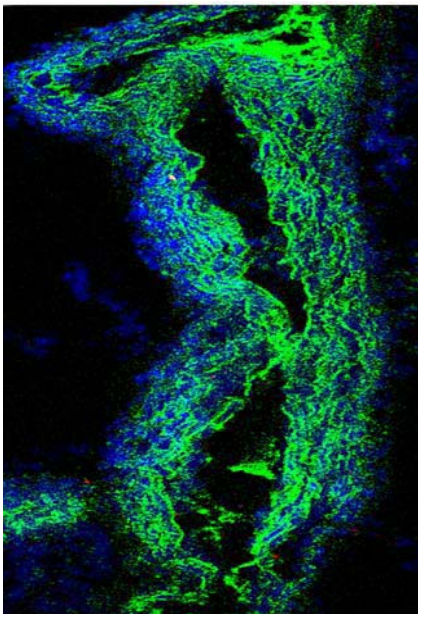
In order to detect transgenic GFP and LacZ (in neural crest cells) as well as muscle myosin we performed immunohistochemistry on slides. Sections were washed 3x in PBS for 10 minutes at room temperature. To block endogenous peroxidase, the samples were incubated in PBS containing 0.1% hydrogen peroxide for 1-2 h. After washing 3x in PBS primary antibody (goat anti-beta-galactosidase, Biogenesis; dilution of 1/400) (rabbit anti-GFP, ab290, abcam ltd.; dilution of 1/1500) (mouse monoclonal anti-skeltal myosin, clone MY-32, Sigma; dilution of 1/400) solutions were applied for overnight at 4 °C. After extensive washing, the samples were then incubated with donkey anti-goat IgG, donkey anti-rabbit IgG, and donkey anti-mouse IgG conjugated to peroxidase, Cy2, Cy3 or Cy5 (Jackson Immuno Research), respectively, for 60 min at room temperature. For the samples stained by the peroxidase-conjugated secondary antibodies, tyramide signal amplification kits (Alexa Fluor 488 dye, Alexa Fluor 546 dye, Molecular Probes) were used following their application protocols. Samples were mounted in Vectashield with DAPI (Vector Laboratories), covered with a cover glass and sealed with nail polish. Samples were analysed with a Leica TSC confocal laser scanning microscope using a 10× or 20× lens. Alexa488 or Cy2 labelled antibodies were excited at 488 nm and emission was measured above 500 nm (FITC channel). Alexa546 or Cy3 labelled antibodies were excited at 543 nm and emission measured above 555 nm (TRITC channel). Cy5 labelled antibodies were excited at 633 nm and emission measured above 655 nm. DAPI staining was excited at 364nm and emission measured above 400 nm. A total of 20 sections were taken for each sample and assembled as extended focus views in a Z-series.

C. Analysis

A detailed time course of double transgenic offspring was examined in daily intervals between E9.5 and P7. Two investigators (TM and GK) were independently scoring positions of GFP/LacZ+ labeled cells. The full extent of GFP+ and LacZ+ cell populations in *Wnt-1* and *Sox10-Cre/R26R-LacZ* and *–GFP* double transgenic offspring was identical which suggested to us that in mice – despite its early emigration from the neural tube-ectoderm transition zone all Sox10+ cells are derived from precursors that previously express *Wnt1*. All fate map data shown in this paper are therefore corroborated by two different independent experiments, employing genes in unconnected pathways and providing a firm experimental grounding. Similarly, labeled populations in offspring of 4 independent HoxD4-Cre founders had identical distribution. A fifth founder was transgenic by PCR but no LacZ stain was detectable in offspring.

1. Controls for accuracy of transgenes

a. Wnt-1, Sox10 transgenics



In order to ascertain that Wnt-1 and Sox10- activation of the Cre transgene accurately reflects endogenous gene expression we performed double in situ hybridization for Cre- and Sox10 on E9-E10.5 embryos (data not shown). Wnt-1 transgene accuracy has been extensively checked [17,51,52] and we can fully confirm these previous results. We could not detect endogenous expression of Sox10 in neural crest at stages of skeletogenic differentiation (from E13 onwards), confirming that transgene activation must have happened in postmigratory neural crest prior to differentiation. Expression of Sox10 and the Sox10-Cre transgene as well as the eGFPCre reporter in oligodendrocytes of double transgenics proved to be a safe internal labeling control (see green cells in spinal cord sc of Fig. 2d (ref Wegner, scotting). Sox 10 is also expressed in pre-otic mouse neural crest between 4 and 20 somites, but gets downregulated in all branchial arches [53]. However, it is retained in cranial ganglia, DRGs and the entirety of postotic neural crest. Apparently this expression time window of 16 somites is not sufficient for transgenic eGFP marker activation in the pre-otic region as later eGFP expression in known preotic neural crest derivatives such as lower jaws are patchy. We assayed this behaviour by double antibody-staining for the pan-neural crest marker AP2 (kind gift of Trevor Williams) and GFP (at E9-13) and only postotic neural crest persists in being doubly positive (data not shown), thus confirming previous reports [17,53,68]. As a positive example, the tunica media inside heart outflow tracts that has been independently shown to be postotic neural crest derivatives by several groups are GFP+ in Sox10-Cre/ reporter double transgenics (see adjacent figure) [2].

b. HoxD4 construct

In order to test the accuracy of HoxD4 transgene expression we assayed transgenic F_1 offspring from E10.5 onwards for LacZ and V5/His in wholemounts and on sections and could confirm the previous results by Zhang et al. (data not shown) that the entire 12.5kb fragment around HoxD4 is needed for controlling HoxD4 expression in somite 5 and posterior [48]. As HoxD4 is also expressed in the neural tube posterior to r6 we were concerned that transgene activation might occur in vagal neural crest derived from these regions. We therefore assayed all previously known neuronal vagal and more posterior neural crest derivatives (Auerbach and Meissner's gut plexus, vagal, superior cervical ganglion, parasympathetic system and DRGs) [2], that cannot be mesodermal of origin as well as known mesenchymal derivatives such as heart outflow tracts for transgene activity but could not find any lacZ positive cells in these structures. From this we concluded that self-excision of Cre under HoxD4 control is mesoderm specific and neuronal but excludes neural crest. The axial levels of transgene activity are in line with embryological studies in chick showing that somite 5 spans the occipital plane [2,11,18].

2. Controls for embryonic tissue specificity

In order to ensure that we did not fall victim to hitherto unknown *wnt1* and *Sox10* gene (and transgene)

expression domains in developing skeletal elements we investigated in all transgenic specimens of the time course between E12 and P7 23 skeletal regions each that have been mapped safely as being entirely mesodermal in origin: the pelvic bone, the ilium the humerus, ulna, tibia, fibula, digits, sacral vertebrae at different stages of osteogenesis and chondrogenesis or generally in osteocytes (data not shown). None of these regions were GFP or LacZ positive while DRGs and Schwann cells surrounding motor axons were labelled. Moreover, in the vertebral column GFP-labeled connective tissue was confined to the trapezius attachment region around cervical somites $C_{1\text{-}7}$ and did not reach more posteriorly, corroborating previous studies that *Wnt1* or *Sox10* are not generally expressed in developing or mature periosteum, perichondrium, cartilage or bone. Thus, labeling in Wnt1, Sox10 transgenics was lineage-specific. Conversely, lateral plate mesodermal derivatives (such as limbs) were unlabelled in HoxD4 transgenics as HoxD4expression in lateral plate mesoderm is controlled by different genomic regions [62,63].