Nucleic Acids Research Advance Access published February 28, 2013

Nucleic Acids Research, 2013, 1–11
doi:10.1093/nar/gkt140

A far-upstream (−70 kb) enhancer mediates *Sox9* auto-regulation in somatic tissues during development and adult regeneration

Timothy J. Mead[1], Qiuqing Wang[1], Pallavi Bhattaram[1], Peter Dy[1], Solomon Afelik[2], Jan Jensen[2] and Véronique Lefebvre[1,*]

[1]Department of Cellular and Molecular Medicine, and Orthopaedic and Rheumatologic Research Center, Cleveland Clinic Lerner Research Institute, Cleveland, OH 44195, USA and [2]Department of Stem Cell Biology and Regenerative Medicine, Cleveland Clinic Lerner Research Institute, Cleveland, OH 44195, USA

Received December 21, 2012; Revised February 10, 2013; Accepted February 11, 2013

ABSTRACT

***SOX9* encodes a transcription factor that presides over the specification and differentiation of numerous progenitor and differentiated cell types, and although *SOX9* haploinsufficiency and overexpression cause severe diseases in humans, including campomelic dysplasia, sex reversal and cancer, the mechanisms underlying *SOX9* transcription remain largely unsolved. We identify here an evolutionarily conserved enhancer located 70-kb upstream of mouse *Sox9* and call it *SOM* because it specifically activates a *Sox9* promoter reporter in most *Sox9*-expressing somatic tissues in transgenic mice. Moreover, *SOM*-null fetuses and pups reduce *Sox9* expression by 18–37% in the pancreas, lung, kidney, salivary gland, gut and liver. Weanlings exhibit half-size pancreatic islets and underproduce insulin and glucagon, and adults slowly recover from acute pancreatitis due to a 2-fold impairment in *Sox9* upregulation. Molecular and genetic experiments reveal that Sox9 protein dimers bind to multiple recognition sites in the *SOM* sequence and are thereby both necessary and sufficient for enhancer activity. These findings thus uncover that Sox9 directly enhances its functions in somatic tissue development and adult regeneration through *SOM*-mediated positive auto-regulation. They provide thereby novel insights on molecular mechanisms controlling developmental and disease processes and suggest new strategies to improve disease treatments.**

INTRODUCTION

Heterozygous mutations within and around *SOX9* cause campomelic dysplasia (CD), a severe skeletal malformation syndrome that is often lethal at birth (1–3). Moreover, two-thirds of XY CD patients show sex reversal or abnormal genitalia (4,5), and surviving patients have been reported to experience hearing loss, seizures, global developmental delay and heart and pancreas malformations (6). These clinical features and mouse genetic and molecular studies have proven that *SOX9* specifies cell fate and differentiation in many lineages, including chondrocytes, Sertoli cells, neural stem cells, pancreas progenitor cells and neural crest, neuronal, glial, heart valve, gut and kidney cells (7–14). While CD and associated malformations are due to *SOX9* haploinsufficiency, *SOX9* duplication causes XX sex reversal (15), and increased or ectopic expression of *SOX9* has been linked to liver fibrosis, melanoma and colon, pancreas and prostate cancer (16–19). To better understand and manage *SOX9*-dependent diseases, it is imperative to uncover the mechanisms that underlie *SOX9* qualitative and quantitative expression. Currently, however, our knowledge of these mechanisms is still in its infancy.

Sequence alterations in the 2-Mb gene desert that surrounds *SOX9* have been shown to cause CD and have thereby implied that critical *SOX9 cis*-regulatory elements reside within or beyond this region (1–3). Accordingly, experiments with yeast (YAC) and bacteria (BAC) artificial chromosomes have revealed that the ~70-kb sequence lying 5′ of *Sox9* is sufficient to mimic most of the *Sox9* expression pattern in mouse embryos, but more robust expression is achieved with 350 kb of 5′ sequence (20,21). Experiments with smaller transgenes

*To whom correspondence should be addressed. Tel: +1 216 445 0762; Fax: +1 216 444 9404; Email: lefebvv@ccf.org

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

have shown that the *Sox9* promoter can drive gene expression in the spinal cord and hindbrain (22); a −1.5-Mb enhancer in the mandible region (23); a −251-kb enhancer in the cranial neural crest and inner ear; a −28-kb enhancer in the node, notochord, gut, bronchial epithelium and pancreas; a +95-kb enhancer in the telencephalon and midbrain (22) and a −14-kb enhancer, named *TES*, in the undifferentiated embryonic gonad and testis (21). The transcription factor Sf1 was proposed to cooperate with Sry to activate *TES* in the undifferentiated gonad, and with Sox9 to maintain *TES* activity in Sertoli cells after SRY expression has ceased. While these enhancers can activate transgenes, their actual contribution to endogenous *Sox9* expression remains unknown, as are the factors that mediate the activity of most of them.

We newly uncover here an enhancer that is highly conserved and located 70-kb upstream of *Sox9* and we show that this enhancer mediates a feedback loop of *Sox9* auto-regulation in multiple somatic tissues. The increase in *Sox9* expression generated by this feedback loop results in optimization of *Sox9* function in development and adult organ regeneration. These findings provide novel insights on molecular mechanisms controlling major processes and suggest new strategies to improve disease treatments.

MATERIALS AND METHODS

Genome analyses and construction of plasmids

Sox9 sequence conservation and histone modifications were analyzed using the University of California, Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/). *Sox9* elements were amplified using mouse 129 DNA and specific primers (Supplementary Table S1A) and sequence-verified on cloning. Upstream elements were multimerized using *Bam*HI and *Bgl*II flanking sites. Reporters were assembled in pBluescript (Stratagene) using *Sox9* sequences, a pACT intron (Promega), the *βGeo* and enhanced green fluorescent protein gene (*EGFP*) reporters separated by internal ribosome entry site (*IRES*; Clontech) and followed by a bovine *Gh1* polyadenylation site. *SOM* was mutated using Quick Change Mutagenesis (Stratagene). Expression plasmids for Sox2, Sox8 and Sox10 were made as reported for SOX9 (24) and Sox11 (25).

Transfection and protein assays

Rat chondrosarcoma (RCS), 10T1/2 and COS1 cells were cultured and transfected as described (26). Transfection mixtures contained 1.2 µl FuGENE6 (Roche), 200 ng *Sox9* reporter, 80 ng pGL3 control plasmid (Promega) and 120 ng empty or Sox expression plasmid. Cells extracts were prepared 40 h later and assayed for luciferase and β-galactosidase activities (Applied Biosystems). Reporter activities were normalized for transfection efficiency and are presented as the average with standard deviation of triplicates in one experiment representative of all other (two to five) independent experiments. Western blots of cell extracts were hybridized with anti-FLAG antibody as reported (25).

Mice

Mice were used according to federal guidelines and as approved by the Cleveland Clinic Institutional Animal Care and Use Committee. Transgenic founders were generated by DNA injection into fertilized mouse eggs. Founders and progeny were identified by *lacZ* polymerase chain reaction (PCR). *Sox9*-null embryos were generated using *Sox9* conditional null alleles (27), as published (28). Data were reproduced with at least two pairs of control and tester littermates. $SOM^{-/-}$ mice were generated as described (Supplementary Figure S1). Acute pancreatitis was induced in 7–13-week-old mice by intraperitoneal injection of 100 µg/kg body weight of cerulein (Sigma) eight times at hourly intervals for 2 days (29). The second day of injection was defined as day 0. Buprenorphine (Reckitt Benckiser Healthcare) was injected subcutaneously at 30 µg/ml (100 µl) as an analgesic twice a day for 3 days during and after cerulein injections.

X-gal staining and histology analysis

Standard protocols were used to stain whole embryos with X-gal, make frozen sections of paraformaldehyde-fixed embryos and tissues, stain sections with X-gal and nuclear fast red or with hematoxylin and eosin and process sections for immunostaining and counterstaining with 4',6-diamidino-2-phenylindole (DAPI). Rabbit Sox9 antibody (1:500; Millipore AB5535) was detected with Alexa Fluor® 594-conjugated goat anti-rabbit antibody (Invitrogen); goat Sox10 antibody (1:200; Santa Cruz Biotechnology, sc-17342) with Alexa Fluor® 488-conjugated donkey anti-goat antibody (Invitrogen); guinea pig insulin antibody (1:500; Dako) with Texas red-conjugated donkey anti-guinea pig antibody (1:100; Jackson ImmunoResearch) and mouse glucagon antibody (1:500; Sigma) with Cy3-conjugated donkey anti-mouse antibody (1:100; Jackson ImmunoResearch). Data were visualized with Leica DM2500 microscope, captured with Qimaging Micropublisher 5.0 RTV digital camera, and processed with Adobe Photoshop 7.0 software. Pancreas parameters, including percentages of Sox9-positive and EGFP-positive cells, were determined using NIH ImageJ software. Unpaired two-tailed Student's *t*-tests were used to assess statistical significance.

RNA assays

Samples were isolated in RNAlater (Invitrogen), transferred to TRIzol (Invitrogen) and homogenized (Power Gen 125; Fisher Scientific). Total RNA was further purified using chloroform and Rneasy Mini Kit (Qiagen), treated with Rnase-Free Dnase (Qiagen) and reverse transcribed (SuperScript III First Strand, Invitrogen). Real-time quantitative reverse transcription (qRT)-PCR was performed using SYBR Green and StepOne Plus (Applied Biosystems). Primers are listed (Supplementary Table S1B). Data were calculated by the delta-delta-Ct method using *Gapdh* or *Actb* values as references. Unpaired, two-tailed Student's *t*-tests were used to assess statistical significance.

Chromatin immunoprecipitation and electrophoretic mobility shift assay

Chromatin immunoprecipitation (ChIP) was performed as described (30) using RCS cells, rabbit non-immune IgG (Sigma) and Sox9 antibody (Millipore AB5535). Electrophoretic mobility shift assay (EMSA) and SOX9-expressing COS1 cell extracts were prepared as described (25). ChIP primers and EMSA probes are listed (Supplementary Table S1C and D).

RESULTS

Identification of a *Sox9* enhancer at −70 kb

The goal of this study was to delineate the enhancer(s) that activate a −70/0-kb BAC transgene in *Sox9*-expressing somatic tissues in mouse embryos (21). We hypothesized that these enhancers have been evolutionarily preserved, and we therefore used the UCSC genome browser to identify regions conserved over >500 bp in vertebrate genomes and exhibiting logarithm-of-the-odds (LOD) scores >200. We found such regions at −14 (*TES*), −19, −64 and −70 kb (Figure 1A). To test these regions for regulatory function, we cloned 1 copy or 4 tandem copies 5′ of the mouse *Sox9* −307/+364-bp sequence in a *βGeo*-IRES-*EGFP* reporter (Figure 1B). We transiently transfected the reporters in RCS cells, which highly express *Sox9*, and in mouse mesenchymal 10T1/2 cells, which weakly express *Sox9* (24). The *Sox9* promoter-only reporter was minimally active in both cell lines, but was robustly activated in RCS cells by a cartilage-specific *Col2a1* enhancer used as a positive control (24,31) (Figure 1C). The −14, −19 and −64 kb elements displayed weak if any activity in either cell line, even as tetramers. Interestingly, the −70 kb element activated the *Sox9* promoter 10-fold as a monomer and 70-fold as a tetramer in RCS cells, but was inactive in 10T1/2 cells (Figure 1C and D). This element is conserved from human to lizard and its location varies between −18 and −102 kb (Supplementary Figure S2A). In addition, a 44-bp sequence located 270 kb 5′ of the zebrafish *Sox9a* gene shows 73% of identity with the enhancer segment that is most conserved in higher vertebrates (Supplementary Figure S2B). ENCODE/LICR high-throughput ChIP data showed histone modifications indicative of enhancer activity at −70 kb in mouse limb and brain at embryonic day 14.5 (E14.5) (Supplementary Figure S3). We thus concluded that the −70 kb element could be a potent enhancer of *Sox9*.

The −70-kb enhancer activates a *Sox9* transgene in most *Sox9*-expressing somatic tissues

As a first test of the activity of the −70-kb element *in vivo*, we generated transgenic mice with the promoter-only and 4-copy-enhancer reporters. When stained with X-gal (β-galactosidase assay), E14.5 founder embryos harboring the promoter-only reporter showed no transgene activity in any tissue or showed weak activity in neuronal tissue (data not shown), as reported (22). In contrast, three E14.5 founder embryos (out of seven) and one mouse line (out of one) carrying the 4-copy-enhancer reporter showed identical transgene expression patterns (data not shown). We therefore analyzed in depth progeny from the mouse line. X-gal staining of E9.5 embryos revealed transgene activity in characteristic sites of *Sox9* expression, including otic vesicles, notochord,

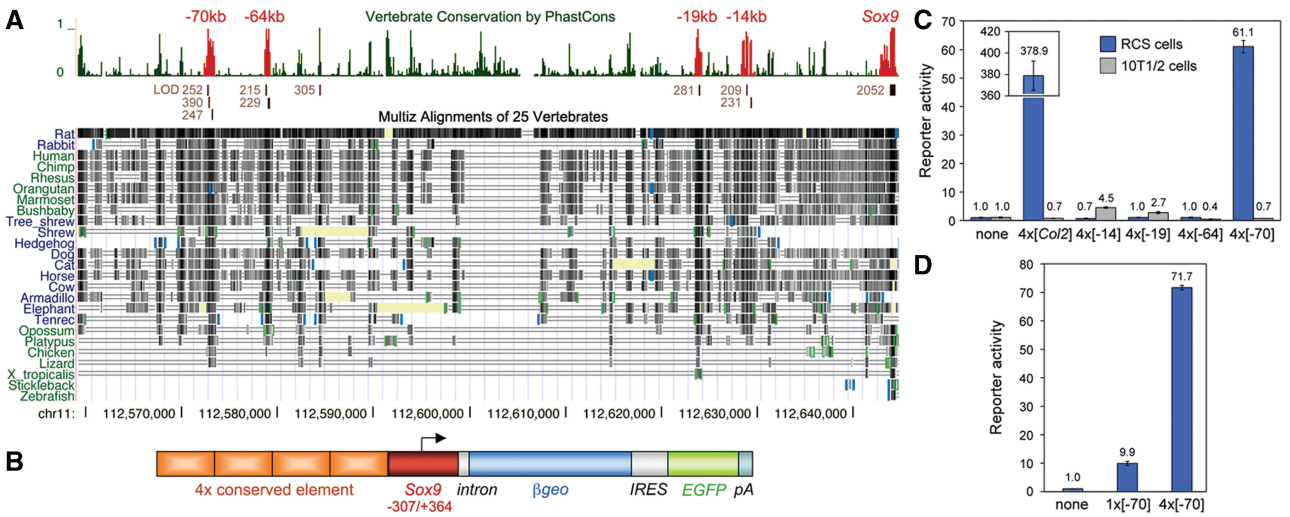

Figure 1. Identification of a *Sox9* enhancer. (**A**) UCSC genome browser analysis of the 85-kb sequence upstream of mouse *Sox9* and first 2 kb of transcribed sequence (chr11:112 559 200–112 645 199). The upper schematic shows peaks of conservation in vertebrate genomes relative to the mouse genome. The most conserved regions are highlighted in red, and segments with a LOD score >200 in brown. The lower schematic is an alignment of conserved regions (black boxes) in 25 genomes. (**B**) Schematic of *Sox9* transgenes. Up to 4 tandem copies of any conserved element were cloned 5′ of the *Sox9* proximal promoter and 5′ untranslated region (−307/+364 bp). The latter was linked to an exogenous intron, a reporter encoding a fusion protein of *Escherichia coli* β-galactosidase and neomycin resistance *βgeo*), an *IRES*, the *EGFP* and a bovine polyadenylation site (*pA*). (**C**) Activity of *Sox9* reporters in transiently transfected RCS and 10T1/2 cells. Reporters contained no upstream element (none) or 4 copies of a 48-bp *Col2a1* enhancer or *Sox9* conserved element. Normalized reporter activities are presented relative to the activity of the *Sox9* promoter-only reporter. (**D**) Activity of reporters harboring 0, 1 or 4 copies of the −70-kb element in RCS cells.

neural crest, foregut and head and branchial arch mesenchyme (Figure 2A). At E14.5, both X-gal staining and EGFP fluorescence also revealed transgene activity in many sites of *Sox9* expression, namely cartilage, pancreas, kidney, gut, lung, choroid plexus, hypothalamus, olfactory epithelium and salivary glands (Figure 2B and C). The transgene was inactive in heart valves and testis, despite expression of *Sox9*, and it was active in dorsal root ganglia, which express *Sox9*'s closest relative, *Sox10*, but not *Sox9*. Three-week-old pups expressed the transgene in the same tissues as embryos, as well as in liver bile ducts, which express *Sox9* (Figure 2D). Hence, the −70-kb element is able to activate a *Sox9* promoter transgene in most *Sox9*-positive somatic tissues. We therefore named this element *SOM* and the transgene *TgSOM*.

SOM enhances endogenous *Sox9* expression in somatic tissues

We next used DNA homologous recombination to delete *SOM* from the mouse genome and assess its contribution to *Sox9* expression (Supplementary Figure S1A and C). Although $SOM^{+/-}$ and $SOM^{-/-}$ mice were viable and developed normally (Supplementary Figure S1D), a significant reduction in the *Sox9* RNA level was detected in the pancreas, lung, salivary gland, kidney, gut and liver of $SOM^{-/-}$ fetuses (E14.5), newborns (postnatal day 0, P0) and weanlings (P23), but not in adults (P90) (Figure 3). The reduction range was 18–37%, with a mean of 23%. A similar reduction was observed in fetal brain and adult cartilage, but none in heart and testis. Thus, *SOM* enhances *Sox9* expression in selective somatic tissues, primarily during development.

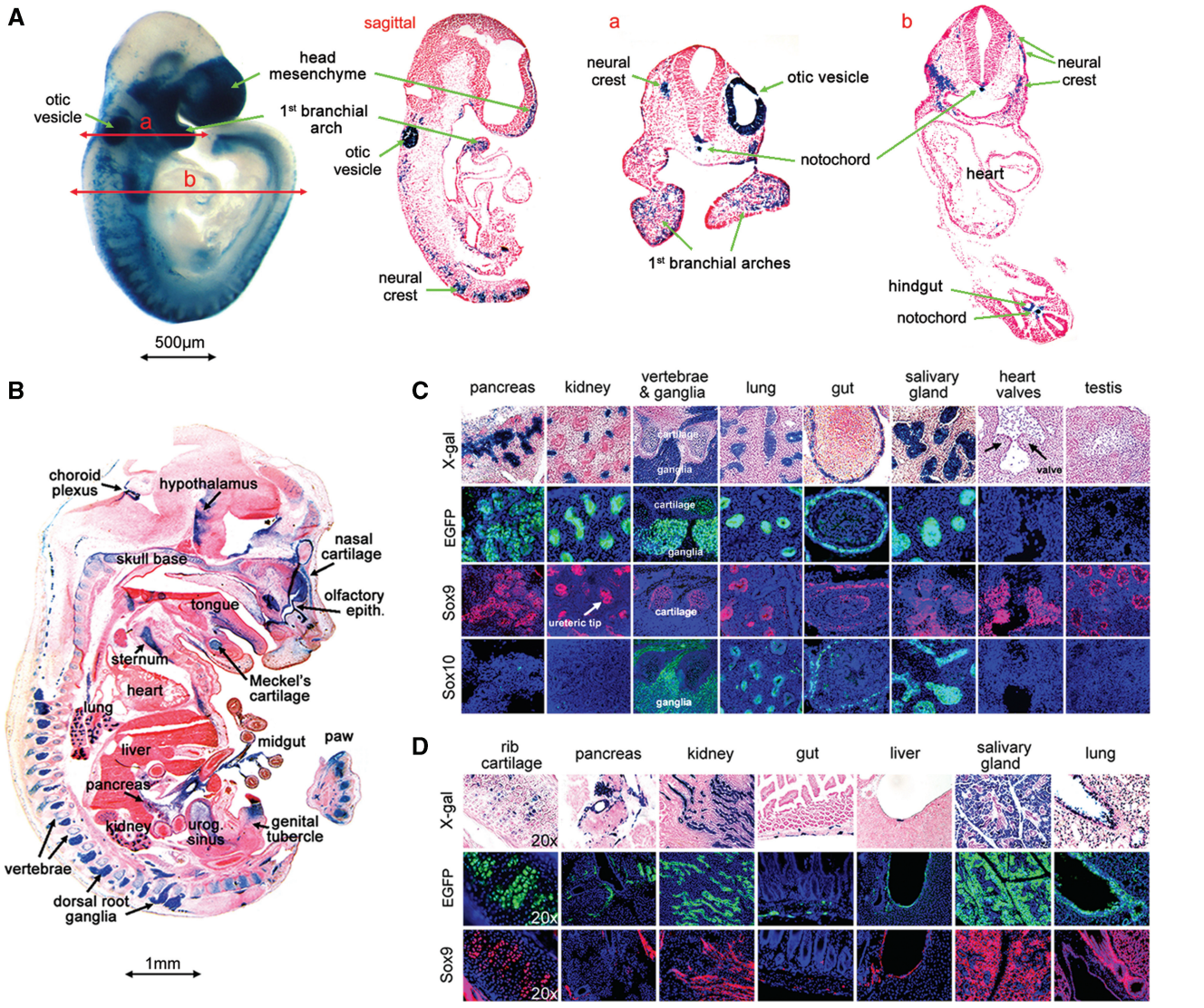

Figure 2. Analysis of a *TgSOM* mouse line. (**A**) E9.5 *TgSOM* embryo stained with X-gal (blue). Left, side view of whole-mount-stained embryo. Double red arrows indicate the level of transverse sections (a and b) shown in the same panel, next to a sagittal section. Sections are counterstained with nuclear fast red. (**B**) X-gal staining of a mid-sagittal section through an E14.5 embryo. (**C**) High-magnification pictures of tissue sections from the same embryo stained with X-gal staining, analyzed for EGFP fluorescence and immunostained for Sox9 and Sox10. (**D**) *TgSOM* activity and Sox9 protein expression in P23 mouse tissues.

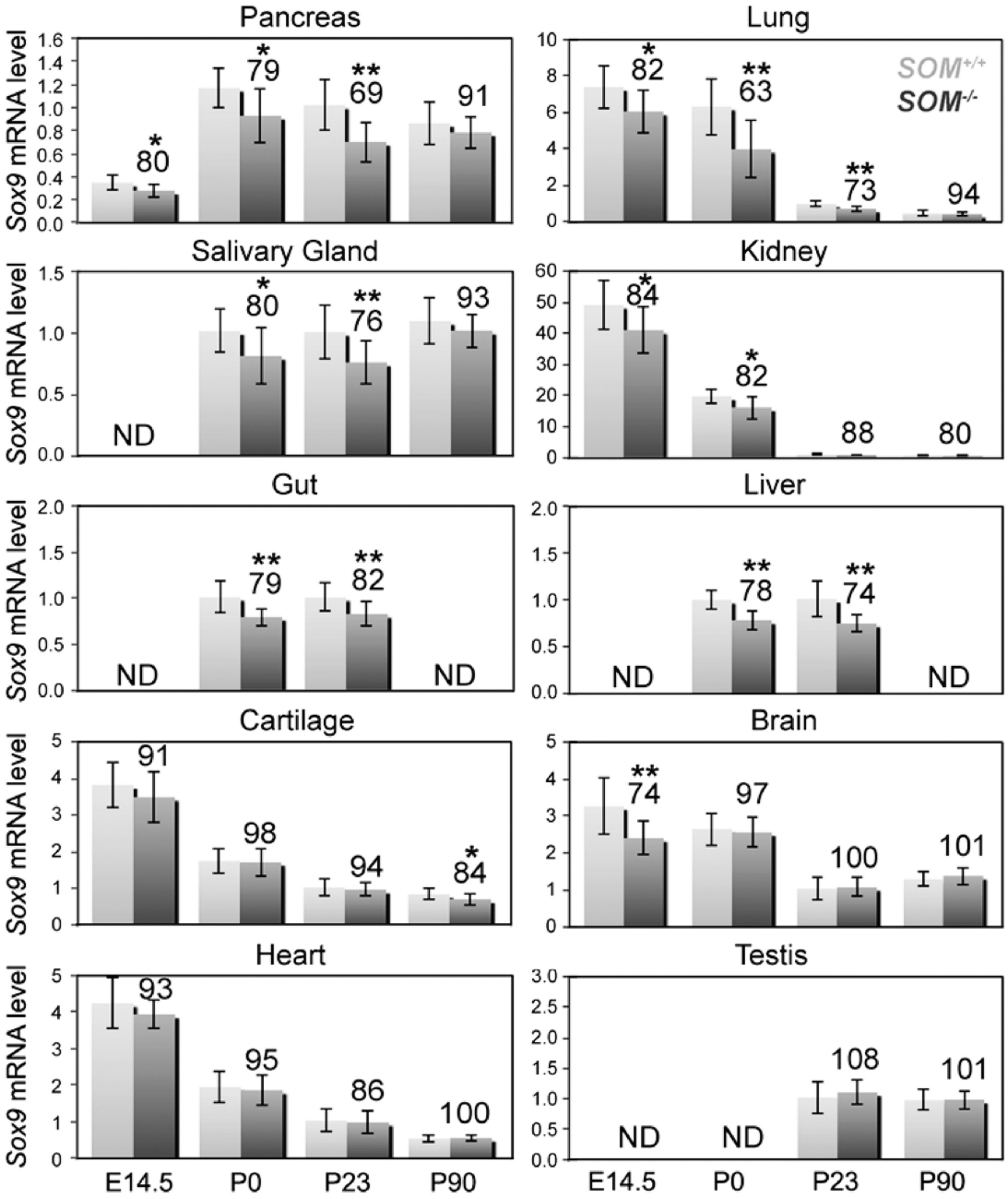

Figure 3. *Sox9* RNA levels in $SOM^{+/+}$ and $SOM^{-/-}$ mice. Various tissues were collected at E14.5 ($n = 13$ and 10, respectively), P0 ($n = 7$ and 11, respectively), P23 ($n = 9$ and 12, respectively) and P90 ($n = 14$ and 14, respectively). *Sox9* RNA levels were measured by qRT-PCR and normalized to *Gapdh* RNA levels. For each tissue, data are presented as average with standard deviation relative to the value obtained in P23 $SOM^{+/+}$ mice. The percentage of *Sox9* RNA level in mutants compared with controls is indicated for each type of sample. $*P < 0.05$; $**P < 0.01$. ND, not determined.

SOM is necessary for normal pancreas development

Like $Sox9^{fl/+}Pdx1Cre$ newborns, which are *Sox9* haploinsufficient in the pancreas (32), $SOM^{-/-}$ weanlings exhibited half-size pancreas endocrine islets, but no change in the pancreas' overall size and number of islets (Figure 4A and B). Accordingly, the insulin and glucagon RNA levels were reduced by 15–20% at P0 and P23 (Figure 4C). The RNA levels of *Pdx1* and *Ngn3*, which encode transcription factors essential for pancreas development downstream of *Sox9*, were similarly reduced in $SOM^{-/-}$ fetuses and newborns, whereas the RNA level of *Hnf6*, which controls pancreas development upstream of *Sox9*, was unchanged (Figure 4D). Thus, *SOM* is necessary for normal pancreas development.

SOM is necessary for prompt recovery from acute pancreatitis

To determine the importance of *SOM* in adult tissue regeneration, we administered $SOM^{+/+}$ and $SOM^{-/-}$ mice with cerulein, a cholecystokinin-like oligopeptide.

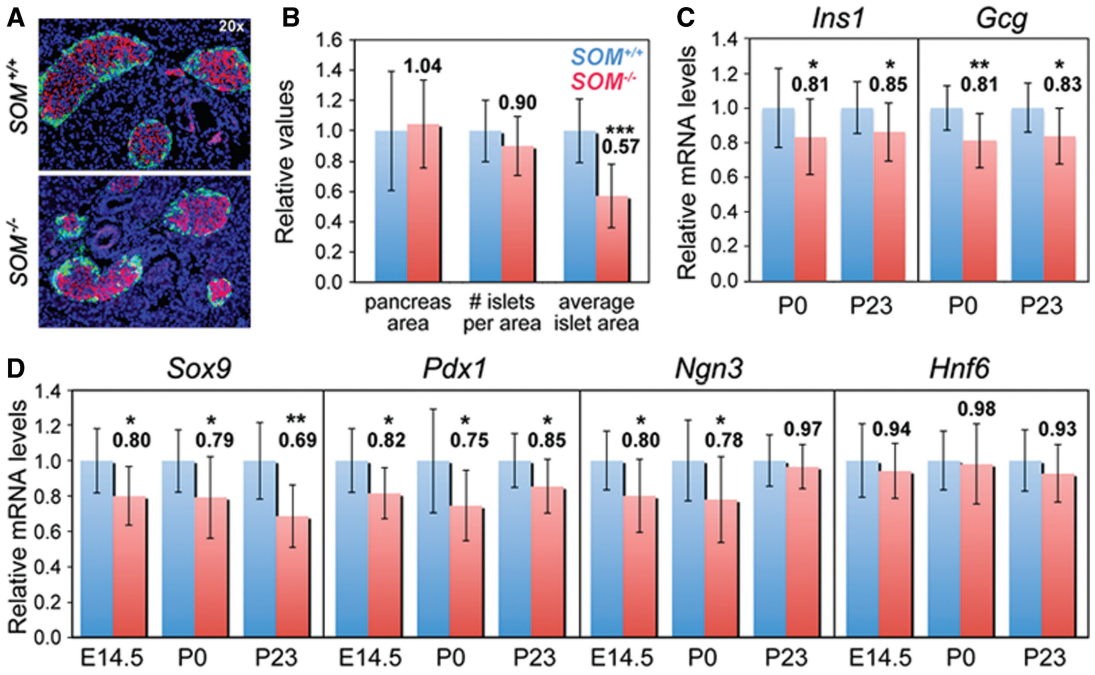

Figure 4. Pancreas developmental defects in $SOM^{-/-}$ pups. (**A**) Pancreas sections from P23 $SOM^{+/+}$ and $SOM^{-/-}$ mice were immunostained for insulin (red) and glucagon (green) to mark endocrine islets and counterstained with DAPI to identify cell nuclei (blue). (**B**) Number and size of endocrine islets in P23 $SOM^{-/-}$ mice relative to controls. Data are presented as the average with standard deviation of values obtained for four non-adjacent tissue sections in each of four mice per genotype. Mutant values are indicated. (**C**) Relative levels of insulin (*Ins1*) and glucagon (*Gcg*) messenger RNA (mRNAs) at P0 ($n = 7$ and 10, respectively) and P23 ($n = 9$ and 12, respectively). (**D**) Relative mRNA levels of pancreas regulatory genes at E14.5 ($n = 13$ and 10, respectively), P0 ($n = 7$ and 10, respectively) and P23 ($n = 9$ and 12, respectively). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

As expected (29), both types of mice developed severe pancreatitis, by the end of a 2-day regimen of drug injections, as visible by loss of H&E staining in many acini (Figure 5A). Control tissues were already recovering one day later (day 1) and appeared normal at day 7, whereas mutant tissues were still severely damaged at day 1 and still recovering at day 7. Control pancreas increased its percentage of Sox9-positive cells by >7-fold by the end of injections (day 0) and returned to baseline by day 7 (Figure 5B and C). *TgSOM* mice subjected to the same treatment exhibited a 5- to 20-fold increase in transgene activity during this time period, suggesting a role for *SOM* in generating this effect (Supplementary Figure S4A–C). Proving this role, $SOM^{-/-}$ mice increased their percentage of Sox9-positive cells in pancreas only half as much as control littermates by day 0, and maintained this reduced proportion stationary for 1 week (Figure 5B and C). Control pancreas raised its global *Sox9* RNA level by 5.0-fold by day 1 and returned to normalcy by day 7, whereas $SOM^{-/-}$ tissue raised this level by only 2.7-fold by day 1 and kept it stable through day 7 (Figure 5C). The fact that the percentage of Sox9-positive cells in control and mutant tissues was already maximal at day 0, whereas the global levels of *Sox9* RNA peaked only at day 1, suggests that most cells initiated *Sox9* expression by day 0 and then continued to accumulate Sox9 RNA and protein through day 1. The tempered upregulation of *Sox9* expression in $SOM^{-/-}$ pancreas was functionally consequential, as *Pdx1* and *Ngn3* were also mildly upregulated, and expression of the acinar tissue differentiation marker *Amy1* (amylase 1) was more slowly recovered. Thus, *SOM* is necessary to forcefully upregulate *Sox9* expression upon pancreas injury, and thereby allow timely regeneration of the tissue.

Sox9 is necessary and sufficient for *TgSOM* activity

The prevalent activity of *TgSOM* in somatic tissues that share few similarities beside *Sox9* or *Sox10* expression raised the possibility that *SOM* could be under the control of the Sox9 and Sox10 proteins themselves. Supporting this hypothesis, forced expression of human SOX9 protein in 10T1/2 cells led to 10- and 30-fold activation of the 1- and 4-copy-*SOM* reporters, respectively (Figure 6A). This effect was as robust as for the *Col2a1* enhancer, a known target of SOX9 (24,33). Sox10 was half as potent as SOX9, whereas Sox8 (a SoxE protein, like Sox9 and Sox10), Sox2 (SoxB) and Sox11 (SoxC) showed little if any activity compared with SOX9 and Sox10 (Figure 6B). Thus, Sox9 and, to a lesser degree, Sox10 have the specific ability and are sufficient to activate *SOM*. We next generated $Sox9^{+/+}$ and $Sox9^{-/-}$ littermates carrying *TgSOM* and analyzed them at E11.5, just before $Sox9^{-/-}$ embryos die (28), to test whether Sox9 is necessary to activate *SOM in vivo*. Unlike $Sox9^{+/+}$ embryos, $Sox9^{-/-}$ embryos were unable to activate *TgSOM* in cartilage primordia, pancreas and kidney, which normally express *Sox9* but not *Sox10* (Figure 6C). They also lost the ability to express *TgSOM* in the lung, likely because this tissue no longer expresses *Sox10* in the absence of

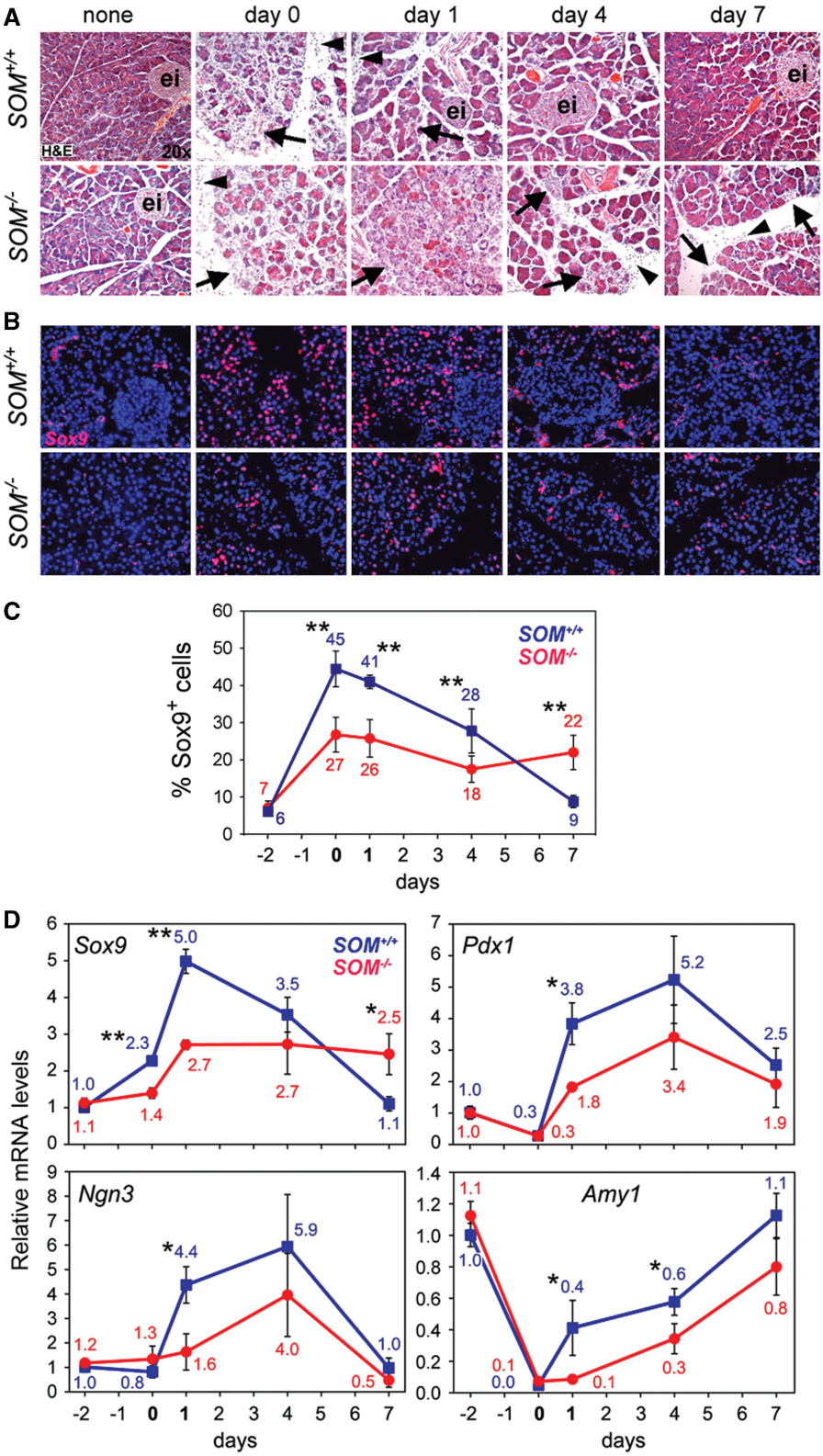

Figure 5. Pancreas regeneration defects in $SOM^{-/-}$ adult mice. (**A**) Pancreas sections from $SOM^{+/+}$ and $SOM^{-/-}$ mice untreated (none) or treated with cerulein and analyzed at days 0, 1, 4 and 7. Dark red H&E staining identifies healthy exocrine acini. Arrows, severely damaged acini. Ei, endocrine islets. Arrowheads, cell debris. (**B**) Sox9 immunostaining (red) in sections generated as in A and counterstained with DAPI (blue). (**C**) Quantification of the percentage of Sox9-positive cells in tissue sections similar to those shown in B. Data are presented as the average with standard deviation of values obtained for two to three non-adjacent tissue sections in each of three mice per genotype and time point. (**D**) qRT-PCR of *Sox9*, *Ngn3*, *Pdx1* and *Amy1* RNA levels ($n = 3$). $^{*}P < 0.05$; $^{**}P < 0.01$.

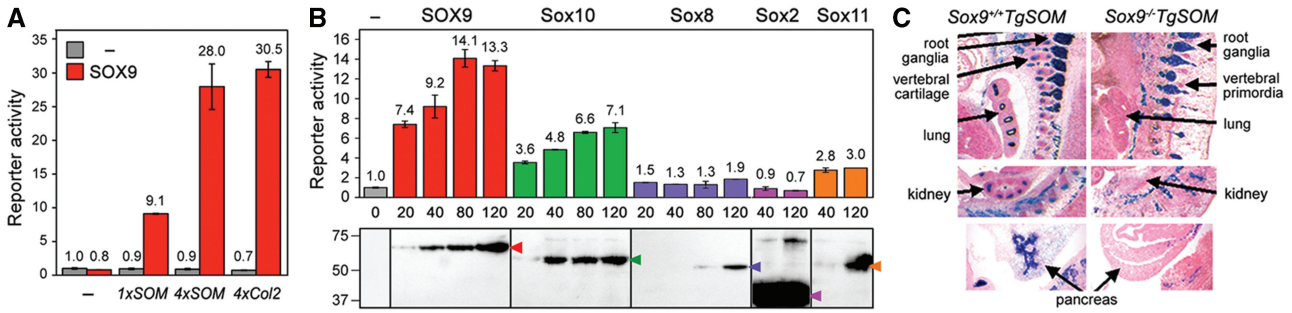

Figure 6. Sox9 is sufficient and necessary to activate *TgSOM*. (**A**) 10T1/2 cells were transfected with empty (grey bars) or SOX9 (red bars) expression plasmid along with *Sox9* reporters harboring no enhancer (−), 1 or 4 copies of *SOM* or 4 copies of *Col2a1* enhancer. Reporter activities are presented relative to the activity of the *Sox9* promoter-only reporter. (**B**) 10T1/2 cells were transfected with the 4-copy *SOM* reporter and 0–120 ng of FLAG-Sox expression plasmid. Top, relative reporter activities. Bottom, western blot of cell extracts hybridized with FLAG antibody. The migration level and Mr of protein standards are indicated. Arrowheads, Sox proteins. (**C**) X-gal-stained sections from E11.5 $Sox9^{+/+}TgSOM$ and $Sox9^{-/-}TgSOM$ littermates.

Sox9 (34). In contrast, they maintained *TgSOM* activity in dorsal root ganglia, a tissue that strongly expresses *Sox10*. By demonstrating that Sox9 and Sox10 are sufficient to activate a *SOM*-driven reporter and that Sox9 is necessary for this activation in multiple tissues, these data suggest that the proteins may similarly mediate the activity of the endogenous *SOM* enhancer.

Sox9 directly activates *SOM*

Sox9 binds its known genomic targets at sites matching the Sox consensus sequence $C^A/_TTTG^A/_T{}^A/_T$ or presenting 1 or 2 nucleotide variations in this sequence (26,35,36). It binds single sites as a monomer in Sertoli cells, and binds site pairs as a homodimer in somatic tissues, provided that these pairs are oriented head to head and separated by 3–5 bp. Applying these parameters, we found 32 evolutionarily conserved putative binding sites for Sox9, including seven pairs, in the *SOM* sequence (Figure 7A and Supplementary Figure S5). ChIP revealed that Sox9 specifically binds to *SOM* in RCS cells (Figure 7B), and EMSAs showed that SOX9 is able to bind most predicted sites *in vitro*, but prefers pairs (Figure 7C and Supplementary Figure S6A). The mutation of any Sox9-dimer site reduced *SOM* activity by 33–92% in RCS cells and SOX9-expressing 10T1/2 cells (Figure 7D and Supplementary Figure S6B). The amplitude of these effects was larger than expected for additive contributions of the sites, strongly suggesting site cooperativity. Finally, the combined mutation of multiple pairs abrogated enhancer activity *in vitro* as well as in transgenic embryos (Figure 7D and E). We thus concluded that Sox9 dimers directly bind to multiple sites in the *SOM* sequence and thereby cooperatively activate the enhancer.

DISCUSSION

This study newly identified an enhancer of the *Sox9* gene and demonstrated its underlying mechanisms and roles *in vivo*. The enhancer is located 70 kb 5′ of *Sox9* and is highly conserved in vertebrates. It is active in multiple somatic tissues and we therefore call it *SOM*. It is directly targeted by the Sox9 protein, which thereby amplifies its own gene expression to benefit the development and adult regeneration of at least one organ, the pancreas.

SOX9 was proposed to be critically controlled by remote *cis*-elements almost 20 years ago when alterations in its surrounding 2-Mb gene desert were found to cause CD and associated malformations (1–3). Subsequent transgenic studies lent support to this concept by identifying putative tissue-specific enhancers across the gene desert, but the actual roles of these enhancers *in vivo* remain untested. The present study thus represents a departure from previous studies by newly identifying *SOM* and by demonstrating that this enhancer quantitatively contributes to *Sox9* expression in multiple somatic tissues during development and adult tissue regeneration. The fact that *SOM* is not absolutely required for *Sox9* expression supports the notion that other enhancers also participate in *Sox9* transcription. *Sox9* is not unique in this regard, as many other genes also are controlled by multiple, often remote, enhancers (37). The −1-Mb limb-specific enhancer of the Sonic Hedgehog (*Shh*) gene works solo (38), but multiple enhancers in the >800-kb-wide transcriptional archipelago of the *HoxD* cluster work in concert (39). Some enhancers have qualitative roles, conferring gene-specific spatiotemporal transcription, while others have quantitative roles, conferring robustness to transcription. Further studies are thus necessary to reveal the roles of putative *Sox9* enhancers previously reported, unearth more enhancers and demonstrate how these enhancers functionally interact with one another and with *SOM*.

SOM features multiple evolutionarily conserved sequences that mediate its cooperative activation by Sox9 dimers. This cooperative mechanism fits with the observation that *SOM* makes a difference only when *Sox9* expression is high, in development and following injury, thus presumably when enough Sox9 molecules are present to cooperatively activate *SOM*. It also fits with the dimeric mode of action of Sox9 that has been demonstrated in several somatic tissues, and the fact that mutations in the SOX9 dimerization domain cause CD (36,40). In contrast, these mutations do not cause XY sex reversal, and Sox9 works as a monomer in the gonad, including in

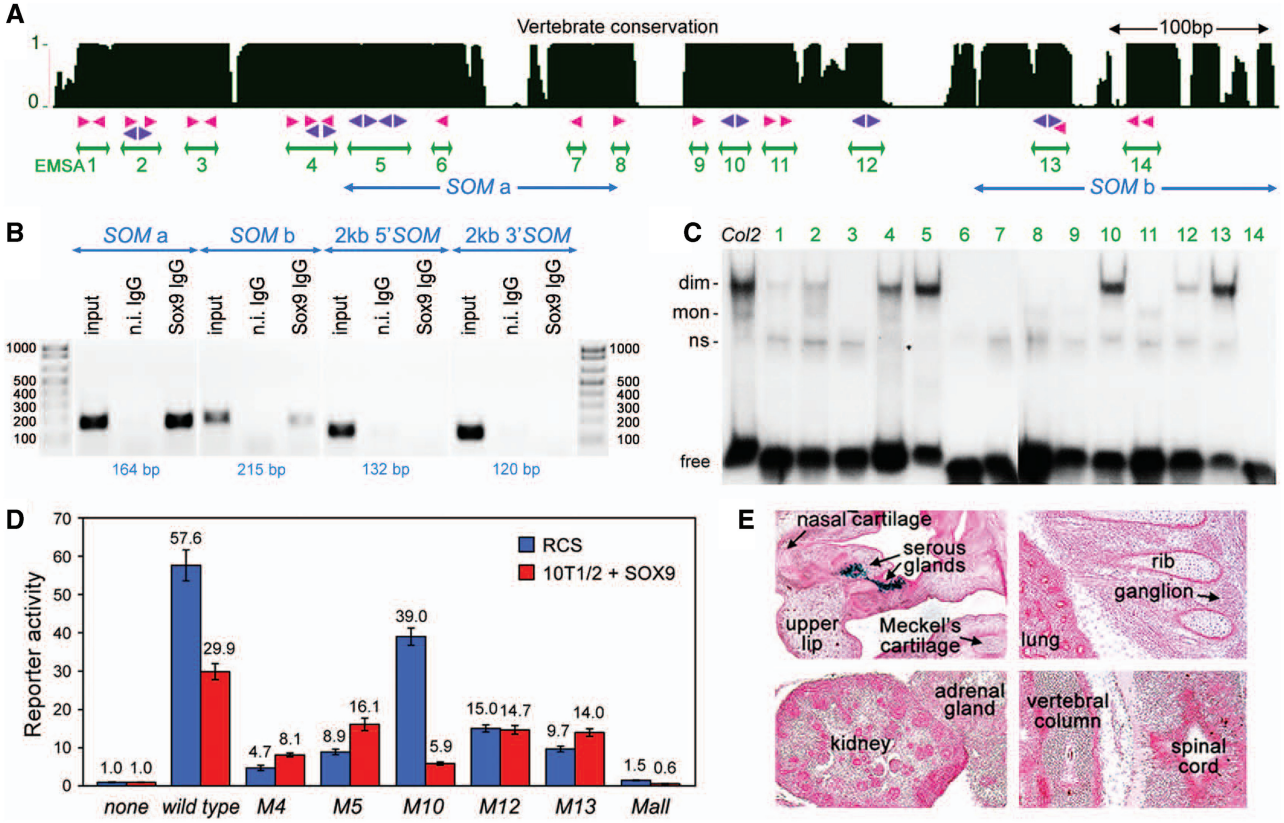

Figure 7. Sox9 directly activates *SOM*. (**A**) Identification of putative Sox binding sites in *SOM*. Underneath the UCSC vertebrate conservation plot of *SOM*, pink and purple arrowheads represent predicted sites for Sox9 monomers and dimers, respectively; green double arrows, EMSA probes and blue double arrows, *SOM* regions amplified by PCR in ChIP assay. (**B**) ChIP assay of Sox9 binding to *SOM* in RCS cells. The *SOM* a and b regions, and regions located 2-kb upstream or downstream of *SOM* were amplified by PCR using input material, chromatin precipitated with non-immune (n.i.) IgG or chromatin precipitated with Sox9 antibodies. PCR products are visualized following resolution in electrophoresis gels. Sox9 antibodies precipitated chromatin containing the *SOM* a and b regions, but not sequences flanking *SOM*, demonstrating specific binding of Sox9 to the enhancer. DNA markers and their length (in bp) are shown on either side of the gel. (**C**) EMSA of SOX9-expressing COS1 cell extracts incubated with probes 1–14. The migration levels of free probe and probe bound by SOX9 dimers (dim), SOX9 monomers (mon) and non-specific protein (ns) are shown. (**D**) Activity of 4-copy *SOM* reporters mutated (M) in the Sox9 paired sites 4–13 or in all these sites. (**E**) Activity of *TgSOM* mutated in the Sox9 binding pairs 4–13 in an E14.5 transgenic founder embryo. Ectopic X-gal staining was seen in serous glands in one of five transgenic embryos, but no staining was seen anywhere else in any embryo.

TES activation (21). Thus, although Sox9 may control its own expression in the gonad, as it does in somatic tissues, it may use a distinct enhancer and mode of action to achieve this function in each tissue type. By proving that Sox9 regulates its own gene expression, our study allows to more definitively identify Sox9 as a master transcription factor, i.e. a factor that crucially contributes to determining cell fate and that ensures this function at least in part by sustaining its own expression. Other factors of this caliber include the stem cell factor Sox2 and myogenic factor MyoD (41,42). It may also include Sox10, as this factor was recently proposed to orchestrate the activity of a neural crest-specific enhancer in the vicinity of its gene (43). Interestingly, Sox9 can activate this *Sox10* enhancer, and Sox10 can activate *SOM*. Because Sox9 and Sox10 are co-expressed in several cell types (9), it is thus possible that these proteins upregulate each other's gene in addition to their own gene.

While *TgSOM* activity matched *Sox9* expression and the consequences of *SOM* deletion in many somatic tissues, some discrepancies were observed: *TgSOM* was

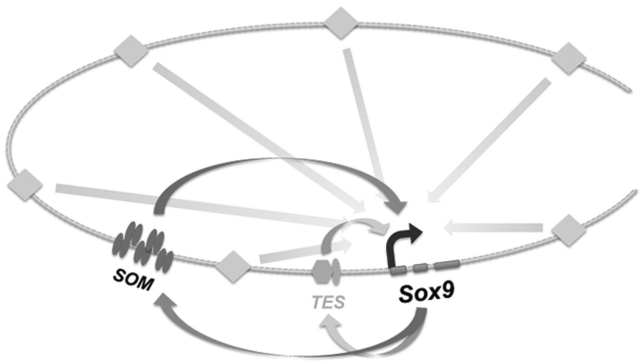

Figure 8. Model of *Sox9* transcriptional regulation. The 2-Mb *Sox9* locus is shown as a loop; the three *Sox9* exons as dark grey blocks and transcription as an angled arrow. *SOM* is bound by Sox9 dimers (paired ovals), which upregulate *Sox9* expression in a positive feedback loop (large, curved arrows) in somatic tissues. *TES* is bound by a Sox9 monomer (single oval), which cooperates with Sf1 (hexagon) to maintain *Sox9* expression in testis (small, curved arrows). Other putative enhancers are bound by transcription factors (diamonds) that interact with the basal transcriptional machinery (straight arrows) to contribute to *Sox9* expression in various tissues.

not active and *SOM* deletion had no effect in heart valves, despite expression of *Sox9*; *TgSOM* was active at all ages in brain and cartilage, but *SOM* deletion affected *Sox9* expression only in fetal brain and adult cartilage; *TgSOM* was active in tissues that express *Sox10* but not *Sox9*, such as dorsal root ganglia. One reason could be that Sox9 molecules are rare or kept inactive in heart valves, precluding any *SOM* activity. An explanation for cartilage, brain and dorsal root ganglia could be that the threshold of Sox9 or Sox10 protein needed to activate *SOM* is lower in *TgSOM* than in the *Sox9* locus because *SOM* is present as 4 copies directly linked to the *Sox9* promoter in *TgSOM*, but present as a single copy and 70 kb away from the *Sox9* promoter in the endogenous locus. Moreover, the number of copies and the site of integration of *TgSOM* in the mouse genome may also facilitate transgene activity. Finally, the lack of consequences of *SOM* deletion in cartilage and brain is likely due to other enhancers exerting compensatory effects.

Although *SOM* deletion had no major impact on mice under standard conditions, RNA assays revealed decreases from 18 to 37% in *Sox9* expression in several tissues during development. While such decreases would likely be inconsequential for many genes, we reasoned that they could be consequential for *Sox9*, as *SOX9* haploinsufficiency causes CD and associated malformations. Accordingly, pancreas development was as affected in $SOM^{-/-}$ mice as it is in mice lacking one *Sox9* allele in pancreas progenitors (32,44). Moreover, $SOM^{-/-}$ mice showed a 2-fold impairment in *Sox9* upregulation when subjected to acute pancreatitis, and recovered about twice as slowly as control mice. Based on these data, the importance of *Sox9* dosage and the evolutionary conservation of *SOM*, we predict that *SOM* may have additional roles in development, adult physiology and such disease settings as tissue repair, fibrosis and cancer, but that these roles may become apparent only upon co-deletion of co-acting enhancers. In addition, our data raise the attracting possibility that *SOM* be used as a new genetic tool for various types of scientific and clinical applications. It could be used for instance to amplify *SOX9* expression in CD and other diseases due to *SOX9* deficiency or to drive expression of tumor suppressors in *SOX9*-positive cancers.

In conclusion, the findings from this study and previous studies suggest a model whereby *Sox9* expression is controlled by multiple enhancers spread over a large genomic region (Figure 8). *SOM*, located at −70 kb, enhances *Sox9* expression in developing and regenerating somatic tissues, and thereby allows *Sox9* to properly achieve its functions. *TES*, located at −14 kb, may drive *Sox9* expression in the undifferentiated gonad and testis. Both *SOM* and *TES* mediate a positive feedback loop of Sox9 auto-regulation. *TES* binds Sox9 as a monomer together with its partner Sf1, whereas *SOM* binds multiple, cooperatively acting, Sox9 dimers. Additional enhancers, yet to identified and characterized, must contribute along with *TES* and *SOM* to achieve qualitative and quantitative expression of *Sox9* in multiple cell types.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1–6.

ACKNOWLEDGEMENTS

We thank R.A. Conlon and the Case University Transgenic Facility for generating *TgSOM* and *SOM*-null mice, G. Scherer for providing $Sox9^{fl/fl}$ mice and A. Hulin for assistance with ChIP assay.

FUNDING

National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases [AR46249 and AR60016 to V.L.]; Morgenthaler (to T.J.M.); Arthritis Foundation postdoctoral fellowships (to T.J.M. and P.B.). Funding for open access charge: Cleveland Clinic.

Conflict of interest statement. None declared.

REFERENCES

1. Foster,J.W., Dominguez-Steglich,M.A., Guioli,S., Kwok,C., Weller,P.A., Stevanovic,M., Weissenbach,J., Mansour,S., Young,I.D., Goodfellow,P.N. *et al.* (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*, **372**, 525–530.
2. Wagner,T., Wirth,J., Meyer,J., Zabel,B., Held,M., Zimmer,J., Pasantes,J., Bricarelli,F.D., Keutel,J., Hustert,E. *et al.* (1994) Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell*, **79**, 1111–1120.
3. Gordon,C.T., Tan,T.Y., Benko,S., Fitzpatrick,D., Lyonnet,S. and Farlie,P.G. (2009) Long-range regulation at the SOX9 locus in development and disease. *J. Med. Genet.*, **46**, 649–656.
4. Barrionuevo,F. and Scherer,G. (2010) SOX E genes: SOX9 and SOX8 in mammalian testis development. *Int. J. Biochem. Cell Biol.*, **42**, 433–436.
5. Jakob,S. and Lovell-Badge,R. (2011) Sex determination and the control of Sox9 expression in mammals. *FEBS J.*, **278**, 1002–1009.
6. Mansour,S., Offiah,A.C., McDowall,S., Sim,P., Tolmie,J. and Hall,C. (2002) The phenotype of survivors of campomelic dysplasia. *J. Med. Genet.*, **39**, 597–602.
7. Akiyama,H. and Lefebvre,V. (2011) Unraveling the transcriptional regulatory machinery in chondrogenesis. *J. Bone Miner. Metab.*, **29**, 390–395.
8. Haldin,C.E. and LaBonne,C. (2010) SoxE factors as multifunctional neural crest regulatory factors. *Int. J. Biochem. Cell Biol.*, **42**, 441–444.
9. Stolt,C.C. and Wegner,M. (2010) SoxE function in vertebrate nervous system development. *Int. J. Biochem. Cell Biol*, **42**, 437–440.
10. Harris,M.L., Baxter,L.L., Loftus,S.K. and Pavan,W.J. (2010) Sox proteins in melanocyte development and melanoma. *Pigment Cell Melanoma Res.*, **23**, 496–513.
11. Chakraborty,S., Combs,M.D. and Yutzey,K.E. (2010) Transcriptional regulation of heart valve progenitor cells. *Pediatr. Cardiol.*, **31**, 414–421.
12. Furuyama,K., Kawaguchi,Y., Akiyama,H., Horiguchi,M., Kodama,S., Kuhara,T., Hosokawa,S., Elbahrawy,A., Soeda,T., Koizumi,M. *et al.* (2011) Continuous cell supply from a Sox9-expressing progenitor zone in adult liver, exocrine pancreas and intestine. *Nat. Genet.*, **43**, 34–41.
13. Scott,C.E., Wynn,S.L., Sesay,A., Cruz,C., Cheung,M., Gomez Gaviro,M.V., Booth,S., Gao,B., Cheah,K.S., Lovell-Badge,R. *et al.* (2010) SOX9 induces and maintains neural stem cells. *Nat. Neurosci.*, **13**, 1181–1189.

14. Seymour,P.A., Freude,K.K., Tran,M.N., Mayes,E.E., Jensen,J., Kist,R., Scherer,G. and Sander,M. (2007) SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proc. Natl Acad. Sci. USA*, **104**, 1865–1870.
15. Huang,B., Wang,S., Ning,Y., Lamb,A.N. and Bartley,J. (1999) Autosomal XX sex reversal caused by duplication of SOX9. *Am. J. Med. Genet.*, **87**, 349–353.
16. Pritchett,J., Athwal,V., Roberts,N., Hanley,N.A. and Hanley,K.P. (2011) Understanding the role of SOX9 in acquired diseases: lessons from development. *Trends Mol. Med.*, **17**, 166–174.
17. Matheu,A., Collado,M., Wise,C., Manterola,L., Cekaite,L., Tye,A.J., Canamero,M., Bujanda,L., Schedl,A., Cheah,K.S. *et al.* (2012) Oncogenicity of the developmental transcription factor Sox9. *Cancer Res.*, **72**, 1301–1315.
18. Huang,Z., Hurley,P.J., Simons,B.W., Marchionni,L., Berman,D.M., Ross,A.E. and Schaeffer,E.M. (2012) Sox9 is required for prostate development and prostate cancer initiation. *Oncotarget*, **3**, 651–663.
19. Prevot,P.P., Simion,A., Grimont,A., Colletti,M., Khalaileh,A., Van den Steen,G., Sempoux,C., Xu,X., Roelants,V., Hald,J. *et al.* (2012) Role of the ductal transcription factors HNF6 and Sox9 in pancreatic acinar-to-ductal metaplasia. *Gut*, **61**, 1723–1732.
20. Wunderle,V.M., Critcher,R., Hastie,N., Goodfellow,P.N. and Schedl,A. (1998) Deletion of long-range regulatory elements upstream of SOX9 causes campomelic dysplasia. *Proc. Natl Acad. Sci. USA*, **95**, 10649–10654.
21. Sekido,R. and Lovell-Badge,R. (2008) Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. *Nature*, **453**, 930–934.
22. Bagheri-Fam,S., Barrionuevo,F., Dohrmann,U., Gunther,T., Schule,R., Kemler,R., Mallo,M., Kanzler,B. and Scherer,G. (2006) Long-range upstream and downstream enhancers control distinct subsets of the complex spatiotemporal Sox9 expression pattern. *Dev. Biol.*, **291**, 382–397.
23. Benko,S., Fantes,J.A., Amiel,J., Kleinjan,D.J., Thomas,S., Ramsay,J., Jamshidi,N., Essafi,A., Heaney,S., Gordon,C.T. *et al.* (2009) Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nat. Genet.*, **41**, 359–364.
24. Lefebvre,V., Huang,W., Harley,V.R., Goodfellow,P.N. and de Crombrugghe,B. (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Mol. Cell. Biol.*, **17**, 2336–2346.
25. Dy,P., Penzo-Mendez,A., Wang,H., Pedraza,C.E., Macklin,W.B. and Lefebvre,V. (2008) The three SoxC proteins—Sox4, Sox11 and Sox12—exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.*, **36**, 3101–3117.
26. Han,Y. and Lefebvre,V. (2008) L-Sox5 and Sox6 drive expression of the aggrecan gene in cartilage by securing binding of Sox9 to a far-upstream enhancer. *Mol. Cell Biol.*, **28**, 4999–5013.
27. Kist,R., Schrewe,H., Balling,R. and Scherer,G. (2002) Conditional inactivation of Sox9: a mouse model for campomelic dysplasia. *Genesis*, **32**, 121–123.
28. Akiyama,H., Chaboissier,M.C., Behringer,R.R., Rowitch,D.H., Schedl,A., Epstein,J.A. and de Crombrugghe,B. (2004) Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. *Proc. Natl Acad. Sci. USA*, **101**, 6502–6507.
29. Jensen,J.N., Cameron,E., Garay,M.V., Starkey,T.W., Gianani,R. and Jensen,J. (2005) Recapitulation of elements of embryonic development in adult mouse pancreatic regeneration. *Gastroenterology*, **128**, 728–741.
30. Dy,P., Wang,W., Bhattaram,P., Wang,Q., Wang,L., Ballock,R.T. and Lefebvre,V. (2012) Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev. Cell*, **22**, 597–609.
31. Lefebvre,V., Zhou,G., Mukhopadhyay,K., Smith,C.N., Zhang,Z., Eberspaecher,H., Zhou,X., Sinha,S., Maity,S.N. and de Crombrugghe,B. (1996) An 18-base-pair sequence in the mouse proalpha1(II) collagen gene is sufficient for expression in cartilage and binds nuclear proteins that are selectively expressed in chondrocytes. *Mol. Cell Biol.*, **16**, 4512–4523.
32. Dubois,C.L., Shih,H.P., Seymour,P.A., Patel,N.A., Behrmann,J.M., Ngo,V. and Sander,M. (2011) Sox9-haploinsufficiency causes glucose intolerance in mice. *PLoS One*, **6**, e23131.
33. Bell,D.M., Leung,K.K., Wheatley,S.C., Ng,L.J., Zhou,S., Ling,K.W., Sham,M.H., Koopman,P., Tam,P.P. and Cheah,K.S. (1997) SOX9 directly regulates the type-II collagen gene. *Nat. Genet.*, **16**, 174–178.
34. Perl,A.K., Kist,R., Shan,Z., Scherer,G. and Whitsett,J.A. (2005) Normal lung development and function after Sox9 inactivation in the respiratory epithelium. *Genesis*, **41**, 23–32.
35. Bridgewater,L.C., Walker,M.D., Miller,G.C., Ellison,T.A., Holsinger,L.D., Potter,J.L., Jackson,T.L., Chen,R.K., Winkel,V.L., Zhang,Z. *et al.* (2003) Adjacent DNA sequences modulate Sox9 transcriptional activation at paired Sox sites in three chondrocyte-specific enhancer elements. *Nucleic Acids Res.*, **31**, 1541–1553.
36. Sock,E., Pagon,R.A., Keymolen,K., Lissens,W., Wegner,M. and Scherer,G. (2003) Loss of DNA-dependent dimerization of the transcription factor SOX9 as a cause for campomelic dysplasia. *Hum. Mol. Genet.*, **12**, 1439–1447.
37. Spitz,F. and Furlong,E.E. (2012) Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.*, **13**, 613–626.
38. Lettice,L.A., Heaney,S.J., Purdie,L.A., Li,L., de Beer,P., Oostra,B.A., Goode,D., Elgar,G., Hill,R.E. and de Graaff,E. (2003) A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.*, **12**, 1725–1735.
39. Montavon,T., Soshnikova,N., Mascrez,B., Joye,E., Thevenet,L., Splinter,E., de Laat,W., Spitz,F. and Duboule,D. (2011) A regulatory archipelago controls Hox genes transcription in digits. *Cell*, **147**, 1132–1145.
40. Bernard,P., Tang,P., Liu,S., Dewing,P., Harley,V.R. and Vilain,E. (2003) Dimerization of SOX9 is required for chondrogenesis, but not for sex determination. *Hum. Mol. Genet.*, **12**, 1755–1765.
41. Zingg,J.M., Pedraza-Alva,G. and Jost,J.P. (1994) MyoD1 promoter autoregulation is mediated by two proximal E-boxes. *Nucleic Acids Res.*, **22**, 2234–2241.
42. Kondoh,H. and Kamachi,Y. (2010) SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *Int. J. Biochem. Cell Biol.*, **42**, 391–399.
43. Wahlbuhl,M., Reiprich,S., Vogl,M.R., Bosl,M.R. and Wegner,M. (2012) Transcription factor Sox10 orchestrates activity of a neural crest-specific enhancer in the vicinity of its gene. *Nucleic Acids Res.*, **40**, 88–101.
44. Seymour,P.A., Freude,K.K., Dubois,C.L., Shih,H.P., Patel,N.A. and Sander,M. (2008) A dosage-dependent requirement for Sox9 in pancreatic endocrine cell formation. *Dev. Biol.*, **323**, 19–30.