Development 133, 7761-772 doi:10.1242/dev.02239

A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers

Yongsu Jeong[1], Kenia El-Jaick[2], Erich Roessler[2], Maximilian Muenke[2] and Douglas J. Epstein[1,*]

The secreted protein sonic hedgehog (Shh) plays an integral role in forming the ventral midline of the vertebrate central nervous system (CNS). In the absence of Shh function, ventral midline development is perturbed resulting in holoprosencephaly (HPE), a structural malformation of the brain, as well as in neuronal patterning and path finding defects along the length of the anteroposterior neuraxis. Central to the understanding of ventral neural tube development is how *Shh* transcription is regulated in the CNS. To address this issue, we devised an enhancer trap assay to systematically screen 1 Mb of DNA surrounding the *Shh* locus for the ability to target reporter gene expression to sites of *Shh* transcription in transgenic mouse embryos. This analysis uncovered six enhancers distributed over 400 kb, the combined activity of which covered all sites of *Shh* expression in the mouse embryonic CNS from the ventral forebrain to the posterior extent of the spinal cord. To evaluate the relative contribution of these enhancers to the overall pattern of *Shh* expression, individual elements were deleted in the context of a transgenic Bac reporter assay. Redundant mechanisms were found to control *Shh*-like reporter activity in the ventral spinal cord, hindbrain and regions of the telencephalon, whereas unique elements regulated *Shh*-like expression in the ventral midbrain, the majority of the ventral diencephalon and parts of the telencephalon. Three ventral forebrain enhancers locate on the distal side of translocation breakpoints that occurred upstream of *Shh* in human cases of HPE, suggesting that displacement of these regulatory elements from the *Shh* promoter is a likely cause of HPE in these individuals.

KEY WORDS: *Shh*, Gene expression, Forebrain, Holoprosencephaly, Bac modification, Mouse

INTRODUCTION

Shh is expressed in the axial mesoderm (prechordal plate and notochord) and ventral midline (floor plate) of the overlying neural tube throughout most of the anteroposterior (AP) neuraxis (Echelard et al., 1993). It is from these sources that a morphogenic gradient of Shh is generated to promote distinct neuronal progenitor subtypes in the ventral neural tube. In the absence of Shh function, ventral CNS development is severely impaired because of neuronal patterning and pathfinding defects (Chiang et al., 1996; Charron et al., 2003). In humans, the ventral forebrain is particularly sensitive to the level of *Shh* expression, given that a 50% reduction in its normal levels causes holoprosencephaly (HPE), a structural brain malformation resulting in the fusion of the cerebral hemispheres, optic vesicles and other craniofacial anomalies (Roessler et al., 1996). In mice, loss of both *Shh* alleles is required to generate a similar HPE phenotype (Chiang et al., 1996). Over recent years, significant progress has been made in deciphering the molecular mechanisms used by Shh to pattern the ventral neural tube (Jessell, 2000; Patten and Placzek, 2000; Briscoe and Ericson, 2001; Ruiz i Altaba et al., 2003). By contrast, much less is known of the transcription factors acting upstream in the pathway that control *Shh* expression in key signaling centers, including the floor plate and axial mesoderm.

[1]Department of Genetics, University of Pennsylvania School of Medicine, Clinical Research Building, Room 470, 415 Curie Boulevard, Philadelphia, PA 19104, USA. [2]Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD 20892-3717, USA.

*Author for correspondence (e-mail: epsteind@mail.med.upenn.edu)

Accepted 6 December 2005

The etiology of HPE is complex, with multiple genetic and environmental factors contributing to the phenotype. To date, mutations in seven genes, several of which disrupt Shh or Nodal signaling, have been shown to account for nearly 20% of HPE cases in humans (Ming and Muenke, 2002). Additional causative loci have been identified in mice (Hayhurst and McConnell, 2003; Zoltewicz et al., 2004). Surprisingly, these genetic studies have yet to uncover the transcription factors directly responsible for regulating *Shh* expression in the rostral forebrain and prechordal plate. Although *Shh* expression is preferentially lost in the ventral forebrain of mouse embryos with attenuated Nodal signaling, the phenotype appears to result from a failure in the induction of prechordal plate mesoderm, rather than a direct effect on *Shh* transcription per se (Lowe et al., 2001; Dunn et al., 2004). This contrasts with the situation in zebrafish, where Nodal signaling is thought to directly regulate the expression of *Shh* in the ventral midline of the neural tube (Muller et al., 2000; Rohr et al., 2001).

Our efforts to identify the genes regulating *Shh* transcription in the ventral midline of the mouse CNS commenced with the analysis of 35 kb of genomic DNA surrounding *Shh* for regulatory potential in a transgenic mouse reporter assay (Epstein et al., 1999). Multiple enhancers were shown to activate *lacZ* expression at discrete positions along the AP axis of the neural tube (Epstein et al., 1999). Two *Shh* floor-plate enhancers, SFPE1 and SFPE2, were identified that regulated reporter activity in a *Shh*-like pattern in the ventral midline of the spinal cord and hindbrain. A third regulatory element, *Shh* brain enhancer 1 (SBE1), directed reporter activity to the ventral midbrain and caudal region of the diencephalon. The absence of reporter activity in rostral regions of the ventral forebrain in the 35 kb surveyed suggested that the enhancers controlling *Shh* expression in this domain must operate over greater distance. Two additional hints suggested that the *Shh* forebrain enhancers were long-range

acting. First, a translocation breakpoint mapping 250 kb upstream of *Shh* was identified in an individual exhibiting a mild form of HPE, raising the possibility that the translocation separated a forebrain enhancer from the *Shh* transcription start site (Belloni et al., 1996). Second, *Shh* expression in the limb bud is regulated by sequences mapping ~1 Mb upstream of *Shh* (Lettice et al., 2003; Sagai et al., 2005).

In this study, we describe a functional genomic approach to the analysis of *Shh* regulatory sequences. Using modified Bac clones as reporter constructs, we employed an enhancer trap assay to systematically screen 1 Mb of DNA surrounding the *Shh* locus for the ability to target reporter gene expression to sites of *Shh* transcription in transgenic mouse embryos. Highly specific patterns of reporter expression were identified in the axial mesoderm, gut endoderm, limb bud and CNS, all known sites of *Shh* transcription. By coupling the Bac reporter assay with comparative sequence analysis, we identified three novel enhancers located over 400 kb from the *Shh* transcription start site that directed *Shh*-like expression to the ventral forebrain. This brings the total number of *Shh* CNS enhancers to six, the combined activity of which covers all regions of *Shh* transcription along the AP axis of the mouse neural tube. We also assessed the relative contribution of each CNS enhancer to the overall pattern of *Shh* expression by deleting individual elements in the context of the Bac reporter assay. Redundant mechanisms regulated *Shh*-like expression in the ventral spinal cord, hindbrain and subventricular zone of the medial ganglionic eminence (mge) in the telencephalon. Whereas *Shh* reporter activity in the ventral midbrain, much of the ventral diencephalon and the ventricular zone of the mge was found to be dependent on unique elements. The multitude of enhancers controlling *Shh* transcription in the CNS is further indication of the complex mechanisms operating to regionalize the ventral midline of the neural tube along the AP axis.

MATERIALS AND METHODS

Bac modification by homologous recombination

The eGFP reporter gene was introduced into Bacs RPCI-23 429M20 and RPCI-23 389P3 (CHORI, BacPac Resources) by homologous recombination in *E. coli* according to the method of Gong et al. (Gong et al., 2002). All references to nucleotide positions for mouse Chromosome 5 listed below were obtained from the mm5-May'04 assembly. Homology arms A and B flanking mouse *Shh* exon 1 (corresponding to the following nucleotide positions Chr5: 26864937-26865255) were synthesized by PCR using the primer sets E169/E170 and E239/E243, respectively (see Table S1 in the supplementary material for the list of primer sequences described throughout this section). The fragments were cloned into the *Asc*I and *Pac*I restriction sites of the pLD53SCAEB shuttle vector (provided by N. Heintz, Rockefeller University). A detailed description of the Bac modification protocol is available (Jeong and Epstein, 2005). The correct positioning of eGFP within the individual Bacs was verified by probing *Xba*I-digested DNA with external probes on Southern blots.

Bac reporter constructs harboring deletions of SFPE1, SFPE2 and SBE1 in Bac 429M20eGFP were generated according to the same method described above. Homology arms A and B flanking the sequences to be deleted corresponding to mouse SFPE1 (Chr5: 26874573-26875234), SFPE2 (Chr5: 26857299-26858064) and SBE1 (Chr5: 26858065-26858607) were generated by PCR using the following primer sets: E260/E261, E262/E263 (SFPE1); E268/E269, E270/E271 (SFPE2); and E264/E265, E266/E267 (SBE1), and cloned into the pLD53SCAEB shuttle vector after removing the eGFP cassette. The construct carrying deletions of SFPE1 and SFPE2 was generated by deleting SFPE2 from Bac 429M20eGFP$^{\Delta SFPE1}$. Bac reporter constructs harboring deletions of SBE2, SBE3 and SBE4 in Bac 447L17βlacZ were generated by the same approach. Homology arms A and B flanking the sequences to be deleted corresponding to mouse SBE2 (Chr5: 27274453-27275587), SBE3 (Chr5: 27291667-27292041) and SBE4 (Chr5: 27184220-27185333) were generated by PCR using the following primer sets: E289/E290, E291/E292 (SBE2); E295/E296, E297/E298 (SBE3); and E360/E361, E362/E363 (SBE4), and cloned into the pLD53SCAEB shuttle vector after removal of the eGFP cassette. The constructs carrying combined deletions of SBE2 and SBE3, SBE2 and SBE4, SBE3 and SBE4, and SBE2, SBE3 and SBE4 were generated sequentially.

Transposon-based Bac modification

The Tn7βlacZ targeting vector (provided by D. Duboule, University of Geneva) containing Tn7 transposable elements surrounding the β-globin promoter, *lacZ* gene and SV40 poly(A) signal was introduced into overlapping Bacs obtained from the RPCI-23 (208K5, 447L17, 265M1, 214O17) and RPCI-24 (265M10) libraries (CHORI, BacPac Resources) according to the method of Spitz et al. (Spitz et al., 2003). Tn7βlacZ vector (40 ng) was mixed with 200 ng of Bac, GPS buffer and TnsABC (NEB). After incubation at 37°C for 10 minutes, Start Solution was added and the reaction was prolonged for 1 hour. After heat inactivation at 75°C for 10 minutes, electrocompetent DH10B cells were transformed with 1 μl of this incubate and grown overnight on LB plates containing 20 μg/ml Kanamycin and 12.5 μg/ml Chl. Positive colonies were identified by PCR using a primer pair corresponding to sequences from the β-globin promoter and *lacZ* gene. Modified Bac clones were further analyzed by restriction enzyme digestion and pulsed-field gel electrophoresis to determine the number and location of integration sites. Only intact Bacs with a single Tn7βlacZ integration were used in the transgenic reporter assay.

Reporter constructs

All regulatory sequences were cloned into a reporter vector containing the *Shh* minimal promoter, *lacZ* gene and SV40 poly(A) signal. Each of the conserved regions (ECR1-ECR7) from the overlap between Bacs 447L17 and 265M1 was amplified by PCR (ECR1, Chr5: 27274453-27275587; ECR3, Chr5: 27291667-27292041; and data not shown) using primer sets E278/E279 (ECR1) and E282/E283 (ECR3). SBE4 (Chr5: 27184220-27185333) was amplified by PCR using primers E337 and E338. Conserved SBE2 sequences from human (Chr7: 155560539-155561314), chicken (Chr2: 7701492-7702259), frog (Scaffold57: 2747939-2748753) and tetraodon (Chr6: 4066565-4067332) were amplified from genomic DNA by PCR using primer sets: E311/E313 (human); E309/E310 (chicken); E319/E320 (frog); E321/E322 (tetraodon). Conserved SBE4 sequences from chicken (Chr2: 7641242-7642395), frog (Scaffold57: 2689775-2690912) and tetraodon (Chr6: 4063308-4064467) were amplified from genomic DNA by PCR using primer sets: E372/E373 (chicken); E368/E369 (frog); E374/E375 (tetraodon). The frog DNA derived from *X. tropicalis* was kindly provided by Dr Frank Conlin (UNC, Chapel Hill, NC).

Electromobiliy shift assays (EMSA)

pCMV or pCMV-Nkx2.1 plasmids were transfected into NIH3T3 cells using FuGENE 6 transfection reagent (Roche). After 48 hours, whole-cell lysates were prepared in a buffer containing 50 mM Tris-HCl (pH 7.4), 1 mM EDTA, 1 mM DTT, protease inhibitor cocktail and 25% glycerol. For EMSA, 10 μg of protein from the cell lysates was incubated for 10 minutes at room temperature in a DNA-binding buffer containing 10 mM Tris-HCl (pH 7.4), 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 5% glycerol, 200 ng poly (dI-dC), 1 μg BSA in the presence or absence of competitor double stranded oligonucleotides. After 0.1 ng (5×10^4 to 10×10^4 cpm) of probe was added to the mixture, incubation was continued for an additional 20 minutes. Protein-bound DNA complex was separated from free probe on a 6.5% acrylamide gel run in 1×TBE (Tris-borate-EDTA) buffer. The sequence of probe and competitors were as follows: probe (also used as wild-type competitor), 5′-ACAGAAATTGTTTTTTAAGTAGTTGCC-CTCTGAAAATAT-3′; mutant competitor, 5′-ACAGAAATTGTTT-TTTGGAGCATTGCCCTCTGAAAATAT-3′. Underlined regions of the sequence indicate wild-type and mutant Nkx2-binding sites, respectively.

Production and genotyping of transgenic mice

Conventional plasmid transgenes were prepared for microinjection as described (Epstein et al., 1999). Bac transgenes were linearized with the PI-*Sce*I (NEB) restriction enzyme prior to injection (Jeong and Epstein, 2005). Transient transgenic embryos or mouse lines were generated by pronuclear

injection into fertilized eggs derived from the (BL6×SJL) F1 mouse strain (Jackson Labs). The primers listed in parentheses were used for PCR genotyping of mice carrying: 429M20eGFP and 389P3eGFP (E128 and E129); 208K5βlacZ, 447L17βlacZ, 265M1βlacZ, 214O17βlacZ and 265M10βlacZ (E276 and E277).

Whole-mount β-galactosidase staining, in situ hybridization and immunohistochemistry

The assessment of β-galactosidase activity was performed as described (Epstein et al., 1999). Whole-mount RNA in situ hybridization was performed using a digoxigenin-UTP-labeled *Shh* riboprobe (Echelard et al., 1993). Stained embryos were photographed after dehydration in methanol and clearing in benzyl alcohol:benzyl benzoate (1:1). Representative embryos were rehydrated, immersed in 30% sucrose overnight, embedded and frozen in OCT and sectioned at 20 μm on a cryostat. Primary antibodies used for immunohistochemistry and dilutions were as follows: eGFP (Molecular Probes), 1:1000; Shh (5E1, Developmental Studies Hybridoma Bank), 1:100. Detection of primary antibodies was achieved using Cy3- (Jackson ImmunoResearch Laboratories) or Alexa 488- (Molecular Probes) conjugated goat anti-mouse secondary antibodies.

RESULTS

Requirement of locally acting *Shh* regulatory sequences

To determine whether *Shh* forebrain enhancer(s) resided on sequences beyond the 35 kb previously analyzed, we isolated two Bac clones (429M20 and 389P3) from the RPCI-23 library (CHORI) that extended 180 kb upstream and 160 kb downstream of the *Shh* promoter, respectively. The first *Shh* exon in each Bac was replaced with an eGFP reporter gene by homologous recombination in bacteria according to the method of Gong et al. (Gong et al., 2002) (see Materials and methods for a description of the targeting strategy). Transgenic embryos were then generated with the modified Bacs and assessed for reporter activity. At 9.5 dpc, embryos carrying either of the 429M20eGFP or 389P3eGFP transgenes displayed eGFP expression in the ventral midline of the CNS in a similar pattern to that generated from the locally acting *Shh* regulatory elements (Fig. 1A,B; Table 1) (Epstein et al., 1999). In addition, *Shh*-like reporter activity was detected in the hindgut of embryos carrying 429M20eGFP but not 389P3eGFP, indicating that cis-acting sequences controlling *Shh* expression in the gut endoderm are located on Bac 429M20 (Fig. 1A,B).

It was clear from this experiment that regulatory sequences controlling *Shh* expression in the rostral forebrain were not located within 180 kb upstream or 160 kb downstream of the *Shh* promoter. However, as the previously identified local enhancers were active in embryos carrying 429M20eGFP, it provided us with an opportunity to assess their requirement in a genomic context more in keeping with the endogenous *Shh* locus. To address the contribution of SFPE1, SFPE2 and SBE1 in regulating *Shh*-like expression in the CNS, the sequences mediating their respective activities were deleted from Bac 429M20eGFP by homologous recombination in bacteria. Transgenic embryos carrying 429M20eGFP$^{\Delta SFPE1}$ or 429M20eGFP$^{\Delta SFPE2}$ showed no loss of eGFP expression in the floor plate of the hindbrain or spinal cord (compare Fig. 1B,I with 1C,K and 1D,L). By contrast, embryos carrying 429M20eGFP$^{\Delta SFPE1,\Delta SFPE2}$, a Bac transgene in which both floor plate enhancers were deleted, showed barely detectable levels of eGFP staining in the ventral midline of the hindbrain and spinal cord, despite normal expression in the ventral midbrain (Fig. 1E,M). Conversely, embryos carrying 429M20eGFP$^{\Delta SBE1}$, a Bac with a deletion of SBE1, showed an absence of eGFP expression in the ventral midbrain and caudal diencephalon, despite normal staining in the hindbrain and spinal cord (compare Fig. 1B,I,J with F,N,O). These results suggest that *Shh*-like transcription in the floor plate of the hindbrain and spinal cord is dependent on the redundant activities of SFPE1 and SFPE2, and that *Shh*-like expression in the ventral midbrain and caudal diencephalon is solely dependent on SBE1.

Positioning long-range *Shh* regulatory sequences on Bacs using an enhancer trap assay

The absence of reporter activity in the rostral forebrain of embryos carrying 429M20eGFP or 389P3eGFP suggested that the sequences regulating *Shh* expression in this tissue must operate over considerable distance. To further explore this possibility, we had to develop an efficient strategy of screening large amounts of DNA for regulatory activity. Comparative analysis of genomic DNA from divergent organisms is an effective method of identifying evolutionary conserved regions (ECR) that may function as regulatory elements (Nobrega et al., 2003; Thomas et al., 2003; Woolfe et al., 2005). The problem with applying this approach over large genomic intervals is that the number of ECRs is often too numerous to easily identify the ones of interest. For example, in the 800 kb interval between *Shh* and *Rnf32*, the next neighboring gene upstream, there are ~170 ECRs (>70% identity over 100 bp) shared between human and mouse, of which 57 are also found in chicken (ECR browser: http://ecrbrowser.dcode.org/). Rather than testing each of the ECRs independently for *Shh* reporter activity, we implemented an enhancer-trap assay similar to the one designed by Spitz et al. (Spitz et al., 2003) in their study of the HoxD global control region. In this way, modified Bacs could be used as reporter constructs to first determine whether a given genomic interval possessed *Shh* enhancer activity. Five overlapping Bacs extending ~1 Mb upstream of *Shh* were isolated (CHORI) and modified to incorporate a single copy reporter cassette containing a minimal β-globin promoter and *lacZ* gene (Fig. 2). The reporter cassette was randomly inserted into each of the Bacs by Tn7 transposition in *E. coli* (Spitz et al., 2003). Two independent transposition events were generated for each Bac and subsequently assayed for reporter activity in transgenic embryos.

Remarkably, of the five Bacs tested, four directed highly specific and reproducible patterns of reporter expression to the axial mesoderm, ventral forebrain and posterior limb bud, all known sites of *Shh* transcription (Fig. 2). Of particular interest was the observation that X-gal staining in the ventral forebrain of embryos carrying either of two overlapping Bacs (447L17βlacZ and 265M1βlacZ) was similar to each other and highly reflective of the pattern of *Shh* expression in the ventral diencephalon and telencephalon (compare Fig. 2A with C,D).

In addition to the ventral forebrain, X-gal staining was observed in the notochord in embryos carrying any of three overlapping Bacs (208K5, 447L17, 265M1), suggesting the existence of at least two long-range notochord enhancers (Fig. 2B,C,D). The early expression of *Shh* in the node, notochord, prechordal plate and ventral forebrain was also recapitulated in embryos carrying 447L17βlacZ (see Fig. S1 in the supplementary material). However, in contrast to the normal expression of *Shh*, X-gal staining was not maintained at rostral levels of the notochord in older embryos (Fig. 2B-D).

The most distant acting enhancer identified in our study mapped to Bac 265M10βlacZ and directed *Shh*-like reporter activity to the zone of polarizing activity (ZPA) in the posterior region of the fore- and hindlimb buds between 10.5 and 12.5 dpc (Fig. 2F; data not shown). This *Shh* ZPA enhancer is probably the same as the one reported by Lettice et al. (Lettice et al., 2003), located ~850 kb upstream of *Shh*.

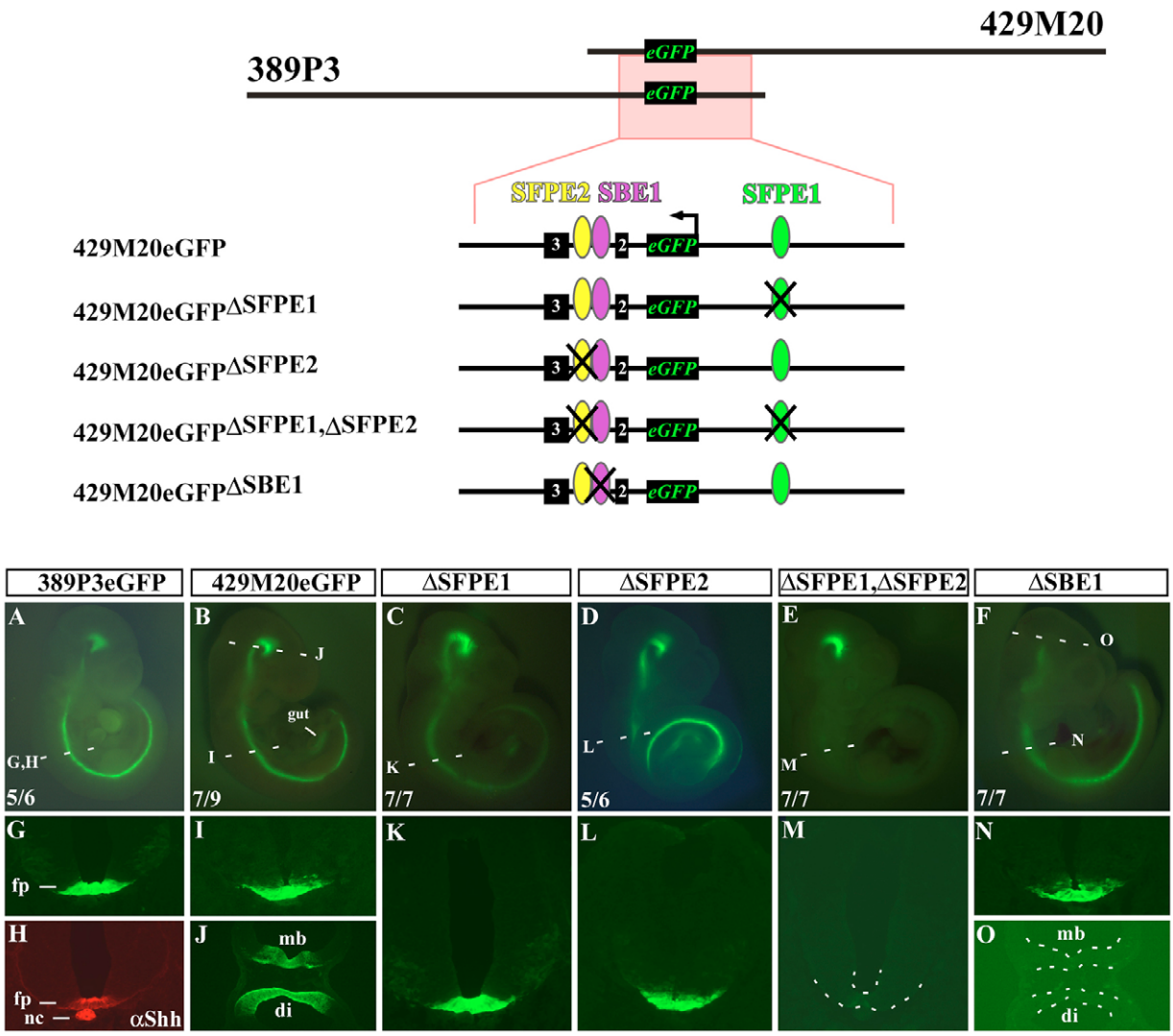

Fig. 1. Functional analysis of locally acting *Shh* enhancers in the context of Bac 429M20eGFP. Schematic of Bac clones 429M20eGFP and 389P3eGFP, which contain an eGFP insertion into *Shh* exon 1 and extend 180 kb upstream and 160 kb downstream of *Shh*, respectively. Reporter constructs carrying enhancer deletions generated in Bac 429M20eGFP are listed. (**A-F**) Whole-mount views of embryos carrying (A) Bac 389P3eGFP, (B) 429M20eGFP and (C-F) 429M20eGFP deletion constructs. (**G-I**) In embryos carrying 389P3eGFP or 429M20eGFP, strong eGFP expression was observed in the floor plate (G,I), in a pattern comparable with endogenous Shh protein (H). Embryos carrying a deletion of SFPE1 (C,**K**) or SFPE2 (D,**L**) show similar patterns of eGFP fluorescence to each other and to embryos carrying the wild type Bac (B). (**J**) The ventral midbrain and caudal diencephalon also show strong eGFP expression in embryos carrying 429M20eGFP and 389P3eGFP. Embryos carrying deletions of both SFPE1 and SFPE2 showed a dramatic reduction in eGFP expression in the floor plate (E,**M**). Embryos carrying a deletion of SBE1 showed a complete loss of eGFP staining in the ventral midbrain and caudal diencephalon (compare F,**N**,**O** with B,I,J). The absence of hindgut staining in embryos depicted in E and F is a consequence of the variable nature of the hindgut enhancer and does not reflect a dependency on sequences mediating Shh floor-plate enhancer activity. eGFP staining on sections was performed by immunohistochemistry using an anti eGFP antibody. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct (A-F). di, diencephalon; fp, floor plate; mb, midbrain; nc, notochord.

Table 1. Sites of *Shh* enhancer activity in the CNS

	Sc	Hb	Mb	Dienc			Tel (mge)	
Shh enhancer (mm5-May'04)	(floor plate)			(p1-p2)	(p3)	(p4-p6)	(vz)	(svz)
SFPE1 (chr5:26874570-26875234)	+	+	–	–	–	–	–	–
SFPE2 (chr5:26857296-26858042)	+	+	–	–	–	–	–	–
SBE1 (chr5:26858062-26858595)	–	–	+	+	–	–	–	–
SBE2 (chr5:27274453-27275587)	–	–	–	–	+	+	–	–
SBE3 (chr5:27291667-27292041)	–	–	–	–	–	–	–	+
SBE4 (chr5:27184220-27185333)	–	–	–	–	+	–	+	+

The nucleotide position (according to the mm5-May'04 assembly) for each *Shh* regulatory element is listed in parentheses. The presence (+) or absence (–) of reporter activity in a given location is indicated.
Dienc, diencephalon; Hb, hindbrain; Mb, midbrain; mge, medial ganglionic eminence; p, prosomere; Sc, spinal chord; svz, subventricular zone; Tel, telencephalon; vz, ventricular zone.

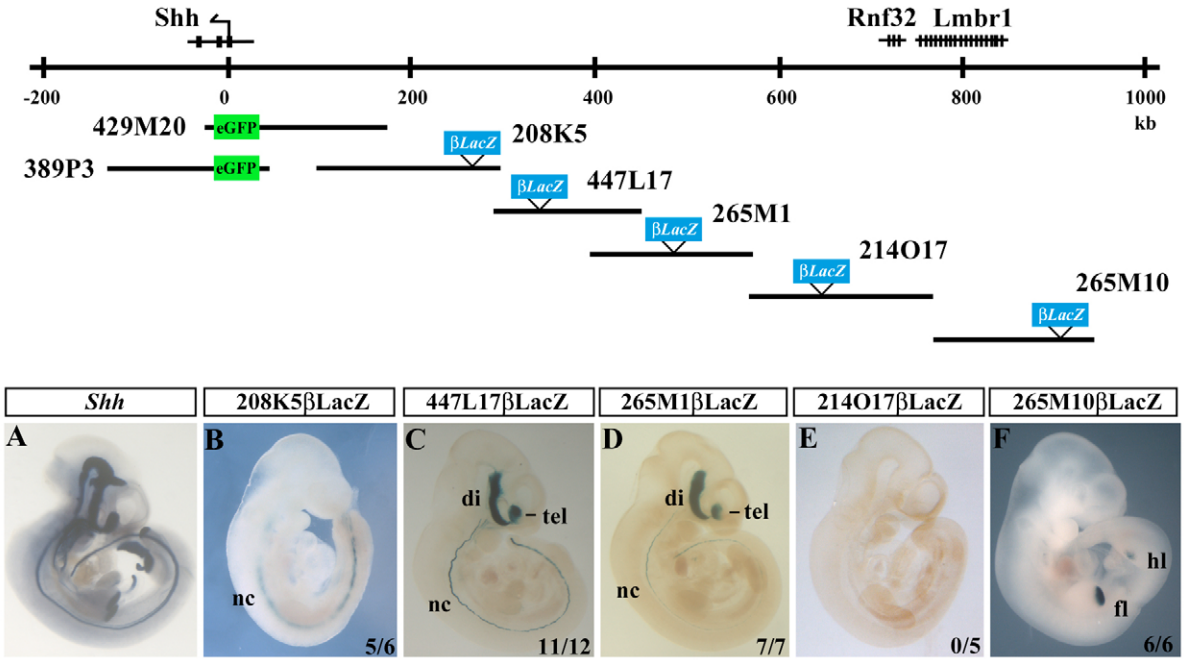

Fig. 2. A screen for long-range acting *Shh* regulatory sequences. Physical map displaying the distribution of genes and Bac clones spanning 1 Mb upstream of *Shh*. Reporter cassettes (*lacZ*, blue rectangle) were introduced into overlapping Bacs by random transposon insertion. (**A-F**) Whole-mount views of embryos stained for (A) *Shh* mRNA and (B-F) *lacZ* expression (X-gal) in representative Bac transgenic embryos harvested at 10.5 dpc. X-gal staining in the notochord was observed in embryos carrying any of three overlapping Bacs (208K5βlacZ, 447L17βlacZ, 265M1βlacZ), suggesting the existence of at least two long-range acting notochord enhancers (B-D). X-gal staining in the ventral forebrain was observed in embryos carrying 447L17βlacZ (B) and 265M1βlacZ (D). X-gal staining in the posterior limb bud was observed in embryos carrying 265M10βlacZ (F). None of the five embryos carrying 214O17βlacZ exhibited reporter activity (E). di, diencephalon; fl, forelimb; hl, hindlimb; nc, notochord; tel, telencephalon. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct (B-F).

Identifying long range *Shh* enhancers by comparative sequence analysis

The similar pattern of X-gal staining in the forebrain of embryos carrying the 447L17βlacZ and 265M1βlacZ transgenes suggested that the cis-acting sequences responsible for this expression localized to the 70 kb region of overlap between the two clones. The pair-wise alignment (ECR browser: http://ecrbrowser.dcode.org/) of genomic sequence from this interval between mouse and human identified seven highly conserved regions of DNA showing at least 75% nucleotide identity that ranged in size from 0.37 to 1.4 kb (Fig. 3). To address their regulatory potential, each of the ECRs in this interval was amplified by PCR from mouse genomic DNA, cloned into a reporter cassette containing a minimal *Shh* promoter and *lacZ* gene and assayed for forebrain enhancer activity in transgenic embryos. Of the seven ECRs (ECR1-7) tested, embryos carrying ECR1 and ECR3 showed consistent X-gal staining in the ventral forebrain. The 1.1 kb ECR1 fragment, located 410 kb distal to *Shh*, directed *lacZ* expression to the rostral region of the ventral diencephalon in two lateral stripes that converged in the ventral midline at the level of the optic vesicles in a manner consistent with *Shh* and 447L17βlacZ reporter activity (compare Fig. 3D,J,P,V with 3A,G,M,S and 3B,H,N,T). Embryos carrying the 370 bp ECR3 element, which mapped 426 kb distal to *Shh*, directed reporter activity to only part of the *Shh* expression domain in the ventral telencephalon, encompassing the subventricular zone (svz) of the medial ganglionic eminence (mge) (Fig. 3E,K,Q,W). As the expression of *Shh* and 447L17βlacZ reporter activity was also detected in the ventricular zone (vz) of the mge, we presumed that a second telencephalic enhancer must regulate *Shh* transcription at this site (Fig. 3S,T). None of the other ECR constructs directed reporter activity to *Shh*-expressing tissues, including the mge and axial mesoderm, suggesting that these enhancers must form on other sequences contained in Bac 447L17 (data not shown). In keeping with the nomenclature of previously identified *Shh* regulatory elements, we refer to the rostral diencephalic enhancer as Shh brain enhancer-2 (SBE2) and the telencephalic svz enhancer as SBE3 (Table 1).

The pattern of X-gal staining in the ventral telencephalon of embryos carrying 447L17βlacZ and 265M1βlacZ appeared similar when viewed as whole embryos (Fig. 3B,C); however, upon their sectioning, subtle yet significant differences were realized. 447L17βlacZ-containing embryos showed X-gal staining in both the vz and svz of the mge in a *Shh*-like pattern, whereas, 265M1βlacZ containing embryos only stained positively in the svz of the mge (Fig. 3S-U). This indicated that the missing mge enhancer must reside somewhere on the 100 kb region of Bac 447L17 that did not overlap with Bac 265M1. In analyzing the whole genome alignment (ECR browser) of human, mouse and chicken covering this 100 kb interval, we identified only one highly conserved region (77% identity over 350 bp) positioned ~319 kb upstream of the *Shh* promoter (Fig. 3). A 1.1 kb fragment overlapping this ECR was amplified from the mouse genome and tested for reporter activity. Transgenic embryos carrying the 1.1 kb fragment showed X-gal staining in both the vz and svz of the mge in a manner comparable with 447L17βlacZ and *Shh* expression (Fig. 3F,X). X-gal staining was also detected in the ventral diencephalon in the vicinity of prosomere 3 (Fig. 3F,L). This enhancer was designated SBE4 (Table 1).

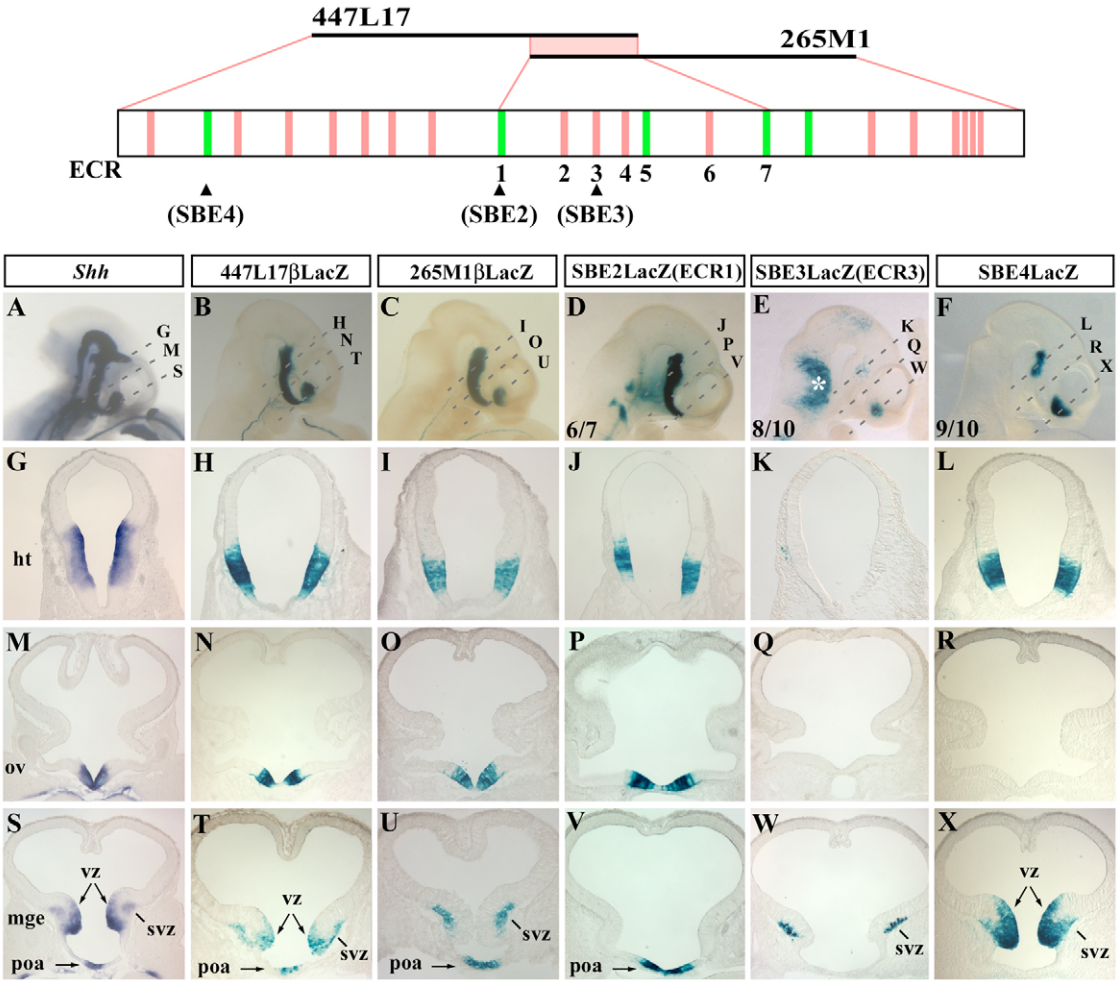

Fig. 3. Isolation of regulatory sequences controlling *Shh* reporter activity in the forebrain. Schematic of Bac clones 447L17 and 265M1. Whole-genome alignments were performed using the ECR Browser with human sequence (hg17-May'04, chr7:155395662-155783102) as the base. ECRs (>75% similarity over a minimum of 350 bp) shared between human and mouse (pink) or between human, mouse and chicken (green) are represented by vertical bars. ECRs in the overlap between 447L17 and 265M1 are numbered ECR1 to ECR7. (**A**) *Shh* mRNA expression in the head of a 10.5 dpc embryo with accompanying transverse sections through the forebrain at the level of the (G) hypothalamus, (M) optic vesicles and (S) telencephalon. (**B-X**) X-gal staining in the forebrain of transgenic embryos carrying (B,H,N,T) 447L17βlacZ; (C,I,O,U) 265M1βlacZ; (D,J,P,V) ECR1/SBE2; (E,K,Q,W) ECR3/SBE3; and (F,L,R,X) SBE4 reporter constructs at 10.5 dpc. *Shh* forebrain reporter activity is divided into ventral diencephalic (SBE2, SBE4) and ventral telencephalic (SBE3, SBE4) regulatory elements. Unlike *Shh* and 447L17βlacZ, 265M1βlacZ transgene was not expressed in ventricular zone (vz) of the medial ganglionic eminence (mge) (compare U with S,T). SBE3 activity is restricted to the subventricular zone (svz) of the mge (W). The asterisk in E indicates ectopic X-gal staining in the hindbrain. ht, hypothalamus; ov, optic vesicle; poa, preoptic area; vz, ventricular zone; svz, subventricular zone. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct (D-F).

Requirement of long range *Shh* forebrain enhancers

We next addressed whether the sequences mediating SBE2, SBE3 and SBE4 were required for the *Shh*-like pattern of expression in the ventral forebrain of embryos carrying 447L17βlacZ. We reasoned that if SBE2-4 were necessary then deleting their sequences from 447L17βlacZ would abrogate reporter activity in the ventral forebrain. However, if ventral forebrain expression persisted, in whole or in part, in the absence of SBE2-4, then additional regulatory elements would have to be implicated.

Transgenic embryos carrying 447L17βlacZ$^{\Delta SBE2}$ were devoid of reporter activity throughout most of the rostral-ventral diencephalon with the exception of a small region corresponding to prosomere 3 (p3) that showed weak bilateral patches of X-gal staining (compare Fig. 4A-C with Fig. 3B,H,N). Reporter activity in the ventral telencephalon and notochord were unaffected by the deletion of SBE2 (Fig. 4A,D). Interestingly, embryos carrying 447L17βlacZ$^{\Delta SBE2, \Delta SBE3}$ showed a similar pattern of X-gal staining to those carrying 447L17βlacZ$^{\Delta SBE2}$ (Fig. 4E-H). This finding indicates that SBE3 is not required for svz expression in the ventral telencephalon and that another forebrain element in Bac 447L17 must compensate in its absence (Fig. 4D,H). As SBE3 and SBE4 were each capable of directing reporter expression to the svz of the ventral telencephalon, it is likely that the two enhancers function in a partially redundant manner. To test this hypothesis, we engineered deletions in Bac 447L17βlacZ of either SBE4 on its own or in combination with SBE3 and tested the respective transgenes (447L17βlacZ$^{\Delta SBE4}$, 447L17βlacZ$^{\Delta SBE3, \Delta SBE4}$) for reporter activity. Embryos carrying 447L17βlacZ$^{\Delta SBE4}$ showed no X-gal staining in the vz of the ventral telencephalon, but did retain staining in the svz

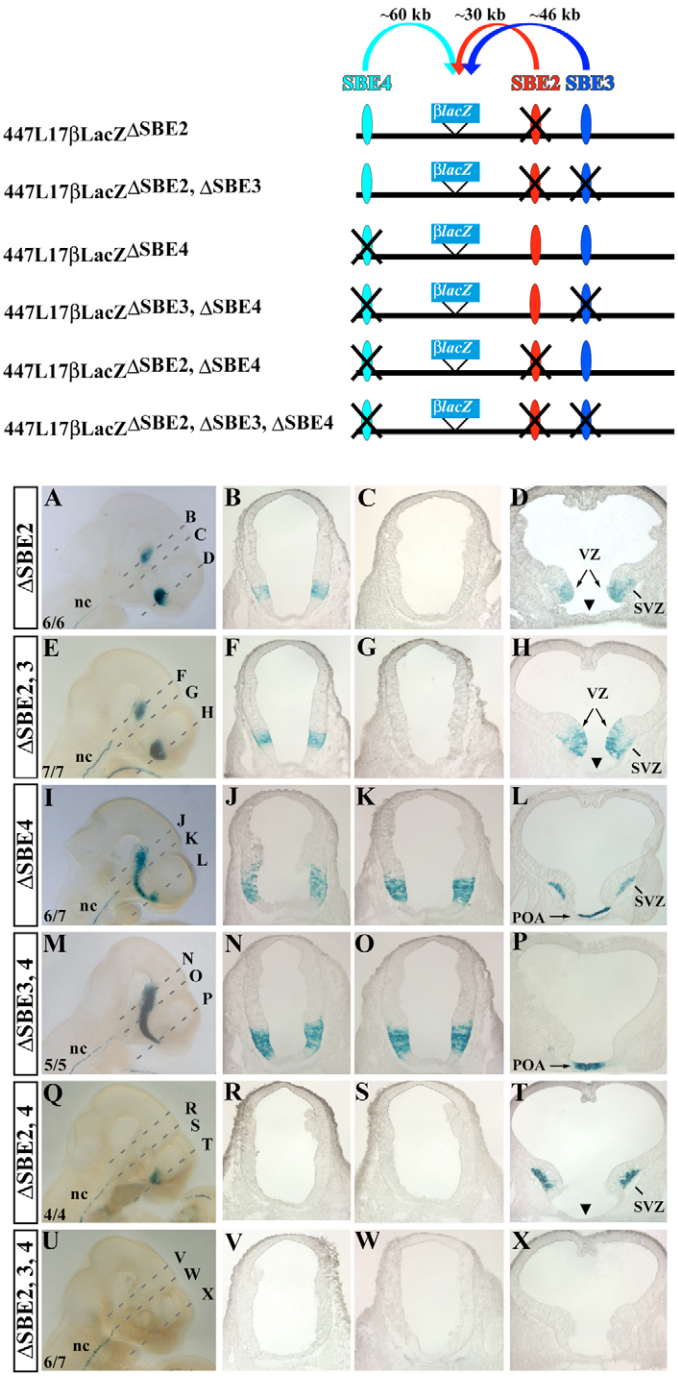

Fig. 4. Requirement of *Shh* forebrain enhancers in the context of Bac 447L17βlacZ. Schematic representation of the SBE2, SBE3 and SBE4 deletion constructs. X-gal staining in the forebrain of embryos carrying (**A-D**) 447L17βlacZ$^{\Delta SBE2}$, (**E-H**) 447L17βlacZ$^{\Delta SBE2, \Delta SBE3}$, (**I-L**) 447L17βlacZ$^{\Delta SBE4}$, (**M-P**) 447L17βlacZ$^{\Delta SBE3, \Delta SBE4}$, (**Q-T**) 447L17βlacZ$^{\Delta SBE2, \Delta SBE4}$ and (**U-X**) 447L17βlacZ$^{\Delta SBE2, \Delta SBE3, \Delta SBE4}$. Deletion of SBE2 resulted in the loss of reporter activity at most levels of the diencephalon, including the preoptic area (POA; arrowhead in D). Compare A-D with the pattern of X-gal staining in embryos carrying the wild-type 447L17βlacZ transgene (Fig. 3B,H,N,T). Embryos carrying 447L17βlacZ$^{\Delta SBE2, \Delta SBE3}$ (E-H) showed patterns of X-gal staining that were similar to those carrying 447L17βlacZ$^{\Delta SBE2}$ (A-D). Embryos carrying 447L17βlacZ$^{\Delta SBE4}$ (I-L) showed an absence of staining in the ventricular zone (vz) of the mge (L). In embryos carrying 447L17βlacZ $^{\Delta SBE3, \Delta SBE4}$ (M-P), X-gal staining was not detected in the vz or subventricular zone (svz) of the medial ganglionic eminence (mge) (P). Deletion of SBE2 and SBE4 resulted in a loss of expression in the vz of the mge and the entire diencephalon (Q-T). Embryos carrying 447L17βlacZ$^{\Delta SBE2, \Delta SBE3, \Delta SBE4}$ (U-X) showed no expression in the diencephalon or telecephalon. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct (A,E,I,M,Q,U).

(compare Fig. 4L with Fig. 3T). Only when SBE3 and SBE4 were both deleted (447L17βlacZ$^{\Delta SBE3, \Delta SBE4}$) was X-gal staining completely eliminated from the ventral telencephalon (Fig. 4M,P). These results confirm that the *Shh*-like expression derived from Bac 447L17βlacZ in the svz of the ventral telencephalon is dependent on both SBE3 and SBE4, and that the vz expression in the mge is solely dependent on SBE4.

Redundant mechanisms were also identified in the regulation of *Shh*-like reporter activity in the p3 region of the ventral diencephalon. Only when SBE2 and SBE4 were deleted from Bac 447L17βlacZ (447L17βlacZ$^{\Delta SBE2, \Delta SBE4}$) was reporter activity abrogated from the p3 domain (Fig. 4Q,R). Interestingly, transgenic embryos carrying 447L17βlacZ$^{\Delta SBE2, \Delta SBE3, \Delta SBE4}$, in which all three forebrain enhancers were deleted, showed no X-gal staining anywhere in the ventral forebrain, despite persistent notochord staining, indicating that all *Shh*-like expression generated from Bac 447L17βlacZ is regulated by SBE2-4 (Fig. 4U-X).

Conservation of *Shh* forebrain enhancer activity

The sequences mediating the activities of SBE2-SBE4 were identified based on their high degree of conservation between mouse and human (80% for SBE2 and SBE4 and 72% for SBE3). To determine the extent to which the conservation of sequence reflected the conservation of enhancer function, we performed blat and blast searches using the UCSC (http://www.genome.ucsc.edu/) and Ensemble (http://www.ensembl.org/) genome browsers, respectively, and identified enhancer sequences from a number of vertebrate species. SBE2 sequences were identified in most vertebrate phyla at various distances from the *Shh* promoter (Fig. 5A,B). However, SBE3 sequences were only found in closely related organisms, including chimp and rat, raising the possibility that the SBE3 element arose relatively recently in the clade containing rodents and primates (data not shown). Sequences mediating SBE4 were identified in most organisms surveyed; however, the degree of conservation was reduced in comparison with SBE2 (Fig. 6A,B).

The SBE2 sequences from human, chicken and frog were each sufficient to drive *lacZ* expression into the ventral diencephalon of transgenic mice, in keeping with the significant preservation of nucleotide identity across much of this element (Fig. 5C-E). By contrast, the SBE2 sequence from puffer fish, which differed significantly from the other vertebrates, was incapable of regulating transcription in the ventral diencephalon of mouse embryos (Fig. 5F). Similarly, the SBE4 element from chicken showed conserved reporter activity in mice, but the more divergent frog and puffer fish SBE4 sequences were not active (Fig. 6C-E). Whether these findings indicate that puffer fish and frog use other cis-acting sequences to regulate distinct aspects of *Shh* expression in the forebrain, or that the sequences tested can only accommodate enhancer function in the puffer fish and frog brains remains to be determined.

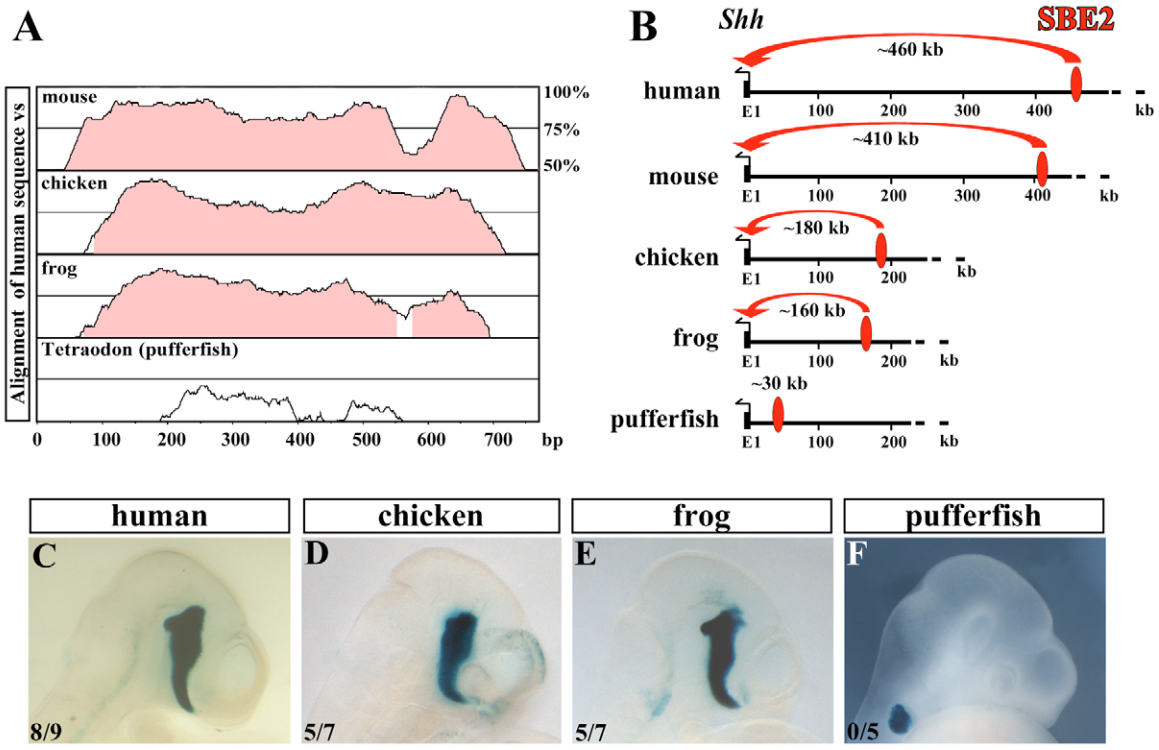

Fig. 5. Conservation of SBE2 sequence and function. (**A**) Vista plots comparing the alignment of human SBE2 sequences with mouse, chicken, frog and tetraodon. (**B**) Comparison of the distance between SBE2 and the *Shh* transcription start site in human, mouse, chicken, frog and tetraodon. (**C-F**) SBE2 reporter activity derived from (C) human, (D) chicken, (E) frog and (F) pufferfish sequences. X-gal staining in the otic vesicle of the embryo shown in F is ectopic. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct (C-F).

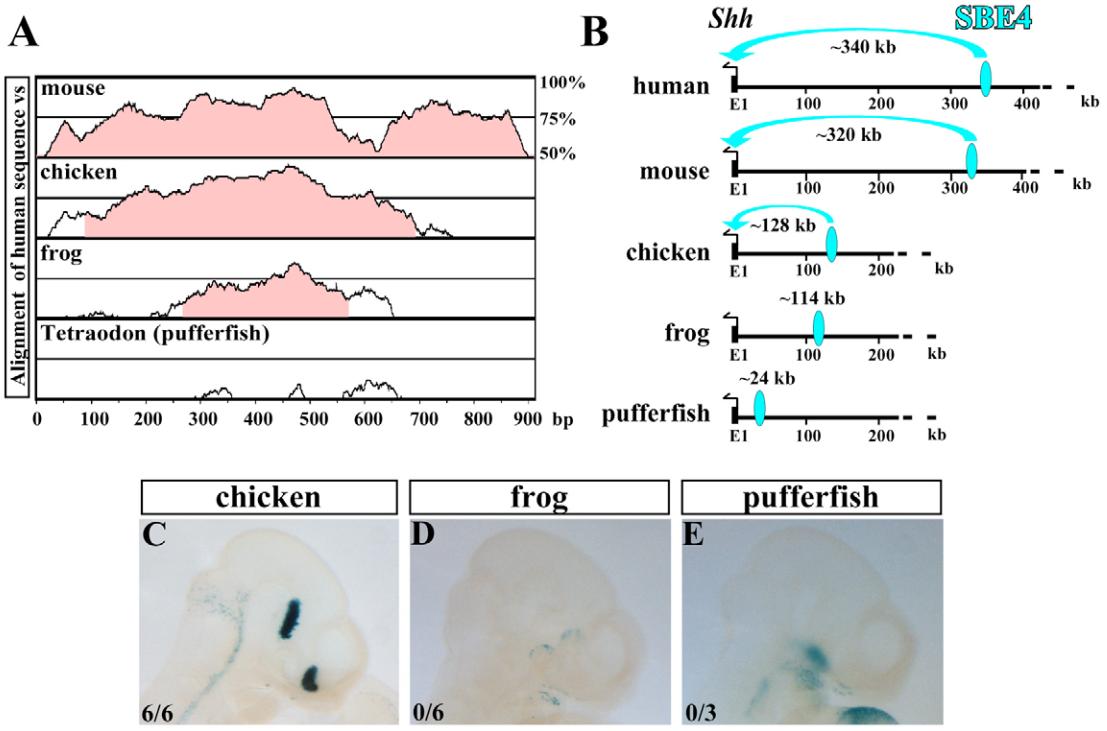

Fig. 6. Conservation of SBE4 sequence and function. (**A**) Vista plots comparing the alignment of human SBE4 sequences with mouse, chicken, frog and tetraodon. (**B**) Comparison of the distance between SBE4 and the *Shh* transcription start site in human, mouse, chicken, frog and tetraodon. (**C-E**) SBE4 reporter activity derived from (C) chicken, (D) frog and (E) pufferfish sequences. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct (C-E).

Dependency of SBE3 reporter activity on an Nkx2 binding site

To begin to understand the molecular mechanisms governing *Shh* forebrain enhancer activity, we surveyed the sequence of one of the enhancers, SBE3, for known transcription factor-binding sites and considered only those sites showing a correlation between the expression of their cognate DNA-binding protein and that of *Shh*. A binding site matching the consensus for Nkx2 proteins (T(T/C)AAGT(A/G)(G/C)TT) (Watada et al., 2000) was identified in the sequence of the SBE3 enhancer in a region that was 100% conserved between human and mouse (Fig. 7A). The expression of *Nkx2.1* overlaps with that of *Shh* in the ventral forebrain including the svz of the mge (Shimamura et al., 1995). Moreover, the telencephalic expression of *Shh* is downregulated in *Nkx2.1*$^{-/-}$ embryos (Sussel et al., 1999). However, as the mge fails to form in these mutants, it is unclear whether the regulation of *Shh* by Nkx2.1 is direct. To address this issue, we performed electromobility shift assays, the results of which suggested that Nkx2.1 protein was able to bind to its target recognition sequence in the SBE3 enhancer (Fig. 7B). We next determined the requirement of the Nkx2-binding site in the context of the SBE3 reporter assay. Transgenic embryos carrying an SBE3 reporter construct that lacked the core Nkx2 binding site (AAGTAG) failed to activate *lacZ* expression in the svz of the mge (Fig. 7D,F). Based on these findings, we conclude that Nkx2.1 is a direct regulator of SBE3 activity.

DISCUSSION

In this report, we describe a functional genomic approach to the analysis of *Shh* regulatory sequences. An enhancer trap assay was used to systematically screen 1 Mb of DNA surrounding the *Shh* locus for the ability to target reporter gene expression to ventral cells of the mouse neural tube. The strength of our technique in analyzing large amounts of DNA for regulatory potential relied on coupling a transgenic Bac reporter assay with comparative sequence analysis. By performing the Bac reporter assay first, a reasonable genomic interval could be defined in which to search for conserved regulatory

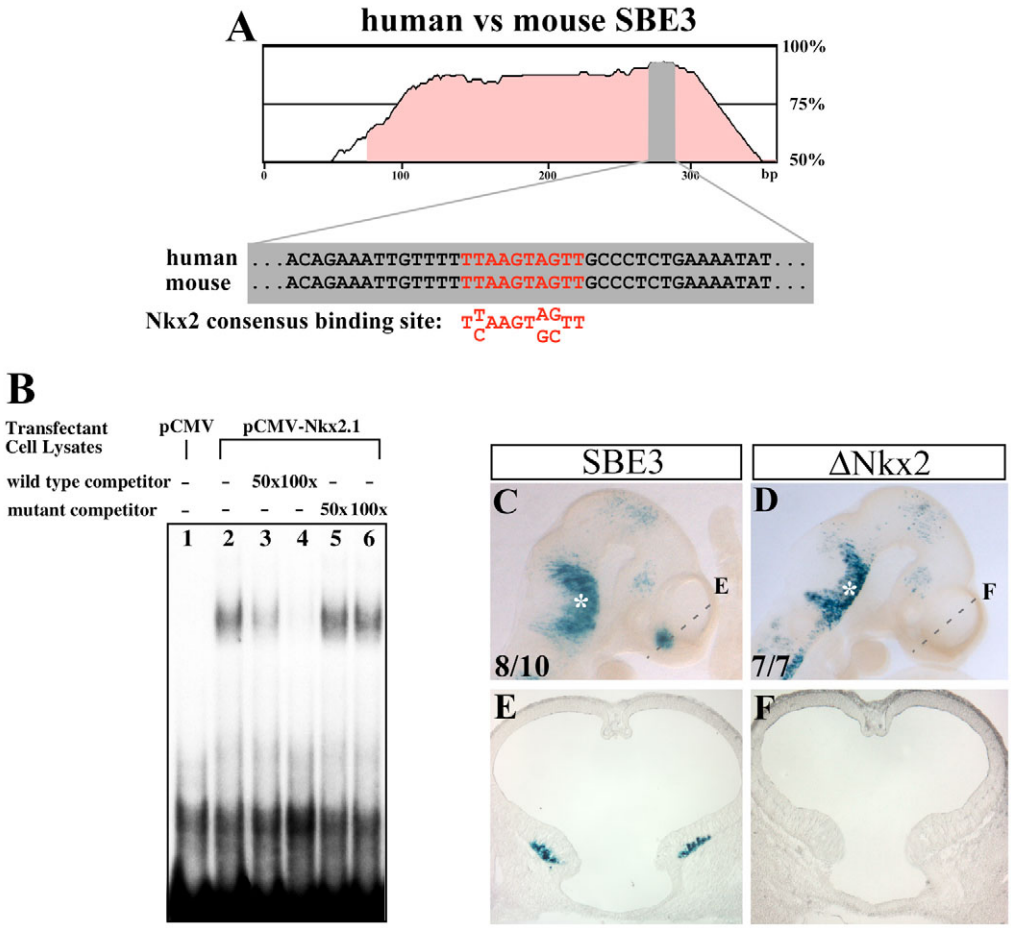

Fig. 7. An Nkx2 binding site is required for SBE3 enhancer function. (**A**) Vista plot of the alignment between human and mouse SBE3 sequences. The target recognition sequence for the Nkx2 homeodomain protein (red) is present in both human and mouse SBE3 sequences. (**B**) Binding of Nkx2.1 protein to a site in the SBE3 sequence. Cell lysates transfected with pCMV (lane 1) or pCMV-Nkx2.1 (lanes 2-6) were analyzed for binding to a 39 bp probe overlapping the Nkx2 recognition sequence in SBE3. Wild-type cold competitor (lanes 3, 4) interfered with binding of Nkx2.1 protein, while mutant competitor with nucleotide substitutions in the core binding site (AAGTAG→GGAGCA) did not alter the shifted complex (lanes 5, 6). (**C-F**) X-gal staining of transgenic embryos carrying a wild-type (C,E) or mutant SBE3 reporter construct in which the Nkx2 core recognition sequence (AAGTAG) was deleted (D,F). Embryos carrying the wild-type SBE3 reporter construct show consistent X-gal staining in the svz of the mge (C,E). By contrast, embryos carrying an SBE3 reporter construct lacking the Nkx2 site (ΔNkx2) showed no staining in the ventral forebrain (D,F). Ectopic X-gal staining in the hindbrain (asterisk) of embryos carrying either wild-type or mutant SBE3 transgenes was detected in equal frequency and thus served as an internal staining control. The ratio of embryos exhibiting reproducible *Shh*-like reporter activity over the total number of transgenic embryos is indicated for each construct.

elements. Other methods that rely solely on sequence conservation are often less efficient at identifying enhancers because the process of evaluating evolutionary conserved regions (ECRs) for regulatory activity is random (Nobrega et al., 2003; Woolfe et al., 2005). This is particularly problematic when the enhancers are spread out across a large genomic landscape, as with *Shh*, because the number of highly conserved sequences can be large. Using our methodology, we were able to rapidly identify several novel enhancers distributed over 400 kb from the *Shh* promoter that regulate *Shh* transcription in the embryonic forebrain. Together with our previous analysis of locally acting *Shh* regulatory sequences, our studies have uncovered a total of six enhancers that are capable of directing reporter activity to all sites of *Shh* transcription in the ventral neural tube of the mouse embryo (Fig. 8; Table 1).

One limitation for all approaches that use multi-species sequence comparison to uncover cis-acting regulatory elements is the potential to overlook enhancers whose sequence or position in the genome has not been conserved across phyla (Frazer et al., 2004). This may explain why we were unable to identify the precise location of the *Shh* notochord element on Bac 447L17. In such instances, deletion mapping may be a more suitable strategy to narrow down the location of the enhancer of interest.

Two enhancers activate *Shh* floor plate transcription via distinct mechanisms

Shh expression in the floor plate is induced by Shh signaling from the underlying notochord (reviewed by Placzek and Briscoe, 2005). According to our results, this homeogenetic activation of *Shh* transcription within the floor plate of the spinal cord and hindbrain is dependent on the redundant activities of SFPE1 and SFPE2. Both floor plate enhancers are reliant on Gli2, as *Shh* expression and consequently floor-plate formation are absent in *Gli2*$^{-/-}$ embryos (Ding et al., 1998; Matise et al., 1998). However, Gli2 is unlikely to be regulating SFPE1 and SFPE2 directly as Gli-binding sites were not identified in sequences mediating these enhancers (Epstein et al., 1999; Jeong and Epstein, 2003). Instead, Foxa2 and Nkx6 proteins are probably acting as transcriptional intermediaries downstream of Gli2 as conserved binding sites matching their consensus are required for SFPE2 activity (Jeong and Epstein, 2003). To determine whether SFPE1 and SFPE2 are under the influence of the same transcription factors, we screened the 300 bp segment mediating SFPE1 for Nkx6- and Foxa2-binding sites. However, no sequences even remotely matching their consensus binding sites were identified (data not shown). Taken together, these findings suggest that although SFPE1 and SFPE2 possess equivalent functions, the transcription factors responsible for their activation appear to be different.

Multiple enhancers regulate *Shh* expression in the ventral forebrain

Ventral midline cells of the forebrain exhibit distinct properties from floor-plate cells in more posterior regions of the CNS and are thought to emerge from a discrete pool of progenitors through slightly different inductive signaling events (Placzek and Briscoe, 2005). Lineage tracing studies performed in the chick indicate that ventral midline cells of the forebrain derive from a small domain, termed 'area a', situated immediately rostral to Hensen's node (Garcia-Martinez et al., 2004; Patten et al., 2003). The prechordal plate that underlies 'area a' cells is the source of Shh and Nodal, the combination of which is thought to convert 'area a' progenitors into ventral midline cells (Patten et al., 2003). This contrasts with the origin of floor-plate precursors in more posterior regions of the embryo, which derive from the node and are dependent on notochord-derived Shh signaling for their formation (Placzek and Briscoe, 2005). Studies in zebrafish suggest that in addition to Nodal and Shh, suppression of Wnt signaling in the ventral midline of the

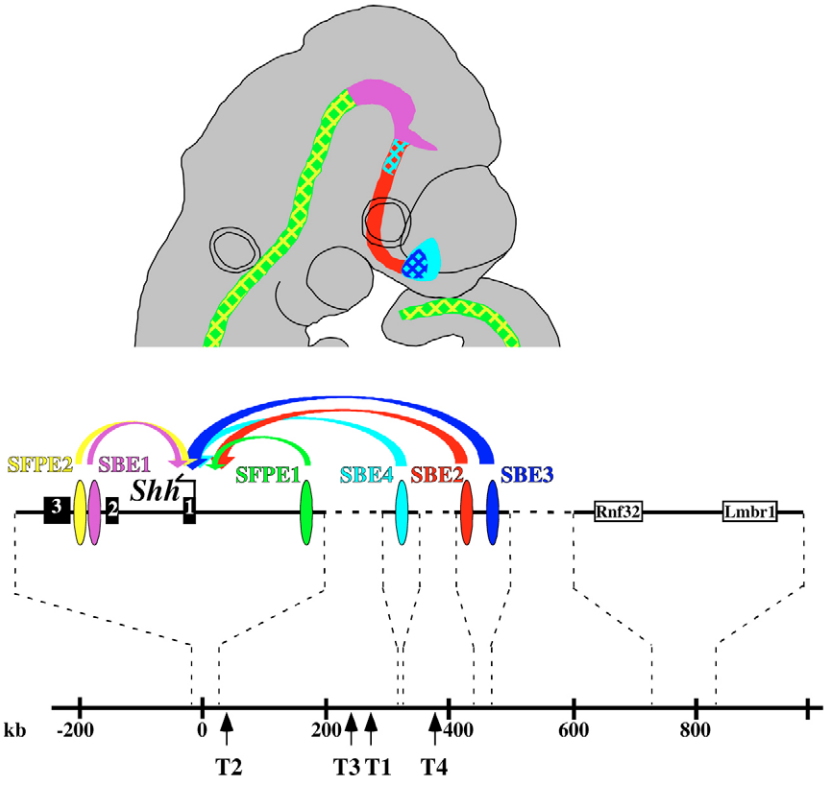

Fig. 8. *Shh* expression in the CNS is controlled by multiple regulatory modules. Schematic view of *Shh* expression in the CNS, color-coded to depict the distinct regulatory elements governing *Shh* transcription along the AP axis of the mouse neural tube. Hatched patterns in the floor plate of the spinal cord and hindbrain, in the p3 domain of the diencephalon and in the svz of the telencephalon represent the sites of *Shh* expression regulated by more than one enhancer. Solid patterns in the ventral midbrain, diencephalon and telencephalon represent sites of *Shh* expression controlled by single regulatory elements. The location of the six CNS enhancers with respect to the *Shh* transcription start site is also indicated. The arrows indicate the position of the translocation breakpoints (T1, T2, T3, T4) identified in individuals with HPE (Roessler et al., 1997).

rostral diencephalon plays a role in distinguishing the hypothalamic infundibulum from floor-plate cells that occupy more posterior regions of the neural tube (Kapsimali et al., 2004). It remains to be confirmed if 'area a' cells and attenuated Wnt signaling are involved in ventral midline development in the mouse forebrain. Nevertheless, the distinct signaling properties of the prechordal plate versus the notochord are likely to play a significant part in distinguishing *Shh* expression in the ventral forebrain from other regions of the CNS as evidenced by the separate enhancers controlling *Shh* transcription in rostral versus caudal region of the neural tube.

The finding that multiple enhancers control *Shh* expression in the forebrain argues that multiple transcription factors must be involved in this regulation. One likely candidate, the Nkx2.1 homeodomain protein, is expressed in the basal telencephalon and rostral region of the ventral diencephalon in a pattern that overlaps with *Shh* (Shimamura et al., 1995). Importantly, *Shh* transcripts were significantly reduced from these domains in *Nkx2.1*$^{-/-}$ embryos (Sussel et al., 1999). Indication that at least one aspect of *Shh* regulation by Nkx2.1 is direct comes from our identification of a functional Nkx2.1-binding site in sequences mediating SBE3. In the absence of this binding site, SBE3 reporter activity was abrogated from the svz of the mge (Fig. 7). SBE2 and SBE4 also direct *Shh*-like reporter activity in a manner that overlaps with Nkx2.1, raising the possibility that Nkx2.1 binds to these regulatory sequences as well. However, as SBE2 and SBE4 control different aspects of *Shh* expression in the forebrain, Nkx2.1 would have to be acting in concert with other transcription factors.

Although the identities of the transcription factors that cooperate with Nkx2.1 to positively regulate *Shh* expression in the forebrain remain unclear, at least one negative regulator, Tbx2, has been proposed. *Shh* is not expressed in the ventral midline of the diencephalon corresponding to the hypothalamic infundibulum, despite the presence of Nkx2.1 in this domain. Instead, *Shh* transcripts are found in two bilateral stripes adjacent to the ventral midline. Recent data in the chick suggest that Tbx2 is activated specifically in the region of the hypothalamic infundibulum in response to Bmp signaling from the prechordal plate and is responsible for the repression of *Shh* in this domain (E. Manning and M. Placzek, personal communication). There is some evidence to suggest that the repression of *Shh* by Tbx2 is direct as the removal of a Tbx-binding site in sequences mediating SFPE2 caused ectopic *lacZ* expression in the ventral midline of the diencephalon (Jeong and Epstein, 2003). Analysis of the crucial sequences and cognate transcription factors required for the activities of SBE2 and SBE4 should provide additional insight to the mechanisms regulating *Shh* transcription in the mouse forebrain.

Blocking access of *Shh* enhancers to their promoter: a potential cause of HPE

HPE is a genetically heterogeneous condition caused predominantly by mutations in *Shh* (Ming and Muenke, 2002). The initial indication for *Shh* as an HPE candidate gene was based on the analysis of two individuals with HPE who were carrying cytogenetically detectable rearrangements involving chromosome 7q36, where *Shh* was shown to map (Belloni et al., 1996). Interestingly, the translocation breakpoints in these individuals did not disrupt *Shh*-coding regions but instead mapped 15 kb and 250 kb distal to the *Shh* gene. It is unlikely that these translocations interrupted other genes in the area as the interval between *Shh* and *Rnf32*, the next distal gene over, is separated by an 840 kb gene desert. Rather, our findings suggest that the cause of HPE in these translocation cases was due to the displacement of *Shh* enhancers from its promoter (Fig. 8). A similar explanation may also account for the HPE phenotype exhibited by the short digits (*Dsh*) mouse mutant, which contains an 11.7 Mb inversion on mouse chromosome 5 with a distal breakpoint 13 kb upstream of *Shh* (Niedermaier et al., 2005).

The positioning of each of the forebrain and axial mesoderm enhancers on the distal side of the translocation breakpoints in the individuals with HPE does little to resolve which elements in particular are required for proper forebrain development. In principle, a reduction in *Shh* transcription from any of the non-redundant enhancers could cause specific phenotypes associated with HPE. The expression of *Shh* in the prechordal plate and ventral forebrain is responsible for establishing patterns of growth and differentiation of neuronal and glial precursors in the rostral CNS, including bifurcation of the telencephalic vesicles and craniofacial structures (Chiang et al., 1996; Rallu et al., 2002; Fucillo et al., 2004). Gain- and loss-of-function studies in several organisms have implicated the source of Shh in the prechordal plate as requisite for ventral forebrain development (reviewed by Hayhurst and McConnell, 2003; Wilson and Houart, 2004). However, as the ventral forebrain is missing in mutants lacking a prechordal plate, the specific contribution of Shh signaling from the ventral forebrain itself remains unclear. Interestingly, in chick embryos, the severity of craniofacial defects resulting from a block in Shh signaling correlated with where and when the signal was disrupted (Marcucio et al., 2005). These studies showed that the expression of *Shh* in the ventral diencephalon and telencephalon had a significant impact on craniofacial morphogenesis. The inactivation of specific enhancers in the mouse that direct *Shh* expression to the ventral diencephalon (SBE2) and ventral telencephalon (SBE4) will provide a unique opportunity to investigate more thoroughly the contribution of Shh from a given source with respect to its role in forebrain and craniofacial development.

The objective of the current study was to identify enhancers controlling *Shh* transcription in the mouse CNS. Coincidently, reporter activity from Bac transgenic embryos was also detected in the gut, axial mesoderm and posterior limb bud, all known sites of *Shh* expression. Previous studies have shown that the *Shh* limb enhancer is located ~1 Mb away from *Shh*-coding sequences and is mutated in cases of preaxial polydactyly in humans, mice and chickens (Lettice et al., 2003; Sagai et al., 2004; Maas and Fallon, 2005). The types of defects caused by alterations in *Shh* limb enhancer activity are representative of the growing number of developmental anomalies that result from the disruption of cis-acting regulatory sequences, either by mutation or position effect resulting from chromosomal rearrangements (Kleinjan and van Heyningen, 2005; Lettice and Hill, 2005). Undoubtedly, this number will continue to rise with improved methods of identifying functional regulatory elements. Approaches such as ours will be useful in identifying other long-range enhancers that regulate tissue-specific expression of genes during development.

We thank Dr Jean Richa and his staff at the University of Pennsylvania Transgenic and Mouse Chimeric Facility for their assistance in transgenic mouse production, and Drs Haig Kazazian, Steve Liebhaber and Tom Kadesch for helpful comments on the manuscript. The authors are grateful to Drs Nat Heintz and Denis Duboule for providing the pLD53SCAEB and Tn7βLacZ plasmids, respectively. This work was supported by NIH grant R01 NS39421 from NINDS (D.J.E.), a Pew Scholar Award in the Biomedical Sciences (D.J.E.) and the Division of Intramural Research, National Human Genome Research Institute, National Institutes of Health (M.M.).

Supplementary material

Supplementary material for this article is available at http://dev.biologists.org/cgi/content/full/133/3/761/DC1

References

Belloni, E., Muenke, M., Roessler, E., Traverso, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H. F., Donis-Keller, H., Helms, C., Hing, A. V. et al. (1996). Identification of *Sonic hedgehog* as a candidate gene responsible for holoprosencephaly. *Nat. Genet.* **14**, 353-356.

Briscoe, J. and Ericson, J. (2001). Specification of neuronal fates in the ventral neural tube. *Curr. Opin. Neurobiol.* **11**, 43-49.

Charron, F., Stein, E., Jeong, J., McMahon, A. P. and Tessier-Lavigne, M. (2003). The morphogen sonic hedgehog is an axonal chemoattractant that collaborates with netrin-1 in midline axon guidance. *Cell* **113**, 11-23.

Chiang, C., Litingtung, Y., Lee, E., Young, K. E., Corden, J. L., Westphal, H. and Beachy, P. A. (1996). Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383**, 407-413.

Ding, Q., Motoyama, J., Gasca, S., Mo, R., Sasaki, H., Rossant, J. and Hui, C.-C. (1998). Diminished Sonic Hedgehog signaling and lack of floor plate differentiation in *Gli2* mutant mice. *Development* **125**, 2533-2543.

Dunn, N. R. , Vincent, S. D., Oxburgh, L., Robertson, E. J. and Bikoff, E. K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development* **131**, 1717-1728.

Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A. and McMahon, A. P. (1993). Sonic hedgehog, a member of a family of putative signaling molecules is implicated in the regulation of CNS and limb polarity. *Cell* **75**, 1417-1430.

Epstein, D. J., McMahon, A. P. and Joyner, A. L. (1999). Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3-dependent and -independent mechanisms. *Development* **126**, 281-292.

Frazer, K. A., Tao, H., Osoegawa, K., de Jong, P. J., Chen, X., Doherty, M. F. and Cox, D. R. (2004). Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res.* **14**, 367-372.

Fuccillo, M., Rallu, M., McMahon, A. P. and Fishell, G. (2004). Temporal requirement for hedgehog signaling in ventral telencephalic patterning. *Development* **131**, 5031-5040.

Garcia-Lopez, R., Vieira, C., Echevarria, D. and Martinez, S. (2004). Fate map of the diencephalon and the zona limitans at the 10-somites stage in chick embryos. *Dev. Biol.* **268**, 514-530.

Gong, S., Yang, X. W., Li, C. and Heintz, N. (2002). Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res.* **12**, 1992-1998.

Hayhurst, M. and McConnell, S. K. (2003). Mouse models of holoprosencephaly. *Curr. Opin. Neurol.* **16**, 135-141.

Jeong, Y. and Epstein, D. J. (2003). Distinct regulators of *Shh* transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. *Development* **130**, 3891-3902.

Jeong, Y. and Epstein, D. J. (2005). Modification and Production of Bac transgenes. Chapter **23**, Unit 11. *Current Protocols in Molecular Biology.* Hoboken (NJ): John Wiley.

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nat. Rev. Genet.* **1**, 20-29.

Kapsimali, M., Caneparo, L., Houart, C. and Wilson, S. W. (2004). Inhibition of Wnt/Axin/beta-catenin pathway activity promotes ventral CNS midline tissue to adopt hypothalamic rather than floorplate identity. *Development* **131**, 5923-5933.

Kleinjan, D. A. and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8-32.

Lettice, L. A. and Hill, R. E. (2005). Preaxial polydactyly: a model for defective long-range regulation in congenital abnormalities. *Curr. Opin. Genet. Dev.* **15**, 294-300.

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.

Lowe, L. A., Yamada, S. and Kuehn, M. R. (2001). Genetic dissection of nodal function in patterning the mouse embryo. *Development* **128**, 1831-1843.

Maas, S. A. and Fallon, J. F. (2005). Single base pair change in the long-range Sonic hedgehog limb-specific enhancer is a genetic basis for preaxial polydactyly. *Dev. Dyn.* **232**, 345-348.

Marcucio, R. S., Cordero, D. R., Hu, D. and Helms, J. A. (2005). Molecular interactions coordinating the development of the forebrain and face. *Dev. Biol.* **284**, 48-61.

Matise, M. P., Epstein, D. J., Park, H. L., Platt, K. P. and Joyner, A. L. (1998). *Gli2* is required for induction of floor plate but not most ventral neurons in the mouse central nervous system. *Development* **125**, 2759-2770.

Ming, J. E. and Muenke, M. (2002). Multiple hits during early embryonic development: digenic diseases and holoprosencephaly. *Am J. Hum. Genet.* **71**, 1017-1032.

Müller, F., Albert, S., Blader, P., Fischer, N., Hallonet, M. and Strahle, U. (2000). Direct action of the nodal-related signal cyclops in induction of sonic hedgehog in the ventral midline of the CNS. *Development* **127**, 3889-3897.

Niedermaier, M., Schwabe, G. C., Fees, S., Helmrich, A., Brieske, N., Seemann, P., Hecht, J., Seitz, V., Stricker, S., Leschik, G. et al. (2005). An inversion involving the mouse Shh locus results in brachydactyly through dysregulation of Shh expression. *J. Clin. Invest.* **115**, 900-909.

Nobrega, M. A., Ovcharenko, I., Afzal, V. and Rubin, E. M. (2003). Scanning human gene deserts for long-range enhancers. *Science* **302**, 413.

Patten, I. and Placzek, M. (2000). The role of Sonic hedgehog in neural tube patterning. *Cell Mol. Life Sci.* **57**, 1695-1708.

Patten, I., Kulesa, P., Shen, M. M., Fraser, S. and Placzek, M. (2003). Distinct modes of floor plate induction in the chick embryo. *Development* **130**, 4809-4821.

Placzek, M. and Briscoe, J. (2005). The floor plate: multiple cells, multiple signals. *Nat. Rev. Neurosci.* **6**, 230-240.

Rallu, M., Machold, R., Gaiano, N., Corbin, J. G., McMahon, A. P. and Fishell, G. (2002). Dorsoventral patterning is established in the telencephalon of mutants lacking both Gli3 and Hedgehog signaling. *Development* **129**, 4963-4974.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S. W., Tsui, L. C. and Muenke, M. (1996). Mutations in the human *Sonic Hedgehog* gene cause holoprosencephaly. *Nat. Genet.* **14**, 357-360.

Roessler, E., Ward, D.E., Gaudenz, K., Belloni, E., Scherer, S. W., Donnai, D., Siegel-Bartelt, J., Tsui, L. and Muenke, M. (1997). Cytogenetic rearrangements involving the loss of the Sonic Hedgehog gene at 7q36 cause holoprosencephaly. *Hum. Genet.* **100**, 172-181.

Rohr, K. B., Barth, K. A., Varga, Z. M. and Wilson, S. W. (2001). The nodal pathway acts upstream of hedgehog signaling to specify ventral telencephalic identity. *Neuron* **29**, 341-351.

Ruiz i Altaba, A., Nguyen, V. and Palma, V. (2003). The emergent design of the neural tube: prepattern, SHH morphogen and GLI code. *Curr. Opin. Genet. Dev.* **13**, 513-521.

Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. and Shiroishi, T. (2005). Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797-803.

Shimamura, K., Hartigan, D. J., Martinez, S., Puelles, L. and Rubenstein, J. L. (1995). Longitudinal organization of the anterior neural plate and neural tube. *Development* **121**, 3923-3933.

Spitz, F., Gonzalez, F. and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405-417.

Sussel, L., Marin, O., Kimura, S. and Rubenstein, J. L. (1999). Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359-3370.

Thomas, J. W., Touchman, J. W., Blakesley, R. W., Bouffard, G. G, Beckstrom-Sternberg, S. M., Margulies, E. H., Blanchette, M., Siepel, A. C., Thomas, P. J., McDowell J. C. et al. (2003). Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788-793.

Watada, H., Mirmira, R. G., Kalamaras, J. and German, M. S. (2000). Intramolecular control of transcriptional activity by the NK2-specific domain in NK-2 homeodomain proteins. *Proc. Natl. Acad. Sci. USA* **97**, 9443-9448.

Wilson, S. W. and Houart, C. (2004). Early steps in the development of the forebrain. *Dev. Cell* **6**, 167-181.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J., Cooke, J. E. and Elgar, G. (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PloS Biol.* **3**, 116-130.

Zoltewicz, J. S., Stewart, N. J., Leung, R. and Peterson, A. S. (2004). Atrophin 2 recruits histone deacetylase and is required for the function of multiple signaling centers during mouse embryogenesis. *Development* **131**, 3-14.