Cis-regulatory architecture of a brain signaling center predates the origin of chordates

Yao Yao, Paul J Minor, Ying-Tao Zhao, Yongsu Jeong, Ariel M Pani, Anna N King, Ariel M Pani, Anna N King, Lin Gan, Wellington V Cardoso, François Spitz, Christopher J Lowe & Douglas J Epstein

Genomic approaches have predicted hundreds of thousands of tissue-specific cis-regulatory sequences, but the determinants critical to their function and evolutionary history are mostly unknown. Here we systematically decode a set of brain enhancers active in the zona limitans intrathalamica (ZLI), a signaling center essential for vertebrate forebrain development via the secreted morphogen Sonic hedgehog (Shh). We apply a de novo motif analysis tool to identify six position-independent sequence motifs together with their cognate transcription factors that are essential for ZLI enhancer activity and Shh expression in the mouse embryo. Using knowledge of this regulatory lexicon, we discover new Shh zli enhancers in mice and a functionally equivalent element in hemichordates, indicating an ancient origin of the Shh zli regulatory network that predates the chordate phylum. These findings support a strategy for delineating functionally conserved enhancers in the absence of overt sequence homologies and over extensive evolutionary distances.

Organization of the vertebrate brain into discrete structural and functional regions begins early during embryonic development in response to signaling molecules secreted from localized brain organizing centers. The zli is one such signaling hub in the posterior diencephalon of all vertebrates that specifies the thalamic and prethalamic territories through the release of Shh but is divergent or absent in invertebrate chordate lineages. Central to the understanding of zli formation is how Shh transcription is regulated in this crucial brain signaling center and the extent to which this regulatory mechanism is shared across species.

Comparative sequence analysis is often used to identify conserved noncoding regulatory elements. However, it has become increasingly apparent that not all functionally conserved regulatory elements show clear evidence of DNA sequence homology, a phenomenon that may confound interpretations of their evolutionary origin. Moreover, conventional methods of phylogenetic footprinting do not always capture individual binding sites at nucleotide resolution, especially when long stretches of regulatory sequence are under strong positive selection.

To decipher the regulatory logic of Shh expression in the zli, we adapted a strategy that does not rely on DNA sequence conservation alone but, instead, follows the premise that enhancers with similar spatiotemporal profiles often have common cis-regulatory features. We surveyed the collection of experimentally validated regulatory elements in the VISTA Enhancer Browser for patterns of reporter activity that overlapped those of SBE1, an enhancer located in the second intron of the Shh gene that directs expression to the ventral midbrain, ventroposterior diencephalon and zli (Supplementary Fig. 1a–c). We identified 52 distinct SBE1-like enhancers scattered throughout the mouse and human genomes (Supplementary Fig. 2). Each of these enhancers is near at least one gene transcribed in the region of the mid-diencephalic organizer in mice according to the RNA-seq profile of SBE1-positive cells at embryonic day E10.5 (Supplementary Fig. 3 and Supplementary Table 1).

To determine whether the SBE1-like enhancers possess a common cis-regulatory signature, we applied the Weeder algorithm and performed an unbiased search for shared DNA sequence motifs in seven enhancers that most closely match the SBE1 activity pattern, including SBE1 (Fig. 1a). This approach identified five motifs (motifs 1–5) that were enriched in the seven enhancers in comparison to random genomic sequence (Fig. 1b). The motifs were highly conserved across vertebrate phyla, suggestive of their functional importance. In an alignment of the SBE1 sequences from different species, the five motifs were clustered in a 116-bp homology block extending from human to zebrafish that was both necessary and sufficient for full enhancer activity (Fig. 1c and Supplementary Fig. 1).

We also searched the set of SBE1-like enhancers for over-represented transcription factor binding sites present in the JASPAR and UniPROBE databases. Three of the five motifs identified by Weeder matched consensusbinding sites for transcription factors, several of which are expressed in mouse embryos where SBE1 is active (Fig. 1b and Supplementary Fig. 4). This analysis also uncovered an additional over-represented motif (motif 6) that was missed by Weeder, presumably because of its more unknown function.

1Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. 2Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, California, USA. 3Department of Genetic Engineering, College of Life Sciences and Graduate School of Biotechnology, Kyung Hee University, Yongin-si, Republic of Korea. 4Department of Developmental Biology and European Molecular Biology Laboratory, Heidelberg, Germany. 5Department of Ophthalmology, University of Rochester Medical Center, Rochester, New York, USA. 6Columbia Center for Human Development, Department of Medicine, Pulmonary Allergy Critical Care, Columbia University Medical Center, New York, New York, USA. 7These authors contributed equally to this work. Correspondence should be addressed to D.J.E. (epsteind@mail.med.upenn.edu).

Received 28 January; accepted 11 March; published online 11 April 2016; corrected after print 16 May 2016; doi:10.1038/ng.3542
stringent criteria for finding DNA sequence matches. The six motifs did not display any consistent order or spacing in the SBE1-like enhancers, suggesting that they follow a flexible arrangement model observed for other tissue-specific enhancers.\textsuperscript{19,24,25} The co-occurrence of motifs 1–6 was significantly higher in a larger set of SBE1-like enhancers (\(n = 46\)) than in random genomic sequence matched for GC content and length (\(P < 0.05\), Welch’s two-sample \(t\) test) and in a set of 172 heart enhancers from the VISTA Enhancer Browser (\(P < 2.2 \times 10^{-16}\)).

\[
\begin{align*}
\text{mmSBE1} & \quad \text{hs194} & \quad \text{hs593} & \quad \text{hs779} & \quad \text{hs1093} & \quad \text{hs1180} & \quad \text{hs1391} \\
\text{VM} & \quad \text{ZLI} & \quad \text{VPD} & \quad \text{Chr. 2: 103,792,328–103,793,819} & \quad \text{Chr. 18:} & \quad \text{Chr. 14: 37,726,340–37,727,348} & \quad \text{Chr. 6: 3,349,367–3,352,257} \\
\text{Human} & \quad \text{Mouse} & \quad \text{Opossum} & \quad \text{Chicken} & \quad \text{Zebrafish} & \quad \text{Coelacanth} & \quad \text{X. tropicalis} \\
1 & \quad \text{AGA(A/T)GTTTAA} & \quad \text{AGAATT} & \quad \text{TTGCGAGATA(AA)} & \quad \text{AGACAT} & \quad \text{G(A/G)GAT} & \quad \text{CAATTA(A/G/T)} \\
2 & \quad \text{AGA(A/T)GTTTAA} & \quad \text{AGAATT} & \quad \text{TTGCGAGATA(AA)} & \quad \text{AGACAT} & \quad \text{G(A/G)GAT} & \quad \text{CAATTA(A/G/T)} \\
3 & \quad \text{AGA(A/T)GTTTAA} & \quad \text{AGAATT} & \quad \text{TTGCGAGATA(AA)} & \quad \text{AGACAT} & \quad \text{G(A/G)GAT} & \quad \text{CAATTA(A/G/T)} \\
4 & \quad \text{AGA(A/T)GTTTAA} & \quad \text{AGAATT} & \quad \text{TTGCGAGATA(AA)} & \quad \text{AGACAT} & \quad \text{G(A/G)GAT} & \quad \text{CAATTA(A/G/T)} \\
5 & \quad \text{AGA(A/T)GTTTAA} & \quad \text{AGAATT} & \quad \text{TTGCGAGATA(AA)} & \quad \text{AGACAT} & \quad \text{G(A/G)GAT} & \quad \text{CAATTA(A/G/T)} \\
6 & \quad \text{AGA(A/T)GTTTAA} & \quad \text{AGAATT} & \quad \text{TTGCGAGATA(AA)} & \quad \text{AGACAT} & \quad \text{G(A/G)GAT} & \quad \text{CAATTA(A/G/T)} \\
\end{align*}
\]

\[\text{Otx2 + Barhl2} \quad \text{Fold change in luciferase activity} \quad \text{Otx2 + Barhl2} \quad \text{Fold change in luciferase activity} \quad \text{Yap1 + Tead2} \quad \text{Fold change in luciferase activity} \]

\[\text{Barhl2/Hmx1, Hmx2, Hmx3, Max3} \quad \text{Barhl2/Hmx1, Hmx2, Hmx3, Max3} \quad \text{Barhl2/Hmx1, Hmx2, Hmx3, Max3} \]

Figure 1 A common cis-regulatory signature in SBE1-like enhancers. (a) Heads of transgenic embryos (E11.5) showing X-gal staining from reporter constructs for mouse SBE1 and selected human SBE1-like enhancers from the VISTA Enhancer Browser (VM, ventral midbrain; VPD, ventroposterior diencephalon). Chromosome positions are indicated (NCBI37/mm9 (mouse) and GRCh37/hg19 (human)). (b) Shared motifs identified by Weeder (motifs 1–5) and by JASPAR and UniPROBE (motif 6) that are significantly enriched in SBE1-like enhancers. Transcription factors matching a particular motif that are also expressed in the SBE1 domain appear in blue. (c) DNA sequence alignment (MAFFT version 7) of a core region of SBE1 from representative vertebrate species showing the positions of shared motifs. Deeply conserved SBE1 nucleotides are shaded in gray. (d–g) Luciferase reporter assays performed in COS-1 cells transfected with reporter constructs encoding mouse homologs of SBE1-like enhancers (x axes) and expression vectors encoding Otx2 (d), Barhl2 (e), Otx2 and Barhl2 (f), and Yap1 and Tead2 (g). Comparison of luciferase activity is to the activity observed in cells transfected with empty vector (dashed lines). Bar graphs are color-coded to match transcription factors with their binding sites (motifs). Black bars represent mutant SBE1 reporter constructs in which a particular motif was deleted. Lower concentrations of the Otx2 and Barhl2 expression vectors were used in f than in d and e to demonstrate the synergy between these transcription factors. Each bar represents the average of at least three experiments performed in triplicate. (h–j) ChIP-qPCR performed with chromatin isolated from E10.5 embryonic mouse brain (B) and limb bud (L) for Otx2 (h) and from transfected COS-1 cells for Barhl2 (i) and Tead2 (j). qPCR results represent an average of at least three biological replicates. The negative-control (NC) and positive-control (PC) primers in h amplify sequence upstream of SBE1 and within an Emx2 forebrain enhancer, respectively. Error bars in all graphs represent standard deviation of the mean (s.d.m.): *\(P < 0.05\), **\(P < 0.01\), ***\(P < 0.001\), Student’s \(t\) test.
Our criteria for pairing motifs with their candidate transcription factors included expression of the transcription factors in the SBE1 domain and previous indication that they have a role in Shh regulation and/or zli formation. Motifs 1 and 6 correspond to recognition sequences for homeodomain proteins of the Prd and NKL subclasses, respectively.26 Otx1 and Otx2 are the best candidates for recruitment by motif 1 given their roles in mid-diencephalic development and Shh expression27–29. Similarly, Barhl2 is a potential transcription factor for motif 6 on the basis of its requirement for zli formation in Xenopus laevis embryos.30 Foxa2 was previously shown to bind motif 3, which is necessary for SBE1 activity in the ventral midbrain but not the zli22 and, therefore, will not be discussed further here. Of several candidate binding factors for motif 2, we hypothesize that TEA-domain family member 2 (Tead2), a key mediator of Hippo signaling, is recruited to this site. Tead and its co-transcriptional activation partner Yap are dependent on Wnt and/or Shh signaling in various biological contexts.31,32

To determine whether a candidate transcription factor was capable of regulating SBE1-like enhancers through a given motif, we performed luciferase reporter assays in COS-1 cells. Otx2 induced significant luciferase expression from all seven SBE1-like reporter constructs but not when motif 1 was deleted from mouse SBE1 (Fig. 1d). The core Otx1- and Otx2-binding site in motif 1 (underlined sequence in AAGATTAAG) is preferentially flanked on each side by adenine nucleotides, which when mutated blocked Otx2 binding and activation of the SBE1-luciferase construct, suggesting a context-dependent role in zli gene regulation (Supplementary Fig. 5).

Although Barhl2 triggered only a modest response from SBE1-like enhancers (Fig. 1e), the combined action of Barhl2 and Otx2 resulted in a synergistic induction of reporter activity from most enhancers containing motifs 1 and 6 (Fig. 1f). Therefore, crosstalk between Otx2 and Barhl2 may mobilize a subset of SBE1-like enhancers. We also observed that the Tead2–Yap1 co-activation complex stimulated transcriptional responses for most SBE1-like enhancers, including mouse SBE1, which depended on motif 2 and a second Tead2-binding site (motif 2.1), located 141 bp downstream of motif 2 (Fig. 1g). No other transcription factor combinations tested showed synergistic interactions (Supplementary Fig. 6).

We next performed chromatin immunoprecipitation (ChiP) to examine the occupancy of candidate transcription factors at their respective binding sites in SBE1-like enhancers. Chromatin isolated from embryonic brain but not limb bud extracts was enriched for Otx2 at all seven SBE1-like enhancers in their native genomic context (Fig. 1h). Barhl2 and Tead2 were also recruited to a subset of SBE1-like enhancers containing the corresponding motifs in COS-1 cells (Fig. 1i,j). These findings suggest that SBE1-like enhancers are directly regulated by a transcription factor collective comprising Otx2, Barhl2 and Tead2.

To assess the in vivo requirement for the SBE1 transcription factor collective, we performed transgenic mouse reporter assays with SBE1-lacZ constructs containing mutations in motif 1 (Otx), motifs 2 and 2.1 (Tead), or motif 6 (Barhl). X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) staining was greatly compromised in the zli of embryos carrying the SBE1–motif1-lacZ transgene (96% reduction in staining along the zli length as compared to embryos with the SBE1-lacZ transgene, P < 0.0001; Fig. 2a,b,c). Deletions of motif 6 and motif 2/2.1 also resulted in a significant loss of staining in the zli as compared to control embryos with the SBE1-lacZ transgene (46% reduction, P < 0.001 and 32% reduction, P < 0.01, respectively; Fig. 2a,c,d). A similar reduction in zli staining was observed for constructs with deletions in the two orphan motifs, motifs 4 and 5 (Supplementary Fig. 7). These results are further supported by genetic studies, which showed a selective reduction in zli Shh expression in Barhl2−/− embryos and embryos with conditional loss of Yap1 (loxP-flanked Yap1 and

Figure 2 In vivo requirement for the SBE1 transcription factor collective. (a–d) X-gal staining of transgenic embryos expressing wild-type SBE1-lacZ (a) or the mutant versions SBE1-motif1-lacZ (b), SBE1-motif6-lacZ (c) and SBE1-motif2/2.1-lacZ (d) at E10.5. The extent of zli staining is indicated by the length of each red bracket. The number of stained embryos out of the total number of embryos carrying a given transgene is indicated. Scale bars, 1 mm. (e) Schematic of an E10.5 embryo demonstrating the approach for measuring the spatial distribution of X-gal staining or Shh expression in the zli with respect to head size. (f–j) Whole-mount in situ hybridization for Shh in control embryos (f), Barhl2−/− embryos (n = 4) (g), and embryos with conditional disruption of one (h) or both (i) alleles of Yap1 (n = 4) at E10.5. Scale bars, 1 mm. (j) Quantification of the spatial distribution of X-gal staining (blue bars) or Shh expression (purple bars) in the zli normalized to head size. WT, wild type. Error bars represent s.d.m.: **P < 0.01, ***P < 0.001, ****P < 0.0001, Student’s t test.
Cre recombinase expression directed from the Shh promoter, or Shh<sup>cre</sup>/Yap1<sup>fl</sup>, along with earlier results in Otx1–Otx2 mutant embryos<sup>27–29</sup> (Fig. 2f–j). Together, these data validate the in vivo contribution of the SBE1 transcription factor collective in the direct control of Shh expression in the zli.

In earlier work, we reported that mouse embryos homozygous for a targeted deletion of SBE1 (Shh<sup>ΔSBE1/ΔSBE1</sup>) failed to maintain Shh transcription in the basal plate of the rostral midbrain and the caudal diencephalon after E10.0 yet retained expression in the zli<sup>33</sup> (Fig. 3h). This observation implies the existence of another enhancer that functions independently of, or redundantly with, SBE1 to promote Shh expression in the zli. We sought to identify the missing Shh regulatory sequence in zli using knowledge of the shuffled motif arrangement typified by SBE1-like enhancers.

We surveyed a 1-Mb interval encompassing Shh for histone modifications (monomethylation of histone H3 at lysine 4 (H3K4me1) and acetylation of histone H3 at lysine 27 (H3K27ac)) associated with active regulatory sequences using Encyclopedia of DNA Elements (ENCODE) data from E14.5 brain<sup>34</sup> (Fig. 3a). Most of our previously identified Shh brain enhancers, including SBE1, showed H3K4me1 and H3K27ac enrichment. We searched the remaining peaks for evidence of the SBE1 motif signature and identified a single region located 784 kb upstream of Shh, within the penultimate intron of the Lmbr1 gene, that contained a cluster of permuted motifs as compared to SBE1 in the absence of any other overt sequence homology (Fig. 3a and Supplementary Fig. 8a). We tested the 1.9-kb sequence under the peak, designated SBE5, in a transgenic reporter assay and observed embryos with consistent X-gal staining in the ventral midbrain, ventroposterior diencephalon and zli reminiscent of SBE1 activity (Fig. 3f). SBE5 performed equivalently to SBE1 in all cell-based reporter and ChIP assays using components of the SBE1 transcription factor collective, demonstrating that SBE5 is also directly controlled by Otx2, Barhl2 and Tead2 (Fig. 3b–e).

Notably, Shh expression was only partially attenuated in the zli of mouse embryos homozygous for a 228-kb deletion encompassing SBE5 (Shh<sup>ΔSBE5/ΔSBE5</sup>; Fig. 3g,i). Yet, in mutants lacking both SBE1 and SBE5 (Shh<sup>ΔSBE1/ΔSBE1ΔSBE5/ΔSBE5</sup>), Shh transcription was completely eliminated from the ventral midbrain, ventroposterior diencephalon and zli (Fig. 3g,j). Rescue of Shh expression with an ShhP1 transgene comprising the Shh ORF and several Shh regulatory sequences, including SBE1, suggests that this phenotype was caused by loss of Shh enhancers rather than deletion of other coding or noncoding sequence elements potentially involved in Shh regulation (Supplementary Fig. 9). In further support of this claim, we observed that a 2-kb deletion of SBE5 generated by CRISPR/Cas9 technology had the same effect on Shh expression in the zli as the larger SBE5 deletion allele (Supplementary Fig. 9). From these results, we conclude that SBE1 and SBE5 function in a partially redundant manner to regulate Shh zli expression and that the activities of these two enhancers are achieved through similar cis and trans determinants (see the model in Supplementary Fig. 10).

Some studies have proposed that the origins of vertebrate brain signaling centers, including the zli, date to early vertebrates, concurrent with increases in brain complexity, whereas others have contended a more ancient deuterostome origin predating the diversification of chordates<sup>35,36</sup>. Studies performed in the hemichordate <i>Saccoglossus kowalevskii</i> support the latter hypothesis, as they showed patterns of gene expression for many signaling ligands and transcription factors similar to those in vertebrate brain signaling centers<sup>36</sup>. <i>S. kowalevskii</i> <i>hedgehog</i> (<i>hh</i>) is expressed in a narrow band of cells at the proboscis–collar boundary that appear zli-like in character in relation to surrounding genes<sup>36</sup>. To determine whether this pattern of <i>hh</i> expression is governed by a cis-regulatory mechanism similar to the vertebrate counterpart, we searched the <i>hh</i> locus for evidence of the SBE1 motif signature and identified a 1.1-kb region in the second intron that contained all six motifs in the absence of any
The ancient origin of SBE1 predates the chordate phylum. (a) Schematic of the Shh and hh gene structures in mouse and S. kowalevskii (acorn worm), respectively, showing the position of SBE1 (blue oval) in the second intron of both species. The shuffled arrangement of the SBE1-like motifs (colored boxes) within the 1.1-kb skSBE1 sequence is shown. (b) Otx2 and Yap–Tead2 but not Barhl2 were sufficient to activate skSBE1-driven luciferase activity in cotransfection assays performed in COS-1 cells. Error bars represent s.d.m.; 

Ascidian Zebrafish

4

3

lacZ

VOLUME 48 | NUMBER 5 | MAY 2016

We next tested the activity of mmSBE1 and mmSBE5 in S. kowalevskii embryos. The skSBE1, mmSBE1 and mmSBE5 constructs each drove mosaic expression of mNeonGreen in a narrow line of cells at the prospective proboscis–collar boundary, partially recapitulating the endogenous domain of hh expression (Fig. 4c–h) that is proposed to have a patterning role homologous to that of Shh expression in the zli of vertebrates. Embryos injected with a negative-control construct lacking an enhancer showed no reporter activity (Fig. 4i). These data demonstrate that SBE1 elements from mice and hemichordates possess functionally conserved, species-specific regulatory activity in non-homologous structures.

Our finding that SBE1-like enhancers have shuffled binding sites prompted us to reevaluate SBE1 motif conservation in basal chordates. We screened the second intron of hh for evidence of the shuffled motif arrangement in amphioxus (cephalochordate), ascidian (tunicate) and lamprey (basal vertebrate). Organisms that possess the SBE1 motif cluster (lamprey and all jawed vertebrates) express Shh in a delineated domain of the central nervous system (CNS) that defines the zli, whereas organisms without this motif cluster (amphioxus and ascidians) lack hh expression in a homologous region, suggesting secondary loss of the zli (Fig. 4j). This observation is consistent with data showing that the second intron of the amphioxus hh gene lacks enhancer activity in the zli.

These results support the hypothesis of deep homology of a zli regulatory cassette that used SBE1 in an ancient deuterostome to activate hh in a narrow band of ectodermal cells in the anterior half of the embryo. Our results suggest that early chordates would have inherited SBE1 from this deuterostome ancestor but SBE1 was subsequently lost in the invertebrate chordate lineages. Following the diversification of hedgehog ligands in vertebrates, SBE1 was maintained in the second intron of the Shh gene and used to activate its transcription in a narrow band of cells in the caudal forebrain, thus establishing the zli as a brain signaling center. The gain of SBE5 in other sequence homology (Fig. 4a and Supplementary Fig. 8b). The motif arrangement in S. kowalevskii SBE1 (skSBE1) was once again shuffled in comparison to that in mouse SBE1 (mmSBE1), yet the sequence of a given motif differed by no more than a single nucleotide from its mouse equivalent. In cell-based reporter assays, the skSBE1 element was activated by Otx2 but by the Tead2–Yap1 but not the Barhl2–Otx2 combination, for unknown reasons (Fig. 4b). Remarkably, mouse embryos expressing the skSBE1–lacZ transgene displayed X-gal staining in the ventral midbrain, ventroposterior diencephalon and zli (arrows). Ectopic staining outside of these domains is likely due to the site of transgene integration. Scale bars, 1 mm. (e) hh expression in a hemichordate embryo at 48 hours post-fertilization. Scale bar, 100 µm. (f–i) Transgenic S. kowalevskii embryos expressing mNeonGreen in a narrow band of cells at the prospective proboscis–collar boundary from skSBE1 (n = 6/60 injected) (f); mmSBE1 (n = 5/55 injected) (g); mmSBE5 (n = 5/15 injected) (h); and control (gbx promoter only; n = 0/60 injected) (i) reporter constructs. Scale bars, 100 µm. (j) The evolutionary trajectory of Shh expression in the zli and hh expression in zli-like structures (black arrows) correlates with presence of the SBE1 motif cluster.

Figure 4 The ancient origin of SBE1 predates the chordate phylum. (a) Schematic of the Shh and hh gene structures in mouse and S. kowalevskii (acorn worm), respectively, showing the position of SBE1 (blue oval) in the second intron of both species. The shuffled arrangement of the SBE1-like motifs (colored boxes) within the 1.1-kb skSBE1 sequence is shown. (b) Otx2 and Yap–Tead2 but not Barhl2 were sufficient to activate skSBE1-driven luciferase activity in cotransfection assays performed in COS-1 cells. Error bars represent s.d.m.; *** P < 0.001, Student’s t test; NS, not significant. Each bar represents the average of at least three experiments performed in triplicate. (c, d) Transgenic mouse embryos expressing mouse SBE1–lacZ (c) and skSBE1–lacZ (d) reporter constructs at E10.5 show similar patterns of X-gal staining in the ventral midbrain, ventroposterior diencephalon and zli (arrows). Ectopic staining outside of these domains is likely due to the site of transgene integration. Scale bars, 1 mm. (e) hh expression in a hemichordate embryo at 48 hours post-fertilization. Scale bar, 100 µm. (f–i) Transgenic S. kowalevskii embryos expressing mNeonGreen in a narrow band of cells at the prospective proboscis–collar boundary from skSBE1 (n = 6/60 injected) (f); mmSBE1 (n = 5/55 injected) (g); mmSBE5 (n = 5/15 injected) (h); and control (gbx promoter only; n = 0/60 injected) (i) reporter constructs. Scale bars, 100 µm. (j) The evolutionary trajectory of Shh expression in the zli and hh expression in zli-like structures (black arrows) correlates with presence of the SBE1 motif cluster.

Our finding that SBE1-like enhancers have shuffled binding sites prompted us to reevaluate SBE1 motif conservation in basal chordates. We screened the second intron of hh for evidence of the shuffled motif arrangement in amphioxus (cephalochordate), ascidian (tunicate) and lamprey (basal vertebrate). Organisms that possess the SBE1 motif cluster (lamprey and all jawed vertebrates) express Shh in a delineated domain of the central nervous system (CNS) that defines the zli, whereas organisms without this motif cluster (amphioxus and ascidians) lack hh expression in a homologous region, suggesting secondary loss of the zli. This observation is consistent with data showing that the second intron of the amphioxus hh gene lacks enhancer activity in the zli.

These results support the hypothesis of deep homology of a zli regulatory cassette that used SBE1 in an ancient deuterostome to activate hh in a narrow band of ectodermal cells in the anterior half of the embryo. Our results suggest that early chordates would have inherited SBE1 from this deuterostome ancestor but SBE1 was subsequently lost in the invertebrate chordate lineages. Following the diversification of hedgehog ligands in vertebrates, SBE1 was maintained in the second intron of the Shh gene and used to activate its transcription in a narrow band of cells in the caudal forebrain, thus establishing the zli as a brain signaling center. The gain of SBE5 in...
vertebrates, whether by duplication and subsequent rearrangement of SBE1, binding site turnover of preexisting sequence or some other means, is thought to buffer Shh expression in the zili.

In summary, our study provides a framework for decoding coordinate enhancers that is generally applicable to other tissue-specific regulatory sequences\(^1\)–\(^4\). We demonstrate the feasibility of identifying enhancers with similar functions that lack obvious sequence conservation, either in the same organism or in ones with disparate anatomies and separated by hundreds of millions of years of evolution. Applying our approach to other well-characterized cis-regulatory modules in diverse taxa may provide additional insights into genomic mechanisms underlying evolutionary change or stasis in gene regulation.

**METHODS**

Methods and any associated references are available in the **online version of the paper**.

**Accession codes.** RNA-seq data presented in this study have been deposited in the Gene Expression Omnibus (GEO) under accession GSE78005.

**Note:** Any Supplementary Information and Source Data files are available in the **online version of the paper**.

**ACKNOWLEDGMENTS**

We thank J. Richa and his staff at the Transgenic and Chimeric Mouse Facility (Perelman School of Medicine, University of Pennsylvania) for assistance with transgenic mouse generation. We also thank S. Liebhaber, K. Kaestner, C. Brown, K. Zaret and members of the Epstein laboratory for helpful discussions and comments on the manuscript. This work was funded by grants from the National Institutes of Health, KO1 NS039421 (D.E.) and R21 EY023104 (L.G.), and the National Science Foundation, 1258169 (C.J.L.), and by a predoctoral fellowship from the Louis-Jeantet Foundation (O.S.).

**AUTHOR CONTRIBUTIONS**

Y.Y. and D.J.E. conceived the project, designed the experiments and wrote the manuscript. Y.Y. performed the cotransfection, transgenic mouse, gene expression and ChIP assays. P.J.M. performed the transgenic hemichordate reporter assays. Y.J. performed the transgenic mouse reporter assays with core region constructs. Y.-T.Z. performed the statistical analysis. Y.Y. and A.N.K. performed the motif analysis, A.M.P. and C.I.L. provided reagents and advice on the hemichordate experiments. Y.Y., L.G., O.S., W.V.C. and F.S. generated mutant mouse lines and provided embryos.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at [http://www.nature.com/reprints/index.html](http://www.nature.com/reprints/index.html).

---

1. Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
2. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–283 (2011).
3. Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
4. Yue, F. et al. A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 359–364 (2014).
5. Keckler, C. & Lumsden, A. Hedgehog signaling from the ZLI regulates diencephalic specification and development. *Dev. Biol.* **312**, 1242–1249 (2007).
6. Vieira, C. & Martinez, S. Sonic hedgehog from the basal plate and the zona limitans intrathalamica exhibits differential activity on diencephalic molecular regionalization in the early chick embryo. *Development* **134**, 129–139 (2006).
7. Hébert, J.M. & Fishel, G. The genetics of early telencephalon patterning: some assembly required. *Nat. Rev. Neurosci.* **9**, 678–685 (2008).
8. Jessell, T.M. Neuronal specification in the spinal cord: inductive signals and transcriptional cascades. *Nat. Rev. Genet.* **1**, 20–29 (2000).
9. Schlopp, S. & Lumsden, A. Building a bridgelike chamber: development of the thalamus. *Trends Neurosci.* **33**, 373–380 (2010).
10. Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I. & Hardison, R.C. Cross-species sequence comparisons: a review of methods and available resources. *Genome Res.* **10**, 65–122 (2000).
11. Arnold, C.D. et al. Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* **46**, 685–692 (2014).
12. Episcopo, C., Visel, A., Vento-Torres, M.I., Grinnell, A.F. & Bellen, H.J. Comparative enrichment reveals conserved and new Drosophila enhancers. *Genome Res.* **22**, 1600–1613 (2012).
13. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, DB8–D92 (2007).
14. Epstein, D.J., McMahon, A.P. & Joyner, A.L. 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 (1999).
15. Pavesi, G., Meregalli, P., Mauri, G. & Pesole, G. Weeder: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **32**, W199–W203 (2004).
16. Brown, C.D., Johnson, D.S. & Sidow, A. Functional architecture and evolution of transcriptional elements that drive gene coexpression. *Science* **317**, 1557–1560 (2007).
17. Smith, R.P. et al. Massively parallel decoding of mammalian regulatory sequences supports a flexible model of organismal organization. *Nat. Genet.* **45**, 1021–1028 (2013).
18. Nakano, E., Park, B.H., Shyh-Cha, I., Groudine, M. & Felsenfeld, G. Coordination of chromatin remodeling and transcriptional regulation by Hnf3-dependent and -independent mechanisms. *Science* **324**, 1516–1519 (2009).
19. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L.A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, DB8–D92 (2007).
20. Franzen-Lupi, A. et al. YAPI is amplified and up-regulated in hedgehog-associated medulloblastomas and mediates Sonic hedgehog-driven neural precursor proliferation. *Genes Dev.* **23**, 2729–2741 (2009).
21. Schlopp, S. et al. 8-catenin-driven cancers require a YAPI transcriptional complex for survival and tumorigenesis. *Cell* **151**, 1457–1467 (2012).
22. Jeong, Y. et al. Spatial and temporal requirements for sonic hedgehog in the regulation of thalamic interneuron identity. *Development* **138**, 531–541 (2011).
23. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
24. Irimia, M. et al. Conservation developmental expression of Fzd in chordates and Drosophila and the origin of the zona limitans intrathalamica (ZLI) brain organizer. *EvoDevo* **7**, 7 (2010).
25. Pani, A.M. et al. Ancient deuterostome origins of vertebrate brain signalling centres. *Nature* **483**, 289–294 (2012).
26. Irimia, M. et al. Comparative genomics of the Hedgehog loci in chordates and the origins of Shh regulatory novelties. *Sci. Rep.* **2**, 433 (2012).
27. Sugahara, F. et al. Involvement of Hedgehog and FGF signaling in the lamprey telencephalon: evolution of regionalization and dorsoventral patterning of the vertebrate forebrain. *Development* **138**, 1217–1226 (2011).
28. Takahashi, N., Satoh, Y. & Satoh, N. Expression of hedgehog genes in Ciona intestinalis embryos. *Biol. Rev.* **38**, 216, 235–238 (2002).
29. Shubin, N., Tabin, C. & Carroll, S. Deep homology and the origins of evolutionary novelty. *Nature* **457**, 818–823 (2009).
ONLINE METHODS

Cell-based reporter assays. Mouse homologs of SBE1, SBE5 and human SBE1-like enhancers from the VISTA Enhancer Browser (hs194, hs593, hs779, hs1109, hs1180 and hs1391) were cloned into the pGL4.13 Luciferase reporter vector (luc2/minP, Promega). The skSBE1 element was amplified from S. kowaensis genomic DNA by PCR. SBE1 reporter constructs harboring deletions of motif 1, 4 or 6 were generated by ligating two PCR products immediately flanking each motif. The QuickChange II XL Site-Directed Mutagenesis kit was used to introduce the following SBE1 mutations: Amotif2, Amotif1, Amotif5, SBE1M1.1 and SBE1M1.2. Mouse cDNAs encoding Barhl2, HA-Barhl2 and FLAG-Otx2 were cloned into the pClX3 mammalian expression vector (Life Technologies). The pcDNA3-HA-Tead2 and pcDNA3-HA-Yap1 expression vectors were kindly provided by D. Pan (Johns Hopkins University). The primers used to generate each of the reporter constructs are listed in Supplementary Table 2.

Luciferase reporter assays were performed by cotransflecting (FuGENE 6, Promega) COS-1 cells with 250 ng of an enhancer-driven reporter construct and 200–300 ng of a transcription factor expression vector or empty vector, together with 20 ng of pRL-TK (Promega) as an internal control. Cells were collected 48 h after transfection and assayed for firefly and Renilla luciferase activities (Dual-Luciferase Reporter Assay System, Promega). Enhancer activity is presented as fold induction relative to the activity in cells transfected with empty pcDNA3 expression vector. At least three independent experiments were performed for each reporter construct in triplicate. Cells were not tested for mycoplasma.

Chromatin immunoprecipitation. The midbrain, caudal diencephalic region (including the zli) and forelimb buds were dissected in DMEM (with 10% FBS) from approximately 25–30 E10.5 embryos, pooled into separate brain and forelimb fractions, homogenized into small pieces and cross-linked with 1% paraformaldehyde for 15 min at room temperature with shaking. ChIP was performed essentially as described using 6 μg of antibody to Otx2 (Abcam, ab21990) or IgG (Cell Signaling Technology, 2729). qPCR was conducted as described using the primer sequences listed in Supplementary Table 2. The positive-control primers used for Figures 1h and 3c amplify a DNA fragment from an Ems2 forebrain enhancer bound by Otx2 (ref. 43).

A similar protocol was followed when performing ChIP–qPCR from COS-1 cells (10^7–10^8) cultured in 10-cm plates and cotransfected with 3 μg of SBE1-like enhancer construct and 1 μg of pcDNA3-FLAG-Otx2, pcDNA3-HA-Barhl2 or pcDNA3-HA-Tead2 using 6 μg of antibody to FLAG (Sigma, F7425), HA (kindly provided by G. Blobel, Children's Hospital of Philadelphia) or IgG (Sigma, A4312).

Transgenic mouse reporter assays. SBE1-like enhancers were cloned into a vector containing the Shh promoter, lacZ gene and SV40 poly(A) cassette. Transiently transfected embryonic embryos were generated by pronuclear injection into fertilized mouse eggs derived from the B6SJLF1/J mouse strain (Jackson Laboratories) at the Transgenic and Chimeric Mouse Facility (Perelman School of Medicine, University of Pennsylvania).

Mouse lines. Experiments were performed in accordance with the ethical guidelines of the National Institutes of Health and with the approval of the Institutional Animal Care and Use Committee of the University of Pennsylvania. The TRACER mouse deletion line Del(C1-Z), with a deletion encompassing chr. 5: 29,413,901–29,642,246 (mm9) including SBE5, was generated by Cre-mediated recombination between loxP sites carried by the insertion alleles ShhsS5B1 (ref. 44) and ZZD (T. Akes and F.S., unpublished data), following a previously described strategy.55 For simplicity, the Del(C1-Z) line is referred to herein as ShhS5B1. To generate ShhS5B1 SBE5 double-mutant embryos, the ShhS5B1/−/− embryos harboring the deletion1,2 mutants53. ShhS5B1/−/− males, carrying the SBE1 and SBE5 deletions in trans, were then bred to wild-type CD1 females. The progeny from this cross were screened for recombination events that placed the SBE1 and SBE5 deletions in cis (1/1600 offspring). ShhS5B1/−/− double-heterozygous mice were then intercrossed to generate ShhS5B1/−/− double-homozygous embryos.

Barhl2^−/− (ref. 46), conditional Yap1 knockout (ShhS5B1/−/−; Yap1^fl/fl)47 and ShhP1 (ref. 48) embryos were described previously. The SBE5A2kb mouse line, referred to here as ShhS5B5A2kb, was generated with CRISPR/Cas9 genome editing tools. The Target Finder platform (E. Zhang) was used to design two pairs of single guide RNAs (sgRNAs) (Supplementary Table 2) flanking SBE5 with the lowest off-target specificity. Complementary guide sequences were annealed, phosphorylated and cloned into the BbsI site of the pX458 or pX459 vector.49 DNA from the two constructs was purified, combined in a 1:1 ratio (2.5 ng per construct) and injected into the male pronucleus of fertilized mouse eggs (B6SJLF1/J, Jackson Laboratories) at the Transgenic and Chimeric Mouse Facility (Perelman School of Medicine, University of Pennsylvania). F1, founder mice were screened by PCR for the expected 2-kb deletion of SBE5 (5/44 positive for the deletion).

Whole-mount β-galactosidase staining and in situ hybridization. For X-gal staining, whole embryos (E10.5) were fixed in 0.2% glutaraldehyde and 1% formaldehyde at 4 °C for 30 min, stained in a solution containing 1 mg/ml X-gal at 37 °C overnight, washed in PBS, dehydrated in methanol and cleared for imaging in a 1:1 ratio of benzyl alcohol to benzyl benzoate. The length of the stained portion of the zli, normalized to the width of the head, was quantified using ImageJ. Whole-mount RNA in situ hybridization, embryos were fixed in 4% paraformaldehyde at 4 °C overnight and hybridized with digoxigenin-UTP-labeled riboprobes according to a previously described protocol.

RNA sequencing. The 429M20eGFp BAC reporter line30 was used to guide the dissection of Shh-expressing cells from the ventral midbrain, ventroposterior diencephalon and zli of E10.5 embryos under a fluorescence stereomicroscope. Total RNA was extracted from GFP-positive brain tissue isolated from approximately 30 embryos using the miRNeasy Micro kit (Qiagen). The RNA-seq library was prepared from 1 μg of total RNA according to the manufacturer’s protocol for the TruSeq RNA Sample Prep kit (Illumina). Paired-end sequencing (100 bp) was performed on an Illumina HiSeq 2000 instrument at the Next-Generation Sequencing Core (Perelman School of Medicine, University of Pennsylvania) to a depth of 62 million reads. Raw sequences were filtered to retain only high-quality reads. Sequences were processed with the RNA-Seq Unified Mapper (RUM)51, which aligns reads to the set of known transcripts in RefSeq, UCSC, Ensembl and the mouse genome (mm9) and outputs feature-level quantification (transcript, exon and intron). To analyze global gene expression profiles, the number of reads uniquely aligning to mRNA transcripts was extracted from RUM output and processed using a custom script.52 Transcripts with FPKM >2 were considered to be expressed.

Motif analysis. De novo motif discovery in mmSBE1 and the six human SBE1-like enhancers (hs194, hs593, hs779, hs1109, hs1180 and hs1391) was performed using Weeder (v1.4.2)33,52 on a Mac terminal. The Weeder parameter ‘HS/MM large S M T20’ was employed to identify the top 20 overrepresented motifs in each length category, ranging from 6 to 12 nt, as the ‘interesting motifs’ (highest ranking). The number of mismatches in each length category is based on the default setting of the algorithm (motifs of 6 nt allow one mismatch; those of 8 nt allow two mismatches; those of 10 nt allow three mismatches; and those of 12 nt allow four mismatches). Interesting motifs with overlapping sequences (24) were merged. Motif enrichment was calculated on the basis of the probability of observing a given motif in 280 random genomic sequences (40 sets of 7 inputted to Weeder) matched for GC content and length, using Fisher’s exact test in R. Each of the motifs was used to query human and mouse transcription factor binding sites using the ‘search for Similar Motifs’ function in UniPROBE53 and ‘JASPAR CORE Vertebrata’ (ref. 54). The candidate transcription factors were further filtered on the basis of their expression level in the SBE1 active region according to their RNA-seq profiles. The seven SBE1-like enhancers were also screened for known human and mouse transcription factor binding sites using web-based tools associated with the UniPROBE and JASPAR databases. In addition to transcription factor binding sites matching motifs 1–3, a Barhl1- and Barhl2-binding site (motif 6) was identified as significantly enriched in SBE1-like enhancers (30/53) as compared to random genomic sequence (P < 0.1, Fisher’s exact test in R).

doi:10.1038/ng.3542

© 2016 Nature America, Inc. All rights reserved.
The enrichment of motifs 1–6 in SBE1-like enhancers was calculated using a random sampling approach to compare the co-occurrence of the six motifs in 46 SBE1-like enhancers with that in random genomic sequences (matched for GC content and length). Briefly, 20 random sequences were sampled from the 46 SBE1-like enhancers, and the number of sequences containing all six motifs was determined. The same calculation was also performed for 20 random sequences sampled from the human genome. After 1,000 iterations of sampling, the two sets of counts were compared using Welch’s two-sample t test in R to determine the statistical significance of enrichment.

Identification of SBE5 and skSBE1. Clustering of motifs 1–6 within a 2-kb DNA sequence (the average length of enhancers in the VISTA Enhancer Browser) was used to predict the location of new SBE1-like enhancers. To identify SBE5, we surveyed 1-Mb regions upstream and downstream of the Shh transcription start site for histone modifications (H3K4me1 and H3K27ac) associated with regulatory sequences using ENCODE data from E14.5 mouse brain34. We next screened these putative regulatory sequences for the presence of motif 1. This approach directed us to a region located 784 kb upstream of Shh, within the penultimate intron of the Lmbr1 gene. A sequence scan of the immediate area identified motifs 2–6 within a 1.9-kb region containing motif 1. To identify the SBE1 ortholog in S. kowalevskii, we searched the hhh locus (Skow_1.1 scaffold44409) for the presence of the SBE1-like motif cluster and identified a ~1.1-kb region within intron 2 (303,512–304,562), close to exon 3, that contained all six motifs in a shuffled arrangement.

Transgenic S. kowalevskii reporter assays. skSBE1, mmSBE1 and mmSBE5 were cloned into an expression vector flanked by I-SceI restriction sites containing an S. kowalevskii gbx basal promoter upstream of the mNeonGreen reporter. The transgenes were digested with the I-SceI meganuclease and introduced into fertilized embryos by microinjection as previously described35. Embryos were cultured at 20 °C and screened for expression beginning at 36 hours post-fertilization.

41. Liu-Chittenden, Y. et al. Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. Genes Dev. 26, 1300–1305 (2012).
42. Zhao, L. et al. Disruption of SoxB1-dependent Sonic hedgehog expression in the hypothalamus causes septo-optic dysplasia. Dev. Cell 22, 585–596 (2012).
43. Suda, Y. et al. The same enhancer regulates the earliest Emx2 expression in caudal forebrain primordium, subsequent expression in dorsal telencephalon and later expression in the cortical ventricular zone. Development 137, 2939–2949 (2010).
44. Symmons, G. et al. Functional and topological characteristics of mammalian regulatory domains. Genome Res. 24, 390–400 (2014).
45. Ruf, S. et al. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat. Genet. 43, 379–386 (2011).
46. Ding, Q. et al. BARHL2 differentially regulates the development of retinal amacrine and ganglion neurons. J. Neurosci. 29, 3992–4003 (2009).
47. Mahoney, J.E., Mori, M., Szymaniak, A.D., Varelas, X. & Cardoso, W.V. The hippo pathway effector Yap controls patterning and differentiation of airway epithelial progenitors. Dev. Cell 30, 137–150 (2014).
48. Riccomagno, M.M., Martinu, L., Mulheisen, M., Wu, D.K. & Epstein, D.J. Specification of the mammalian cochlea is dependent on Sonic hedgehog. Genes Dev. 16, 2365–2378 (2002).
49. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).
50. Jeong, Y., El-Jaick, K., Roessler, E., Muenke, M. & Epstein, D.J. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. Development 133, 761–772 (2006).
51. Grant, G.R. et al. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinformatics 27, 2518–2528 (2011).
52. Pavesi, G. & Pesole, G. Using Weeder for the discovery of conserved transcription factor binding sites. Curr. Protoc. Bioinformatics Chapter 2, Unit 2.11 (2006).
53. Newburger, D.E. & Bulyk, M.L. UniPROBE: an online database of protein binding microarray data on protein–DNA interactions. Nucleic Acids Res. 37, D77–D82 (2009).
54. Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. 32, D91–D94 (2004).
55. Lowe, C.J. et al. Dorsalventral patterning in hemichordates: insights into early chordate evolution. PLoS Biol. 4, e291 (2006).
Erratum: *Cis*-regulatory architecture of a brain signaling center predates the origin of chordates

Yao Yao, Paul J Minor, Ying-Tao Zhao, Yongsu Jeong, Ariel M Pani, Anna N King, Orsolya Symmons, Lin Gan, Wellington V Cardoso, François Spitz, Christopher J Lowe & Douglas J Epstein

*Nat. Genet.* 48, 575–580 (2016); published online 11 April 2016; corrected after print 16 May 2016

In the version of this article initially published, the received date for the manuscript was incorrectly listed as 28 January 2015. The correct date is 28 January 2016. The error has been corrected in the HTML and PDF versions of the article.