Comparative transgenic analysis of enhancers from the human SHOX and mouse Shox2 genomic regions

Jessica M. Rosin, Samuel Abassah-Oppong and John Cobb

Department of Biological Sciences, 2500 University Drive N.W. University of Calgary, Calgary AB T2N 1N4, Canada

Received January 28, 2013; Revised March 13, 2013; Accepted April 5, 2013

Disruption of presumptive enhancers downstream of the human SHOX gene (hSHOX) is a frequent cause of the zeugopodal limb defects characteristic of Léri–Weill dyschondrosteosis (LWD). The closely related mouse Shox2 gene (mShox2) is also required for limb development, but in the more proximal stylopodium. In this study, we used transgenic mice in a comparative approach to characterize enhancer sequences in the hSHOX and mShox2 genomic regions. Among conserved noncoding elements (CNEs) that function as enhancers in vertebrate genomes, those that are maintained near paralogous genes are of particular interest given their ancient origins. Therefore, we first analyzed the regulatory potential of a genomic region containing one such duplicated CNE (dCNE) downstream of mShox2 and hSHOX. We identified a strong limb enhancer directly adjacent to the mShox2 dCNE that recapitulates the expression pattern of the endogenous gene. Interestingly, this enhancer requires sequences only conserved in the mammalian lineage in order to drive strong limb expression, whereas the more deeply conserved sequences of the dCNE function as a neural enhancer. Similarly, we found that a conserved element downstream of hSHOX (CNE9) also functions as a neural enhancer in transgenic mice. However, when the CNE9 transgenic construct was enlarged to include adjacent, non-conserved sequences frequently deleted in LWD patients, the transgene drove expression in the zeugopodium of the limbs. Therefore, both hSHOX and mShox2 limb enhancers are coupled to distinct neural enhancers. This is the first report demonstrating the activity of cis-regulatory elements from the hSHOX and mShox2 genomic regions in mammalian embryos.

INTRODUCTION

Disruptions of enhancers, often at considerable genomic distance from their target genes, are now known to be an important mechanism of human disease (1–3). Most frequently, these cis-regulatory elements control the transcription of genes that function during embryonic development. Among these, the human short stature homeobox gene (hSHOX) appears to have one of the highest rates of enhancer deletions, probably due to the high rate of recombination at its location within the pseudoautosomal region 1 (PAR1) of the sex chromosomes (4,5). Haploinsufficiency of SHOX causes Léri–Weill dyschondrosteosis (LWD) and in some cohorts more than one-third of LWD cases can be attributed to deletions far downstream of the SHOX gene (5–7). Similarly, idiopathic short stature (ISS) can be caused by deletions of putative enhancer sequences downstream of SHOX (8). A homozygous loss of SHOX function results in the more severe limb deformities of Langer syndrome, whereas the short stature of Turner syndrome, like LWD, is due to SHOX haploinsufficiency. In all of these syndromes, the forearms and lower legs (the middle segments of the limbs, or zeugopodal elements) are shortened and malformed, often including the painful wrist abnormality, Madelung deformity. Together these conditions are prominent in the human population affecting approximately 1 in 1000 to 1 in 5000 people (9,10).

All vertebrate genomes contain a closely related paralog of SHOX called SHOX2. Since there have been no reports of human SHOX2 deficiencies, the function of the human gene is uncertain. However, several important developmental functions have been described for the mouse Shox2 ortholog (mShox2) (11–14). Inactivation of mShox2 in the developing limbs of mice causes severe shortening of the humeras and femur (15,16) due to defects in chondrocyte maturation specific to the stylopodal domain (17). Since human SHOX and SHOX2 as well as mouse SHOX2 all have identical DNA-binding homeodomains, they may have similar cellular functions that are...
deployed in different parts of the embryo according to their specific expression patterns. Indeed, a knock-in study revealed that human SHOX could rescue heart and forelimb defects caused by the homozygous loss of mShox2 (18). Interestingly, rodents, unlike all other vertebrates, have lost their Shox gene during evolution along with a large segment of PAR1 (19). Therefore, mice have somehow compensated for the loss of the Shox gene and are able to pattern their limbs with only the single Shox2 paralog. In this context, a comparative study of the enhancers controlling hSHOX and mShox2 is of considerable interest for its evolutionary implications.

Developmental limb enhancers are some of the best-characterized long-range cis-regulatory elements. The genomic scale of such sequences ranges from single, discrete elements of a few hundred base pairs (20) to a vast ‘regulatory archipelago’ with many cooperative elements that can buffer the effects of even large-scale deletions (21). In the former category, deletion of the ~1 kb mouse MFCS1 enhancer (also called the ZRS) was sufficient to eliminate sonic hedgehog (Shh) expression in developing limb buds and to cause severe limb truncations (22). In contrast, deletion of a segment of >800 kb containing many enhancers was required to remove the expression of Hoxd13 from distal limb buds (21). Similarly, five distinct, dispersed enhancers control the expression of Fgf8 in the apical ectodermal ridge during limb development (23). The latter mechanism, in which multiple enhancers control the expression of a developmental gene, may be prevalent since the deletion of enhancers, even those that are ultraconserved, has often resulted in no obvious abnormal phenotype (24,25). However, because relatively small deletions are sufficient to cause the LWD and ISS phenotypes (5,8,26), the SHOX enhancer(s) appear(s) to fit the discrete rather than the archipelago enhancer model. Therefore, we sought to determine whether similar discrete enhancers might control expression of both hSHOX and the mShox2 paralog during limb development.

To identify sequences that may regulate the expression of hSHOX in limbs, Fukami et al. (27) and Sabherwal et al. (28) used sequence comparisons to predict the location of enhancers based on their evolutionary conservation. Such an approach has been fruitful in many cases since a large subset of conserved non-coding elements (CNEs) act as enhancers (29) and these elements often cluster around developmental genes (30–32). Sabherwal et al. identified nine CNEs within a 200 kb interval downstream of hSHOX and tested these sequences for enhancer activity by electroporation into chick limb buds (28). Three of these CNEs, designated CNE4, 5 and 9, drove reporter expression in limbs by this in ovo assay (the location of the positive CNEs is shown on the chromosomal map in Fig. 1A). CNE9 is equivalent to the ECS4 enhancer that Fukami et al. first identified based on its ability to activate a luciferase reporter in U2OS osteosarcoma cells (27). CNE9/ECS4 is of particular interest because it is within a region whose deletion is sufficient to cause LWD or ISS, as first reported by Benito-Sanz et al. (26). However, subsequent studies have identified other deletions associated with LWD and ISS that are near, but do not include CNE9 (5,8,33). Furthermore, when tested in a transgenic zebrafish assay, CNE5 and CNE9 drove expression most frequently in the central nervous system (CNS) and eyes, respectively, with less frequent activity in the fins (34). Therefore, a more thorough characterization of the limb enhancer(s) downstream of hSHOX is needed. In addition, limb regulatory elements may not be restricted to the region downstream of hSHOX since three conserved regions upstream of the gene (CNE2, CNE3, uCNE5, Fig. 1A) have been shown to have limb enhancer activity in an in ovo assay (35).

Here, we report the results of a transgenic mouse study in which we compare the regulatory potential of sequences from both the hSHOX and mShox2 genomic regions. Previous studies have justified the use of the chick as a more suitable model for studying hSHOX enhancers since the bird genome maintains the Shox gene. However, we reasoned that even though the Shox gene has been lost in the rodent lineage, the regulatory, upstream proteins that control transcription of the Shox gene are likely conserved. Many studies have demonstrated that evolutionarily distant or novel enhancers can drive transgene expression in mice (36–38). Therefore, we utilized a similar approach to characterize enhancers from the hSHOX genomic region in transgenic mice.

RESULTS

BAC transgenesis reveals regulatory potential near hSHOX and mShox2

We used recombinating to introduce a lacZ reporter into bacterial artificial chromosomes (BACs) containing either the entire hSHOX or mShox2 genes and flanking sequences. In each case, the reporter was inserted at the start codon of the respective gene without deleting any sequence. The chromosomal location of the three BACs used is shown in Fig. 1A. For the analysis of Shox2, we chose to analyze murine sequences so that the function of identified enhancers could potentially be investigated later by genomic targeting.

Four of four E11.5–E12.5 transgenic embryos produced from a lacZ-targeted BAC containing the hSHOX gene (RP13-391G2-lacZ) showed transgene expression in the first pharyngeal arch (arrowheads, Fig. 1B; Supplementary Material, Fig. S1A). This expression domain is similar to that seen by in situ hybridization for hSHOX (39) and is comparable with expression of mShox2 in the same region (Fig. 1E). The transgene was also consistently expressed in the ventral neural tube and in variable patterns in the hindbrain (Fig. 1B). None of the embryos showed any expression in the developing limbs. The lack of limb expression was not surprising since the BAC lacks the downstream limb enhancers CNE4, 5 and 9/ECS4 identified by Sabherwal et al. (28) and Fukami et al. (27) and is consistent with the human deletion data. However, the BAC does contain an upstream sequence (CNE2) shown to have limb enhancer activity by an in ovo electroporation assay (35). Therefore, our results do not support a function for CNE2 in controlling hSHOX limb expression, but are consistent with a model in which SHOX limb enhancers are located distant from the gene.

The mShox2 gene is centrally located within BAC RP23-103D17 (abbreviated as 103D17 below). Three of four transgenic embryos positive for the 103D17-lacZ BAC showed strong transgene expression in the proximal limbs, dorsal root and trigeminal ganglia and hindbrain (Fig. 1C). These are all regions where mShox2 is expressed in a similar pattern as shown by whole-mount in situ hybridization (Fig. 1E). A second BAC (RP23-105B3-lacZ, abbreviated as
105B3) containing the mShox2 gene but different flanking regions was used to help localize the regulatory sequences in the region. Two permanent transgenic lines made from this lacZ-targeted BAC showed posterior limb, dorsal root and trigeminal ganglia expression (Fig. 1D; Supplementary Material, Fig. S1B). However, these lines did not show hindbrain or anterior limb expression, but they were positive for diencephalon and midbrain expression not seen in the 103D17 embryos. The staggered location of 105B3 relative to 103D17 helped reveal the probable location of specific regulatory elements. The diencephalon/midbrain expression shown in the 105B3 embryos appears to require the more 5′ sequences only found in that BAC, whereas hindbrain expression apparently depends on the 3′ sequences unique to 103D17. The different limb expression patterns generated by the two BACs suggest that each BAC contains unique limb regulatory sequences. Specifically, the 105B3-lacZ transgene drove stronger expression in the hindlimbs, while the 103D17-lacZ transgene most closely recapitulated the endogenous mShox2 pattern in the forelimbs (Fig. 2A).
Evolutionary conservation of sequences on BAC 103D17

To identify candidate sequences that may be responsible for the forelimb expression pattern generated by the 103D17-lacZ transgene, we looked for evolutionarily conserved sequences within the BAC interval (Fig. 2B). A comparison of the mouse sequence with its human, chicken, frog and Fugu orthologs revealed three areas of noncoding sequence that are highly conserved from mice to fish. The human orthologs of two of these elements (636 and 1251 in Fig. 2B) have been analyzed in previous studies. The remaining element (741) has not been tested for enhancer activity. The alignment of the region of homology between the duplicated CNEs downstream of mShox2/hSHOX2 (labeled LHB) and hSHOX (CNE5) is shown in Figure 2C. Seventy-eight of 103 bases (75.7%) are identical between the two sequences (shaded boxes). The yellow shading designates sequences identified by MatInspector as potential binding sites for the indicated transcription factors. Only the basepairs with a high information content as designated by MatInspector (ci-values > 60) are shaded. A second PHOX2A binding site is located within the HOXA3, MSX2 binding site. The evolutionary conservation within a 5478 bp interval downstream of mShox2, with additional species added to illustrate the region at left that is only conserved in mammals, is shown in Figure 2D. The purple track shows the DNase I HS for this sequence as detected in E11.5 forelimb buds by the ENCODE project (45). The schematics below show the location and names of constructs analyzed for enhancer activity in the figures that follow with the separate areas of conservation indicated as colored rectangles. The LHB-Core represents the most highly conserved sequence.
transgenic embryos and the results are posted on the VISTA enhancer browser (40). Element hs636 (as designated on the VISTA enhancer browser) located 5′ to SHOX2 and within an intron of the RSRC1 gene, showed only weak limb expression, whereas element hs1251 drove only neural expression in that study. Notably another sequence near SHOX2 from the VISTA enhancer study (hs741) showed expression mostly restricted to the posterior of the limb bud, very similar to that for BAC 105B3-lacZ. Since the mouse ortholog of hs741 is found on BAC 105B3, but not on BAC 103D17 (positions shown in Figs 1A and 2B), that element is a likely candidate for directing the limb expression from the 105B3-lacZ transgene.

Epigenetic chromatin modification, transcriptional coactivator DNA binding and DNase I hypersensitivity (HS) have been used to identify enhancers that are not necessarily evolutionarily conserved (reviewed in (41)). Chromatin immunoprecipitation (ChIP-Seq) detection of EP300 (p300) binding sites and acetylation of histone H3K27 (H3K27ac) have identified thousands of enhancer candidates in developing mouse limbs (42–44). However, none of the EP300 binding sites in E11.5 limb buds, as identified by Blow et al. and Visel et al. (42,44), are within the intervals of the 105B3 and 103D17 BACs we found to contain limb regulatory activity. A single H3K27 mark, −10 kb downstream of the Shox2 gene in the 103D17 BAC interval, was among those identified by Cotney et al. (43). However, DNase I HS (DNase-Seq) data available through the ENCODE project indicate that this H3K27 mark does not correspond to a region of open chromatin in E11.5 limb buds (Fig. 2B) (45). Instead, DNase I HS sites in E11.5 limb buds mostly correspond to evolutionarily conserved regions in the 103D17 BAC interval (Fig. 2B).

In an attempt to localize the limb enhancer(s) on BAC 103D17, we focused on the region downstream of mShox2 containing two peaks of highly conserved sequence, including the mouse ortholog of hs1251 (Fig. 2B and D). This region intrigued us in part because it contained homology to CNE5 downstream of hSHOX (the sequence homology between the two regions is shown in Fig. 2C). Such duplicated CNEs (dCNEs) are assumed to have existed before duplication events produced paralogs such as SHOX and SHOX2 from a single ancestral gene (46). In addition, the corresponding CNE near mShox2 also contains a sequence that has been termed an ultradelet conserved CNE (defined by 100% conservation between the human, mouse and rat genomes over a span of at least 200 bp; orange box, Fig. 2D) (47). Most notably, the sequence of this ultradelet conserved CNE overlaps with the dCNE common to SHOX and SHOX2 (Fig. 2C and D). We first analyzed the interval containing the two peaks of highly conserved sequences downstream of mShox2 whose evolutionary conservation is shown magnified in Fig. 2D.

Coupled limb and hindbrain enhancers downstream of mShox2

To address the significance of the conserved region downstream of mShox2, we produced a 103D17-lacZ transgene in which the most conserved 4378 bp shown in Fig. 2D was deleted. Five of five transient transgenic embryos produced from this construct showed reduced or absent limb and hindbrain expression, prompting us to name the element the limb hindbrain control region or LHB (Fig. 3A; Supplementary Material, Fig. S1C). We refer to the most highly conserved portion of this interval as the LHB-Core (position indicated in Fig. 2D). The 103D17-lacZ ΔLHB-Core transgenic embryos maintained lacZ expression in the dorsal root and trigeminal ganglia similar to transgenic embryos carrying the intact BAC (compare Fig. 3A with Fig. 1C). In order to determine whether this sequence was sufficient for limb and hindbrain expression, we cloned the LHB-Core upstream of a minimal β-globin promoter-lacZ cassette (lacZβ) (48,49) from which we produced transient transgenic embryos (Fig. 3B; Supplementary Material, Fig. S1D). Five of six transient transgenic embryos (E11.5–E12.5) showed neural and limb expression. However, the limb expression was limited to a spot in the proximal limb that did not fully recapitulate the expression of the BAC transgene. In contrast, the hindbrain expression was similar to that of the 103D17-lacZ BAC.

To further localize the enhancer activities within this element, we subdivided the LHB into two segments designated LHB-A and LHB-B (Fig. 2D). First, we produced transient transgensics from LHB-B, which contains the ortholog of hs1251 and shows homology to CNE5. In four of four LHB-B-lacz transgenic embryos, only hindbrain and neural tube expression was scored, and no limb expression was noted (Fig. 3C; Supplementary Material, Fig. S1E). The LHB-Core sequence showed limb enhancer activity, but it was not sufficient to drive expression in a complete Shox2-like pattern in the limbs (Fig. 3B). Therefore, when we cloned the LHB-A segment, we decided to include a further 773 bp of flanking sequence that was only conserved in mammals (yellow segment in Fig. 2D). When ENCODE data from this region became available for E11.5 limb buds, we noted that the flanking 773 bp region contained the most DNase I hypersensitive segment in the 5478 bp LHB interval (purple track, Fig. 2D) (45). Strikingly, two of two transient transgenic embryos carrying the LHB-A-lacz transgenic embryos showed strong expression in limb buds as well as in the hindbrain and neural tube (Fig. 3D; Supplementary Material, Fig. S1F). To characterize the LHB-A-lacz enhancer over a developmental time course, we produced three permanent transgenic lines, all of which showed a similar expression in limbs and the hindbrain (Fig. 3E; Supplementary Material, Fig. S1F). The proximal limb expression of the LHB-A-lacz transgenic lines was similar to mShox2 from E9.5 to E13.5 (Fig. 3E). The sections of E13.5 limbs showed that LHB-A-lacz expression was strongest in the menenchyme surrounding the skeletal elements, and was relatively downregulated in chondrocytes (Fig. 3F), similar to endogenous mShox2 (17).

**Neural enhancer activity of human CNE5 and CNE9**

While searching for functional homologs of the LHB enhancers near the hSHOX gene, we first focused on the CNE5 and CNE9/EC5 enhancers previously identified by Sabherwal et al. (28) and Fukami et al. (27). As noted above, CNE5 was an intriguing candidate because of its ancient homology with the LHB. However, CNE9 seemed the more likely candidate to be required for hSHOX limb expression, since it is located in a region frequently deleted in LWD and ISS patients. Therefore, we tested both of these sequences for enhancer activity in transgenic embryos by fusing each sequence to a βlacz cassette and...
producing stable transgenic lines from the constructs. Four of four CNE5-βlacZ lines showed strong enhancer activity in the CNS, most prominently in the hindbrain at E11.5 (Fig. 4A–D). However, only one of the lines showed limb expression, barely detectable as a small dot of staining at E11.5 (Fig. 4A).

The same line had considerably more limb expression at E12.5 (Fig. 4F, data not shown). We generated one additional CNE5-βlacZ E12.5 transgenic embryo that was stained with X-gal after sectioning. Scale bar = 0.25 mm. Staining surrounds the humerus (hu), radius (r) and ulna (u) elements, but is lighter in the proliferating chondrocytes of the humerus (hollow arrow) and radius and ulna. Produced stable transgenic lines from the constructs. Four of four CNE5-βlacZ lines showed strong enhancer activity in the CNS, most prominently in the hindbrain at E11.5 (Fig. 4A–D). However, only one of the lines showed limb expression, barely detectable as a small dot of staining at E11.5 (Fig. 4A). The same line had considerably more limb expression at E12.5 (Fig. 4F, data not shown).

We obtained similar results with a CNE9-βlacZ transgene, indicating that it is also a neural enhancer. Eight of eight CNE9-βlacZ transgenic lines showed neural expression from E11.5 to E14.5 (expression at E12.5 is shown in Fig. 4G–J; Supplementary Material, Fig. S1G). All eight CNE9-βlacZ lines showed weak, but reproducible staining in the hindbrain, and variable staining in the ventral neural tube (Fig. 4G–J), although the staining was considerably weaker than that obtained with the CNE5 transgene. Only one line showed limb expression, in this case in a domain in the distal limb (Fig. 4J). However, this line had very strong staining throughout the CNS and in the somites, a pattern not seen in any of the other seven lines. Because the overall transgene expression of this one line was radically different from the seven others, its tissue specificity may be attributable to its genomic insertion site.

**Figure 3.** Identification of a limb enhancer ≈45 kb downstream of mShox2. (A) Four transient transgenic embryos (ranging from ≈E10.5 to E12.5) produced with the RP23-103D17- lacZ BAC with the LHB-Core deleted (ΔLHB-Core). The staining in the sensory neurons is similar to that of the same BAC with an intact LHB-Core sequence, but the forelimb (arrows), hindlimb (h) and hindbrain (hollow arrowheads) expression is absent or greatly diminished (compare with Fig. 1C). (B–D) Representative transient transgenic embryos positive for the LHB-Core-, LHB-B-, LHB-A-βlacZ constructs. (B) The LHB-Core is sufficient for strong hindbrain (hollow arrowhead) and neural tube staining, but it produces only a spot of activity in the limbs (arrow, h). (C) The LHB-B construct drives lacZ expression only in the hindbrain (hollow arrowhead) and neural tube. (D) The LHB-A transgene drives strong expression in the hindbrain (hollow arrowhead) and fore and hindlimbs (arrow, h). (E) Forelimb staining in a permanent LHB-A-βlacZ transgenic line at the indicated stages. All images are at the same magnification, scale bar = 1 mm. Expression is similar to endogenous mShox2 (compare with Figs 1E and 2A) including similar distal borders of expression (arrows) and an area that is negative for expression in the anterior-proximal region (hollow arrow at E10.5). Some staining is seen in the distal limb at later stages (hollow arrowhead at E13.5). (F) Cryosection through an E13.5 LHB-A-βlacZ forelimb that was stained with X-gal after sectioning. Scale bar = 0.25 mm. Staining surrounds the humerus (hu), radius (r) and ulna (u) elements, but is lighter in the proliferating chondrocytes of the humerus (hollow arrow) and radius and ulna.

**Functional homology between hSHOX and mShox2 hindbrain enhancers**

The hindbrain expression pattern obtained with CNE5-βlacZ was strikingly similar to that of the LHB-Core transgene (compare Fig. 3B with Fig. 4A–D). Sectioning revealed that the hindbrain patterns generated by both the CNE5 and LHB-Core transgenes were similar to the endogenous expression of mShox2 (Fig. 5A–E). However, the CNE5-βlacZ staining was more continuous through the hindbrain, whereas the LHB-βlacZ transgene generated an expression pattern with distinct medial and lateral domains, much like endogenous mShox2 (arrows, Fig. 5A–E, H, I).

When we compared the hindbrain activity of the LHB-Core with its constituent components, we noted that the LHB-B element controlled lacZ expression mostly in the medial domain (Fig. 5F and J), while LHB-A drove transgene expression in the lateral domain (Fig. 5G and K). In addition, inclusion of the flanking domain necessary for strong limb expression did not appear to augment hindbrain expression, since the expression pattern of the LHB-A transgene was also generated by the LHB-Core transgene (compare Fig. 5E, I and G, K). Therefore, the full LHB hindbrain expression pattern was generated in a
modular fashion through the combination of the activities of its A and B domains, which contain separate peaks of conservation separated by a non-conserved sequence (Fig. 2D). Interestingly, our results indicate that the LHB-Core is a functional homolog of CNE5 in terms of its ability to direct hindbrain gene expression, yet CNE5 maintains sequence homology only with the B domain of the LHB.

A zeugopodal limb enhancer linked to CNE9 downstream of hSHOX

As described above, in our analysis of the LHB region, it was necessary to include less-conserved sequences in our transgene to identify an enhancer that recapitulates the mShox2 expression pattern in limbs (Fig. 2 and 3). We decided to try a similar strategy to search for limb enhancer activity in the human CNE9 region. This approach seemed particularly promising, since human LWD and ISS patients have been identified with deletions that are near, but do not include CNE9 (Fig. 6A) (5,8,33). Therefore, we designed a transgene that included the shortest region of overlapping (SRO) deletion associated with an LWD or ISS phenotype, as reported by Huber et al. (8). To maximize our chances of capturing limb enhancer activity, we included the sequence spanning from the 10.5 kb SRO up to and including CNE9, resulting in an ~19 kb transgene, hereafter referred to as 19 kbCNE9-βlacZ (diagramed in Fig. 6A). We also examined the available ENCODE data for clues to the location of enhancers in this genomic region (45). Among the 22 human cell types with data posted for epigenetic chromatin marks, only two [normal human osteoblasts (NHOst) and A549 carcinoma cells] showed significant positive H3K27ac signatures within the 19 kbCNE9 region as shown for osteoblasts in Fig. 6A. We also examined the available ENCODE data for clues to the location of enhancers in this genomic region (45). Among the 22 human cell types with data posted for epigenetic chromatin marks, only two [normal human osteoblasts (NHOst) and A549 carcinoma cells] showed significant positive H3K27ac signatures within the 19 kbCNE9 region as shown for osteoblasts in Fig. 6A. The presence of H3K27ac marks in cells of skeletal origin was intriguing, and indicated that enhancer activity may be present within the fragment outside of its conserved region.

Three independent transgenic lines were produced from the 19 kbCNE9-βlacZ construct. Staining of E13.5 embryos from two of those lines is shown in Fig. 6B and C. Both lines...
showed a similar strong X-gal staining pattern in the developing fore and hindlimbs, in a domain mostly restricted to the developing zeugopodium, as would be expected for a hSHOX enhancer. These lines also showed weak hindbrain expression (Fig. 6B and C). Transgene expression was strongest in the mesenchyme surrounding the developing radius and ulna, as can be seen in sections of E13.5 and E14.5 limbs (Fig. 6D and E). The transgene was more weakly expressed in proliferating chondrocytes of the radius, ulna and distal humerus, and absent from hypertrophic chondrocytes at the center of the elements. This expression domain is reminiscent of the down-regulation of mShox2 in differentiating chondrocytes of the proximal limb (17) and similar to that reported for hSHOX (39). The line shown in Fig. 6C also showed staining in the lateral plate mesoderm, portions of the somites and in the facial mesenchyme. Since these additional tissues are not stained in the line shown in Fig. 6B, these unique expression domains may be attributable to a position effect of the transgene. A third line showed overall very weak lacZ expression, perhaps due to a repressive insertion site, but limb and hindbrain expression was also detectable in that line (Supplementary Material, Fig. S11).

A developmental time course of stained forelimbs from 19 kbCNE9-βlacZ transgenic embryos is shown in Fig. 6 G–K. At E10.5, enhancer activity was localized to a proximal, posterior region of the forelimb buds and by E11.5 a more centrally located domain developed (arrows, Fig. 6G and H). By E12.5, the expression became more specific to the zeugopodial domain although some proximal expression remained (Fig. 6I). At E13.5 and E14.5, staining outlined the forelimb zeugopodial elements (Fig. 6J and K). In a close-up of an E13.5 hindlimb (Fig. 6F), staining can be seen surrounding the developing tibia and fibula, similar to the corresponding expression domain in the forelimbs.

**DISCUSSION**

Our results demonstrate the activity of enhancers in the hSHOX and mShox2 genomic neighborhoods, specifically identifying limb and neural cis-regulatory sequences within similar gene deserts downstream of each of the genes. Unexpectedly, the limb elements we describe are each closely associated with more highly conserved neural enhancers. By utilizing a comparative approach, we have shown that duplicated, conserved neural enhancers near mShox2 and hSHOX drive similar expression patterns in the hindbrain. We suggest that this expression may indicate the location of an ancient, shared function for the genes. In addition, we demonstrate that transgenic mice can be used as important tools for characterizing the enhancer deletions that cause LWD and ISS phenotypes.

**Evolutionary implications**

The two Shox gene paralogs are present in all vertebrate genomes for which the sequence is available, with the notable exception of rodents, which only maintain Shox2. The duplication event that created the paralogs and associated dCNEs probably coincided with the whole genome-duplication events of approximately...
500 million years ago (55). Because fish and tetrapods emerged after the genome duplications, the cis-regulatory elements directing paired fin and tetrapod limb expression presumably appeared after this time as well. However, limb/fin enhancers could have been generated by modification or elaboration of previously existing elements, perhaps when the transcription factors binding to those elements made them favorable or accessible for activity in limbs, a process that has been dubbed ‘regulatory priming’ (56). Duplicated CNEs such as those found within CNE5/LHB are intriguing in this context. On the other hand, we cannot exclude the possibility that the close association of the neural and limb enhancers is fortuitous.

Because ancient enhancers near the mShox2 and hSHOX genes are active in the CNS, this implies that Shox genes had an ancestral function in these tissues. In support of this idea, the single Shox gene of the basal chordate amphioxus is expressed in the neural tube during development (57). Therefore, we hypothesize that the Shox genes have a conserved function in the CNS of vertebrates. However, CNS abnormalities have not been reported in hSHOX patients, perhaps due to the redundancy of the SHOX2 paralog. Studies in mice may help resolve this issue because of the absence of Shox gene redundancy in this species. Our preliminary results indicate that mShox2 is indeed required for CNS development in mice since conditional knockout of the gene in the developing brain and spinal cord is lethal and results in hindbrain malformation (J.M.R., unpublished results).

**Limb enhancers near short-stature genes**

In our transgenic mouse model, CNE5 and CNE9 by themselves did not show any reproducible limb expression. This contrasts with the limb activity found for these sequences by Sabherwal et al. using a chick limb model (28). This is not surprising since other studies have noted variable results when different species, zebrafish and chick for example, were used for...
transgenic studies of enhancers, including those from the hSHOX region (34). We suggest that the limb activity of CNE5 and CNE9 transgenes detected by in ovo electroporation may have revealed a weak ability to direct limb expression that may not necessarily reflect significant activity of the enhancers when integrated into the genome. In certain contexts, electroporation has been reported to result in broader expression than expected, which was suggested to occur because of the high copy number of the episomal transgenes used in such experiments (58). Our results indicate, at least in the case of CNE9, that the limb enhancer activity is not robust without the inclusion of less conserved sequences. One might predict that the addition of flanking sequences to a CNE5 transgene might also enhance limb activity as it did for CNE9—this is an interesting possibility that should be tested in the future.

The identification of a downstream limb enhancer for mShox2 establishes this as a model with interesting parallels to hSHOX regulation. The common regulatory features of the paralogs may extend upward of the genes as well, since limb enhancers have been identified 5′ of hSHOX using the same in ovo electroporation assay used to identify the downstream enhancers (35), although the clinical relevance of the upstream sequences is not clear (59). Interestingly, our BAC transgenesis results implicate an upstream enhancer, most likely element 741, as another element that may control mShox2 limb expression (Fig. 1D). Therefore, both Shox paralogs may have upstream and downstream regulatory inputs that can be further characterized by transgenesis. The existence of multiple elements controlling mShox2 limb expression is supported by our recent data, showing that limb patterning appears to be normal with a genomic deletion of the LHB enhancer (S.A.O, unpublished results).

A mouse model for studying sequences deleted in hSHOX patients

Our results establish transgenic mice as a suitable model for the study of hSHOX enhancers. The characterization of the limb enhancer activity in the CNE9 region will have clinical implications, since the assignment of the specific sequences required for limb activity should be possible in this system. Future studies will identify the minimal sequences necessary for limb expression within the 19 kb fragment, and mutational analysis should be able to identify the specific sequences required for this activity. The availability of the mouse as a model could also make it possible to identify the proteins that bind and regulate the activity of the enhancer(s).

To our knowledge, all downstream deletions associated with a SHOX-deficient phenotype disrupt at least some portion of the 19 kbCNE9 fragment. This includes the interesting, recurrent 47 kb deletion found in 30 unrelated individuals with ISS or LWD (33), which deletes the distal 11.1 kb of the 19 kb fragment (Fig. 6A). Since this recurrent deletion does not include CNE9, this raises the interesting possibility that CNE9 may not be required for limb enhancer activity, a possibility that can be tested with transgenic mice. Of course, limb enhancer activity might also be present beyond the 19 kb element tested here. For example, deletions described in one study implicated the CNE8 region as a region with limb regulatory activity (5), although the deletion in that study did include ~1 kb of the 19 kbCNE9 fragment. Also a CNE distal to the 19 kbCNE9 element (given the name ECR1, Fig. 1A), which lies within the recurrent deletion described by Benito-Sanz et al., was shown to associate with the Shox promoter in chick limb bud cells, as measured by 3C analysis (33). A scan using transgenic constructs through the region of the recurrent 47 kb deletion up to CNE9 should help define the activity of enhancers in the region. These data should eventually contribute to a more precise clinical screen to detect deletions causing LWD and ISS phenotypes.

MATERIALS AND METHODS

BACs

Human genomic BACs RP13-391G2, RP13-76L22 and RP13-167H21 were obtained from the Wellcome Trust Sanger Institute. The Children’s Hospital Oakland Research Institute (CHORI) supplied mouse genomic BACs RP23-103D17 and RP23-105B3. The identity and integrity of each BAC were confirmed by restriction enzyme fingerprinting. RP13-391G2 required further confirmation because the BAC supplied was much larger than the 69 861 bp sequence available through the NCBI website (GenBank: BX004827.19). BAC end sequencing revealed a genomic location and size as indicated in Fig. 1A (insert size = 190 492 bp). Extensive restriction enzyme characterization, pulsed-field electrophoresis and sequencing at the hSHOX locus confirmed the genomic location, integrity and size of the BAC on which the complete hSHOX gene is centrally located.

Recombineering

A recombination cassette was engineered for introduction of a lacZ reporter into the hSHOX and mShox2 BACs as follows. A PCR was performed with the following PCR primer pairs using plasmid pGPS2 (NEB) as a template. pGPS2 was used for cloning since this plasmid will not replicate in BAC host strains of E. coli, thus reducing background.

Oligonucleotides for PCR shown below, with italics indicating restriction sites for cloning (NcoI or NotI), bold indicating homologous arms from hSHOX and mShox2 recombinase and underlined sequences that are homologous to pGPS2 sequences. hSHOX-5′: gatCCATGGCTGGGCGGGGCTG TGGACCAGCCGCCGCGGCTGTGATTGCCGCTGCCGCCG TGGCCGCTGCCGCCG AGATCTATTTTGTTCAGTTTAAAG; hSHOX-3′: atcGGCGCG CATGGAAGAGGCTACGGGCTTTTGTATCCTAACAGTTT TTGACCAGAAAAAGCAAGGACGTTACCCCTGTAAGTGC GCAAAAAC; mShox2-5′: gcattGTTGGCCTGGGCCTGGCGCCGG CCCTACACCCCTGCGCTCAAGCTCCTCCCTCCGGCAGC CCCCAGAGATCTATTTTGTTCAGTTTAAAG; mShox2-3′: atcGGCGCGCATGGAAGAGGCTACGGGCTTTTGTATCCTAACAGTTT TTGACCAGAAAAAGCAAGGACGTTACCCCTGTAAGTGC GCAAAAAC.

The NcoI/NotI digested PCR fragment/plasmid was ligated with an NcoI/NotI digested cassette from the pβlacFrtKanFrt plasmid (a gift from F. Spitz) containing a human minimal β-globin promoter upstream of a lacZ reporter, SV40 polyadenylation sequence and a kanamycin resistance cassette flanked by FRT sites. The ligation was transformed into PIR1
cells (Invitrogen) followed by selection with kanamycin. A recombination cassette was liberated from positive clones by 
KpnI/BglII digestion and electroporated into DY380 (or derivative strain EL350) cells containing the appropriate BAC and induced for recombination by incubation at 42°C as described (60). Clones containing targeted BACs were selected by kanamycin resistance and verified by restriction enzyme fingerprinting.

In order to purify BACs for pronuclear injection we used either a Nucleobond BAC-100 Midiprep kit (Macherey-Nagel) or Qiagen Plasmid Midi kit starting from 500 ml of bacterial culture. In our early injections, BACs were linearized with PI-SceI, but all later injections were performed with circular BACs with similar results as reported (62). All BACs were suspended in BAC injection buffer (63). Cloned fragments were verified by DNA sequencing.

### Cloning of enhancer sequences

All blacZ enhancer constructs were produced by cloning into the pBlacZ-FrtKanFrt plasmid or pBlacZ (an identical plasmid without the kanamycin resistance cassette). CNE5 and CNE9 sequences were generated by PCR from BACs RP13-76L22 or RP23-103D17. All PCR products were digested with Acc65I/Sallı and ligated into plasmid pBlacZ-FrtKanFrt. Fragments were excised for injection with Acc65I/XbaI digestion and gel purified using a GeneClean Turbo kit (MP Biomedicals) and resuspended in standard injection buffer (63). Cloned fragments were verified by DNA sequencing.

The larger enhancer fragments (LHB-Core (4705 bp) and 19 kbCNE9 (18 796 bp)) were cloned by gap repair (61). Recombination arms for the LHB-Core were produced by PCR (primers listed in Table 1), digested with NotI and HindIII and

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence, 5'–3'</th>
<th>Band size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNE5F</td>
<td>TAAGGTACCAGCCCTCCTCCGGGACGGATTG</td>
<td>1026</td>
</tr>
<tr>
<td>CNE5R</td>
<td>AATAGTTGAATCTCTATCATCTTCCTGGGA</td>
<td>822</td>
</tr>
<tr>
<td>CNE9F</td>
<td>TAAGGTACCAGCCCTCCTCCGGGACGGATTG</td>
<td>1026</td>
</tr>
<tr>
<td>CNE9R</td>
<td>AATAGTTGAATCTCTATCATCTTCCTGGGA</td>
<td>822</td>
</tr>
<tr>
<td>LHBAF</td>
<td>TAAAGGTACCCAGCCCTCCTCCGGGACGGATTG</td>
<td>1026</td>
</tr>
<tr>
<td>LHBAR</td>
<td>AATAGTTGAATCTCTATCATCTTCCTGGGA</td>
<td>822</td>
</tr>
</tbody>
</table>

**PCR fragments for 19 kbCNE9 fragment retrieval**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Band size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 kbCNE9A</td>
<td>ATAGGTACCAGCCCTCCTCCGGGACGGATTG</td>
<td>389</td>
</tr>
<tr>
<td>19 kbCNE9B</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>312</td>
</tr>
<tr>
<td>19 kbCNE9C</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>312</td>
</tr>
<tr>
<td>19 kbCNE9D</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>312</td>
</tr>
</tbody>
</table>

**PCR fragments for LHB-Core fragment retrieval**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Band size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHBDELA</td>
<td>ATAGGTACCAGCCCTCCTCCGGGACGGATTG</td>
<td>469</td>
</tr>
<tr>
<td>LHBDELB</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>319</td>
</tr>
<tr>
<td>LHBDELC</td>
<td>ATAGGTACCAGCCCTCCTCCGGGACGGATTG</td>
<td>319</td>
</tr>
<tr>
<td>LHBDELD</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>319</td>
</tr>
</tbody>
</table>

**PCR fragments for LHB-Core deletion**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Band size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHBDELA</td>
<td>ATAGGTACCAGCCCTCCTCCGGGACGGATTG</td>
<td>164</td>
</tr>
<tr>
<td>LHBDELB</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>484</td>
</tr>
<tr>
<td>LHBDELBC</td>
<td>TGTAGATCCAGCCCTCCTCCGGGACGGATTG</td>
<td>484</td>
</tr>
</tbody>
</table>

**Primers used for enhancer cloning, recombineering, confirming BAC integrity and mouse genotyping**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Band size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHOX2LACZF</td>
<td>GGCTAGCCCCTTCTCTTCTTTG</td>
<td>392</td>
</tr>
<tr>
<td>SHOXLACZF</td>
<td>AGCAAGGGCTCAGACCCCTTCTTCTTTG</td>
<td>371</td>
</tr>
<tr>
<td>LACZ</td>
<td>GAATCCGATGCTTCTTCTTTG</td>
<td>578</td>
</tr>
</tbody>
</table>

*Sequences in italics are restriction enzyme sites for cloning; bases 5' to these are to facilitate cutting. Nucleotides with homology to the target sequence are in bold.*
HindIV/SpeI, respectively, and ligated into NotI/SpeI digested pL253. The resulting retrieval plasmid was linearized with HindIII and electroporated into induced EL350 cells containing BAC RP23-103D17 as described (61). The retrieved fragment was cloned into the SalI site of pBlacZ and the fragment excised and purified for injection as described above. The LHB-B construct was generated by a NotI/BglII digest and relocation of the LHB-Core–βlacZ construct which removed the 2561 bp at the telomeric side of the construct.

The 19 kbCNE9 fragment was retrieved from BAC RP13-167H21 by a similar protocol, except the PCR fragments (primers listed in Table 1) were digested with Acc65I/BglII and BglII/SalI, respectively, and ligated directly into the pBlacZ plasmid. The plasmid with the cloned PCR products was linearized for retrieval with BglII and the fragment for injection was excised from the resulting clone by digestion with Acc65I/NotI and purified using the Zymoclean Large Fragment DNA recovery kit (Zymo Research).

### Generation of transgenic mice

Transgenic mice were produced by pronuclear injection into C57BL/6 X DBA F1 or CD-1 single-cell embryos using standard techniques (63). Transient transgenic embryos were produced in our laboratory and permanent lines were produced at the University of Calgary Center for Mouse Genomics. Transgenic mice and embryos were genotyped by transgene-specific primers: examples for the BAC transgenes using a common reverse primer are shown in Table 1.

The integrity of human BAC RP23-391G2A in transgenic embryos/mice was confirmed by a PCR using primers from each extremity of the BAC (primers 391G2A/B and C/D, Table 1).

### X-gal staining and in situ hybridization

Whole-mount in situ hybridization with a mShox2 riboprobe was previously described (64). In situ hybridization on 16–20 μm cryosections was previously described (15). Embryos with lacZ transgenes were stained with X-gal according to the standard techniques (63).

### Sequence analysis

Sequences from human genome hg19 and mouse genome mm9 assemblies and homologous sequences from the indicated species were retrieved from the UCSC genome browser uploaded to the mVISTA server, aligned using MLAGAN and compared with mVISTA (65,66). Sequence alignment and primer design were done with MacVector.

Transcription factor binding sites were identified with MatInspector (67).

### SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

### ACKNOWLEDGEMENTS

We thank D. Duboule for his support during the early stages of this project and F. Spitz for training in pronuclear injection, recombineering advice, strategies and reagents. We are grateful for the technical assistance of L. Vickerman and N. Fraudeau, to S. Neufeld for BAC targeting, to B. Bobick and D. Kurrasch for assistance with sectioning and to C. Geary-Joo, and C. Fielding of the University of Calgary Center for Mouse Genomics for pronuclear injections. We thank L. Selleri for providing limb bud samples to ENCODE; G. Stamatoyannopoulos for the DNA-seq data and B. Bernstein for the osteoblast H3K27ac data available through ENCODE. J.C. is supported as an Alberta Heritage Foundation for Medical Research (AHFMR) Scholar. J.M.R. was supported by fellowships from NSERC and the CIHR Training Program in Genetics, Child Development and Health at the University of Calgary.

### Conflict of Interest statement

None declared.

### FUNDING

This work was supported by the Canadian Institutes of Health Research (CIHR) (MOP-93562).

### REFERENCES


