

SCIENTIFIC REPORTS



OPEN

Identification of a limb enhancer that is removed by pathogenic deletions downstream of the *SHOX* gene

Isabella Skuplik¹, Sara Benito-Sanz^{2,3}, Jessica M. Rosin¹, Brent E. Bobick¹, Karen E. Heath^{2,3} & John Cobb¹

Haploinsufficiency of the human *SHOX* gene causes Léri-Weill dyschondrosteosis (LWD), characterized by shortening of the middle segments of the limbs and Madelung deformity of the wrist. As many as 35% of LWD cases are caused by deletions of non-coding sequences downstream of *SHOX* that presumably remove an enhancer or enhancers necessary for *SHOX* expression in developing limbs. We searched for these active sequences using a transgenic mouse assay and identified a 563 basepair (bp) enhancer with specific activity in the limb regions where *SHOX* functions. This enhancer has previously escaped notice because of its poor evolutionary conservation, although it does contain 100 bp that are conserved in non-rodent mammals. A primary cell luciferase assay confirmed the enhancer activity of the conserved core sequence and demonstrated that putative HOX binding sites are required for its activity. This enhancer is removed in most non-coding deletions that cause LWD. However, we did not identify any likely pathogenic variants of the enhancer in a screen of 124 LWD individuals for whom no causative mutation had been found, suggesting that only larger deletions in the region commonly cause LWD. We hypothesize that loss of this enhancer contributes to the pathogenicity of deletions downstream of *SHOX*.

In addition to disturbing coding sequences, chromosomal aberrations can cause genetic disease when they remove, modify or displace *cis*-regulatory elements necessary for the control of gene expression¹. Within this context, the enhancers that regulate tetrapod limb development are among the best studied², including the zone of polarizing activity regulatory sequence (ZRS) that controls sonic hedgehog (*Shh*) expression in developing limbs³. The ZRS is uniquely required for the expression of *Shh* for anterior-posterior patterning of the distal portion of the limbs. Several point mutations cause a gain-of-function of the ZRS resulting in polydactyly in many species including humans³. While the discrete ZRS has served as a prototype for understanding the function of enhancers during limb development, in many cases numerous enhancers dispersed over large genomic regions control the expression of genes active in limb patterning^{4,5}. For example, the *HoxD* genes that pattern the distal limb fit this dispersed model of regulation⁶. Similarly, multiple enhancers control the expression of *Gdf5* in the developing joints of the limbs⁷. Overall, genes coding for transcription factors important during limb development have been estimated to have a median number of eight enhancers regulating their expression, indicating that genes controlled by a single discrete enhancer might be exceptions to an otherwise general rule⁵. Animal models have been crucial for our understanding of how mutations and chromosomal rearrangements result in limb deformities in humans, helping to define which genes are regulated in a discrete or dispersed fashion⁸.

The *cis*-regulatory control of the human short stature homeobox gene (*SHOX*) is an emerging paradigm in this field^{9,10}. Deletions of non-coding sequences that overlap in an interval approximately 250 kb downstream of *SHOX* on the pseudoautosomal region 1 (PAR1) of the X and Y chromosomes cause a large number of Léri-Weill dyschondrosteosis (LWD, MIM 127300) cases, as many as 35% in some cohorts and a small proportion of

¹Department of Biological Sciences, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, T2N 1N4, Canada. ²Instituto de Genética Médica y Molecular (INGEMM), IdiPAZ and Skeletal dysplasia multidisciplinary unit (UMDE), Hospital Universitario La Paz, Universidad Autónoma de Madrid, Pº Castellana 261, 28046, Madrid, Spain. ³CIBERER, ISCIII, Madrid, Spain. Correspondence and requests for materials should be addressed to K.E.H. (email: karen.heath@salud.madrid.org) or J.C. (email: jacobb@ucalgary.ca)

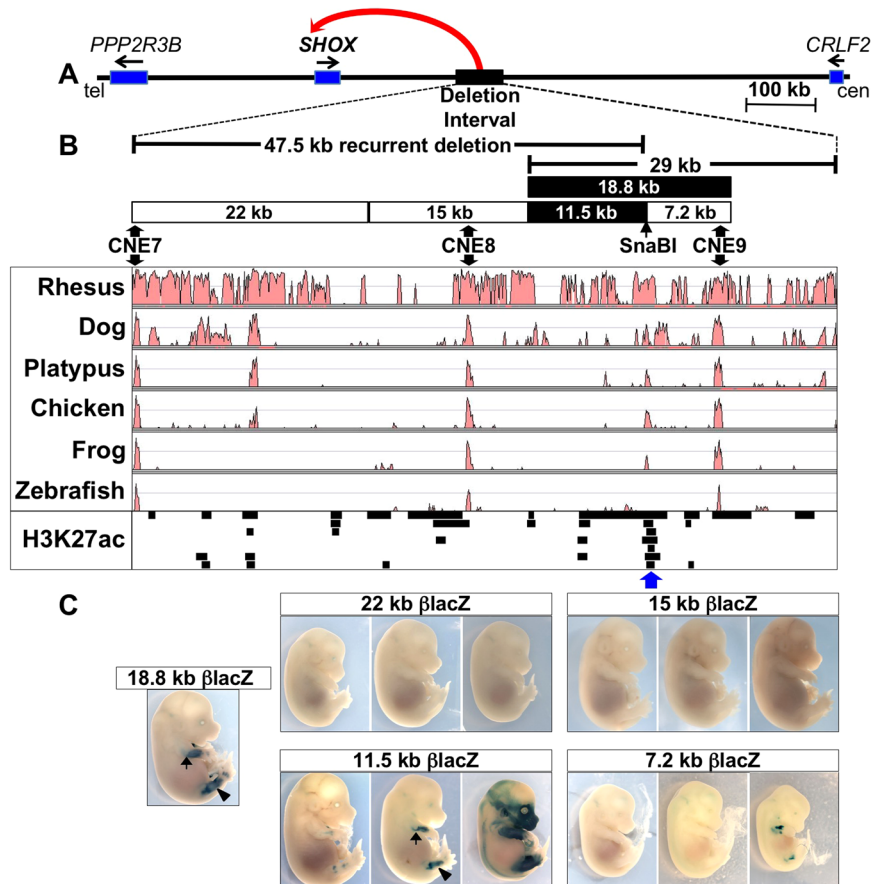


Figure 1. ZED activity is located within an 11.5 kb sequence that does not include CNE9. (A) Schematic of ~1 Mb of the human X and Y chromosomes at the *SHOX* locus with a red arrow indicating the location of a presumed enhancer or enhancers within an interval frequently deleted in LWD and ISS patients. (tel = telomeric, cen = centromeric, blue rectangles indicate genes) (B) Expanded view of the 65.7 kb deletion interval extending from the telomeric breakpoint of the ~47.5 kb recurrent deletion¹⁷ to the centromeric boundary of the 29 kb smallest region of overlapping (SRO) deletion from Benito-Sanz *et al.*⁹. The rectangles indicate the fragments tested in the transgenic assay, which are aligned with the deletion intervals and with a plot of the evolutionary conservation of the region (below), as generated with VISTA for the indicated species⁶³. The fragments found to have limb enhancer activity are in black; white rectangles indicate fragments that did not have reproducible limb activity. The H3K27ac track shows the location of regions enriched for H3K27 acetylation in human embryonic limbs at E33, E41, E44, E47 (two replicates per stage, except for E41 for which only one replicate showed enrichment in this region) as reported by Cotney *et al.*²⁹. Each row of this track represents a different sample arranged from E33, replicate 1 at the top, to E47 replicate 2 at the bottom. The blue arrow indicates the only region with H3K27ac in all of these samples. (C) Sets of three E13.5/E14.5 embryos carrying the indicated transgenes; each embryo represents an independent genomic insertion. Only the 11.5 kb β lacZ transgene has reproducible limb staining, which appears similar to that of the previously reported 18.8 kb transgene staining as shown in the embryo at left (in two representative embryos, the arrow points to staining in the forelimbs and the arrowhead indicates hindlimb staining).

cases with Langer mesomelic dysplasia (LMD, MIM 249700) and idiopathic short stature (ISS, MIM 300582) (Fig. 1A)^{9,11–13}. Presumably these deletions remove an enhancer or enhancers necessary for *SHOX* expression in developing limbs resulting in *SHOX* haploinsufficiency, although the precise location of these elements has not been determined.

Individuals with either downstream deletions or disruptions of *SHOX* coding sequences have a similar shortening of the middle (zeugopodal) limb elements⁹. LWD is a dominantly inherited skeletal dysplasia that is characterized by disproportionate short stature, mesomelic limb shortening, and Madelung deformity of the forearm: the bowing of the radius, distal dislocation of the ulna and triangulation of the carpal bones¹⁴. A more severe skeletal dysplasia (LMD) is caused by homozygous or compound heterozygous mutations of *SHOX* or its enhancer regions. LMD is characterized by severe disproportionate short stature due to both mesomelic and rhizomelic shortening of the limbs. ISS is classified as individuals with a height below -2 SDS in the absence of known specific causative disorders.

Others have used evolutionary sequence conservation as a guide to locate potential enhancers downstream of the *SHOX* gene. In this way, an Evolutionarily Conserved Sequence designated ECS4 was identified as a limb

enhancer candidate within the critical deletion interval downstream of *SHOX*¹⁵. ECS4, now commonly known as CNE9 (Conserved Non-coding Element 9), was shown to activate transcription from an episomal reporter construct when electroporated into chick limb buds and in a luciferase assay in U2OS cells^{15,16}. However, in these experiments enhancer activity was only tested in the specific cells that were electroporated. Therefore, it was not clear whether the enhancer was tissue-specific since reporter activity was not tested in all tissues of a developing embryo from a construct integrated into the genome. Furthermore, many of the deletions identified in individuals with LWD and ISS do not include CNE9^{17,18} indicating that the limb-specific enhancer(s) may lie elsewhere. Most notably, an identical, recurrent 47,543 base pair (bp) deletion (abbreviated 47.5 kb) identified by array CGH and breakpoint PCR sequencing appears to be a major cause of LWD (location shown in Fig. 1B)¹⁷. Since this deletion does not remove CNE9, the pathogenic mechanism was thought to be due to the loss of another conserved sequence called CNE7 (also known as ECR1), although the limb-specific enhancer activity of this sequence was not demonstrated¹⁷. The identical 47.5 kb deletion was found in 15.3% (19/124) and 12.9% (17/132) of LWD patients in two separate cohorts^{17,18}. Another study found that 19 of 28 LWD patients with non-coding deletions downstream of *SHOX* carried the 47.5 kb deletion¹⁹. The penetrance of the LWD phenotype is variable in people carrying the 47.5 kb deletion, including a few that do not have short stature¹⁹. However, it is unclear if the variability is any greater than that of patients with disruptions of the *SHOX* coding sequence, which also show variable penetrance^{19,20}.

Candidate enhancer sequences from the human genome have been analyzed extensively using transgenic mouse assays²¹. This is complicated in the case of *SHOX*, since mice, the most common model for such studies, lack a *SHOX* ortholog. However, since the fundamental gene regulatory networks controlling limb development are common to all tetrapods^{22,23}, we predicted that *SHOX* regulatory elements would be faithfully active in a mouse transgenic model. Many studies support such an approach. For example, human-specific enhancer activities have been revealed in transgenic mice²⁴ and a recent study showed that the ZRS of the coelacanth fish has similar activity and can functionally substitute for the mouse ZRS despite an evolutionary separation of approximately 400 million years²⁵. In these studies, candidate sequences are cloned upstream of a minimal promoter and a reporter gene (usually lacZ). Embryos carrying the reporter transgene constructs are produced by pronuclear injection of one-cell mouse embryos using standard techniques. Although these constructs are incorporated randomly into the genome, position effects can be distinguished from specific and reproducible patterns by collecting multiple transgenic embryos, with each representing an independent genomic insertion. Typically, an enhancer is scored as having a specific activity when three embryos display a similar pattern of expression²¹. Such studies have proven to be an effective method to identify the precise sequences whose modification or removal cause human disease. For example, the ZRS was identified by this technique³. Therefore, we have used a similar strategy to search for the enhancer(s) controlling *SHOX* limb expression.

Previously, we used a transgenic mouse assay to show that an 18,796 bp (abbreviated 18.8 kb) genomic fragment containing CNE9 has enhancer activity in developing limbs, specifically in the zeugopodal region where *SHOX* function is critical (genomic location and activity is shown in Fig. 1B,C)²⁶. This result supported the validity of using a mouse transgenic model to study *SHOX* enhancers. We also found that a reporter transgene containing only CNE9 did not have reproducible limb expression, suggesting that the limb-specific activity lies elsewhere within the 18.8 kb fragment²⁶. We will refer to the activity within the 18.8 kb transgene as the ZED, or zeugopodal enhancer downstream of *SHOX*. Here we report that ZED activity is located within a previously overlooked 563 bp fragment that contains sequences conserved in non-rodent mammals, but not in other vertebrates. We also demonstrate that putative HOX binding sites are required for the enhancer's activity *in vitro*.

Results

The ZED is found within an 11.5 kb genomic fragment that does not contain CNE9. We sought to find the precise genomic location of the ZED and to determine if it is removed by the 47.5 kb recurrent deletion (Fig. 1B). We first used a transgenic mouse assay to screen 11.5 and 7.2 kb sub-fragments of the 18.8 kb transgene previously shown to contain ZED activity²⁶. The subdivision of the 18.8 kb fragment was guided by the breakpoint of the recurrent deletion (genomic coordinate ChrX:828,092 (GRCh37/hg19)). For cloning, we used a SnaBI restriction site located 426 bp centromeric to the deletion breakpoint, so that all sequences of the 18.8 kb fragment that are removed by the recurrent deletion are found in the 11.5 kb fragment (Fig. 1B). In order to determine whether there were other limb enhancers within the 47.5 kb deletion, we also used the transgenic assay to test the enhancer activity of 22 kb and 15 kb fragments that spanned sequences of the 47.5 kb deletion that are not within the 11.5 kb fragment (Fig. 1B). In each case, the assay was performed by cloning the genomic fragments upstream of a human β -globin minimal promoter adjacent to lacZ (β lacZ) as previously described²⁶. The resulting constructs were injected into pronuclei of one-cell stage embryos and transient transgenic embryos were isolated 13 or 14 days later²⁶. In one case, permanent lines were created (the 7.2 kb β lacZ transgene). Embryos of this developmental range were chosen because of the previously determined timing of the 18.8 kb β lacZ transgene activity²⁶ and because *SHOX* is expressed at a comparable stage in human fetuses²⁷. Furthermore, *SHOX* mutations are known to cause shortened, malformed limbs as early as 12 weeks gestation, indicating an early function for *SHOX* in developing limbs²⁸.

In total, we used the transgenic assay to screen 55 kb of the human genome for enhancer activity, including all of the sequences removed in the recurrent 47.5 kb deletion, plus 7.2 kb extending further centromeric to and including CNE9 (Fig. 1B). Three representative E13.5/E14.5 transgenic embryos for each fragment are shown in Fig. 1C with the complete results summarized in Table 1. Among the transgenes tested, only the 11.5 kb β lacZ transgene showed reproducible limb enhancer activity with staining similar to that of 18.8 kb β lacZ (Fig. 1C). Although one of six lines carrying the 7.2 kb β lacZ transgene (at right, Fig. 1C) showed limb staining, this was judged to be a position effect since its activity was distinct from that of the ZED (see Supplementary Fig. S1) and it was present in only that single line. Therefore, the 11.5 kb fragment contains the ZED and this is the only limb

Fragment abbreviation	Precise Size of fragment tested (bp)	Total number of transgenic embryos [†]	Number with any limb expression	Number with 'ZED-like' limb staining [‡]	Cloning by: endogenous restriction site, retrieval or PCR [§]	Genomic coordinates ^{***}
11.5 kb	11,514	N/D*	6	5	SnaBI	816,903–828,416
7.2 kb	7,232	6	1	0	SnaBI	828,467–835,698
22 kb	22,421	3	0	0	Retrieval	780,255–802,675
15 kb	15,028	3	0	0	Retrieval	802,264–817,291
5.2 kb	5,288	7	0	0	HindIII	816,903–822,190
6.2 kb	6,230	N/D	7	7	HindIII	822,187–828,416
2.9 kb	2,868	4	1	0	NruI	822,187–825,054
3.3 kb	3,362	N/D	8	8	NruI	825,055–828,416
1.3 kb	1,375	9	3	3	PCR	827,128–828,502
930 bp	930**	7	5	5	PCR	827,128–828,057
461 bp	461**	6	3	1 [§]	PCR	828,042–828,502
563 bp	563	12	7	6	PCR	827,128–827,691
2,774 bp	2,774	5	5	5	PCR	827,128–829,901
CNE7	656	12	3	0	PCR	780,580–781,235

Table 1. Summary of Transgenic Analysis. [†]As determined by PCR on DNA from yolk sacs. [‡]Refers to staining that gave the same pattern in the limbs as the 18.8 kb transgene. [§]Subcloning was from the 18.8 kb β lacZ clone, using these restrictions sites and bordering sites from the multiple cloning site. The SnaBI digest removed 51 bp, but this sequence was retained in the 1,375 and 461 bp fragments. [§]We scored this single embryo as “ZED-like” since dark staining covered the entire ZED domain (embryo at extreme lower right in Fig. 2E). ^{*}Genotypes of all embryos were not determined (N/D) during intermediate screening steps when obvious ‘ZED-like’ staining was obtained in multiple embryos. ^{**}The 930 and 461 bp PCR products had 16 bp of overlap. ^{***}All coordinates are for Chromosome X, Genome version GRCh37/hg19.

enhancer detected within the 55 kb as tested in our E13.5/E14.5 transgenic embryo assay. Notably, the 11.5 kb ZED fragment does not contain any of the highly conserved sequences previously suspected to function as *SHOX* limb enhancers (CNE7, CNE8 or CNE9) (Fig. 1B)^{16,17}.

A 563 bp fragment is sufficient for ZED activity in a transgenic mouse assay. We subsequently subdivided the 11.5 kb sequence into progressively smaller fragments while tracking ZED activity with our transgenic embryo assay (Fig. 2). This strategy ultimately identified a 563 bp sequence that was sufficient to generate the ZED expression pattern in transgenic embryos (Fig. 2E,F). The genomic coordinates of this fragment are ChrX:827,128–827,691 in GRCh37/hg19, which is approximately 7 kb telomeric to CNE9 and 401 bp telomeric to the 47.5 kb deletion breakpoint. For all positive fragments, ZED transgene expression in the forelimb is characterized by a narrow posterior domain at the elbow (arrows in Fig. 2F) that becomes broader as it extends distally to the wrist. Hindlimb expression is similarly stereotypical for each of the positive fragments, but in this case extending more proximally into the stylopodal domain (femur) as well as the zeugopod. The complete set of embryos with reporter expression in limbs is shown in Supplementary Fig. S1. The 1,987 bp telomeric sub-fragment of the 3,362 bp β lacZ transgene was not tested in the transgenic assay because it mostly consisted of repetitive sequences (Fig. 2A,D) and we successfully predicted that the 1,375 bp fragment would contain the ZED activity.

At its endogenous location, the ZED is found directly adjacent to a region of high sequence conservation that is enriched for the activating acetylation of histone H3 on lysine 27 (H3K27ac) in human embryonic limbs²⁹ (location indicated by the blue arrow in Figs 1B and 2D). Although this adjacent sequence did not have reproducible limb activity when tested as part of the 7.2 kb β lacZ transgene, we wondered if it might modulate ZED activity when combined with the active 563 bp enhancer sequence. Therefore, we created a 2,774 bp β lacZ transgenic construct that spanned the ZED and the more centromeric conserved region and tested it in our transgenic assay. The embryos carrying this “broader ZED” transgene had enhancer activity that was indistinguishable from the other ZED constructs, indicating that this additional sequence did not affect ZED activity in the context of a transgene (Fig. 2E,F and Supplementary Fig. S1). Nonetheless, we hypothesize that the H3K27ac/conserved sequence may participate in ZED function in its remote, endogenous chromosomal location (see discussion).

Another sequence at the telomeric extreme of the 47.5 kb deletion interval has been implicated as a potential limb enhancer based on its interaction with the *SHOX* promoter in a chromosome conformation capture assay (3C) performed on samples from chick limb buds¹⁷. This element contains the CNE7 sequence that was negative for limb expression in chick electroporation assays¹⁶. We tested the human CNE7 sequence in our transgenic assay and found that among 12 transgenic E12.5 embryos, three had detectable limb activity (Supplementary Fig. S1). We chose E12.5 for CNE7 transgenic analysis since this corresponds to the chick stage (HH26) used in the 3C assay¹⁷. Only one CNE7 transgenic embryo showed strong limb expression, and in that case, the expression was in the distal limb where *SHOX* is not expressed, and likely represents a position effect. In addition, the 22 kb transgene containing CNE7 did not show any limb enhancer activity at E14.5 (Fig. 1C). Taken together the data from our assays do not support a role for CNE7 in the control of *SHOX* expression in limbs.

A primary limb bud cell luciferase assay identifies active ZED sequences and implicates putative HOX binding sites in its regulation. In order to facilitate the more efficient screening of sequences

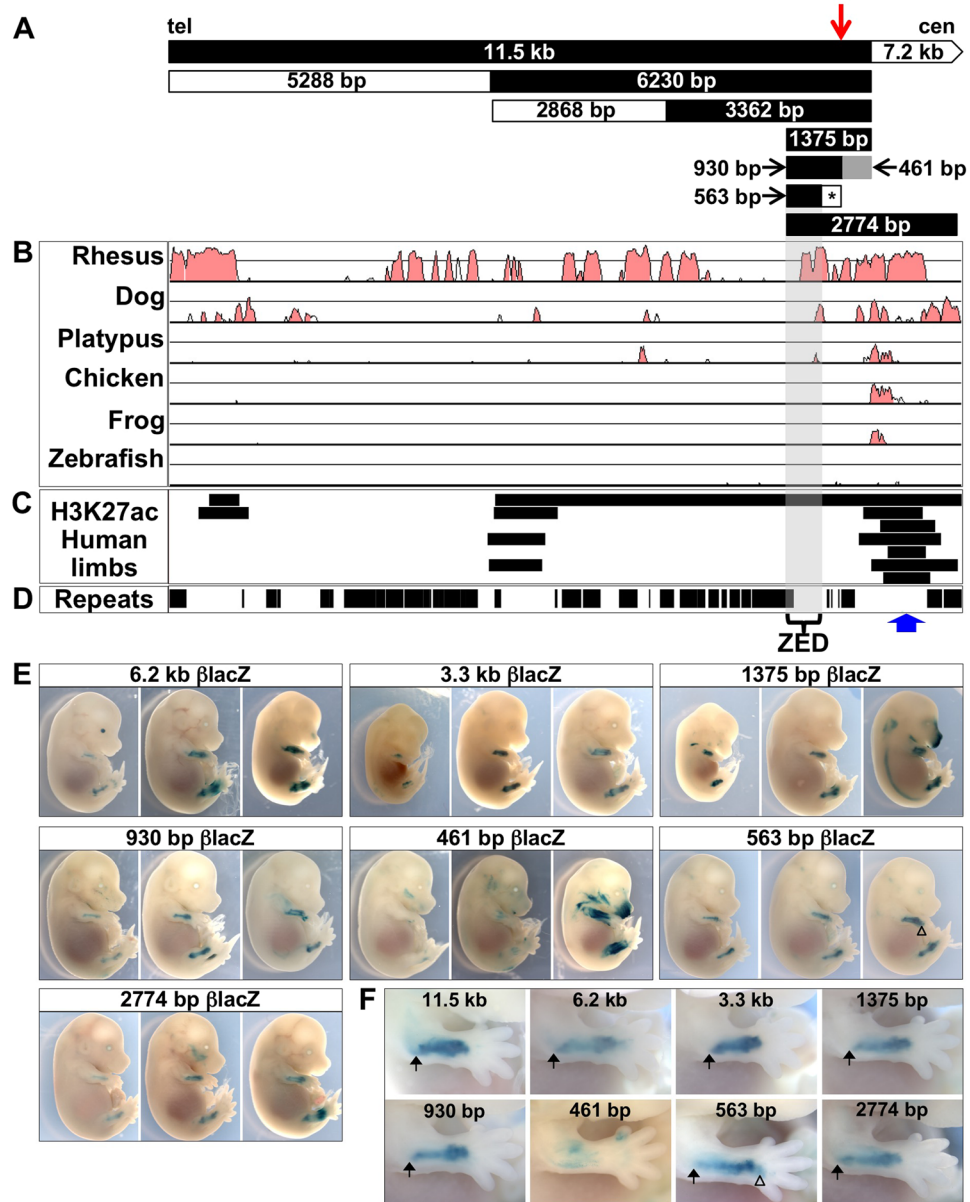


Figure 2. A 563 bp sequence is sufficient for ZED transgene activity. **(A)** Schematic map of nine fragments tested in the transgenic mouse assay, black fragments reproducibly tested positive for activity similar to that of the 11.5 kb βlacZ transgene. The 461 bp fragment is shaded grey to indicate its variable staining. The asterisk marks a 324 bp fragment that tested negative in a luciferase assay (Fig. 3C). **(B)** VISTA plot of evolutionary conservation among the indicated species aligned with the fragments from A. The 563 bp sequence with ZED activity is shaded and contains a small peak of conservation among non-rodent mammals (see Fig. 3). **(C)** Regions of H3K27ac in human embryonic limbs arranged as in Fig. 1B from Cotney *et al.*²⁹. Note that the only region with enriched H3K27ac in all samples is within 1 kb (centromeric (cen)) of the 563 bp ZED and contains sequences conserved in mammals, chicken and frogs. **(D)** Repetitive elements from the UCSC genome browser. The blue arrow indicates the location of enriched H3K27ac in seven human limb samples²⁹. **(E)** Sets of three E13.5/E14.5 embryos carrying the indicated transgenes. For each set, the embryo with the most staining obtained is shown at right, with a more moderately stained embryo in the middle and a weakly stained embryo at left. **(F)** Close-up of the forelimbs from each of the middle transgenic embryos from E, with fragment sizes indicated. The arrows show the location of the elbow where ZED staining is restricted to the posterior of the limb. The open triangles indicate area of gained expression in the autopod in the 563 bp βlacZ transgenics.

for enhancer activity, we transfected luciferase reporter plasmids containing various candidate sequences into dissociated E11.5 mouse limb bud cells, adapting methods previously used for chick limb buds³⁰. After transfection, the primary cells were cultured for 40 to 45 hours in micromass culture before measuring firefly and *Renilla* luciferase activity (the latter is for normalization, see Methods). We first validated this method by measuring

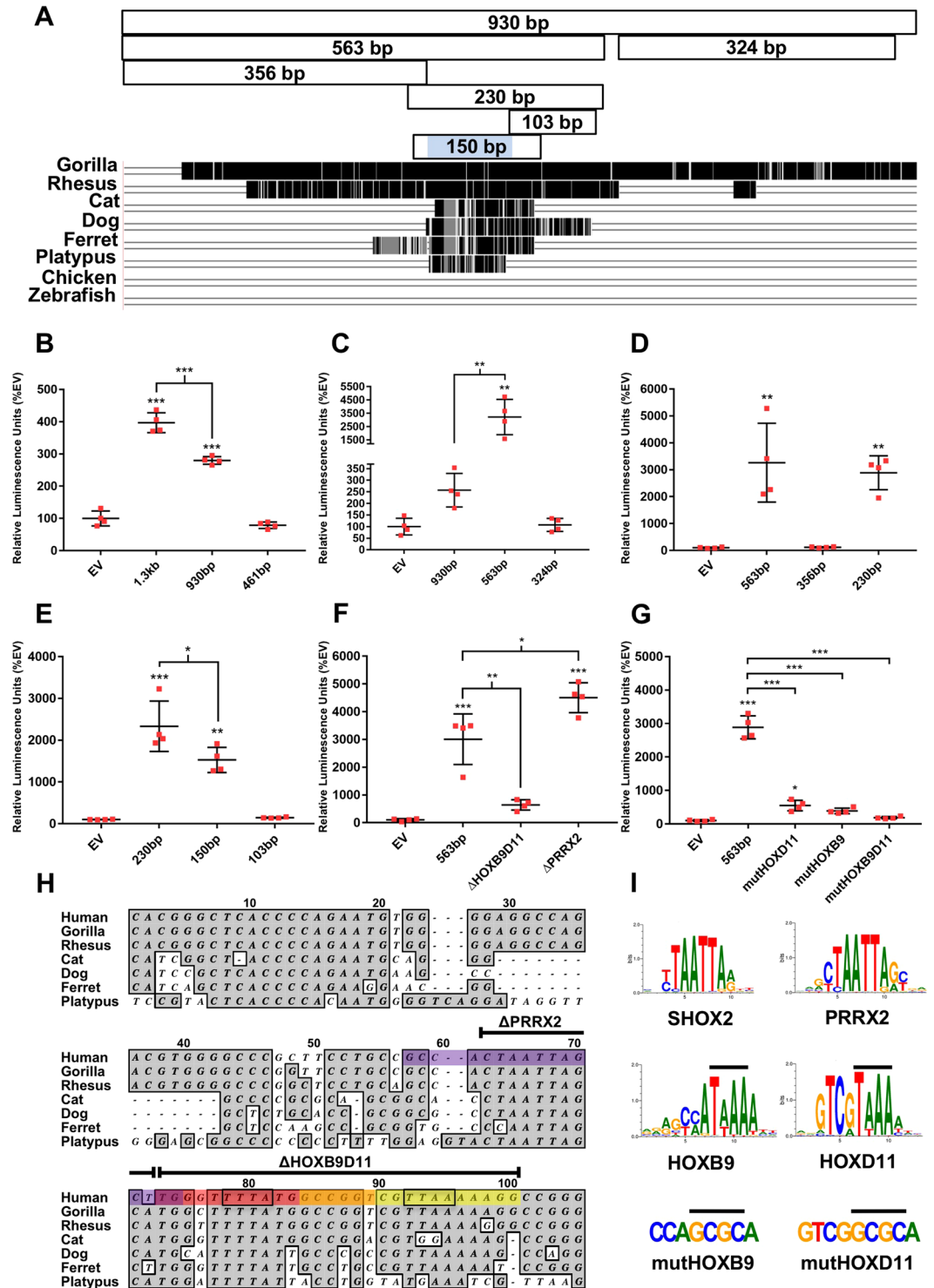


Figure 3. A Primary Limb Bud Cell Luciferase Assay further delineates ZED sequences with enhancer activity and identifies putative HOX binding sites required for ZED activity. (A) Map of sub-fragments of the 930 bp fragment tested with the luciferase assay, aligned above a track from the UCSC genome browser showing evolutionary conservation of the 930 bp in the indicated species. (B–E) Each panel represents a separate luciferase assay with the indicated fragments identified by their size. (F) Luciferase assay with the 563 bp ZED and the same fragment with a 28 bp deletion of the HOXB9/HOXD11 TFBSs (Δ HOXB9D11) or a 10 bp deletion of the core of two overlapping PRRX2 TFBSs (Δ PRRX2). The sequences deleted are shown in H. (G) Luciferase assay with the 563 bp ZED and the same fragment with site-specific mutations of the HOXD11 (mutHOXD11) or HOXB9 (mutHOXB9) TFBSs or mutations in both sites (mutHOXB9D11). The activity of each fragment was measured in duplicate limb bud samples from four E11.5 embryos with individual data points shown. Error bars show the SD. Statistical significance: * $p < 0.03$; ** $p < 0.003$; *** $p < 0.0001$ in one-way ANOVA analysis with a Tukey's post hoc test performed with GraphPad Prism. (H) Alignment of the conserved core sequence of the

ZED in the indicated seven mammals (location of the 100 bp core shaded blue in A). The yellow and red shaded bases indicate the predicted HOXD11 and HOXB9 binding sites respectively, as identified by MatInspector, with orange representing the overlap of the two sites. The HOXB9 binding site is on the complementary strand. The core of two overlapping PRRX2 binding sites (one on each strand) is shown in purple. The sequences deleted in the constructs used in F are indicated. The bases mutagenized in the HOXB9 and HOXD11 sites are boxed. (I) The consensus binding sites for SHOX2, PRRX2, HOXB9 and HOXD11³⁶. Images from Matbase (Genomatix). Lines above the HOXB9 and HOXD11 consensus sites indicate the bases that were replaced by mutagenesis. The mutagenized core sequences are shown below, with the changed bases indicated by lines.

the enhancer activity of the 1,375 bp (abbreviated 1.3 kb) and 930 bp fragments (Fig. 3B). Interestingly, the 930 bp fragment had significantly less activity than the 1,375 bp fragment in this assay, although the remaining sub-fragment (461 bp) did not have any measurable activity (Fig. 3B), perhaps indicating that this fragment contributes to the magnitude of ZED activity.

We then used the primary limb bud cell luciferase assay to analyze two sub-fragments of the 930 bp, which identified the 563 bp fragment as the active sequence (Fig. 3C). As noted above, our transgenic assay confirmed that this 563 bp fragment contains limb enhancer activity *in vivo* (Fig. 2E,F). Surprisingly, the 563 bp fragment was found to have markedly higher activity (approximately 10-fold) in the luciferase assay than the larger 930 bp sequence. We hypothesize that the remaining 367 bp contains repressive sequences that may be involved in fine tuning the expression domain of *SHOX*. This hypothesis is modestly supported by our transgenic results with the 563 bp enhancer, which consistently showed a slightly expanded expression domain into the autopod (open triangles, Fig. 2E,F). Note that the luciferase assay is performed on cells from earlier stage embryos (E11.5) than the E14.5 transgenic assay because micromass cultures are dependent on easily dissociated cells that have not yet terminally differentiated.

We further exploited the luciferase assay to delineate the active sequences within the 563 bp sequence (a map of the fragments tested is shown in Fig. 3A). These results support the conclusion that much of the enhancer activity is found within a 230 bp sub-fragment of the 563 bp (Fig. 3D). This 230 bp sequence contains a 100 bp interval that is well-conserved in non-rodent mammals, but not other tetrapods (Fig. 3A,H). A 150 bp fragment containing this conserved sequence also had high activity in the luciferase assay (Fig. 3E). We used the MatInspector software³¹ to identify candidate transcription factor binding sites (TFBSs) within the conserved 100 bp core. Among these were predicted binding sites for HOX9 and HOX11 proteins (red and yellow respectively in Fig. 3H; see discussion for a description of HOX9/11 proteins), as well as PRRX transcription factors (PRRX1/2) (purple, Fig. 3H). PRRX1/2 are paired-like homeodomain transcription factors that are closely related to SHOX and SHOX2³². These transcription factors all have prominent functions in stylopodal and/or zeugopodal limb development^{33–35}.

We tested the importance of the predicted TFBSs for ZED activity by deleting and mutagenizing the specific sequences from the 563 bp ZED. We then measured the activity of the resulting constructs in primary cell luciferase assays (Fig. 3). Deletion of the 28 bp containing the overlapping HOX9/HOX11 TFBSs dramatically reduced the activity of the ZED by approximately 80% in the luciferase assay (Fig. 3F). Furthermore, we used site-directed mutagenesis to determine the importance of individual HOX TFBS for ZED activity. Mutagenizing the core four bp of either the HOX9 or HOX11 TFBS (from a TAAA or TTAA, respectively to GCGC), reduced the activity of the ZED to a level comparable to deletion of both sites (compare Fig. 3G and F), indicating synergistic function of these sequences in regulating the ZED in this assay. Mutagenesis of both HOX sites in the same construct reduced ZED activity even further (Fig. 3G).

Surprisingly, deletion of the 10 bp core sequence of two overlapping PRRX2 binding sites significantly increased ZED activity by approximately 50% (Fig. 3F), indicating that this site might function to inhibit ZED activity. The putative PRRX2 binding site identified by MatInspector is very similar to the consensus sequence for the SHOX2 TFBS (Fig. 3I) and almost identical to the PRRX1 site³⁶. Furthermore, since the human SHOX homeodomain has an identical amino acid sequence to mouse and human SHOX2, these homeodomains are expected to bind the same TFBS. Therefore, this sequence could be part of negative feedback or fine tuning of ZED activity by SHOX, SHOX2 and/or PRRX1/2.

Screening of LWD patients for mutations within the ZED region. At the point in our study when the ZED had been narrowed down to a 930 bp fragment, we began a screen to determine if inactivating mutations or deletions within this interval existed in suspected LWD patients for whom no causative mutation had been found. We screened for *SHOX* mutations and deletions in the region using commercial Multiplex ligation-dependent probe amplification (MLPA) assays (P018E1, F1 or G1, MRC Holland, The Netherlands), self-designed MLPA assays for the upstream region of *SHOX*³⁷ and sequences flanking *SHOX* exon 6a³⁸, and High resolution melting (HRM) and DNA sequencing of the *SHOXa* transcript (NM_000451.3). A 1026 bp fragment containing all but the initial 12 bp of the 930 bp fragment was PCR amplified from patient DNA and screened for mutations by Sanger sequencing. Eleven variants, nine single nucleotide variants (SNVs) and two single nucleotide insertions were identified in the cohort of 124 suspected LWD patients (Supplementary Table S2). Two of the SNVs, rs60406052 and rs142683771, present in this cohort were absent from the Spanish normal stature control cohort. However, all nine SNVs are now known to be present at different frequencies in some populations (gnomAD³⁹, Supplementary Table S2) and are therefore unlikely to contribute to the LWD phenotype. Deletions and duplications in the region of interest were screened using a custom-designed MLPA assay in 104 of the 124 suspected LWD patients for whom sufficient DNA was available (Supplementary Table S3). No deletions or duplications were detected.

Discussion

The majority of non-coding deletions linked to a *SHOX*-deficient phenotype remove the ZED^{9,13,15,17,18,40}. Several of these, including the recurrent 47.5 kb deletion, leave the nearby CNE9 sequence intact. In contrast, to our knowledge no pathogenic deletions have been reported that remove CNE9 but leave the ZED intact. Therefore, we propose that the ZED is a more likely candidate than CNE9 to be a critical limb enhancer in the analyzed downstream region. In this view, CNE9 was implicated as a limb enhancer in part due to its proximity to the ZED. The ZED sequence was apparently overlooked in earlier studies because of its weak conservation; however, others have reported developmental enhancers with poor evolutionary conservation^{41,42}. Our results demonstrate the utility of methodically analyzing a deletion interval for enhancer activity rather than assuming that an enhancer will be the most conserved sequence in a genomic region. Nonetheless, the other CNEs identified to date may be *SHOX* enhancers in other spatial or temporal contexts.

Importantly, the ZED is located in a position that could potentially explain the pathogenicity of the 47.5 kb recurrent deletion. Specifically, the centromeric breakpoint of the 47.5 kb deletion is located within the 461 bp fragment, 37 bp from the centromeric boundary of the 930 bp fragment such that it removes the enhancer from the genome (breakpoint shown by the red arrow, Fig. 2A). Since the 461 bp fragment displayed some limb enhancer activity in our transgenic assay, its retention after deletion of the 47.5 kb fragment could potentially support some degree of limb expression. Studies of families with this 47.5 kb deletion have shown that while some individuals may have a few or all of the LWD clinical characteristics, other family members with the deletion have normal limbs^{17,18}. Thus, this deletion appears to have an incomplete or reduced penetrance. Despite the identification of a novel limb-specific enhancer within this deleted sequence, we still cannot explain the observed incomplete penetrance and the variable clinical heterogeneity. However, even mutations removing or mutating *SHOX* coding sequences have remarkably variable penetrance. This issue has recently been addressed by Montalbano *et al.* who reported a *SHOX* modifier, retinoic acid catabolising enzyme *CYP26C1*⁴³. *CYP26C1* mutations were identified in three families in which pathogenic *SHOX* mutations were also present. Only the individuals with both mutations presented with the classical LWD phenotype. However, the frequency of mutations in this modifier was very low with only two cases being identified in a cohort of 68 LWD individuals. The estimated frequency of having a mutation in both *SHOX* and *CYP26C1* was 1.8×10^{-5} . Thus, other modifiers remain to be identified.

A core ~100 bp of the ZED sequence is conserved in mammals, with the exception of rodents (Fig. 3A,H). The identification of a relatively short sequence responsible for ZED activity could serve as a gateway for identifying how *SHOX* is integrated into the gene regulatory networks that control limb patterning, especially since the ZED contains conserved, putative TFBSs for the HOXB9 and HOXD11 proteins (Fig. 3). Since HOXB9 does not function in patterning the proximal-distal axis of the limb, it is important to note that the identified HOXB9 TFBS is almost identical to that of the paralogous HOXA9/D9 proteins^{36,44} (collectively referred to as the HOX9 proteins) that are expressed in the developing zeugopod and have an important function in proximal limb development⁴⁵. In addition, HOXA9 has previously been implicated in the regulation of *SHOX*⁴⁶. Similarly, the predicted HOXD11 TFBS is identical to that of its paralog HOXA11, as determined by protein-binding microarrays³⁶. The *HOXA11/HOXD11* genes are particularly intriguing candidates as potential upstream regulators of *SHOX* since their mutation in mice causes zeugopodal truncations similar to that of Langer mesomelic dysplasia patients⁴⁷. Significantly, our results show that predicted HOX9 and HOX11 TFBSs are each required for the majority of ZED activity in a luciferase assay. Therefore, the ZED sequence could be an important tool for deciphering how *SHOX* fits into the HOX limb patterning system. In future experiments beyond the scope of the present study, a transgenic assay will be used to determine whether HOX9 and HOX11 proteins are required for ZED enhancer activity *in vivo*.

Analysis of epigenetic chromatin modifications has been another fruitful strategy for determining the genomic location of enhancers⁴⁸. As noted above, the wild-type mouse genome cannot be used for this study since it lacks *SHOX*. However, Cotney *et al.* reported the genome-wide sites of H3K27ac, which labels active promoters and enhancers, in human embryonic limbs⁴⁹. The regions showing enriched H3K27ac within the *SHOX* enhancer deletion interval are shown for limb tissue samples from E33, E41, E44 and E47 in the lower track in Fig. 1B. Interestingly, the most highly conserved sequences within the interval are either completely negative for H3K27ac enrichment (CNE7) or positive in only one of the seven samples (CNE8/CNE9), further suggesting that these sequences may not be limb enhancers. In contrast, the only sequence positive in all seven samples is within 1 kb of the ZED sequence (Blue arrow, Figs 1B and 2C). We did not detect any change in ZED activity when the more highly conserved/H3K27ac sequences were included in a transgene construct (2,744 bp βlacZ, Fig. 2E,F). However, the epigenetic modification of these neighboring regions could perhaps contribute to ZED function in its endogenous chromosomal context if, as we hypothesize, the enhancer must act at a distance to contact and activate the *SHOX* gene in developing limbs. The ZRS has been shown to contain sequences necessary for spatial activation of gene expression, such as those revealed by transgene assays, and other adjacent, yet distinct sequences required for long-range function⁴⁹. Therefore, the ZED, like the ZRS, may have adjacent sequences that are required for function from its endogenous location. Ultimately, elucidating the definitive function of the ZED and neighboring sequences will likely require genomic targeting experiments in a non-rodent mammalian model whose genome contains both *SHOX* and the ZED. Furthermore, non-mammalian vertebrate genomes could be tested for functional ZED orthologs that have evaded detection because of a lack of obvious sequence conservation.

Although the ZED's location and specific activity make it a promising candidate as a discrete *SHOX* limb enhancer, several reports of copy number variations (CNVs) in LWD and ISS patients implicate other genomic regions in the *cis*-regulation of *SHOX*. Three ISS families/patients have been reported with non-coding deletions upstream of *SHOX*^{37,50} and another study reported a homozygous >500 kb downstream deletion in members of a consanguineous family with LWD⁵¹. This latter deletion begins ~100 kb further downstream of *SHOX* than the ZED. Several studies have also linked duplications of PAR1 regions to LWD and ISS^{50,52–54}. In some cases, these

duplications include the *SHOX* gene itself^{52,54}, whereas several affect only upstream or downstream non-coding regions, some of which include the ZED⁵³. These duplications may interfere with the regulatory interactions necessary for *SHOX* expression, perhaps including perturbation of ZED function. Chromosomal interactions are now known to be confined mostly to discrete ~1 Mb regions called topologically associating domains (TADs)⁵⁵. The disruption of TAD boundaries has been shown to alter interactions of limb enhancers with their target genes resulting in limb abnormalities in mice and humans⁸. *SHOX* is near the center of a TAD that consists of the most telomeric ~1.2 Mb of the X and Y chromosomes, with a boundary just telomeric to the *CRLF2* gene as determined in a variety of cell types^{56,57}. Several CNVs disrupt the boundary of this TAD⁵⁰, which may cause misregulation of the *SHOX* gene resulting in LWD and ISS in these patients. Our screen of 124 suspected LWD patients without a known causative mutation did not reveal any deletions or SNVs of the minimal ZED region that are not found in the unaffected, general population. This suggests that the ~47.5 kb deletion may be the smallest, common pathogenic deletion that removes the ZED. Taken together, these results highlight a need for further studies of the components and extent of the *cis*-regulatory landscape of the *SHOX* gene in order to understand the functional significance of CNVs in this genomic region.

Methods

Cloning of transgene constructs. All genomic fragments tested in transgenic constructs were cloned into the p β lacZ plasmid²⁶, either as restriction fragments of the 18.8 kb β lacZ transgene, PCR products amplified from the same transgene, or by retrieval from bacterial artificial chromosomes (BACs) by gap repair as indicated in Table 1 and as described²⁶. The primers used for cloning are listed in Supplementary Table S1. The 22 kb and 15 kb fragments were cloned by retrieval directly into p β lacZ from BACs RP13-76L22 and RP13-167H21, respectively. BACs were obtained from the Wellcome Trust Sanger Institute. CNE7 was generated by PCR from BAC RP13-76L22. The CNE7 PCR fragment used was identical to the ECR1 fragment from Benito-Sanz *et al.*¹⁷, which includes all of the CNE7 sequence analyzed by Sabherwal *et al.* as well as 135 bp of flanking sequence. All fragments generated by PCR were confirmed by sequencing. Retrieved fragments were confirmed by sequencing at each end and restriction enzyme analysis.

Transgenic mice. All experiments using mice were performed in accordance with Canadian Council on Animal Care guidelines as approved by the University of Calgary Life and Environmental Sciences Animal Care Committee, Protocol # AC13-0053. Transgenic embryos and permanent lines were produced by pronuclear injection of DNA constructs into CD-1 one-cell stage embryos using standard techniques⁵⁸. All transient transgenic embryos and lines were produced at the University of Calgary Centre for Mouse Genomics. Mice were genotyped by forward primers specific for each transgene and a common reverse primer within the lacZ sequence. Transgenic embryos were stained with X-gal using standard techniques⁵⁸.

Patient and control cohorts. A total of 124 patients with suspected LWD, without alterations in *SHOX* or its known enhancers were included in this study which was performed in accordance with the tenets of the Declaration of Helsinki. The inclusion criteria included the presence of short stature and/or mesomelic shortening of the limbs and/or bilateral Madelung deformity in the proband or a direct family member. All participants provided informed consent for the performed studies and ethical approval was obtained from Hospital La Paz, Madrid (PI-1387, PI-2630). Genomic DNA was isolated from peripheral blood using the Blood kit (QIAGEN, Valencia, CA) or Chemagic DNA extraction special kit (Chemagen, Perkin Elmer Inc, Germany). A cohort of 126 healthy controls with heights within the normal range for the Spanish population for age and gender were obtained from the Spanish DNA bank (University of Salamanca, Salamanca, Spain).

Mutation Sequence Analysis. The screening of a 1026 bp fragment that included all but the most telomeric 12 bp of the 930 bp fragment with expression in the limbs of transgenic embryos was performed using direct sequencing. The slight difference in fragment size was due to the difficulties in primer design for amplifying sequences from human genomic DNA. The oligonucleotides for PCR amplification were RCE-P4_F1 5'-ATGGACACACAACACAGTTT-3' and RCENP4_R1 5'-AATCGTGACCATCATACTC-3'. Amplified products were directly sequenced with the amplification primers and four internal sequence primers RCE-P4_SeqF1 5'-ACTTTAAGTTTATAGGTACAGGTG-3', RCE-P4_SeqR1 5'-GGCGTCCTCATGGGGA-3', RCE-P4_SeqF2 5'-CCGAGACGTCGTCACGA-3' and RCE-P4_SeqF3 5'-GGGTTTGCAGCAAAGTG-3'. The PCR product was subsequently sequenced using the BigDye Terminator V3.1 kit (Life Technologies, Thermo Scientific Inc) and electrophoresed on an ABI3730XL sequencer (Applied Biosystems).

Population frequencies of the detected variants were assessed by screening 126 Spanish normal height controls and using the Genome Aggregation Database (gnomAD browser beta, <http://gnomad.broadinstitute.org/>) or dbSNP (<https://www.ncbi.nlm.nih.gov/>).

PAR1 deletion/duplication detection. Screening for the presence of deletions and duplications within the 930 bp region in the *SHOX* downstream region was undertaken using a custom designed Multiplex ligation-dependent probe amplification (MLPA) which encompassed ~11 kb (ChrX:817,716 to 828,838 in GRCh37/hg19, Supplementary Table 3) and analyzed as previously described¹⁷.

Primary Cell Luciferase Assay. Each biological replicate for the luciferase assay consisted of cells isolated from fore- and hindlimb buds dissected from a single CD-1 E11.5 embryo as described⁵⁹, with minor modifications using n = 4 littermates per condition. Limb buds were incubated in 0.8 units/ml dispase (Gibco) for 1.5 hours at room temperature, followed by 0.25% trypsin for 20 minutes at 37 °C. Single-cell suspensions were then created by repeated pipetting of the limb buds in micromass media (3:2 F12:DMEM, 10%FBS, 100 μ g/ml

Normocin) and diluted to approximately 1×10^7 cells/ml for each animal. Cells from 20 μ l of each suspension were transfected as described⁵⁹ using 1 μ g of reporter plasmid (the enhancer construct as cloned into pGL4.23 *luc2*/minP plasmid (Promega)) together with 5 ng pRL-TK *Renilla* (Promega). 10 μ l micromass cultures ($\sim 1 \times 10^5$ cells) were spotted in duplicate for each animal and allowed to adhere for 60–70 minutes at 37 °C, 5% CO₂, before feeding with 600 μ l media and incubating an additional ~40–45 hours. Firefly luciferase luminescence was read for each well ~1 hr after addition of Dual-Glo luciferase reagent (Promega), then incubated a further 20–30 minutes after adding Dual-Glo Stop & Glo reagent before reading *Renilla* luciferase luminescence. The ratios of firefly to *Renilla* luminescence values were calculated for each well, and averaged between the duplicate cultures. The means (+/–standard deviation (SD)) of four of these normalized values (biological replicates) were calculated and presented as a percentage of the values obtained for empty vector (EV) transfections. GraphPad Prism was used for statistical analysis using a one-way ANOVA with Tukey's post hoc test for multiple comparisons.

Deletions and site-directed mutagenesis within ZED luciferase constructs. All deletions and mutations were introduced into the 563 bp ZED fragment. Deletions were created with PCR SOEing^{60,61} using the primers in Supplementary Table S1 and the 563 bp fragment as a template. For each deletion two PCR products were synthesized with Precision Taq (abm, Inc.). The products were mixed and a filling reaction without primers was performed as described⁶¹, followed by the SOEing reaction using the forward primer from Product 1 and the reverse primer from Product 2. The resulting fragment was cloned into pGL4.23 *luc2*/minP after Acc65I/SalI digestion. In order to simultaneously remove predicted HOXB9 and HOXD11 binding sites, 28 bp were deleted as shown in Fig. 3H to produce construct Δ HOXB9D11. Similarly, the core of two predicted and overlapping PRRX2 binding sites were deleted by removing 10 bp (Fig. 3H) resulting in construct Δ PRRX2.

Mutations of specific TFBSs were performed on the 563 bp ZED-pGL4.23 *luc2*/minP plasmid with the primers in Supplementary Table S1 using the QuikChange II site-directed mutagenesis kit (Agilent) according to the manufacturer's instructions. In each case a core sequence of TAAA (HOXB9) or TTAA (HOXD11) (boxed sequences in Fig. 3H) was changed to GCGC as described⁶² (Fig. 3I), to produce the mutHOXB9 and mutHOXD11 site. A second reaction was performed to mutagenize the HOXD11 binding site within the mutHOXB9 construct resulting in the double-mutant mutHOXB9D11 plasmid. All deletions and mutations were confirmed by sequencing and the verified constructs were used in the primary cell luciferase assay as described above.

Sequence analysis. Genomic sequences from all species were retrieved from the UCSC genome browser and uploaded to the VISTA website where they were aligned with MLAGAN and analyzed for conservation with mVISTA^{63,64}. All human sequences and tracks were from genome version GRCh37/h19. Candidate transcription factor binding sites were identified with MatInspector from Genomatix³¹.

References

- Bhatia, S. & Kleinjan, D. A. Disruption of long-range gene regulation in human genetic disease: a kaleidoscope of general principles, diverse mechanisms and unique phenotypic consequences. *Hum Genet* **133**, 815–845, <https://doi.org/10.1007/s00439-014-1424-6> (2014).
- Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat Rev Genet* **18**, 245–258, <https://doi.org/10.1038/nrg.2016.167> (2017).
- Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725–1735 (2003).
- Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res* **27**, 223–233, <https://doi.org/10.1101/gr.213066.116> (2017).
- Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243, <https://doi.org/10.1038/nature25461> (2018).
- Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132–1145, <https://doi.org/10.1016/j.cell.2011.10.023> (2011).
- Chen, H. *et al.* Heads, Shoulders, Elbows, Knees, and Toes: Modular Gdf5 Enhancers Control Different Joints in the Vertebrate Skeleton. *PLoS Genet* **12**, e1006454, <https://doi.org/10.1371/journal.pgen.1006454> (2016).
- Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025, <https://doi.org/10.1016/j.cell.2015.04.004> (2015).
- Benito-Sanz, S. *et al.* A novel class of Pseudoautosomal region 1 deletions downstream of SHOX is associated with Leri-Weill dyschondrosteosis. *Am J Hum Genet* **77**, 533–544, <https://doi.org/10.1086/449313> (2005).
- Marchini, A., Ogata, T. & Rappold, G. A. A Track Record on SHOX: From Basic Research to Complex Models and Therapy. *Endocr Rev* **37**, 417–448, <https://doi.org/10.1210/er.2016-1036> (2016).
- Bertorelli, R. *et al.* The homozygous deletion of the 3' enhancer of the SHOX gene causes Langer mesomelic dysplasia. *Clin Genet* **72**, 490–491, <https://doi.org/10.1111/j.1399-0004.2007.00875.x> (2007).
- Campos-Barros, A., Benito-Sanz, S., Ross, J. L., Zinn, A. R. & Heath, K. E. Compound heterozygosity of SHOX-encompassing and downstream PARI deletions results in Langer mesomelic dysplasia (LMD). *Am J Med Genet A* **143A**, 933–938, <https://doi.org/10.1002/ajmg.a.31676> (2007).
- Chen, J. *et al.* Enhancer deletions of the SHOX gene as a frequent cause of short stature: the essential role of a 250 kb downstream regulatory domain. *J Med Genet* **46**, 834–839, <https://doi.org/10.1136/jmg.2009.067785> (2009).
- Seki, A. *et al.* Skeletal Deformity Associated with SHOX Deficiency. *Clin Pediatr Endocrinol* **23**, 65–72, <https://doi.org/10.1297/cpe.23.65> (2014).
- Fukami, M., Kato, F., Tajima, T., Yokoya, S. & Ogata, T. Transactivation function of an approximately 800-bp evolutionarily conserved sequence at the SHOX 3' region: implication for the downstream enhancer. *Am J Hum Genet* **78**, 167–170, <https://doi.org/10.1086/499254> (2006).
- Sabherwal, N. *et al.* Long-range conserved non-coding SHOX sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients. *Hum Mol Genet* **16**, 210–222, <https://doi.org/10.1093/hmg/dll470> (2007).
- Benito-Sanz, S. *et al.* Identification of the first recurrent PARI deletion in Leri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *J Med Genet* **49**, 442–450, <https://doi.org/10.1136/jmedgenet-2011-100678> (2012).

18. Bunyan, D. J., Baker, K. R., Harvey, J. F. & Thomas, N. S. Diagnostic screening identifies a wide range of mutations involving the SHOX gene, including a common 47.5 kb deletion 160 kb downstream with a variable phenotypic effect. *Am J Med Genet A* **161A**, 1329–1338, <https://doi.org/10.1002/ajmg.a.35919> (2013).
19. Kant, S. G. *et al.* Phenotypic characterization of patients with deletions in the 3'-flanking SHOX region. *PeerJ* **1**, e35, <https://doi.org/10.7717/peerj.35> (2013).
20. Binder, G. *et al.* SHOX haploinsufficiency and Leri-Weill dyschondrosteosis: prevalence and growth failure in relation to mutation, sex, and degree of wrist deformity. *J Clin Endocrinol Metab* **89**, 4403–4408, <https://doi.org/10.1210/jc.2004-0591> (2004).
21. Pennacchio, L. A. *et al.* *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502, <https://doi.org/10.1038/nature05295> (2006).
22. Zeller, R., Lopez-Rios, J. & Zuniga, A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat Rev Genet* **10**, 845–858, <https://doi.org/10.1038/nrg2681> (2009).
23. Zuniga, A. Next generation limb development and evolution: old questions, new perspectives. *Development* **142**, 3810–3820, <https://doi.org/10.1242/dev.125757> (2015).
24. Prabhakar, S. *et al.* Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350, <https://doi.org/10.1126/science.1159974> (2008).
25. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633–642 e611, <https://doi.org/10.1016/j.cell.2016.09.028> (2016).
26. Rosin, J. M., Abassah-Oppong, S. & Cobb, J. Comparative transgenic analysis of enhancers from the human SHOX and mouse Shox2 genomic regions. *Hum Mol Genet* **22**, 3063–3076, <https://doi.org/10.1093/hmg/ddt163> (2013).
27. Clement-Jones, M. *et al.* The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome. *Hum Mol Genet* **9**, 695–702 (2000).
28. Ambrosetti, F., Palicelli, A., Bulfamante, G. & Rivasi, F. Langer mesomelic dysplasia in early fetuses: two cases and a literature review. *Fetal Pediatr Pathol* **33**, 71–83, <https://doi.org/10.3109/15513815.2013.807322> (2014).
29. Cotney, J. *et al.* The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* **154**, 185–196, <https://doi.org/10.1016/j.cell.2013.05.056> (2013).
30. Bobick, B. E. & Kulyk, W. M. The MEK-ERK signaling pathway is a negative regulator of cartilage-specific gene expression in embryonic limb mesenchyme. *J Biol Chem* **279**, 4588–4595, <https://doi.org/10.1074/jbc.M309805200> (2004).
31. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res* **23**, 4878–4884 (1995).
32. Galliot, B., de Vargas, C. & Miller, D. Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. *Dev Genes Evol* **209**, 186–197 (1999).
33. Cobb, J., Dierich, A., Huss-Garcia, Y. & Duboule, D. A mouse model for human short-stature syndromes identifies Shox2 as an upstream regulator of Runx2 during long-bone development. *Proc Natl Acad Sci USA* **103**, 4511–4515, <https://doi.org/10.1073/pnas.0510544103> (2006).
34. ten Berge, D. *et al.* Prx1 and Prx2 are upstream regulators of sonic hedgehog and control cell proliferation during mandibular arch morphogenesis. *Development* **128**, 2929–2938 (2001).
35. Zakany, J. & Duboule, D. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* **17**, 359–366, <https://doi.org/10.1016/j.gde.2007.05.011> (2007).
36. Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276, <https://doi.org/10.1016/j.cell.2008.05.024> (2008).
37. Benito-Sanz, S. *et al.* Identification of the first PAR1 deletion encompassing upstream SHOX enhancers in a family with idiopathic short stature. *Eur J Hum Genet* **20**, 125–127, <https://doi.org/10.1038/ejhg.2011.210> (2012).
38. Benito-Sanz, S. *et al.* Identification of 15 novel partial SHOX deletions and 13 partial duplications, and a review of the literature reveals intron 3 to be a hotspot region. *J Hum Genet* **62**, 229–234, <https://doi.org/10.1038/jhg.2016.113> (2017).
39. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291, <https://doi.org/10.1038/nature19057> (2016).
40. Huber, C., Rosilio, M., Munnich, A., Cormier-Daire, V. & French, S. G. M. High incidence of SHOX anomalies in individuals with short stature. *J Med Genet* **43**, 735–739, <https://doi.org/10.1136/jmg.2006.040998> (2006).
41. Blow, M. J. *et al.* ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* **42**, 806–810, <https://doi.org/10.1038/ng.650> (2010).
42. Fisher, S., Grice, E. A., Vinton, R. M., Bessling, S. L. & McCallion, A. S. Conservation of RET regulatory function from human to zebrafish without sequence similarity. *Science* **312**, 276–279, <https://doi.org/10.1126/science.1124070> (2006).
43. Montalbano, A. *et al.* Retinoic acid catabolizing enzyme CYP26C1 is a genetic modifier in SHOX deficiency. *EMBO Mol Med* **8**, 1455–1469, <https://doi.org/10.15252/emmm.201606623> (2016).
44. Jerkovic, I. *et al.* Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCE. *PLoS Genet* **13**, e1006567, <https://doi.org/10.1371/journal.pgen.1006567> (2017).
45. Fromental-Ramain, C. *et al.* Specific and redundant functions of the paralogous Hoxa-9 and Hoxd-9 genes in forelimb and axial skeleton patterning. *Development* **122**, 461–472 (1996).
46. Durand, C., Decker, E., Roeth, R., Schneider, K. U. & Rappold, G. The homeobox transcription factor HOXA9 is a regulator of SHOX in U2OS cells and chicken micromass cultures. *PLoS One* **7**, e45369, <https://doi.org/10.1371/journal.pone.0045369> (2012).
47. Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S. & Capecchi, M. R. Absence of radius and ulna in mice lacking hoxa-11 and hoxd-11. *Nature* **375**, 791–795, <https://doi.org/10.1038/375791a0> (1995).
48. Cotney, J. *et al.* Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res* **22**, 1069–1080, <https://doi.org/10.1101/gr.129817.111> (2012).
49. Lettice, L. A. *et al.* Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* **141**, 1715–1725, <https://doi.org/10.1242/dev.095430> (2014).
50. Verdin, H. *et al.* Profiling of conserved non-coding elements upstream of SHOX and functional characterisation of the SHOX cis-regulatory landscape. *Sci Rep* **5**, 17667, <https://doi.org/10.1038/srep17667> (2015).
51. Bunyan, D. J., Taylor, E. J., Maloney, V. K. & Blyth, M. Homozygosity for a novel deletion downstream of the SHOX gene provides evidence for an additional long range regulatory region with a mild phenotypic effect. *Am J Med Genet A* **164A**, 2764–2768, <https://doi.org/10.1002/ajmg.a.36724> (2014).
52. Benito-Sanz, S. *et al.* Clinical and molecular evaluation of SHOX/PAR1 duplications in Leri-Weill dyschondrosteosis (LWD) and idiopathic short stature (ISS). *J Clin Endocrinol Metab* **96**, E404–412, <https://doi.org/10.1210/jc.2010-1689> (2011).
53. Bunyan, D. J. *et al.* Duplications upstream and downstream of SHOX identified as novel causes of Leri-Weill dyschondrosteosis or idiopathic short stature. *Am J Med Genet A*, <https://doi.org/10.1002/ajmg.a.37524> (2015).
54. Fukami, M. *et al.* Rare pseudoautosomal copy-number variations involving SHOX and/or its flanking regions in individuals with and without short stature. *J Hum Genet* **60**, 553–556, <https://doi.org/10.1038/jhg.2015.53> (2015).
55. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380, <https://doi.org/10.1038/nature1082> (2012).
56. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336, <https://doi.org/10.1038/nature14222> (2015).

57. Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680, <https://doi.org/10.1016/j.cell.2014.11.021> (2014).
58. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. R. *Manipulating the mouse embryo, a Laboratory Manual, 3rd edition*. 3rd edn, (Cold Spring Harbor Press, 2003).
59. Bobick, B. E. & Cobb, J. Shox2 regulates progression through chondrogenesis in the mouse proximal limb. *J Cell Sci* **125**, 6071–6083, <https://doi.org/10.1242/jcs.111997> (2012).
60. Horton, R. M., Cai, Z. L., Ho, S. N. & Pease, L. R. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* **8**, 528–535 (1990).
61. Thornton, J. A. Splicing by Overlap Extension PCR to Obtain Hybrid DNA Products. *Methods Mol Biol* **1373**, 43–49, https://doi.org/10.1007/7651_2014_182 (2016).
62. Guerreiro, I. *et al.* Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci USA* **110**, 10682–10686, <https://doi.org/10.1073/pnas.1300592110> (2013).
63. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**, W273–279, <https://doi.org/10.1093/nar/gkh458> (2004).
64. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* **13**, 721–731, <https://doi.org/10.1101/gr.926603> (2003).

Acknowledgements

We thank Colleen Geary-Joo and Cameron Fielding of the Centre for Mouse Genomics at the University of Calgary for producing all the transgenic mice used in this study. We are grateful to Madison Peaker for technical assistance and to Andrew Cobb for insight into the ZED. This work was supported by the following grants: OGB-143064 to J.C. from the Institute of Genetics of the Canadian Institutes of Health Research (CIHR) and a sustainability grant from Alberta Innovates-Health Solutions to J.C. and SAF2012-30871 and SAF2015-66831-R from the MINECO to K.E.H. S.B.-S. was a recipient of a postdoctoral CIBERER fellowship.

Author Contributions

K.E.H. and J.C. designed the study; I.S., S.B.-S., B.E.B., J.M.R. and J.C. performed the experiments; I.S., S.B.-S., B.E.B., J.M.R. K.E.H. and J.C. analysed and interpreted the results; I.S., J.C. and K.E.H. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32565-1>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018