The conserved sonic hedgehog limb enhancer consists of discrete functional elements that regulate precise spatial expression

The conserved sonic hedgehog limb enhancer consists of discrete functional elements that regulate precise spatial expression.

Laura A Lettice, Paul Devenney, Carlo De Angelis, and Robert E Hill*

MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine,
University of Edinburgh, Edinburgh, United Kingdom EH4 2XU

*Corresponding author and lead contact:

Robert Hill

MRC Human Genetics Unit
MRC Institute of Genetics and Molecular Medicine
University of Edinburgh
Edinburgh, United Kingdom EH4 2XU

Email Bob.Hill@igmm.ed.ac.uk
Phone +44-(0)131-6518621

SUMMARY

Expression of sonic hedgehog (*Shh*) in the limb bud is regulated by an enhancer called the ZRS, which in evolution belongs to an ancient group of highly conserved *cis*-regulators found in all classes of vertebrates. Here, we examined the endogenous ZRS in mouse using genome editing to establish the relationship between enhancer composition and embryonic phenotype. We show that enhancer activity is a consolidation of distinct activity domains. Spatial restriction of *Shh* expression is mediated by a discrete repressor module; whereas, levels of gene expression are controlled by large overlapping domains containing varying numbers of HOXD binding sites. The number of HOXD binding sites regulate expression levels incrementally. Substantial portions of conserved sequence are dispensable indicating the presence of sequence redundancy. We propose a collective model for enhancer activity in which function is an integration of discrete expression activities and redundant components that drive robust expression.

Key Words: *Shh* expression, limb development, ZRS, enhancer, *HoxD* genes, Werner mesomelic syndrome, genome editing, phenotype

Highlights

- The ancient vertebrate enhancer, the ZRS, shows sequence plasticity.
- Discrete regulatory activities are assigned to specific sites in the enhancer.
- Number of HOXD binding sites determines the level of *Shh* expression.
- Robust expression is a collective of regulatory and redundant information.

52 In brief (eTOC)

53 Lettice et al. examine the composition of a highly conserved limb-specific enhancer,
54 the ZRS, by dissecting the endogenous sequence using genome editing. Analysis of the
55 resulting phenotype gives insights into the complex composition of the enhancer which
56 integrates discrete expression activities and redundant elements to drive accurate
57 spatiotemporal gene expression.

58

59

INTRODUCTION

The basis of embryonic development lies in the spatiotemporal control of gene expression, which is mediated by remote *cis*-regulatory elements. These *cis*-acting elements, or enhancers, are fundamental to evolution and disease. Despite these important roles, major unanswered questions remain about the information encoded by the enhancer sequence and the importance of the overall structural architecture to enhancer activity. One class of enhancers that operates during embryogenesis are those that are highly conserved acting at long distances from their target genes (Visel et al., 2009). Here, we focus on a highly conserved element called the ZRS that is responsible for the spatiotemporal expression of *Shh* during limb bud development (Lettice et al., 2003; Sagai et al., 2005) and is essential for specifying digit identity and number. This enhancer is ~770bp in length and shows a high degree of similarity in vertebrates across a lengthy evolutionary time scale including the sharks and rays (Dahn et al., 2007) and in accord, the mouse shows >70% similarity with the coelacanth (lobe finned fish) sequence (Fig. S1). Hence, the ZRS has remained highly invariant against a backdrop of major evolutionary changes to the anatomy of the appendicular skeleton which includes the transition of fish fins to tetrapod limbs (Gehrke and Shubin, 2016). The structural organisation of this class of deeply conserved vertebrate enhancers is under strong selective constraints and even in light of binding site redundancies exhibited by transcription factors few sequence changes are present.

The ZRS is located 800-1000 kb away from the *Shh* promoter in mouse and human and is necessary and sufficient for accurately activating and maintaining *Shh* expression in the limb (Lettice et al., 2003; Sagai et al., 2005). An enhancer evolves not simply as a regulator that switches gene expression on or off but must also solve the challenges of

regulating expression from a distance (Lettice et al., 2014) while controlling gene activity accurately in space and time and at the appropriate levels. Based on the evolutionary stasis of the ZRS, it is reasonable to expect that the sequence was finely honed during evolution such that that there is little tolerance for sequence change. Indeed point mutations in and duplications of the ZRS result in a spectrum of appendicular skeletal defects (Anderson et al., 2012). Point mutations in well over 20 different positions scattered across the ZRS cause autosomal dominant limb defects, called 'ZRS associated syndromes' (Wieczorek et al., 2010). Some of the conditions associated with ZRS mutations include preaxial polydactyly type 2, triphalangeal thumb polysyndactyly, syndactyly type 4 and Werner mesomelic syndrome (WMS).

To investigate the structural composition of this highly conserved vertebrate enhancer, we used genome editing technology (Dow, 2015) to target deletions in three regions within the ZRS. Since ZRS activity is limb specific, the phenotypes were expected to be overt, accessible and nonlethal. The regions that were targeted contain the 5bp site responsible for Werner's mesomelic syndrome (Anderson et al., 2012), the single mutation responsible for hemimelic extra toes (*Hx*) (Lettice et al., 2008) in mouse and a previously identified site for binding the HAND2 transcription factor (Osterwalder et al., 2014). This approach generated an overlapping series of mutations and deletions that scan across 250bp of the endogenous ZRS. Here, we show that the ZRS encodes multiple, diverse functions that contribute to the enhancer activity. Spatial restriction of expression is, in part, controlled by a small repressor domain which confines Shh expression to the posterior limb bud margin. In contrast, large overlapping domains regulate expression levels contingent on the number of HOXD binding sites. In addition, in response to insertion mutations cryptic, unique phenotypes were generated that revealed the functional plasticity

potentially encoded in an enhancer. Mutational analysis, however, also showed that even though the enhancer is highly conserved it could still tolerate quite substantial losses of sequence information without causing an abnormal phenotype. We propose a collective model for enhancer composition in which discrete activities and redundant sequences in the ZRS accrue to provide for a robust regulatory response during development.

RESULTS

ZRS Mutations in Mouse Mimics Werner Mesomelic Syndrome

The Werner mesomelic syndrome (WMS) is associated with point mutations in a single, short 5 bp stretch of the ZRS (green box, Fig. 1A) which results in preaxial polydactyly of the hands and feet but is uniquely associated with short limb dwarfism due to tibial hypoplasia. WMS results from any heterozygous point mutation at position 404 (in human) (Fig. 1A) (Lettice et al., 2008), a heterozygous A to G change two nucleotides downstream at position 406 (Norbnop et al., 2014) and a homozygous C to T change at position 402 (VanderMeer et al., 2014) (Green bases in Fig. 1A). These three nucleotide positions lie within a highly conserved site and to date is the only site association with this syndrome.

Initially, to examine the nature of these mutations in mouse and to ensure that it is possible to recreate the human abnormality, a G to A replacement in position 404 originally reported in a Cuban family (labelled as Cu; Fig.1A) (Lettice et al., 2003) was generated in mouse using conventional 'knock-in' technology (Lettice et al., 2014). The resulting heterozygous mice exhibited extra preaxial digits on the hindlimbs (Fig.1C) while homozygotes, in addition, had bent legs due to tibial dysplasia (Fig. 1D). Bone stains confirmed the loss of the terminal portion of each tibia (Fig. 1D), which copied the dysplastic

tibias of the WMS patients; however, unlike the patients the forelimbs in mice were unaffected and tibia dysplasia only occurred in the homozygous mutant.

To investigate further the nature of the dominant mutations at the WMS position, we targeted deletions using CRISPR/Cas9. A guide RNA (gRNA) targeted to this region (black box in Fig. 1A) resulted in a number of different deletions and insertions. The most common mutation that was recovered was the precise removal of the five basepairs (called the WMSΔ5 deletion) (Green box in Fig. 1A) implicated as the site of WMS. The hindlimb phenotype of the WMSΔ5 mutant mice is similar to that observed in the homozygous Cu mutant mice; in that, the hind limbs show extra preaxial digits and the tibia is hypoplastic ranging from a partial loss of the distal portion of the bone to its complete absence (Fig. 1E). In contrast to the point mutation, these phenotypes occurred in both heterozygous and homozygous WMSΔ5 mice and both genotypes exhibit PPD in the forelimbs (Fig. 1F). Thus the strength of the WMSΔ5 allele is similar to the point mutation in human. No differences in the severity of the phenotypes were observed between heterozygous and homozygous mice.

Analysis of *Shh* expression in the developing limb buds in the homozygous Cu mutant embryos showed normal expression at E10.5, while by E11.5 ectopic anterior expression of *Shh* (Fig. 1l) was observed in approximately half of the embryos examined (3/7 mice). By E12.5, ectopic *Shh* occurred at the anterior margin (Fig. 1J) in an outgrowth of limb tissue in all embryos examined. Heterozygous embryos showed normal *Shh* expression at all stages examined but analysis of *Ptc1*, a sensitive readout of *Shh* signalling, showed ectopic, anterior expression at both E11.5 and E12.5 (Fig. 1K, L) showing that low but sufficient levels of ectopic *Shh* were present in all mutant limb buds at these stages. Both heterozygous and homozygous WMSΔ5 mutant embryos showed appreciably more *Shh* and *Ptc* expression at

the ectopic site of the hind limb bud at E11.5 (Fig. 1M, N) than detected in the homozygous Cu embryos, with some also showing ectopic expression in the fore limbs. Thus, the levels of ectopic *Shh* signalling detected reflected the final phenotype with long bone abnormalities arising in those limb buds expressing higher levels of ectopic *Shh* earlier in development. The clustering of the human mutations within a short 5bp sequence causing WMS indicates that this is a single important site for transcription factor binding; while, the deletions confirm that WMS is due to the loss of binding of a repressor that actively represses ectopic expression.

Small Insertions Extend the Limb Phenotypic Spectrum

A second set of mutations arose adjacent to the 5bp WMS site resulting in the insertion of either one or two additional adenosines (called +A and +AA in Fig. 2A). The mutant phenotype generated in the WMS+A mutant heterozygotes was a lengthening of the first digit and sometimes the addition of an extra terminal phalange on digit 1 of the hind limbs (Fig. 2B) with normal fore limbs. *Shh* expression appears normal in +A mutant limb buds at E11.5 (Fig. 2C); however, the phenotype suggests a low level of expression at the ectopic, anterior margin of the limb bud. Insertion of +AA resulted in a more severe phenotype which has not been previously described for ZRS associated mutations. This dinucleotide insertion caused typical PPD in the fore limbs (Fig. 2D) but in the hind limbs extra digits occurred centrally in the digital array (Fig 2E), occasionally in conjunction with long bone anomalies (1 in 7 heterozygotes) (Fig. 2F). The +AA hind limb buds showed an extended pattern of ectopic expression covering from the posterior margin all around the distal edge of the limb bud (Fig.2G). The plasticity of a developmental enhancer in producing morphological changes has been investigated in *Drosophila* (Swanson et al., 2010) and in mouse, it is clear that point mutations in the ZRS give rise to additional preaxial digits

and to homeotic transformations of the thumb to a finger (Anderson et al., 2012). The +AA mutant embryo presents an unusual skeletal configuration in the digital ray indicating that further cryptic plasticity is uncovered by mutational events that disrupt the enhancer's organization.

Large Regions within the ZRS are Dispensable

We next focussed our mutation analysis on two highly conserved regions 3' of the WMS site (Fig. 3A, see Fig. S1 for sequence comparison) to delve into the function of previously identified sites that are putatively important for *Shh* gene regulation. Corresponding gRNAs were designed that overlapped these sites (sequences shown as boxes in Fig.3B and J, with the PAM sites in italics). One region contains the conserved E-box that binds the transcription factor HAND2 (Osterwalder et al., 2014) (Fig. 3A, and blue nucleotides in 3B) crucial to the spatial specific activation of *Shh* in the posterior margin of the limb bud. The second region contains the *Hx* mouse point mutation which lies at position 553 (Fig.3A and red nucleotide in 3J) shown by transgenic analysis to operate as a dominant gain-of-function mutation and to encode important structural features crucial for enhancer activity (Lettice et al., 2014). In addition, the *Hx* site is embedded in a large region of the enhancer that is crucial for the long-range activity of the enhancer.

A series of overlapping deletions targeting the Ebox were identified (Fig. 3B), two of these disrupted the Ebox; EboxΔ3 which was Ebox specific removing the 3 nucleotides from the middle, and EboxΔ17 which deleted the Ebox and surrounding nucleotides. Two other deletions EboxΔ8 and EboxΔ16 removed nucleotides at the 3' side of the Ebox. None of these four deletions showed a phenotype either as heterozygotes or homozygotes (number of homozygotes analysed shown as n=, in Fig. 3B) For example, the largest deletion, EboxΔ17, which disrupts the Ebox and removes surrounding nucleotides showed wildtype

skeletal patterns in both the fore and hind limbs (Fig. 3C and D, respectively). *Shh* expression in the limb buds for all four deletions showed the normal posterior pattern in homozygous embryos (Fig. 3F-I). The EboxΔ3 and EboxΔ17 deletions suggest that removal of a single Ebox site has no detectable effect on *Shh* expression. Possibly, a second conserved Ebox site downstream which has a lower affinity for HAND2 (Osterwalder et al., 2014) may compensate for this loss. The EboxΔ8, EboxΔ16 and EboxΔ17 mutations overlap in a conserved region (Fig. 3B and S1) deleting a total of 24bp. No deletions in this region affected the limb phenotype showing that a substantial region of conserved information can be disrupted.

Using gRNA targeted to the *Hx* mutation, we identified four deletions, 3'Δ42, 3'Δ11, 3'Δ12, and 3'Δ8, all of which are encompassed in 56bp including the Hx site (Fig. 3J) and none of these had an effect on limb phenotype. Similar to the deletions created for the Ebox, these removed highly conserved nucleotide stretches; the 3'Δ8 and Δ12 deletion disrupting a HOXD binding site (Hoxsite 4, orange nucleotides in Fig 3J, and Fig. 5) (see below). The 3'Δ11 and 3'Δ42 remove the *Hx* mutant site and no polydactylous phenotype is detected in the heterozygotes confirming that, unlike the WMS mutations, the *Hx* point change is a gain-of-function mutation (Lettice et al., 2014). The two other deletions, 3'Δ8 and 3'Δ12, do not contain the Hx mutant site but do remove the adjacent highly conserved sequences containing the Hoxsite4 and these do not show a heterozygous phenotype. Homozygous mutants were made for all these deletions and no phenotype was detected (*n* numbers are shown in Fig. 3J). The larger 127 bp deletion (3'Δ127) (Fig. 3A and S1) confirmed this tolerance for loss of conserved sequence. The large 3'Δ127allele, showed no dominant effect on digit number and in the homozygous state there was no influence on the limb phenotype (n=7) (Fig. 3K, L) while both *in situ* hybridization and qRT-PCR showed no

appreciable change in the expression profile or levels (Fig. 3M, N). This deletion showed that a large region of conserved sequence can be deleted from this enhancer. Since the mutational analysis was performed at the endogenous locus, the lack of a phenotype suggests that the loss of the 3'Δ127 sequence is compensated for; thus, indicating that there is encoded redundancy within the enhancer.

Large Deletions Encompassing the WMS Site Incrementally Affect Expression Levels

Three other deletions were generated (FIG. 4A, Fig S1) when making the WMS mutations; a 20bp deletion, WMSΔ20, which included the 5bp site of the WMSΔ5 and two other deletions, WMSΔ48 and WMSΔ110; both of which lost 21bp on the 3' side of the WMS site removing the E-box element but extending to different positions at the 5' end. The WMSΔ20 deletion, unexpectedly showed no observable limb phenotype; neither a dominant phenotype displaying extra toes nor in the WMSΔ20/WMSΔ20 homozygote, a loss of activity phenotype displaying skeletal deficiencies (*n*=5) (Fig. 4B, C). The WMSΔ20 mouse was further crossed to the *Shh* null mutation to make the WMSΔ20/Shh^{null} compound heterozygote to expose any subtle loss of activity but these again, showed no abnormal phenotype (*n*=5). Analysis of *Shh* expression in WMSΔ20 homozygotes showed little observable differences in the expression pattern (Fig. 4H) compared to wildtype (Fig. 4K) and levels of expression measured by qRT-PCR were not affected significantly (Fig. 4L). Thus the deleterious phenotypic effects of the WMSΔ5 mutations were lost in the larger WMSΔ20 deletion.

The two deletions, WMSΔ48 and WMSΔ110 (FIG. 4A), were examined and in the homozygote removal of these sequences resulted in loss of digits. The WMSΔ48 mutation showed loss of up to one digit on each on the forepaws (Fig. 4D), with some elements being retained and soft and hard tissue syndactyly and fusion being observed The hindpaws were

mildly affected, some showing only partial loss of of a single digit (digit 3) (Fig.4E). The WMSΔ110 embryos showed a precise loss of one digit on all four paws (Fig. 4F, G), with the rest of the digits apparently unaffected. Since these phenotypes were seen only in the homozygous state, these were loss of activity mutations resulting in a decrease in enhancer activity. Indeed *Shh* expression was lower but was retained at the posterior margin of the limb bud at E11.5 but by *in situ* hybridisation levels in Δ48 (Fig. 4I) appeared appreciably lower than wildtype (Fig. 4K), with further reductions in the WMSΔ110 (Fig. 4J). Levels of RNA measured by qRT-PCR showed a significant reduction in *Shh* both WMSΔ110 and WMSΔ48 compared to wildtype (Fig. 4L).

Multiple, Conserved HOX Binding Sites Control Expression Levels of *Shh*

Within the ZRS, a highly conserved 6bp element composed of the sequence CATAAA was detected at four positions (boxed in Fig. 5, FigS1). This 6bp sequence is embedded in sites that compare well to the consensus motif established for the *5'Hoxd* genes (motif for HOXD9-11 shown Fig. 5,) and these were numbered Hoxsites 1-4**.** Genetic analysis of the HOX complexes previously demonstrated that the 5' *Hoxd* genes (*Hoxd*10-13) and their counterparts in the *Hoxa* locus regulate *Shh* expression in the limb (Tarchini et al., 2006) and chromatin immunoprecipitation (ChIP) showed that at least two Hox proteins, HOXD10 and 13, directly bind to the ZRS (Capellini et al., 2006). Three of these identified sites (Hoxsites1, 2 and 3) are contained within the Δ110 deletion, the Δ48 contained two sites (Hoxsite2 & 3) and the Δ20 contained only Hoxsite 3 (Fig. 5, FigS1). Hoxsite 4 was deleted in the series that included the *Hx* site (Fig. 3J) discussed above. Each of the 5'*Hoxd* genes was cloned into the vector pT7CFE1-CHis for subsequent expression in the human *in vitro* expression system (1-Step Human Coupled IVT Kit, Thermo Fisher Scientific). The 5' HOXD proteins were synthesised (Fig. S2) and used in an electromobility shift assay (EMSA) to establish binding

to double-stranded oligonucleotides containing one of these four sites (Table S1). The *in vitro* synthesized HOXD9, 10, and 11 proteins showed the highest binding activity with these sites (Fig. 5) while HOXD12 and13 showed lower activity across all oligos (Fig. S2B). The HOXD proteins showed different preferences for these sites; HOXD9 and 11 bound all four sites, with D9 showing a preference for sites 3 and 4 while D11 favoured Hoxsites 1, 2 and 3. HOXD10 bound site 3 but bound weakly to Hoxsites 1, 2 and 4 (Fig. 5). Specificity of binding was shown using competitor oligonucleotides with either wildtype sequence or Hox binding site mutations (Fig 5, Table S1). Nuclear extracts from embryonic (E11.5) limb buds were also used in EMSA (Fig. 5) and was found to bind to all sites and the binding was specific for the putative HOX binding motif.

The role played by the three HOX sites (Hoxsites 1-3) contained in the Δ110 deletion on ZRS activity was assayed in a series of transgenic embryos. Each site was mutated by replacing three bases in the CATAAA element (mutations for each site shown in Fig. 6A) in a construct carrying the mutated, full-length ZRS driving the expression of a *LacZ* reporter gene. Mutations in each of these HOXD binding sites were made individually or in combination and expression was examined at E11.5 in each injected transgenic embryo (the transient G_0 embryo). As a measure of the relative extent of expression in each transgenic embryo, the width of expression as a percentage of limb bud width was plotted to show the trends (individual limbs are represented by dots in Fig. 6P). Mutations in individual Hoxsites had no observable effects on transgenic expression in the limb bud (in Fig. 6, compare B with C, D and E; Fig. 5P); however, mutations in the two sites (Hoxsites 2 and 3) that were contained in the Δ48 deletion or mutations in Hoxsites 1 and 3 showed detectably decreased expression (Fig. 6F, G and P). Mutation of all three sites (Hoxsites 1-3) showed even further decreases (Fig. 6H and P) comparable to the ZRS carrying the Δ110 deletion

(Fig. 6I and P). The accumulative decrease in expression of the endogenous *Shh* in the Δ48 and Δ110 deletions correlates with the progressive loss of the HOXD binding sites Hoxsites 1-3.

Deletions of the WMS Domain Restores the Wildtype Phenotype

The deletion in WMSΔ20 removes the WMS repressor site but also includes the Hoxsite 3. Transgenics carrying either the Cu point mutation (Lettice et al., 2003) (Fig. 6J) or the WMSΔ5 deletion (Fig. 6K) drives reporter gene expression to an elevated level in the posterior margin of the limb bud (Fig.6P) with appreciable ectopic expression. The WMSΔ20 deletion appears to return transgenic expression to wildtype levels (Fig. 6L, P). The loss of the WMS repressor in combination with the mutant Hoxsite 3 binding site may be sufficient to nullify the increased and ectopic expression by the WMS mutations. To examine this possibility, transgenic mice carrying the Cu mutation in the presence of the three basepair replacement (see above) that disrupts Hoxsite 3 was used in the transgenic assay, and showed no detectable upregulation of the reporter at E11.5 and importantly, no ectopic expression (Fig. 6M, P). To show that the lack of ZRS upregulation was due to the independent action of the WMS mutations and loss of Hoxsite binding, transgenics carrying the Cu change and a different Hoxsite mutation (Hoxsite 2) also predominantly showed the wildtype pattern of expression (one out of five G_0 embryos retained ectopic expression) (Fig. 6N, P). Reductions in expression were shown in the presence of the WMS point mutation when two Hoxsites are mutated (Fig. 6O, P). The transgenic expression reflects the WMSΔ20 deletion suggesting that the WMS mutation, which affects the binding of a repressor, when in the presence of mutations that disrupt binding of an activator effectively cancel each other's activities giving rise to wildtype expression levels. The independent action of these opposing activities emphasize the combinatorial nature of elements that operate in the ZRS.

DISCUSSION

The aim of this study was to investigate the composition of a vertebrate enhancer that falls into the highly conserved class of elements (Ovcharenko et al., 2004). These vertebrate enhancers represent a class in which the structural architecture is under selective constraints resulting in apparent structural inflexibility in both the redundancy and the positioning of transcription factor binding motifs. These enhancers which range in size from 100bp to >1kb have the capacity to bind a substantial number of transcription factors arguing that within a single functional element there is also a degree of structural complexity. This structural complexity has enabled the dissection of the ZRS into discrete regulatory activities. The expression pattern of the *Shh* gene in the limb bud is a consolidation of activities that control restriction of expression to the posterior margin, spatial and temporal expression, levels of expression and long-range promoter activation (Summarised in Fig 7A).

Other examples of a complex arrangement of components have been reported including an elegant analysis of the *Drosophila spa* enhancer that showed structural organisation underlies correct developmental gene expression (Swanson et al., 2010); for instance, sequence elements were defined that regulate long range activity and the organization of other elements repress expression in the wrong cell type. Our analysis surveying deletions showed, further, a complex organization that included unexpected redundancy incorporated into the element. The model for enhancer action that we propose here is one that relies on consolidation of discrete, discernible activities acting as a collective. This collective model suggests an integration of these discrete activities and redundant elements in delivering robust spatiotemporal developmental expression.

Hox genes function at the ZRS to regulate levels of expression

We show the homotypic clustering of conserved HOXD binding sites (Hoxsites 1-4) in the ZRS. At least three of these sites (Hoxsites 1-3) are clustered in a 110bp domain of the ZRS and regulate levels of *Shh* expression. The *5'HoxD* genes, which include *Hoxd 9-13*, are fundamental to limb patterning and are expressed in a temporal collinear fashion with the *Hoxd9* gene expressing earliest in the limb bud followed in sequence with *Hoxd13* being expressed latest (Tarchini and Duboule, 2006). A clustering of highly conserved sites that contain the core motif for binding the 5'HOXD proteins operate in an accumulative manner to regulate activity levels of the ZRS enhancer. In an *in vitro* assay, we showed that the early 5' HOXD proteins (HOXD9-11) bind this motif suggesting that these play an initial role in establishing the activity levels of the ZRS. The region of the ZRS that contains three of these HOXD motifs is crucial for activity and deletions show decreasing *Shh* expression corresponding to the number of Hox binding sites lost. In addition, loss of a Hox binding site counterbalances the increased and ectopic expression generated by the loss of the WMS repressor site. Thus, multiple HOXD factors coordinate, through the binding at multiple sites, the expression levels of *Shh*.

Additional HOXD binding sites have been identified near the 5' end of the ZRS which have a preference for binding HOXD13 (Leal and Cohn, 2016) (Fig 7B). Two sets of HOX sites, therefore, regulate gene expression reacting to the temporal changes in the expression of the 5'Hoxd genes. We suggest that the sites we identified play a role in establishing the levels of *Shh* expression in the initial stages of limb development by binding the early expressing 5'HOXD proteins (HOXD10, 11) but adjusts to the changing embryonic environment within the developing limb by also interacting with the later expressed HOXD13 at different sites.

371 This establishes a regulatory loop that operates by positive feedback, reinforcing the
372 expression of the *Shh* gene by the *5'HoxD* genes (Fig 7B). The early 5'HOXD proteins
373 interact with the ZRS at the HOX binding sites examined in this study to establish the levels
374 of Shh expression based on the sum of the sites occupied (Arrow 1, Fig.7B). SHH, in turn, is
375 crucial for the shift in the regulation of *HoxD* gene expression from a set of early acting
376 enhancers to the enhancers at the 5' end of the gene cluster (Arrow 2, Fig. 7B) that regulate
377 the late expressing genes, in particular *Hoxd13* (Zákány et al., 2004). HOXD13 subsequently
378 binds to sites at the 5' end of the ZRS established by Leal and Cohn (2016) (Arrow 3, Fig. 7B).
379 We suggest that a temporal response to *HoxD* genes is important for continued *Shh*
380 expression as the regulatory environment in the limb bud changes over the 2 days that *Shh*
381 is expressed in the mouse limb.

382 In accord, the pythons and boa snakes which have lost the HOXD13 binding sites in
383 the ZRS initially express *Shh* in the rudimentary limb buds, presumably dependent on the
384 early 5'HOXD proteins binding sites that we established; however, *Shh* expression is lost
385 later and limb development is prematurely terminated. This loss of the HOXD13 binding
386 sites in combination with loss of an ETS binding site (Leal and Cohn, 2016; Kvon et al., 2016)
387 is responsible for the loss of limbs in these snakes.

388 Homotypic clustering of binding sites in the ZRS appears to play a number of roles in
389 determining the spatial expression pattern of *Shh* expression in the embryonic limb bud. We
390 previously showed multiple binding sites for the ETS factors, ETS1 and GABPα (Lettice et al.,
391 2012). Multiple occupancy of these sites determine the extent of the boundary of *Shh*
392 expression. Mutations in the human ZRS which generate an extra ETS site results in the
393 extension of this expression boundary and ectopic expression in the limb bud resulting in
394 preaxial polydactyly (Lettice et al., 2012; Laurell et al., 2012). Here, occupancy of multiple

HOXD binding sites regulates the levels of expression, sequential loss of these sites result in a gradual decrease in expression levels. Homotypic clustering of binding sites in the ZRS, therefore, operates to adjust incrementally the expression of the *Shh* gene and is therefore, a fundamental mechanism for fine-tuning the regulatory activity of the enhancer.

ZRS Activity and Congenital Abnormalities

Mutations in the human ZRS cause skeletal abnormalities (Anderson et al., 2012). The point mutations act in a dominant fashion to cause digital abnormalities and presumably, most operate by switching restricted posterior expression to expression at both the posterior margin and an ectopic site at the anterior margin. One set of point mutations generates additional binding sites for ETS1/GABPα transcription factors, acting as dominant gain-of-activity mutations (Lettice et al., 2012). WMS, on the other hand, is highlighted by point mutations in three distinct positions in a single 5bp site. The action of these point mutations, confirmed by the WMSΔ5 deletion data, is consistent with the loss of binding of a repressor and thus, an overall loss of functional activity. Hence, point mutations in ZRS have two modes of action, operating as both gain and loss of activity, but both resulting in dominant genetic effects on the phenotype.

Insertions Reveal Cryptic Phenotypes

The WMS+AA insertional mutation reveals an unusual phenotype showing the latent capacity for phenotypic innovation carried by this enhancer. The potential for appreciable morphological change shows that developmental enhancers may have the capacity for change without undergoing large sequence and structural changes in evolution. Selection against such substantial morphological changes may be one of the evolutionary constraints operating on the ZRS but, in contrast, this also highlights the capacity for appreciable change in vertebrate evolution. These additions reveal the plasticity that is potentially hidden

within an enhancer in controlling the phenotype and highlights mechanisms that may be available for phenotypic change during the evolution of an enhancer.

Evolution of the ZRS

The function of a *cis*-regulator is encoded in its molecular architecture. Overlapping deletions in the ZRS that would predictably disrupt this architecture were made near and encompassing the proposed E-Box binding site and the WMS site that removed a total of 44 basepairs of highly conserved sequence and these do not affect the limb phenotype. Moreover, the large 3'Δ127 mutation removes conserved sequence from the 3'half of the ZRS, which overlaps this 44bp and also displays neither a limb phenotype nor a detectable reduction in expression. The ability to compensate for loss of sequence information suggests that there is encoded redundancy within the enhancer. This seemingly redundant activity may contribute to phenotypic robustness during development. Robustness is deemed important to buffer developmental processes from environmental and genetic perturbations and was proposed as canalization by Waddington (Waddington, 1942). For enhancers, such redundancy is widespread in Drosophila (Cannavo et al., 2016). Secondary or 'shadow enhancers' in Drosophila provide redundant activity for the primary enhancer and analysis of specific examples show these can buffer a developmental process against environmental perturbations (Frankel et al., 2010; Perry et al., 2010). It is clear that the ZRS is able to tolerate losses of highly conserved sequence without affecting phenotype under ideal breeding conditions and defined genetic background. In contrast to shadow enhancers, the robustness apparent in the ZRS is encoded within a single enhancer element, since no compensatory activity is apparent in ZRS deletions. Hence, redundancy is an important characteristic of enhancers whether this is encoded in secondary enhancers or is contained within a single element such as in the ZRS.

443 The evolutionary stability of the ZRS sequence raises a number of questions about
444 the evolvability of this, and perhaps other, highly conserved enhancers. In addition, this
445 stability occurs in light of the major morphological changes that have occurred to the limb
446 during vertebrate evolution. Thus the ZRS displays low sequence variability in a
447 morphologically plastic developmental system. The recurrent role that the ZRS plays in the
448 diverse species so far analysed is to ensure that *Shh* is expressed specifically along the
449 posterior margin of the developing appendage whether it is an embryonic fin (Dahn et al.,
450 2007) or a limb bud. Many of the genes and signalling pathways known to regulate *Shh* in
451 mouse; such as the *HoxD* complex, *Hand2*, *Gli3* and the FGF pathway are implicated in chick
452 and fish suggesting that the gene network responsible for *Shh* activation is also conserved
453 (Gehrke and Shubin, 2016). For vertebrate enhancers, which are found in all vertebrate
454 classes from cold-blooded fishes to warm-blooded mammals, it is unlikely that the apparent
455 robustness is a response to environmental factors since these insults would be different for
456 each species. The genetic network of transcription factors and signalling pathways that
457 converge at the ZRS is complex and we suggest that the regulatory robustness observed for
458 the ZRS buffers against variability and perturbations in this genetic network. This network
459 that converges at the ZRS would, therefore, have evolved early in vertebrates operating
460 relatively unchanged in the appendicular skeleton in all classes of vertebrates. The
461 conserved enhancer architecture is a response to this complex network and would be a
462 constant factor that pervades species evolution during the morphological changes that have
463 occurred during the fin to limb transitions
464
465 **EXPERIMENTAL PROCEDURES**
466 **Production and analysis of CrispR mice**

Guide RNAs were designed using the Optimized CrispR design site (http://crispr.mit.edu/) and the exact guides chosen on the basis of their precise location relative to the desired sites in ZRS (Oligonucleotides selected are listed in Table S1). Oligos were cloned into px330 vector (Addgene) (Cong et al., 2013), and DNA prepared using Qiagen Plasmid Maxi kit (manufacturer's protocol).

Transgenic mice were made by pronuclear injection of plasmid DNA at a concentration of 5ng/µl. All resulting pups were screened phenotypically and had their ZRS sequence amplified by PCR and sequenced. All genotyping was performed by direct sequencing.

Skeletal preparations were stained simultaneously with Alizarin Red and Alcian Blue (Nagy et al., 2009a, b). Whole-mount *in situ* hybridisation was performed as previously described (Hecksher-Sorensen et al., 1998) using probes for *Shh* (Echelard et al., 1993) (a kind gift from Andy McMahon) and *Ptc* (Hayes et al., 1998) (a kind gift from Chris Hayes). qRT-PCR for Shh expression was performed on individual pairs of limb buds as described in Lettice et al., (2014). Expression was normalised within a litter to the wildtype level and statistical significance calculated by Prism using the Kruskal-Wallis test with Dunn's multiple comparisons. Mouse studies were approved by the University of Edinburgh AWERB and carried out under the auspices of the UK Home Office.

EMSAs/IVT proteins

The coding regions of mouse *Hoxd 9-13* were amplified by PCR using KOD polymerase (Merck Millipore). The primers used are listed in Table S1. Products were cloned into the expression vector pT7CFE1-CHis for subsequent expression in the human in vitro expression system (1-Step Human Coupled IVT Kit, Thermo Fisher Scientific) following the manufacturer's instructions. Synthesis of each of the HOXD proteins was verified on a

491 western blot using a rabbit anti-His Tag antibody (#2365, Cell Signalling Technology) (Fig.
492 S2), before the protein was used in an electromobility shift assay (EMSA). The double
493 stranded oligonucleotides were biotin labelled by the manufacturers (Sigma) and assayed to
494 ensure that each was labelled to a similar specific activity. EMSAs were conducted as
495 previously described (Lettice et al., 2012) and used either 2ul of a 1/25 dilution of protein
496 from the IVT reaction or 4ug of limb bud extract (prepared using NE-PER® Nuclear and
497 Cytoplasmic Extraction Reagent Kit, Thermo Scientific). The specificity of binding was
498 confirmed by competition with 100x excess of either unlabelled wild type or mutant (mut)
499 HoxSite oligonucleotides. (Table S1)

500 **Mutant ZRS transgenic constructs**

501 Reporter gene transgenic analysis were made as previously described (Lettice et al.,
502 2012). The mutant ZRS deletions (WMS Δ5, Δ20, Δ110) used were generated by PCR using
503 primers ZRSF and R (Table S1) from the appropriate mutant DNA. The Cu point mutation
504 and MutHoxsite constructs were created using primers in Table S1 and a QuikChange II Site-
505 Directed Mutagenesis Kit (Agilent). For combinations of sites, multiple rounds of
506 mutagenesis were conducted and the correctly mutated ZRS subsequently cloned into fresh
507 lacZ containing vector.

508 **AUTHOR CONTRIBUTIONS**

509 Conceptualization, L.L.; Methodology L.L., Validation L.L., Investigation L.L., P.D., C.D.;
510 Writing – Original Draft R.H.; Writing – Review and editing L.L., R.H.; Visualization L.L.;
511 Supervision L.L., R.H.; Funding Acquisition R.H.

512 **ACKNOWLEDGEMENTS**

513 We would like to thank Richard Mort for help with the statistical analysis and MRC Central
514 Services for providing DNA sequence support. We would also like to thank the staff at the

515 Evans Building for expert technical assistance. Invaluable advice on the manuscript was
516 given by David Fitzpatrick, Nick Hastie and Wendy Bickmore. This work was supported by an
517 MRC core program grant.

518

519

520 **REFERENCES**

521 Anderson, E., Peluso, S., Lettice, L.A., and Hill, R.E. (2012). Human limb abnormalities
522 caused by disruption of hedgehog signaling. Trends Genet 28, 364-373.

523 Cannavo, E., Khoueiry, P., Garfield, D.A., Geeleher, P., Zichner, T., Gustafson, E.H., Ciglar,
524 L., Korbel, J.O., and Furlong, E.E. (2016). Shadow Enhancers Are Pervasive Features of
525 Developmental Regulatory Networks. Curr Biol 26, 38-51.

526 Capellini, T.D., Di Giacomo, G., Salsi, V., Brendolan, A., Ferretti, E., Srivastava, D.,
527 Zappavigna, V., and Selleri, L. (2006) Pbx1/Pbx2 requirement for distal limb patterning is
528 mediated by the hierarchical control of Hox gene spatial distribution and Shh expression
529 Development 133, 2263-73.

530 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W.,
531 Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems.
532 Science 339, 819-823.

533 Dahn, R.D., Davis, M.C., Pappano, W.N., and Shubin, N.H. (2007). Sonic hedgehog
534 function in chondrichthyan fins and the evolution of appendage patterning. Nature 445,
535 311-314.

536 Dow, L.E. (2015). Modeling Disease In Vivo With CRISPR/Cas9. Trends Mol Med 21, 609-
537 621.

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

541 Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic
542 robustness conferred by apparently redundant transcriptional enhancers. Nature 466, 490-
543 493.
544 Gehrke, A.R., and Shubin, N.H. (2016). Cis-regulatory programs in the development and
545 evolution of vertebrate paired appendages. Semin Cell Dev Biol.
546 Hayes, C., Brown, J.M., Lyon, M.F., and Morriss-Kay, G.M. (1998). Sonic hedgehog is not
547 required for polarising activity in the Doublefoot mutant mouse limb bud. Development
548 125, 351-357.
549 Hecksher-Sorensen, J., Hill, R.E., and Lettice, L. (1998). Double labeling for whole-mount
550 in situ hybridization in mouse. Biotechniques 24, 914-916, 918.
551 Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J.,
552 Tissieres, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive Loss of Function
553 in a Limb Enhancer during Snake Evolution. Cell 167, 633-642 e611.
554 Laurell, T., Vandermeer, J.E., Wenger, A.M., Grigelioniene, G., Nordenskjöld, A., Arner,
555 M., Ekblom, A.G., Bejerano, G., Ahituv, N., and Nordgren, A.(2012) A novel 13 base pair
556 insertion in the sonic hedgehog ZRS limb enhancer (ZRS/LMBR1) causes preaxial polydactyly
557 with triphalangeal thumb. Hum Mutat. 33, 1063-6.
558 Leal, F. and Cohn, M.J. (2016). Loss and Re-emergence of Legs in Snakes by Modular
559 Evolution of Sonic hedgehog and HOXD Enhancers Curr Biol. 26, 2966-2973.
560 Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G.,
561 Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the
562 developing limb and fin and is associated with preaxial polydactyly. Hum Mol Genet 12,
563 1725-1735.

564 Lettice, L.A., Hill, A.E., Devenney, P.S., and Hill, R.E. (2008). Point mutations in a distant
565 sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial
566 polydactyly. Hum Mol Genet 17, 978-985.
567 Lettice, L.A., Williamson, I., Devenney, P.S., Kilanowski, F., Dorin, J., and Hill, R.E. (2014).
568 Development of five digits is controlled by a bipartite long-range cis-regulator. Development
569 141, 1715-1725.
570 Lettice, L.A., Williamson, I., Wiltshire, J.H., Peluso, S., Devenney, P.S., Hill, A.E., Essafi, A.,
571 Hagman, J., Mort, R., Grimes, G., et al. (2012). Opposing functions of the ETS factor family
572 define Shh spatial expression in limb buds and underlie polydactyly. Dev Cell 22, 459-467.
573 Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2009a). Alcian blue staining
574 of the mouse fetal cartilaginous skeleton. Cold Spring Harb Protoc 2009, pdb prot5169.
575 Nagy, A., Gertsenstein, M., Vintersten, K., and Behringer, R. (2009b). Alizarin red staining
576 of post-natal bone in mouse. Cold Spring Harb Protoc 2009, pdb prot5171.
577 Norbnop, P., Srichomthong, C., Suphapeetiporn, K., and Shotelersuk, V. (2014). ZRS
578 406A>G mutation in patients with tibial hypoplasia, polydactyly and triphalangeal first
579 fingers. J Hum Genet 59, 467-470.
580 Osterwalder, M., Speziale, D., Shoukry, M., Mohan, R., Ivanek, R., Kohler, M., Beisel, C.,
581 Wen, X., Scales, S.J., Christoffels, V.M., et al. (2014). HAND2 targets define a network of
582 transcriptional regulators that compartmentalize the early limb bud mesenchyme. Dev Cell
583 31, 345-357.
584 Ovcharenko, I., Stubbs, L., and Loots, G.G. (2004). Interpreting mammalian evolution
585 using Fugu genome comparisons. Genomics 84, 890-895.
586 Perry, M.W., Boettiger, A.N., Bothma, J.P., and Levine, M. (2010). Shadow enhancers
587 foster robustness of Drosophila gastrulation. Curr Biol 20, 1562-1567.

588 Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M., and Shiroishi, T. (2005). Elimination of a
589 long-range cis-regulatory module causes complete loss of limb-specific Shh expression and
590 truncation of the mouse limb. Development 132, 797-803.
591 Swanson, C.I., Evans, N.C., and Barolo, S. (2010). Structural rules and complex regulatory
592 circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. Dev Cell 18,
593 359-370.
594 Tarchini, B., Duboule, D., and Kmita, M. (2006) Regulatory constraints in the evolution of
595 the tetrapod limb anterior-posterior polarity. Nature 443, 985-8.
596 Tarchini, B., and Duboule, D. (2006) Control of Hoxd genes' collinearity during early limb
597 development Dev Cell 10, 93–103
598 VanderMeer, J.E., Lozano, R., Sun, M., Xue, Y., Daentl, D., Jabs, E.W., Wilcox, W.R., and
599 Ahituv, N. (2014). A novel ZRS mutation leads to preaxial polydactyly type 2 in a
600 heterozygous form and Werner mesomelic syndrome in a homozygous form. Hum Mutat
601 35, 945-948.
602 Visel, A., Rubin, E.M., and Pennacchio, L.A. (2009). Genomic views of distant-acting
603 enhancers. Nature 461, 199-205.
604 Waddington, C.H. (1942). Canalization of development and the inheritance of acquired
605 characters. Nature 150, 563-565.
606 Wieczorek, D., Pawlik, B., Li, Y., Akarsu, N.A., Caliebe, A., May, K.J., Schweiger, B., Vargas,
607 F.R., Balci, S., Gillessen-Kaesbach, G., et al. (2010). A specific mutation in the distant sonic
608 hedgehog (SHH) cis-regulator (ZRS) causes Werner mesomelic syndrome (WMS) while
609 complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD)
610 with or without triphalangeal thumb. Hum Mutat 31, 81-89.

611 Zákány, J., Kmita, M., and Duboule, D. (2004). A dual role for Hox genes in limb anterior-
612 posterior asymmetry. Science 304, 1669-72.

613

614

615 **Figure Legends**

616 **Figure 1.** Mutational analysis of the WMS site in the ZRS. The position of the three
617 sites within the ZRS that were targeted for mutation analysis are depicted in (A) and the
618 WMS site is boxed. The conservation of the region containing the Werner mesomelic
619 syndrome site is shown. The 3 nucleotides (green) mutated in WMS are shown in a green
620 box and labelled WMSΔ5, the Cu mutation is highlighted by the red box and the position of
621 the gRNA is contained in the black box (the PAM site is underlined and in italics). The
622 position of the +A and +AA insertion is also shown. The wildtype and mutant allele
623 sequences are shown at the top of each panel. The wildtype hind limb (B) and expression
624 patterns of *Shh* (G) and *Ptc* (H) at E11.5 hind limb buds are shown for comparison. The
625 hindlimbs of the Cuban mutation (a G > A point change) shows an extra anterior digit in the
626 heterozygote (C) and a polydactylous hindlimb and the hypoplastic tibia in the homozygote
627 (D). *Shh* expression in the Cu homozygote at E11.5 (I) and E12.5 (J) and *Ptc* in the
628 heterozygote at E11.5 (K) and E12.5 (L) are shown. Ectopic expression is highlighted by the
629 black arrows. The heterozygous WMSΔ5 deletion mutants are shown in (E, F, M and N).
630 The hindlimb shows the absence of the tibia and polydactyly (E) and unlike the Cu mutation,
631 polydactyly on the forelimb (F). Strong ectopic expression of *Shh* (M) and *Ptc* (N) are
632 observed (highlighted with arrows). Scale bars = 500μm in (B) and (C); 1mm in (D), (E) and
633 (F); 100μm in (G) to (N).

634

635 **Figure 2.** Insertion mutations disrupt limb development to generate an unusual
636 skeletal phenotype. The position of the adenosine insertions are shown (A) adjacent to the
637 WMS site to create the +A and the +AA mutations. The skeletal features of the +A hindlimb
638 in (B) show a triphalangeal digit 1 (arrowhead). (C) shows expression of *Shh* at E11.5 in the

639 hindlimb of +A mutant. Forelimb of the +AA mutant in (D) shows fusion and duplication of
640 internal digits (asterisks) while the hindlimb in (E) shows bifurcation at the tip of the extra
641 preaxial digit (arrowheads) and the centrally located extra digit (arrow). (F) shows the
642 bending of the hindlimb caused by a shortening of the tibia. The E11.5 WMS+AA hindlimb
643 bud in (G) shows expression of *Shh* along the entire distal edge (black arrowheads). Scale
644 bars = 500μm in (B), (D) and (E); 100μm in (C) and (G); 1mm in (F).
645
646 **Figure 3.** Mutational analysis of the Ebox and the *Hx* sites in the ZRS. The ZRS
647 (yellow rectangle) is depicted in (A) and the relative locations of the WMS site, the Ebox and
648 the Hx mutation are indicated. Boxes highlight the relative positions of the sequences
649 shown in (Ebox, B) and (Hx, J) respectively. Linking these two regions, the position of the
650 3'Δ127 deletion is also shown. The gRNA sequences are boxed in the wildtype sequences in
651 (B and J) with the PAM site (in italics). In (B) the EBox (highlighted in blue font) and the
652 deleted nucleotides for each mutation are shown and the numbers of the homozygous
653 animals analysed are indicated below each mutation (as *n*=). Representative fore limb (C)
654 and hind limb (D) from an EboxΔ17 homozygote demonstrate no detectable deviation from
655 wildtype. The *Shh* expression in hind limbs at E11.5 for the wildtype (E), and EboxΔ3 (F),
656 EboxΔ17 (G), EboxΔ16 (H), EboxΔ8 (I) homozygotes are depicted showing a normal pattern
657 of expression. The mutant sequence affected by the 3' deletions near Hx are shown in (J).
658 The wildtype sequence with the position of the Hx mutation (the red base and box) is
659 indicated. The position of Hoxsite 4 is highlighted in orange. The sequences of all the
660 deletions are shown below and the numbers of the homozygous animals analysed are
661 indicated below each mutation (as *n*=). The apparent unaffected fore limb (J) and hind limb
662 (K) of the large 3'Δ127deletion are shown and The levels of expression of *Shh* at E11.5 hind

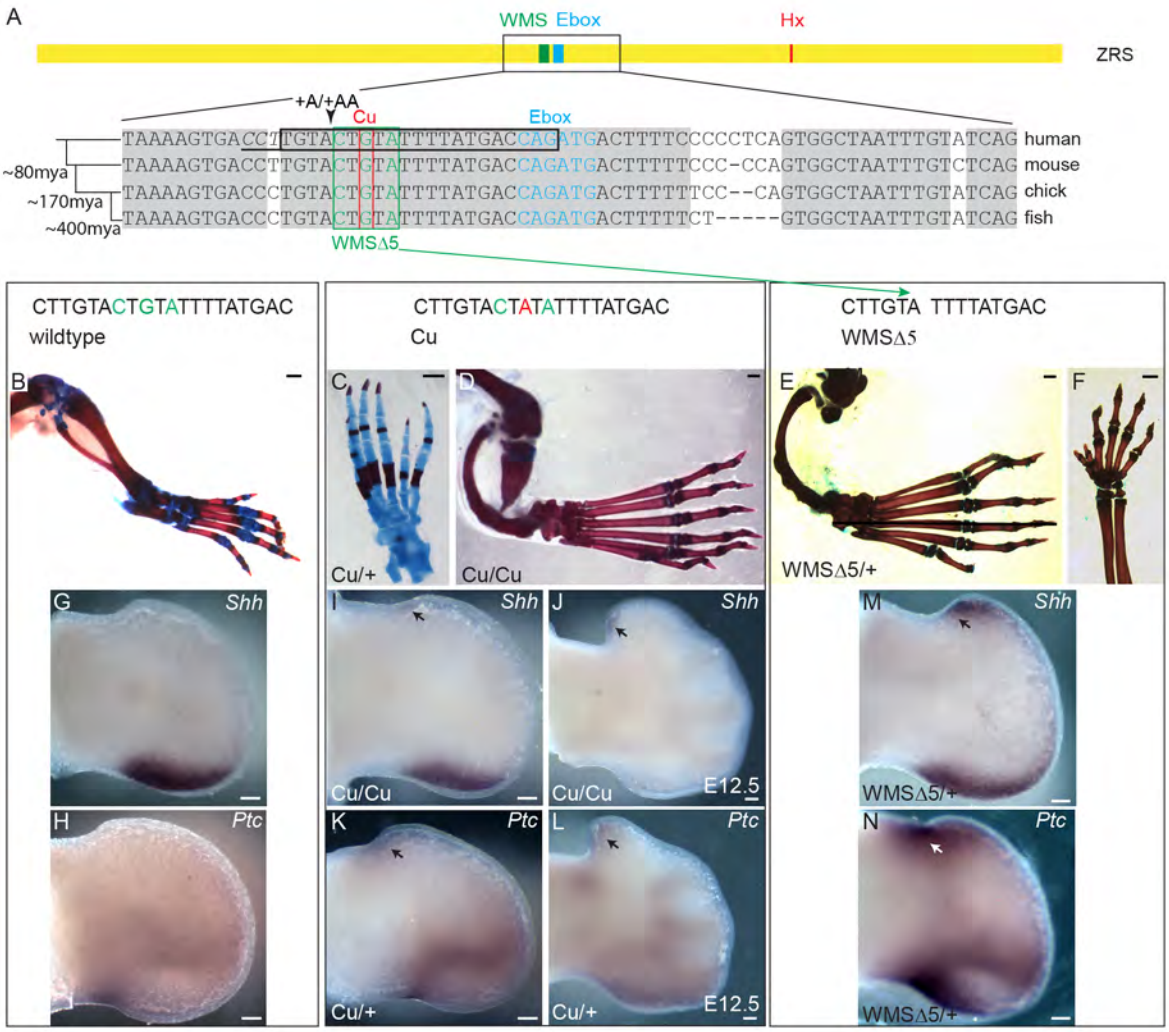
limb buds shown by in situ hybridization (M) and by quantification by qRT-PCR (N) in 3'Δ127 homozygous embryos. Scale bars = 500μm in (C), (D), (K) and (L); 100μm in (C) – (I) and (M).

Figure 4. Deletions near the WMS site reduce the levels of *Shh* expression. (A) The three deletions WMSΔ20, WMSΔ48 and WMSΔ110 are shown relative to the WMS (green line), Ebox sites (blue line) and Hx mutation (red) within the ZRS (yellow rectangle). The homozygous WMSΔ20 mutants are shown in (B, C and H). No limb abnormalities are detected in the forelimb (B) or the hind limb (C). WMSΔ48 limbs are shown in (D) and (E). The forelimb (D) shows loss of a digit and two terminal phalanges on the adjacent digit (arrowheads) and the hindlimb in (G) shows partial loss of digit 3 (arrow). Loss of the middle digit in forelimb (F) and hind limb (G) are shown in the WMSΔ110 deletion. H-K shows *Shh* expression at E11.5 in the hindlimbs of WMSΔ20 (H), WMSΔ48 (I), WMSΔ110 (J) and wildtype (K) embryos. (L) shows the outcome of the quantification by qRT-PCR of *Shh* expression in E11.5 limb buds from wildtype and WMSΔ20, WMSΔ48 and WMSΔ110 homozygous embryos. WMSΔ48 and WMS110 expression levels are significantly ($p < 0.001$) lower than wildtype (Kruskal-Wallis test with Dunn's multiple comparisons). Scale bars = 1mm in (B) and (C); 500μm in (D)- (G); 100μm in (H) - (K).

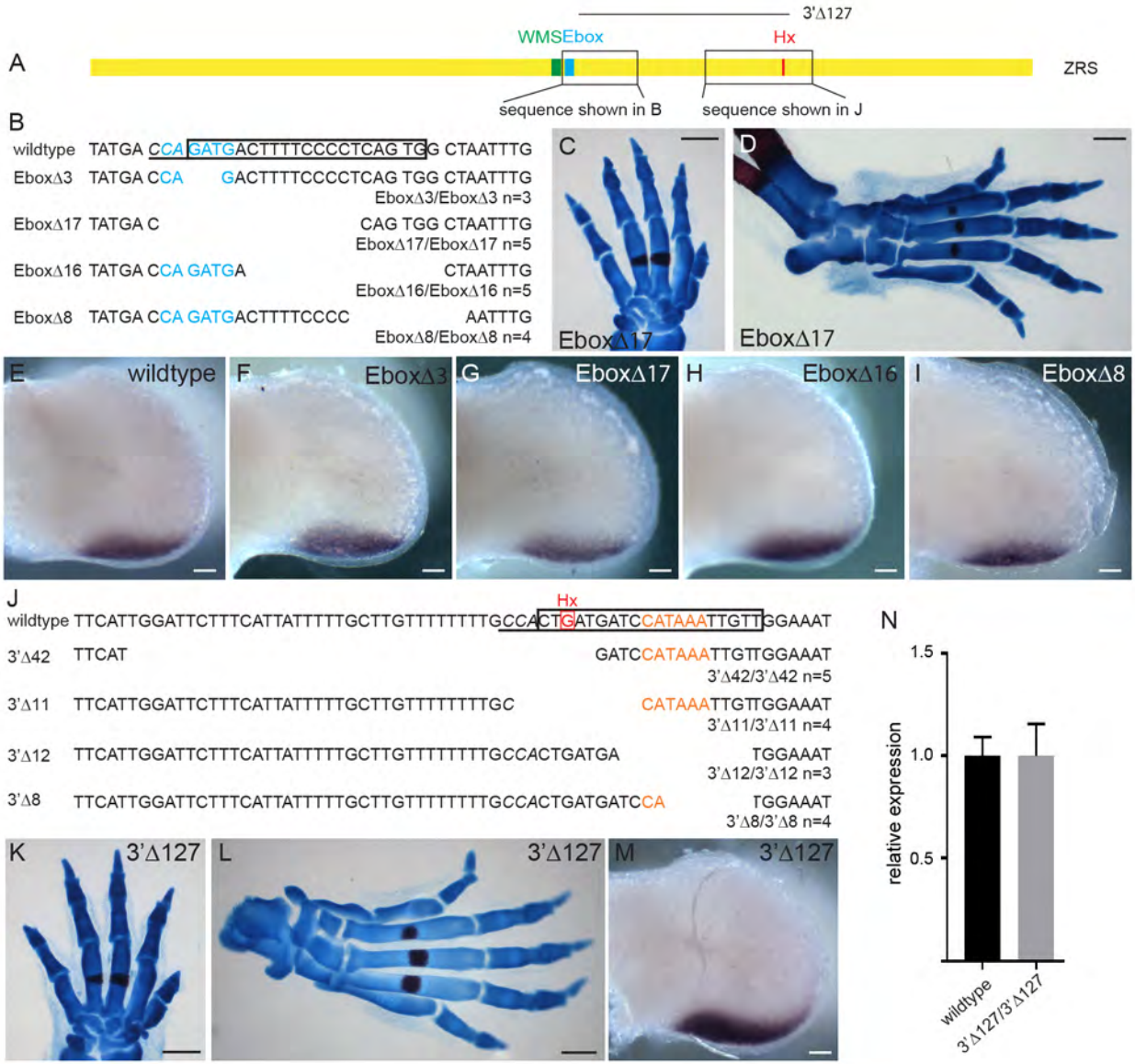
Figure 5. HoxD binding to conserved motifs in the ZRS. The top line shows the genomic sequence round the 4 Hox binding sites (designated Hoxsites1-4), with the position of the WMS region indicated. Also the relative positions of the WMSΔ20, WMSΔ48 and WMSΔ110 deletions are indicated. Below, the sequence of the 4 Hoxsites are shown in the same orientation. The consensus binding sites for each of the proteins HOXD9, D10 and D11 are shown as position weight matrices under their gene names. For each triplet of EMSAs,

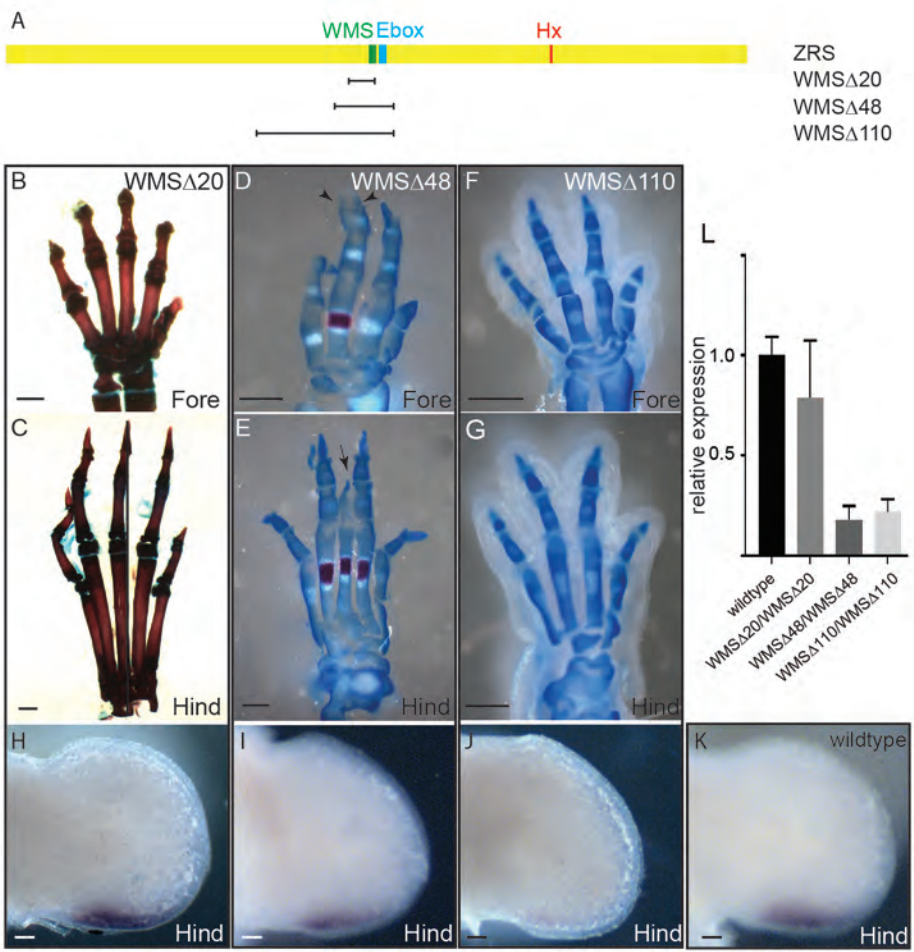
687 the lanes are shown as binding to a labelled Hoxsite oligonucleotide with no competition,
688 with excess of the wildtype oligo as competitor and with the mutated Hoxsite oligo in
689 competition to show specificity of binding to the Hoxsite. The specific binding is indicated by
690 the arrowheads. In the case of the E11.5 limb bud extract binding to Hoxsite2 a higher
691 mobility shift is observed, indicated by the asterisk. The non-specific band (arrow) is marked
692 as a comparison with Fig S2.
693
694 **Figure 6**. Transgenic analysis of embryos carrying mutant ZRS sequences. (A) shows
695 the sequences of the wildtype Hoxsites 1-3 and the mutated sequences (designated
696 MutHoxsite) that were used in the transgenic constructs. (B-E) Limb buds from transgenic
697 embryos (E11.5) carrying the following ZRS sequences driving LacZ expression: (B) the
698 wildtype ZRS sequences while C-E show the no effect on expression of mutating the
699 Hoxsites singly, MutHoxsite 1 (C), MutHoxsite 2 (D) and MutHoxsite 3 (E). Mutating
700 combinations of sites results in lower LacZ expression; MutHoxsite2+3 shown in (F),
701 MutHoxsite1+3 in (G) and MutHoxsite1,2 +3 in (H). The low level of expression in
702 MutHoxsite1, 2 +3 is reproduced in the WMSΔ110 construct (I). Addition of the Cu point
703 mutation (J) or deletion of the WMSΔ5 (K) results in distal and ectopic anterior expression;
704 whereas, deletion of WMSΔ20 (L) returns expression to wildtype levels. M-O show the Cu
705 mutation in combination with mutant Hox sites; MutHoxsite3+Cu (M), MutHoxsite2+Cu (N)
706 and MutHoxsite2+3+Cu (0). (P) Graphical representation of the LacZ expression patterns
707 resulting from mutations within the ZRS. The width of the expression domain was divided by
708 the width of the limb and expressed as a percentage. One spot represents the extent of
709 reporter gene (*LacZ*) expression for each individual limb from a set of transient transgenic
710 embryos. Data was subjected to a one-way ANOVA and TukeyHSD test and those that differ

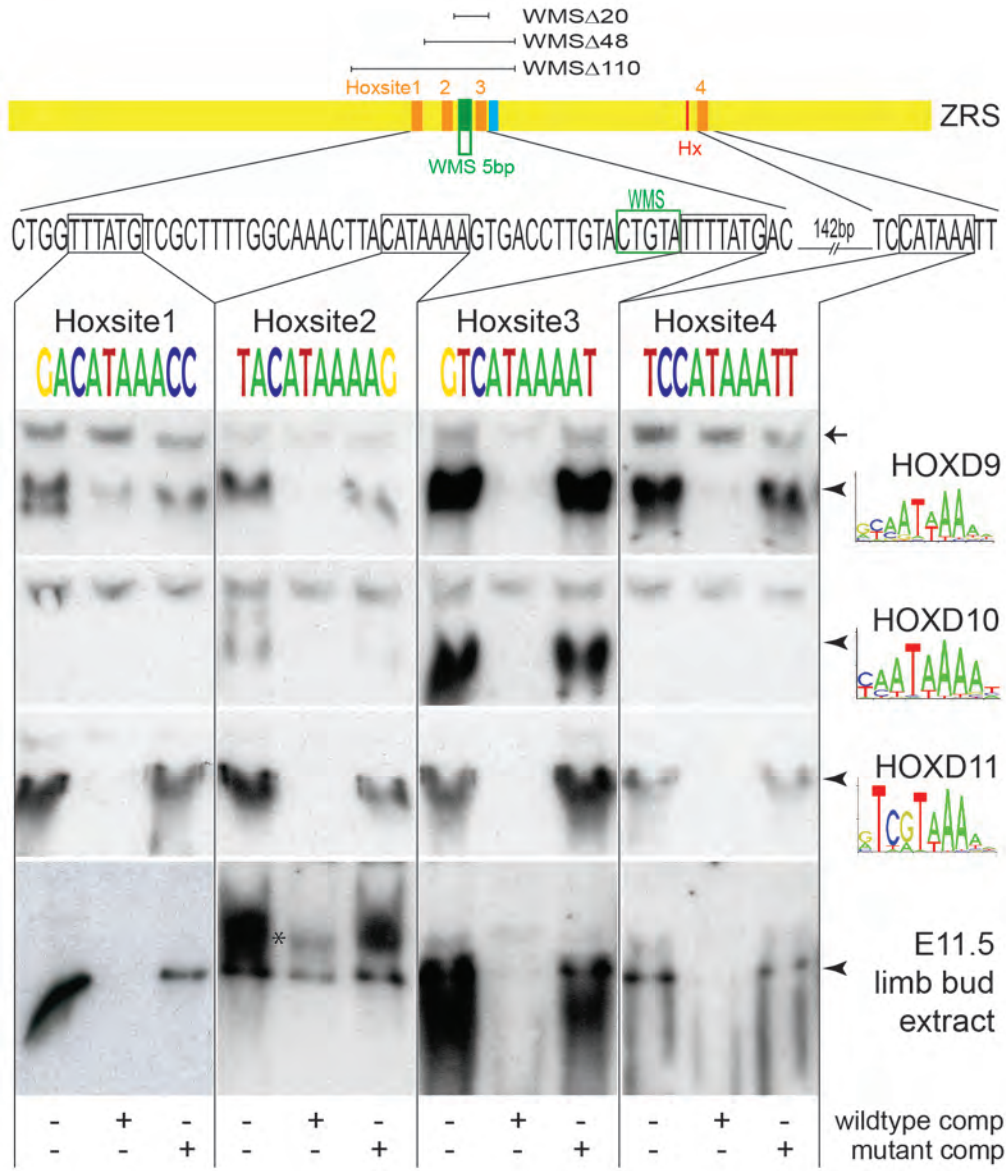
711 significantly from wildtype are indicted. (* P ≤0.05, *** P ≤ 0.001, ****P≤0.0001) Scale bars
712 = 100μm.
713
714 **Figure7**. A representation of different functional regions and sites established for the
715 ZRS is depicted in (A). The ZRS is represented by the yellow rectangle and the position of
716 the WMS 5bp site (green), the Ebox (blue), the Hx mutations (red) and Hoxsites (orange) are
717 indicated in the ZRS. The positions of the 5 ETS sites that control the position of the
718 expression boundary are represented by the red ovals. The region that contributes to
719 regulating levels is in the blue box and the deletions that revealed this activity are shown.
720 The region that mediates long range activity is shown in the black box. The large region of
721 this domain that is redundant is shown by the grey shading**.** The two systems that control
722 posterior restriction are shown below the ZRS rectangle indicating the position of the two
723 ETV binding sites and the position of the WMS 5bp site. (B) summarises the positive
724 feedback loop between the 5' *HoxD* genes and *Shh* to reinforce expression of *Shh.* ZRS
725 (yellow box) and its position relative to *Shh* is shown on the left hand side, while a schematic
726 of the HoxD complex including the two flanking regulatory domains (the early enhancer and
727 the late enhancer) is depicted by green boxes on the right hand side. Early expressing
728 5'HOXD proteins bind Hoxsite 1-3 within the ZRS to establish the levels of *Shh* expression in
729 the initial stages of limb development (arrow 1). The levels of Shh expression is dependent
730 on the number of Hox sites occupied. SHH, in turn, is crucial for the shift in HoxD gene
731 expression to the later genes, in particular *Hoxd13* (arrow 2). HOXD13 subsequently binds
732 to sites at the 5' end of the ZRS established by Leal and Cohn (2016) (arrow 3) to maintain
733 *Shh* expression later in limb development.
734

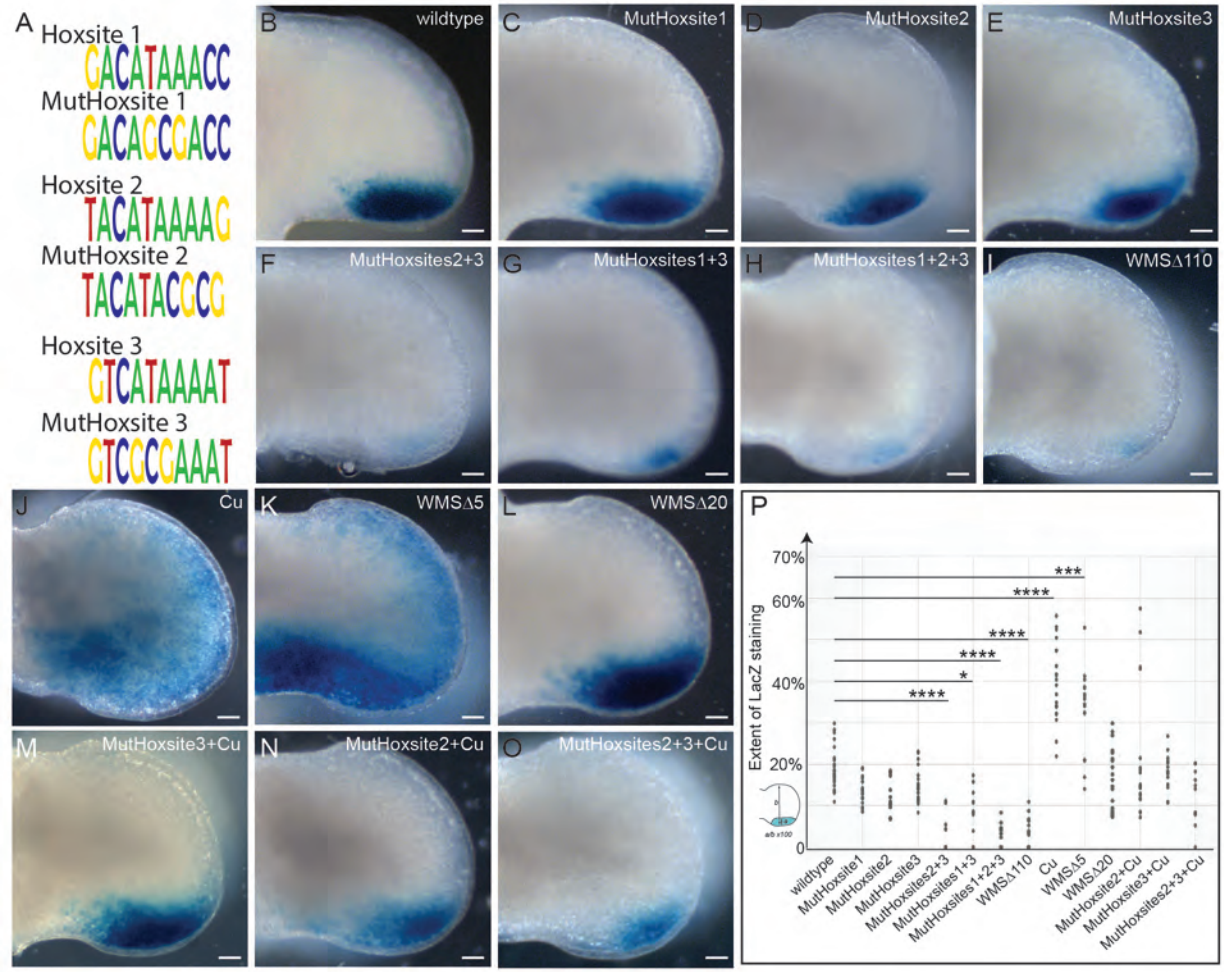

A
WMS Ebox
Hx
ZRS
+A/+AA
Cu
Ebox
TAAAAGTGACCTTGTACTGTATTTTATGACCAGATGACTTTTCCCCCTCAGTGGCTAATTTGTATCAG human
TAAAAGTGACCTTGTACTGTATTTTATGACCAGATGACTTTTTCCC-CCAGTGGCTAATTTGTCTCAG mouse
TAAAAGTGACCCTGTACTGTATTTTATGACCAGATGACTTTTTTCC--CAGTGGCTAATTTGTATCAG chick
TAAAAGTGACCCTGTACTGTATTTTATGACCAGATGACTTTTTCT-----GTGGCTAATTTGTATCAG fish
~80mya
~170mya
~400mya
WMSΔ5
CTTGTACTGTATTTTATGAC
wildtype
CTTGTACTATATTTTATGAC
Cu
CTTGTA TTTTATGAC
WMSΔ5
B
C
D
E
F
Cu/+
Cu/Cu
WMSΔ5/+
G
Shh
H
Ptc
I
J
K
L
M
N
Cu/Cu
E12.5
Cu/+
WMSΔ5/+

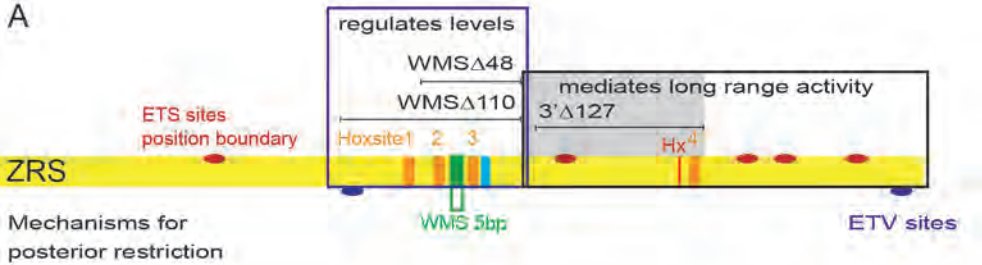

A
+A/+AA
A CCT TGTACTGTATTTTATGACCAGATGA wildtype
ACCTTGTAACTGTATTTTATGACCAGATGA
+A
ACCTTGTAAACTGTATTTTATGACCAGATGA
+AA
B
+A/+
C
+A/+
D
+AA/+
E
+AA/+
F
+AA/+
G
+AA/+


A
3'Δ127
WMSEbox
Hx
ZRS
sequence shown in B
sequence shown in J
B
wildtype TATGA CCA GATGACTTTTCCCCTCAG TGG CTAATTTG
EboxΔ3 TATGA CCA GACTTTTCCCCTCAG TGG CTAATTTG
EboxΔ3/EboxΔ3 n=3
EboxΔ17 TATGA C CAG TGG CTAATTTG
EboxΔ17/EboxΔ17 n=5
EboxΔ16 TATGA CCA GATGA CTAATTTG
EboxΔ16/EboxΔ16 n=5
EboxΔ8 TATGA CCA GATGACTTTTCCCC AATTTG
EboxΔ8/EboxΔ8 n=4
C
EboxΔ17
D
EboxΔ17
E wildtype
F EboxΔ3
G EboxΔ17
H EboxΔ16
I EboxΔ8
J
Hx
wildtype TTCATTGGATTCTTTCATTATTTTTGCTTGTTTTTTTTTGCCACTGATGATCCATAAATTGTTTGGAAAT
3'Δ42 TTCAT GATCCATAAATTGTTTGGAAAT
3'Δ42/3'Δ42 n=5
3'Δ11 TTCATTGGATTCTTTCATTATTTTTGCTTGTTTTTTTTTGC CATAAATTGTTTGGAAAT
3'Δ11/3'Δ11 n=4
3'Δ12 TTCATTGGATTCTTTCATTATTTTTGCTTGTTTTTTTTGCCACTGATGA TGGAAAT
3'Δ12/3'Δ12 n=3
3'Δ8 TTCATTGGATTCTTTCATTATTTTTGCTTGTTTTTTTTGCCACTGATGATCCA TGGAAAT
3'Δ8/3'Δ8 n=4
N
relative expression
1.5
1.0
0.5
wildtype
3'Δ127/3'Δ127
K 3'Δ127
L 3'Δ127
M 3'Δ127


A
WMS Ebox
Hx
ZRS
WMSΔ20
WMSΔ48
WMSΔ110
B
WMSΔ20
Fore
C
Hind
D
WMSΔ48
Fore
E
Hind
F
WMSΔ110
Fore
G
Hind
L
relative expression
1.0
0.5
wildtype
WMSΔ20/WMSΔ20
WMSΔ48/WMSΔ48
WMSΔ110/WMSΔ110
H
Hind
I
Hind
J
Hind
K
wildtype
Hind


WMSΔ20
WMSΔ48
WMSΔ110
Hoxsite1
2
3
4
ZRS
Hx
WMS 5bp
WMS
CTGGTTTATGTCGCTTTTGGCAAACTTACATAAAAGTGACCTTGTACTGTATTTTATGAC
142bp
TCCATAAATT
Hoxsite1
Hoxsite2
Hoxsite3
Hoxsite4
GACATAAACC
TACATAAAAG
GTCATAAAAT
TCCATAAATT
HOXD9
HOXD10
HOXD11
E11.5
limb bud
extract
*
wildtype comp
mutant comp


A
Hoxsite 1
GACATAAACC
MutHoxsite 1
GACACCGACC
Hoxsite 2
TACATAAAAG
MutHoxsite 2
TACATACGCG
Hoxsite 3
GTCATAAAAT
MutHoxsite 3
GTCCCCAAAT
B wildtype
C MutHoxsite1
D MutHoxsite2
E MutHoxsite3
F MutHoxsites2+3
G MutHoxsites1+3
H MutHoxsites1+2+3
I WMSΔ110
J Cu
K WMSΔ5
L WMSΔ20
M MutHoxsite3+Cu
N MutHoxsite2+Cu
O MutHoxsites2+3+Cu
P
Extent of LacZ staining
70%
60%
40%
20%
0
****
*
****
****
****
***
wildtype
MutHoxsite1
MutHoxsite2
MutHoxsite3
MutHoxsites2+3
MutHoxsites1+3
MutHoxsites1+2+3
WMSΔ110
Cu
WMSΔ5
WMSΔ20
MutHoxsite2+Cu
MutHoxsite3+Cu
MutHoxsites2+3+Cu


A
regulates levels
WMSΔ48
WMSΔ110
mediates long range activity
3'Δ127
ETS sites
position boundary
Hoxsite1 2 3
Hx4
ZRS
WMS 5bp
ETV sites
Mechanisms for
posterior restriction

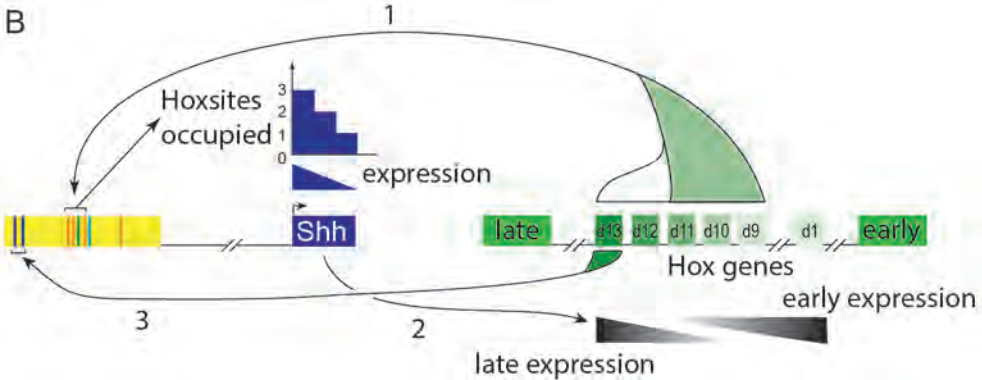

B
1
Hoxsites
occupied
3
2
1
0
expression
Shh
late
d13 d12 d11 d10 d9
d1
early
Hox genes
early expression
3
2
late expression

Table S1- Relates to Figures 5 and 6
Oligonucleotides used.

Px330 oligos	
WMS guideF	CACCGTGGTCATAAAATACAGTACA
WMS guideR	AAACTGTACTGTATTTTATGTCCAC
EBox guideF	CACCGCACTGAGGGGAAAAGTCATC
EBox guideF	AAACGATGACTTTTCCCCTCAGTGC
3' guideF	CACCGAACAATTTATGGATCATCAG
3' guideR	AAACCTGATGATCCATAAATTGTTC
Hox gene IVT constructs	
HoxD9F	CATGAT**CATATG**TCGTCCAGTGGCACCC
HoxD9R	CATGAT**CTCGAG**GTCTCCTTTAGGGCACTTCTC
HoxD10F	CATGAT**CATATG**TCCTTTCCCAACAGCTCTC
HoxD10R	CATGAT**CTCGAG**AGAAAAGGTGAGGTTGGCGGTC
HoxD11F	CATGAT**CATATG**AACGACTTTGACGAGTGCG
HoxD11R	CATGAT**CTCGAG**AAATAAGGGGTTTCCAGTGAAATATTG
HoxD12F	CATGAT**CATATG**TGTGAGCGCAGTCTCTAC
HoxD12R	CATGAT**CTCGAG**ATAGAGGGCCAGTGCTTGCTC
HoxD13F	CATGAT**CATATG**AGCCGCTCGGGACTTGG
HoxD13R	CATGAT**CTCGAG**GGAGACAGTGTCTTTGAGCTTG
Hoxsites EMSA oligos	
Hoxsite 1F	TTGTCCTGG**TTTATG**TCGCTTTTG
Hoxsite 1R	CAAAAGCGA**CATAAA**CCAGGACAA
MutHoxsite 1F	TTGTCCTGG**TTgcgG**TCGCTTTTG
MutHoxsite 1R	CAAAAGCGA**CcgcAA**CCAGGACAA
Hoxsite 2F	CAAACTTA**CATAAAA**GTGACCTTGT
Hoxsite 2R	ACAAGGTCAC**TTTTATG**TAAGTTTG
MutHoxsite 2F	CAAACTTA**CATAcgc**GTGACCTTGT
MutHoxsite 2R	ACAAGGTCAC**gcgTATG**TAAGTTTG
Hoxsite 3F	TGTACTGTAT**TTTATG**ACCAGATGACT
Hoxsite 3R	AGTCATCTGGT**CATAAAA**TACAGTACA
MutHoxsite 3F	TGTACTGTA**TTTcgcG**ACCAGATGACT
MutHoxsite 3R	AGTCATCTGGT**CgcgAAA**TACAGTACA
Hoxsite 4F	CTGATGATC**CATAAA**TTGTTGGAA
Hoxsite 4R	TTCCAACAA**TTTATG**GATCATCAG
MutHoxsite 4F	CTGATGATC**CcgcAA**TTGTTGGAA
MutHoxsite 4R	TTCCAACAA**TTgcgG**GATCATCAG
LacZ Transgenic constructs	
ZRS F	GATCAT**AAGCTT**TACTTTAAGCCATCTTTG
ZRS R	GATCAT**AAGCTT**CACATAGAACACTTAGTGAG
Mutate ZRS oligos	
Cu F	GACCTTGTACTaTA**TTTTATG**ACCAGATGACTTTTCCCTC
Cu R	GAGGGAAAGTCATCTGGT**CATAAAA**TAtAGTACAAGGTC
MutHoxsite 1F	CAGTTTGAGATTGTCCTGG**TcgcTG**TCGCTTTTGGCAAAC
MutHoxsite 1R	GTTTGCCAAAAGCGA**CAgcgA**CCAGGACAATCTCAAACTG
MutHoxsite 2F	GTCGCTTTTGGCAAACTTA**CATAcgc**GTGACCTTGTACTG
MutHoxsite 2R	CAGTACAAGGTCAC**gcgTATG**TAAGTTTGCCAAAGCGAC
MutHoxsite 3F	GACCTTGTACTGTA**TTTcgcG**ACCAGATGACTTTTCCCTC
MutHoxsite 3R	GAGGGAAAGTCATCTGGT**CgcgAAA**TACAGTACAAGGTC
MutHoxsite 3F+Cu	GACCTTGTACTaTA**TTTcgcG**ACCAGATGACTTTTCCCTC
MutHoxsite 3R+Cu	GAGGGAAAGTCATCTGGT**CgcgAAA**TAtAGTACAAGGTC

Figure S1-relates to Figure 1

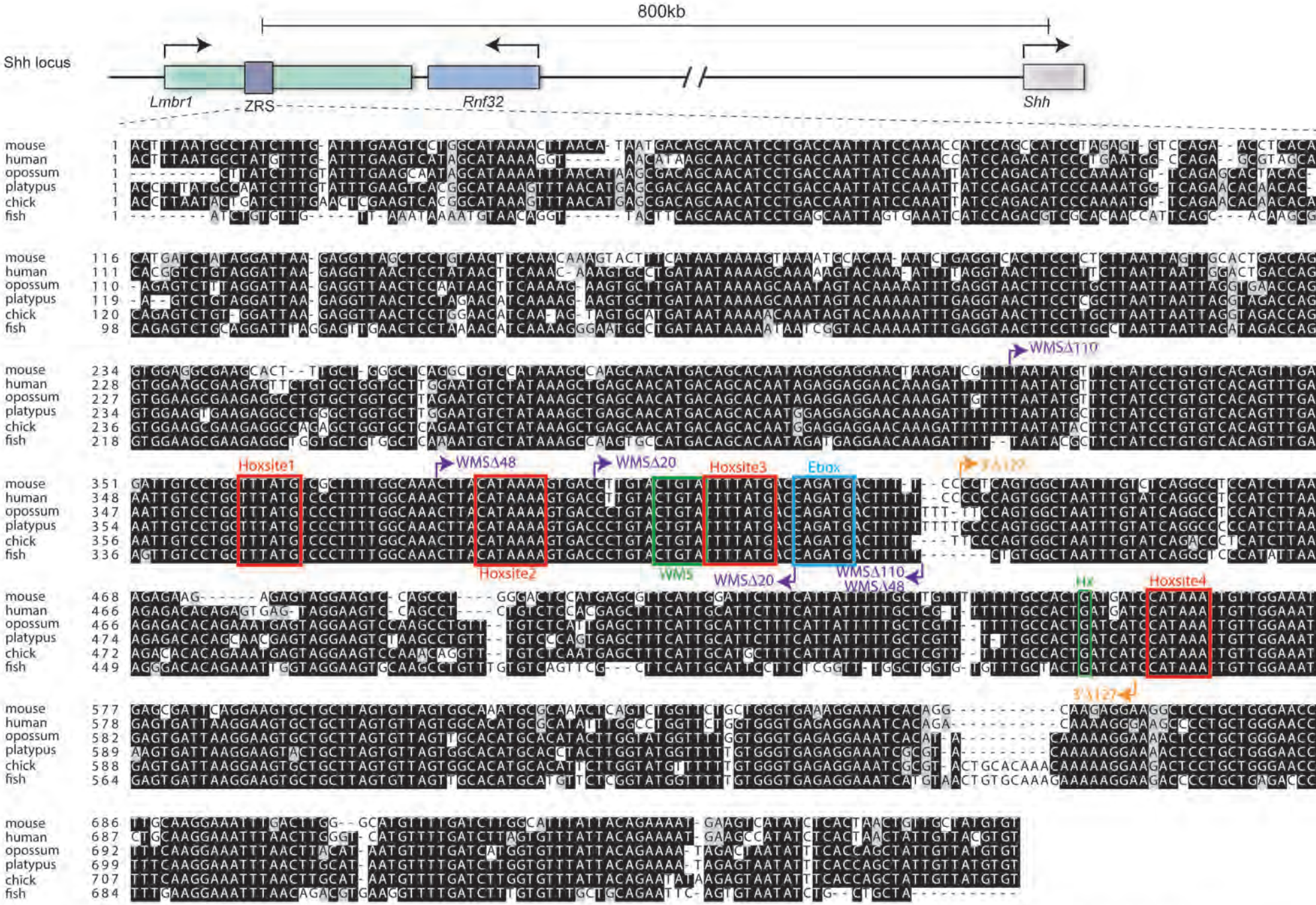
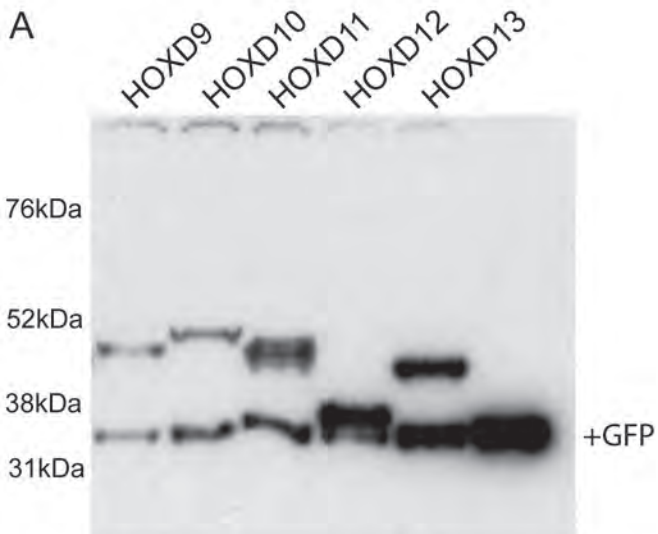
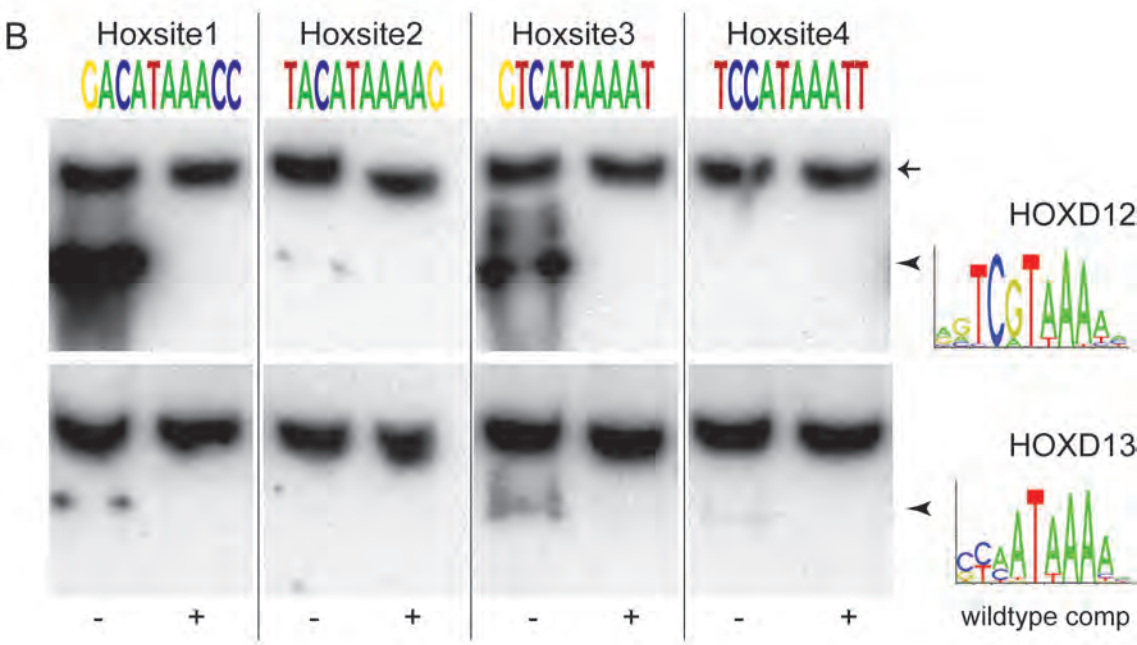


Figure S1. Top line shows a graphical representation of the Lmbr1 to Shh genomic region with the position of ZRS indicated by the grey box. The direction of transcription of the genes is indicated by the arrows. Underneath is the sequence line-up comparing the ZRS in diverse vertebrate species (listed to the left of the sequence) including three different vertebrate classes (mammal, birds and fish) representing >400Myrs of evolution. The sequences for the WMS site and Hx mutation (green) and the Ebox (blue) are boxed. Also boxed in red are the positions of the 4 Hoxsites. The start and end positions of all the large deletions (WMSΔ110, WMSΔ48, WMSΔ20 and 3'Δ127) are indicated.

Figure S2 - relates to Figure 5
Western blots of HOXD proteins and analysis of HOXD12 and D13 binding.





(A) shows a western blot of protein from the IVT reactions probed with an anti HisTag antibody. In addition to the Hox gene containing vector, each reaction included the GFP control vector and a GFP band can be seen in all lanes.
(B) shows the sequence of the Hoxsites1 -4 in the same orientation. The consensus binding sites for the proteins HoxD12 and D13 are shown as position weight matrices under their gene names. For each doublet of EMSAs, the lanes are shown as binding to a labelled Hoxsite oligonucleotide with no competition and with excess of the wildtype oligo as competitor. The specific binding is indicated by the arrowheads, while the non-specific band (arrow) indicates the increased length of exposure time necessary, compare with the equivalent band in Figure 4