ARTICLE

doi:10.1038/nature14010

RNA regulons in *Hox* 5′ UTRs confer ribosome specificity to gene regulation

Shifeng Xue[1,2,3], Siqi Tian[4], Kotaro Fujii[1,2], Wipapat Kladwang[4], Rhiju Das[4,5] & Maria Barna[1,2]

Emerging evidence suggests that the ribosome has a regulatory function in directing how the genome is translated in time and space. However, how this regulation is encoded in the messenger RNA sequence remains largely unknown. Here we uncover unique RNA regulons embedded in homeobox (*Hox*) 5′ untranslated regions (UTRs) that confer ribosome–mediated control of gene expression. These structured RNA elements, resembling viral internal ribosome entry sites (IRESs), are found in subsets of *Hox* mRNAs. They facilitate ribosome recruitment and require the ribosomal protein RPL38 for their activity. Despite numerous layers of *Hox* gene regulation, these IRES elements are essential for converting *Hox* transcripts into proteins to pattern the mammalian body plan. This specialized mode of IRES–dependent translation is enabled by an additional regulatory element that we term the translation inhibitory element (TIE), which blocks cap–dependent translation of transcripts. Together, these data uncover a new paradigm for ribosome–mediated control of gene expression and organismal development.

An important layer of post-transcriptional control of gene expression may be conferred through regulatory functions of the ribosome[1–6]. For example, RPL38, one of 80 ribosomal proteins of the eukaryotic ribosome, helps establish the mammalian body plan by selectively facilitating the translation of subsets of *Hox* mRNAs[1], genes critically required for formation of the body plan[7]. However, how ribosome-mediated regulation of gene expression is encoded within mRNA sequence remains an unanswered question.

All eukaryotic cellular mRNAs are capped, and cap-dependent translation is considered a highly efficient and predominant means for translation of most transcripts in the mammalian genome[8,9]. In many viral mRNAs that are not capped, IRESs provide an alternative mechanism for ribosome recruitment to promote translation initiation[8,10]. Interestingly, IRES elements have also been discovered in a handful of cellular mRNAs, including *c-myc*, *XIAP*, *Apaf-1* and *p53* (refs 11–16). As these cellular mRNAs are capped, their IRES elements act as a 'fail-safe' mechanism to promote translation under stress conditions, such as apoptosis or hypoxia, when cap-dependent translation is downregulated[17,18]. Cellular IRES elements have little sequence or structural homology[19], and their relevance for translational control and gene expression in normal cell physiology, tissue patterning, or organismal development is unclear.

We report below the unexpected identification of IRES elements within *Hox* mRNAs, particularly those regulated by RPL38, which are critically required for accurate gene expression during normal development. We also identify the TIE, an additional regulatory element in these mRNAs, which modulates the dependence of individual transcripts on canonical cap-dependent translation. This mechanism enables the IRES to become the predominant mode of translation initiation and confers greater gene regulatory potential by the ribosome.

RPL38–regulated *Hox* mRNAs possess IRES elements

To address the nature of specific *cis*-regulatory elements within transcripts that confer greater regulation by the ribosome, we studied RPL38-mediated control of *Hox* mRNA translation. We showed previously that although RPL38 haploinsufficiency within the developing mouse embryo does not affect general cap-dependent translation, the translation of a subset of the eleven *HoxA* mRNAs (*Hoxa4*, *Hoxa5*, *Hoxa9*, *Hoxa11*) is perturbed (Fig. 1a)[1]. We employed a bicistronic reporter assay in the murine mesenchymal stem cell line C3H10T1/2, which expresses *Hox* transcripts[20], to delineate whether *Hox* 5′ UTRs direct cap-independent translation (Fig. 1a, b). In this reporter system, the first cistron (*Renilla* luciferase, RLuc) is translated by a cap-dependent mechanism, and the second cistron (Firefly luciferase, FLuc) is translated only if the preceding 5′ UTR element can recruit ribosomes by a cap-independent mechanism (Fig. 1a); the latter is considered to reflect IRES-dependent translation initiation[21].

These experiments revealed that many *Hox* 5′ UTRs possess IRES activity that is as strong as, or even stronger than, the hepatitis C virus (HCV) IRES element, a bona fide viral IRES element (Fig. 1b). To ensure that *Hox* 5′ UTRs do not contain putative promoters or splice sites that may produce a monocistronic capped FLuc transcript, we performed additional controls. As expected for a single transcript containing both Rluc and Fluc: (i) the RNA levels of the two reporters are expressed in the same ratio among all bicistronic constructs; (ii) a short hairpin RNA (shRNA) against RLuc downregulated both Rluc and Fluc expression; (iii) these transcripts do not contain cryptic splice sites; and (iv) no ribosome read-through of the first reporter is evident (Extended Data Fig. 1). There is good correlation between RPL38 regulation and the presence of IRES elements; all RPL38-regulated mRNAs in the *HoxA* locus possess IRES elements. These include *Hoxa4*, which we show by 5′ rapid amplification of cloned ends (RACE) possesses an IRES-containing 5′ UTR of 1.1 kb, which is longer than previously annotated.

To test whether cap-independent translation from the *Hox* 5′ UTRs is controlled by RPL38 in C3H10T1/2 cells, a copy of *Rpl38* was disrupted by transcription activator-like effector nucleases (TALENs), yielding a 40% reduction in RPL38 protein expression (Fig. 1c, Extended Data Fig. 2a). In these cells, although there is no change in cap-dependent translation, there is a specific decrease in IRES-dependent translation of *Hox* target mRNAs that are regulated by RPL38 *in vivo* (Fig. 1c,

[1]Department of Developmental Biology, Stanford University, Stanford, California 94305, USA. [2]Department of Genetics, Stanford University, Stanford, California 94305, USA. [3]Tetrad Graduate Program, University of California, San Francisco, San Francisco, California 94158, USA. [4]Department of Biochemistry, Stanford University, Stanford, California 94305, USA. [5]Department of Physics, Stanford University, Stanford, California 94305, USA.

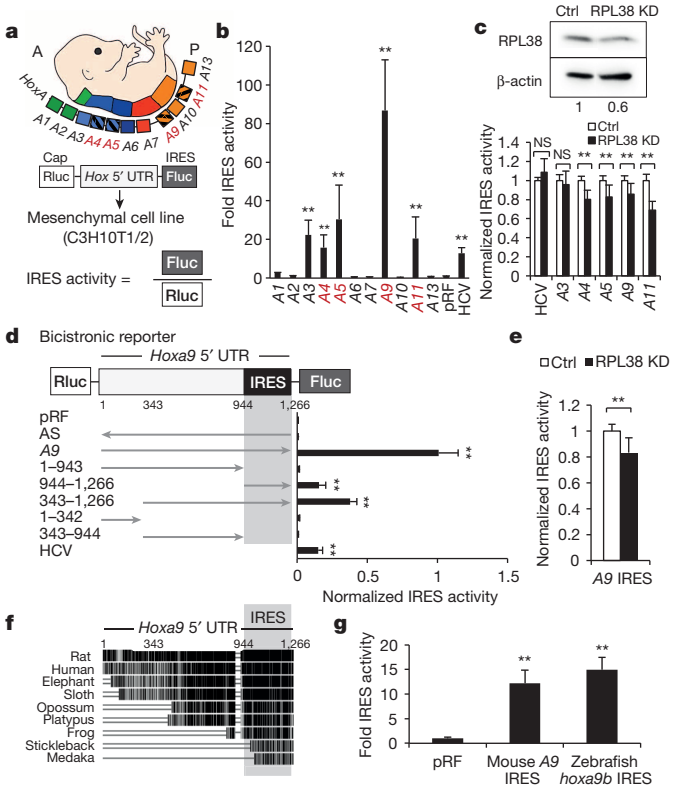


Figure 1 | Select *HoxA* 5′ UTRs contain IRES elements regulated by RPL38. **a**, *HoxA* genes are located in tandem along a chromosomal locus, and their location directs their anterior expression boundaries along the embryo. Striped boxes, *Hox* genes translationally regulated by RPL38. IRES activity is the ratio between Fluc and Rluc. **b**, IRES activity of *HoxA* 5′ UTRs. IRES activity of UTRs was normalized to that of empty vector (pRF) and RPL38-regulated *Hox* genes are denoted in red. $^{**}P < 0.01$ (χ^2 test against pRF), $n = 5$. **c**, RPL38 regulates translation of specific *HoxA* IRES elements. Bicistronic reporters were transfected into control cells (Ctrl) or cells where one copy of *Rpl38* was deleted (RPL38 KD). Top, western blot of RPL38 and β-actin with quantification of RPL38 levels below the blot. Bottom, IRES activity of specific *Hox* 5′ UTRs in RPL38 knockdown cells normalized to that of control cells. $^{**}P < 0.01$ (*t*-test), $n = 3$. **d**, Deletion analysis of the *Hoxa9* 5′ UTR. IRES activity was normalized to that of the full-length *Hoxa9* 5′ UTR (*A9*). AS, antisense. $^{**}P < 0.01$ (χ^2 test against pRF), $n = 3$. **e**, IRES activity of the *Hoxa9* IRES (nt 944–1,266) in RPL38 knockdown cells. $^{**}P < 0.01$ (*t*-test), $n = 3$. **f**, Alignment of the *Hoxa9* 5′ UTR in vertebrates. **g**, The 5′ UTR of one of two zebrafish *Hoxa9* paralogues (*hoxa9b*) has similar IRES activity as the mouse *Hoxa9* IRES. IRES activity was normalized to pRF. $^{**}P < 0.01$ (*t*-test as compared to pRF), $n = 3$. All experiments were performed in triplicates. Data represent mean ± s.d.

Extended Data Fig. 2b). Consistent with the fact that RPL38 control of IRES-dependent translation is transcript-specific, *Hoxa3* IRES-dependent translation, which is not regulated by RPL38 within the embryo[1], and the HCV IRES, are similarly unaffected by RPL38 knockdown (Fig. 1c). These results reveal the unexpected presence of IRES elements within *Hox* 5′ UTRs and a critical function of RPL38 in regulating their IRES-dependent translation.

The *Hoxa9* minimal IRES is evolutionarily conserved

To further understand IRES-dependent regulation of *Hox* mRNAs, we made a series of deletions within the 1.2-kb *Hoxa9* 5′ UTR. We localized the minimum fragment for RPL38-dependent IRES activity to nucleotides (nt) 944–1,266 (which we term the *Hoxa9* IRES element) (Fig. 1d). Although this fragment does not fully recapitulate the IRES activity observed with full-length *Hoxa9* 5′ UTR, it has strong IRES activity on its own and its removal abolishes all IRES activity (Fig. 1d, e). This region of the *Hoxa9* 5′ UTR shows remarkable evolutionary conservation in all vertebrates (from fish to mammals) that is greater than sequence conservation in the remaining 5′ UTR or the 3′ UTR sequence (Fig. 1f, Extended Data Fig. 3). IRES function is also evolutionarily conserved, as the zebrafish *hoxa9b* 5′ UTR shows strong IRES activity in murine C3H10T1/2 cells (Fig. 1g). These findings suggest that the *Hoxa9* IRES element may have arisen early during vertebrate evolution for post-transcriptional regulation of *Hox* expression.

Hoxa9 IRES is an RNA structure that recruits ribosomes

Many viral IRESs possess structures such as conserved helices, asymmetric bulges, and pseudoknots that interact with initiation factors or the ribosome to promote translation initiation[22–24]. We therefore examined whether the *Hoxa9* IRES element has structural properties *in vitro* that are functionally important.

Within the full-length *Hoxa9* 5′ UTR, protections of sub-regions of the 944–1,266 domain from selective 2′-hydroxyl acylation analysed by primer extension (SHAPE)[25] were consistent with the minimal IRES domain forming a specific RNA structure *in vitro* (Extended Data Fig. 4). Automated modelling indicated that this domain formed a four-way junction with two long hairpin arms (P3 and P4; Fig. 2a, Extended Data Fig. 6) and a 'right-angle' asymmetric bulge (between P3b and P3c), but with little interaction with other domains of the 5′ UTR (Extended Data Fig. 4). Analogous chemical mapping and functional studies on the 5′ UTR from a different *Hox* mRNA, *Hoxa5*, also revealed a structured subdomain (nt 77–132) necessary for its full IRES activity (Extended Data Fig. 5a, d, e). Deletion of the subdomain substantially reduces IRES activity (Extended Data Fig. 5b). Structural conservation between different cellular IRES elements has so far been poorly defined[19]. Interestingly, the 'right angle' asymmetric bulges observed in *Hoxa5* and *Hoxa9* (Extended Data Fig. 5c) bear qualitative similarity to asymmetric bulges in, for example, the HCV IRES[26], suggesting at least certain common structural features are present amongst cellular *Hox* IRESs and viral IRESs.

To gain further insight into structural and functional properties of a *Hox* IRES, we applied multidimensional chemical mapping to the *Hoxa9* minimal IRES subdomain (Fig. 2a, Extended Data Fig. 6). Comprehensive mutate-and-map analysis[27] gave strong independent evidence for the same structural topology seen in the context of the full-length 5′ UTR, revealing exposure of RNA segments upon mutation of their predicted pairing partners (Extended Data Fig. 6a–c). Importantly, compensatory mutation analysis[28] unambiguously confirmed specific base pairs in the P3 and P4 segments (green circles, Fig. 2a, and Extended Data Fig. 7). After this *in vitro Hoxa9* structural characterization, we tested the functional significance of these hairpins for IRES-dependent translation through bicistronic assays. Deletion of P3, P4, or both hairpins in the context of the full-length *Hoxa9* 5′ UTR resulted in a significant reduction in IRES activity. Most strikingly, the mutation of nucleotides in the base of P3a reduced IRES activity, whereas compensatory mutations gave quantitative rescue of IRES activity (Fig. 2b). These results demonstrate the importance of the P3 and P4 hairpins in IRES function and connect the detailed *in vitro Hoxa9* IRES structural model to functional activity. Many viral IRES elements contain pseudoknots that may position the ribosome at the initiation codon[29]. A potential pseudoknot was suggested by mutate-and-map analysis (Fig. 2a), but this awaits further analysis. Together, these data confirm a structure defined at base-pair resolution for a *Hox* IRES element and indicate that a specific RNA structure is important for cellular *Hox* IRES activity.

To date, only a small class of viral IRES elements have been shown to interact with both the large and small ribosome subunits to form a translationally competent 80S ribosome[30,31]. However, a biotinylated full-length *Hoxa9* 5′ UTR, as well as the minimal IRES element contained within nt 944–1,266, are able to pull down ribosomal proteins from both the large and small subunits, including RPL38 (Fig. 2c, d). The full-length 5′ UTR also pulls down both 28S and 18S rRNAs (Fig. 2e), suggesting that the 80S ribosome is able to form on the

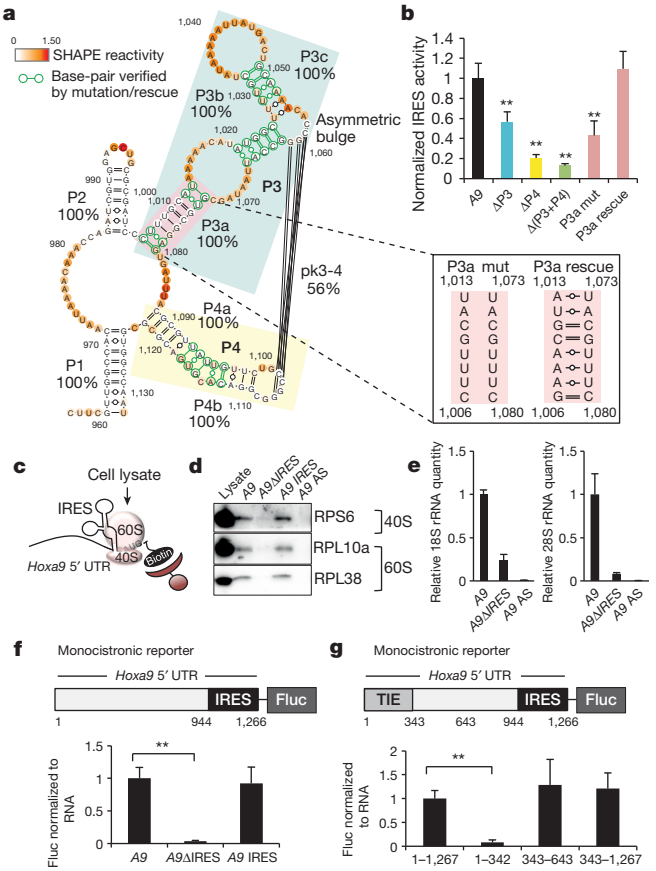


Figure 2 | The *Hoxa9* IRES is an RNA structure that facilitates recruitment of the ribosome and is required for normal translation. **a**, Secondary structure of *Hoxa9* 5′ UTR (nt 957–1,132) by SHAPE and mutate-and-map analysis. Nucleotides are coloured with SHAPE reactivities. Percentages give bootstrap support values for each helix, including a tentative pseudoknot (pk). Base-pairs verified by mutation and rescue are circled green. Inset, mutations made in helix P3a for functional testing in **b**. **b**, Deletions as well as mutation-and-rescue of specific hairpins identified in *Hoxa9* IRES structure shown in **a** affects IRES activity in bicistronic reporter assays. IRES activity was normalized to full-length *Hoxa9* 5′ UTR (*A9*, set to 1). $^{**}P < 0.01$ (*t*-test as compared to *A9*). $n = 3$ and experiments performed in triplicates. **c**, Schematic of RNA pull-down experiments. **d**, **e**, Representative western blot for RPs (**d**) and qPCR for rRNA (**e**) from eluted material of two independent RNA pull-down experiments. The 18S and 28S rRNA amounts of all *Hoxa9* deletion mutants were normalized to amount of full-length *Hoxa9* 5′ UTR (set to 1) eluted from beads, $n = 2$. **f**, Deletions within the *Hoxa9* 5′ UTR in a monocistronic reporter assay. Fluc activity was normalized to amount of luciferase mRNA determined by quantitative PCR (qPCR). All values are normalized to *Hoxa9* full-length 5′ UTR (*A9*), set to 1; $n = 4$. **g**, A TIE that inhibits cap-dependent translation was identified in the *Hoxa9* 5′ UTR (nt 1–342); $n = 3$ and experiments were performed in duplicates. Data represent mean ± s.d.

uncapped *Hoxa9* 5′ UTR. Importantly, the full-length *Hoxa9* 5′ UTR lacking the nt 944–1,266 IRES element (*A9ΔIRES*) or an antisense *Hoxa9* 5′ UTR (*A9* AS) is no longer able to efficiently recruit ribosomal proteins or ribosomal RNA (Fig. 2d, e), consistent with the fact that these fragments do not possess IRES activity (Fig. 1d). These data support the idea that a structured element from a *Hox* IRES facilitates 80S ribosome recruitment to the mRNA, either directly or through intermediary RNA-binding proteins.

TIE inhibits cap-dependent translation

Because all eukaryotic mRNAs are capped and translated by a cap-dependent mechanism[8], the function of an IRES-element in normal translational control remains unclear. Surprisingly, removing the IRES element from a *Hoxa9* 5′ UTR monocistronic reporter construct dramatically decreases translational efficiency by 70-fold, despite the fact that the mRNA remains capped (Fig. 2f). We therefore hypothesized that the *Hoxa9* 5′ UTR may contain additional element(s) that intrinsically inhibit cap-dependent translation. The *Hoxa9* 5′ UTR contains 14 putative upstream open reading frames (uORFs), which are generally thought to inhibit the translation of the main open reading frame[32]. However, mutating all 14 ATGs to TTGs in the *Hoxa9* 5′ UTR (*A9ΔIRESΔuORF*) (Extended Data Fig. 8a) does not affect cap-dependent translation of *Hoxa9*. We therefore placed various regions of the *Hoxa9* 5′ UTR upstream of a monocistronic reporter to assay for effects on cap-dependent translation. Strikingly, nt 1–342 of the *Hoxa9* 5′ UTR strongly inhibit cap-dependent translation (Fig. 2g) despite this region not being particularly GC-rich (55%) and therefore unlikely to impede ribosome scanning. Other similarly sized pieces of the 5′ UTR (nt 343–643), or the very long remaining portion of the 5′ UTR (nt 343–1,266), do not inhibit cap-dependent translation and are probably translated by a combination of cap and IRES-mediated translation. To our knowledge, the 1–342 fragment of the *Hoxa9* 5′ UTR is a unique example of a 5′ UTR regulon that is sufficient to strongly inhibit cap-dependent translation under physiological conditions. We termed this 5′ UTR mRNA element the translation inhibitory element (TIE). SHAPE analysis of the full-length *Hoxa9* 5′ UTR lacks evidence for structural interactions between the TIE and the minimal IRES (Extended Data Fig. 4a), suggesting they function as independent modules. These findings reveal that the TIE enables a more specialized mode of translation initiation directed by the IRES.

5′ UTR topology allows for specialized translation

We next asked whether the unique 5′ UTR topology of both a TIE close to the 5′ end of the mRNA and an IRES element is a more generalizable mechanism of translational regulation. Our findings reveal that all the minimal IRES domains of *HoxA* 5′ UTRs lie within the 100–300 nt closest to the AUG, with the exception of *Hoxa5* (Fig. 3). Deletions of these minimal IRES domains in monocistronic reporter constructs reveal that they are critically required for normal translation (Fig. 4a–c). In addition, within these *HoxA* 5′ UTRs, a TIE is also present near the mRNA cap (Fig. 4a–c). Moreover, although the *Hoxa3*, *Hoxa4*, *Hoxa9* and *Hoxa11* TIEs do not share obvious sequence similarity, they all strikingly block translation from the β-globin (*Hbb*) 5′ UTR (Fig. 4d), which initiates translation exclusively by a cap-dependent mechanism[33]. *Hoxa5* is the only 5′ UTR that does not appear to possess a TIE and a putative uORF also does not decrease cap-dependent translation (Extended Data Fig. 8b). However, *Hoxa5* has the shortest 5′ UTR (237 nt) and almost the entire UTR is required for IRES activity, leaving little space for an additional TIE (Fig. 3c). Interestingly, the *Hoxa5* transcript can be produced as a bicistronic mRNA with the neighbouring *Hoxa6* gene[34] and therefore may not need a TIE to inhibit cap-dependent translation in this context. These findings reveal a unique topology to *Hox* 5′ UTRs that suppresses a more generic mode of cap-dependent translational control and facilitates specialized, ribosome-mediated control of protein expression.

Hoxa9 IRES is essential for translation *in vivo*

To determine whether IRES-dependent translation is critically required for the conversion of the *Hoxa9* transcript into protein *in vivo*, we generated the first targeted mouse knockout of a cellular IRES by removing nt 944–1,193 of the *Hoxa9* 5′ UTR, leaving the Kozak sequence and the rest of the 5′ UTR intact (Fig. 5a, Extended Data Fig. 8c, d). A caveat to the previously generated HOXA9 targeted knockout is the retention of a neomycin (Neo) cassette[35,36], which can have indirect effects on the expression of neighbouring *Hox* genes[37]. Indeed, our current genetic studies reveal that the presence of the Neo cassette within the *Hoxa9* targeting locus is sufficient to produce phenotypes previously attributed to *Hoxa9* loss-of-function (Extended Data Fig. 9).

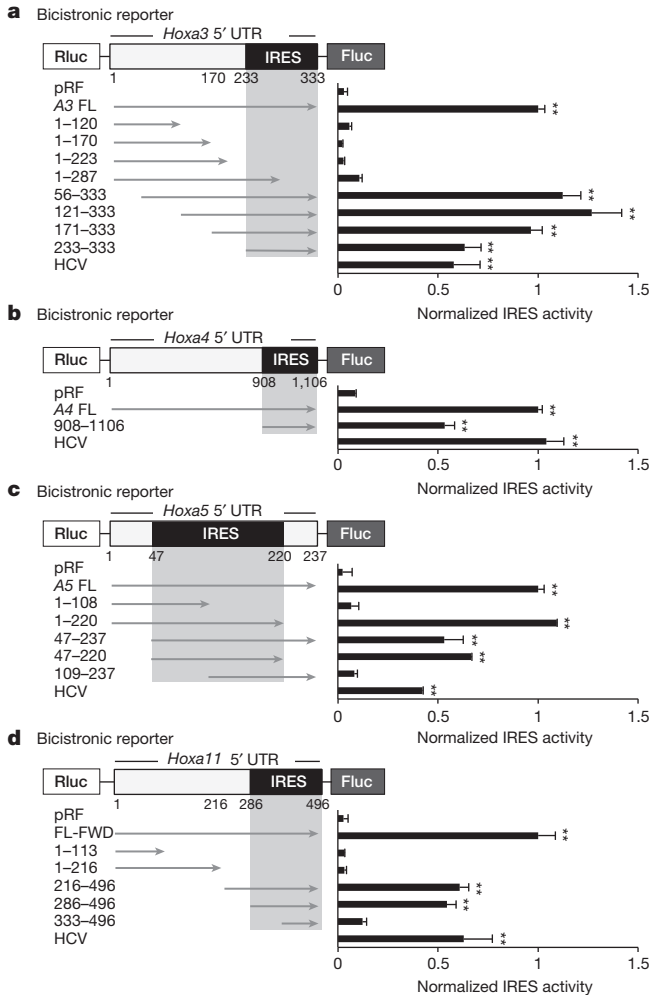


Figure 3 | Identification of minimal IRES domains in all IRES-containing *HoxA* 5′ UTRs. **a**–**d**, Deletion analysis of *HoxA* 5′ UTRs performed in bicistronic constructs maps the IRES element in *Hoxa3* (**a**), *Hoxa4* (**b**), *Hoxa5* (**c**) and *Hoxa11* (**d**). IRES activity was normalized to full-length 5′ UTRs (FL). $^{**}P < 0.01$ (χ^2 test against pRF). $n = 3$ and experiments were performed in triplicate. Data represent mean ± s.d.

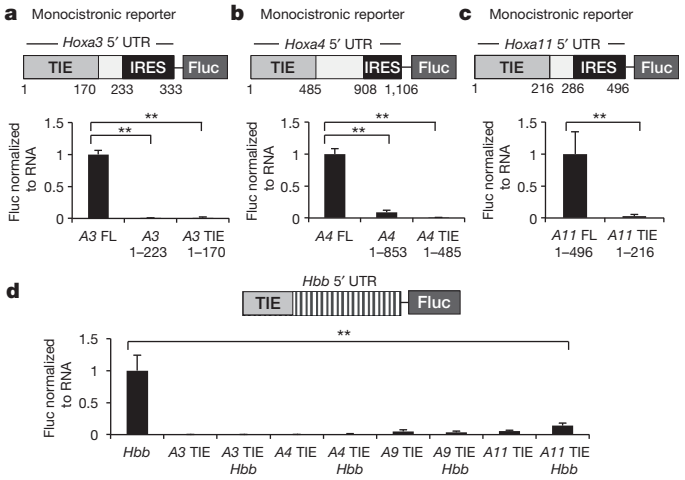


Figure 4 | Identification of a common TIE that blocks cap-dependent translation of *Hox* IRES-mRNAs. **a**–**c**, Deletion of the minimal IRES elements within monocistronic reporters of *HoxA* 5′ UTR and identification of TIEs close to the mRNA cap in *Hoxa3* (**a**), *Hoxa4* (**b**) and *Hoxa5* (**c**). All values are normalized to the full-length constructs (FL); $n = 3$. **d**, *HoxA* TIEs were placed upstream of the β-globin 5′ UTR (*Hbb*) in a monocistronic reporter to assay inhibition of cap-dependent translation, $n = 3$. All experiments were performed in duplicate. Data represent mean ± s.d. $^{**}P < 0.01$ (*t*-test).

Therefore, in our $Hoxa9^{\Delta IRES/\Delta IRES}$ mouse, we were careful to remove the Neo cassette from the targeted locus by Cre-mediated excision.

$Hoxa9^{\Delta IRES/\Delta IRES}$ embryos do not show any major change in *Hoxa9* transcript boundaries or expression levels (Fig. 5b, c). Strikingly, however, these mice show a fully penetrant homeotic transformation of the thirteenth thoracic vertebra (T13) to the first lumbar vertebra (L1), in which a full set of ribs normally present on T13 is missing (Fig. 5d). Although we cannot at present ascertain whether the *Hoxa9* IRES deletion completely phenocopies a *Hoxa9* knockout mouse, this dramatic homeotic transformation is precisely at the anterior boundary of *Hoxa9* expression at the thoracic to lumbar transition[35]. This transformation is also evident in ~50% of $Hoxa9^{\Delta IRES/+}$ mice, although small rudimentary ribs can still form on T13. Consistently, there is a striking decrease in HOXA9 protein levels in the neural tube and somites of $Hoxa9^{\Delta IRES/\Delta IRES}$ embryos from the thoracic to sacral level (Fig. 6a) revealing a block in HOXA9 protein production. Furthermore, in $Hoxa9^{\Delta IRES/\Delta IRES}$ embryos there is a dramatic decrease in ribosome association of the *Hoxa9* transcript. The majority of the *Hoxa9* transcript in $Hoxa9^{\Delta IRES/\Delta IRES}$ embryos, but not control transcripts such as β-actin, accumulates in pre-polysomal fractions, reflecting an accumulation of mRNA not bound by translationally active ribosomal subunits (Fig. 6b, Extended Data Fig. 10). These findings reveal that the *Hoxa9* IRES element is crucial in recruiting the ribosome to this transcript to pattern the mammalian body plan *in vivo*.

Discussion

Our studies reveal how expression information encoded within the mRNA template is sufficient to confer greater gene regulatory potential by the ribosome. In particular, our findings suggest that IRES elements are key regulatory elements within the mammalian genome that facilitate important, ribosome-mediated control of gene expression in development (Fig. 6c). Cellular IRES elements have been controversial because of their weak IRES activity as compared to viral

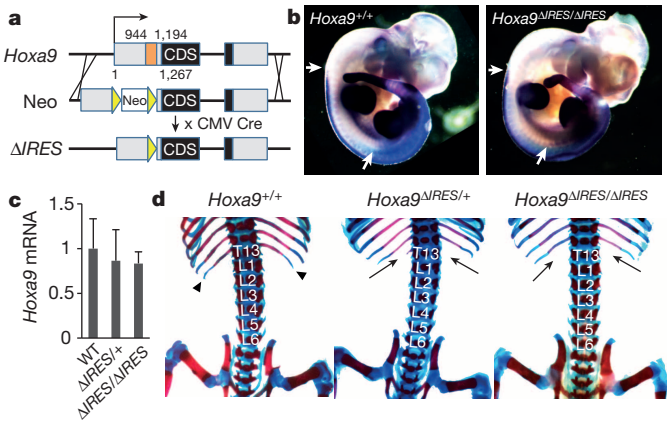


Figure 5 | The *Hoxa9* IRES is required for axial skeleton patterning but not for *Hoxa9* mRNA expression. **a**, Scheme of *Hoxa9* IRES targeted deletion in the mouse. *LoxP* sites are depicted by yellow triangles. The region deleted is highlighted in orange. **b**, Representative *in situ* hybridization of *Hoxa9* mRNA expression at E11.5, $n = 3$ embryos of each genotype. Arrows indicate the anterior boundaries of expression in the neural tube and somites. **c**, qPCR shows little difference in *Hoxa9* transcript levels in neural tube and somites between $Hoxa9^{+/+}$, $Hoxa9^{\Delta IRES/+}$ and $Hoxa9^{\Delta IRES/\Delta IRES}$ embryos. $Hoxa9^{\Delta IRES/+}$, $P = 0.06$; $Hoxa9^{\Delta IRES/\Delta IRES}$, $P = 0.03$, *t*-test as compared to $Hoxa9^{+/+}$. $n = 5$ embryos of each genotype. **d**, Representative skeletons of $Hoxa9^{+/+}$, $Hoxa9^{\Delta IRES/+}$ and $Hoxa9^{\Delta IRES/\Delta IRES}$ mice. Arrows point to the missing rib(s) on T13, normally present in WT mice (arrowheads), revealing a posterior homeotic transformation to L1. $n = 5$ mice of each genotype.

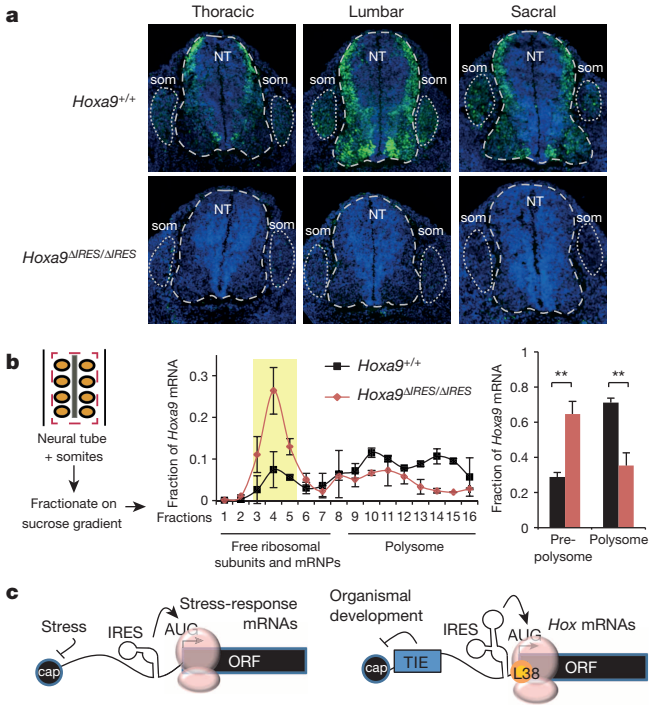


Figure 6 | The *Hoxa9* IRES is critically required for *Hoxa9* translation *in vivo*. **a**, Representative immunostaining of HOXA9 (green) in cross-section of E11.5 $Hoxa9^{+/+}$ (top) and $Hoxa9^{\Delta IRES/\Delta IRES}$ (bottom) embryos from the thoracic to sacral levels. DAPI (4′,6-diamidino-2-phenylindole) staining is in blue. NT, neural tube; som, somite. $n = 3$ embryos of each genotype. **b**, Somites and neural tubes of E11.5 $Hoxa9^{+/+}$ and $Hoxa9^{\Delta IRES/\Delta IRES}$ embryos were microdissected and fractionated on a sucrose gradient. Middle, qPCR of *Hoxa9* mRNA from each fraction. Early pre-polysome fractions are highlighted. Right, quantification of *Hoxa9* mRNA in fractions. Fractions 1–8 are pre-polysome and 9–16 polysome fractions. $^{**}P < 0.01$ (*t*-test compared to $Hoxa9^{+/+}$). $n = 3$ embryos of each genotype. **c**, Model for how ribosome-mediated regulation of gene expression is encoded within the mRNA sequence. Stress-responsive cellular IRES elements are only active when cap-dependent translation is downregulated through a decrease in eIF4F activity. *HoxA* mRNAs possess a TIE, which normally inhibits cap-dependent translation. The TIE enables ribosome-mediated control of *HoxA* IRES expression during embryonic development.

IRES elements under non-stress conditions. Their relevance to normal gene regulation is also poorly understood, as all eukaryotic mRNAs are capped and undergo cap-dependent translation[21,38]. Importantly, the requirement for *Hox* IRES elements is explained by our identification of the TIE within these transcripts, which strongly inhibits general, cap-dependent translation (Fig. 6c). We speculate that such IRES elements may be more widespread in the genome and suggest a two-state mechanism for gene regulation, controlled by transcription and an additional layer of *cis*-acting RNA regulons within 5′ UTRs that direct protein production. Both IRES and TIE elements may be differentially regulated, providing a versatile toolbox for controlling the ultimate expression of transcripts. Although *Hox* IRES elements are highly conserved, and may already be present in *Drosophila* orthologues[39], the TIE lacks conservation and is absent even in vertebrate species such as fish and amphibians (Fig. 1f). These species often lay their eggs in harsh environments[40] and may have originally used IRES elements as a fail-safe mechanism to ensure proper expression of key transcripts under stress conditions. Therefore, the acquisition of a TIE later in evolution may have enabled an additional, independent level of regulatory control to IRES-dependent translation and gene expression during normal development. A mechanism that regulates whether a pre-existing transcript is converted into a protein offers a fast and dynamic means to provide greater variation and control to gene expression.

It will be interesting to determine if additional ribosomal proteins may promote specialized translation through control of unique subsets of IRES-containing mRNAs, either directly or through RNA-binding proteins. For example, RPS25 is required for IRES-dependent translation of certain viral IRES mRNAs[41]. Moreover, rRNA modifications both at the level of pseudouridylation and RPL13a-dependent methylation also appear to regulate the translation of certain cellular IRES-containing mRNAs[14,16,42,43]. We therefore speculate that similar to the complex and highly regulated system of transcriptional control, in which specific DNA sequences and histone marks regulate gene expression, *cis*-acting RNA regulons, in conjunction with more specialized ribosome activity, provide newfound regulatory control to gene expression critical for mammalian development.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 March; accepted 30 October 2014.
Published online 19 November 2014.

1. Kondrashov, N. *et al.* Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. *Cell* **145,** 383–397 (2011).
2. Lee, A. S., Burdeinick-Kerr, R. & Whelan, S. A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs. *Proc. Natl Acad. Sci. USA* **110,** 324–329 (2013).
3. Vesper, O. *et al.* Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in *Escherichia coli*. *Cell* **147,** 147–157 (2011).
4. Xue, S. & Barna, M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. *Nature Rev. Mol. Cell Biol.* **13,** 355–369 (2012).
5. Dinman, J. D. The eukaryotic ribosome: current status and challenges. *J. Biol. Chem.* **284,** 11761–11765 (2009).
6. Komili, S., Farny, N. G., Roth, F. P. & Silver, P. A. Functional specificity among ribosomal proteins regulates gene expression. *Cell* **131,** 557–571 (2007).
7. Alexander, T., Nolte, C. & Krumlauf, R. Hox genes and segmentation of the hindbrain and axial skeleton. *Annu. Rev. Cell Dev. Biol.* **25,** 431–456 (2009).
8. Sonenberg, N. & Hinnebusch, A. G. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* **136,** 731–745 (2009).
9. Livingstone, M., Atas, E., Meller, A. & Sonenberg, N. Mechanisms governing the control of mRNA translation. *Phys. Biol.* **7,** 021001 (2010).
10. Plank, T.-D. M. & Kieft, J. S. The structures of nonprotein-coding RNAs that drive internal ribosome entry site function. *Wiley Interdiscip. Rev. RNA* **3,** 195–212 (2012).
11. Stoneley, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A. & Willis, A. E. C-Myc 5′ untranslated region contains an internal ribosome entry segment. *Oncogene* **16,** 423–428 (1998).
12. Riley, A., Jordan, L. E. & Holcik, M. Distinct 5′ UTRs regulate XIAP expression under normal growth conditions and during cellular stress. *Nucleic Acids Res.* **38,** 4665–4674 (2010).
13. Ungureanu, N. H. *et al.* Internal ribosome entry site-mediated translation of Apaf-1, but not XIAP, is regulated during UV-induced cell death. *J. Biol. Chem.* **281,** 15155–15163 (2006).
14. Yoon, A. *et al.* Impaired control of IRES-mediated translation in X-linked dyskeratosis congenita. *Science* **312,** 902–906 (2006).
15. Ray, P. S., Grover, R. & Das, S. Two internal ribosome entry sites mediate the translation of p53 isoforms. *EMBO Rep.* **7,** 404–410 (2006).
16. Bellodi, C., Kopmar, N. & Ruggero, D. Deregulation of oncogene-induced senescence and p53 translational control in X-linked dyskeratosis congenita. *EMBO J.* **29,** 1865–1876 (2010).
17. Pyronnet, S., Dostie, J. & Sonenberg, N. Suppression of cap-dependent translation in mitosis. *Genes Dev.* **15,** 2083–2093 (2001).
18. Holcik, M., Sonenberg, N. & Korneluk, R. G. Internal ribosome initiation of translation and the control of cell death. *Trends Genet.* **16,** 469–473 (2000).
19. Xia, X. & Holcik, M. Strong eukaryotic IRESs have weak secondary structure. *PLoS ONE* **4,** e4136 (2009).
20. Phinney, D. G., Gray, A. J., Hill, K. & Pandey, A. Murine mesenchymal and embryonic stem cells express a similar Hox gene profile. *Biochem. Biophys. Res. Commun.* **338,** 1759–1765 (2005).
21. Thompson, S. R. So you want to know if your message has an IRES? *Wiley Interdiscip. Rev. RNA* **3,** 697–705 (2012).
22. Fraser, C. S. & Doudna, J. A. Structural and mechanistic insights into hepatitis C viral translation initiation. *Nature Rev. Microbiol.* **5,** 29–38 (2007).
23. Kieft, J. S., Zhou, K., Jubin, R. & Doudna, J. A. Mechanism of ribosome recruitment by hepatitis C IRES RNA. *RNA* **7,** 194–206 (2001).
24. Spahn, C. M. T. *et al.* Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell* **118,** 465–475 (2004).
25. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature Protocols* **1,** 1610–1616 (2006).
26. Lukavsky, P. J., Kim, I., Otto, G. A. & Puglisi, J. D. Structure of HCV IRES domain II determined by NMR. *Nature Struct. Biol.* **10,** 1033–1038 (2003).

27. Kladwang, W., VanLang, C. C., Cordero, P. & Das, R. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat. Chem.* **3,** 954–962 (2011).
28. Tian, S., Cordero, P., Kladwang, W. & Das, R. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **20,** 1815–1826 (2014).
29. Berry, K. E., Waghray, S. & Doudna, J. A. The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA* **16,** 1559–1569 (2010).
30. Jan, E. & Sarnow, P. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.* **324,** 889–902 (2002).
31. Stupina, V. A., Yuan, X., Meskauskas, A., Dinman, J. D. & Simon, A. E. Ribosome binding to a 5′ translational enhancer is altered in the presence of the 3′ untranslated region in cap-independent translation of turnip crinkle virus. *J. Virol.* **85,** 4638–4653 (2011).
32. Calvo, S. E., Pagliarini, D. J. & Mootha, V. K. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc. Natl Acad. Sci. USA* **106,** 7507–7512 (2009).
33. Fletcher, L., Corbin, S. D., Browning, K. S. & Ravel, J. M. The absence of a m^7G cap on β-globin mRNA and alfalfa mosaic virus RNA 4 increases the amounts of initiation factor 4F required for translation. *J. Biol. Chem.* **265,** 19582–19587 (1990).
34. Coulombe, Y. *et al.* Multiple promoters and alternative splicing: *Hoxa5* transcriptional complexity in the mouse embryo. *PLoS ONE* **5,** e10600 (2010).
35. 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).
36. Chen, F. & Capecchi, M. R. Targeted mutations in *Hoxa-9* and *Hoxb-9* reveal synergistic interactions. *Dev. Biol.* **181,** 186–196 (1997).
37. Boulet, A. M. & Capecchi, M. R. Targeted disruption of *hoxc-4* causes esophageal defects and vertebral transformations. *Dev. Biol.* **177,** 232–249 (1996).
38. Komar, A. A. & Hatzoglou, M. Internal ribosome entry sites in cellular mRNAs: mystery of their existence. *J. Biol. Chem.* **280,** 23425–23428 (2005).
39. Oh, S. K., Scott, M. P. & Sarnow, P. Homeotic gene *Antennapedia* mRNA contains 5′-noncoding sequences that confer translational initiation by internal ribosome binding. *Genes Dev.* **6,** 1643–1653 (1992).
40. Johnston, I. A. Environment and plasticity of myogenesis in teleost fish. *J. Exp. Biol.* **209,** 2249–2264 (2006).
41. Landry, D. M., Hertz, M. I. & Thompson, S. R. RPS25 is essential for translation initiation by the *Dicistroviridae* and hepatitis C viral IRESs. *Genes Dev.* **23,** 2753–2764 (2009).
42. Bellodi, C. *et al.* Loss of function of the tumor suppressor DKC1 perturbs p27 translation control and contributes to pituitary tumorigenesis. *Cancer Res.* **70,** 6026–6035 (2010).
43. Chaudhuri, S. *et al.* Human ribosomal protein L13a is dispensable for canonical ribosome function but indispensable for efficient rRNA methylation. *RNA* **13,** 2224–2237 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank members of the Barna laboratory and D. Ruggero for discussion and critical reading of the manuscript. We thank E. Sarinay Cenik for advice with RNA pull-down experiments; and Y. Rim, A. Sapiro and A. Mateo for technical assistance. This work was supported by the Agency of Science, Technology and Research of Singapore (S.X.), Stanford Graduate Fellowship (S.T.), Human Frontier Science Program Fellowship (K.F.), NIH R01 GM102519 (R.D.), NIH Director's New Innovator Award, 7DP2OD00850902 (M.B.), Alfred P. Sloan Research Fellowship (M.B.) and Pew Scholars Award (M.B.).

Author Contributions M.B. conceived and supervised the project; S.X. and M.B. designed experiments; S.X. performed all experiments with help from K.F. for 5′ RACE and sucrose gradient fractionation; S.T. and R.D. designed the RNA structure experiments; S.T. and W.K. performed the RNA structure experiments and analysed results; S.X. and M.B. wrote the manuscript with contributions from all authors.

Author Information *Hoxa4* 5′ UTR sequence has been deposited in GenBank under accession number KM596709. All chemical mapping datasets have been deposited at the RNA Mapping Database (http://rmdb.stanford.edu) under the following accession codes: (1) Full-length: HOXA5_STD_0000, HOXA9_STD_0000; (2) *Hoxa9* TIE: HOXA9A_STD_0001; (3) *Hoxa9* IRES: HOXA9D_STD_0001, HOXA9D_STD_0002, HOXA9D_1M7_0001, and HOXA9D_RSQ_0001. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (mbarna@stanford.edu).

METHODS

Plasmids. *Hox* 5′ UTRs were amplified from mouse E10.5–12.5 cDNA and cloned into the bicistronic pRF vector[14] using EcoRI and NcoI sites. For monocistronic assays, *Hox* 5′ UTRs are cloned into pGL3 control vector (Promega) using HindIII and NcoI sites. 5′ UTRs of *Hoxa3* and *Hoxa10* were made by gene synthesis (IDT). pRF-HCV was a kind gift from D. Ruggero (UCSF). Zebrafish *hoxa9b* 5′ UTR was amplified from zebrafish 24 hpf cDNA. Sequences were based off the Ensembl database[44]. See Supplementary Table 1 for a list of PCR primers used.

To ensure no ribosome read-through of the first cistron in bicistronic assays, a hairpin with a linker (GAATTCAGATCTGGTACCGAGCTCCCCGGGCTGC AGGATATCCTGCAGCCCGGGGACCTCGGTACCAGATCTAGGATCCGG ATCCGGCGACACTCCACCATGAATCACTCCCCTGTGAGGAACTACTGT CTTCACGCAGAAAGCGTCTAGCCATGG) was synthesized and ligated into the pRF vector at the EcoRI and NcoI sites. *Hox* 5′ UTRs and HCV IRES were cloned into this new vector pRF-hpC in the NcoI site. Clones were sequenced to ensure directionality.

5′ RACE. RNA was purified from dissected E11.5 neural tubes and somites using TRIzol (Invitrogen). The reverse transcription template was produced using ExactSTART Eukaryotic mRNA 5′-& 3′-RACE Kit (Epicentre) and reverse transcribed using Superscript III (Invitrogen) with a primer in the coding region of *Hoxa4*. The 5′ UTR fragment was amplified using Phusion High-Fidelity DNA polymerase (Thermo Scientific) with a gene specific reverse primer and a forward primer at the 5′ linker. PCR products were adenylated, gel-extracted and subcloned into TOPO TA cloning vector (Invitrogen) for sequencing.

TALEN. 2 sets of TALEN pairs were designed against the second and third exons of *Rpl38*. TALENs were generated towards the following sequences: F1 5′-GC GGACGCCAGAGCCGA, R1 5′- CCCTCGTCCCGCACTCA, F2 5′- CATCTT CTGTCCCTC, R2 5′- TGATCTCCTCAATTTTC. Intermediate arrays were produced for each TALEN-pair and assembled into pC-GoldyTALEN[45]. The 2 sets of TALEN pairs were transfected together into C3H10T1/2 cells. Cells were trypsinized into single cell suspension and plated for clonal expansion 2 days after transfection. Positive clones were identified by PCR followed by restriction digest. *Rpl38* disruption was confirmed by Sanger sequencing and western blotting.

Luciferase assay. C3H10T1/2 cells were transfected in 24-well plates with various pRF constructs using Lipofectamine 2000 (Invitrogen) as per manufacturer instructions. Cells were collected 24 h post-transfection and assayed using Dual luciferase kit (Promega). IRES activity is expressed as a ratio between Fluc and Rluc. To verify the integrity of the bicistronic construct, an shRNA plasmid against Rluc or scrambled plasmid[16] was cotransfected with the pRF constructs. Cells were harvested 72 h post-transfection for luciferase assays. For monocistronic luciferase assays, various pGL3 constructs were transfected as above into 12-well plates. Cells were collected 24 h later. Then 50% of the cells were used for qPCR analysis with Fluc primers to quantify total transcript abundance, the rest for luciferase assays. Fluc activity was normalized to total luciferase mRNA expression levels.

RNA pull down. Template DNA was made by PCR from pRF-HoxA9 using a 5′ primer containing the T7 promoter and a 3′ primer containing a linker sequence T2 (Supplementary Table 1). Biotinylated RNA was *in vitro* transcribed with Megascript T7 kit (Ambion) using a final concentration of 6.8 mM each of ATP, CTP and GTP, 6.5 mM UTP and 0.3 mM Biotin-16-UTP (Roche). Synthesized RNAs were treated with TURBO DNase (Ambion) and purified by phenol/chloroform extraction. Unincorporated nucleotides were removed by running the RNA through a G50 column (GE Healthcare). The RNA was analysed by electrophoresis for quality and Nanodrop for quantity.

A total of four 15-cm dishes of subconfluent C3H10T1/2 cells were collected for analysis using trypsinization. Cell pellets were washed and lysed on ice with 1.5 ml lysis buffer (10 mM Tris pH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, 0.1% TritonX, 0.25% NP40, 10 mM DTT, RNaseOUT, Halt Protease inhibitor) for 30 min. Lysate was centrifuged at 14,000 rpm for 10 min, 4 °C. Supernatant was transferred to new tube and cleared with 400 µl avidin-agarose beads (Thermo) for 50 min, 4 °C. Cleared lysate was transferred to new tube and quantified by spectrometry.

Folding buffer was added to 100 pmol of biotinylated RNA of each species to a final concentration of 100 mM Tris pH8, 100 mM NaCl, 10 mM $MgCl_2$. RNA was heated to 95 °C for 1.5 min, ice for 2 min and 37 °C for 20 min. Then 100 pmol of folded RNA and 1 mg cleared lysate were incubated together with 100 µg ml^{-1} heparin and 100 µg ml^{-1} yeast tRNA in 2 ml lysis buffer for 1 h at room temperature.

Streptavidin M280 Dynabeads (Invitrogen) were washed according to manufacturer's instructions. 100 µl beads were incubated with each RNA per lysate sample for 1.5 h at 4 °C. Beads were then washed with lysis buffer with increasing concentrations of salt up to 600 mM NaCl. One-tenth of the bead sample was kept for RNA and the rest was used for protein analysis. RNA was eluted with 50 µl RNA denaturing buffer (95% formamide, 10 mM EDTA) at 65 °C for 5 min. Eluted RNA was purified by phenol/chloroform extraction. RNA was converted to cDNA and analysed by qPCR using primers *HoxA9* 5′ UTR F and *HoxA9* 342 R (Supplementary Table 1). Proteins were eluted with 40 µl protein elution buffer (5 mM biotin, 3× SDS sample buffer) at 95 °C for 5 min. Proteins were analysed by western blotting.

Western blotting. Western blots were performed with antibodies against RPS6 (Cell Signaling no. 2217) and RPL10a (Santa Cruz sc-100827). Polyclonal RPL38 antibodies were generated against C-EKAEKLKQSLPPGLAVKDLK.

RNA structural probing and data analysis. Double-stranded DNA templates were prepared by PCR assembly of DNA oligomers with maximum length of 60 nt ordered from IDT (Integrated DNA Technologies). DNA templates for mutate-and-map (M^2) and mutation/rescue were designed as previously described using NA_Thermo (https://github.com/DasLab/NA_thermo) and HiTRACE[27,28]. DNA template and RNA transcript preparation and quality checks were carried out as previously described[27,28].

All chemical mapping were carried out in 96-well format as described previously[27,28,46]. Briefly, before chemical modification, RNA was heated to 90 °C for 30 seconds, cooled on the bench top to room temperature to remove secondary structure heterogeneity, and folded for 20 min at 37 °C in 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0. RNA was modified by adding 1/4 volume freshly made 5 mg ml^{-1} 1-methyl-7-nitroisatoic anhydride (1M7)[47] in anhydrous DMSO or neat anhydrous DMSO as control. Modification reactions were incubated at room temperature and then quenched with 0.5 M Na-MES, pH 6.0. Quenches also included poly(dT) magnetic beads (Ambion) and FAM-labelled Tail2-A20 primer for reverse transcription. Samples were separated and purified using magnetic stands, washed with 70% ethanol twice, and air-dried. Beads were resuspended in ddH_2O and reverse transcription mix, then incubated at 55 °C for 30 min. RNA was degraded by addition of 1 volume of 0.4 M NaOH and incubation at 90 °C for 3 minutes, then cooled and neutralized with an additional volume of acid quench (prepared as 2 volumes of 5 M NaCl, 2 volumes of 2 M HCl, and 3 M NaOAc pH 5.2). Fluorescent labelled cDNA was recovered by magnetic bead separation, rinsed twice with 70% ethanol and air-dried. The beads were resuspended in Hi-Di formamide containing ROX-350 ladder (Applied Biosystems), then loaded on capillary electrophoresis sequencer (ABI3100).

For full-length *Hox* transcripts, an Illumina MiSeq-based ligSHAPE (ligation-based SHAPE) on long RNAs was carried out, using a protocol analogous to Mod-seq[48]. RNA was folded similarly and probed by SHAPE as above, then recovered by ethanol precipitation. To obtain 3′ ends for ligating primer-binding sites at unstructured regions, modified RNA was folded again and fragmented by 5 mM $TbCl_3$ for 5 min (or no Tb^{3+} as control)[49], quenched by 0.5 M Na-EDTA pH 8.0, then purified by RNA cleanup and concentration columns (Zymo Research). Fragmented RNA was prepared through 3′ end phosphate removal by treatment with T4 polynucleotide kinase (New England Biolabs), T4 RNA-ligase-based ligation of a 5′-adenylated universal miRNA cloning linker (New England Biolabs) to the RNA 3′ end, reverse transcription to cDNA using sequencing primers, and circLigase-based Illumina adapter ligation to the cDNA 3′ end as described[50].

The HiTRACE 2.0 package was used to analyse CE data, available as MATLAB toolbox at (https://github.com/hitrace)[51] and web server at (http://hitrace.org)[52]. Electrophoretic traces were aligned and baseline subtracted using linear and non-linear alignment routines as previously described[53]. Band intensities were fitted to Gaussian peaks, then background subtracted, signal attenuation corrected and normalized. MiSeq results were analysed using the MAPseeker software package[50,54]. The SHAPE reactivity profile was generated by projection of pair-ended reads. Briefly, a correction factor of 0.25 was used for full-length read ligation bias correction, accounting for the empirically observed low circLigase ligation efficiency for full-length cDNAs ending in CC[55]. Reads shorter than 8 nt were discarded due to ambiguity in their positional alignment in the full-length RNA. Modification fractions were determined by the number of cDNAs in which reverse transcription stopped at a given nucleotide divided by the number of cDNAs in which reverse transcription stopped or proceeded beyond that nucleotide[55]. Traces were background subtracted (SHAPE minus no-SHAPE data; and SHAPE-Tb minus Tb-alone data) and normalized through a boxplot-based heuristic[56] for *Hoxa5*, or using a scanning window of 350 nt for *Hoxa9*. Errors were estimated as the standard deviation between SHAPE/no-SHAPE and SHAPE-Tb/Tb-alone measurements.

Data-guided secondary structure models were obtained using the Fold executable of the RNAstructure package[57]. Pseudo-energy parameters of RNAstructure version 5.4 were used by default. To obtain 2D-data-guided secondary structure models, *Z* score matrices for mutate-and-map data sets were calculated and used as base-pair-wise pseudo-energies with a slope and intercept of 1.0 kcal mol^{-1} and 0 kcal mol^{-1} as previously described[27]. The ShapeKnots executable of RNA structure version 5.6 (ref. 58) enables modelling of pseudoknots but with different parameters for using SHAPE data in the modelling; this newer method was used for pseudoknot prediction of *Hoxa9* nt 957–1,132 when using 1D and 2D SHAPE data combined. Extension of ShapeKnots to accept 2D data are being incorporated into the next version of RNAstructure. Helix-wise confidence values were calculated via bootstrapping as described previously[27,28]. All predictions were

modelled in context with full-length sequences. Secondary structures were visualized by the VARNA software[59].

Mice. A 2.8-kb 5′ arm and a 4.9-kb 3′ arm of *Hoxa9* were cloned into PGKneolox2DTA.2 (P. Soriano, Addgene plasmid 13449). This results in a replacement of nt 945–1,194 of *Hoxa9* 5′ UTR with a *LoxP*-flanked Neo cassette in the same transcriptional direction as *Hoxa9* (Extended Data Fig. 8c). The plasmid was linearized with XhoI and electroporated into JM8NA3.N1 ESCs for homologous recombination by the UCSF ES cell core. The correctly targeted ESC clones were identified by long-range PCR and confirmed by Southern blotting. The clones were injected into C57BL/6 blastocysts by the Gladstone Institute Transgenic Core Facility. Resulting chimaeras were crossed with C57BL/6 mice to generate F1 ($Hoxa9^{Neo/+}$). $Hoxa9^{Neo/+}$ were crossed with CMV-Cre mice (Jackson Laboratory) to precisely remove the *LoxP*-flanked Neo cassette within the 5′ UTR to generate $HoxA9^{\Delta IRES/+}$ mice. All mice used for experiments were between 2 and 6 months old. All animal studies were performed in accordance with Stanford University Animal Care and Use guidelines.

Sucrose gradient fractionation. E11.5 neural tubes and somites were dissected and trypsinized as previously described[1]. The neural tube and somites of each embryo were lysed on ice in 200 µl of lysis buffer (20 mM Tris pH 7.5, 200 mM NaCl, 15 mM $MgCl_2$, 1 mM DTT, 8% glycerol, 1% Triton X-100, 100 µg ml^{-1} cycloheximide, 100 µg ml^{-1} heparin, 200 U ml^{-1} SUPERase In (Ambion)). After lysis, nuclei and membrane debris was removed by centrifugation (1,300*g*, 5 min at 4 °C, and then 14,000 rpm, 5 min, at 4 °C). The supernatant was layered onto a linear sucrose gradient (10–40% sucrose(w/v), 20 mM Tris, pH 7.5, 100 mM NaCl, 15 mM $MgCl_2$, 100 µg ml^{-1} cycloheximide) and centrifuged in a SW41Ti rotor (Beckman) for 2.5 h at 40,000 rpm at 4 °C. Fractions were collected by Density Gradient Fraction System (Brandel).

qPCR analysis. For cells, RNA was isolated using NucleoSpin RNA kit (Clontech) and DNase-treated on column. For embryos, E11.5 neural tubes and somites were dissected and trypsinized as previously described[1]. RNA was isolated using TRIzol (Invitrogen) and treated with TURBO DNA-free kit (Ambion). For sucrose gradient fractions, RNA was purified by acid phenol/chloroform followed by isopropanol precipitation. 0.5 µg of RNA was converted to cDNA using iScript supermix (BioRad). cDNA was diluted twofold and 1 µl used to run a SYBR green detection qPCR assay (SsoAdvanced SYBR Green supermix and CFX384, BioRad). Data was analysed and converted to relative RNA quantity using CFX manager (BioRad). For sucrose gradient fractions, amount of RNA from individual fractions was expressed as a fraction of the total RNA collected from all fractions. Primers were used at 250 nM per reaction. β-actin and *Hoxa9* qPCR primers were previously described[1]. Fluc F 5′-CATCACGTACGCGGAATACTT, Fluc R 5′-AAGAGATA CGCCCTGGTTC. 18S F 5′-ACATCCAAGGAAGGCAGCAG, 18S R 5′-CATTC CAATTACAGGGCCTC. 28S F 5′-GGGGAGAGGGTGTAAATCTC, 28S R 5- T CCTTATCCCGAAGTTACGG.

Whole mount *in situ* hybridization. Whole mount *in situ* hybridization was carried out using digoxygenin-labelled antisense RNA probes as previously described[1]. *Hoxa9* probe was amplified from mouse embryo cDNA. Primers sequences are described in Supplementary Table 1.

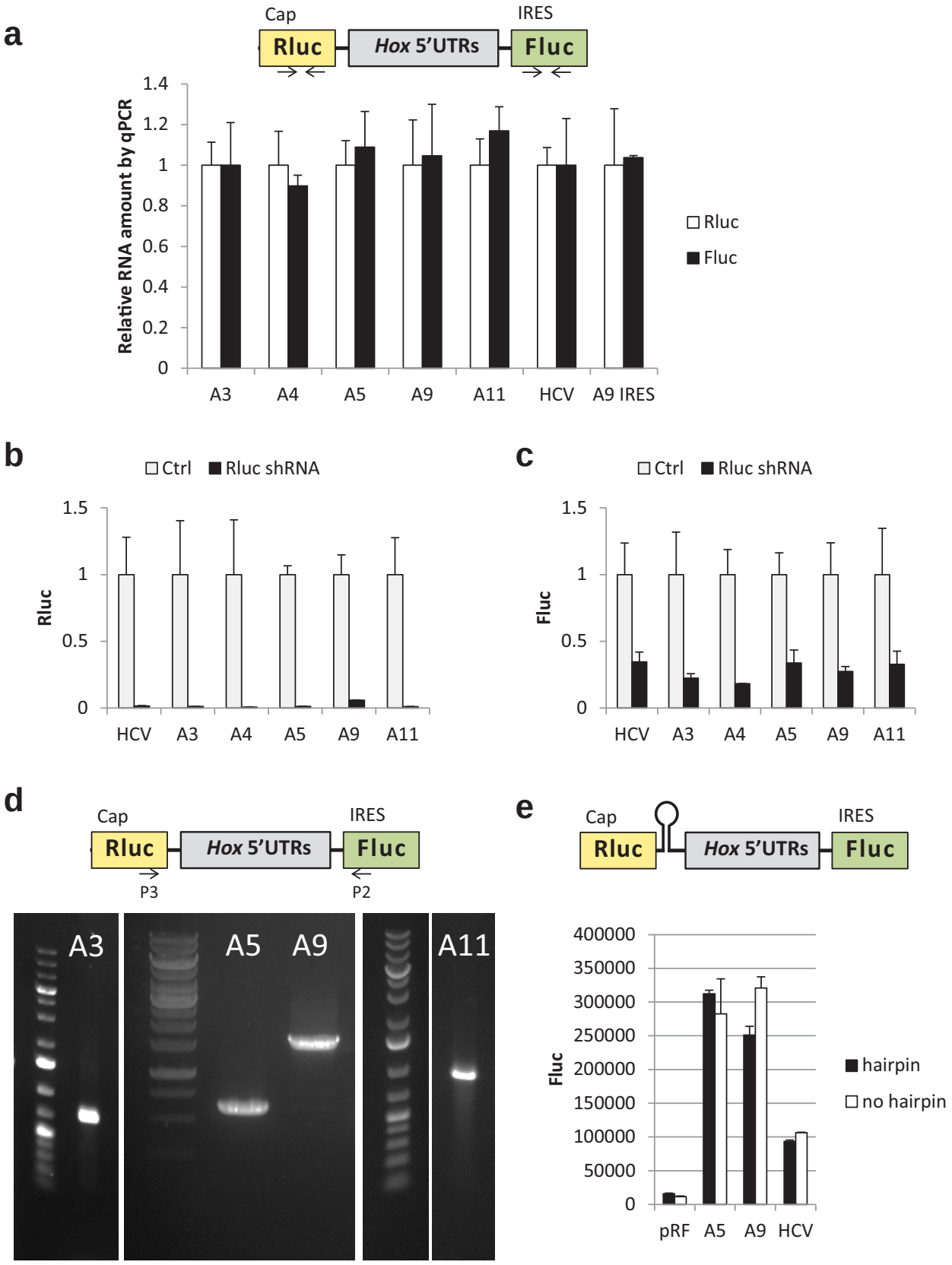
Skeletal staining. Alcian blue and alizarin red staining of cartilage and bone were performed on neonates as previously described[1].

Immunofluorescence. E11.5 embryos were dissected, fixed and cryosectioned as previously described[1]. WT and $Hoxa9^{\Delta IRES/\Delta IRES}$ littermates sections were placed onto the same slide for direct comparison and identical processing. Sections were equilibrated to room temperature and rehydrated in blocking buffer (1% heat-inactivated goat serum, 0.1% Triton X-100 in PBS) for 1 h at room temperature. Sections were incubated with anti-HOXA9 antibody (Abcam ab140631) at 1:1,000 in blocking buffer overnight at 4 °C, washed three times, 30 min each and incubated with Alexa 488 anti-rabbit secondary antibodies (Invitrogen) for 1 h at room temperature. Slides were washed with blocking buffer three times, 30 min each and incubated with 1 µg ml^{-1} DAPI (Cell Signaling) for 15 min at room temperature. They were washed again in blocking buffer for 20 min and mounted with VectaShield Mounting Media (Vector Laboratories).

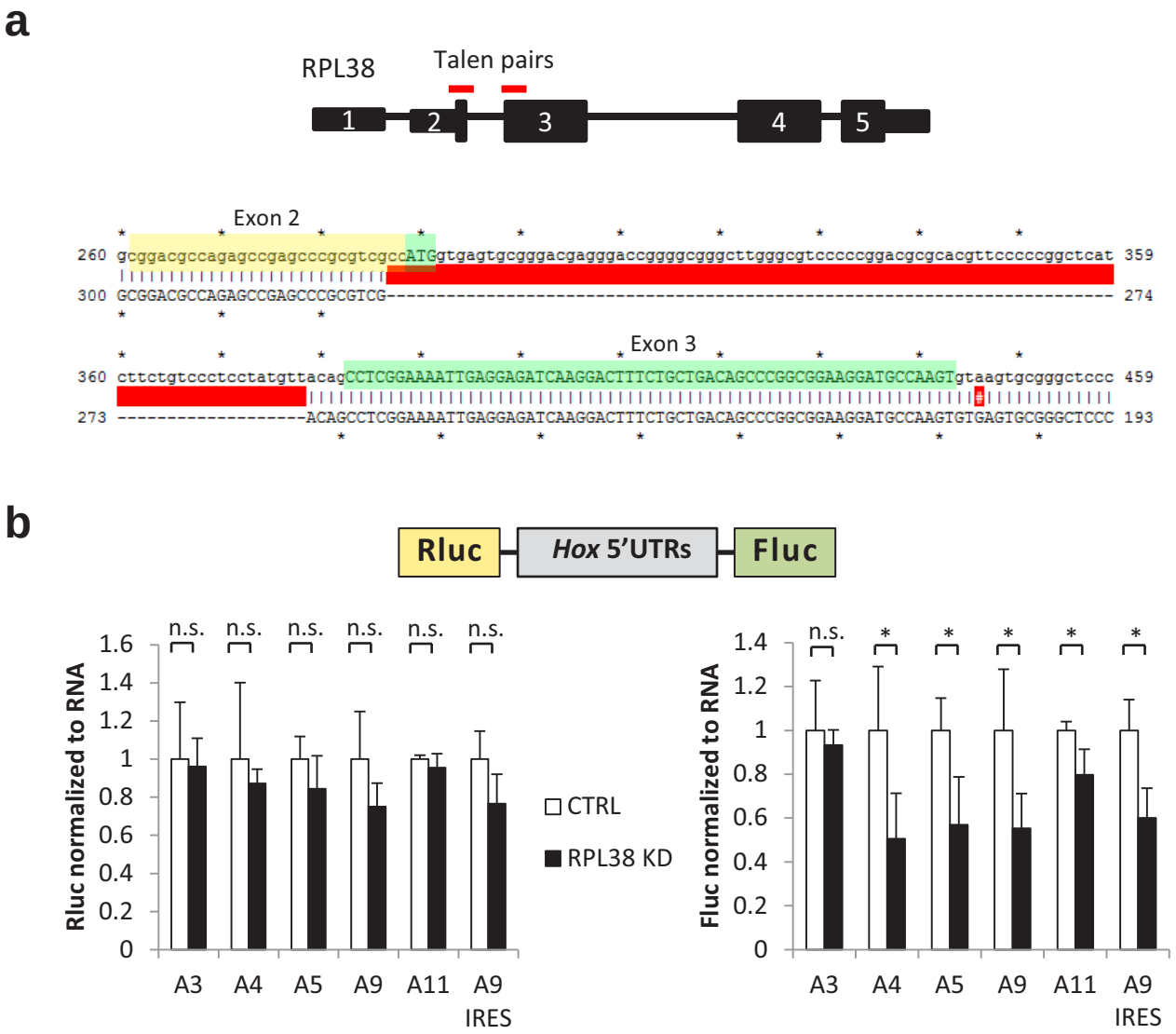
Statistical analysis. No normalization, blinding or exclusion was used in any experiments. No statistical method was used to predetermine sample size. In all cases, multiple independent experiments were performed on different days to verify the reproducibility of experimental findings. For mice experiments, embryos from multiple litters were used to avoid litter-specific bias.

Each variable was analysed using the two-tailed unpaired Student's *t* test. For all analyses, a *P* value of less than 0.05 was considered significant. Results are given as means ± s.d.

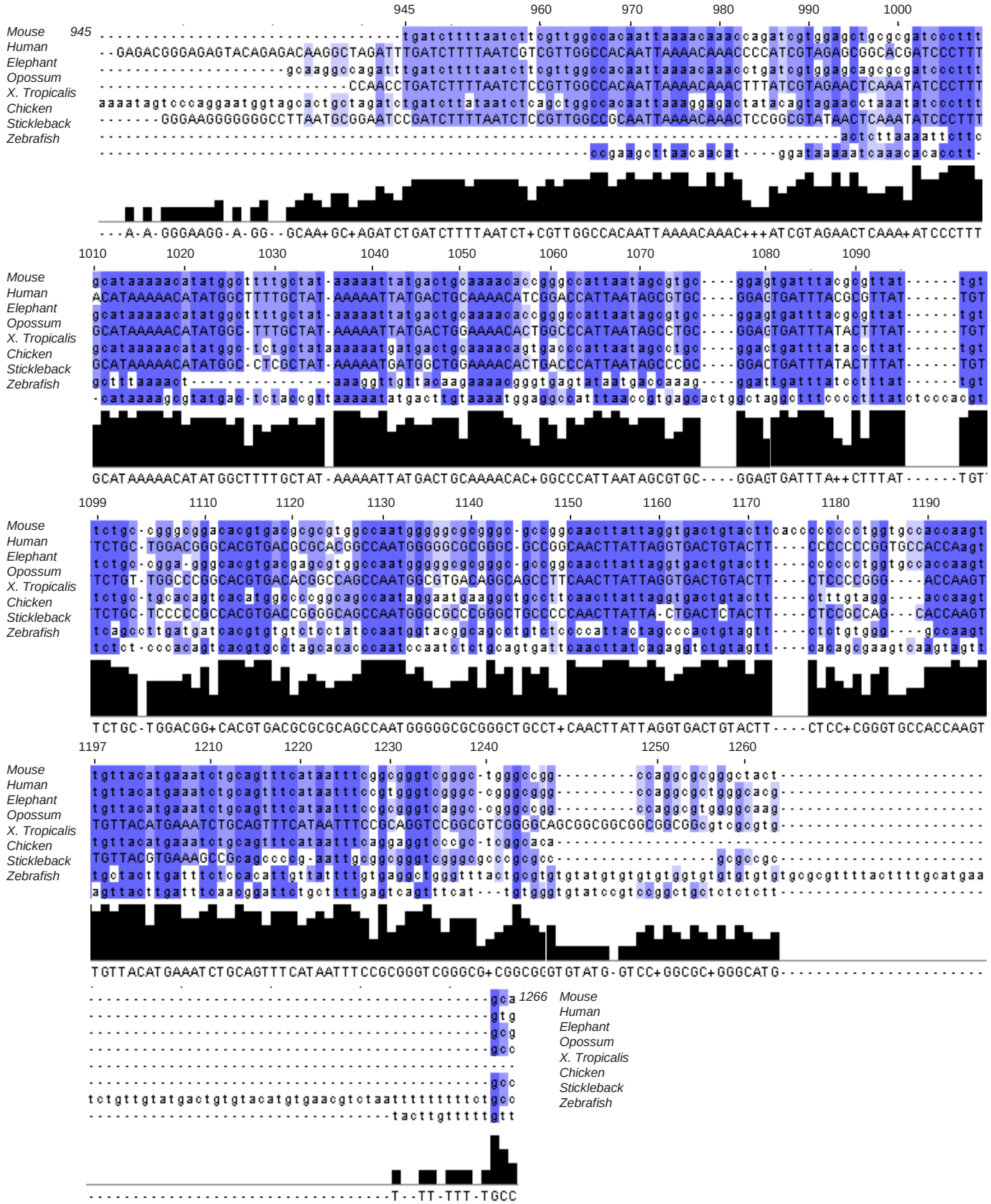
44. Flicek, P. *et al.* Ensembl 2014. *Nucleic Acids Res.* **42,** D749–D755 (2014).
45. Carlson, D. F. *et al.* Efficient TALEN-mediated gene knockout in livestock. *Proc. Natl Acad. Sci. USA* **109,** 17382–17387 (2012).
46. Kladwang, W., Cordero, P. & Das, R. A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA. *RNA* **17,** 522–534 (2011).
47. Mortimer, S. A. & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129,** 4144–4145 (2007).
48. Talkish, J., May, G., Lin, Y., WoolFord J. L. Jr & McManus, C. J. Mod-seq : high-throughput sequencing for chemical probing of RNA structure. *RNA* **20,** 713–730 (2014).
49. Harris, D. A. & Walter, N. G. Probing RNA structure and metal-binding sites using Terbium (III) footprinting. *Curr. Protoc. Nucleic Acid Chem.* http://dx.doi.org/10.1002/0471142700.nc0608s13 (2003).
50. Cheng, C., Chou, F. & Kladwang, W. MOHCA-seq : RNA 3D models from single multiplexed proximity-mapping experiments. Preprint at *bioRxiv* http://dx.doi.org/10.1101/004556 (2014).
51. Yoon, S. *et al.* HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* **27,** 1798–1805 (2011).
52. Kim, H., Cordero, P., Das, R. & Yoon, S. HiTRACE-Web: an online tool for robust analysis of high-throughput capillary electrophoresis. *Nucleic Acids Res.* **41,** W492–W498 (2013).
53. Kim, J. *et al.* A robust peak detection method for RNA structure inference by high-throughput contact mapping. *Bioinformatics* **25,** 1137–1144 (2009).
54. Seetin, M., Kladwang, W., Bida, J. & Das, R. in *RNA Folding: Methods and Protocols* (ed. Waldsich, C.) **1086,** 95–117 (Humana Press, 2014).
55. Kladwang, W. *et al.* Standardization of RNA chemical mapping experiments. *Biochemistry* **53,** 3063–3065 (2014).
56. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure. *Proc. Natl Acad. Sci. USA* **106,** 97–102 (2009).
57. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11,** 129 (2010).
58. Hajdin, C. E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl Acad. Sci. USA* **110,** 5498–5503 (2013).
59. Darty, K., Denise, A. & Ponty, Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25,** 1974–1975 (2009).



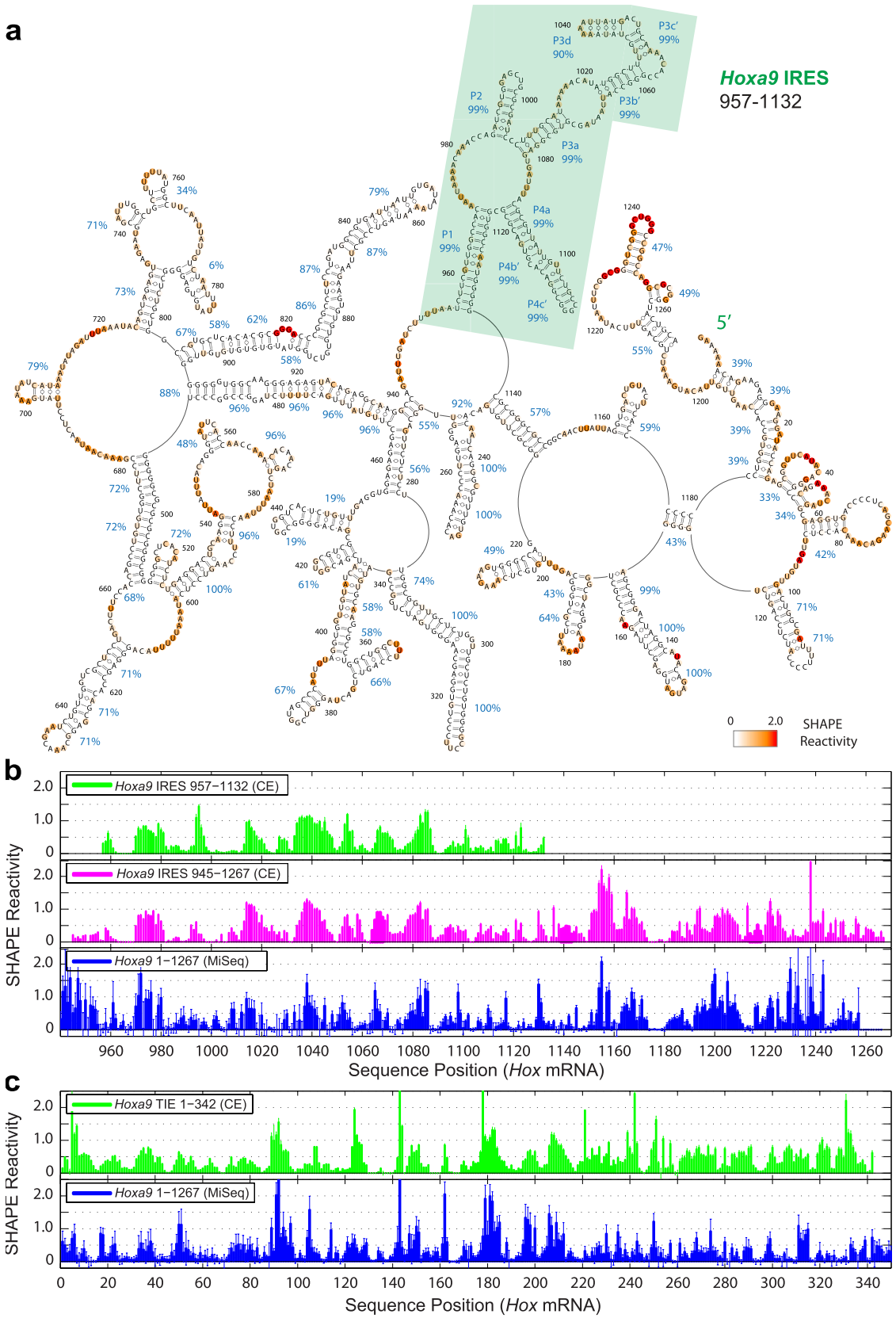
Extended Data Figure 1 | *HoxA* IRES controls confirming that Fluc activity from bicistronic vector is due to IRES activity. **a**, qPCR of both Rluc and Fluc from transfected cells shows that Rluc and Fluc are produced at the same ratio. All Rluc and Fluc values are normalized to that of HCV (set to 1). $n = 3$, individual experiments performed in duplicate. **b**, **c**, shRNA against Rluc decreased reporter activity of both Rluc (**a**) and Fluc (**b**), confirming that Rluc and Fluc were transcribed on the same mRNA. $n = 3$, individual experiments performed in triplicate. **d**, RT–PCR using primers in Rluc and Fluc show that there is no cryptic splice site in the cloned *Hox* 5′ UTR. Primer locations are shown as arrows in the diagram. **e**, Inserting a strong hairpin (-67 kcal mol^{-1}) after the Rluc reporter did not affect Fluc activity, suggesting that Fluc activity was not due to ribosome read-through.



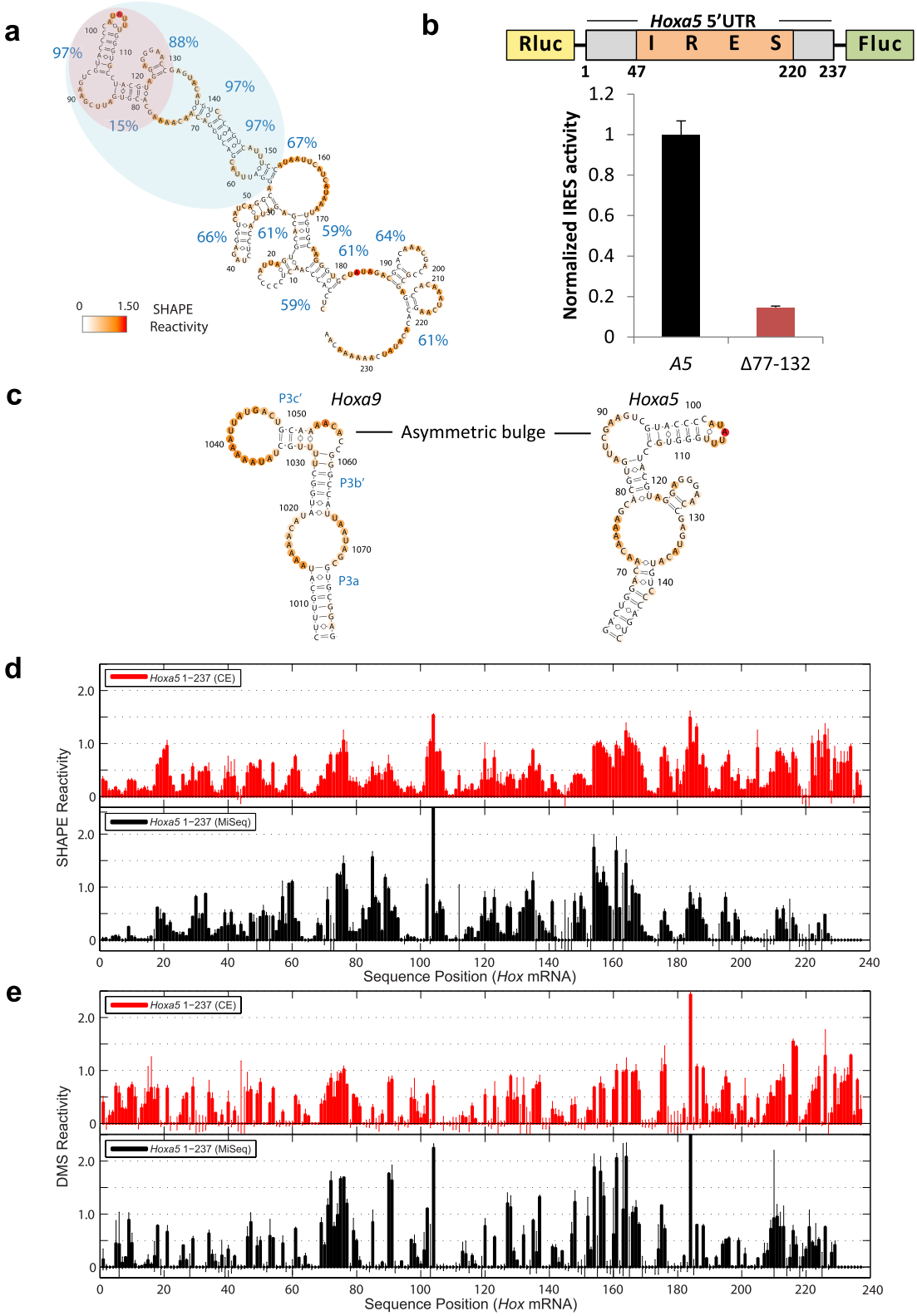
Extended Data Figure 2 | Disruption of *Rpl38* in C3H10T1/2 by TALEN nucleases. a, Location of TALEN pairs. Two pairs of TALENs were designed to bind at the end of exon 2 and the beginning of exon 3 to make a genomic break close to the ATG. Sequencing of a positive clone shows a deletion of the ATG and most of the intron after it. Coding sequence is highlighted in green. **b,** *Rpl38* knockdown does not change cap-dependent translation (Rluc) but decreases IRES-dependent translation (Fluc) from specific *Hox* 5′ UTRs. Luciferase activity was normalized to amount of Fluc RNA in the cells as quantified by qPCR. $*P < 0.05$ (*t*-test compared to control). $n = 2$, individual experiments performed in duplicate.



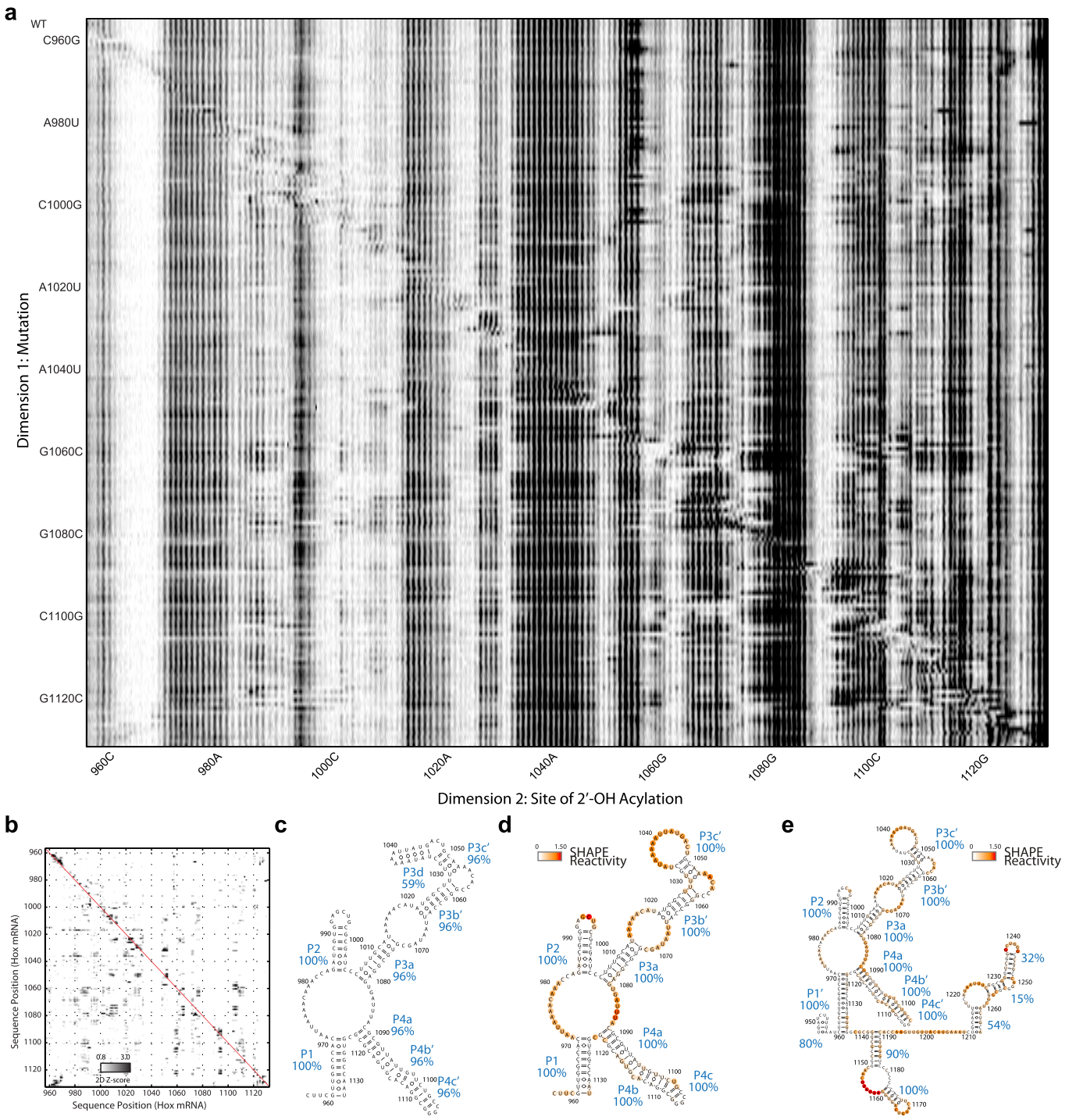
Extended Data Figure 3 | Alignment of the *Hoxa9* IRES element between vertebrate species. Nucleotides 945 to 1,266 of the mouse *Hoxa9* 5′ UTR were aligned with sequences from other vertebrates and show high sequence homology. Nucleotides are coloured based on their homology, with darker colours representing higher conservation.



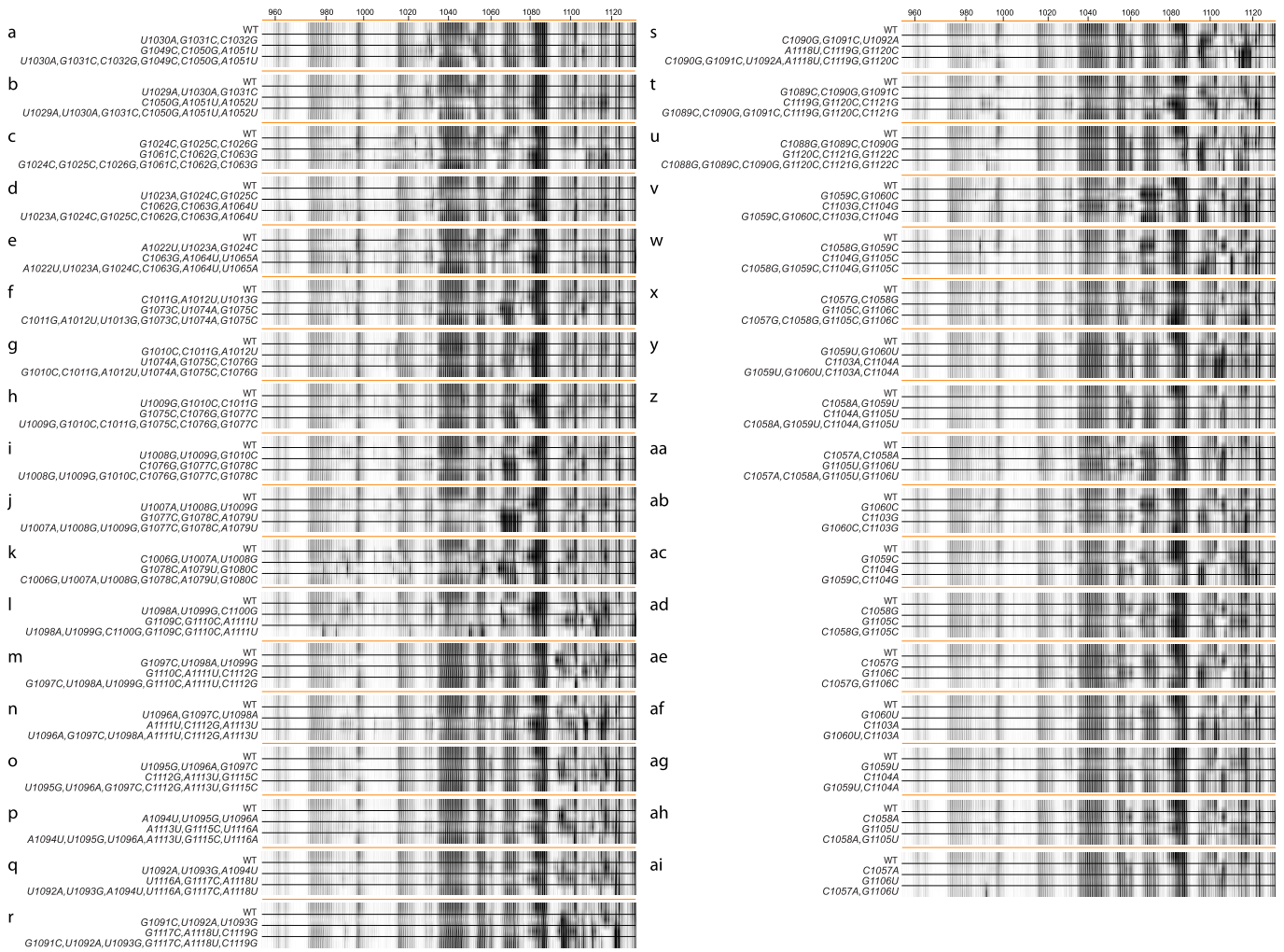
Extended Data Figure 4 | Chemical mapping and secondary structure prediction of full-length *Hoxa9* 5′ UTR. **a**, Secondary structure modelling of full-length *Hoxa9* using ligSHAPE data. The *Hoxa9* IRES element (nt 957–1,132, shaded in green) is modelled as the same secondary structure shown in Fig. 2a. Confidence values from bootstrapping (bulge percentages) exceed 90% for this element, suggesting a well-determined subdomain, but are generally low outside this region, indicating poor certainty in other regions. **b**, Normalized SHAPE reactivity of *Hoxa9* IRES (nt 957–1,132 and 944–1,266 from one-dimensional SHAPE read out through capillary electrophoresis (CE), full-length 1–1,266 from MiSeq-based ligSHAPE). **c**, Normalized SHAPE reactivity of *Hoxa9* TIE (nt 1–342 from CE-based one-dimensional SHAPE, full-length 1–1,266 from MiSeq-based ligSHAPE).



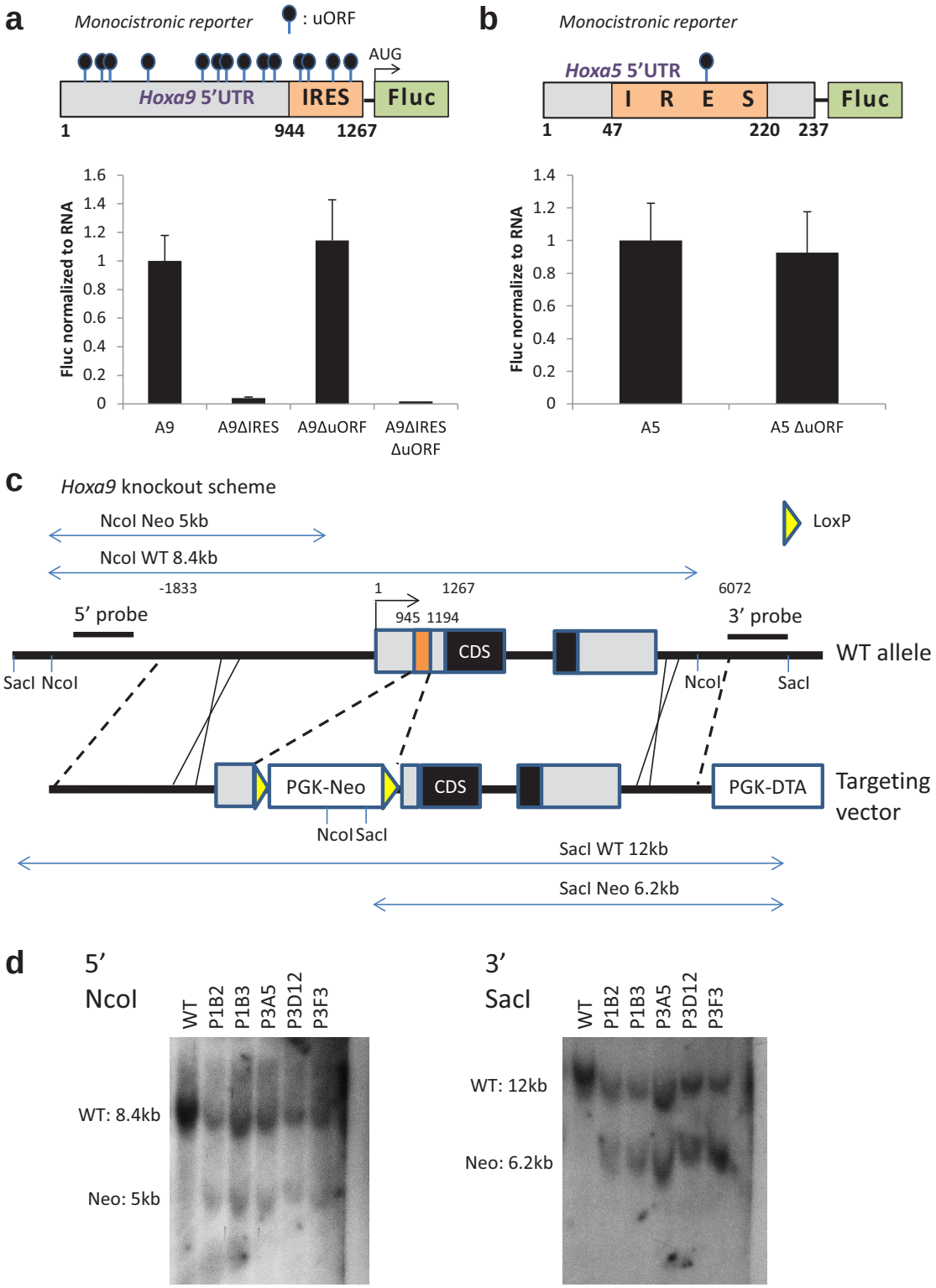
Extended Data Figure 5 | Chemical mapping and secondary structure model of full-length *Hoxa5* 5′ UTR. **a**, Secondary structure modelling of *Hoxa5* using one-dimensional SHAPE data. Nucleotides are coloured with SHAPE reactivities. Percentage labels give bootstrap support values for each helix. The feature highlighted in blue resembles P3 in *Hoxa9* and the tip highlighted in pink is deleted in **b**. **b**, The deletion of the tip identified in *Hoxa5* IRES structure shown in **a** decreases IRES activity in bicistronic reporter assays. IRES activity was normalized to full length *Hoxa5* 5′ UTR (*A5*, set to 1). $^{**}P < 0.01$ (*t*-test as compared to *A5*). $n = 2$ experiments, performed in triplicate. **c**, Both *Hoxa9* and *Hoxa5* contain an asymmetric bulge in a region important for IRES activity. **d**, **e**, Normalized SHAPE (**d**) and DMS (**e**) reactivity of *Hoxa5* (CE-based and MiSeq-based).



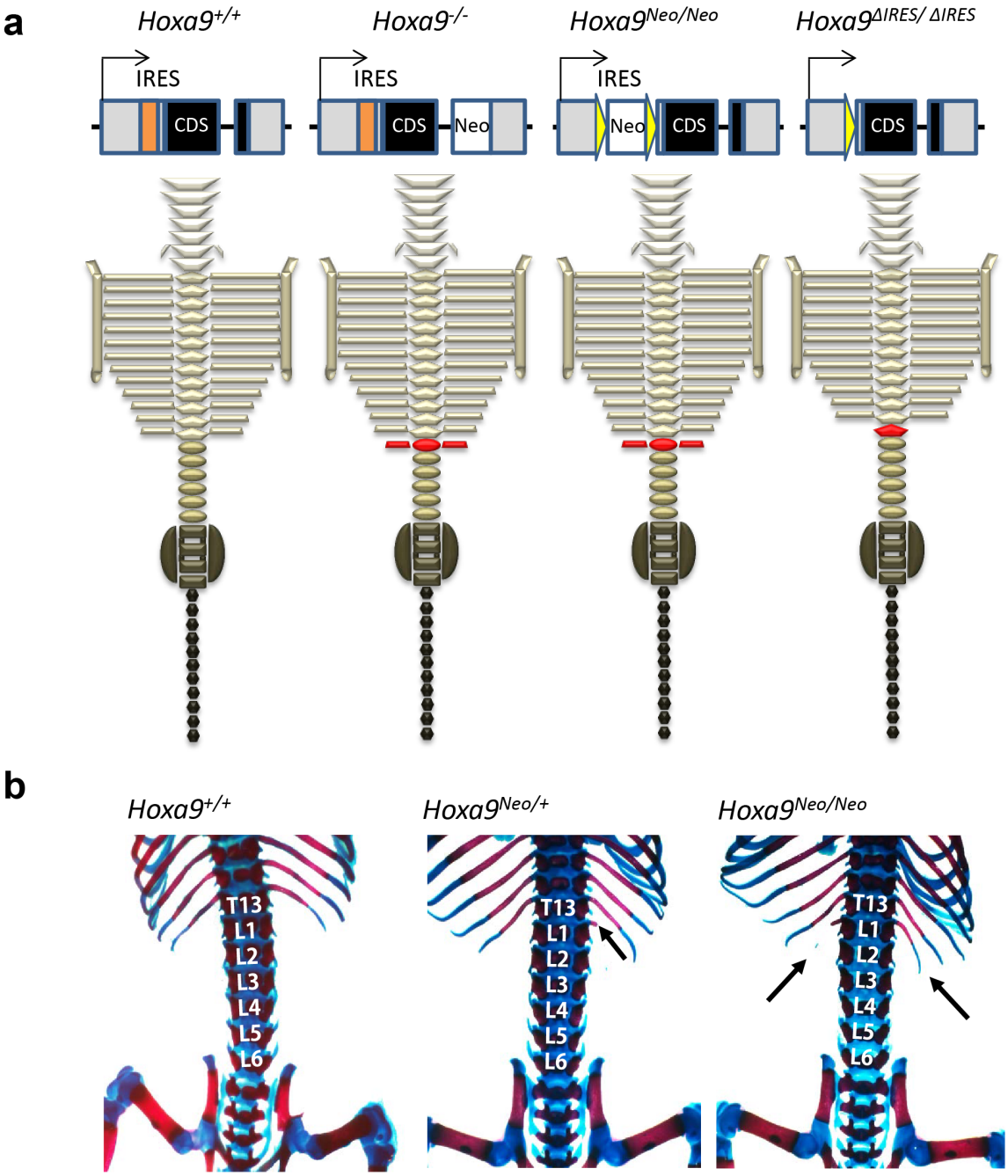
Extended Data Figure 6 | Secondary structure model and mutate-and-map (M^2) data set of *Hoxa9* IRES element. **a**, **b**, Entire M^2 data set and *Z*-score contact-map of *Hoxa9* nt 957–1,132 across 177 single mutants probed by 1M7. **c**, Secondary structure model of *Hoxa9* nucleotides 957–1,132 using M^2 data alone. **d**, Secondary structure model of *Hoxa9* nt 957–1,132 using one-dimensional SHAPE data alone. Nucleotides are coloured with SHAPE reactivity. **e**, Secondary structure model of *Hoxa9* nt 944–1,266 using one-dimensional SHAPE data. The models in **c**–**e** contain the same helices as the model from combined SHAPE/M^2 analysis in Fig. 2a, up to register shifts and edge base pairs; the small rearrangements are labelled P3b', P3c', P3d, P4b' and P4c'.



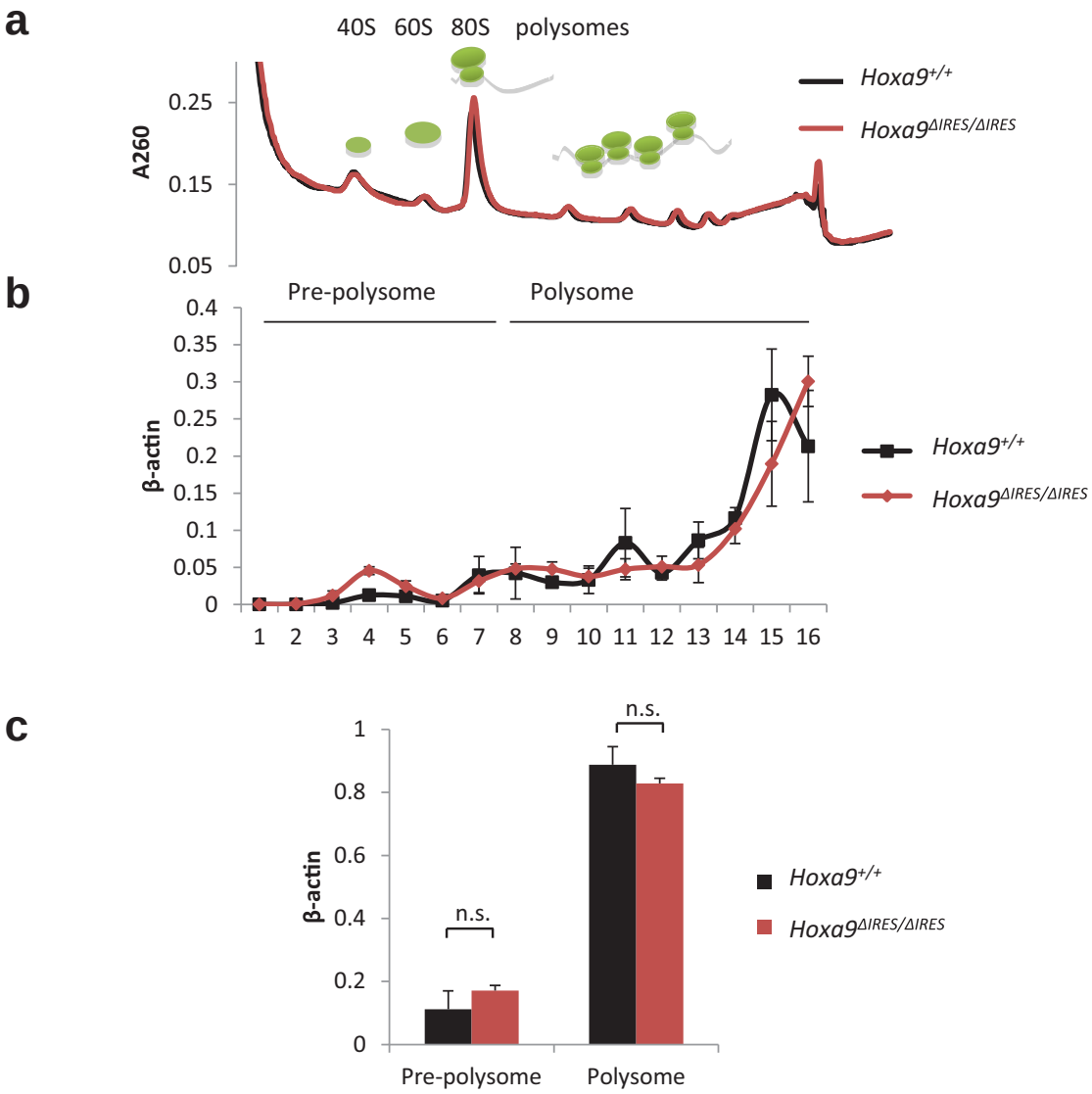
Extended Data Figure 7 | Mutation/rescue results of *Hoxa9* IRES structure (nt 944–1,266) probed by 1M7. Electropherograms of mutation/rescue to test base-pairings in P3c (**a, b**), P3b (**c**–**e**), P3a (**f**–**k**), P4b (**l**–**o**), P4a (**p**–**u**) and pk3-4 (**v**–**ai**). Perturbation of the chemical mapping reactivities by mutations of one strand and restoration by mutations in the other strand provide strong evidence for the tested pairings in P3c (**a**, **b**), P3b (**c**–**e**), P3a (**f**–**k**), P4b (**m**–**o**) and P4a (**p**–**q**, **t**). Near-perfect restoration by compensatory mutations in (**x**) and (**ad**) support pseudoknot pk3-4. Lack of rescue in other tested pairings is consistent with either absence of those pairings or higher-order structure (for example, base triples) interacting with those pairings.



Extended Data Figure 8 | Putative uORFs within the 5′ UTRs of *Hoxa9* and *Hoxa5* do not inhibit cap-dependent translation and *Hoxa9*$^{\Delta IRES}$ targeting strategy. uORFs are marked by black circles on the diagram of monocistronic reporter for the *Hoxa9* (**a**) and *Hoxa5* (**b**) 5′ UTR. All the ATGs in the 5′ UTR were mutated to TTG in *A9ΔuORF* construct and GTG in *A5ΔuORF*. The IRES element (944–1,266) was removed in *A9ΔIRES* construct. The IRES element was removed from the *A9ΔuORF* construct in *A9ΔIRESΔuORF*. $n = 3$ individual experiments in duplicates. Data represent mean ± s.d. **c**, Diagrams of the *Hoxa9* locus and the targeting vector. Boxes represent exons, grey boxes represent UTRs, and black boxes represent the coding sequence. Nucleotides 944–1,145 were replaced by a floxed Neo cassette in the targeting vector. Locations of Southern blot probes, restriction enzymes used for Southern analysis and expected sizes are marked on the diagrams. **d**, Southern blot analysis of targeted cells using both the 5′ and 3′ probes showing that both arms integrated correctly into the *Hoxa9* locus. Mice were generated from clone P3A5.



Extended Data Figure 9 | The presence of a Neo cassette in the *Hoxa9* locus is linked to the presence of an L1 → T13 homeotic transformation. **a**, Diagram of the *Hoxa9* locus (top) and axial skeleton phenotype (bottom) in different *Hoxa9* mouse mutants. The original $Hoxa9^{-/-}$ was made by replacing the homeodomain with a Neo cassette. Vertebra with homeotic transformation is coloured red. **b**, Representative skeletons of $Hoxa9^{+/+}$, $Hoxa9^{Neo/+}$ and $Hoxa9^{Neo/Neo}$. Arrows point to the additional rib(s) on L1, revealing a homeotic transformation to T13. These results show that it is the presence of Neo in the targeting locus, which may affect the expression of neighbouring Hox gene, that is sufficient to cause the L1 → T13 phenotype. When the Neo cassette is removed from the targeting locus by crossing the $Hoxa9^{Neo/+}$ mouse with a CMV Cre line, the L1 → T13 phenotype is no longer present. $n = 3$ skeletons of each genotype.



Extended Data Figure 10 | Sucrose gradient fractionation shows no difference in β-actin association with polysomes in *Hoxa9*$^{+/+}$ and *Hoxa9*$^{\Delta IRES/\Delta IRES}$ embryos. **a**, Overlay of A_{260} trace during fractionation showing no difference in polysome profiles between E11.5 *Hoxa9*$^{+/+}$ and *Hoxa9*$^{\Delta IRES/\Delta IRES}$ embryos. **b**, qPCR from each fraction reveals no difference in β-actin mRNA accumulation between *Hoxa9*$^{+/+}$ and *Hoxa9*$^{\Delta IRES/\Delta IRES}$ embryos. **c**, Quantification of β-actin mRNA in fractions. Fractions 1–8 are pre-polysomes and 9–16 are polysome fractions. $n = 3$ embryos of each genotype.