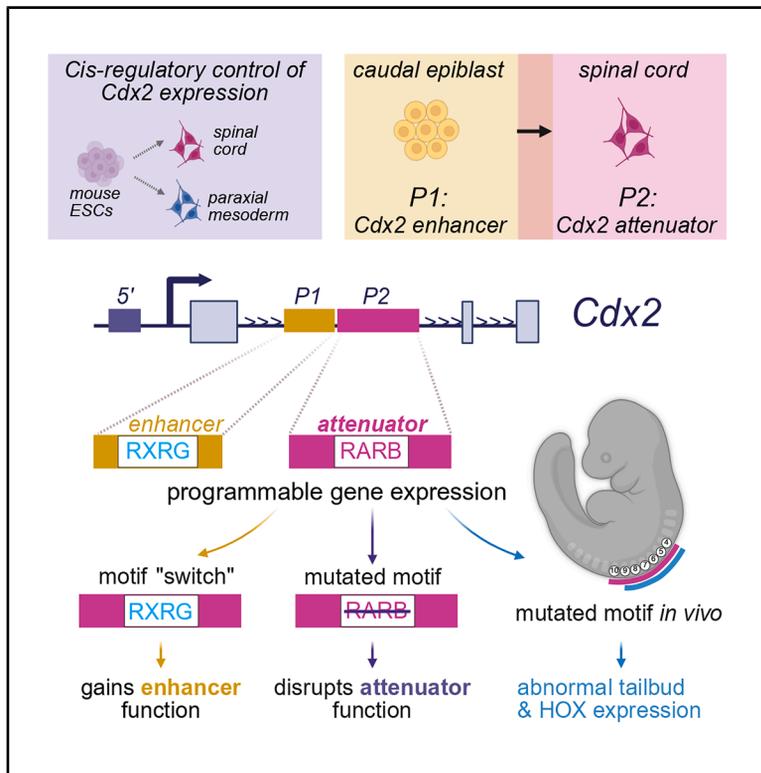


Developmental Cell

A dual enhancer-attenuator element ensures transient *Cdx2* expression during mouse posterior body formation

Graphical abstract



Authors

Irène Amblard, Damir Baranasic, Sheila Q. Xie, Benjamin Moyon, Michelle Percharde, Boris Lenhard, Vicki Metzis

Correspondence

v.metzis@lms.mrc.ac.uk

In brief

Amblard et al. dissect the function of *cis*-regulatory elements regulating transient *Cdx2* expression during mouse caudal body formation. They highlight the requirement of an attenuator, a transiently repressive element, which can be converted into an enhancer through a single nuclear receptor motif substitution.

Highlights

- Transient *Cdx2* expression depends on a dual enhancer-attenuator element
- A single nuclear receptor motif substitution converts an attenuator into an enhancer
- The *Cdx2* attenuator is required *in vivo* for mouse posterior body formation



Short article

A dual enhancer-attenuator element ensures transient *Cdx2* expression during mouse posterior body formation

Irène Amblard,¹ Damir Baranasic,^{1,2,3} Sheila Q. Xie,² Benjamin Moyon,² Michelle Percharde,^{1,2} Boris Lenhard,^{1,2} and Vicki Metzis^{1,2,4,*}

¹Institute of Clinical Sciences, Faculty of Medicine, Imperial College London, London W12 0HS, UK

²Medical Research Council (MRC) Laboratory of Medical Sciences (LMS), London W12 0HS, UK

³Division of Electronics, Ruder Boskovic Institute, Bijenicka cesta 54, 10000 Zagreb, Croatia

⁴Lead contact

*Correspondence: v.metzis@lms.mrc.ac.uk

<https://doi.org/10.1016/j.devcel.2025.06.006>

SUMMARY

During development, cells express precise gene expression programs to assemble the trunk of the body plan. Appropriate control over the duration of the transcription factor *Cdx2* is critical to achieve this outcome, yet how cells control the onset, maintenance, or termination of *Cdx2* has remained unclear. Here, we delineate the *cis*-regulatory logic orchestrating dynamic *Cdx2* expression in mouse caudal epiblast progenitors and their derivatives. Combining CRISPR-mediated deletion of regulatory elements with *in vitro* models and *in vivo* validation, we demonstrate that distinct enhancers and a repressive element embedded at the *Cdx2* locus act sequentially to drive transient *Cdx2* expression. We pinpoint an “attenuator”: a minimal region relying on a nuclear receptor to extinguish *Cdx2*. Changing this single motif converts the attenuator to an enhancer with the opposite regulatory behavior. Our findings establish a dual *cis*-regulatory logic ensuring precise spatiotemporal control over gene expression for vertebrate body patterning.

INTRODUCTION

In the mammalian body, a striking array of cell types emerges during development in response to extrinsic cues. This immense diversity results from the activity of gene regulatory networks that define cell identity.¹ Yet, how cells interpret extrinsic signaling in a context-dependent manner to ensure the generation of different cell types remains a major open question.

CDX (Caudal Type Homeobox; CDX1, 2, and 4) transcription factors (TFs) play a central role in the development of the caudal part of the body plan.^{2–5} Removal of these partially redundant factors^{6,7} results in the loss of most post-occipital tissues, in part due to their regulation of homeobox (*Hox*) genes.^{3–5,8–10} Reduced or prolonged expression of CDX factors respectively truncates or expands the territory that forms the spinal cord (SC), at the expense of hindbrain fates, in multiple species.^{6,11–15} Unlike *Cdx1* and *Cdx4*, however, genetic removal of *Cdx2* alone demonstrates its indispensable role in posterior body formation.^{3,6–8,16,17}

The expression of *Cdx2* is detected during gastrulation in the mouse caudal epiblast (CEpi).^{18–20} This region of the embryo harbors neuromesodermal progenitors (NMPs), a progenitor source that contributes to the developing SC and somites.^{21–23} Although *Cdx2* is detected in NMPs^{24–26} and is later maintained

in the hindgut,^{18,26} it is only transiently expressed in tissues derived from NMPs, such as the SC and somites.^{3,9,26,27} WNT and fibroblast growth factor (FGF) signaling promote caudal embryo development and the expression of *Cdx2*.^{12,13,28,29} Similar regulation is observed *in vitro* using the directed differentiation of mouse or human embryonic stem cells (ESCs).^{30–34} In addition, retinoic acid (RA) signaling is a critical determinant of posterior body formation and differentiation^{35,36} and restricts the expression of *Cdx2* in the SC *in vivo*²⁷ and *in vitro*.^{24,37} The transition from a CEpi to an SC progenitor coincides with a switch from FGF to RA signaling.^{35,36} How cells coordinate the duration of *Cdx2* expression in response to extrinsic cues remains unresolved.

Extrinsic cues are interpreted by *cis*-regulatory elements (CREs) to control gene expression. CREs encompass a broad range of elements that include promoters, enhancers,³⁸ silencers,³⁹ and insulators.⁴⁰ Recent findings suggest that enhancers may function in concert with additional elements, such as tethers⁴¹ and, more recently, facilitators,⁴² yet how common such elements are in the genome remains unclear. Although many classes of CREs exist, the ability to predict the cellular context and function of individual CREs remains challenging.⁴³ Several *Cdx2* enhancers have been identified that partially recapitulate the trophectoderm, CEpi, or intestinal expression



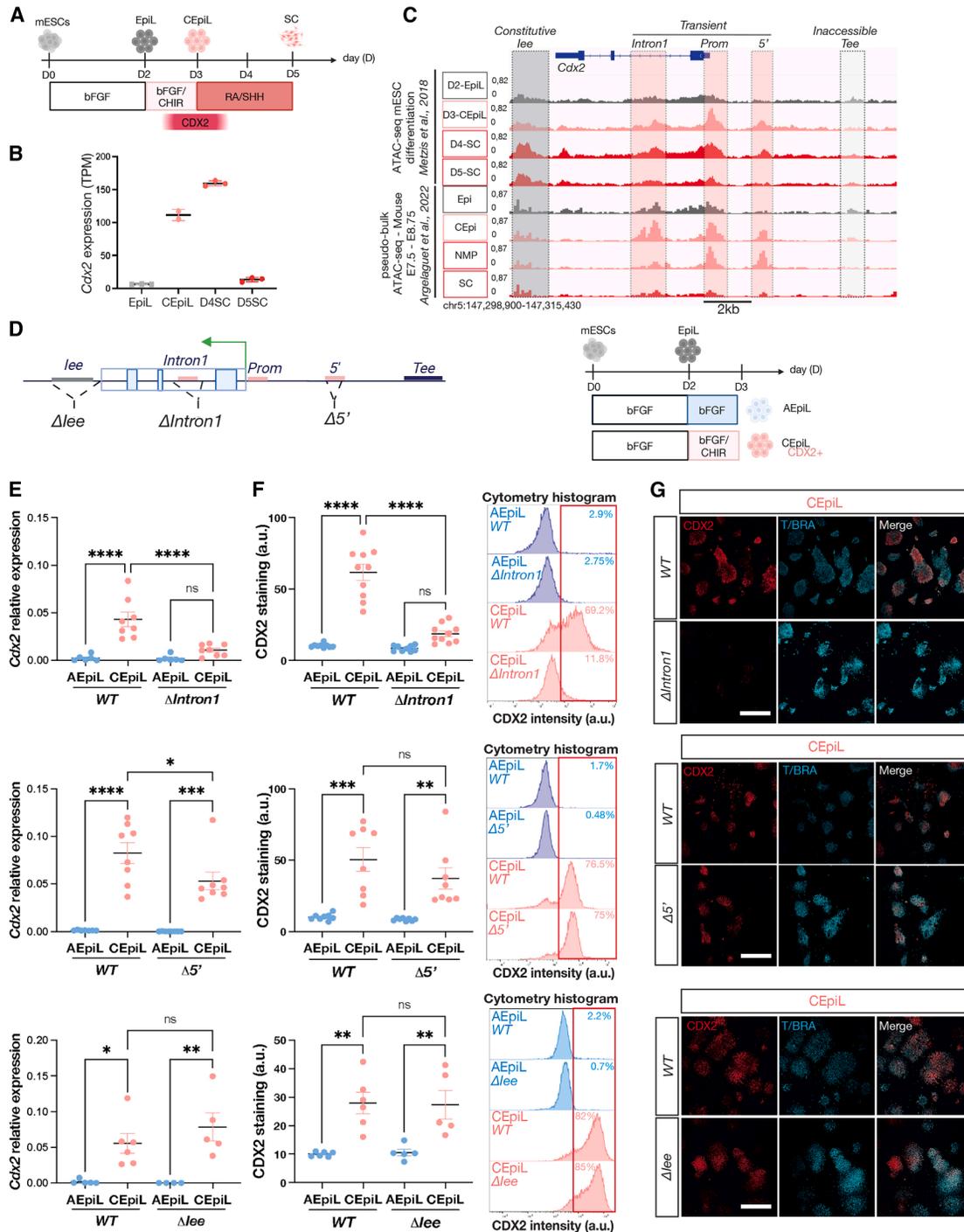


Figure 1. An intronic enhancer is indispensable for the onset of *Cdx2* in CEpi conditions

(A) Simplified schematic of the 5-day SC progenitor differentiation from mESCs.

(B) Reprocessed mRNA-seq data demonstrate *Cdx2* expression is transiently induced.³¹

(C) Bulk ATAC-seq signal at the *Cdx2* locus between days 2 and 5⁵⁵ and single-nucleus pseudobulk ATAC-seq signals at the *Cdx2* locus in indicated *in vivo* cell types.⁵⁶ Pink shading highlights the promoter (*Prom*), *Intron1*, and *5'* elements. The intestinal (*lee*) and trophectoderm enhancer (*Tee*) are highlighted in dark and light gray shading, respectively.

(D) Schematic of the *Cdx2* regulatory elements targeted to remove the *lee* (Δlee), first intron ($\Delta Intron1$) or *5'* element ($\Delta 5'$) and conditions used to induce *Cdx2* *in vitro* in CEpiL cells versus AEpiL cells used as a control.

(E) Relative expression (RT-qPCR) for *Cdx2* in AEpiL (blue) versus CEpiL (pink) conditions collected from $\Delta Intron1$, $\Delta 5'$, and Δlee showing *Cdx2* is not induced in $\Delta Intron1$ cells, while $\Delta 5'$ cells display a slight decrease.

(legend continued on next page)

pattern of *Cdx2* *in vivo*.^{19,44–48} These studies suggest that a subset of *Cdx2* CREs play tissue-specific roles. Although long-range interactions can play a vital role in regulating developmental genes,^{49–52} the expression of *Cdx2* during posterior body formation appears to be regulated by elements located within the *Cdx2* locus. In particular, an 11 kb region flanking *Cdx2* recapitulates the caudal tailbud expression pattern of *Cdx2* between embryonic day (E)7.5 and E10.5.⁵³ Strikingly, several CREs located within this region demonstrate enhancer activity in transgenic reporter embryos, yet, as individual elements, they do not recapitulate the full expression pattern of *Cdx2*.^{46,53} How multiple CREs within their native genomic context facilitate *Cdx2* initiation, maintenance, or termination remains unclear.

In this study, we investigate the regulatory mechanisms responsible for controlling the duration of *Cdx2* during the formation of posterior body derivatives: SC and paraxial presomitic mesoderm (PSM) progenitors. To dissect the molecular mechanisms that control *Cdx2*, without compromising axial elongation or trophoderm specification,⁵⁴ we used genome engineering approaches, combined with the directed differentiation of pluripotent ESCs, to model posterior body formation *in vitro*. Using this strategy, we provide evidence that the transient expression of *Cdx2* in cells relies on the sequential usage of CREs that perform discrete, nonredundant functions during development. We identify the location of a CRE that limits the duration of *Cdx2*. We demonstrate that this repressive element can be converted into an enhancer through a single RA nuclear receptor motif substitution. Furthermore, we validate that its function is critically dependent on the presence of the motif *in vitro* and during caudal body development *in vivo*. Taken together, we provide evidence that the composition and number of RA nuclear receptor motifs dictate regulatory element function and underpin the context-specific regulation of *Cdx2* during posterior body development.

RESULTS

Cdx2 CREs display transient accessibility during SC development

To identify putative CREs regulating *Cdx2* expression during posterior body formation, we examined the chromatin accessibility landscape inferred from assay for transposase-accessible chromatin with sequencing (ATAC-seq) experiments using two different approaches. We complemented an *in vitro* time course of bulk ATAC-seq data from mouse ESCs differentiated into SC progenitors, which transiently express *Cdx2*⁵⁵ (Figures 1A–1C), with pseudobulk ATAC-seq profiles obtained from the corresponding cell types present *in vivo*, extracted from 10× multiome single-nucleus (sn)ATAC-seq performed on E7.5–E8.75 whole embryos⁵⁶ (Figure 1C). The transient expression of *Cdx2* *in vitro* corresponds to CEpi-like (CEpiL) cells, which, upon differentiation to SC, lose *Cdx2*^{31,55,57} (Figures 1A and 1B). As genome-wide changes in chromatin accessibility take place in CEpiL versus

SC progenitors,^{55,57,58} we hypothesized that changes in the availability of CREs may contribute to the regulation of *Cdx2*.

Consistent with this, known *Cdx2* CREs^{19,44–47,53,59} (Figure S1A) exhibited distinct patterns of chromatin accessibility (Figures 1C and S1B). In particular, a set of three CREs were transiently accessible *in vitro* when comparing CEpiL (*Cdx2* positive; Figures 1B, 1C, and S1B) to SC progenitors (*Cdx2* lacking; Figures 1B, 1C, and S1B). These corresponded to the *Cdx2* promoter, an *Intron1* CRE,^{19,53} and a CRE located upstream of the transcriptional start site of *Cdx2*,⁵³ termed 5' (Figure 1C, “transient,” boxed in pink). Similarly, *in vivo*, these regions are accessible in the CEpi (Figure 1C) but appear inaccessible in SC progenitors (Figure 1C). Both the *Intron1* and 5' CRE exhibit enhancer activity in posterior tissues, although their onset and specificity differ from *Cdx2*.⁵³ By contrast, CREs that regulate *Cdx2* in other tissues such as the intestine (*lee*)^{46,47} or trophoderm enhancer (*Tee*)^{44,45} appear either continuously accessible (“constitutive,” Figure 1C; box shaded in gray) or largely lacking accessibility, respectively, in the same cellular conditions examined *in vitro* and *in vivo* (“inaccessible,” Figure 1C; white box). In summary, we show that a defined set of *Cdx2* CREs are transiently accessible at the time *Cdx2* is expressed in an *in vitro* model of SC development.

Separate CREs control the onset versus maintenance of *Cdx2* expression

Having established that a distinct set of *Cdx2* CREs are transiently accessible, we set out to test directly the function of each CRE on the regulation of *Cdx2* during posterior body formation using a previously established ESC *in vitro* system to model SC or paraxial mesoderm development.^{24,57} We engineered a suite of ESC lines that lacked individual CREs corresponding to either *transient*, *constitutive*, or *inaccessible* regions using CRISPR-Cas9-mediated genome editing (Figures 1D and S1C). Wild-type (WT) versus CRISPR mutant ESC lines were then differentiated toward CEpiL cells that express *Cdx2* in response to a brief pulse of basic fibroblast growth factor (bFGF) and the glycogen synthase kinase-3 inhibitor, CHIR99021 (CHIR).^{24,31,55,60} Anterior epiblast-like (AEpiL) cells, which do not express *Cdx2*, were used as a control and induced by exposure to bFGF alone^{31,32,55,57} (Figures 1D and S1C). *Cdx2* induction was assayed by RT-qPCR and immunofluorescence (IF), together with flow cytometry, to investigate CDX2 in a quantitative and single-cell manner. The expression of CDX2 remained indistinguishable between WT cells and cells lacking either the 5', the *lee* (Figures 1E–1G), or the *Tee* (Figures S1D–S1F) CRE (referred to as $\Delta 5'$, Δlee , and ΔTee cells, respectively). By contrast, removal of the *Intron1* CRE severely impaired the induction of *Cdx2*, as demonstrated at the transcript (Figure 1E) and protein level (Figures 1F and 1G). Impaired induction of CDX2 was also recapitulated by the removal of the promoter for *Cdx2* (Figures S1D–S1F). This demonstrates that the

(F) Flow cytometry quantification of CDX2 levels in AEpiL (blue) versus CEpiL (pink) conditions collected from $\Delta Intron1$, $\Delta 5'$, and Δlee cells showing CDX2 is not induced in $\Delta Intron1$ cells and representative flow cytometry histograms for CDX2 indicating the proportion of CDX2-positive cells.

(G) IF demonstrates that $\Delta Intron1$ cells maintain expression of TBXT (T/BRA) in CEpiL conditions despite loss of CDX2. Scale bar, 500 μ m. $n = 3$.

(A) and (D) created with BioRender.com. Data are represented as mean \pm SEM. CEpiL, caudal epiblast-like; EpiL, epiblast-like; mESCs, mouse embryonic stem cells; SC, spinal cord; TPM, transcripts per million.

* p value < 0.05, ** p value < 0.01, *** p value < 0.001, **** p value < 0.0001.

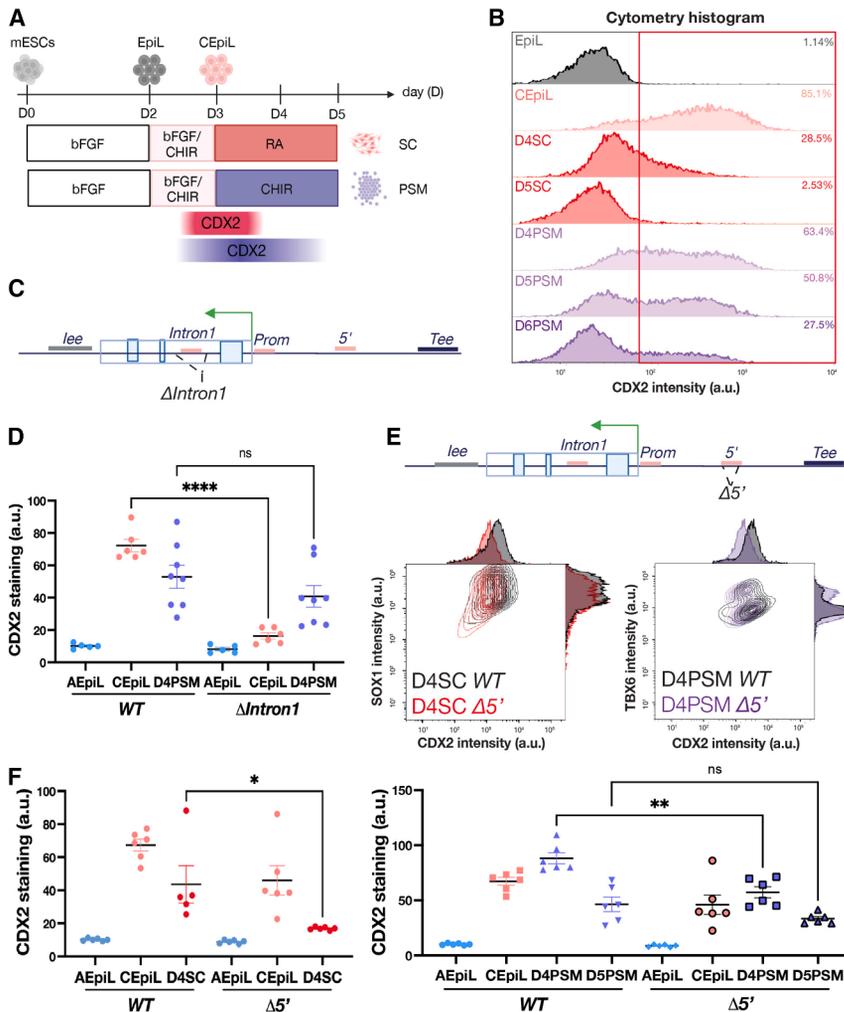


Figure 2. *Cdx2* is transiently maintained in CEpi derivatives via an alternative 5' CRE

(A) Schematic of the SC or PSM progenitor differentiation highlighting the transient expression of *Cdx2*.

(B) Representative flow cytometry histogram for CDX2 in WT cells showing CDX2 detected in PSM conditions until day 5, versus day 4 in SC.

(C) Schematic of the CRISPR-targeted element lacking in Δ Intron1 ESCs.

(D) Flow cytometry for CDX2 demonstrates WT and Δ Intron1 cells display no differences by day 4 in PSM conditions.

(E) Schematic of the CRISPR-targeted element lacking in Δ 5' ESCs and representative flow cytometry contour and histogram plot for CDX2 and SOX1 or TBX6 in indicated conditions from WT (black) and Δ 5' (red and purple) cells.

(F) CDX2 levels assessed by flow cytometry show a significant decrease of CDX2 in day 4 SC (red) and in day 4 PSM (purple) Δ 5' cells.

(A), (C), and (E) created with BioRender.com. Data are represented as mean \pm SEM. AEpiL, anterior epiblast-like; CEpiL, caudal epiblast-like; EpiL, epiblast-like; PSM, presomitic mesoderm; SC, spinal cord. **p* value < 0.05, ***p* value < 0.01, *****p* value < 0.0001.

cells displayed a reduction in CDX2 expression at day 4 in both SC and PSM conditions, relative to WT cells (Figures 2E and 2F). RT-qPCR primers designed to detect spliced versus nascent *Cdx2* transcripts confirmed that a reduction in *Cdx2* was detectable at the level of transcription (Figure S2C). This suggests that *Cdx2* expression in derivatives of the CEpiL cells is dependent on the 5' CRE, and, upon its removal, *Cdx2* is rapidly downregulated, decreasing the duration of expression (Figure S2C). Taken together, transient chromatin accessibility changes can be used to predict key regulatory elements that play nonredundant roles in the onset of *Cdx2* expression in CEpiL cells (*Intron1*) versus SC or paraxial mesoderm (5') progenitors.

intronic CRE plays an indispensable role in the induction of *Cdx2* in CEpiL cells. These data demonstrate that the removal of the *intron1* CRE is sufficient to block *Cdx2* induction in CEpiL cells, despite the presence of several alternative and accessible regulatory regions at the *Cdx2* locus, such as the *lee* or the 5' CRE (Figure 1C).

Having established that *Intron1* is indispensable for *Cdx2* induction in CEpiL cells, we tested whether *Intron1* was required for *Cdx2* expression in alternative cell types. To assess this, we differentiated ESCs under paraxial PSM conditions (Figure 2A), in which CDX2 is sustained for a longer time period (days 3–5) relative to SC progenitors (days 3–4) (Figures 2A and 2B). Despite an initial loss of *Cdx2* in CEpiL cells lacking *Intron1*, the expression of CDX2 is recovered at day 4 to levels comparable to WT cells in PSM conditions (Figures 2C and 2D; day 4 PSM condition). These data demonstrate that *Intron1* plays a cell-type-specific role in regulating *Cdx2* and suggest that an alternative CRE role may be responsible for *Cdx2* expression in PSM conditions.

The 5' CRE demonstrates transient accessibility and is also preferentially bound by CDX2 in CEpiL cells (Figure S2D). Under CEpiL conditions, ESCs lacking the 5' CRE (Figure 2E) expressed CDX2 at comparable levels to WT cells (Figure 2F). However, Δ 5'

The duration of *Cdx2* is regulated by separate subregions within the intronic CRE

To investigate what determines CRE usage in different cellular conditions, we sought to define what factors are responsible for CRE activity. CEpiL cells require active WNT signaling conditions to express *Cdx2*^{31,32,55} (Figure 1D). We therefore examined the chromatin immunoprecipitation sequencing (ChIP-seq) signal of several WNT effectors in naive mouse pluripotent ESCs versus CEpiL cells (Figure 3A), distinct cellular conditions that respectively repress or promote *Cdx2* expression.⁵⁷ Analysis of these data revealed that despite the presence of TCF/LEF sites in multiple, accessible CREs (Figure 1C), the occupancy of WNT effectors is highly selective. CTNNB1 (β CAT), LEF1, and TCF3 are exclusively occupying the intronic CRE in both conditions (Figure 3A). Further examination of the intronic CRE revealed that WNT effectors occupy two distinct subregions within the *Intron1* CRE (labeled *P1* and *P2*; Figure 3A), depending on the cellular

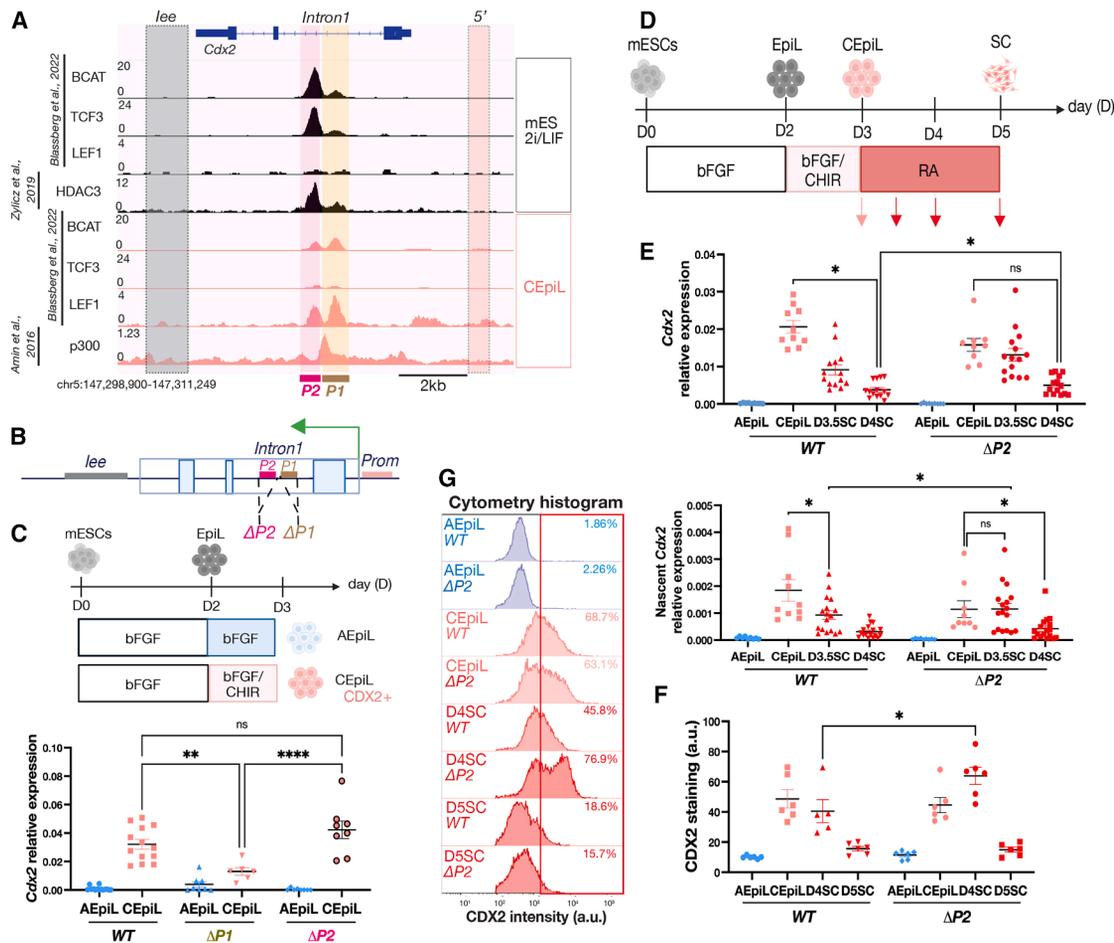


Figure 3. Cells lacking *P2* transiently prolong *Cdx2* in SC progenitors

(A) ChIP-seq signal in naive ESCs (black) versus CEpiL cells (pink) from indicated studies, highlighting context-specific binding at *P1* and *P2*, adjacent regions located within the *Intron1* element but not the *lee* or *5'* element. *P1* is occupied by LEF1 and p300, while CTNNB1 (β CAT), TCF3, and HDAC3 occupy *P2*. (B) Schematic illustrating the deleted region in $\Delta P1$ or $\Delta P2$ cells. (C) Relative expression (RT-qPCR) for *Cdx2* in indicated conditions from $\Delta P1$, $\Delta P2$, and WT cells demonstrates that *Cdx2* is induced in the absence of *P2* but not *P1*. (D) Schematic of SC differentiation highlighting the time points assayed (red arrows). (E) RT-qPCR for spliced and nascent *Cdx2* levels demonstrates that $\Delta P2$ cells fail to downregulate nascent transcription at day 3.5. (F and G) CDX2 levels (F) assessed by flow cytometry and representative cytometry histogram (G) for CDX2 show a population of CDX2-positive cells in $\Delta P2$, unlike WT SC progenitors. (B), (C), and (D) created with BioRender.com. Data are represented as mean \pm SEM. AEpiL, anterior epiblast-like; CEpiL, caudal epiblast-like; EpiL, epiblast-like; SC, spinal cord. **p* value < 0.05, ***p* value < 0.01, *****p* value < 0.001.

conditions. In naive pluripotency conditions, in which *Cdx2* is repressed, *P2* is occupied by CTNNB1 and TCF3. By contrast, the expression of *Cdx2* in CEpiL cells coincides with LEF1 preferentially occupying *P1*, while TCF3 and, to a lesser extent, CTNNB1 are depleted at *P2*.

We hypothesized from these data that *P1* and *P2* may mediate opposing regulatory functions that favor activation (at *P1*) versus repression (at *P2*) of *Cdx2*. Consistent with this hypothesis, the histone deacetylase 3 (HDAC3) preferentially accumulates at *P2* in naive pluripotency conditions,⁶¹ in contrast to the transcriptional co-activator p300, which occupies *P1* in CEpiL cells.⁹ These defined subregions contrast with the relatively broad deposition of H3K27me3 detected at *Cdx2* in both anterior and caudal (*Cdx2* expressing) epiblast tissues *in vivo*⁶² (Figure S3A), in addition

to both ESCs and SC progenitors engineered *in vitro*^{33,63} (Figure S3B). To test the possibility that *P1* and *P2* perform separate regulatory functions, we generated ESCs lacking either *P1* or *P2* (Figure 3B) and directed their differentiation into CEpiL cells (Figure 3C). ESCs lacking *P1*, a region encompassing \sim 225 bp, recapitulated the effect of removing the entire intronic CRE (990 bp): *Cdx2* induction was impaired despite exposure to active WNT signaling conditions (Figure 3C). By contrast, in the absence of *P2*, cells maintain the ability to induce *Cdx2* (Figure 3C). Strikingly, in SC conditions (Figure 3D), *P2*-lacking ESCs prolong *Cdx2* expression, with higher levels of spliced and nascent transcript (Figure 3E) detected transiently at day 3.5. By contrast, nascent transcription was comparable in the presence or absence of *P2* at day 4, while *P2*-lacking ESCs showed an increase in spliced

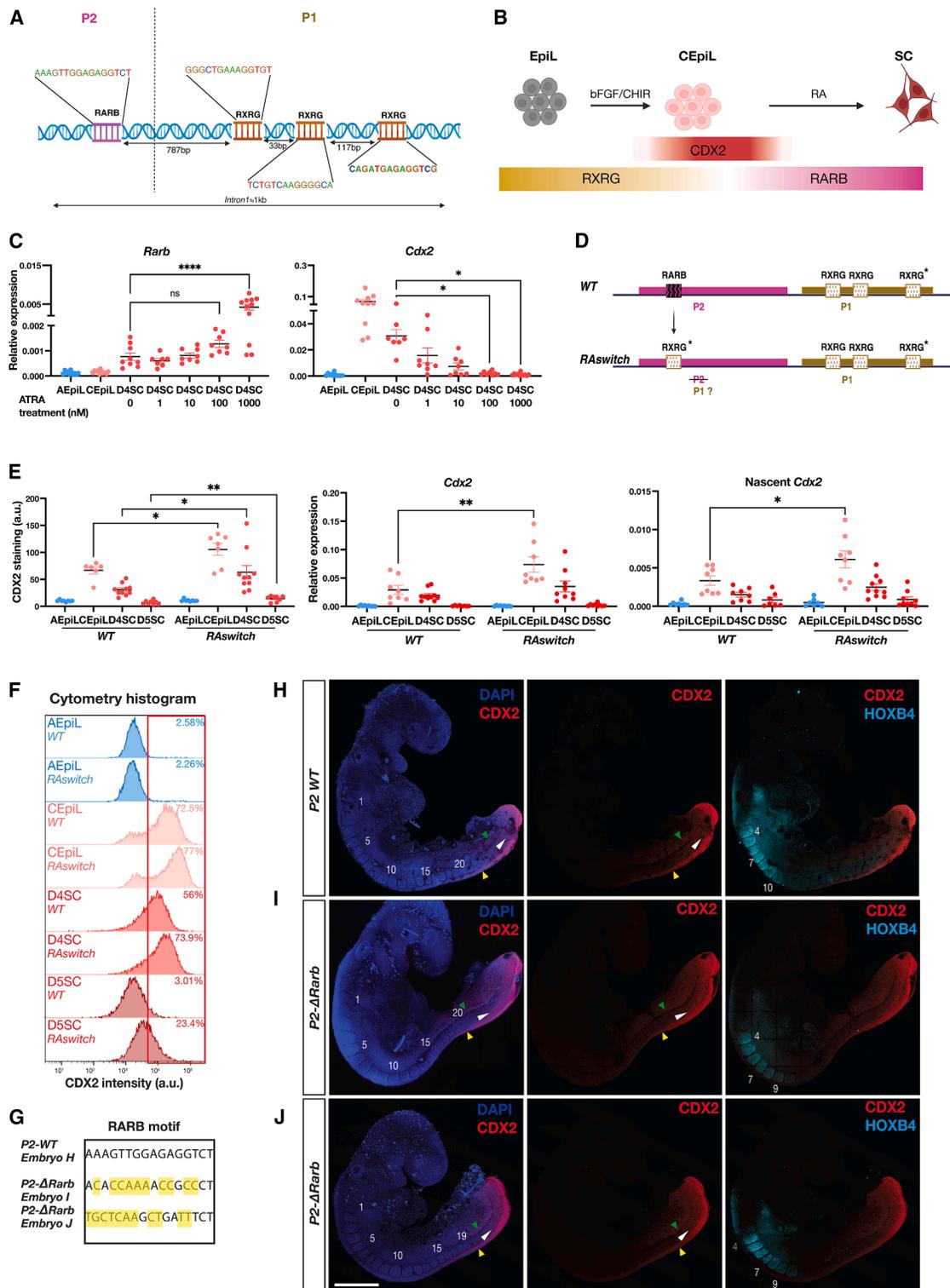


Figure 4. RARE motif switching controls *Cdx2* attenuator activity

(A and B) Schematic of the RARE binding motifs distributed within P1 and P2 at *Intron1* and summary of the RA nuclear receptor expression changes observed as mESCs differentiate into SC progenitors.^{24,31}

(C) Relative expression (RT-qPCR) for *Rarb* and *Cdx2* in cells exposed to increasing amounts of ATRA shows an increase in *Rarb* at 1 μM coinciding with a significant reduction in *Cdx2* in day 4 SC cells versus no ATRA treatment control.

(D) *RAswitch* ESCs harbor a single copy of RXRG (asterisk) instead of an RARB motif at P2.

(legend continued on next page)

transcript (Figure 3E) and higher levels of protein relative to control cells (Figures 3F and 3G). In summary, the data indicate that defined subregions within a single intron mediate opposing regulatory outcomes on *Cdx2* and identify *P2* as a critical region involved in the extinction of *Cdx2* in SC progenitors.

RA nuclear receptor composition dictates regulatory element activity

Having identified two adjacent regions harboring similar TCF/LEF motifs (Table S3), with opposing regulatory outcomes on *Cdx2*, we next asked what factors recruited to *P1* and *P2* could explain their functional differences. We performed motif analysis to predict TFs occupying *P1* and *P2* (Table S3). From this analysis, we recovered SOX and TCF/LEF sites at both *P1* and *P2*, consistent with their known occupancy at these sites⁵⁷ (Figures 3A and S4E). In addition, we detected a striking difference in the composition of RAR versus RXR RA response elements (RAREs), recognized by the RA family of nuclear receptor TFs (Figures 4A, S4A, and S4B). Within *P1*, three separate, conserved RXRG motifs were detected (Figures 4A and S4A). By contrast, *P2* lacked any recognizable RXRG motifs and instead contained a single, rodent-specific RARB motif (Figures 4A, S4A, and S4B). In contrast to *Rarg*, which is detected throughout the differentiation, *Rxrg* levels drop while *Rarb* levels rise in single cells as they progress from a CEpiL to SC identity²⁴ (Figure 4B). This raises the possibility that the combination of nuclear receptor subtypes present in different cell types^{24,35,64–66} plays a central role in regulating *Cdx2* expression.

Consistent with this view, CEpiL cells treated with increasing amounts of all-*trans*-RA (ATRA) display increased levels of *Rarb* in the resulting SC progenitors, while *Cdx2* was reduced (Figure 4C). Having established that increased levels of *Rarb* correspond to lower levels of *Cdx2* in SC progenitors, we sought to verify whether the repressive activity of *P2* relies on the presence of the RARB motif in the region. We generated ESCs in which 75% of the RARB motif present in *P2* was disrupted (Figure S4C highlighted in yellow). Mimicking the $\Delta P2$ cells (Figure 3E), the resulting $\Delta Rarb$ day 4 SC cells showed an increase in *Cdx2* transcript and protein (Figure S4D) levels relative to WT day 4 SC cells.

These findings raise the possibility that variation in RARE motifs dictates regulatory element function. To test the hypothesis that motif composition underpins the functional differences between *P1* and *P2*, we generated ESCs in which the single RARB motif present in *P2* was replaced with a single copy of the RXRG motif, identical in sequence to the highest-affinity RXRG site detected in *P1* (Figure 4D; RXRG site labeled with an asterisk). The resulting *RAswitch* ESCs, which harbor an additional RXRG site compared with control cells, displayed an in-

crease in *Cdx2* levels in CEpiL cells, detected at both the transcript and protein level (Figures 4E and 4F, in light pink). In addition, as *RAswitch* cells progressed to an SC identity, CDX2 expression was maintained, in contrast to control cells, which instead began downregulating *Cdx2*, demonstrating the vital role of the RARE subtype at *P2* to ensure the appropriate offset of *Cdx2* in SC progenitors (Figures 4E and 4F, in red). Taken together, the data demonstrate that the total number and composition of RAREs within *P2* determine its regulatory effect on *Cdx2*.

As the onset and termination of *Cdx2* is mediated through the intronic CRE, also bound by several WNT effectors, how do alternative RA nuclear receptors facilitate *Cdx2* transcriptional control (Figure 3A)? Tight control over the level of SOX2 in CEpiL cells is required for *Cdx2* expression and correlates with genome-wide redistribution of several TCF/LEF effectors, both at the *Cdx2* locus and at a genome-wide scale⁵⁷ (Figure S4E). To probe potential interactions between RARB or RXRG and LEF1 or SOX2, we used AlphaFold-Multimer⁶⁷ to perform *in silico* predictions of possible protein-protein complexes (Figure S4F). These simulations provide evidence that both RARB and RXRG can interact with LEF1 and SOX2. In line with this, ChIP-qPCR experiments showed that the enrichment of RARB at *P2* is reduced, together with LEF1 and SOX2 (Figures S4H and S4I), in day 4 SC *RAswitch* cells that lack the RARB motif at *P2*. By contrast, similar levels of LEF1 and SOX2 are detected across control and *RAswitch* cells (Figure S4I) at *P1*. Taken together, these findings support the hypothesis that, as cells adopt an SC identity, the loss of *Cdx2* coincides with the accumulation of LEF1 and SOX2 at *P2* in an RARB-dependent manner.

P2 is indispensable for posterior body formation

Having established a model of *Cdx2* regulatory control during posterior body formation *in vitro*, we sought to validate these findings *in vivo*. To this end, we attempted to perturb *Cdx2* by disrupting the RARB site within *P2* in mouse embryos. We reasoned that if *P2* limits the duration of *Cdx2* during posterior body formation (Figures 3E–3G), disruptions to the RARB site alone would be expected to transiently prolong *Cdx2* expression, an effect known to perturb tailbud morphology and *Hox* gene expression boundaries in mouse embryos.^{2,11} We used CRISPR-Cas9, together with the same gRNA pairs used *in vitro*, to remove *P2* ($\Delta P2$) or to disrupt directly the RARB site in fertilized zygotes and collected resulting somite-matched transient transgenic embryos at mid-gestation (Table S2). *P2*- $\Delta Rarb$ CRISPR mutants recovered at ~E9 harbored mutations and deletions within the *P2* region (Figure 4G). The RARB site was disrupted in mutants (Figures 4G, 4I, and 4J) relative to control littermates that harbored an entirely intact *P2* CRE (WT *P2*; Figures 4G and 4H).

(E) Quantification of CDX2 levels assessed by flow cytometry shows a significant increase in expression levels in *RAswitch* cells compared with WT, while relative expression (RT-qPCR) for *Cdx2* or nascent *Cdx2* shows a significant increase only in *RAswitch* CEpiL cells compared with WT.

(F) Representative flow cytometry histogram for CDX2, related to (E).

(G) Resulting genotypes in embryos presented in (H)–(J) following CRISPR-Cas9 in fertilized zygotes highlighting mismatches to the WT RARB motif.

(H)–(J) Whole-mount IF shows broader detection of CDX2 (red) in the tailbud of *P2*- $\Delta Rarb$ mutants (I and J) versus the control (H) at somite-matched stages. WT *Rarb* embryos display a caudal limit of HOXB4 that extends to somite 10 (H), versus a caudal limit at somite 9 in *P2*- $\Delta Rarb$ mutants (I and J). Abnormal tailbud morphology is most apparent in the *P2*- $\Delta Rarb$ mutant presented in (J). Somites are numbered in white; green, white, and yellow arrows indicate the approximate anterior limit of CDX2 expression in the lateral plate mesoderm, PSM, and neural tube, respectively. Scale bar represents 500 μ m. (A), (B), and (D) were created with BioRender.com. Data are represented as mean \pm SEM. **p* value < 0.05, ***p* value < 0.01, *****p* value < 0.0001.

Whole embryo IF and imaging confirmed that *P2-WT* embryos recapitulated the known endogenous expression pattern of CDX2,^{3,9} which is restricted to the tailbud and the caudal-most aspect of the neural tube at this stage (Figure 4H). Strikingly, the *P2-ΔRarb* CRISPR mutant embryos displayed a rostral expansion in CDX2, most notably in the mesoderm (Figures 4I and 4J; green and white arrows). In addition, regional identity was disrupted in the somites. Mutant embryos displayed a caudal limit of HOXB4 that extended to somite 9, in contrast to the control, which reached a more caudal position, up to somite 10 (Figure 4; compare Figures 4H–4J). Mutant embryos also displayed abnormal tailbud morphology, reminiscent of the phenotype resulting from prolonged *Cdx2* expression in mouse embryos.² These data provide evidence that *P2* is required to restrict the activity of CDX2 *in vivo* and pinpoint the RARB motif as a critical site that operates during posterior body formation.

DISCUSSION

Multiple, nonredundant CREs coordinate *Cdx2* expression during development

Using an *in vitro* model of posterior body development, we demonstrate that multiple, functionally discrete CREs convert extrinsic cues into a finite window of expression. These data support the idea proposed by previous enhancer reporter experiments that multiple CREs located proximally to the promoter regulate *Cdx2*.⁵³ Furthermore, our data extend these findings by demonstrating the functional specificity of CREs during development: *Intron1* is required for induction in CEpiL conditions (Figures 1E–1G), in contrast to the 5' CRE, which is not required in this context but is indispensable at later stages to maintain *Cdx2* transiently in SC or paraxial mesoderm progenitors (Figure 2F). Furthermore, the occupancy of CDX2 at the 5' CRE suggests that this element may perform its maintenance function at least in part via a positive feedback loop (Figure S2D). Previous studies also proposed that a silencer may regulate the caudal expression pattern of *Cdx2*.⁵³ Here, we resolve a minimal region residing within *Intron1* (*P2*; Figure 3A) required to limit the duration of *Cdx2* in SC progenitors (Figures 3E–3G), and validate its requirements for appropriate CDX2 activity *in vivo* (Figures 4H–4J). By contrast, previous studies demonstrate that a fragment containing *P2* displays enhancer activity in transgenic mouse reporter assays.^{53,59} Although we cannot exclude the possibility that *P2* displays enhancer activity in an alternative context and thus may act as a bifunctional element,^{68,69} our findings highlight that targeted base pair substitutions at CREs in their native context can aid in the identification of regulatory regions that include repressive elements.

Our results confirm that individual CREs perform indispensable roles since single CRE deletions are sufficient to perturb the expression window and cannot be compensated for by the presence of alternative and accessible CREs (Figures 1E–1G, 2E, 2F, and 3C–3G). The requirement for several, functionally distinct CREs may ensure robustness in gene expression.^{70–72} Consistent with this view, removal of the 5' or *P1* region has a clear but limited effect on *Cdx2* expression, potentially due to the presence of additional CREs that are yet to be resolved. In addition, *P2* extends the expression window in SC progenitors, yet *Cdx2* is eventually ex-

tinguished in these cells (Figures 3E–3G). The presence of additional, potentially long-range CREs likely explains this effect. Evidence of long-range CREs impacting *Cdx2* expression has been previously described in an alternative cellular context, in B cell acute lymphoblastic leukemia patients.^{51,52} Such long-range CREs may also buffer fluctuations of extrinsic signaling, as observed in the zebrafish neural plate border for *Zic3* expression⁷³ and more recently in mouse embryos.^{70,74}

Motif composition dictates attenuator function

In this study, *P2* represses *Cdx2* in a context-specific manner, as its removal transiently alters the level of nascent transcription in SC progenitors but not CEpi conditions (Figure 3E). Whether *P2* mediates any part of its repressive function at the level of splicing or transcript stability remains to be determined. As *P2* represses *Cdx2* in a limited manner, we refer to this element as an “attenuator.” In contrast to enhancers, relatively few repressive CREs have been identified and functionally validated during development, especially in mammals.⁷⁵ As functional validation of repressive elements is challenging to perform at scale,⁷⁶ the mechanisms that distinguish repressive versus activating elements remain to be elucidated.

Cdx2 is expressed in response to active WNT signaling conditions, and ChIP-seq against LEF1 demonstrates its preferential accumulation at *P1* (Figure 3A). However, TCF/LEF binding motifs are also present in multiple *Cdx2* CREs, yet, despite their accessibility, these sites do not compensate for the intronic CRE upon its removal in *ΔIntron1* or *ΔP1* cells. These findings suggest that chromatin accessibility is not sufficient to predict enhancer function at CREs,⁵⁸ which indicates that an additional mechanism is involved. Recent findings indicate that the specificity of TCF/LEF binding in the genome is driven by context-specific TFs⁷⁷ and their level of expression.⁵⁷ In CEpiL cells, SOX2 levels dictate the genome-wide occupancy of several WNT effectors, including the occupancy of LEF1 and CTNNB1 at the *Cdx2* intronic CRE.⁵⁷ Moreover, the activity of the intronic CRE requires SOX2 binding sites,⁵⁷ consistent with the view that the recruitment of WNT effectors is driven by cooperation between cell identity-specific TFs. As *P1* and *P2* both harbor SOX2 binding sites and can be occupied by SOX2,⁵⁷ an additional molecular determinant must govern *P1* versus *P2* function.

In this study, we provide evidence that regulatory element function is dependent on motif composition. We demonstrate that *P2* can be converted into an enhancer through a single motif switch from RARB to RXRG (Figures 4E and 4F). This switch disrupts RARB enrichment and the recruitment of SOX2/LEF1 at *P2*. As RARB and RXRG can each interact with SOX2 and LEF1 *in silico*, these findings suggest that the regulatory function of *P2* is driven by RA nuclear receptor subtypes (Figure S4F). Dissection of individual elements in different species demonstrates that motif composition can dictate silencer function in *Drosophila*.^{78,79} In mammalian genomes, the same CRE can operate as an enhancer or silencer, a function that changes depending on the cellular context.^{80–84} Our results demonstrate that a single nuclear receptor motif switch is sufficient to change the function of a given CRE without altering the cellular conditions. These findings support the view that TF engagement contributes to the functional versatility of CREs across different cell types.^{58,81,84}

Repressive mechanisms operating during posterior body development

Although RA is a known major determinant of posterior body formation,^{3,24,27,35,36,85–87} its mechanism of action is not fully understood. Among the predicted 14,000 potential RAREs in the mouse genome,⁸⁸ only a handful have been experimentally validated, including both enhancers^{89,90} and silencers.^{91,92} However, what underpins enhancer or silencer activity at individual RAREs remains unclear.⁹⁰ Here, we identify activating and repressive RAREs for *Cdx2* that suggest RARB and RXRG may exert opposing regulatory roles. Furthermore, we validate that the RARB motif we have identified is occupied by the nuclear receptor RARB and demonstrate that variation in RARE motifs impacts RARB occupancy and regulatory element function. The data suggest that differences in nuclear receptor recruitment at individual RAREs may underpin the pleiotropic role of RA during development.⁹³

Recruitment of cofactors is likely to impart distinct functions at RAREs. Consistent with this view, the occupancy of distinct RA nuclear receptors is associated with changes in cofactor recruitment. RARG recruits the transcriptional co-activators pCIP/p300 to an RARE regulating *Hoxa1*, whereas the occupancy of RARB/RARA abolishes p300 and instead promotes the recruitment of the Polycomb subunit SUZ12.⁸⁹ In addition, NCOR1/2,⁹⁴ HDAC,^{92,95} and Polycomb repressive complexes (PRCs)^{89,92} are recruited in the vicinity of RAREs associated with transcriptional repression in response to RA signaling. PRC recruitment can also impact chromatin compaction,^{96,97} yet how these events relate to nuclear receptor engagement at defined CREs remains unclear. Our findings indicate that HDAC3 accumulates at *P2* in repressive conditions (Figure 3A); how this relates to RARB is unresolved. By contrast, PRC is deposited in a widespread manner at *Cdx2* (Figure S3), as commonly observed at developmental genes.^{33,98,99} In summary, multiple mechanisms are likely to explain how nuclear receptor subtypes exert regulatory roles during development.

Since previously predicted^{75,100} and experimentally validated^{101–103} silencers are located relatively close to the transcription start site of genes and can be found adjacent to enhancers,^{100,101} short-range gene silencing mechanisms may represent a more general principle of gene regulation during development. Consistent with this view, *Cdx1* is regulated by a silencer located ~400 bp upstream of the promoter and ~500 bp away from an enhancer.¹⁰¹ Physical obstruction of individual enhancers or their interaction with the promoter could silence gene expression.³⁹

CDX factors display a graded expression profile along the rostrocaudal axis,¹⁹ which, in turn, plays a central role in constraining regional identity through the regulation of *Hox* genes.² That CDX factors contain conserved regulatory elements and play a caudalizing role in several species^{8,104–108} suggests that the regulatory principles governing their transient expression may underpin body plan organization across multiple bilaterian animals.

Limitations of the study

In the current work, how *P2* attenuates *Cdx2* is not resolved. Future studies will elucidate how the occupancy of RARB at *P2* impacts transcription, broadening our understanding of the gene regulatory mechanisms operating during mammalian development.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Vicki Metzis.

Materials availability

Murine ESC lines generated in this study are available upon request.

Data and code availability

Accession numbers of the reanalyzed sequencing dataset are provided in the STAR Methods section. All original code has been deposited at https://github.com/da-bar/cdx2_transient_expression. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank James Briscoe, Joaquina Delas, Matthias Merckenschlager, Teresa Rayon, Kate Storey, Juan M. Vaquerizas, and all lab members for comments on the manuscript. We are grateful to Tristan Rodriguez for advice on the generation of transgenic embryos. For support, training, and access to equipment, we thank James Elliot from the LMS/NIHR Imperial Biomedical Research Centre Flow Cytometry Facility, the staff at the Central Biomedical Services unit at Imperial College London, Zoe Webster at the LMS transgenic facility, and Dirk Dormann at the LMS microscopy facility. This work was supported by core funding to the Laboratory of Medical Sciences from the Medical Research Council, a Sir Henry Dale Fellowship awarded to V.M., jointly funded by the Wellcome Trust and the Royal Society (grant number 218536/Z/19/Z), and a Small-scale Researcher Mobility Scheme between the UK and Croatia (011-31/1-2024/pm) to D.B. This work was funded by the European Union – NextGenerationEU grant NPOO.C3.2.R2-11.06.0060.

AUTHOR CONTRIBUTIONS

I.A. and V.M. conceived the project, designed the experiments, interpreted the data, and wrote the manuscript. I.A. performed the experiments and data analysis, with assistance from S.Q.X. D.B. performed data analysis and together with B.L. interpreted the data. B.M. performed microinjections and embryo transfers. M.P. performed data analysis. All authors revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
 - Cells lines
 - ESC culture and differentiation
 - CRISPR mutant embryos
- METHOD DETAILS
 - Immunofluorescence on cells
 - Flow cytometry
 - RNA extraction, cDNA synthesis and RT-qPCR analysis
 - Embryo wholemount immunofluorescence
 - ChIP-qPCR
 - ChIP-seq, ATAC-seq and mRNA-seq data and processing
 - Identification of Transcription Factor Binding Sites (TFBS)
 - Prediction of Protein-Protein interaction Complexes
 - Experimental design
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.devcel.2025.06.006>.

Received: April 8, 2024
Revised: January 24, 2025
Accepted: June 4, 2025
Published: June 27, 2025

REFERENCES

- Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. *Nature* 468, 911–920. <https://doi.org/10.1038/nature09645>.
- Gaunt, S.J., Drage, D., and Trubshaw, R.C. (2008). Increased Cdx protein dose effects upon axial patterning in transgenic lines of mice. *Development* 135, 2511–2520. <https://doi.org/10.1242/dev.015909>.
- Young, T., Rowland, J.E., Van De Ven, C., Bialecka, M., Novoa, A., Carapuco, M., Van Nes, J., De Graaff, W., Duluc, I., Freund, J.-N., et al. (2009). Cdx and Hox Genes Differentially Regulate Posterior Axial Growth in Mammalian Embryos. *Dev. Cell* 17, 516–526. <https://doi.org/10.1016/j.devcel.2009.08.010>.
- Van Rooijen, C., Simmini, S., Bialecka, M., Neijts, R., Van De Ven, C., Beck, F., and Deschamps, J. (2012). Evolutionarily conserved requirement of Cdx for post-occipital tissue emergence. *Development* 139, 2576–2583. <https://doi.org/10.1242/dev.079848>.
- Neijts, R., Amin, S., Van Rooijen, C., and Deschamps, J. (2017). Cdx is crucial for the timing mechanism driving colinear Hox activation and defines a trunk segment in the Hox cluster topology. *Dev. Biol.* 422, 146–154. <https://doi.org/10.1016/j.ydbio.2016.12.024>.
- Van Den Akker, E., Forlani, S., Chawengsaksophak, K., De Graaff, W., Beck, F., Meyer, B.I., and Deschamps, J. (2002). *Cdx1* and *Cdx2* have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* 129, 2181–2193. <https://doi.org/10.1242/dev.129.9.2181>.
- Savory, J.G.A., Pilon, N., Grainger, S., Sylvestre, J.-R., Béland, M., Houle, M., Oh, K., and Lohnes, D. (2009). *Cdx1* and *Cdx2* are functionally equivalent in vertebral patterning. *Dev. Biol.* 330, 114–122. <https://doi.org/10.1016/j.ydbio.2009.03.016>.
- Chawengsaksophak, K., De Graaff, W., Rossant, J., Deschamps, J., and Beck, F. (2004). *Cdx2* is essential for axial elongation in mouse development. *Proc. Natl. Acad. Sci. USA* 101, 7641–7645. <https://doi.org/10.1073/pnas.0401654101>.
- Amin, S., Neijts, R., Simmini, S., Van Rooijen, C., Tan, S.C., Kester, L., Van Oudenaarden, A., Creighton, M.P., and Deschamps, J. (2016). Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. *Cell Rep.* 17, 3165–3177. <https://doi.org/10.1016/j.celrep.2016.11.069>.
- Needham, J., and Metzis, V. (2022). Heads or tails: making the spinal cord. *Dev. Biol.* 485, 80–92. <https://doi.org/10.1016/j.ydbio.2022.03.002>.
- Charité, J., de Graaff, W.D., Consten, D., Reijnen, M.J., Korving, J., and Deschamps, J. (1998). Transducing positional information to the *Hox* genes: critical interaction of *cdx* gene products with position-sensitive regulatory elements. *Development* 125, 4349–4358. <https://doi.org/10.1242/dev.125.22.4349>.
- Bel-Vialar, S., Itasaki, N., and Krumlauf, R. (2002). Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the *HoxB* genes in two distinct groups. *Development* 129, 5103–5115. <https://doi.org/10.1242/dev.129.22.5103>.
- Shimizu, T., Bae, Y.-K., and Hibi, M. (2006). Cdx-Hox code controls competence for responding to Fgfs and retinoic acid in zebrafish neural tissue. *Development* 133, 4709–4719. <https://doi.org/10.1242/dev.02660>.
- Skromne, I., Thorsen, D., Hale, M., Prince, V.E., and Ho, R.K. (2007). Repression of the hindbrain developmental program by Cdx factors is required for the specification of the vertebrate spinal cord. *Development* 134, 2147–2158. <https://doi.org/10.1242/dev.002980>.
- Sturgeon, K., Kaneko, T., Biemann, M., Gauthier, A., Chawengsaksophak, K., and Cordes, S.P. (2011). *Cdx1* refines positional identity of the vertebrate hindbrain by directly repressing *Mafb* expression. *Development* 138, 65–74. <https://doi.org/10.1242/dev.058727>.
- Van Nes, J., De Graaff, W., Lebrin, F., Gerhard, M., Beck, F., and Deschamps, J. (2006). The *Cdx4* mutation affects axial development and reveals an essential role of Cdx genes in the ontogenesis of the placental labyrinth in mice. *Development* 133, 419–428. <https://doi.org/10.1242/dev.02216>.
- Young, T., and Deschamps, J. (2009). Chapter 8. Hox, Cdx, and Anteroposterior Patterning in the Mouse Embryo. *Curr. Top. Dev. Biol.* 88, 235–255. [https://doi.org/10.1016/S0070-2153\(09\)88008-3](https://doi.org/10.1016/S0070-2153(09)88008-3).
- Beck, F., Erler, T., Russell, A., and James, R. (1995). Expression of *Cdx-2* in the mouse embryo and placenta: Possible role in patterning of the extra-embryonic membranes. *Dev. Dyn.* 204, 219–227. <https://doi.org/10.1002/aja.1002040302>.
- Gaunt, S.J., Drage, D., and Trubshaw, R.C. (2005). *cdx4/lacZ* and *cdx2/lacZ* protein gradients formed by decay during gastrulation in the mouse. *Int. J. Dev. Biol.* 49, 901–908. <https://doi.org/10.1387/ijdb.052021sg>.
- McDole, K., and Zheng, Y. (2012). Generation and live imaging of an endogenous *Cdx2* reporter mouse line. *Genesis* 50, 775–782. <https://doi.org/10.1002/dvg.22049>.
- Tzouanacou, E., Wegener, A., Wymeersch, F.J., Wilson, V., and Nicolas, J.-F. (2009). Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Dev. Cell* 17, 365–376. <https://doi.org/10.1016/j.devcel.2009.08.002>.
- Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. *Development* 142, 2864–2875. <https://doi.org/10.1242/dev.119768>.
- Binagui-Casas, A., Dias, A., Guillot, C., Metzis, V., and Saunders, D. (2021). Building consensus in neuromesodermal research: Current advances and future biomedical perspectives. *Curr. Opin. Cell Biol.* 73, 133–140. <https://doi.org/10.1016/j.ceb.2021.08.003>.
- Gouti, M., Delile, J., Stamatakis, D., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2017). A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Dev. Cell* 41, 243–261.e7. <https://doi.org/10.1016/j.devcel.2017.04.002>.
- Guibentif, C., Griffiths, J.A., Imaz-Rosshandler, I., Ghazanfar, S., Nichols, J., Wilson, V., Göttgens, B., and Marioni, J.C. (2021). Diverse Routes toward Early Somites in the Mouse Embryo. *Dev. Cell* 56, 141–153.e6. <https://doi.org/10.1016/j.devcel.2020.11.013>.
- Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* 566, 490–495. <https://doi.org/10.1038/s41586-019-0933-9>.
- Zhao, X., and Duester, G. (2009). Effect of retinoic acid signaling on Wnt/β-catenin and FGF signaling during body axis extension. *Gene Expr. Patterns* 9, 430–435. <https://doi.org/10.1016/j.gep.2009.06.003>.
- Ikeya, M., and Takada, S. (2001). Wnt-3a is required for somite specification along the anteroposterior axis of the mouse embryo and for regulation of *cdx-1* expression. *Mech. Dev.* 103, 27–33. [https://doi.org/10.1016/s0925-4773\(01\)00338-0](https://doi.org/10.1016/s0925-4773(01)00338-0).
- Keenan, I.D., Sharrard, R.M., and Isaacs, H.V. (2006). FGF signal transduction and the regulation of Cdx gene expression. *Dev. Biol.* 299, 478–488. <https://doi.org/10.1016/j.ydbio.2006.08.040>.
- Frith, T.J., Granata, I., Wind, M., Stout, E., Thompson, O., Neumann, K., Stavish, D., Heath, P.R., Ortmann, D., Hackland, J.O., et al. (2018).

- Human axial progenitors generate trunk neural crest cells in vitro. *eLife* 7, e35786. <https://doi.org/10.7554/eLife.35786>.
31. Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In Vitro Generation of Neuromesodermal Progenitors Reveals Distinct Roles for Wnt Signalling in the Specification of Spinal Cord and Paraxial Mesoderm Identity. *PLoS Biol.* 12, e1001937. <https://doi.org/10.1371/journal.pbio.1001937>.
 32. Lippmann, E.S., Williams, C.E., Ruhl, D.A., Estevez-Silva, M.C., Chapman, E.R., Coon, J.J., and Ashton, R.S. (2015). Deterministic HOX Patterning in Human Pluripotent Stem Cell-Derived Neuroectoderm. *Stem Cell Rep.* 4, 632–644. <https://doi.org/10.1016/j.stemcr.2015.02.018>.
 33. Mazzoni, E.O., Mahony, S., Peljto, M., Patel, T., Thornton, S.R., McQuine, S., Reeder, C., Boyer, L.A., Young, R.A., Gifford, D.K., et al. (2013). Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. *Nat. Neurosci.* 16, 1191–1198. <https://doi.org/10.1038/nn.3490>.
 34. Wind, M., Gogolou, A., Manipur, I., Granata, I., Butler, L., Andrews, P.W., Barbaric, I., Ning, K., Guarracino, M.R., Placzek, M., et al. (2021). Defining the signalling determinants of a posterior ventral spinal cord identity in human neuromesodermal progenitor derivatives. *Development* 148, dev194415. <https://doi.org/10.1242/dev.194415>.
 35. Diez del Corral, R.D., Olivera-Martinez, I., Goriely, A., Gale, E., Maden, M., and Storey, K. (2003). Opposing FGF and Retinoid Pathways Control Ventral Neural Pattern, Neuronal Differentiation, and Segmentation during Body Axis Extension. *Neuron* 40, 65–79. [https://doi.org/10.1016/S0896-6273\(03\)00565-8](https://doi.org/10.1016/S0896-6273(03)00565-8).
 36. Olivera-Martinez, I., and Storey, K.G. (2007). Wnt signals provide a timing mechanism for the FGF-retinoid differentiation switch during vertebrate body axis extension. *Development* 134, 2125–2135. <https://doi.org/10.1242/dev.000216>.
 37. Cooper, F., Gentsch, G.E., Mitter, R., Bouissou, C., Healy, L.E., Rodriguez, A.H., Smith, J.C., and Bernardo, A.S. (2022). Rostrocaudal patterning and neural crest differentiation of human pre-neural spinal cord progenitors in vitro. *Stem Cell Rep.* 17, 894–910. <https://doi.org/10.1016/j.stemcr.2022.02.018>.
 38. Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* 167, 1170–1187. <https://doi.org/10.1016/j.cell.2016.09.018>.
 39. Ogbourne, S., and Antalis, T.M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem. J.* 331, 1–14. <https://doi.org/10.1042/bj3310001>.
 40. Herold, M., Bartkuhn, M., and Renkawitz, R. (2012). CTCF: insights into insulator function during development. *Development* 139, 1045–1057. <https://doi.org/10.1242/dev.065268>.
 41. Batut, P.J., Bing, X.Y., Sisco, Z., Raimundo, J., Levo, M., and Levine, M. S. (2022). Genome organization controls transcriptional dynamics during development. *Science* 375, 566–570. <https://doi.org/10.1126/science.abi7178>.
 42. Blayney, J.W., Francis, H., Rampasekova, A., Camellato, B., Mitchell, L., Stolper, R., Cornell, L., Babbs, C., Boeke, J.D., Higgs, D.R., et al. (2023). Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* 186, 5826–5839.e18. <https://doi.org/10.1016/j.cell.2023.11.030>.
 43. Kim, S., and Wysocka, J. (2023). Deciphering the multi-scale, quantitative cis-regulatory code. *Mol. Cell* 83, 373–392. <https://doi.org/10.1016/j.molcel.2022.12.032>.
 44. Rayon, T., Menchero, S., Nieto, A., Xenopoulos, P., Crespo, M., Cockburn, K., Cañon, S., Sasaki, H., Hadjantonakis, A.-K., de la Pompa, J.L., et al. (2014). Notch and Hippo Converge on Cdx2 to Specify the Trophoblast Lineage in the Mouse Blastocyst. *Dev. Cell* 30, 410–422. <https://doi.org/10.1016/j.devcel.2014.06.019>.
 45. Rayon, T., Menchero, S., Rollán, I., Ors, I., Helness, A., Crespo, M., Nieto, A., Azuara, V., Rossant, J., and Manzanera, M. (2016). Distinct mechanisms regulate Cdx2 expression in the blastocyst and in trophoblast stem cells. *Sci. Rep.* 6, 27139. <https://doi.org/10.1038/srep27139>.
 46. Watts, J.A., Zhang, C., Klein-Szanto, A.J., Kormish, J.D., Fu, J., Zhang, M.Q., and Zaret, K.S. (2011). Study of FoxA Pioneer Factor at Silent Genes Reveals Rfx-Repressed Enhancer at Cdx2 and a Potential Indicator of Esophageal Adenocarcinoma Development. *PLoS Genet.* 7, e1002277. <https://doi.org/10.1371/journal.pgen.1002277>.
 47. Benahmed, F., Gross, I., Gaunt, S.J., Beck, F., Jehan, F., Domon-Dell, C., Martin, E., Kedinger, M., Freund, J.N., and Duluc, I. (2008). Multiple Regulatory Regions Control the Complex Expression Pattern of the Mouse Cdx2 Homeobox Gene. *Gastroenterology* 135, 1238–1247.e3. <https://doi.org/10.1053/j.gastro.2008.06.045>.
 48. Chen, Y., Tan, F., Fang, Q., Zhang, L., Liao, J., Qian, Y., Wen, M., Song, R., Fu, Y., Xu, H.J., et al. (2025). Gastrula-premarked posterior enhancer primes posterior tissue 2 development through cross-talk with TGF- β signaling pathway. Preprint at bioRxiv. <https://doi.org/10.1101/2024.04.14.589453>.
 49. Friman, E.T., Flyamer, I.M., Marenduzzo, D., Boyle, S., and Bickmore, W. A. (2023). Ultra-long-range interactions between active regulatory elements. *Genome Res.* 33, 1269–1283. <https://doi.org/10.1101/gr.277567.122>.
 50. Montavon, T., and Duboule, D. (2013). Chromatin organization and global regulation of *Hox* gene clusters. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368, 20120367. <https://doi.org/10.1098/rstb.2012.0367>.
 51. Kimura, S., Montefiori, L., Iacobucci, I., Zhao, Y., Gao, Q., Paietta, E.M., Haferlach, C., Laird, A.D., Mead, P.E., Gu, Z., et al. (2022). Enhancer re-targeting of *CDX2* and *UBTF::ATXN7L3* define a subtype of high-risk B-progenitor acute lymphoblastic leukemia. *Blood* 139, 3519–3531. <https://doi.org/10.1182/blood.2022015444>.
 52. Passet, M., Kim, R., Gachet, S., Sigaux, F., Chaumeil, J., Galland, A., Sexton, T., Quentin, S., Hernandez, L., Larcher, L., et al. (2022). Concurrent *CDX2 cis*-deregulation and *UBTF::ATXN7L3* fusion define a novel high-risk subtype of B-cell ALL. *Blood* 139, 3505–3518. <https://doi.org/10.1182/blood.2021014723>.
 53. Wang, W.C.H., and Shashikant, C.S. (2007). Evidence for positive and negative regulation of the mouse *Cdx2* gene. *J. Exp. Zool. Pt. B* 308B, 308–321. <https://doi.org/10.1002/jez.b.21154>.
 54. Strumpf, D., Mao, C.-A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* 132, 2093–2102. <https://doi.org/10.1242/dev.01801>.
 55. Metzis, V., Steinhäuser, S., Pakanavicius, E., Gouti, M., Stamatakis, D., Ivanovitch, K., Watson, T., Rayon, T., Mousavy Gharavy, S.N., Lovell-Badge, R., et al. (2018). Nervous System Regionalization Entails Axial Allocation before Neural Differentiation. *Cell* 175, 1105–1118.e17. <https://doi.org/10.1016/j.cell.2018.09.040>.
 56. Argelaguet, R., Lohoff, T., Li, J.G., Nakhuda, A., Drage, D., Krueger, F., Velten, L., Clark, S.J., and Reik, W. (2022). Decoding Gene Regulation in the Mouse Embryo Using Single-Cell Multi-omics. Preprint at bioRxiv. <https://doi.org/10.1101/2022.06.15.496239>.
 57. Blassberg, R., Patel, H., Watson, T., Gouti, M., Metzis, V., Delás, M.J., and Briscoe, J. (2022). Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat. Cell Biol.* 24, 633–644. <https://doi.org/10.1038/s41556-022-00910-2>.
 58. Delás, M.J., Kalaitzis, C.M., Fawzi, T., Demuth, M., Zhang, I., Stuart, H.T., Costantini, E., Ivanovitch, K., Tanaka, E.M., and Briscoe, J. (2023). Developmental cell fate choice in neural tube progenitors employs two distinct cis-regulatory strategies. *Dev. Cell* 58, 3–17.e8. <https://doi.org/10.1016/j.devcel.2022.11.016>.
 59. Coutaud, B., and Pilon, N. (2013). Characterization of a novel transgenic mouse line expressing Cre recombinase under the control of the *Cdx2* neural specific enhancer. *Genesis* 51, 777–784. <https://doi.org/10.1002/dvg.22421>.

60. Tsakiridis, A., Huang, Y., Blin, G., Skylaki, S., Wymeersch, F., Osorno, R., Economou, C., Karagianni, E., Zhao, S., Lowell, S., et al. (2014). Distinct Wnt-driven primitive streak-like populations reflect *in vivo* lineage precursors. *Development* *141*, 1209–1221. <https://doi.org/10.1242/dev.101014>.
61. Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., Da Rocha, S.T., Schwab, B., Syx, L., Dingli, F., Loew, D., et al. (2019). The Implication of Early Chromatin Changes in X Chromosome Inactivation. *Cell* *176*, 182–197.e23. <https://doi.org/10.1016/j.cell.2018.11.041>.
62. Yang, X., Hu, B., Liao, J., Qiao, Y., Chen, Y., Qian, Y., Feng, S., Yu, F., Dong, J., Hou, Y., et al. (2019). Distinct enhancer signatures in the mouse gastrula delineate progressive cell fate continuum during embryo development. *Cell Res.* *29*, 911–926. <https://doi.org/10.1038/s41422-019-0234-8>.
63. Kanellopoulou, C., Gilpatrick, T., Kilaru, G., Burr, P., Nguyen, C.K., Morawski, A., Lenardo, M.J., and Muljo, S.A. (2015). Reprogramming of Polycomb-Mediated Gene Silencing in Embryonic Stem Cells by the miR-290 Family and the Methyltransferase Ash1l. *Stem Cell Rep.* *5*, 971–978. <https://doi.org/10.1016/j.stemcr.2015.10.001>.
64. Ang, H.L., and Duester, G. (1997). Initiation of retinoid signaling in primitive streak mouse embryos: Spatiotemporal expression patterns of receptors and metabolic enzymes for ligand synthesis. *Dev. Dyn.* *208*, 536–543. [https://doi.org/10.1002/\(SICI\)1097-0177\(199704\)208:4<536::AID-AJA9>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0177(199704)208:4<536::AID-AJA9>3.0.CO;2-J).
65. Ruberte, E., Dolle, P., Krust, A., Zelent, A., Morriss-Kay, G., and Chambon, P. (1990). Specific spatial and temporal distribution of retinoic acid receptor gamma transcripts during mouse embryogenesis. *Development* *108*, 213–222. <https://doi.org/10.1242/dev.108.2.213>.
66. Dollé, P., Ruberte, E., Leroy, P., Morriss-Kay, G., and Chambon, P. (1990). Retinoic acid receptors and cellular retinoid binding proteins: I. A systematic study of their differential pattern of transcription during mouse organogenesis. *Development* *110*, 1133–1151. <https://doi.org/10.1242/dev.110.4.1133>.
67. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Židek, A., Bates, R., Blackwell, S., Yim, J., et al. (2021). Protein complex prediction with AlphaFold-Multimer. Preprint at bioRxiv. <https://doi.org/10.1101/2021.10.04.463034>.
68. Gisselbrecht, S.S., Palagi, A., Kurland, J.V., Rogers, J.M., Ozadam, H., Zhan, Y., Dekker, J., and Bulyk, M.L. (2020). Transcriptional Silencers in *Drosophila* Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Mol. Cell* *77*, 324–337.e8. <https://doi.org/10.1016/j.molcel.2019.10.004>.
69. Erceg, J., Pakozdi, T., Marco-Ferreres, R., Ghavi-Helm, Y., Girardot, C., Bracken, A.P., and Furlong, E.E.M. (2017). Dual functionality of *cis*-regulatory elements as developmental enhancers and Polycomb response elements. *Genes Dev.* *31*, 590–602. <https://doi.org/10.1101/gad.292870.116>.
70. Ahituv, N., Zhu, Y., Visel, A., Holt, A., Afzal, V., Pennacchio, L.A., and Rubin, E.M. (2007). Deletion of Ultraconserved Elements Yields Viable Mice. *PLoS Biol.* *5*, e234. <https://doi.org/10.1371/journal.pbio.0050234>.
71. Dickel, D.E., Ypsilanti, A.R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B.J., Khin, Y.S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Ultraconserved Enhancers Are Required for Normal Development. *Cell* *172*, 491–499.e15. <https://doi.org/10.1016/j.cell.2017.12.017>.
72. Duarte, P., Brattig Correia, R., Nóvoa, A., and Mallo, M. (2023). Regulatory changes associated with the head to trunk developmental transition. *BMC Biol.* *21*, 170. <https://doi.org/10.1186/s12915-023-01675-2>.
73. Garnett, A.T., Square, T.A., and Medeiros, D.M. (2012). BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. *Development* *139*, 4220–4231. <https://doi.org/10.1242/dev.081497>.
74. Exelby, K., Herrera-Delgado, E., Perez, L.G., Perez-Carrasco, R., Sagner, A., Metzis, V., Sollich, P., and Briscoe, J. (2021). Precision of tissue patterning is controlled by dynamical properties of gene regulatory networks. *Development* *148*, dev197566. <https://doi.org/10.1242/dev.197566>.
75. Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. *Nat. Commun.* *11*, 1061. <https://doi.org/10.1038/s41467-020-14853-5>.
76. Halfon, M.S. (2020). Silencers, Enhancers, and the Multifunctional Regulatory Genome. *Trends Genet.* *36*, 149–151. <https://doi.org/10.1016/j.tig.2019.12.005>.
77. Mukherjee, S., Luedeke, D.M., McCoy, L., Iwafuchi, M., and Zorn, A.M. (2022). SOX transcription factors direct TCF-independent WNT/β-catenin responsive transcription to govern cell fate in human pluripotent stem cells. *Cell Rep.* *40*, 111247. <https://doi.org/10.1016/j.celrep.2022.111247>.
78. Jiang, J., Cai, H., Zhou, Q., and Levine, M. (1993). Conversion of a dorsal-dependent silencer into an enhancer: evidence for dorsal corepressors. *EMBO J.* *12*, 3201–3209. <https://doi.org/10.1002/j.1460-2075.1993.tb05989.x>.
79. Kirov, N., Zhelnin, L., Shah, J., and Rushlow, C. (1993). Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in *Drosophila*. *EMBO J.* *12*, 3193–3199. <https://doi.org/10.1002/j.1460-2075.1993.tb05988.x>.
80. Kallunki, P., Edelman, G.M., and Jones, F.S. (1998). The neural restrictive silencer element can act as both a repressor and enhancer of L1 cell adhesion molecule gene expression during postnatal development. *Proc. Natl. Acad. Sci. USA* *95*, 3233–3238. <https://doi.org/10.1073/pnas.95.6.3233>.
81. Kehayova, P., Monahan, K., Chen, W., and Maniatis, T. (2011). Regulatory elements required for the activation and repression of the protocadherin-α gene cluster. *Proc. Natl. Acad. Sci. USA* *108*, 17195–17200. <https://doi.org/10.1073/pnas.1114357108>.
82. Koike, S., Schaeffer, L., and Changeux, J.P. (1995). Identification of a DNA element determining synaptic expression of the mouse acetylcholine receptor delta-subunit gene. *Proc. Natl. Acad. Sci. USA* *92*, 10624–10628. <https://doi.org/10.1073/pnas.92.23.10624>.
83. Weintraub, S.J., Prater, C.A., and Dean, D.C. (1992). Retinoblastoma protein switches the E2F site from positive to negative element. *Nature* *358*, 259–261. <https://doi.org/10.1038/358259a0>.
84. Bessis, A., Champiaux, N., Chatelin, L., and Changeux, J.P. (1997). The neuron-restrictive silencer element: A dual enhancer/silencer crucial for patterned expression of a nicotinic receptor gene in the brain. *Proc. Natl. Acad. Sci. USA* *94*, 5906–5911. <https://doi.org/10.1073/pnas.94.11.5906>.
85. Cunningham, T.J., Kumar, S., Yamaguchi, T.P., and Duester, G. (2015). *Wnt8a* and *Wnt3a* cooperate in the axial stem cell niche to promote mammalian body axis extension. *Dev. Dyn.* *244*, 797–807. <https://doi.org/10.1002/dvdy.24275>.
86. Ribes, V., Le Roux, I., Rhinn, M., Schuhbaur, B., and Dollé, P. (2009). Early mouse caudal development relies on crosstalk between retinoic acid, Shh and Fgf signalling pathways. *Development* *136*, 665–676. <https://doi.org/10.1242/dev.016204>.
87. Savory, J.G.A., Edey, C., Hess, B., Mears, A.J., and Lohnes, D. (2014). Identification of novel retinoic acid target genes. *Dev. Biol.* *395*, 199–208. <https://doi.org/10.1016/j.ydbio.2014.09.013>.
88. Chatagnon, A., Veber, P., Morin, V., Bedo, J., Triqueneaux, G., Sémon, M., Laudet, V., d'Alché-Buc, F., and Benoît, G. (2015). RAR/RXR binding dynamics distinguish pluripotency from differentiation associated cis-regulatory elements. *Nucleic Acids Res.* *43*, 4833–4854. <https://doi.org/10.1093/nar/gkv370>.
89. Gillespie, R.F., and Gudas, L.J. (2007). Retinoic Acid Receptor Isotype Specificity in F9 Teratocarcinoma Stem Cells Results from the Differential Recruitment of Coregulators to Retinoic Acid Response Elements. *J. Biol. Chem.* *282*, 33421–33434. <https://doi.org/10.1074/jbc.M704845200>.

90. Berenguer, M., Meyer, K.F., Yin, J., and Duester, G. (2020). Discovery of genes required for body axis and limb formation by global identification of retinoic acid–regulated epigenetic marks. *PLoS Biol.* *18*, e3000719. <https://doi.org/10.1371/journal.pbio.3000719>.
91. Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A., and Krumlauf, R. (1994). Role of a Conserved Retinoic Acid Response Element in Rhombomere Restriction of *Hoxb-1*. *Science* *265*, 1728–1732. <https://doi.org/10.1126/science.7916164>.
92. Kumar, S., and Duester, G. (2014). Retinoic acid controls body axis extension by directly repressing *Fgf8* transcription. *Development* *141*, 2972–2977. <https://doi.org/10.1242/dev.112367>.
93. Ghyselinck, N.B., and Duester, G. (2019). Retinoic acid signaling pathways. *Development* *146*, dev167502. <https://doi.org/10.1242/dev.167502>.
94. Kumar, S., Cunningham, T.J., and Duester, G. (2016). Nuclear receptor corepressors *Ncor1* and *Ncor2* (*Smrt*) are required for retinoic acid-dependent repression of *Fgf8* during somitogenesis. *Dev. Biol.* *418*, 204–215. <https://doi.org/10.1016/j.ydbio.2016.08.005>.
95. Urvalek, A.M., and Gudas, L.J. (2014). Retinoic Acid and Histone Deacetylases Regulate Epigenetic Changes in Embryonic Stem Cells. *J. Biol. Chem.* *289*, 19519–19530. <https://doi.org/10.1074/jbc.M114.556555>.
96. Patel, N.S., Rhinn, M., Semprich, C.I., Halley, P.A., Dollé, P., Bickmore, W.A., and Storey, K.G. (2013). FGF Signalling Regulates Chromatin Organisation during Neural Differentiation via Mechanisms that Can Be Uncoupled from Transcription. *PLoS Genet.* *9*, e1003614. <https://doi.org/10.1371/journal.pgen.1003614>.
97. Semprich, C.I., Davidson, L., Amorim Torres, A., Patel, H., Briscoe, J., Metzis, V., and Storey, K.G. (2022). ERK1/2 signalling dynamics promote neural differentiation by regulating chromatin accessibility and the polycomb repressive complex. *PLoS Biol.* *20*, e3000221. <https://doi.org/10.1371/journal.pbio.3000221>.
98. Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T. I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* *441*, 349–353. <https://doi.org/10.1038/nature04733>.
99. Kraft, K., Yost, K.E., Murphy, S.E., Magg, A., Long, Y., Corces, M.R., Granja, J.M., Wittler, L., Mundlos, S., Cech, T.R., et al. (2022). Polycomb-mediated genome architecture enables long-range spreading of H3K27 methylation. *Proc. Natl. Acad. Sci. USA* *119*, e2201883119. <https://doi.org/10.1073/pnas.2201883119>.
100. Kreimer, A., Ashuach, T., Inoue, F., Khodaverdian, A., Deng, C., Yosef, N., and Ahituv, N. (2022). Massively parallel reporter perturbation assays uncover temporal regulatory architecture during neural differentiation. *Nat. Commun.* *13*, 1504. <https://doi.org/10.1038/s41467-022-28659-0>.
101. Hu, Y., Kazenwadel, J., and James, R. (1993). Isolation and characterization of the murine homeobox gene *Cdx-1*. Regulation of expression in intestinal epithelial cells. *J. Biol. Chem.* *268*, 27214–27225. [https://doi.org/10.1016/S0021-9258\(19\)74240-9](https://doi.org/10.1016/S0021-9258(19)74240-9).
102. Sawada, S. (1994). A lineage-specific transcriptional silencer regulates CD4 gene expression during T lymphocyte development. *Cell* *77*, 917–929. [https://doi.org/10.1016/0092-8674\(94\)90140-6](https://doi.org/10.1016/0092-8674(94)90140-6).
103. Siu, G., Wurster, A.L., Duncan, D.D., Soliman, T.M., and Hedrick, S.M. (1994). A transcriptional silencer controls the developmental expression of the CD4 gene. *EMBO J.* *13*, 3570–3579. <https://doi.org/10.1002/j.1460-2075.1994.tb06664.x>.
104. Clark, E., and Peel, A.D. (2018). Evidence for the temporal regulation of insect segmentation by a conserved sequence of transcription factors. *Development* *145*, dev155580. <https://doi.org/10.1242/dev.155580>.
105. Copf, T., Schröder, R., and Averof, M. (2004). Ancestral role of *caudal* genes in axis elongation and segmentation. *Proc. Natl. Acad. Sci. USA* *101*, 17711–17715. <https://doi.org/10.1073/pnas.0407327102>.
106. Faas, L., and Isaacs, H.V. (2009). Overlapping functions of *Cdx1*, *Cdx2*, and *Cdx4* in the development of the amphibian *Xenopus tropicalis*. *Dev. Dyn.* *238*, 835–852. <https://doi.org/10.1002/dvdy.21901>.
107. Martin, B.L., and Kimelman, D. (2009). Wnt Signaling and the Evolution of Embryonic Posterior Development. *Curr. Biol.* *19*, R215–R219. <https://doi.org/10.1016/j.cub.2009.01.052>.
108. Morales, A.V., De La Rosa, E.J., and De Pablo, F. (1996). Expression of the *Cdx-B* homeobox gene in chick embryo suggests its participation in rostrocaudal axial patterning. *Dev. Dyn.* *206*, 343–353. [https://doi.org/10.1002/\(SICI\)1097-0177\(199608\)206:4<343::AID-AJA1>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-0177(199608)206:4<343::AID-AJA1>3.0.CO;2-I).
109. Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targetted correction of a mutant Hprt gene in mouse embryonic stem cells. *Nature* *330*, 576–578. <https://doi.org/10.1038/330576a0>.
110. Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* *8*, 2281–2308. <https://doi.org/10.1038/nprot.2013.143>.
111. Ewels, P.A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M.U., Di Tommaso, P., and Nahsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* *38*, 271. <https://doi.org/10.1038/s41587-020-0435-1>.
112. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* *9*, 676–682. <https://doi.org/10.1038/nmeth.2019>.
113. Tan, G., and Lenhard, B. (2016). TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* *32*, 1555–1556. <https://doi.org/10.1093/bioinformatics/btw024>.
114. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26. <https://doi.org/10.1038/nbt.1754>.
115. Truett, G.E., Heeger, P., Mynatt, R.L., Truett, A.A., Walker, J.A., and Warman, M.L. (2000). Preparation of PCR-Quality Mouse Genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *BioTechniques* *29*, 52–54. <https://doi.org/10.2144/00291bm09>.
116. Castro-Mondragon, J.A., Riudavets-Puig, R., Raulusevičute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., et al. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* *50*, D165–D173. <https://doi.org/10.1093/nar/gkab1113>.
117. Mirdita, M., Schütze, K., Moriawaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* *19*, 679–682. <https://doi.org/10.1038/s41592-022-01488-1>.
118. Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., et al. (2023). InterPro in 2022. *Nucleic Acids Res.* *51*, D418–D427. <https://doi.org/10.1093/nar/gkac993>.
119. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* *596*, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-CDX2 – clone 88	Abcam	Cat#ab157524; RRID: AB_2721036
Mouse monoclonal anti-CDX2 – clone 88	Gentaur	Cat#MU392A-5UC; RRID: AB_2923402
Rabbit monoclonal anti-CDX2	Abcam	Cat#ab76541; RRID: AB_1523334
Goat polyclonal anti-SOX1	R&D	Cat#AF3369; RRID: AB_2239879
Rat monoclonal I12 anti-HOXB4	DHSB	Cat#AB_2119288; RRID: AB_2119288
Goat polyclonal anti-BRACHYURY	R&D	Cat#AF2085; RRID: AB_2200235
Goat polyclonal anti-SOX2	R&D	Cat#AF2018; RRID: AB_355110
Goat polyclonal anti-TBX6	R&D	Cat#AF4744; RRID: AB_2200834
Rabbit polyclonal anti-RARB	Invitrogen	Cat#PA1-811; RRID: AB_2253602
Mouse monoclonal anti-LEF1	Millipore	Cat#17-604; RRID: AB_916350
anti-mouse AlexaFluor 488	Thermo Fisher	Cat#A21202; RRID: AB_101607
anti-rabbit AlexaFluor 488	Thermo Fisher	Cat#A21206; RRID: AB_2535792
anti-goat AlexaFluor 488	Thermo Fisher	Cat#A11055; RRID: AB_2534102
anti-rabbit AlexaFluor 568	Thermo Fisher	Cat#A10042; RRID: AB_2534017
anti-mouse AlexaFluor 568	Thermo Fisher	Cat#A10037; RRID: AB_11180865
anti-goat AlexaFluor 647	Thermo Fisher	Cat#A21447; RRID: AB_2535864
Anti-rat AlexaFluor 488	Thermo Fisher	Cat#A21208; RRID: AB_2535794
Anti-rat AlexaFluor 647	Thermo Fisher	Cat#A78947; RRID: AB_2910635
Bacterial and virus strains		
Electro-competent DH5alpha cells	Thermo Scientific	Cat#EC0112
Chemicals, peptides, and recombinant proteins		
Dulbecco's Modified Eagle Medium (DMEM) - Knock OUT	Gibco	Cat#10829-018
ESGRO Mouse LIF Medium	Merck Millipore	Cat#ESG1107
DMEM F12	Gibco	Cat#21331-020
Neurobasal medium	Gibco	Cat#21103-049
L-Glutamine	Gibco	Cat#25030-024
Trypsin-EDTA 0.05%	Gibco	Cat#25300-054
StemPro Accutase	Gibco	Cat#A11105-01
BSA	Sigma	Cat#A7979
N2 Supplement	Gibco	Cat#17502-001
B27 Supplement	Gibco	Cat#A35828-01
B27 Supplement, minus vitamin A	Gibco	Cat#A3353501
Gelatin	Sigma	Cat#G1393-20ML
b-mercaptoethanol	Gibco	Cat#21985-023
Recombinant bFGF	PeproTech	Cat#100-18B-10uG
CHIR99021	Axon	Cat#1386
All-trans-Retinoic Acid (ATRA or RA)	Sigma	Cat#R2625-50MG
Live/dead fixable blue dead cells stain kit for UV excitation	Thermo Fisher	Cat#L34961A
RNAse-free DNase I	Qiagen	Cat#79254
Superscript III reverse transcriptase	Thermo Fisher	Cat#18080-051
PowerUp SYBR-Green Mastermix	Thermo Fisher	Cat#A25742
Puromycin	Gibco	Cat#A11138-03
ES Fetal Bovine Serum (FBS)	Pan Biotech	Cat#P30-2602; Lot P200304ES

(Continued on next page)

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Penicillin/Streptomycin	Gibco	Cat#15140122
Non-essential amino acids	Gibco	Cat#11140-050
Phosphate Buffer Saline (PBS)	Pan Biotech	Cat#P0436500
Paraformaldehyde (PFA)	ThermoScientific	Cat#28908
Triton-X100	Sigma	Cat#T8787-250ML
ProLong Gold antifade reagent	Invitrogen	Cat#P36930
GlutaMAX	Gibco	Cat#35050-061
Di(N-succinimidyl) glutarate (DSG)	Sigma	Cat#80424-5MG-F
Glycine	Sigma	Cat#G7126-1KG
PBS with CaCl ₂ /MgCl ₂	Sigma	Cat#D8662
DMSO	Sigma	Cat#D2650-100ML
Dynabeads Protein G	Thermo Fisher	Cat#10004D
SDS	Thermo Fisher	Cat#BP1311-200
EDTA	Corning	Cat#46-034-CI
EGTA	Thermo Fisher	Cat#J60767.AD
Hepes	Sigma	Cat#H0887-100ML
NaCl 5M solution	Lonza	Cat#51202
Protease Inhibitor cocktail (PI)	Sigma	Cat#P8340-5ML
Sodium Bicarbonate (NaHCO ₃)	Sigma	Cat#S6297-250G
Sodium Deoxycholate	Sigma	Cat#30970-25G
NP-40	Sigma	Cat#I8896-100ML
LiCl	Sigma	Cat#L4408-100G
PureLink RNase A (20mg/mL)	Invitrogen	Cat #12091021
Proteinase K, ChIP-grade	Thermo Fisher	Cat#26160
Critical commercial assays		
QIAGEN RNeasy	Qiagen	Cat#74106
PureLink genomic DNA extraction Kit	Invitrogen	Cat#K182002
Mouse Embryonic Stem Cell nucleofactor kit	Lonza	Cat#VPH-1001
Qiaquick PCR Purification Kit	Qiagen	Cat#28106
Experimental models: Cell lines		
Mus musculus (Male): HM1 <i>WT</i>	Doetschman et al. ¹⁰⁹	N/A
Mus musculus (Male): HM1 Δ <i>lee</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>5'</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>Tee</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>P1</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>P2</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>Intron1</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>Prom</i>	This study	N/A
Mus musculus (Male): HM1 Δ <i>Rarb</i>	This study	N/A
Mus musculus (Male): HM1 <i>RASwitch</i>	This study	N/A
Oligonucleotides		
qPCR primers used in this study, see Table S1	This study	N/A
sgRNA sequence, genotyping primers and sequencing primers used to generate cell lines see Table S2	This study	N/A
sgRNA sequence used for in vivo targeting of the P2 CRE, see Table S2	This study	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
PX459; SpCas9-2A-Puro and single guide RNA	Ran et al. ¹¹⁰	Cat#62988 (Addgene)
Software and algorithms		
nf-core chipseq pipeline	https://nf-co.re/chipseq Ewels et al. ¹¹¹	https://nf-co.re/chipseq https://doi.org/10.5281/zenodo.3240506
nf-core rnaseq pipeline	https://nf-co.re/rnaseq Ewels et al. ¹¹¹	https://doi.org/10.5281/zenodo.1400710
Fiji	Schindelin et al. ¹¹²	https://imagej.net/software/fiji/
TFBStools version 1.40.0	Tan and Lenhard ¹¹³	http://bioconductor.org/packages/TFBStools/
IGV	Robinson et al. ¹¹⁴	https://igv.org/
Other		
Donkey serum	Abcam	Cat#ab7475
35 mm high glass bottom imaging dish	Ibidi	Cat#81158
CellBind 6-well plate	Corning	Cat#3335
35 mm CellBind dish	Corning	Cat#3294
32 mm coverslips, no. 1.5 thickness	SLS	Cat#631-0162
100 mm CellBind dish	Corning	Cat#3296
Cell scrapper	VWR	Cat #734-2602
RNase-free tube	Ambion	Cat#AM12450
Diagenode tube for sonication	Diagenode	Cat#C30010010-300

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cells lines

All mouse ESC lines were cultured at 37 °C with 5% CO₂, and were visually inspected on a daily basis. All mouse ESC lines were subject to mycoplasma testing on a monthly basis. All ESC lines used were derived from the XY HM1 line (129/Ola strain),¹⁰⁹ which was used as the WT control. $\Delta Intron1$, $\Delta 5'$, Δlee , $\Delta Prom$, ΔTee , $\Delta P1$ and $\Delta P2$ lines were generated by electroporating pairs of CRISPR targeted to both extremities of the regions of interest (Table S2). After puromycin selection (at a concentration of 1.5 ug/mL), 10 clones were picked and expanded. gDNA was extracted using the PureLink kit (PureLink™ Genomic DNA) according to the manufacturer's instructions. Clones were then genotyped (Table S2), validated for the deletion of the targeted region in both alleles by DNA sequencing. For most of the generated KO cell lines, we obtained similar results with a second clone ($\Delta Intron1$, ΔTee , $\Delta Prom$, $\Delta P2$, Δlee).

The *RAswitch* and the $\Delta Rarb$ lines were created using HDR recombinant oligos electroporated with the sgRNA guide (Table S2) and the Cas9 protein (#1081058) supplemented with the Alt-R Cas9 electroporation enhancer (#1075915) into HM1 cells. After recovery using the Alt-R HDR EnhancerV2 (#10007910), 10 clones were picked, expanded, and, genotyped (Table S2) and validated by DNA sequencing in a similar method as described above.

ESC culture and differentiation

All mouse ESCs were expanded on mitotically inactivated mouse embryonic fibroblasts (feeders) in ESC medium (DMEM knockout medium supplemented with 1.000U/ml LIF, 10% cell-culture-validated foetal bovine serum, and 2mM L-Glutamine). Data were obtained using low-passage cells (i.e. passaged no more than 10 times after thawing for a total number of 32 passages since derivation).

To differentiate mESCs into neural or paraxial presomitic mesoderm progenitors, ESCs were differentiated as previously described.^{24,55} Briefly, ESCs were dissociated with 0.05% trypsin, and plated on tissue-culture-treated plates for two sequential 20 minutes (mins) periods in ESC medium to separate them from their feeder layer cells, which adhere to the plastic. To start the differentiation, cells remaining in the supernatant were pelleted by centrifugation, counted, and resuspended in N2B27 medium containing 10 ng/ml bFGF, and 40 000 cells per 35 mm gelatin-coated CellBIND dish or 6-well plate (Corning) were plated. N2B27 medium contained a 1:1 ratio of DMEM/F12:Neurobasal medium (Gibco) supplemented with 0.5% N2 (Gibco), 1% B27 (Gibco), 2mM L-glutamine (Gibco), 40mg/ml BSA (Sigma), and 0.1mM 2-mercaptoethanol.

To generate Epiblast-like (EpiL) or Anterior Epiblast-like (AEpiL) cells, the cells were grown respectively for 2 and 3 days in N2B27 + 10 ng/ml bFGF. To generate Caudal Epiblast-like (CEpiL) cells, cells were cultured with N2B27 + 10 ng/ml bFGF for 2 days, then N2B27 + 10 ng/ml bFGF + 5 μM CHIR99021 for a further day. CEpiL cells were differentiated to spinal cord neural progenitors by continuing the differentiation up to day 5 in N2B27 media containing 10nM all-trans-retinoic acid (ATRA, 10nM). In the experiments

presented in Figure 4C, CEpiL were either exposed to N2B27 alone for one day or exposed to concentrations of ATRA ranging from 1nM to 1 μ M.

To generate paraxial mesoderm progenitors, CEpiL cells (generated as described above) were exposed to 5 μ M CHIR (GSK3 β inhibitor) for a further 2 days. For the paraxial mesoderm differentiation, a CEpiL were grown in N2B27 using B27 devoid of Vitamin A as previously described.²⁴ Media was refreshed every day from day 2-5 for all experiments.

Details of key compounds are provided in the STAR Methods.

CRISPR mutant embryos

For generation of the *P2- Δ Rarb* CRISPR mutants, gRNA sequences (Table S2) were ordered as oligonucleotides (Integrated DNA Technologies) together with recombinant Cas9 protein (Alt-R™ S.p. Cas9 Nuclease V3; 1081058). The sgRNA (at 25 ng/ μ l) and the Cas9 (at 75 ng/ μ l) were combined and microinjected into the pronuclei of one-cell embryos in two separate rounds of injection (Figure 4).

One-cell embryos obtained by super-ovulating 10 C57Bl/6 females with 50 IU of PMSG 48h hours before mating and with 50 IU of HCG on the day of mating were mated with C57Bl/6 stud males. 24h after mating embryos were harvested, cleaned and placed in culture media (KSOM) at 37°C. Each zygote was then microinjected into the pronuclei with the CRISPR/Cas9 complex. Microinjected zygotes were transferred back into recipient females (B6CBAF1; previously mated and plugged by vasectomised CD1 males) by embryo transfer procedure at 0.5dpc. All females were monitored daily in a Biological Support Unit. The recipient females were humanely killed and embryos were harvested at 9.5dpc.

After dissection and collection of amniotic tissue for genotyping, embryos were fixed in 4% paraformaldehyde in PBS for 90 mins at 4°C under gentle agitation, followed by two washes in PBS.

Embryos were genotyped using the HotSHOT DNA¹¹⁵ extraction protocol. Briefly, amnions were incubated at 95°C for 30 mins in 25 μ L alkaline lysis buffer before addition of 25 μ L neutralizing buffer and storage at 4°C. The *Intron1* fragment was amplified by PCR using the primers used for the genotyping of the Δ *Intron1* cell line (Table S2). After purification using the Qiagen PCR purification kit, the fragment was sequenced using the primer outlined in Table S2.

All animal procedures were performed by certified staff in the Imperial College London Central Biomedical Services Facility and all experiments were performed with ethical approval in accordance with the Animal (Scientific Procedures) Act 1986 with ethical approval under the UK Home Office project license PP2904879. Animals were housed in a 10-hour light, 14-hour dark cycle with access to food and water ad libitum in individually ventilated cages. The temperature was maintained at 21-24°C and 45-65% humidity. No distinction was made between male and female embryos during analysis.

METHOD DETAILS

Immunofluorescence on cells

Cells were washed in PBS and fixed in 4% paraformaldehyde in PBS for 30 mins at 4°C, followed by three washes in PBS. Primary antibodies (STAR Methods) were applied overnight at 4°C diluted in filtered blocking solution (2% BSA diluted in PBST – 0.1% Triton X-100 diluted in PBS). Cells were washed for 5 mins three times in PBST and incubated with secondary antibodies (STAR Methods) at room temperature, for 90 mins. Cells were washed for 5 mins three times in PBST, incubated with DAPI for 15 mins in PBS and washed twice before mounting with a glass coverslip using Prolong Gold (Invitrogen) or kept in PBS for further imaging.

Cells were imaged on an inverted SP5 or upright SP5 II confocal microscope (Leica). Z stacks were acquired using the Leica LAS AF software and represented as maximum intensity projections using ImageJ software. The same settings were applied to all images. Images presented in Figures 1G and S1F are representative images of a minimum of three biological replicates.

Flow cytometry

Cells were washed in PBS and dissociated with Accutase (Gibco). Once detached, cells were collected, washed with PBS, and pelleted. Cells were resuspended in PBS supplemented with live dye (1/1000, Thermo Fisher) and kept in dark at 4°C for 30 mins. Cells were pelleted, washed in PBS, pelleted, and resuspended in 4% paraformaldehyde in PBS. Following 15 mins incubation at 4°C, cells were centrifuged, resuspended in PBS, and stored at 4°C for future analysis.

On the day of flow cytometry, cells were transferred for staining in U-bottom 96-well plates. Samples were pelleted and resuspended in 50 μ l block media (2% BSA diluted in PBST). After 30 mins incubation at room temperature in the platform rocker, antibodies were added to the sample and incubated overnight at 4°C on a platform rocker. Details of primary and secondary antibodies are described in the STAR Methods. Cells were pelleted for 4 mins, washed in PBST, pelleted, and incubated in 50 μ l PBST supplemented with secondary antibodies (concentration: 1/500) in the dark for 2h at room temperature in the platform rocker. One additional wash was performed before acquisition on a SymphonyA3 (BD Biosciences) using FACSDiva. Analysis was performed using FlowJo.

RNA extraction, cDNA synthesis and RT-qPCR analysis

RNA used for real time quantitative PCR (RT-qPCR) was extracted from cells using a QIAGEN RNeasy kit in RLT buffer, following the manufacturer's instructions. Extracts were digested with DNase I to eliminate genomic DNA.

First-strand cDNA synthesis was performed using Superscript III (Invitrogen) using random hexamers and was amplified using PowerUp SYBR-Green Mastermix (Applied Biosystems). RT-qPCR was performed using the Applied Biosystems QuantStudio

Real Time PCR system and analysed with Applied Biosystems QuantStudio 12K Flex software. PCR primers were designed using the online PrimerBLAST design tool and validated (standard curve and melting curve) or taken from previously published papers. Primer sequences are detailed in [Table S1](#). Two technical replicates were obtained for each sample and averaged before normalization and statistical analysis. Relative expression values for each gene were calculated by normalization against β -actin, using the delta-delta CT method. RT-qPCR analysis was performed on samples obtained from a minimum of three independent experiments for every primer pair analysed.

Embryo wholemount immunofluorescence

Embryos were permeabilized in 0.5% Triton X-100 diluted in PBS for 30 mins at room temperature under gentle agitation. After permeabilization, embryos were incubated in filtered block media (2% BSA and 4% donkey serum diluted in PBST) at room temperature for 2 hrs under gentle agitation. Primary antibodies ([STAR Methods](#)) were applied overnight at 4°C diluted in filtered block media under gentle agitation. The following morning, embryos were washed for 2 hrs in PBST, 4–5 times, at room temperature under gentle agitation and incubated in filtered block media overnight at 4°C under gentle agitation. Secondary antibodies ([STAR Methods](#)) were applied diluted in PBST (1/500) at room temperature for 90 mins in the dark under gentle agitation. After 10 mins PBST washes, embryos were incubated with DAPI (1/1000) in PBST in the dark under gentle agitation at room temperature for 30 mins.

After PBST washes, embryos were mounted in 1.5% low-melt agarose in p35 Ibidi plates before imaging using an inverted Leica DLS. Z stacks were acquired using the Leica LAS AF software and represented as maximum intensity projections using ImageJ software.

ChIP-qPCR

Adherent cells were washed three times with PBS, fixed with gentle agitation for 45 mins at room temperature with fresh 2mM di(N-succinimidyl) glutarate in PBS+ (DPBS with $\text{CaCl}_2/\text{MgCl}_2$), washed an additional three times with PBS+, then fixed for 10 mins at room temperature with 1% molecular-biology-grade paraformaldehyde in PBS+. Fixation was quenched by addition of 250 mM glycerine for 5 mins, followed by three additional washing with PBS+. Plates were cooled, and cells were scraped into tubes in a low volume of PBS+ 0.02% Triton X-100 and pelleted by centrifugation at 100g for 5 mins at 4°C before snap freezing in liquid nitrogen and storing at -80°C . Approximately 5×10^6 cells were transferred to a Diagenode TPX tube and resuspended in ice-cold shearing buffer (1% Triton-X 100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES/pH 7.6) containing 0.3% SDS and protease inhibitors. Chromatin was sheared using a Diagenode Bioruptor *Plus*: 20 cycles of 30sec on/30sec off on the high setting, and lysates were then diluted to 0.15% SDS and cleared by centrifugation at 14,000 RPM for 5 mins at 4°C. Then, 1/20 of the chromatin from $\sim 1 \times 10^7$ cells was set aside and frozen for subsequent use as input control, and the remainder was incubated overnight at 4°C under rotation with 100 μl of protein G dynabeads pre-loaded for 4 hrs at room temperature with 5 μg of ChIP antibodies diluted in shearing buffer containing 0.15% SDS. Beads were magnetically immobilized, unbound supernatant was discarded and beads were sequentially washed under rotation twice with Wash Buffer 1 (0.1% SDS, 0.1% sodium deoxycholate, 1% Triton-X 100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES/pH 7.6), once with Wash Buffer 2 (0.1% SDS, 0.1% sodium deoxycholate, 1% Triton-X 100, 0.5 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES/pH 7.6), once with Wash Buffer 3 (0.5% sodium deoxycholate, 0.5% NP-40, 0.25 M LiCl, 1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES/pH 7.6) and twice with Wash Buffer 4 (1 mM EDTA, 0.5 mM EGTA, 20 mM HEPES/pH 7.6) for 5 mins each, magnetically capturing beads between each wash. Chromatin was eluted from beads by incubating twice at 65°C for 10 mins in 100 μl elution buffer (1% SDS, 0.1 M NaHCO_3) on a shaking heat block, capturing beads between each elution step and then pooling each eluted fraction. Input samples were made up to 200 μl with elution buffer, 6.4 μl of 5 M NaCl was added to each input or immunoprecipitated sample, and all samples were de-crosslinked overnight at 65°C. Samples were incubated for 2 hrs at 37°C with 0.2 $\mu\text{g}/\text{ml}$ PureLink RNase A, then supplemented with 5 mM EDTA and incubated for an additional 2 hrs at 45°C with 0.2 $\mu\text{g}/\text{ml}$ proteinase K before purifying DNA with Qiagen PCR clean-up columns.

Immunoprecipitated DNA was analysed by qPCR using the Applied Biosystems QuantStudio Real Time PCR system and analysed with Applied Biosystems QuantStudio 12K Flex software. PCR primers were designed using the online PrimerBLAST design tool and validated (standard curve and melting curve) or taken from previously published papers. Primer sequences are detailed in [Table S1](#). Enrichment values for each region of interest were calculated by normalization against a no antibody control. Each ChIP analysis was repeated in at least in three independent experiments.

ChIP-seq, ATAC-seq and mRNA-seq data and processing

ATAC-seq data from day 2 epiblast-like (D2-EpiL), day 3 caudal epiblast-like (D3-CEpiL), day 4 (D4) and day 5 (D5) spinal cord (SC) were obtained from Metzis et al.⁵⁵ (accession number E-MTAB-6337). Pseudo-bulk ATAC-seq generated from mouse embryo 10x multiome experiments were obtained from Argelaguet et al.⁵⁶ (accession number GSE205117). ChIP-seq data from naïve mouse ESCs, caudal epiblast-like cells and *Sox2* over-expressing caudal-epiblast-like cells ([Figures 3A](#) and [S4E](#)) were obtained from Blassberg et al.⁵⁷ (accession number GSE162774). ChIP-seq data against histone marks were obtained from Yang et al.⁶² (accession number GSE98101) for embryos and from Mazzoni et al.³³ (accession number GSE39433) for differentiated mouse ESCs. ChIP-seq data against HDAC3 and p300 were respectively obtained from Zyllicz et al.⁶¹ (accession number GSE116480) and Amin et al.⁹ (accession number GSE84899). ChIP-seq data against H3K27me3, SUZ12 and JARID2 in mESCs were obtained from Kanellopoulou et al.⁶³ (accession number GSE60397). Big Wig tracks were visualised using IGV.¹¹⁴

ATAC-seq signals were overlapped with the regulatory elements: *lee*, *Intron1*, *Prom*, 5', and *Tee*. The overall signal was calculated as the mean signal within the region, and the mean signal was normalised to 1 over the intensity of all analysed regions. Relative fold change to the final condition (D5SC for *in vitro* samples and SC for *in vivo* samples) was calculated for each region.

For mRNA-seq data³¹ (accession number E-MTAB-2268), the nf-core/rnaseq pipeline (version 2.0)¹¹¹ was used with default parameters. Briefly, the pipeline performs quality control, trimming (using TrimGalore!), (pseudo-)alignment (using Salmon), and produces a gene expression matrix. All data were processed relative to the mouse UCSC mm10 genome (UCSC) downloaded from AWS iGenomes (<https://github.com/ewels/AWS-iGenomes>).

Identification of Transcription Factor Binding Sites (TFBS)

To identify transcription factor binding sites (TFBS) in genomic sequences, we used a comprehensive bioinformatics approach with publicly available databases and specialised software tools. The matrices representing the binding preferences of transcription factors were obtained from the JASPAR2022 database.¹¹⁶

To detect TFBS instances within our specific sequences, we used TFBSstools version 1.40.0, as described in Tan and Lenhard.¹¹³ We set TFBSstools to search for matches to the JASPAR matrices within our sequences, specifying an identity match threshold of 80%. The threshold was selected to identify low-affinity sites for TFs of interest in our specified regions. Then, we filtered our list to remove any non-expressed transcription factors based on mRNA-seq data generated *in vitro* in EpiL, CEpiL, D4 SC and D5 SC.³¹

The conservation profiles of *P1* and *P2*, as well as the multiple sequence alignment of the RARB motif in *P2* were visualised in the UCSC genome browser. The tracks used are available in the following session: https://genome-euro.ucsc.edu/s/da_bar/cdx2_control_conservation.

Prediction of Protein-Protein Interaction Complexes

To further investigate the functional implications of identified TFBS instances, we aimed to predict the structures of protein-protein complexes involving our TFs of interest. To do this, we retrieved the complete TF protein sequences from InterPro, a comprehensive database of protein families, domains, and functional sites.¹¹⁷ We then used ColabFold version 1.3.0, which is an interface for AlphaFold-multimer program.¹¹⁸ AlphaFold-multimer is a state-of-the-art method for predicting protein complex structures, leveraging deep learning to estimate the three-dimensional arrangements of protein subunits.^{67,119} By inputting the TF protein sequences into ColabFold, we were able to obtain high-confidence predictions of their potential interactions and complex formations.

Experimental design

No statistical method was used to pre-determine sample size. No data were excluded from the analyses. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. For each experiment, data were obtained from a minimum of three independent experiments.

QUANTIFICATION AND STATISTICAL ANALYSIS

For all statistical analyses, data were obtained from a minimum of three independent experiments. Technical and biological replicates were pooled. The number of biological and technical replicates are indicated in Table S4. Bars denote mean \pm s.e.m and statistical significance was calculated using GraphPad Prism (GraphPad Software). Each dot represents a replicate. Details of the statistical analyses performed for each experiment are specified in Table S4. No methods were used to determine whether the data followed a normal distribution.