Tgfbr1 controls developmental plasticity between the hindlimb and external genitalia by remodeling their regulatory landscape

Anastasiia Lozovska[1], Artemis G. Korovesi[1], André Dias[1#], Alexandre Lopes[1], Donald A. Fowler[1], Gabriel G. Martins[1], Ana Nóvoa[1], and Moisés Mallo[1*]

1 Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

Current address: Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain.

* Lead author: mallo@igc.gulbenkian.pt

Key words: Tgfbr1, hindlimb, external genitalia, chromatin conformation, mouse

The hindlimb and external genitalia of present-day tetrapods are thought to derive from an ancestral common primordium that evolved to generate a wide diversity of structures adapted for efficient locomotion and mating in the ecological niche conquered by the species [1]. We show that despite long evolutionary distance from the ancestral condition, the early primordium of the mouse external genitalia preserved the capacity to take hindlimb fates. In the absence of *Tgfbr1*, the pericloacal mesoderm generates an extra pair of hindlimbs at the expense of the external genitalia. It has been shown that the hindlimb and the genital primordia share many of their key regulatory factors [2–5]. *Tgfbr1* controls the response to those factors acting in a pioneer-like mode to modulate the accessibility status of regulatory elements that control the gene regulatory networks leading to the formation of genital or hindlimb structures. Our work uncovers a remarkable tissue plasticity with potential implications in the evolution of the hindlimb/genital area of tetrapods, and identifies a novel mechanism for *Tgfbr1* activity that might also contribute to the control of other physiological or pathological processes.

The hindlimbs and the external genitalia both develop at the posterior end of the trunk as an integral component of the trunk to tail transition [6]. Formation of the external genitalia results from reciprocal interactions between the pericloacal mesoderm and the endodermal cloaca that generate the genital tubercle (GT), the precursor of the external genitalia, while also organizing the exit channels of the digestive and excretory systems with the formation of the rectus/anus, the bladder and the urethra [2,7–10]. The hindlimbs emerge from the somatic layer of the lateral plate mesoderm [11] close to the position of the developing GT. Although the mature hindlimb and external genitalia are anatomically very different, molecular studies indicate that they might be closely connected developmentally, as they share many of their regulatory processes [2–5]. Indeed, it has been reported that in some tetrapod species hindlimbs and genitals share a common primordium, which might represent the ancestral condition [1]. In mammals, it has been suggested that the GT became separated from the hindlimb by the posterior displacement of the cloaca relative to the hindlimb bud, that provides signals promoting genital fates to cells originating from the ventral and tail bud mesenchyme [1].

Genetic analyses showed that *Tgfbr1* plays a key role during the induction stage of these structures. *Tgfbr1* null mutant embryos fail to induce hindlimb or cloacal structures, as shown by the absence of early markers for these tissues [12]. Conversely, premature activation of Tgfbr1 signaling in the axial progenitors resulted in earlier induction of these structures, as estimated by gene marker expression and the more anterior position of the hindlimb buds, adjacent to the forelimbs [6]. Given the essential role of Tgfbr1 signaling in the induction of both hindlimb and GT primordia despite generating morphologically and functionally distinct structures, we wanted to understand whether it also played a role at later stages of hindlimb and/or GT development.

We used a conditional approach to inactivate *Tgfbr1* after the hindlimb and GT primordia were laid down relying on the *Cdx2-Cre*^{ERT} transgenic driver [6], a *Tgfbr1* null allele [12] and either *Tgfbr1*^{flox}, carrying LoxP sites flanking Exon 3 [13], or *Tgfbr1*$^{3ex3\text{-}flox}$, containing a triplicated Exon 3 and associated LoxP sites that codes for a hypomorphic receptor (Supplementary Fig. 1A-D). Experiments using either *Tgfbr1* floxed allele gave similar phenotypes, but *Tgfbr1*$^{3ex3\text{-}flox}$ turned out to facilitate the analysis of the conditional *Tgfbr1* mutants, as *Tgfbr1*$^{3ex3\text{-}flox/-}$*:Cdx2Cre*$^{ERT+/0}$ embryos required half of the dose of tamoxifen than *Tgfbr1*$^{flox/-}$*:Cdx2Cre*$^{ERT+/0}$ embryos to obtain the same

mutant phenotype, thus reducing the frequency of tamoxifen-derived embryo miscarriage. Our tamoxifen treatment scheme produced *Tgfbr1* inactivation in the tissues caudal from the forelimb bud, as confirmed by RT-qPCR analyses on RNA extracted from the posterior part of E10,5 embryos, including the hindlimbs (Supplementary Fig. 1E). Tamoxifen-treated embryos and fetuses with either the *Tgfbr1*$^{flox/-}$*:Cdx2Cre*$^{ERT+/0}$ or the *Tgfbr1*$^{3ex3\text{-}flox/-}$*:Cdx2Cre*$^{ERT+/0}$ genotype will be hereafter referred to as *Tgfbr1-cKO*.

The most prominent feature of *Tgfbr1-cKO* mutants was a duplication of their hindlimbs [4/5 at embryonic age (E)16.5] (Fig. 1A, B). The fifth fetus analyzed at this stage lacked overt hindlimb duplication but still contained small protrusions posterior to the hindlimbs, a phenotype likely derived from less efficient recombination. The limb identity of the duplicated structures was confirmed by the presence of skeletal structures that, although variable in morphology, were clearly identified as belonging to limbs (Fig. 1C; Supplementary Video 1). Another prominent phenotype of *Tgfbr1-cKO* fetuses was the presence of an omphalocele (Fig. 1B). This phenotype, but not the duplicated hindlimbs, was also observed in *Tgfb2*/*Tgfb3* double mutants [14]. This indicates both that Tgfbr1 might mediate Tgfb2 and Tgfb3 activity in the body wall and that different ligands should trigger Tgfbr1 signaling in the body wall and the pericloacal/hindlimb mesoderm. *Tgfbr1-cKO* mutants lacked kidneys (Fig. 1D-E'; Supplementary Videos 2 and 3), a trait also observed in mutant fetuses for *Gdf11* [6,15], another of the ligands shown to activate Tgfbr1 signaling during embryonic development [16]. The cloacal-derived tissues were also strongly compromised in these mutant embryos (Fig. 1D-E'; Supplementary Videos 2 and 3). The gut failed to generate normal rectal-anal structures. Instead, the caudal end of the intestinal tube merged with a structure that could be identified as a small bladder connected to a hypomorphic urethra, thus resembling a persistent cloaca. The external genitalia were also virtually absent from *Tgfbr1-cKO* fetuses. These phenotypes are likely to result from interference with the functional interactions between the cloacal endoderm and the pericloacal mesoderm driving normal morphogenesis of the cloacal region [2,3,5,9,17]. Indeed, endodermal malformations in this area were already evident in *Tgfbr1-cKO* mutants at E10.5-E11.5, including an enlarged cloaca lacking features associated with normal septation, and often containing protrusions entering the intra-cloacal lumen (Supplementary Fig. 2D, D').

104 Morphological alterations in the hindlimb/pericloacal area of *Tgfbr1-cKO* embryos
105 were already visible at mid-gestation (Fig. 2). *Fgf8* expression in E10.0 embryos
106 revealed a posterior and medial expansion of the apical ectodermal ridge (AER) of the
107 hindlimb bud almost reaching the ventral midline of the embryo (Fig. 2A, A'),
108 suggesting an extension of the hindlimb field into the peri-cloacal mesoderm of *Tgfbr1-*
109 *cKO* embryos. This was associated with several expression changes in the cloacal
110 region. For instance, hindlimb specific markers, like *Pitx1* and *Lin28a* were ectopically
111 activated in the pericloacal region (Fig. 2B-C'). Conversely, *Tbx5* expression, normally
112 observed in the pericloacal mesoderm but not in the hindlimb, could not be detected
113 in this region of *Tgfbr1-cKO* embryos (Fig 2G, G'), and *Fgf8* expression was absent
114 from the cloacal endoderm of the mutant embryos. Together these observations
115 suggest a change of identity in the pericloacal mesoderm, becoming incorporated into
116 the hindlimb field instead of entering its normal GT developmental fate. Additional
117 observations were consistent with this hypothesis. At E11.0 when the presence of two
118 independent buds started to become apparent, thus anticipating the duplicated
119 hindlimb, the more caudal of these buds seemed to emerge from the pericloacal area.
120 The limb identity of this caudal bud was further supported by the expression of limb
121 dorso-ventral patterning markers, including *En1* and *Lmx1b* [18,19], which followed
122 typical limb patterns, although they were not as well defined as in control hindlimb
123 buds (Supplementary Fig. 2A-b'). Expression of genes active in both the hindlimb and
124 the pericloacal region/GT, was also consistent with the pericloacal tissues taking a
125 hindlimb fate. For instance, *Wnt5a*, which in wild type embryos is expressed in limb
126 buds beneath the AER and in the pericloacal region and GT, was not detected next to
127 the cloacal endoderm of *Tgfbr1-cKO* embryos, being instead observed beneath the
128 AER in the two buds of the extended hindlimb domain (Fig. 2E, E'). Also, *Hand2* and
129 *Bmp4* were expressed in the mutants as a continuous domain with the more posterior
130 hindlimb bud invading the peri-cloacal tissue, instead of showing the separation
131 between the hindlimb and GT region observed in wild type embryos (Fig. 2F, F';
132 Supplementary Fig. 2C, C'). Additional support for the conversion of pericloacal
133 mesoderm into hindlimb identity was provided by genes showing dynamic expression
134 patterns in these structures. In wild type embryos, *Isl1* is initially activated in the
135 hindlimb and GT primordia, becoming later downregulated in the hindlimb while
136 keeping strong expression in the GT [20](Fig 2H). In *Tgbfr1-cKO* mutants *Isl1* showed
137 residual expression next to the expanded endodermal cloaca (note that *Isl1* is also

138 expressed in the cloacal endoderm [21]) but became downregulated in the adjacent
139 mesoderm in the area likely to generate the extra hindlimb bud (Fig 2H'). Conversely,
140 *Fgf10*, an essential gene for limb induction [22], whose expression in the GT is observed
141 only at later stages [23], was strongly activated already at E10,5 following limb-like
142 patterns, reaching mesodermal tissue adjacent to the endoderm (Fig. 2D, D').
143 It has been shown that regulation of GT development relies to a large extent on signals
144 from the cloacal endoderm, with Shh playing a central role in this process [17,24]. The
145 observation that *Wnt5a*, a known Shh target in the GT [17], was not activated in the
146 pericloacal mesoderm of *Tgfbr1-cKO* despite strong endodermal *Shh* expression (Fig.
147 3A, A', a, a'; Supplementary Fig. 2D, D', d, d'), suggested that either this mesenchyme
148 became refractory to Shh activity or that it changed its response profile. The
149 expression of the Shh target *Gli1* was consistent with the second possibility, as it
150 showed equivalent patterns in the pericloacal mesenchyme in both wild type and
151 *Tgfbr1-cKO* embryos (Fig. 3B, B'), thus indicating that this tissue still responds to the
152 Shh signal from the endoderm.
153 We explored the mechanisms modulating the pericloacal mesoderm response to Shh
154 (and most likely also to other regulatory factors) by comparing the genomic
155 accessibility profiles in the extra hindlimbs and pericloacal region of *Tgfbr1-cKO* with
156 those of hindlimb buds and GT of control embryos by ATAC-seq [25]. The global
157 chromatin accessibility patterns in the extra hindlimb of the mutant embryos were
158 closer to those obtained from the control GT than to those from control hindlimbs (Fig.
159 3C; Supplementary Fig. 3A), consistent with the idea of its pericloacal origin. Despite
160 this, we identified two changes in accessibility patterns fitting with their potential
161 involvement in the control of cell fate decisions of the pericloacal mesoderm. Pattern
162 1 included 4922 regions that in control tissues were accessible in the GT but not in the
163 hindlimb, and that lost accessibility in the tissues from *Tgfbr1-cKO* embryos (Fig. 3D,
164 E). These elements could represent enhancers involved in the activation of genes
165 expressed in the pericloacal region of wild type embryos but inactive in the *Tgfbr1-*
166 *cKO* mutants. Pattern 2 contains 2013 chromatin regions representing elements that
167 in control embryos were accessible in the hindlimb but not in the GT and became
168 accessible in the extra hindlimb of the mutant embryos (Fig 4A, A'). These elements
169 could be involved in activating the limb program in the pericloacal mesoderm of *Tgfbr1-*
170 *cKO* embryos.

171 We identified pattern 1 elements within the *Tbx5, Isl1* and *Wnt5a* genomic regions,
172 three of the pericloacal mesoderm markers down-regulated in the pericloacal region
173 of *Tgfbr1-cKO* mutants (Fig. 3F-G'; Supplementary Fig. 3B, B'). These elements
174 mapped to highly conserved areas, suggesting their possible involvement in regulatory
175 processes. When tested using a transgenic reporter assay, one of the elements
176 associated with *Tbx5* showed activity in the pericloacal region (n=6/14) (Fig 3G, G').
177 Interestingly, it did not activate overt forelimb expression (Supplementary Fig. 3D), the
178 major expression domain of this gene [26], further indicating specificity for the pericloacal
179 region. Similarly, the element identified downstream of *Isl1* also reproduced to a large
180 extent *Isl1* expression in the GT when tested in transgenic reporter assays (n=5/12).
181 An element associated with *Wnt5a* was also able to activate expression in the GT,
182 although much less frequently as the other elements (n=2/18) (Supplementary Fig.
183 3B, B'). This element was also active in the tail, but it did not promote expression in
184 the hindlimb. Together, these observations are consistent with these elements indeed
185 representing enhancers potentially involved in the expression of the relevant genes in
186 the pericloacal area.
187 We found pattern 2 elements in regions associated with genes playing essential roles
188 during the earliest stages of limb development, which could thus play a relevant role
189 in promoting limb fates from the pericloacal mesoderm (Fig. 4A, A'). As for pattern 1
190 elements, these regions were also highly conserved among vertebrates. We found
191 pattern 2 elements in the *Fgf10* genomic region (Supplementary Fig. 3C), a gene
192 activated in the mesenchyme adjacent to the cloaca entering hindlimb fate in the
193 mutant embryo (Fig. 2D, D'). Interestingly, published ChIP-seq data from forelimb buds
194 [27,28] showed binding of Gli3 and Hoxa13 to one of these elements, located within an
195 intronic region of this gene (Supplementary Fig. 3C), indicating that it might respond
196 to Shh and/or Hoxa13 activities in the developing limb buds. The lack of accessibility
197 of this element in the GT of wild type embryos might thus suggest that Tgfbr1 renders
198 this enhancer blind to endodermal Shh and/or the strong pericloacal Hoxa13 (and
199 maybe also Hoxd13) expression [29], consistent with absent *Fgf10* at early stages in this
200 tissue. Rather surprisingly, this element failed to generate reporter activity in
201 transgenic embryos (n=10). We still do not understand whether this reflects real lack
202 of activity or the absence of proper genomic context for their activity.
203 Another potentially relevant pattern 2 element was located within the *Fmn1* locus,
204 matching the position of one of the enhancers (element 7 in Malkmus et al., 2021)

205 controlling *Grem1* in the limb buds (Supplementary Fig. 3E) that had been shown to
206 respond to Shh activity from the ZPA [30,31]. In wild type embryos *Grem1* expression is
207 not detected in the pericloacal mesoderm or in the GT (Fig. 4D), indicating that in these
208 tissues *Tgfbr1* might render the *Grem1* regulatory elements blind to endodermal Shh.
209 The locked configuration of element 7 in the GT (Fig. 4C) is consistent with this
210 hypothesis. Conversely, the gain of accessibility of element 7 in the extra hindlimb
211 suggests that in the absence of *Tgfbr1* the pericloacal mesoderm could become
212 responsive to the endodermal Shh. Considering the capacity of *Grem1* to expand the
213 expression of AER markers [32–34], this gene is a candidate to contribute to the observed
214 AER extension in the hindlimb primordium of *Tgfbr1-cKO* embryos. *Grem1* expression
215 analyses in *Tgfbr1-cKO* embryos revealed unexpected complexities. Some features
216 of the expression pattern were consistent with *Grem1* becoming responsive to
217 endodermal Shh, as expression was detected in the caudal-most end of the expanded
218 hindlimb next to the endoderm (Fig. 4D'). Rather surprisingly, however, *Grem1*
219 expression was strongly reduced or absent throughout the mesoderm of the extended
220 hindlimb bud of *Tgfbr1-cKO* embryos (Fig. 4D'). *Grem1* limb expression depends on
221 Shh activity from the zone of polarizing activity (ZPA) [30,35,36], suggesting that the
222 absent *Grem1* expression resulted from the abnormal *Shh* expression observed in the
223 hindlimb buds of *Tgfbr1-cKO* mutants, which followed spotty patterns beneath the AER
224 instead of forming a well-defined ZPA in the posterior limb bud mesenchyme (Fig. 3A,
225 A').
226 We explored whether this abnormal *Shh* expression resulted from a functional impact
227 of the absence of Tgfbr1 activity on the ZRS, the regulatory element controlling *Shh*
228 expression in the ZPA [37]. As expected, this region was accessible in wild type
229 hindlimbs (Fig. 4E). Conversely, it was in a locked configuration in the wild type GT,
230 consistent with the absence of the *Shh* expression in the pericloacal mesoderm or the
231 GT, despite the presence of relevant activators of this enhancer, like Hand2, Hoxa13
232 and Hoxd13 [4,29,38,39] (the endodermal *Shh* expression is under the control of a different
233 enhancer[40]). Remarkably, the ZRS was also non-accessible in the hindlimb region of
234 *Tgfbr1-cKO* embryos, consistent with the observed abnormal *Shh* limb expression.
235 Therefore, Tgfbr1 seems to also control ZRS accessibility to their regulatory
236 transcription factors. Interestingly, this enhancer is part of an additional pattern
237 (pattern 3) consisting of elements requiring Tgfbr1 activity to become accessible in the
238 limb bud. Analysis of the ATAC-seq data revealed that 13471 elements fit the criteria

239 of this pattern (Fig. 4B, B'). The complementary tissue specificities of patterns 1 and
240 3 indicates context dependency for Tgfbr1 activity in the modulation of genomic
241 configuration.
242
243 **Discussion**
244 The capacity of the pericloacal mesoderm of mouse embryos to generate both external
245 genitalia and hindlimb structures reveals a remarkable degree of developmental
246 plasticity in this tissue. This finding provides a new dimension to the developmental
247 connection between the hindlimb and the GT, previously suggested by the
248 considerable overlap in their regulatory networks [2–5]. It has been shown that the paired
249 external genitalia of snakes develop at least to some extent from mesodermal tissues
250 in the prospective hindlimb domain. It had been suggested that the proximity of this
251 mesoderm to the cloacal signals could play a role in this process [1]. However, such
252 proximity was also observed in lizards that develop both hemipenises and hindlimbs
253 [1]. It will be therefore interesting to determine whether a mechanism related to the
254 control of developmental plasticity uncovered in our work plays a role in the absence
255 of hindlimbs in snakes. In addition, understanding the relative weight of the different
256 factors whose expression was lost or ectopically activated in the pericloacal region of
257 the *Tgfbr1-cKO* mutants might help to uncover the core regulatory network able to
258 promote specific and stable fates from an intrinsically plastic tissue. This will require
259 generating different combinations of gain and loss of function experiments.
260 An important finding from our work is that *Tgfbr1* controls pericloacal mesoderm
261 plasticity by modulating its regulatory landscape. It has been reported that BMP
262 signaling can affect the chromatin status of Smad-responsive enhancers by interacting
263 with chromatin remodelers like Foxh1 or Smarca4, eventually rendering these
264 regulatory elements competent to convey the Smad transcriptional activity [41,42]. What
265 we have observed in the pericloacal/hindlimb region does not easily fit within this
266 mechanism. For instance, pattern 2 elements are rendered non-accessible by *Tgfbr1*,
267 compatible with a repressive rather than potentially activating function. This could
268 result from interference with a positive activity provided by other members of the
269 Tgfb/BMP family on those enhancers. However, at least for the *Grem1* enhancer 7,
270 this is not a likely explanation, as ChIP-seq analyses failed to detect Smad4 binding
271 on limb tissues, where *Grem1* is normally expressed [30]. In addition, we failed to detect
272 enrichment of Smad binding to enhancers rendered accessible by *Tgfbr1* using a

HINT-ATAC transcription factor footprinting analysis [43] on the ATAC-seq profiles (Supplementary Table 1). Even considering the limitations of the bioinformatic approach, these observations suggest that effectors of the Tgbfr1/BMP signaling pathways are not among the main regulators of the activity of those elements. Consistent with this, Smads have so far not been identified among the factors involved in the activation of the ZRS enhancer.

Our data suggest a general model according to which Tgfbr1 signaling acts in a pioneer-like activity, determining the accessibility of regulatory elements to patterning transcription factors. As many of the patterning factors regulating hindlimb and GT development are common to both structures, the global output of Tgfbr1 activity would be establishing how their primordia respond to those factors. While extensive analysis of limb development over the last decades might already guide the identification of factors interacting with different enhancers involved in limb-related processes, understanding the control of pericloacal gene expression will require focused ChIP-seq profiles from this tissue and integrating them with the ATAC-seq data. This information will also provide a direct estimation of the involvement of effectors of the Tgfβ/Bmp pathways in these processes.

It should be noted, however, that although the global output resembles some hallmarks of a pioneer activity, the mechanism is unlikely to follow routes similar to those described for classical pioneer factors [44], as the Tgfbr1-mediated chromatin remodeling process does not seem to involve Smad binding to the regulated elements, and controls accessibility in opposite ways for different enhancers. Identifying the mechanism of this activity, and whether similar or different processes operate for the different patterns in the regulatory landscape, will be essential to fully understand how Tgfbr1 signaling regulates morphogenetic processes. In this regard, it will be important to determine whether a similar pioneer-like mechanism also operates in other physiological and pathological processes under the control of members of the Tgfβ/BMP signaling family.

1. Tschopp, P. *et al.* A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* **516**, 391–394 (2014).
2. Yamada, G. *et al.* Molecular genetic cascades for external genitalia formation: An emerging organogenesis program. *Developmental Dynamics* **235**, 1738–1752 (2006).

3. Cohn, M. J. Development of the external genitalia: Conserved and divergent mechanisms of appendage patterning. *Developmental Dynamics* **240**, 1108–1115 (2011).
4. Kondo, T., Zakany, J., Innis, J. W. & Duboule, D. Of fingers, toes and penises [1]. *Nature* **390**, 29 (1997).
5. Lin, C. *et al.* Delineating a Conserved Genetic Cassette Promoting Outgrowth of Body Appendages. *PLoS Genet* **9**, e1003231 (2013).
6. Jurberg, A. D., Aires, R., Varela-Lasheras, I., Nóvoa, A. & Mallo, M. Switching axial progenitors from producing trunk to tail tissues in vertebrate embryos. *Dev Cell* **25**, 451–462 (2013).
7. Seifert, A. W., Harfe, B. D. & Cohn, M. J. Cell lineage analysis demonstrates an endodermal origin of the distal urethra and perineum. *Dev Biol* **318**, 143–152 (2008).
8. Miyagawa, S. *et al.* Disruption of the temporally regulated cloaca endodermal β-catenin signaling causes anorectal malformations. *Cell Death Differ* **21**, 990–997 (2014).
9. Miyagawa, S. *et al.* Dosage-dependent hedgehog signals integrated with Wnt/β-catenin signaling regulate external genitalia formation as an appendicular program. *Development* **136**, 3969–3978 (2009).
10. Huang, Y. C., Chen, F. & Li, X. Clarification of mammalian cloacal morphogenesis using high-resolution episcopic microscopy. *Dev Biol* **409**, 106–113 (2016).
11. Tickle, C. How the embryo makes a limb: Determination, polarity and identity. *J Anat* **227**, 418–430 (2015).
12. Dias, A. *et al.* A Tgfbr1/Snai1-dependent developmental module at the core of vertebrate axial elongation. *Elife* **9**, e56615 (2020).
13. Pineault, K. M., Novoa, A., Lozovska, A., Wellik, D. M. & Mallo, M. Two CRISPR/Cas9-mediated methods for targeting complex insertions, deletions, or replacements in mouse. *MethodsX* **6**, 2088–2100 (2019).
14. Dünker, N. & Krieglstein, K. Tgfβ2-/-Tgfβ3-/- double knockout mice display severe midline fusion defects and early embryonic lethality. *Anat Embryol (Berl)* **206**, 73–83 (2002).
15. McPherron, A. C., Lawle, A. M. & Lee, S.-J. Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. *Nat Genet* **22**, 260–264 (1999).
16. Andersson, O., Reissmann, E. & Ibáñez, C. F. Growth differentiation factor 11 signals through the transforming growth factor-beta receptor ALK5 to regionalize the anterior-posterior axis. *EMBO Rep* **7**, 831–837 (2006).
17. Perriton, C. L., Powles, N., Chiang, C., Maconochie, M. K. & Cohn, M. J. Sonic hedgehog signaling from the urethral epithelium controls external genital development. *Dev Biol* **247**, 26–46 (2002).
18. Davis, C. A., Holmyard, D. P., Millen, K. J. & Joyner, A. L. Examining pattern formation in mouse, chicken and frog embryos with an En-specific antiserum. *Development* **111**, 287–298 (1991).
19. Riddle, R. D. *et al.* Induction of the LIM Homeobox Gene Lmx7 by WNT7a Establishes Dorsoventral Pattern in the Vertebrate Limb. *Cell* **83**, 631–640 (1995).
20. Yang, L. *et al.* Isl1 Cre reveals a common Bmp pathway in heart and limb development. *Development* **133**, 1575–1585 (2006).

21. Su, T. *et al.* LIM homeodomain transcription factor Isl1 affects urethral epithelium differentiation and apoptosis via Shh. *Cell Death Dis* **10**, 713 (2019).
22. Min, H. *et al.* Fgf-10 is required for both limb and lung development and exhibits striking functional similarity to Drosophila branchless. *Genes Dev* **12**, 3156–3161 (1998).
23. Satoh, Y. *et al.* Regulation of external genitalia development by concerted actions of FGF ligands and FGF receptors. *Anat Embryol (Berl)* **208**, 479–486 (2004).
24. Haraguchi, R. *et al.* Unique functions of sonic hedgehog signaling during external genitalia development. *Development* **128**, 4241–4250 (2001).
25. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962 (2017).
26. Rallis, C. *et al.* Tbx5 is required for forelimb bud formation and continued outgrowth. *Development* **130**, 2741–2751 (2003).
27. Sheth, R. *et al.* Distal Limb Patterning Requires Modulation of cis-Regulatory Activities by HOX13. *Cell Rep* **17**, 2913–2926 (2016).
28. Lex, R. K. *et al.* GLI transcriptional repression regulates tissue-specific enhancer activity in response to hedgehog signaling. *Elife* **9**, e50670 (2020).
29. Scott, V., Morgan, E. A. & Stadler, H. S. Genitourinary functions of Hoxa13 and Hoxd13. *Journal of Biochemistry* vol. 137 671–676 Preprint at https://doi.org/10.1093/jb/mvi086 (2005).
30. Malkmus, J. *et al.* Spatial regulation by multiple Gremlin1 enhancers provides digit development with cis-regulatory robustness and evolutionary plasticity. *Nat Commun* **12**, 5557 (2021).
31. Zhu, J., Patel, R., Trofka, A., Harfe, B. D. & Mackem, S. Sonic hedgehog is not a limb morphogen but acts as a trigger to specify all digits in mice. *Dev Cell* **57**, 2048–2062 (2022).
32. Michos, O. *et al.* Gremlin-mediated BMP antagonism induces the epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. *Development* **131**, 3401–3410 (2004).
33. Nissim, S., Hasso, S. M., Fallon, J. F. & Tabin, C. J. Regulation of Gremlin expression in the posterior limb bud. *Dev Biol* **299**, 12–21 (2006).
34. Farin, H. F. *et al.* Tbx2 Terminates Shh/Fgf Signaling in the Developing Mouse Limb Bud by Direct Repression of Gremlin1. *PLoS Genet* **9**, e1003467 (2013).
35. Zuniga, A. *et al.* Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev* **18**, 1553–1564 (2004).
36. Li, Q. *et al.* A Gli silencer is required for robust repression of gremlin in the vertebrate limb bud. *Development (Cambridge)* **141**, 1906–1914 (2014).
37. Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725–1735 (2003).
38. Galli, A. *et al.* Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. *PLoS Genet* **6**, e1000901 (2010).
39. Leal, F. & Cohn, M. J. Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Report Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Enhancers. *Current Biology* **26**, 1–8 (2016).

40. Sagai, T. *et al.* A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development* **136**, 1665–1674 (2009).
41. Coda, D. M. *et al.* A network of transcription factors governs the dynamics of NODAL/Activin transcriptional responses. *J Cell Sci* **135**, jjcs259972 (2022).
42. Coda, D. M. *et al.* Distinct modes of SMAD2 chromatin binding and remodeling shape the transcriptional response to NODAL/Activin signaling. *Elife* **6**, 1–31 (2017).
43. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* **20**, 45 (2019).
44. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: Establishing competence for gene expression. *Genes Dev* **25**, 2227–2241 (2011).

Acknowledgements

We would like to thank the members of the Mallo lab, especially Ana Casaca for continuous support at different stages of this project, the IGC mouse facility for their help with animal housing and genomics facility for helping with the ATAC-seq procedures. This project was funded by Fundação para a Ciência e a Tecnologia (FCT) grants 2022.01629.PTDC to MM, PTDC/BII-BTI/32375/2017 to GGM, and PhD fellowships PD/BD/128437/2017 to AL and PD/BD/128426/2017 to AD, by the PhD fellowship 36/8I-D/21 from Fundação Calouste Gulbenkian to AK, and the research infrastructures PPBI-POCI-01-0145-FEDER-022122 to the Advanced Imaging Facility, and Congento LISBOA-01-0145-FEDER-022170 to the animal facility, both co-financed by Lisboa Lisboa 2020/FEDER and FCT (Portugal).

Author contribution

Conceptualization: A.L. and M.M.; Formal Analysis: A.L. and A.K.**;** Funding Acquisition: M.M.; Investigation: A.L., A.K. and M.M.; Methodology: A.L., A.D. G.G.M. A.N., A.Lo. and D.F.; Writing – Original Draft Preparation: A.L. and M.M.; Writing – Review & Editing: A.L., A.K. and M.M.

Competing interest declaration

The authors declare no competing interests.

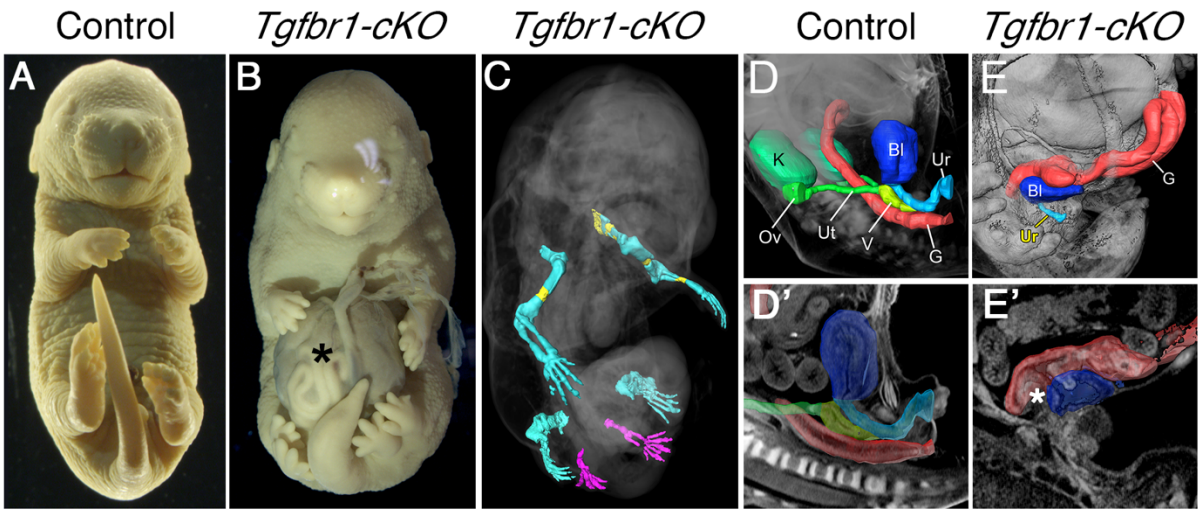
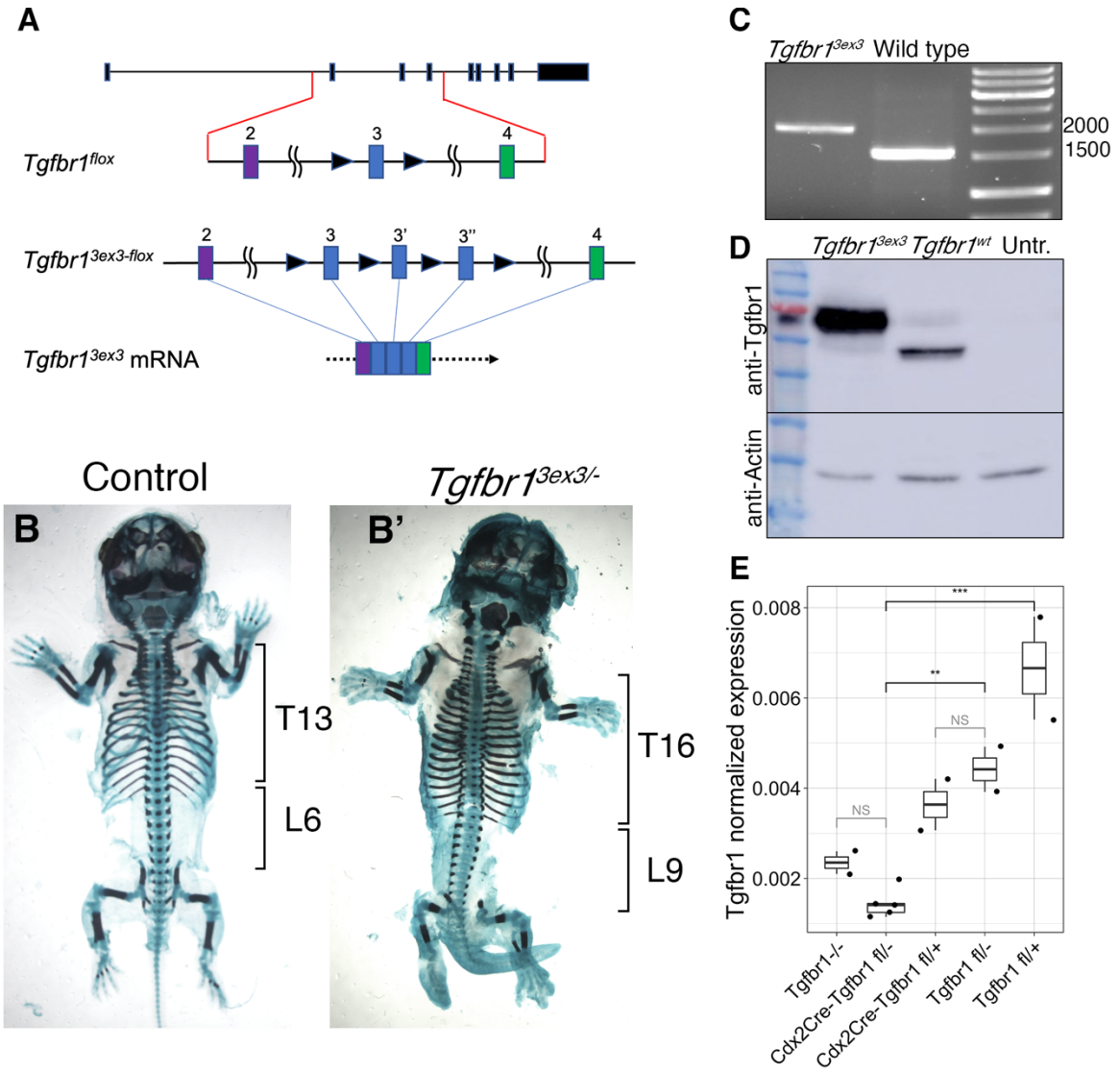


Figure 1. Malformations in the E16,5 *Tgfbr1-cKO* fetuses. (A, B) Images of the fixed WT (A) and *Tgfbr1-cKO* (B) fetuses. The mutant fetus shows the presence of an omphalocele (asterisk) and hindlimb duplication. (C) 3D reconstruction of the limb skeleton of a *Tgfbr1-cKO* fetus obtained by OPT and after segmentation of the limb skeleton. Extra hindlimbs are in magenta. Ossification shown in yellow. (D–E') 3D reconstruction of organs and the excretory outlets of the control (D, D') and *Tgfbr1-cKO* (E, E') fetuses. Images were obtained by OPT, followed by segmentation of the relevant structures. D' E' show virtual sections of the segmented 3D specimen. Asterisk in E' shows the gut-bladder connection in the mutant. K: kidney; G: gut; Bl: bladder; Ov: ovary Ut: uterus; V: vagina; Ur: urethra.



Supplementary Figure 1. Generation on the *Tgfbr1-cKO*.(A) Scheme showing *Tgfbr1*^{flox} and *Tgfbr1*$^{3ex3\text{-}flox}$ alleles. Highlighted is the region containing exons 2, 3 and 4 (colored boxes) and the LoxP sites (black triangles). Also represented is the generation of a mRNA with a triplicated exon 3 after splicing of the transcript from the *Tgfbr1*$^{3ex3\text{-}flox}$ allele. (B, B') Skeletal staining of a E17.5 wild type (B) and *Tgfbr1*$^{3ex3\text{-}flox/-}$ fetus (B'), showing the presence of additional thoracic (T) and lumbar (L) vertebrae, resembling a *Gdf11* mutant phenotype. (C) RT-PCR of the *Tgfbr1* coding region from wild type (1,5 kb) and *Tgfbr1*$^{3ex3\text{-}flox/3exe\text{-}flox}$ (2 kb) embryos. (D). Western blot analysis of protein extracts from 293T cells transfected with mammalian expression vectors carrying either the Tgfbr1^{3ex3} or the Tgfbr1 coding regions, showing that the *Tgfbr1*^{3ex3} transcript produces a protein with a higher molecular mass than its wild type counterpart. (E) RT-qPCR showing recombination efficiency in the *Tgfbr1*$^{flox/-}$*:Cdx2Cre*$^{ERT+/0}$ embryos. Normalized *Tgfbr1* expression level in *Tgfbr1*$^{flox/-}$*:Cdx2Cre*$^{ERT+/0}$ is not different from *Tgfbr1*$^{-/-}$. Level of normalized expression in *Tgfbr1*$^{+/flox}$*:Cdx2Cre*$^{ERT+/0}$ equals that of *Tgfbr1*$^{flox/-}$ and half of *Tgfbr1*$^{flox/+}$ control, consistent with it carrying only one functioning allele. NS: nonsignificant, **p<0.01, ***p<0.001

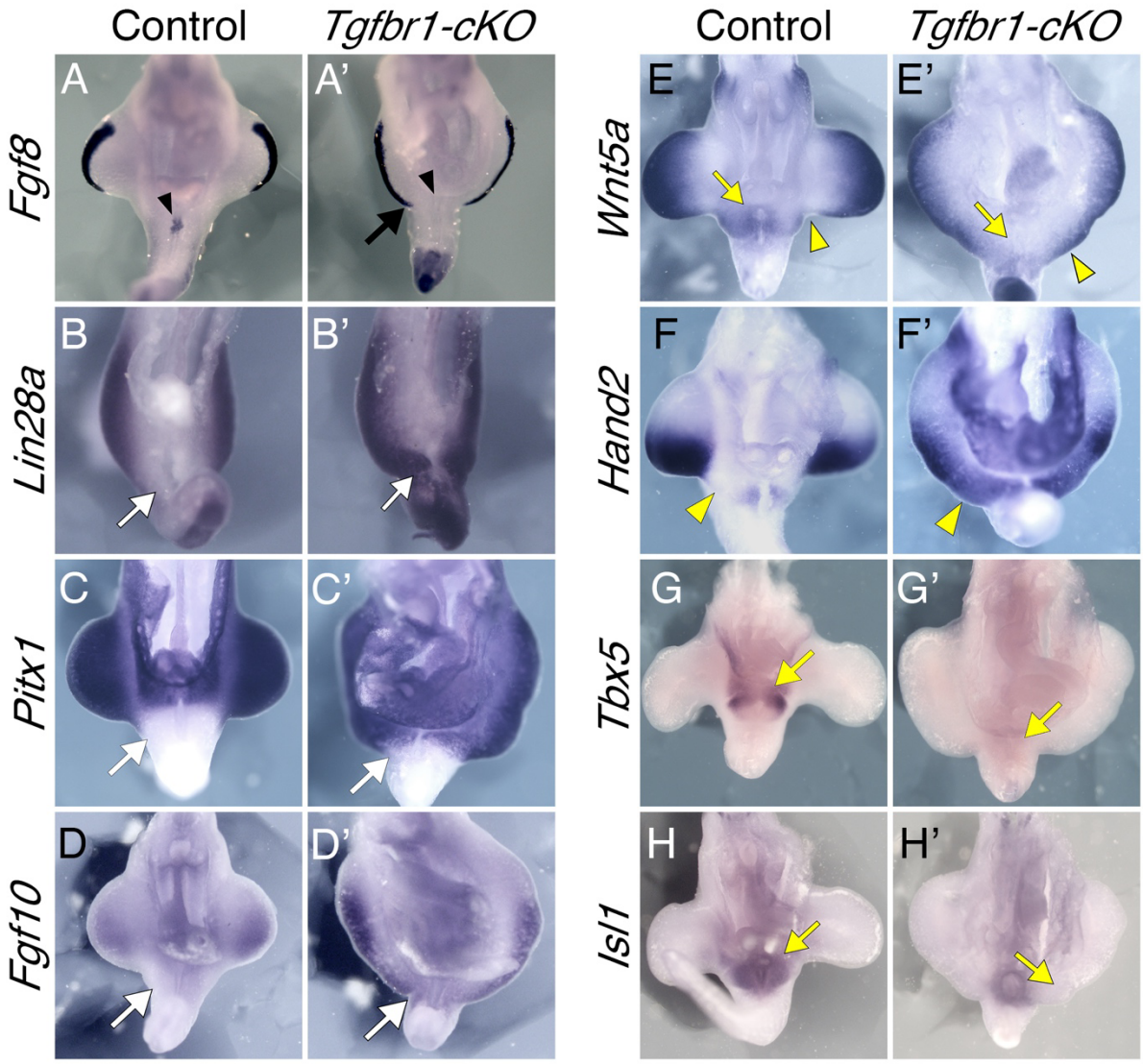
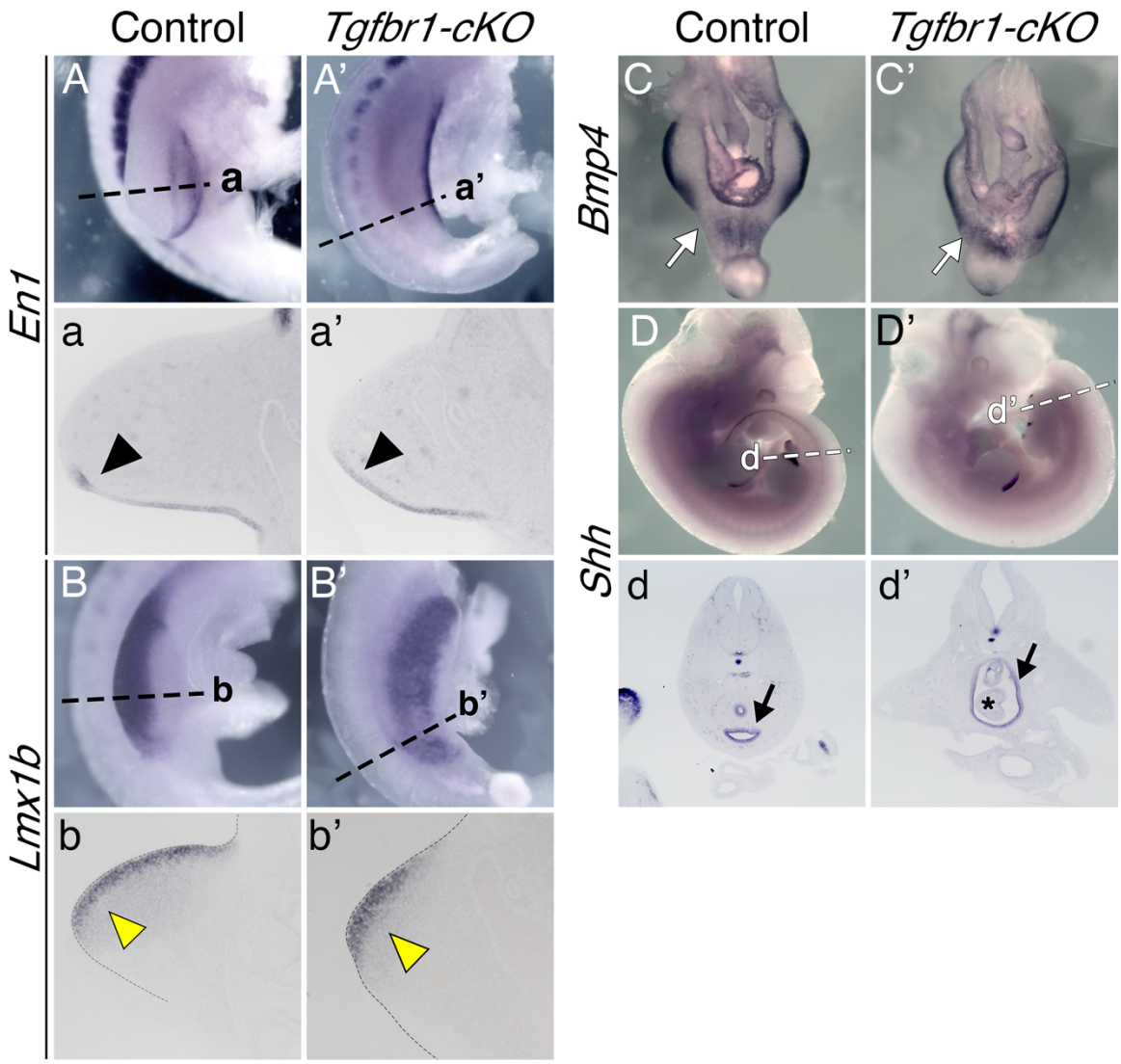


Figure 2. The pericloacal mesenchyme of the *Tgfbr1-cKO* embryos adopts hindlimb fate. Ventral views of the hindlimb/cloacal region of wild type (A-H) or *Tgfbr1-cKO* embryos (A'-H') stained for *Fgf8* (A, A'), *Lin28a* (B, B'), *Pitx1* (, C'), *Fgf10* (D, D'), *Wnt5a* (E, E'), *Hand2* (F, F'), *Tbx5* (G, G') or *Isl1* (H, H'). *Fgf8* expression show posterior and medial elongation of the AER (black arrow) and the absence of expression in the cloacal endoderm (black arrowhead). Hindlimb markers *Lin28a*, *Pitx1* and *Fgf10* extend into the pericloacal region in E10,5 *Tgfbr1-cKO* embryos (white arrows). *Wnt5a* and *Hand2* switch their expression profiles in the pericloacal mesoderm (yellow arrowheads), now following a limb-like pattern. In addition, pericloacal expression of *Wnt5a*, *Tbx5* and *Isl1* is lost in the mutant embryos (yellow arrows).



Supplementary Figure 2. Molecular characterization of the *Tgfbr1-cKO* embryos. (A-b') Dorsal-ventral polarity of the *Tgfbr1-cKO* mutant's hindlimb. Expression of the ventral limb marker *En1* in E10.5 control (A, a) and *Tgfbr1-cKO* (A', a') embryos. (A, A') show lateral views of the hindlimb region in whole mounted embryos; (a, a') show transversal sections through the region indicated in A, A'. Black arrowheads show *En1* expression in the ventral ectoderm of the limb bud. Expression of the dorsal mesodermal limb marker *Lmx1b* in E10.5 control (B, b) and *Tgfbr1-cKO* (B', b') embryos. (B, B') show lateral views of the hindlimb region in whole mounted embryos; (b, b') show transversal sections through the region indicated in B, B'. Yellow arrowheads show *Lmx1b* expression in the dorsal mesenchyme of the limb bud. (C, C') Ventral views of E10.5 wild type (C) and *Tgfbr1-cKO* (C') embryos stained for *Bmp4* showing the extension of the signal into the pericloacal region in the mutant embryos (white arrow). (D-d') Expression of *Shh* in E11,5 control (D, d) and *Tgfbr1-cKO* (D', d') embryos. (D, D') show lateral views of whole mounted embryos; (d, d') show transversal section through the region indicated in D, D' to show the enlarged cloaca (black arrows) with presence of the intralumenal protrusions in the mutant embryo (asterisk), still keeping endodermal *Shh* expression.

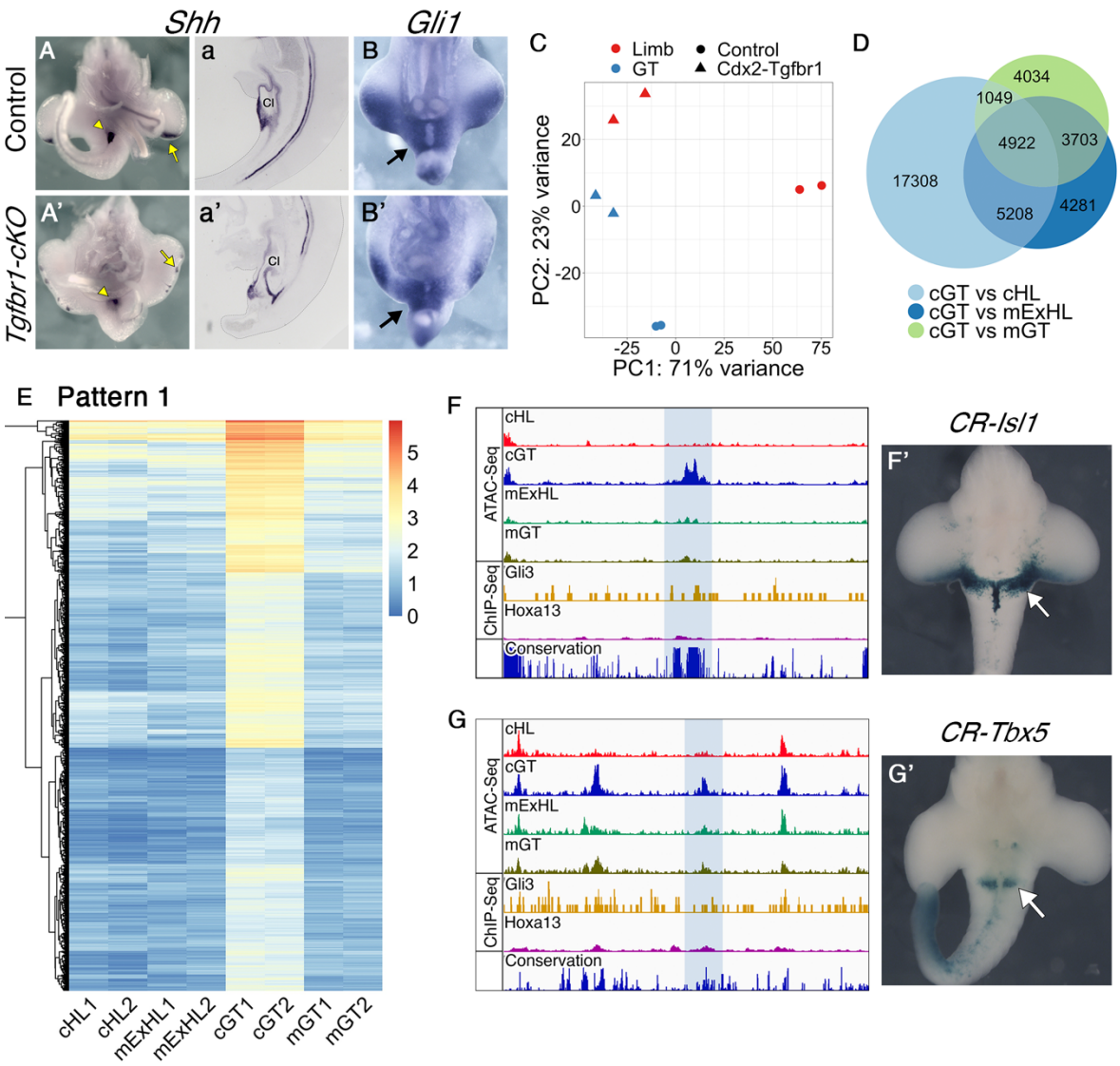
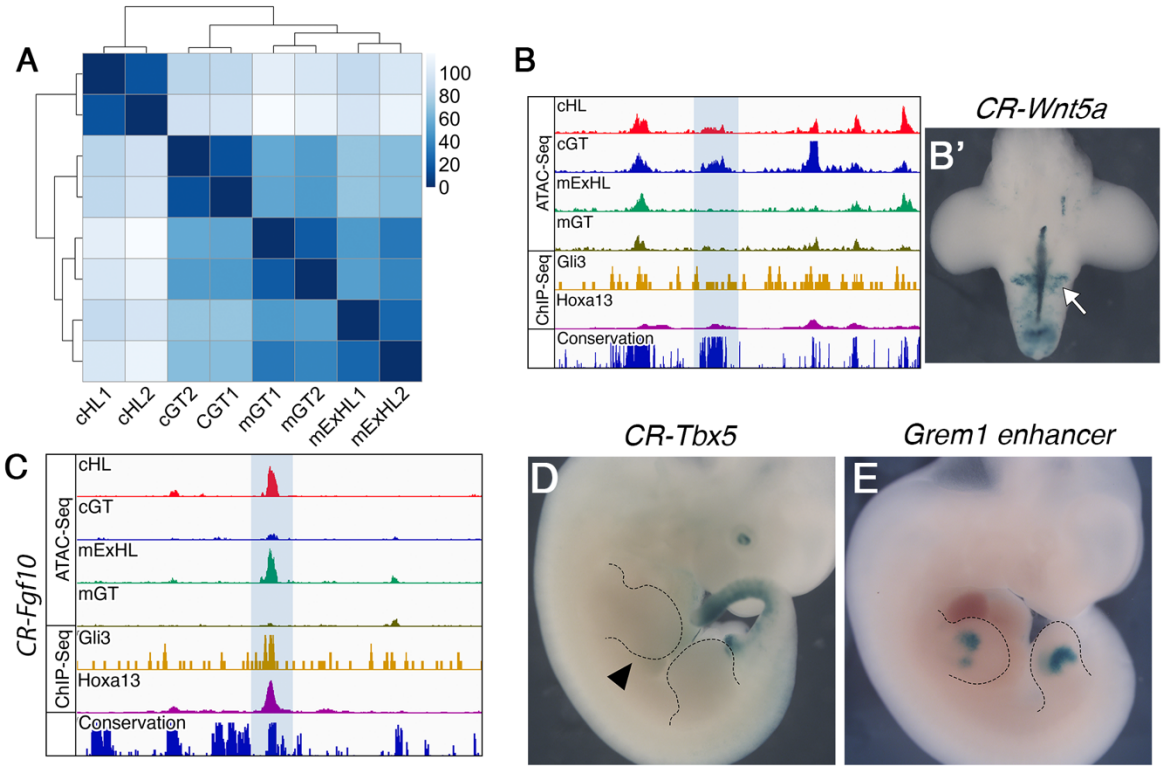


Figure 3. Tgfbr1 regulates the chromatin landscape in the pericloacal region. (A–B') The pericloacal mesenchyme of *Tgfbr1-cKO* embryos responds to Shh signaling from the cloacal endoderm. *Shh* expression in E11,5 control (A, a) and *Tgfbr1-cKO* (A', a') embryos. (A, A') show ventral views of the stained embryos: (a, a') show sagittal sections through cloacal (Cl) region. *Gli1* expression in E10,5 control (B) and *Tgfbr1-cKO* (B') embryos. (C) PCA analysis reveals closer proximity of the chromatin accessibility pattern of the *Tgfbr1-cKO* extra hindlimb to that of *Tgfbr1-cKO* and control GT than to the control hindlimb. (D, E) Chromatin regions following pattern 1. (D) Venn diagram showing interception of the regions tested more accessible in control GT than in other tissues by pairwise comparisons [FDR<0.001, logFC>1.5]. (E) Heatmap showing log2 normalized counts across samples of the regions following pattern 1. (F, G') ATAC-seq tracks showing accessible chromatin in control GT, but not in other samples in regions 24 kb downstream of the *Isl1* TSS (F) and 93 kb upstream of the *Tbx5* TSS (G). The regions highlighted with the blue shade are characterized by high conservation level among placental animals (bottom track). (F', G') The highlighted elements drive b-galactosidase expression in the GT region (white arrows). Hoxa13 and Gli3 ChiP-Seq tracks from forelimb buds (GSE81356 and GSE133710, respectively) are shown in yellow and magenta, respectively. cHL:

516 control hindlimb; mExHL: mutant extra hindlimb; cGT: control GT; mGT: mutant GT;
517 CR: conserved region.
518



Supplementary Figure 3. (A) Distance heatmap of the chromatin accessibility patterns showing that the *Tgfbr1-cKO* extra hindlimb clusters closer to mutant and control GT than to control limb. (B) ATAC-seq tracks showing a chromatin region (highlighted with a blue shadow) accessible in control GT, but not in other tissues of a region 93 kb upstream of the *Wnt5a* TSS. This highly conserved among placental animals (bottom track). (B') The highlighted element in (B) drives b-galactosidase expression in the GT region (white arrow) in a transgenic reporter assay. (C) A region located within intron 1 of the *Fgf10* gene (highlighted with a blue shadow), is accessible in the hindlimb samples of the control and *Tgfbr1-cKO* embryos, but not in GT samples. ChIP data from forelimb buds show that this genomic region is bound by Hoxa13 and Gli3. (D) The putative *Tbx5* GT enhancer shown in Figure 3 does not drive b-galactosidase expression in the forelimb (black arrowhead, forelimb outlined with the dashed line). (E) b-galactosidase expression driven by the *Grem1* enhancer in Figure 4 using a transgenic reporter assay (n=2/8) (limbs outlined with the dashed line). CR: conserved region.

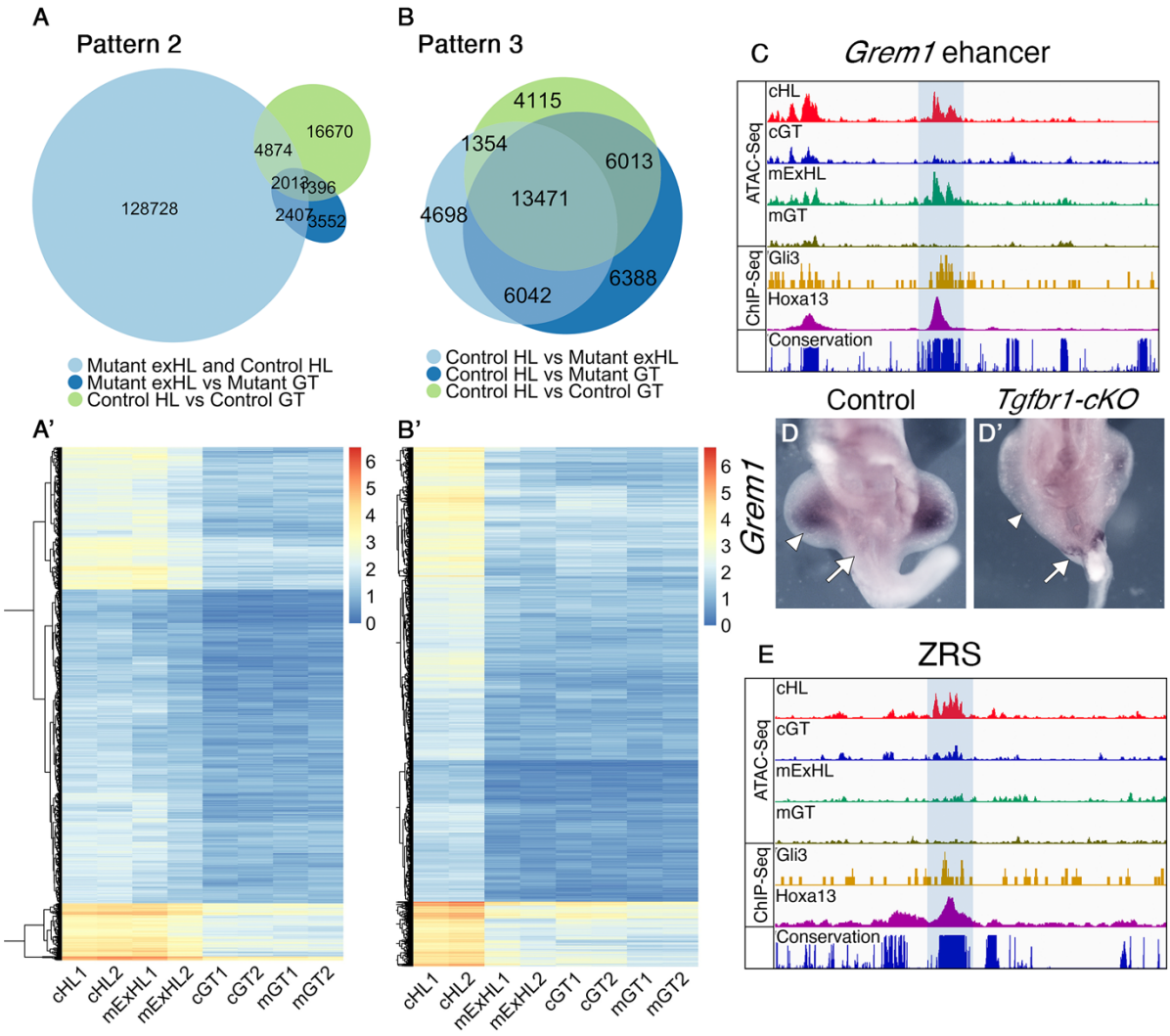


Figure 4. (A, A') Chromatin regions following pattern 2. (A) Venn diagram showing interception between the regions more accessible in mutant extra limb than in mutant GT [FDR<0.001, logFC>1.5] and more accessible in control hindlimb than in control GT [FDR<0.001, logFC>1.5], but at the same time not differentially accessible between mutant extra hindlimb and control hindlimb [FDR>0.01, logFC<0.5]. (A') Heatmap showing log2 normalized counts from genomic regions following pattern 2 across the samples. (B, B') Chromatin regions following pattern 3. (B) Venn diagram showing interception between regions tested more accessible in control hindlimb than in other tissues by pairwise comparisons [FDR<0.001, logFC>1.5]. (E) Heatmap showing log2 normalized counts for the chromatin regions following pattern 3 across samples. (C) ATAC-Seq profiles showing a *Grem1* enhancer that gains accessibility in the mutant extra hindlimb (highlighted with the blue shadow). Also shown are ChiP-Seq tracks of ChIP seq data from wild type forelimbs for Gli3 (yellow) and Hoxa13 (magenta) ((GSE81356 and GSE133710, respectively), as well as the conservation in placental mammals. (D, D') *Grem1* expression in E10.5 wild type (D) or *Tgfbr1-cKO* (D') embryos showing the absence of pericloacal expression in wild type embryos and ectopic activation in the caudal-most region of the pericloacal mesoderm of the mutant embryo (white arrow). *Grem1* is expressed in the developing limb bud of the control embryo, but it is almost non detectable in the

557 developing limb buds of the mutant embryo (white arrowheads). (E) ATAC-seq data
558 through the ZRS (highlighted with the blue shadow), showing that it is accessible in
559 the control limb but not in any of the other samples. cHL: control hindlimb; mExHL:
560 mutant extra hindlimb; cGT: control GT; mGT: mutant GT.
561
562
563 **Supplementary movie 1.** OPT image from a E16.5 *Tgfbr1-cKO* fetus highlighting
564 the limbs. The extra hindlimbs are shown in magenta.
565
566 **Supplementary movie 2.** OPT image from a E16.5 wild type fetus highlighting the
567 caudal end of the intestine (red), the kidneys (cold green), the ovaries and uterus
568 (bright green), the vagina (yellow), the urinary bladder (dark blue) and the urethra
569 (light blue).
570
571 **Supplementary movie 3.** OPT image from a E16.5 *Tgfbr1-cKO* fetus highlighting
572 the caudal end of the intestine (red), the urinary bladder (dark blue) and the urethra
573 (light blue).
574
575 **Supplementary Table 1.** Transcription factor footprint analysis of pattern 1
576 elements. Column A: number of pattern 1 regions with at least 1 occurrence of the
577 specified transcription factor; column B: number of pattern1 regions with no
578 occurrence of the specified transcription factor; C: number of background regions
579 with at least 1 occurrence of the specified transcription factor; D: number of
580 background regions with no occurrence of the specified transcription factor.
581
582 **Materials and methods**
583 *Ethics statement*
584 All animal procedures were performed in accordance with Portuguese (Portaria
585 1005/92) and European (directive 2010/63/EU) legislations and guidance on animal
586 use in bioscience research. The project was reviewed and approved by the Ethics
587 Committee of “Instituto Gulbenkian de Ciência” and by the Portuguese National Entity
588 “Direcção Geral de Alimentação Veterinária” (license reference: 014308).
589
590 *Mice*

591 Generation of the *Tgfbr1*^{Flox} mouse lines was previously reported [13]. *Tgfbr1*$^{3ex3\text{-}flox}$
592 resulted from the same experiments in which *Tgfbr1*^{flox} was generated, identified upon
593 sequencing through the Tgfbr1 locus which revealed the presence of three exon 3
594 together with LoxP sites in the introns between the triplicated exons and in those
595 connecting with exon 2 and exon 4. Characterization of mice homozygous for the
596 *Tgfbr1*$^{3ex3\text{-}flox}$ allele revealed the presence of a transcript containing the three exon 3
597 in frame with the flanking exons 2 and 4. Expression of this mRNA in HEK293T cells
598 confirmed that it can generate a protein product with the expected size of Tgfbr1
599 containing a triplication of exon 3 (supplementary Fig. 1). The *Cdx2-cre*^{ERT} mice were
600 previously reported [6]. Mice were genotyped from ear of digit biopsies. Samples were
601 incubated in 50 μL of the PBND buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 2.5 mM
602 $MgCl_2$, 0.1 mg/mL gelatin, 0.45% NP40, 0.45% Tween 20) supplemented with 100
603 μg/mL of proteinase K at 55˚C overnight. Proteinase K was inactivated at 95˚C for 15
604 minutes and 1 μL of genomic DNA was used for genotyping by PCR. *Cdx2-*
605 *cre*^{ERT}*::Tgfbr1*$^{+/-}$ mice were genotyped for the *Cre* allele and *Tgfbr1* null allele;
606 *Tgfbr1*$^{flox/+}$ and *Tgfbr1*$^{3ex3\text{-}flox/+}$ mice were genotyped with primers surrounding 3'-LoxP
607 (primers listed in supplementary table 2).
608 *Tgfbr1-cKO* (*Cdx2-cre*^{ERT}*::Tgfbr1*$^{flox/-}$ or *Cdx2-cre*^{ERT}*::Tgfbr1*$^{3ex3\text{-}flox/-}$) embryos were
609 obtained by crossing *Tgfbr1*$^{flox/flox}$ or *Tgfbr1*$^{3ex3\text{-}flox/+}$ females with *Tgfbr1*$^{+/-}$:*Cdx2-cre*^{ERT}
610 males. Noon of the day of the plug was considered E0.5. To induce recombination
611 pregnant females were treated with tamoxifen (Sigma, T5648) dissolved in corn oil.
612 Tamoxifen was administered at 0.1 mg per gram of body weight by oral gavage either
613 once at E6.75 when *Tgfbr1*$^{3ex3\text{-}flox}$ females were involved, or twice, at E6.75 and E7.25
614 when *Tgfbr1*^{flox} females were involved.
615 Embryos were obtained by cesarean section and processed for further analyses (see
616 below). They were genotyped from their yolk sacs for Cre, *Tgfbr1*$^{-}$ and *Tgfbr1*^{flox} when
617 analyzing *Tgfbr1-cKO* embryos or for β-gal when analyzing the reporter transgenic
618 embryos (primers listed in supplementary table 2). Yolk sacs were collected in 50 μL
619 of yolk sac lysis buffer (50 mM KCl, 10 mM Tris-HCl, pH8.3, 2 mM $MgCl_2$, 0.45%
620 Tween-20, 0.45% NP40) supplemented with 100 μg/mL of proteinase K and incubated
621 at 55˚C overnight. Samples were heat-deactivated as previously described and used
622 for PCR.
623

624 *Transgenic embryos*

625 To generate transgenic constructs, the relevant enhancers (genomic coordinates
626 shown in supplementary table 3) were amplified by PCR (primers in supplementary
627 table 2) from genomic DNA and cloned into a vector containing the adenovirus major
628 late promoter, the coding region of the β-galactosidase gene and the SV40
629 polyadenylation signal [6]. The enhancers were confirmed by direct sequencing. The
630 constructs were isolated from the plasmid backbone, gel-purified using the
631 NZYGelpure (NZYTech #MB01102) and eluted from the columns with 50 μL of the
632 kit's elution buffer. The purified constructs were diluted in microinjection buffer (10 mM
633 Tris.HCl, 0.25 mM EDTA, pH 7.5) at 2 ng/μL and microinjected into the pronucleus of
634 fertilized FVB/N oocytes according to standard protocols [45]. Microinjected oocytes
635 were transferred into the uteri of pseudopregnant NMRI females and embryos were
636 recovered at E10.5 or E11.5 and stained for β-galactosidase activity.

637

638 *RT-qPCR*

639 To assess recombination efficiency, the posterior part of the E10.5 *Tgfbr1-cKO*
640 embryos, including their hindlimbs, were dissected in PBS (137 mM NaCl, 2.7 mM
641 KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4) on ice. Whole E10.5 *Tgfbr1*$^{-/-}$ embryos
642 were used as controls. RNA was extracted from fresh tissue using TRI reagent (Sigma,
643 #T9424) according to manufacturer's protocol. 200 ng of RNA from each sample was
644 used to prepare complementary DNA using a random hexamer mix (NZYTech
645 #MB12901) and following the protocol of the NZY Reverse Transcriptase enzyme
646 (NZYTech #MB12401). All cDNA samples were diluted 1:5 and 1 ml was used in
647 quantitative PCR (qPCR) reactions with iTaq Universal SYBR Green Supermix (Bio-
648 Rad, #1725124), according to manufacturer's instruction and using the Applied
649 Biosystems QuantStudio 7 flex qPCR System. Primers were localized in exon 3 and
650 exon 4 of *Tgfbr1* mRNA (listed in supplementary table 2), so that in case of successful
651 recombination the product is not amplified.

652 PCR product concentration in each sample was calculated using the standard curve
653 with following formula *Quantity = 10^(mean Ct-b/m)* (where "m" is slope and "b" is an
654 intercept of the standard curve), and *Tgfbr1* expression level was normalized to *b-*
655 *Actin* expression. Linear model was fitted (Tgfbr1 normalized expression ~ Genotype)
656 with R built-in function (lm). Analysis of variance (ANOVA) was performed on the linear

657 model with the built-in function in R (anova). Significance levels between groups of
658 samples were assessed by Tukey method using the “glht” function from “multicomp”
659 package in R. Differences were considered significant at *p<0.05, **p<0.01 and
660 ***p<0.001. Boxplot was generated with “ggplot2” and “ggdignif” packages in R.
661
662 *Cell culture and transfection*
663 HEK 293T cells (ATCC #CRL-3216) were cultured in Dulbecco’s modified Eagle’s
664 medium (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (FBS,
665 Life Technologies) and 1% penicillin/streptomycin at 37 °C in a 5% CO_2 atmosphere.
666 Cells were seeded in 35 mm plates and when they reached 70% confluency, the
667 medium was changed to transfection medium (DMEM supplemented with 10% FBS).
668 Expression vectors were constructed by amplifying the coding region of the *Tgfbr1*
669 transcript isolated from wild type or *Tgfbr1*$^{3ex3\text{-}flox}$ homozygous embryos by RT-PCR
670 (primers in supplementary table 2) and cloned into the pRK5 expression vector.
671 Constructs were verified by sequencing. 3 μg of the plasmids were transfected into
672 the 393T cells using Lipofectamine 2000 (Thermo Fisher Scientific #11668027)
673 according to the manufacturer's instructions. After 24 hours incubation, cells were
674 processed for western blot analysis
675
676 *Protein extraction and western blot*
677 The transfected cells were washed twice with ice-cold PBS and then 150 μL of lysis
678 buffer (50 mM Tris-HCl, 150 mM NaCl and 1% Triton X-100, pH8.0) was added to
679 each plate and incubated on ice for 10 minutes. Lysates were scraped into
680 microcentrifuge tubes on ice and centrifuged at 20000 rcf for 10 min at 4˚C. The
681 supernatant was collected, frozen on dry ice and stored at -80˚C. Protein lysates were
682 mixed (2:1) with 3X loading buffer (150 mM Tris.HCl, pH 6.8, 6% SDS, 30% glycerol,
683 7.5% β-mercaptoethanol, 0.03% bromophenol blue). Samples were incubated for 15
684 minutes at 65˚C and resolved by SDS-PAGE in a 10% polyacrylamide gel. Proteins
685 were transferred to PVDF membranes in 20% methanol, 25 mM Tris, 200 mM glycine
686 at 200 mA for 1 hour. The membranes were blocked in blocking solution [5% dry milk
687 dissolved in PBS containing 0.1% Tween-20 (PBT)] for 1 hour at room temperature
688 and then incubated with primary antibodies overnight at 4˚C. Primary antibodies were
689 anti-Tgfbr1 (Sigma-Aldrich #HPA056473, 1:1000 in blocking solution) and anti-actin

690 (Abcam #ab179467, 1:1000 in blocking solution). Membranes were then washed in
691 PBT and incubated with HRP conjugated-anti-rabbit IgG (GE Healthcare # NA9340,
692 1:5000 in blocking solution) at room temperature for 1 hour and washed twice in PBT.
693 Signals were developed by chemiluminescence using Amersham ECL Prime Western
694 Blotting Detection Reagent (GE Healthcare #RPN2232) and images were captured
695 using the GE Amersham Imager 600.
696
697 *Skeletal preparation*
698 E17,5 fetuses were recovered from pregnant females by cesarean section and
699 dissected from the extraembryonic membranes, eviscerated, the skin removed and
700 then fixed in 100% ethanol for 2 days. Cartilages are then stained by incubation with
701 450 mg/l of alcian blue (Sigma, #A5268) in 80% ethanol/20% acetic acid for one day.
702 Fetuses were then postfixed in 100% ethanol overnight. Tissue was cleared by
703 incubation in 2% KOH for 6 hours, followed by staining of ossified bones with a 50 mg/l
704 of alizarin red S (Sigma, #130-22-3) solution in 2% KOH for 3 hours. Specimens were
705 then further incubated in 2% KOH until tissues were fully cleared. Skeletons were
706 stored in 25% glycerol in water. All incubations were performed at room temperature
707 with rolling. Genotyping was performed on gut tissue by incubating in Laird's buffer
708 (100 mM Tris-HCl pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl) containing 10 μg/ml
709 proteinase K at 55ºC overnight. Genomic DNA was then recovered by precipitation
710 with isopropanol (1:1, vol:vol) and transferred to TE pH 8.0. Genotyping was then
711 performed by PCR using oligos specified in supplementary table 2.
712
713 *Optical projection tomography*
714 Optical projection tomography (OPT) was used to image E16.5 fetuses as previously
715 described [46] with minor modifications. Briefly, fetuses were recovered by cesarean
716 section in ice cold PBS, washed several times in PBS and fixed in 4%
717 paraformaldehyde (PFA) in PBS at 4ºC for several days. After several washes with
718 demineralized water fetuses were dehydrated by sequential incubation in
719 demineralized water with increasing concentrations of methanol (10% increases) and
720 then twice in 100% methanol. Fetuses were then bleached in a three day process with
721 a sequence of 2.5%, 5% and 10% hydrogen peroxide in methanol at room
722 temperature. Fetuses were then rehydrated through a reverse

723 methanol/demineralized water series, and embedded in a 0,7% low-melting agarose.
724 Clearing was done with a 1:2 solution of Benzoic Alcohol:Benzyl Benzoate (BABB)
725 using a BABB/methanol series (25% BABB increases). 100% BABBwas introduced
726 on day 3 and replaced every day for the next 4 days, until the fetuses became
727 completely transparent. Anatomical datasets were then obtained using a custom built
728 OPT scanner and procedures; briefly, green auto-fluorescence was acquired on 1600
729 sequential angles for a full revolution, the raw dataset was pre-processed with a
730 custom-built ImageJ macro, and then back-projection reconstructed using SkyScan's
731 nrecon as in [47], and then post-processed to reduce noise and enhance contrast in
732 ImageJ. 3D reconstruction and visualization and manual segmentation of the limb
733 skeleton and internal organs was performed using the Amira software (Thermo Fisher)
734
735 *Whole mount in situ hybridization and sectioning*
736 Embryos were fixed in 4% PFA overnight, then dehydrated through 25%, 50% and
737 75% series of methanol in PBT (PBS, 0,1% Tween 20), then incubated in 100%
738 methanol. Embryos were then rehydrated through a reverse methanol/PBT series and
739 incubated three times in PBT for at least 5 minutes each at room temperature.
740 Embryos were then bleached for 1 hour in 6% hydrogen peroxide in PBT and
741 permeabilized in 10 μg/mL of proteinase K in PBT for a time period that depended on
742 the embryo size. The reaction was then quenched with a 2 mg/mL solution of glycine
743 in PBT, washed twice in PBT and postfixed in a 4% PFA and 0,2% glutaraldehyde mix
744 for 20 minutes, followed by two washes in PBT. Hybridization was performed at 65˚C
745 overnight in hybridization solution (50% formamide, 1.3 x SSC pH 5.5 [20 x SSC is 3M
746 NaCl, 300 mM sodium citrate], 5 mM EDTA, 0.2 % Tween 20, 50 µg/mL yeast tRNA,
747 100 µg/mL heparin) containing the relevant digoxigenin-labelled antisense RNA
748 probes. The probes used in this work are listed in supplementary table 2. RNA probes
749 were transcribed in vitro from the vector for 1.5 hours at 37˚C with relevant RNA
750 polymerase and the DIG RNA Labeling Mix (Roche #11277073910). The reaction
751 product was verified in 0,8% agarose gel, precipitated with ethanol in the presence of
752 0.3M sodium acetate, pH 5.3 and resuspended in 100 μL of TE pH 8.0. Probes were
753 diluted in hybridization solution (6 μL in 1 ml of solution). After hybridization, embryos
754 were washed twice with hybridization solution without tRNA, heparin, at 65˚C for 30
755 min each, then in a 1:1 mix of hybridization solution and TBST (25 mM Tris.HCl, pH

756 8.0, 140 mM NaCl, 2.7 mM KCl, 0.1% Tween 20) for 30 minutes at 65ºC, and finally
757 washed three times with TBST at room temperature. Embryos were then equilibrated
758 in MABT (100 mM Maleic acid, 150 mM NaCl, 0.1 % Tween-20, pH 7.5) and blocked
759 in MABT blocking buffer [MABT containing 1% blocking reagent (Roche
760 #11096176001)] with 10% sheep serum for 2,5 hours at room temperature. Embryos
761 were then incubated overnight at 4ºC with a 1:2000 dilution of alkaline phosphatase-
762 conjugated anti-digoxigenin antibody (Roche #11093274910) in MABT blocking buffer
763 with 1% sheep serum. After extensive washes with MABT at room temperature,
764 embryos were equilibrated in NTMT buffer (100 mM Tris HCl, pH 9.5, 50 mM $MgCl_2$,
765 100 mM NaCl, 0.1% Tween-20) and developed with a 1:50 dilution of a NBT/BCIP
766 solution (Roche #11681451001) in NTMT at room temperature in the dark. The
767 reaction was stopped with PBT, and embryos postfixed with 4% PFA overnight at room
768 temperature, and stored in PBT. At least 3 embryos were used per probe. Pictures
769 were taken with a SWIFTCAM SC1803 connected to a stereoscope. Stained embryos
770 were embedded in a mix containing 0,45% gelatin, 27% bovine serum albumin (BSA),
771 18% sucrose, jellified with 1,75% glutaraldehyde and sectioned at 35 μm on a Leica
772 Vibratome VT 1000 S.
773
774 *β-galactosidase staining staining*
775 Embryos were dissected out in ice cold PBS and fixed in 4% PFA at 4ºC for 30
776 minutes. They were then washed three times in β-gal wash buffer (PBS plus 0.02%
777 Tween 20) for 10 minutes at room temperature and β-galactosidase activity developed
778 in PBS containing 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$, 0.02% Tween 20,
779 0.4 mg/ml X-gal (Promega #V3941) overnight in the dark at 37ºC. The reaction was
780 stopped with β-gal wash buffer, embryos postfixed in 4% PFA overnight at room
781 temperature and stored in PBS.
782
783 *ATAC-Seq*
784 ATAC-seq was performed as in [25] with minor modifications. E11.25 mouse embryos
785 were dissected on ice in DMEM High Glucose medium (Biowest #L0102-500)
786 containing 10% FBS (this will be referred to as medium in the rest of the protocol).
787 Hindlimbs and genital tubercles were dissected out and collected on ice-cold media.
788 Single cell suspension was prepared by treating the tissue with 500 μL of Accutase

789 (Sigma #A6964-500ML) for 30 minutes at 37˚C with shaking at 600 rpm. Single cells
790 were pelleted at 6000 rcf for 5 minutes at 4˚C, resuspended in 200 μL of media and
791 counted. 50000 viable cells from each sample were used for nuclei extraction. Cells
792 were incubated with 50 μL of ATAC resuspension buffer (10 mM Tris HCl pH 7.4, 10
793 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20 (Sigma-Aldrich, P7949), 0.01%
794 Digitonin (Target Mol #282T2721-1ml/10mM in DMSO) for 3 minutes on ice. Lysis was
795 quenched by adding 1 ml of ATAC resuspension buffer without NP-40 and Digitonin.
796 Nuclei were pelleted by centrifugation at 500 rcf for 10 minutes at 4˚C. The pellet was
797 then resuspended in 50 μL of transposition buffer [for 50 μL: 25 μL 2x TD buffer, 2.5
798 μL Tn5 transposition enzyme (100 nM final) (Illumina #15028212), 16.5 μL PBS, 0.5
799 μL 1% digitonin, 0.5 μL 10% Tween-20, 5 μL H_2O] and incubated at 37˚C for 30
800 minutes with shaking at 1000 rpm. The DNA was purified with Qiagen MinElute PCR
801 purification kit (Qiagen #50928004) and eluted in 20 μL of the kit's elution buffer.
802 Libraries were amplified by PCR with NEBNext High-Fidelity 2X PCR Master Mix (New
803 England Biolabs #174M0541S) for 9 cycles in 96-well Thermal cycler (Biorad) and
804 purified with Qiagen MinElute PCR purification kit. Tagmentation efficiency was
805 assessed on TapeStation 4200 (Agilent). Double size selection to remove primer
806 dimers and fragments exceeding 1 kb was performed using SPRIselect beads
807 (Beckman Coulter). Another quality control was performed with High Sensitivity DNA
808 assay using the Fragment Analyzer (Agilent). The 4nM libraries pool was sequenced
809 with Illumina NextSeq 2000 (100 cycles, Pair-end 50 bp). Two biological replicates
810 were performed per tissue.
811 The data obtained in this work has been deposited in the GEO repository, (accession
812 number GSE231592).
813
814 *ATAC-Seq bioinformatic and statistical analysis*
815 Data bioinformatic analysis was performed on the Galaxy server [48]. Raw sequencing
816 fastq files for each library were assessed for quality, adapter content and duplication
817 rates with FastQC (v0.11.9)
818 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapters were trimmed
819 and reads with length < 20 bp were removed using cutadapt (v1.16.5)[49] (3´ adapter
820 sequence: CTGTCTCTTATACACATCT). The trimmed reads were aligned to the
821 mouse reference genome (GRCm38/mm10 Dec. 2011) using Bowtie2 (v2.4.2)[50]

822 [parameters paired-end options (-X/–maxins 1000, --fr, --dovetail), --very-sensitive].
823 The aligned reads were filtered using BamTools Filter (v2.4.1) [51] (parameters --
824 isProperPair true, --mapQuality ≥ 30, and --reference!=chrM) and the duplicate reads
825 were removed using Picard MarkDuplicates (v2.18.2.2)
826 (http://broadinstitute.github.io/picard/). The resulting BAM files were converted to BED
827 and the reads that overlap to the blacklisted regions listed in
828 https://www.encodeproject.org/files/ENCFF547MET/ were removed using Bedtools
829 (v2.30.0) [52]. The filtered BED files were used as inputs for peak calling. Peaks were
830 called from individual replicates using MACS2 callpeak (v2.1.1) [53] (parameters --
831 format single-end BED, --nomodel, --extsize 200, --shift -100, --qvalue 0.05).
832 BAM files were then converted to BigWig with "deeptools" [54] normalizing to 1x effective
833 genome size (2652783500 for GRCm38/mm10). BigWig files were used for data
834 visualization in IGV with the addition of the conservation scoring by phyloP
835 (phylogenetic p-values)[55] for 60 vertebrate genomes from the UCSC genome browser
836 (mm10.60way.phyloP60way.bw file downloaded from
837 http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phyloP60way/) and ChIP-seq data
838 from Hoxa13 (GSE81356) [27] and Gli3 (GSE133710) [28].
839 The raw count matrix was created by merging peaks across samples as in [56,57].
840 DESeq2 object was built from the raw count matrix and blind dispersion estimated with
841 the "vst" function from the "DESeq2" package in R [58]. Top 2500 most variable peaks
842 estimated with the "rowVars" function in R were selected for the principal component
843 analysis (PCA) and sample distance matrix. PCA was performed using the built-in
844 "prcomp" function in R. Sample distance matrix was made using the built-in "dist"
845 function in R. For statistical analysis raw reads were analyzed using EdgeR [59].
846 ANOVA-like test was performed on all samples and pairwise comparisons between
847 the samples were made by analyzing individual contrasts. Venn diagrams were
848 created from the lists of selected differentially accessible peaks with "eulerr" package
849 in R. Heat Maps were made from dataframes containing log2 normalized counts
850 across samples with the "pheatmap" package (https://CRAN.R-
851 project.org/package=pheatmap.)
852
853 *Footprinting analysis of ATAC-Seq data*
854 We further used the program RGT HINT-ATAC [43] and motif matching to identify, plot
855 and compare transcription factors (TF) footprints in the different samples. In detail, the

BAM files containing the filtered aligned reads for each biological replicate were merged and used as a matrix for the footprinting analysis for the regions corresponding to the peaks called by MACS2 (BED file) (described above). A subset of the BAM and BED files corresponding to the regions that were found accessible only in GT (pattern 1 peaks) were used for the footprint analysis and then tested for motif matching to the Hocomoco database[60] for finding the motif-predicted binding sites (mpbs). Motif enrichment analysis was performed on the footprinting results for the peaks of pattern 1, using as a background the footprinting results from all the peaks called in GT.

Additional references

45. Hogan, B., Beddington, R., Constantini, F. & Lacy, E. *Manipulating the mouse embryo: a laboratory manual*. (Cold Spring Harbour Laboratory Press, 1994).
46. Dias, A., Martins, G. G., Lopes, A. & Mallo, M. Three and Four-Dimensional Visualization and Analysis Approaches to Study Vertebrate Axial Elongation and Segmentation. *Journal of Visualized Experiments* **168**, e62086 (2021).
47. Martins, G. G. *et al.* Optical projection tomography. in *Imaging Modalities for Biological and Preclinical Research: A Compendium, Volume 1: Part I: Ex vivo biological imaging* (eds. Walter, A., Mannheim, J. G. & Caruana, C. J.) i.2.f1-i.2.f11 (IOP Publishing, 2021). doi:10.1088/978-0-7503-3059-6ch12.
48. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**, W3–W10 (2016).
49. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–12 (2011).
50. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
51. Barnett, D. W., Garrison, E. K., Quinlan, A. R., Stŕmberg, M. P. & Marth, G. T. Bamtools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* **27**, 1691–1692 (2011).
52. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
53. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728–1740 (2012).
54. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
55. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* **20**, 110–121 (2010).
56. Rausch, T., Hsi-Yang Fritz, M., Korbel, J. O. & Benes, V. Alfred: Interactive multi-sample BAM alignment statistics, feature counting and feature annotation for long- and short-read sequencing. *Bioinformatics* **35**, 2489–2491 (2019).
57. Erarslan-Uysal, B. *et al.* Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors. *EMBO Mol Med* **12**, e12104 (2020).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

59. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
60. Kulakovskiy, I. V. *et al.* HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res* **46**, D252–D259 (2018).
61. Jones, C. M., Lyons, K. M. & Hogan, B. L. M. Involvement of Bone Morphogenetic Protein-4 (BMP-4) and Vgr-1 in morphogenesis and neurogenesis in the mouse. *Development* **111**, 531–542 (1991).
62. Davis, C. A. & Joyner, A. L. Expression patterns of the homeo box-containing genes and the proto-oncogene int-1 diverge during mouse development. *Genes Dev* **2**, 1736–1744 (1988).
63. Crossley, P. H. & Martin, G. R. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**, 439–451 (1995).
64. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* **16**, 154–160 (1997).
65. Aires, R. *et al.* Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev Cell* **48**, 383–395 (2019).
66. Szeto, D. P. *et al.* Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev* **13**, 484–494 (1999).
67. Echelard, Y. *et al.* Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**, 1417–1430 (1993).
68. Chapman, D. L. *et al.* Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Developmental Dynamics* **206**, 379–390 (1996).
69. Yamaguchi, T. P., Bradley, A., McMahon, A. P. & Jones, S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211–1223 (1999).

Supplementary table 2. Oligonucleotides used in this work.

Genotyping primers		
Tgfbr1-flox	Forward	TTGAGCTTGCTGTCTGACTGGATAG
	Reverse	CTGTGGTTGGCAGGCATGTG
Cre	Forward	CGAGTGATGAGGTTCGCAAG
	Reverse	CCTGATCCTGGCAATTTCGGCT
Tgfbr1-null	Forward	CTACTGTGTTTCAAATGGGAGGGC
	Reverse	GGCCTGTCGGATCCTATCATC
β-galactosidase	Forward	AGCAGTTTTTCCAGTTCCGTTTATC
	Reverse	AGCGGCGTCAGCAGTTGTTTTTTAT
Cloning of regulatory elements		
CR-Isl1	Forward	GACTCGAGTCTGTGATACAAAACAATATATC
	Reverse	GAGGATCCAATCTTTCAAAGACATGGAGGG
CR-Tbx5	Forward	GACTCGAGCTTTATGTATCTGAGCACACTG

	Reverse	GAAGATCTTCCTTCAACAAACCAT CCAC
CR-Fgf10	Forward	CAGCTCGAGACTAGTAATCTGAGGTGGTTGCTCACTC
	Reverse	CATCCATGGCTATGGAAGGCTTATGTATCAC
Grem1-Enh	Forward	GACGTCGACACTAGTGTCATGTGTTCCTGTGATTCTG
	Reverse	GAGCCTGTATAAGAAGTTCAGGGCCA TGGCAG
Other primers		
qPCR-Tgfbr1	Forward	ACCGTGTGCCAAATGAAGAGG
	Reverse	CATCTAGATCTTGTAACACAATGGTCCTGGC
qPCR-Actin	Forward	ATGAAGATCCTGACCGAGCG
	Reverse	TACTTGCGCTCAGGAGGAGC
Tgfbr1-CDS	Forward	GAGAATTCCGGGCCACAAACAGTGGC
	Reverse	GAGTCGACCATTTTGATGCCTTCCTGT TGGC
Oligonucleotides for generating in situ probes		
Lmx1b	Forward	CTGCTGTGCAAGGGTGACTATGAG
	Reverse	GAGGCAAAGTAGGAGCTCTGCATG
Fgf10	Forward	TTCCTAGAAGTTATGGATGTTG
	Reverse	GTACTGCATCCACCAACAGTG
Grem1	Forward	AATGAATCGCACCGCATACAC
	Reverse	GACTAATACGACTCACTATAGGGAAGCAACTGCTGGTTCTTCTG
In situ probes		
Bmp4	Jones et al 1991, ref [61]	
En1	Davis and Joyner, 1988, ref [62]	
Fgf10	Whole coding region cloned into TOPO. Oligonucleotides listed above	
Fgf8	Crossley and Martin, 1995, ref [63]	
Grem1	PCR fragment with the whole coding region and T7 promoter at the 3' end. Oligonucleotides listed above	
Hand2	Srivastava et al, 1997, ref [64]	
Isl1	Jurberg et al 2013, ref [6]	
Lin28a	Aires et al., 2019, ref [65]	
Lmx1b	PCR fragment encompassing nucleotides 504 to 1202 of the mRNA cloned into pKS bluescript. Oligonucleotides listed above.	
Pitx1	Szeto et al, 1999, ref [66]	
Shh	Echelard et al, 1993, ref [67]	
Tbx5	Chapman et al, 1996, ref [68]	
Wnt5a	Yamaguchi et al, 1999, ref [69]	

930

931 **Supplementary table 3.** Coordinates of the genomic elements used in the reporter
932 assays (from the GRCm38/mm10 reference genome).

933

Element	Coordinates
CR-Isl1	chr13:116,284,024-116,286,017
CR-Tbx5	chr5:119,700,950-119,703,188
CR-Fgf10	chr14:28,412,202-28,414,123
Grem1 enhancer	chr2:113,580,726-113,582,264

934