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. 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. *Tgfbr1* controls the response to those factors by modulating 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.

Introduction

The vertebrate body is built sequentially in a head to tail progression. While this is a continuous process, it involves two transitions entailing major changes in gene regulatory mechanisms. The first transition results in the switch from a head to a trunk developmental program [1], marking the start of axial extension and the layout of the primordia for most of the organ systems involved in vital and reproductive functions. The second transition organizes the end of trunk structures and activates the tail developmental program [2]. The transition from trunk to tail entails significant reorganization of embryonic structures involving derivatives from all germ layers. The neuromesodermal competent (NMC) progenitors [3,4] relocate from the epiblast to the tail bud, from where they keep extending the body axis [5,6]. The progenitors for the lateral plate mesoderm, involved in the formation and vascularization of the trunk-resident organs [7], undergo terminal differentiation leading to the formation of the hindlimbs and of the ventral lateral mesoderm that will then become the pericloacal mesoderm [2]. Reciprocal interactions between the pericloacal mesoderm and the endodermal cloaca will then 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 [8–12]. Although the hindlimb shares many of the patterning and morphogenetic processes with the forelimb [13], genetic analyses indicate major regulatory differences controlling the earliest stages of development of the two appendages [14–16]. Interestingly, molecular studies revealed that the hindlimb and the GT share many of their regulatory processes [9,17–19], indicating that they might be closely connected developmentally. Indeed, it has been reported that in some tetrapod species hindlimbs and genitals share a common primordium, which might represent the ancestral condition [20]. In mammals, it has been suggested that the posterior displacement of the cloaca relative to the hindlimb bud disconnected the limb primordium from the influence of the endodermal signals promoting genital fates, thus allowing formation of both legs and external genitalia [20].

Genetic studies, mostly in mouse embryos, revealed that the control of the early stages of hindlimb and cloacal/pericloacal development requires functional input from several regulatory factors. These include signaling pathways, like those of the transforming growth factor β/bone morphogenetic protein (Tgfβ/BMP)

superfamily [21–25], fibroblast growth factors (FGFs) [26–28], sonic hedgehog (Shh) [29–31] and both canonical and non-canonical WNTs [10,11,32], as well as several transcription factors, including Hoxa13, Hoxd13 [33,34] and Isl1[35]. Importantly, while the same regulatory factors play essential roles in both structures [9,18], their functional outcome differ substantially in the hindlimb and the GT, indicating the existence of mechanisms canalizing distinct patterns of response in the hindlimb and the pericloacal mesoderm when exposed to a particular regulatory factor. Shh provides paradigmatic example of such functional duality. It determines anterior/posterior limb bud polarity from the posterior limb bud mesenchyme regulating the identity of the limb skeletal elements [36]. Conversely, from its expression domain in the endodermal cloaca it controls the morphogenesis of the GT, as well as the organization of the different outlets for the urogenital and intestinal systems [29,30]. The mechanisms regulating this distinct tissue response to common regulatory factors are largely unknown.

Genetic analyses have shown that the Tgfβ receptor 1 (*Tgfbr1*) (also known as *Alk5*) plays a key role in the activation of the trunk to tail transition, also initiating the regulatory sequence controlling formation of the hindlimb and external genitalia [2,37,38]. Indeed, *Tgfbr1* null mutant embryos fail to induce the primordia of the hindlimb or cloacal/pericloacal structures, as shown by the absence of early markers or morphological landmarks for these tissues [37]. 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 substantial anterior displacement of the hindlimb buds [2].

It has been shown that the Growth and Differentiation factor 11 (Gdf11) is a physiological ligand for Tgfbr1 in its role at the trunk to tail transition [38]. Indeed, *Gdf11* mouse mutant embryos show a significant delay in the transition to tail development [2,39]. Expression and functional analyses in several other vertebrates indicate that this role for Gdf11 is not restricted to the mouse but might be shared by most other tetrapods [40,41]. However, contrary to *Tgfbr1* mutant embryos, *Gdf11* mutants still activate the transition to tail development, although with a significant delay, indicating that other ligands in the Tgfβ/BMP family might cooperate with Gdf11 in triggering Tgfbr1 activity in axial tissues. One such factor is Gdf8 as revealed by the enhanced trunk size observed in *Gdf11*/*Gdf8* compound mutants

104 [42]. Bmp4 and Bmp7 also stand out as potential candidates to mediate some of the
105 Tgfbr1 activities in the caudal body. Both *Bmp7/Tsg* mutants and compound *Bmp7*
106 homozygous/*Bmp4* heterozygous conditional inactivation in the region of the trunk
107 to tail transition resulted in the fusion of the hindlimbs (sirenomelia) due to the
108 absence of ventral lateral mesoderm [22,23] and, thus, of the pericloacal mesoderm
109 that builds the GT [20], which is also absent from *Tgfbr1* null mutant embryos [21].
110 Although BMPs are not considered as canonical Tgfbr1 ligands [43], given the high
111 complexity of interactions between different receptors of the Tgfβ/BMP superfamily
112 [44–47], functional interactions between Tgfbr1 and Bmp4/Bmp7 cannot be ruled out,
113 most particularly because the bona fide receptors for those BMPs in the
114 pericloacal/hindlimb area await identification.
115 The essential role of Tgfbr1 signaling in the induction of both hindlimb and GT
116 primordia despite generating morphologically and functionally distinct structures,
117 prompted us to evaluate whether it also played a role at later stages of hindlimb
118 and/or GT development. For this we inactivated *Tgfbr1* in the caudal embryonic
119 region after the primordia for the hindlimb and GT were already laid down. We found
120 that in the absence of *Tgfbr1* the pericloacal mesoderm is converted into a second
121 set of hindlimbs. By comparing the chromatin accessibility profiles of wild type and
122 mutant tissues, we could observe that the mutant pericloacal mesoderm acquires a
123 significant number of limb-type signatures, while losing many of those specifically
124 associated with the GT. Our results indicate that Tgfbr1 modulates the type of
125 response of the pericloacal mesoderm to patterning signals from the endoderm by
126 remodeling the regulatory landscape of the chromatin, thus defining a limb or a GT-
127 type of gene activation.
128
129 **Results**
130 *Generation of Tgfbr1-cKO developmental model.*
131 For the conditional *Tgfbr1* inactivation approach we used three *Tgfbr1* alleles,
132 namely the *Tgfbr1* null allele [37] and either $Tgfbr1^{flox}$, carrying LoxP sites flanking
133 Exon 3 [48], or $Tgfbr1^{3ex3\text{-}flox}$, containing a triplicated Exon 3 each surrounded by LoxP
134 sites (supplementary Fig. 1). Characterization of mice homozygous for the
135 $Tgfbr1^{3ex3\text{-}flox}$ allele revealed that this allele generates a transcript containing a
136 tandem of three Exon 3 sequences in frame with the flanking Exons 2 and 4

137 (supplementary Fig. 1C). Expression in HEK293T cells of the *Tgfbr1*^{3ex3} mRNA
138 isolated from the mouse embryos confirmed that it generates a protein product with
139 the expected size of Tgfbr1 containing triplicated Exon 3 sequences (supplementary
140 Fig. 1D). Exon 3 encodes the transmembrane domain and major part of the GS
141 domain of the protein, including the serin-rich region phosphorylated upon ligand
142 binding and implicated in signal transduction [43,49]. Homozygous *Tgfbr1*$^{3ex3\text{-}flox}$
143 animals were not viable, and *Tgfbr1*$^{3ex3\text{-}flox/-}$ fetuses phenocopied the knock out of
144 the Tgfbr1 ligand *Gdf11* [2,39] (supplementary Fig. 1B, B'). Together, the above results
145 strongly suggest that *Tgfbr1*$^{3ex3\text{-}flox}$ produces a hypomorphic receptor.
146 *Tgfbr1* inactivation was promoted using the *Cdx2Cre*^{ERT} transgenic driver [2],
147 triggering Cre activity in caudal embryonic tissues by tamoxifen administration
148 (supplementary Fig. 1E). Experiments using either *Tgfbr1* floxed allele gave similar
149 phenotypes, but while *Tgfbr1*$^{flox/-}$*;Cdx2Cre*$^{ERT+/0}$ embryos required two consecutive
150 tamoxifen doses at embryonic stage (E) 6.75 and E7.25, a single tamoxifen injection
151 at E7.25 was sufficient to obtain the same mutant phenotype from for *Tgfbr1*$^{3ex3\text{-}flox/-}$
152 *;Cdx2Cre*$^{ERT+/0}$ embryos, thus reducing the frequency of tamoxifen-derived late fetal
153 miscarriage. Our tamoxifen treatment scheme inactivated *Tgfbr1* in the tissues
154 caudal from the forelimb bud, as confirmed by RT-qPCR analyses on RNA
155 extracted from the posterior part of E10.5 embryos, including the hindlimbs
156 (Supplementary Fig. 1F, G). Tamoxifen-treated embryos and fetuses with either the
157 *Tgfbr1*$^{flox/-}$*;Cdx2Cre*$^{ERT+/0}$ or the *Tgfbr1*$^{3ex3\text{-}flox/-}$*;Cdx2Cre*$^{ERT+/0}$ genotype will be
158 hereafter referred to as *Tgfbr1-cKO*.
159
160 *Tgfbr1-cKO exhibit multiple developmental malformations*
161 The most prominent feature of *Tgfbr1-cKO* mutants was a duplication of their
162 hindlimbs [4/5 at E16.5] (Fig. 1A, B). The fifth fetus analyzed at this stage lacked
163 overt hindlimb duplication but still contained small protrusions posterior to the
164 hindlimbs, a phenotype likely derived from less efficient recombination. The limb
165 identity of the duplicated structures was confirmed by the presence of skeletal
166 structures that, although variable in morphology, were clearly identified as
167 belonging to limbs (Fig. 1C; supplementary Fig. 2; Supplementary movie 1). For
168 instance, a close look at one such duplicated hindlimb showed the presence
169 stylopod, zeugopod and autopod elements. When these last two sections were
170 analyzed in more detail using light sheet microscopy, in addition to the presence of

171 two skeletal elements in the zeugopod, we could identify individual phalanxes in
172 several of the fingers, although with somewhat abnormal patterns, as well as
173 polydactyly (supplementary Fig. 2).
174 Another prominent phenotype of *Tgfbr1-cKO* fetuses was the presence of an
175 omphalocele [5/5 at E16.5] (Fig. 1B). This phenotype, but not the duplicated
176 hindlimbs, was also observed in *Tgfb2*/*Tgfb3* double mutants [50]. This indicates both
177 that Tgfbr1 might mediate Tgfb2 and Tgfb3 activity in the body wall and that different
178 ligands should trigger Tgfbr1 signaling in the body wall and the pericloacal/hindlimb
179 mesoderm. The three *Tgfbr1-cKO* mutants analyzed either by optical projection
180 tomography (OPT) or by histological sections lacked kidneys (Fig. 1D-E';
181 Supplementary movies 2 and 3), a trait also observed in mutant fetuses for *Gdf11*
182 [2,39]. The cloacal-derived tissues were also strongly compromised in these mutant
183 embryos (Fig. 1D-E'; supplementary Fig. 3C,C'; Supplementary movies 2 and 3).
184 The gut failed to generate normal rectal-anal structures. Instead, the caudal end of
185 the intestinal tube merged with a structure that could be identified as a small bladder
186 connected to a hypomorphic urethra, thus resembling a persistent cloaca. The
187 external genitalia were also virtually absent from *Tgfbr1-cKO* fetuses. These
188 phenotypes are likely to result from interference with the functional interactions
189 between the cloacal endoderm and the pericloacal mesoderm driving normal
190 morphogenesis of the cloacal region [9,11,18,19,31]. Indeed, endodermal malformations
191 in this area were already evident in *Tgfbr1-cKO* mutants at mid gestation. At E10.5
192 and E11.5, the mutant cloaca was larger than in control embryos, lacked features
193 associated with normal septation, and often contained protrusions entering the
194 intra-cloacal lumen (2/4 embryos, supplementary Fig. 3a-B'). Remains of these
195 protrusions persisted later in development at E16.5 in the malformed bladder
196 (supplementary Fig. 3C, C').
197
198 *The extra set of hindlimbs in Tgfbr1-cKO mutants derives from the pericloacal*
199 *mesoderm*
200 Morphological alterations in the hindlimb/pericloacal area of *Tgfbr1-cKO*
201 embryos were already visible at mid-gestation (Fig. 2). *Fgf8* expression in E10.5
202 embryos revealed a posterior and medial expansion of the apical ectodermal ridge
203 (AER) of the hindlimb bud almost reaching the ventral midline of the embryo (Fig.
204 2A, A'), suggesting an extension of the hindlimb field into the pericloacal mesoderm

205 of *Tgfbr1-cKO* embryos. This was associated with several expression changes in
206 the pericloacal region. For instance, hindlimb specific markers, like *Pitx1* and *Lin28a*
207 were at E10.5 ectopically activated in the pericloacal region (Fig. 2B-C').
208 Conversely, *Tbx5* expression, normally observed in the pericloacal mesoderm, and
209 later in the GT, but not in the hindlimb, could not be detected in this region of E11.5
210 *Tgfbr1-cKO* embryos (Fig 2G, G'), and *Fgf8* expression was absent from the cloacal
211 endoderm of the mutant embryos (Fig 2A, A'). Together these observations suggest
212 a change of identity in the pericloacal mesoderm, becoming incorporated into the
213 hindlimb field instead of entering its normal GT developmental fate.
214 Expression of genes active in both the hindlimb and the pericloacal region/GT,
215 was also consistent with the pericloacal tissues taking a hindlimb fate. For instance,
216 *Wnt5a*, which in wild type embryos is expressed in the limb buds beneath the AER
217 and in the pericloacal region and GT, was not detected next to the cloacal endoderm
218 of E10.5 *Tgfbr1-cKO* embryos, being instead observed beneath the AER along the
219 extended hindlimb domain into the pericloacal region (Fig. 2E, E'). This pattern also
220 revealed that the hindlimb domain of *Tgfbr1-cKO* embryos become split in two
221 prominences, the first indication of the generation of two individual hindlimb
222 structures from extended hindlimb bud of the mutant embryos. *Fgf8* expression in
223 this region of *Tgfbr1-cKO* mutants at E11.5 was also consistent with the split of the
224 hindlimb bud in two fields, as it was observed through the apical border of the
225 hindlimb region but demarcating two apparent prominences separated by a valley
226 (Fig. 3A, A'). Additional support for the conversion of pericloacal mesoderm into
227 hindlimb identity was provided by genes showing dynamic expression patterns in
228 these structures. In wild type embryos, *Isl1* is initially activated in the hindlimb and
229 GT primordia, becoming later downregulated in the hindlimb while keeping strong
230 expression in the GT [51] (Fig 2H). At E11.5 *Tgbfr1-cKO* mutants showed residual
231 *Isl1* expression mainly in the expanded endodermal cloaca (note that *Isl1* normal
232 expression profile includes the cloacal endoderm [35]) but became downregulated in
233 the adjacent mesoderm in the area likely to generate the extra hindlimb bud (Fig
234 2H'). Conversely, *Fgf10*, an essential gene for limb induction [52], whose expression
235 in the GT is observed only at later stages [26], was strongly activated already at E10.5
236 following limb-like patterns, reaching mesodermal tissue adjacent to the endoderm
237 (Fig. 2D, D').

238 Although clearly not as well defined as in wild type embryos, a degree of
239 anterior-posterior patterning could still be observed in the hindlimb region of *Tgfbr1-*
240 *cKO* embryos, consistent with changes in digit identity observed in the autopod
241 skeleton of the E16.5 fetuses (supplementary Fig. 2). For instance, at E10.5, *Hand2*
242 was expressed in the mutants as a continuous domain that run medially into
243 pericloacal tissue, showing a slightly stronger expression in the posterior area of
244 the two emerging prominences of the hindlimb bud, although not as clearly
245 restricted to the posterior limb bud mesenchyme observed in wild type embryos
246 (Fig. 2F, F'). A similar trend was observed at E11.5, when *Hand2* expression
247 followed a slight anterior-posterior distribution, being excluded from the anterior-
248 most region of the anterior limb bud prominence (Fig. 3B, B') and showed a slight
249 stronger expression in the posterior part of the caudal prominence of the mutant
250 limb bud (Fig. 3B, B'). Expression of the anterior marker *Pax9* was observed in the
251 anterior border of the anterior prominence of the limb bud in late E11.5 *Tgfbr1-cKO*
252 embryos (Fig. 3C, C'). However, we could not detect *Pax9* signal associated with
253 the posterior prominence. Whether this indicates real absence of *Pax9* expression
254 in this region or just delayed onset of expression remains to be determined.
255 Interestingly, the posterior prominence of the mutant hindlimb bud retained a
256 considerable degree of dorso-ventral patterning, as revealed by *En1* and *Lmx1b*
257 expression at E10.5 that followed typical limb patterns [53,54], although they were not
258 as well defined as in control hindlimb buds (Fig. 3D-e').
259 It has been shown that regulation of GT development relies largely on signals
260 from the cloacal endoderm, with Shh playing a central role in this process [30,31]. The
261 observation that *Wnt5a*, a known Shh target in the GT [31], was not activated in the
262 pericloacal mesoderm of *Tgfbr1-cKO* despite strong endodermal *Shh* expression
263 (Fig. 4A, A'; supplementary Fig. 3a-B'), suggested that either this mesenchyme
264 became refractory to Shh activity or that it changed its response profile. The
265 expression of the Shh target *Gli1* was consistent with the second possibility, as it
266 showed equivalent patterns in the pericloacal mesenchyme in both wild type and
267 *Tgfbr1-cKO* embryos (Fig. 4B, B'), thus indicating that this tissue still responds to
268 the Shh signal from the endoderm.
269
270 *Reorganization of the chromatin regulatory landscape in the pericloacal mesoderm*
271 *of Tgfbr1-cKO embryos*

272 We explored the mechanisms modulating the pericloacal mesoderm response
273 to Shh (and most likely also to other regulatory factors) by comparing the genomic
274 accessibility profiles in the posterior hindlimb prominences (extra hindlimbs) and
275 region adjacent to the cloaca (we will refer to it as mutant GT) of *Tgfbr1-cKO* with
276 those of hindlimb buds and GT of control embryos by ATAC-seq [55] (Fig. 4C).
277 Principal component (PC) analysis of those profiles revealed two main features.
278 PC1 grouped the tissues from *Tgfbr1-cKO* embryos together with the wild type GT
279 and clearly apart from the control hindlimb, consistent with the pericloacal origin of
280 the extra hindlimb of the mutant embryos. PC2, on the contrary, grouped the extra
281 hindlimb of the mutant embryos together the wild type hindlimb, separated from the
282 control GT (Fig. 4D).
283 We further explored the PC1 and PC2 by annotating 500 top peaks
284 contributing for each of the PCs (top loadings). For both PC1 and PC2, most of the
285 top loadings were in distal intergenic regions (46,2% and 48,6%, respectively), while
286 only small proportion of the peaks (6,8% for PC1 and 8,6% for PC2) were attributed
287 to potential promoter regions (<= 5 kb upstream of TSS) (Fig. 4E). This suggested
288 that most of the genomic regions whose accessibility patterns are influenced by the
289 Tgfbr1 might represent distal regulatory elements.
290 Hierarchical clustering of the top loadings contributing to the PC1 reveled the
291 existence of two distinct clusters: cluster 1 represent peaks less accessible in the
292 control limb than in other conditions, while cluster 2 contains the peaks more
293 accessible in the control limb than in other conditions (Fig 4F). These two clusters
294 represent limb specific regions regulated by Tgfbr1 signaling. Top loadings of PC2
295 formed 4 clusters based of their accessibility patterns (Fig. 4G). Clusters 1 and 3
296 are particularly interesting, because they represent the peaks with accessibility
297 patterns common between control and mutant limbs, and different from wild type
298 and mutant GT. These clusters, therefore, are likely to contain the regulatory
299 elements specifying limb fates and being responsible for canalizing the mutant
300 pericloacal mesenchyme towards the limb fate.
301 To further elaborate on the potential significance of conserved patterns within
302 top PC loadings clusters, we performed a series of complementary differential
303 analyses on our samples. We identified two accessibility patterns suggestive of their
304 involvement in the differential control of cell fate decisions of the pericloacal
305 mesoderm. First, we identified a set of peaks more accessible in the GT than in the

306 hindlimb samples from wild type embryos, and that showed significantly reduced
307 accessibility in the tissues from *Tgfbr1-cKO* embryos (1650 regions with
308 $Log_2FC>1.5$; FDR< $1e^{-3}$) (Fig. 4H, H'). These elements might represent enhancers
309 involved in the regulation of genes differentially expressed in the pericloacal region
310 of wild type embryos. The loss of accessibility to these elements in the *Tgfbr1-cKO*
311 mutants could thus contribute significantly to the inability of the pericloacal
312 mesoderm of the mutant embryos to enter their normal GT fate.

313 Another group of chromatin regions were significantly more accessible in
314 hindlimb-generating tissues of wild type and *Tgfbr1-cKO* embryos than in GT
315 tissues (526 peaks with $Log_2FC>1.5$; FDR<$1e^{-3}$ are shown in Fig. 5A, A'). These
316 regions represent a pattern analog to that observed in clusters 1 and 3 of PC2
317 whose accessibility patterns did not diverge between mutant and control hindlimb
318 ($Log_2FC<0.5$; FDR>0.1), indicating that they might represent the most relevant
319 elements driving the mutant pericloacal mesenchyme towards limb identity. For
320 further description we will refer to the GT-specific elements as pattern 1 elements
321 and to those associated with hindlimb tissues pattern 2 elements.

322 We identified pattern 1 elements within the *Tbx5, Isl1* and *Wnt5a* genomic
323 regions, three of the pericloacal mesoderm markers down-regulated in the
324 pericloacal region of *Tgfbr1-cKO* mutants (Fig. 4I, J, K), which will serve as a proof
325 of principle for the involvement of pattern 1 elements in GT expression. These
326 elements mapped to highly conserved genomic areas, suggesting their possible
327 involvement in regulatory processes. When tested using a transgenic reporter
328 assay, one of the elements associated with *Tbx5* showed activity in the pericloacal
329 region [n=6/14] (Fig 4I'). Interestingly, it did not activate overt forelimb expression
330 (supplementary Fig. 4A), the major expression domain of this gene [14], further
331 indicating specificity for the pericloacal region. Similarly, the element identified
332 downstream of *Isl1* also largely reproduced *Isl1* expression in the GT when tested
333 in transgenic reporter assays [n=5/12] (Fig. 4J', Fig. 6B, and supplementary Fig.
334 4C). An element associated with *Wnt5a* was also able to activate expression in the
335 GT, although much less frequently than the other elements [n=2/18] (Fig 4K'). This
336 element was also active in the tail, but it did not promote expression in the hindlimb.
337 Together, these observations are consistent with these elements indeed
338 representing enhancers potentially involved in the expression of the relevant genes
339 in the pericloacal area.

340 We found pattern 2 elements in regions associated with genes playing
341 essential roles during the earliest stages of limb development, which could thus play
342 a relevant role in promoting limb fates from the pericloacal mesoderm (Fig. 5C, D).
343 As for pattern 1 elements, these regions were also highly conserved among
344 vertebrates. We found pattern 2 elements in the *Fgf10* genomic region (Fig. 5C), a
345 gene activated in the mesenchyme adjacent to the cloaca entering hindlimb fate in
346 the mutant embryo (Fig. 2D, D'). Interestingly, published ChIP-seq data from
347 forelimb buds [56,57] showed binding of Gli3 and Hoxa13 to one of these elements,
348 located within an intronic region of this gene (Fig. 5C), indicating that it might
349 respond to Shh and/or Hoxa13 activities in the developing limb buds. The lack of
350 accessibility of this element in the GT of wild type embryos might thus suggest that
351 Tgfbr1 signaling renders this enhancer blind to endodermal Shh and/or the strong
352 pericloacal Hoxa13 (and maybe also Hoxd13) expression [34], consistent with absent
353 *Fgf10* at early stages in this tissue (Fig. 2D). Rather surprisingly, this element failed
354 to generate reporter activity in transgenic embryos [n=10]. We still do not
355 understand whether this reflects real lack of activity or the absence of proper
356 genomic context for their activity.
357 Another potentially relevant pattern 2 element was located within the *Fmn1*
358 locus, matching the position of one of the enhancers (element 7 in [58]) controlling
359 *Grem1* in the limb buds (Fig 5D) that had been shown to respond to Shh activity
360 from the zone of polarizing activity (ZPA) [58,59]. In wild type embryos *Grem1*
361 expression is not detected in the pericloacal mesoderm or in the GT (Fig. 5E),
362 indicating that in these tissues *Tgfbr1* might render the *Grem1* regulatory elements
363 blind to endodermal Shh. The inaccessible state of this *Grem1* enhancer in the GT
364 (Fig. 5D) is consistent with this hypothesis. Conversely, the gain of accessibility of
365 the *Grem1* enhancer in the extra hindlimb suggests that in the absence of *Tgfbr1*
366 the pericloacal mesoderm could become responsive to the endodermal Shh.
367 Considering the capacity of *Grem1* to expand the expression of AER markers [60–62],
368 this gene is a candidate to contribute to the observed AER extension in the hindlimb
369 primordium of *Tgfbr1-cKO* embryos. *Grem1* expression analyses in *Tgfbr1-cKO*
370 embryos revealed unexpected complexities. Some features of the expression
371 pattern were consistent with *Grem1* becoming responsive to endodermal Shh, as
372 expression was detected in the caudal-most end of the expanded hindlimb next to
373 the endoderm (Fig. 5E'). Rather surprisingly, however, *Grem1* expression was

374 strongly reduced or absent throughout the mesoderm of the extended hindlimb bud
375 of *Tgfbr1-cKO* embryos (Fig. 5E'). *Grem1* limb expression depends on Shh activity
376 from the ZPA [58,63,64], suggesting that the absent *Grem1* expression resulted from
377 the abnormal *Shh* expression observed in the hindlimb buds of *Tgfbr1-cKO*
378 mutants, which followed spotty patterns beneath the AER instead of forming a well-
379 defined ZPA in the posterior limb bud mesenchyme (Fig. 4A', supplementary Fig.
380 3A').

381 We explored whether this abnormal *Shh* expression resulted from a functional
382 impact of the absence of Tgfbr1 activity on the ZRS, the regulatory element
383 controlling *Shh* expression in the ZPA [65]. As expected, this region was accessible
384 in wild type hindlimbs (Fig. 5F). Conversely, it was inaccessible in the wild type GT,
385 consistent with the absence of the *Shh* expression in the pericloacal mesoderm or
386 the GT mesoderm, despite the presence of relevant activators of this enhancer, like
387 Hand2, Hoxa13 and Hoxd13 [17,34,66,67] [the endodermal *Shh* expression is under the
388 control of a different enhancer [68], which was accessible in both GT samples
389 (supplementary Fig. 3D), consistent with conserved endodermal *Shh* expression in
390 the *Tgfbr1-cKO* embryos]. Remarkably, the ZRS was also non-accessible in the
391 hindlimb region of *Tgfbr1-cKO* embryos, consistent with the observed abnormal
392 *Shh* limb expression (Fig. 5F). Therefore, Tgfbr1 seems to also control ZRS
393 accessibility to their regulatory transcription factors. Interestingly, this enhancer is
394 part of an additional pattern (pattern 3) consisting of elements requiring Tgfbr1
395 activity to become accessible in the limb bud. Analysis of the ATAC-seq data
396 revealed that 5764 ($Log_2FC>1.5$; $FDR<1e^{-3}$) elements fit the criteria of this pattern
397 (Fig. 5B, B'). The complementary tissue specificities of patterns 1 and 3 indicates
398 context dependency for Tgfbr1 activity in the modulation of genomic configuration.
399

400 *Footprinting analysis on elements differentially accessible in the GT.*

401 The canonical Tgfbr1 signaling pathway involves activation of Smad2 or
402 Smad3 that recruit Smad4 and enter the nucleus where they interact with Smad-
403 responsive elements to regulate gene expression [69]. It has been shown that Smads
404 can interact with chromatin remodelers, like Foxh1 or components of the SWI/SNF
405 complex to promote accessibility of enhancers that are later regulated by Smad
406 transcriptional activity [70–72]. To understand whether a similar mechanism might be
407 involved in the reorganization of the regulatory landscape in the pericloacal

408 mesoderm we performed a transcription factor footprint analysis and subsequently
409 motif enrichment on the ATAC-seq profiles using HINT-ATAC [73]. We did not find
410 enrichment of Smad footprints in the accessible regions of the pericloacal
411 mesoderm of wild type embryos (supplementary Table 1). Similarly, the
412 transcription factor Foxh1, previously shown to interact with Smad proteins was also
413 not found enriched in these elements. These results indicate that the mechanism
414 by which Tgfbr1 signaling modulates accessibility of regulatory regions is likely to
415 be different from that described for the regulation by Tgfβ or Nodal in cell lines [70–
416 72]. The lack of enrichment for Smad footprints in the elements specifically accessible
417 in the GT of wild type embryos indicate that Smad-dependent signaling is unlikely
418 to be a major component participating in the promotion of genital fates in the
419 pericloacal mesoderm.

420 Additional analysis of these footprints revealed enrichment in signals
421 associated with factors known to play significant roles in the development of the
422 external genitalia, including Hoxa13, Hoxd13, Isl1, Six2 or effectors of the canonical
423 Wnt pathway like Lef1, Tcf7 or Tcf7l. As a proof of principle to understand the
424 relevance of these footprints in the regulation of pattern 1 elements, we focused on
425 a Lef1 binding motif identified in the TF footprint analysis within the *CR-Isl1*
426 chromatin region analyzed above (green rectangle in Fig. 6A). We tested the
427 functional relevance of this potential Lef1 binding site, by generating a mutant
428 version of this enhancer lacking this site (*CR-Isl1*$^{\Delta Lef1}$) and testing its activity in the
429 β-galactosidase transgenic reporter assay (Fig. 6C and supplementary Fig. 4D).
430 When compared to the wild type enhancer (Fig. 6B, Fig. 4J' and supplementary Fig.
431 4C) *CR-Isl1*$^{\Delta Lef1}$ transgenic embryos showed reduced activity in the GT
432 mesenchyme, and no activity in the posterior part of the hindlimb typically observed
433 in the control embryos [6/15] (Fig. 6C and supplementary Fig. 4D). In addition, *CR-*
434 *Isl1*$^{\Delta Lef1}$ activated β-galactosidase expression in the otic vesicle (Fig. 6C and
435 supplementary Fig. 4D), likely derived from a conserved Gata3 motif within the TF
436 footprint (red rectangle in Fig. 6A) at level comparable to that observed in *CR-Isl1*
437 transgenics (Fig. 6B, supplementary Fig. 4C), which served as an internal control
438 to evaluate reporter expression. This is consistent with potential relevance of this
439 Lef1 binding site for the activity of this *Isl1* associated regulatory element, and thus,
440 to respond to canonical Wnt signaling in the GT. It is possible that the residual

441 activity of *CR-Isl1*$^{\Delta Lef1}$ in the GT derived from the presence of an additional Lef1
442 binding motif slightly upstream on the TF footprint (blue rectangle in Fig. 6A). We
443 therefore tested the activity of a deletion mutant of this enhancer including both Lef1
444 binding motifs and the region between them (*CR-Isl1*$^{\Delta\Delta Lef1}$). Transgenic embryos for
445 this construct had no sign of activity in the GT, limbs, or the otic vesicle [2/7],
446 promoting instead extensive reporter activation throughout the embryo (Fig. 6D,
447 supplementary Fig. 4E). This pattern contrasts with those obtained with *CR-Isl1* or
448 *CR-Isl1*$^{\Delta Lef1}$, which were consistently almost restricted to the GT and otic vesicle.
449 Together, these results are compatible with the interpretation that Tgfbr1 signaling
450 renders the *CR-Isl1* enhancer accessible to its regulation by canonical Wnt
451 signaling in the pericloacal region.
452
453 **Discussion**
454 The capacity of the pericloacal mesoderm of mouse embryos to generate both
455 external genitalia and hindlimb structures reveals a remarkable degree of
456 developmental plasticity in this tissue. This finding provides a new dimension to the
457 developmental connection between the hindlimb and the GT, previously suggested
458 by the considerable overlap in their regulatory networks [9,17–19]. It should be noted
459 that *Tgfbr1-cKO* embryos develop two independent hindlimb structures despite
460 being first induced as a single bud extending into the pericloacal mesoderm. This
461 suggests differential segregation of the tissues originating from the pericloacal
462 mesoderm and the somatopleure-derived limb field, despite both entering limb
463 differentiation paths, which might indicate that limb and pericloacal mesodermal
464 tissues contain distinct cell affinity properties that keep each primordium as discrete
465 developmental units independently of the differentiation path they enter to generate
466 mature structures. Consistent with this interpretation the PC analysis of ATAC-seq
467 profiles showed that the tissue eventually generating the extra hindlimb keeps a
468 high degree of similarity with the GT of wild type embryos (PC1 in our analysis),
469 despite the simultaneous loss of some GT signatures and the acquisition of several
470 limb-like characteristics (PC2 in our analysis). Further experiments will be required
471 to explore the validity of this hypothesis.
472 The Tgfbr1-dependent differentiation plasticity of the pericloacal mesoderm,
473 together with its apparent inability to mix with the primordial limb mesoderm has

important implications for both the development and evolution of the limb and genital area of tetrapods. It has been suggested that in mice the posterior displacement of the developing cloaca relative to the hindlimb was key to disconnect the hindlimb field from the influence of the cloacal signals that induce the GT, thus allowing the somatic lateral plate mesoderm and the ventral lateral mesoderm entering hindlimb and genital fates, respectively [20]. This contrasts with the position of those structures in squamates, in which the cloaca is formed just adjacent to the hindlimb field, a configuration that has been suggested to promote development of paired external genitalia (hemipenises) instead of hindlimbs from these primordia in snakes [20]. However, similar proximity was also observed in lizards that develop both paired external genitalia and hindlimbs [20]. In addition, and most importantly, the phenotype of the *Tgfbr1-cKO* embryos shows that the mesoderm adjacent to the cloaca can form limb structures, indicating that the intrinsic properties of the mesoderm are key determinants of the differentiation route it will enter. It will be therefore interesting to determine whether a mechanism related to the developmental plasticity uncovered by our work, including changes in cell affinity properties of prospective hindlimb and genital primordia, could help explaining the absence of hindlimbs in snakes but their presence in most lizards.

We showed that Tgfbr1 modifies the profile of chromatin accessibility in the pericloacal mesoderm. The involvement of Tgfβ/BMP signaling in chromatin remodeling has been reported before for Tgfβ and Nodal signals [70–72]. In those cases, this resulted from the interaction of Smads with chromatin remodelers like Foxh1 or components of the SWI/SNF complex, which would turn Smad-responsive enhancers accessible and enable the Smad transcriptional activity [70–72]. Although a similar mechanism could contribute to Tgfbr1 activity in the pericloacal and limb mesoderm, it is unlikely to represent a major component of this activity. First, many chromatin regions gain accessibility in the absence of *Tgfbr1*, indicating that Tgfbr1 activity can also render chromatin regions non-accessible. In addition, we did not detect enrichment of Smad binding to elements rendered accessible by *Tgfbr1* using HINT-ATAC [73] on the ATAC-seq profiles. Instead, this analysis showed some enrichment in footprints for factors known to play important roles in GT development. Even considering the limitations of the bioinformatic approach, these observations indicate that Tgfbr1 modulates the catalog of regions accessible to

507 regulatory factors, not restricted to the downstream effectors of the Tgfbr1 pathway.
508 Our work thus suggests a model according to which *Tgfbr1* controls the differential
509 morphogenesis of the limb and GT primordia by determining the set of enhancers
510 accessible or inaccessible to mediate the activity of patterning inputs. This
511 regulatory mode gains relevance when considering that many of these inputs are
512 common to both structures, despite generating mature structures as anatomically
513 different as the hindlimbs and external genital structures [9,17–19]. This idea can be
514 illustrated by two examples. One of them concerns the *Grem1* enhancer, involved
515 in *Grem1* activation in the limb bud in response to Shh from the ZPA [58]. This
516 enhancer is non accessible in the pericloacal mesoderm of wild type embryos, fitting
517 with the lack of Grem1 expression in the pericloacal mesoderm of wild type embryos
518 despite the strong Shh expression in the endodermal cloaca. It however became
519 accessible in the Tgbfr1-negative mesoderm generating extra hindlimb, and *Grem1*
520 expression was detected adjacent to the Shh source in the endoderm. The second
521 example refers to the *Isl1* GT enhancer identified in this work. While we found no
522 evidence of direct involvement of effectors of the Tgfbr1 pathway in its activation,
523 our data indicate that Tgfbr1 renders this enhancer accessible to regulation by the
524 canonical Wnt signaling, a pathway that has been shown to participate in GT
525 development [11].

526 The identification of chromatin elements uniquely associated with hindlimb
527 and GT fates might guide the search for the molecular signature that specifies
528 hindlimb or GT development. The large number of elements in each group suggest
529 that these signatures might be quite complex, likely involving the combined activity
530 of several factors. Identification of those factors will be challenging, considering that
531 in mammals enhancers are typically located far from the genes they regulate, which
532 are often not even the closest transcription unit [74].

533 An intriguing observation from our experiments was that *Tgfbr1-cKO* embryos
534 were still able to generate mature limb structures that, although malformed, still
535 contained most limb elements despite *Shh* being downregulated in the hindlimb
536 bud. This phenotype contrasts with the absence of most distal elements in *Shh*
537 mutant embryos [75]. It has been shown that a brief pulse of Shh activity during the
538 initial bud stage is enough to promote normal limb development [59]. As the ZRS
539 seems to be silent in *Tgfbr1-cKO* mutants, it is possible that the initial Shh input
540 required to set limb development in motion is provided by the endoderm, facilitated

541 by the extension of the hindlimb field into the pericloacal mesoderm. This would
542 argue in favor of limb polarizing activity intrinsic to cloacal endoderm [31], and against
543 exclusively pro-genital regulatory potential of cloaca [20].
544 A challenging question left open by our work is deciphering the mechanism(s)
545 by which Tgfbr1 activity controls such large-scale remodeling of the regulatory
546 landscape in the target tissues. The lack of enrichment in footprints for Tgfbr1
547 effectors in the target regions makes a direct activity of the canonical Tgfbr1
548 pathway in this process unlikely. Instead, it is more probable that this regulation is
549 indirect. Whatever the mechanism or mechanisms, it requires a high degree of
550 coordination, which would still be more challenging if several mechanisms are
551 involved in the generation of the different patterns (e.g. if the control of accessibility
552 and inaccessibility relies on independent mechanisms). Identification of those
553 mechanisms and determining whether they also operate in other physiological and
554 pathological processes under the control of members of the Tgfβ/BMP signaling
555 family might have far reaching implications for our understanding of morphogenetic
556 processes and disease.
557
558 **Materials and methods**
559 *Ethics statement*
560 All animal procedures were performed in accordance with Portuguese
561 (Portaria 1005/92) and European (directive 2010/63/EU) legislations and guidance
562 on animal use in bioscience research. The project was reviewed and approved by
563 the Ethics Committee of "Instituto Gulbenkian de Ciência" and by the Portuguese
564 National Entity "Direcção Geral de Alimentação Veterinária" (license reference:
565 014308).
566
567 *Mice*
568 Generation of the *Tgfbr1*^{Flox} mouse lines was previously reported [48]. *Tgfbr1*$^{3ex3-}$
569 ^{flox} resulted from the same experiments in which *Tgfbr1*^{flox} was generated, identified
570 upon sequencing through the *Tgfbr1* locus, which revealed that the exon 3 had been
571 triplicated and the presence of LoxP sites in the introns between the triplicated
572 exons and in those connecting them with exon 2 and exon 4. The *Cdx2Cre*^{ERT} mice
573 were previously reported [2]. Mice were genotyped from ear or digit biopsies.

574 Samples were incubated in 50 μL of the PBND buffer (50 mM KCl, 10 mM Tris-HCl,
575 pH 8.3, 2.5 mM $MgCl_2$, 0.1 mg/mL gelatin, 0.45% NP40, 0.45% Tween 20)
576 supplemented with 100 µg/mL of proteinase K at 55˚C overnight. Proteinase K was
577 inactivated at 95˚C for 15 minutes and 1 μL of genomic DNA was used for
578 genotyping by PCR. *Cdx2Cre^{ERT};Tgfbr1$^{+/-}$* mice were genotyped for the *Cre* allele
579 and *Tgfbr1* null allele; *Tgfbr1$^{flox/+}$* and *Tgfbr1$^{3ex3\text{-}flox/+}$* mice were genotyped with
580 primers surrounding 3'-LoxP (primers listed in supplementary table 2).
581 *Tgfbr1-cKO* (*Cdx2Cre^{ERT};Tgfbr1$^{flox/-}$* or *Cdx2Cre^{ERT};Tgfbr1$^{3ex3\text{-}flox/-}$*) embryos
582 were obtained by crossing *Tgfbr1$^{flox/flox}$* or *Tgfbr1$^{3ex3\text{-}flox/+}$* females with *Tgfbr1$^{+/-}$*
583 *;Cdx2Cre^{ERT}* males. Noon of the day of the plug was considered E0.5. To induce
584 recombination pregnant females were treated with tamoxifen (Sigma, T5648)
585 dissolved in corn oil. Tamoxifen was administered at 0.1 mg per gram of body
586 weight by oral gavage either once at E6.75 when *Tgfbr1$^{3ex3\text{-}flox}$* females were
587 involved, or twice, at E6.75 and E7.25 when *Tgfbr1^{flox}* females were involved.
588 Embryos were obtained by cesarean section and processed for further
589 analyses (see below). They were genotyped from their yolk sacs for Cre, *Tgfbr1^{-}*
590 and *Tgfbr1^{flox}* when analyzing *Tgfbr1-cKO* embryos or for β-galactosidase when
591 analyzing the reporter transgenic embryos (primers listed in supplementary table
592 2). Yolk sacs were collected in 50 μL of yolk sac lysis buffer (50 mM KCl, 10 mM
593 Tris-HCl, pH8.3, 2 mM $MgCl_2$, 0.45% Tween-20, 0.45% NP40) supplemented with
594 100 µg/mL of proteinase K and incubated at 55˚C overnight. Samples were heat-
595 deactivated as previously described and used for PCR.
596
597 *Transgenic embryos*
598 To generate transgenic constructs, the relevant enhancers (genomic
599 coordinates shown in supplementary table 3) were amplified by PCR (primers in
600 supplementary table 2) from genomic DNA and cloned into a vector containing the
601 adenovirus major late promoter, the coding region of the β-galactosidase gene and
602 the SV40 polyadenylation signal [2]. The enhancers were confirmed by direct
603 sequencing. The constructs were isolated from the plasmid backbone, gel-purified
604 using the NZYGelpure (NZYTech #MB01102) and eluted from the columns with 50
605 μL of the kit's elution buffer. The purified constructs were diluted in microinjection
606 buffer (10 mM Tris.HCl, 0.25 mM EDTA, pH 7.5) at 2 ng/μL and microinjected into

607 the pronucleus of fertilized FVB/N oocytes according to standard protocols [76].
608 Microinjected oocytes were transferred into the uteri of pseudopregnant NMRI
609 females and embryos were recovered at E10.5, E11.5 or E12.0 and stained for β-
610 galactosidase activity.
611
612 *RT-qPCR*
613 To assess recombination efficiency, the posterior part of the E10.5 *Tgfbr1-*
614 *cKO* embryos, including their hindlimbs, were dissected in PBS (137 mM NaCl, 2.7
615 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4) on ice. Whole E10.5 *Tgfbr1*$^{-/-}$
616 embryos were used as controls. RNA was extracted from fresh tissue using TRI
617 reagent (Sigma, #T9424) according to manufacturer's protocol. 200 ng of RNA from
618 each sample was used to prepare complementary DNA using a random hexamer
619 mix (NZYTech #MB12901) and following the protocol of the NZY Reverse
620 Transcriptase enzyme (NZYTech #MB12401). All cDNA samples were diluted 1:5
621 and 1 μL was used in quantitative PCR (qPCR) reactions with iTaq Universal SYBR
622 Green Supermix (Bio-Rad, #1725124), according to manufacturer's instruction and
623 using the Applied Biosystems QuantStudio 7 flex qPCR System. Primers were
624 localized in exon 3 and exon 4 of *Tgfbr1* mRNA (listed in supplementary table 2),
625 so that in case of successful recombination the product is not amplified.
626 PCR product concentration in each sample was calculated using the standard
627 curve with following formula *Quantity = 10^(mean Ct-b/m)* (where "m" is slope and
628 "b" is an intercept of the standard curve), and *Tgfbr1* expression level was
629 normalized to *β-Actin* expression. Linear model was fitted (Tgfbr1 normalized
630 expression ~ Genotype) with R built-in function (lm). Analysis of variance (ANOVA)
631 was performed on the linear model with the built-in function in R (anova).
632 Significance levels between groups of samples were assessed by Tukey method
633 using the "glht" function from "multicomp" package in R. Differences were
634 considered significant at *p<0.05, **p<0.01 and ***p<0.001. Boxplots were
635 generated with "ggplot2" and "ggdignif" packages in R.
636
637 *Cell culture and transfection*
638 HEK 293T cells (ATCC #CRL-3216) were cultured in Dulbecco's modified
639 Eagle's medium (DMEM, Life Technologies) supplemented with 10% fetal bovine

640 serum (FBS, Life Technologies) and 1% penicillin/streptomycin at 37 °C in a 5%
641 CO_2 atmosphere. Cells were seeded in 35 mm plates and when they reached 70%
642 confluency, the medium was changed to transfection medium (DMEM
643 supplemented with 10% FBS). Expression vectors were constructed by amplifying
644 the coding region of the *Tgfbr1* transcript isolated from wild type or *Tgfbr1*$^{3ex3\text{-}flox}$
645 homozygous embryos by RT-PCR (primers in supplementary table 2) and cloned
646 into the pRK5 expression vector. Constructs were verified by sequencing. 3 μg of
647 the plasmids was transfected into the 393T cells using Lipofectamine 2000 (Thermo
648 Fisher Scientific #11668027) according to the manufacturer's instructions. After 24
649 hours incubation, cells were processed for Western blot analysis.
650
651 *Protein extraction and western blot*
652 The transfected cells were washed twice with ice-cold PBS and then 150 μL
653 of lysis buffer (50 mM Tris-HCl, 150 mM NaCl and 1% Triton X-100, pH8.0) was
654 added to each plate and incubated on ice for 10 minutes. Lysates were scraped into
655 microcentrifuge tubes on ice and centrifuged at 20000 rcf for 10 min at 4˚C. The
656 supernatant was collected, frozen on dry ice and stored at -80˚C. Protein lysates
657 were mixed (2:1) with 3X loading buffer (549 mM Tris HCl, 563 mM Tris Base, 2.05
658 mM EDTA, 8 % LDS (Lithium dodecyl sulfate), 10 % Glycerol, 0.75 % SERVA Blue
659 G250, 0.25% Phenol red, 160 mM DTT). Samples were incubated for 15 minutes
660 at 65˚C and resolved by SDS-PAGE in a 10% polyacrylamide gel. Proteins were
661 transferred to PVDF membranes in 20% methanol, 25 mM Tris, 200 mM glycine at
662 200 mA for 1 hour. The membranes were blocked in blocking solution [5% dry milk
663 dissolved in PBS containing 0.1% Tween-20 (PBT)] for 1 hour at room temperature
664 and then incubated with primary antibodies overnight at 4˚C. Primary antibodies
665 were anti-Tgfbr1 (Sigma-Aldrich #HPA056473, 1:1000 in blocking solution) and
666 anti-actin (Abcam #ab179467, 1:1000 in blocking solution). Membranes were then
667 washed in PBT and incubated with HRP conjugated-anti-rabbit IgG (GE Healthcare
668 # NA9340, 1:5000 in blocking solution) at room temperature for 1 hour and washed
669 twice in PBT. Signals were developed by chemiluminescence using Amersham
670 ECL Prime Western Blotting Detection Reagent (GE Healthcare #RPN2232) and
671 images were captured using the GE Amersham Imager 600.
672

673 *Skeletal preparation*

674 E17.5 fetuses were recovered from pregnant females by cesarean section and
675 dissected from the extraembryonic membranes, eviscerated, the skin removed and
676 then fixed in 100% ethanol for 2 days. Cartilages are then stained by incubation
677 with 450 mg/L of alcian blue (Sigma, #A5268) in 80% ethanol/20% acetic acid for
678 one day. Fetuses were then postfixed in 100% ethanol overnight. Tissue was
679 cleared by incubation in 2% KOH for 6 hours, followed by staining of ossified bones
680 with a 50 mg/l of alizarin red S (Sigma, #130-22-3) solution in 2% KOH for 3 hours.
681 Specimens were then further incubated in 2% KOH until tissues were fully cleared.
682 Skeletons were stored in 25% glycerol in water. All incubations were performed at
683 room temperature with rolling. Genotyping was performed on gut tissue by
684 incubating in Laird's buffer (100 mM Tris-HCl pH 8.5, 5 mM EDTA, 0.2% SDS, 200
685 mM NaCl) containing 10 μg/mL proteinase K at 55ºC overnight. Genomic DNA was
686 then recovered by precipitation with isopropanol (1:1, vol:vol) and transferred to TE
687 pH 8.0. Genotyping was then performed by PCR using oligos specified in
688 supplementary table 2.

689

690 *OPT and SPIM imaging*

691 Optical projection tomography (OPT) and SPIM (Selective Illumination Plane
692 Microscopy) were used to image E16.5 fetuses as previously described [77] with
693 minor modifications. Briefly, fetuses were recovered by cesarean section in ice cold
694 PBS, washed several times in PBS and fixed in 4% paraformaldehyde (PFA) in PBS
695 at 4ºC for several days. After several washes with demineralized water fetuses were
696 dehydrated by sequential incubation in demineralized water with increasing
697 concentrations of methanol (10% increases) and then twice in 100% methanol.
698 Fetuses were then bleached in a three-day process with a sequence of 2.5%, 5%
699 and 10% hydrogen peroxide in methanol at room temperature. Fetuses were then
700 rehydrated through a reverse methanol/demineralized water series and embedded
701 in a 0,7% low-melting agarose. Clearing was done with a 1:2 solution of Benzoic
702 Alcohol:Benzyl Benzoate (BABB) using a BABB/methanol series (25% BABB
703 increases). 100% BABB was introduced on day 3 and replaced every day for the
704 next 4 days, until the fetuses became completely transparent. Anatomical datasets
705 were then obtained using a custom built OPT/SPIM scanner and procedures;

706 briefly, for OPT green auto-fluorescence was acquired on 1600 sequential angles
707 for a full revolution, the raw dataset was pre-processed with a custom-built ImageJ
708 macro, and then back-projection reconstructed using SkyScan's "nrecon" as in [78],
709 and then post-processed to reduce noise and enhance contrast in ImageJ. 3D
710 reconstruction and visualization of internal organs was performed using the Amira
711 software (Thermo Fisher Scientific). The autopods were also imaged with SPIM at
712 higher magnification to be able to reconstruct the digits and bones, also using
713 Amira. To facilitate the segmentation of the bones we used the Biomedisa online
714 tool for machine-learning-assisted interpolation after manual segmentation of a few
715 sparse sections [79]. The correlated OPT and SPIM datasets, and their segmented
716 volumes, were then assembled into a correlative dataset in Amira, to produce the
717 final 3D reconstructions.
718
719 *Histological analysis*
720 E16.5 fetuses were fixed in Bouin's fixative at room temperature for 2 days,
721 dehydrated by repeat washes in 100% ethanol for 5 days and embedded in paraffin.
722 Paraffin blocks were then sectioned at 10 µm and stained with Masson's trichrome
723 according to standard methods.
724
725 *Whole mount in situ hybridization and sectioning*
726 Embryos were fixed in 4% PFA overnight, then dehydrated through 25%, 50%
727 and 75% series of methanol in PBT (PBS, 0,1% Tween 20), then incubated in 100%
728 methanol. Embryos were then rehydrated through a reverse methanol/PBT series
729 and incubated three times in PBT for at least 5 min. each at room temperature.
730 Embryos were then bleached for 1 hour in 6% hydrogen peroxide in PBT and
731 permeabilized in 10 µg/mL of proteinase K in PBT for a period that depended on
732 the embryo size. The reaction was then quenched with a 2 mg/mL solution of
733 glycine in PBT, washed twice in PBT and postfixed in a 4% PFA and 0,2%
734 glutaraldehyde mix for 20 minutes, followed by two washes in PBT. Hybridization
735 was performed at 65˚C overnight in hybridization solution (50% formamide, 1.3 x
736 SSC pH 5.5 [20 x SSC is 3M NaCl, 300 mM sodium citrate], 5 mM EDTA, 0.2 %
737 Tween 20, 50 µg/mL yeast tRNA, 100 µg/mL heparin) containing the relevant
738 digoxigenin-labelled antisense RNA probes. The probes used in this work are listed

739 in supplementary table 2. RNA probes were transcribed in vitro from the vector for
740 1.5 hours at 37˚C with relevant RNA polymerase and the DIG RNA Labeling Mix
741 (Roche #11277073910). The reaction product was verified in 0,8% agarose gel,
742 precipitated with ethanol in the presence of 0.3M sodium acetate, pH 5.3 and
743 resuspended in 100 μL of TE pH 8.0. Probes were diluted in hybridization solution
744 (6 μL in 1 ml of solution). After hybridization, embryos were washed twice with
745 hybridization solution without tRNA and heparin, at 65˚C for 30 min each, then in a
746 1:1 mix of hybridization solution and TBST (25 mM Tris.HCl, pH 8.0, 140 mM NaCl,
747 2.7 mM KCl, 0.1% Tween 20) for 30 minutes at 65ºC, and finally washed three times
748 with TBST at room temperature. Embryos were then equilibrated in MABT (100 mM
749 Maleic acid, 150 mM NaCl, 0.1 % Tween-20, pH 7.5) and blocked in MABT blocking
750 buffer [MABT containing 1% blocking reagent (Roche #11096176001)] with 10%
751 sheep serum for 2,5 hours at room temperature. Embryos were then incubated
752 overnight at 4ºC with a 1:2000 dilution of alkaline phosphatase-conjugated anti-
753 digoxigenin antibody (Roche #11093274910) in MABT blocking buffer with 1%
754 sheep serum. After extensive washes with MABT at room temperature, embryos
755 were equilibrated in NTMT buffer (100 mM Tris HCl, pH 9.5, 50 mM $MgCl_2$, 100 mM
756 NaCl, 0.1% Tween-20) and developed with a 1:50 dilution of NBT/BCIP solution
757 (Roche #11681451001) in NTMT at room temperature in the dark. The reaction was
758 stopped with PBT, and embryos postfixed with 4% PFA overnight at room
759 temperature and stored in PBT. Pictures were taken with a SWIFTCAM SC1803
760 connected to a stereoscope. Stained embryos were embedded in a mix containing
761 0,45% gelatin, 27% bovine serum albumin (BSA), 18% sucrose, jellified with 1,75%
762 glutaraldehyde and sectioned at 35 μm on a Leica Vibratome VT 1000 S.
763
764 *β-galactosidase staining*
765 Embryos were dissected out in ice cold PBS and fixed in 4% PFA at 4ºC for
766 30 minutes. They were then washed three times in β-gal wash buffer (PBS plus
767 0.02% Tween 20) for 10 minutes at room temperature and β-galactosidase activity
768 developed in PBS containing 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$,
769 0.02% Tween 20, 0.4 mg/ml X-gal (Promega #V3941) overnight in the dark at 37ºC.
770 The reaction was stopped with β-gal wash buffer, embryos postfixed in 4% PFA
771 overnight at room temperature and stored in PBS.

772

773 *ATAC-Seq*

774 ATAC-seq was performed as in [55] with minor modifications. E11.25 mouse
775 embryos were dissected on ice in DMEM High Glucose medium (Biowest #L0102-
776 500) containing 10% FBS (this will be referred to as medium in the rest of the
777 protocol). Hindlimbs and genital tubercles were dissected out and collected on ice-
778 cold media. Single cell suspension was prepared by treating the tissue with 500 μL
779 of Accutase (Sigma #A6964-500ML) for 30 minutes at 37˚C with shaking at 600
780 rpm. Single cells were pelleted at 6000 rcf for 5 minutes at 4˚C, resuspended in 200
781 μL of media and counted. 50000 viable cells from each sample were used for nuclei
782 extraction. Cells were incubated with 50 μL of ATAC resuspension buffer (10 mM
783 Tris HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 0.1% Tween-20 (Sigma-
784 Aldrich, P7949), 0.01% Digitonin (Target Mol #282T2721-1ml/10mM in DMSO) for
785 3 minutes on ice. Lysis was quenched by adding 1 ml of ATAC resuspension buffer
786 without NP-40 and Digitonin. Nuclei were pelleted by centrifugation at 500 rcf for 10
787 minutes at 4˚C. The pellet was then resuspended in 50 μL of transposition buffer
788 [for 50 μL: 25 μL 2x TD buffer, 2.5 μL Tn5 transposition enzyme (100 nM final)
789 (Illumina #15028212), 16.5 μL PBS, 0.5 μL 1% digitonin, 0.5 μL 10% Tween-20, 5
790 μL H_2O] and incubated at 37˚C for 30 minutes with shaking at 1000 rpm. The DNA
791 was purified with Qiagen MinElute PCR purification kit (Qiagen #50928004) and
792 eluted in 20 μL of the kit's elution buffer. Libraries were amplified by PCR with
793 NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs #174M0541S)
794 for 9 cycles in 96-well Thermal cycler (Biorad) and purified with Qiagen MinElute
795 PCR purification kit. Tagmentation efficiency was assessed on TapeStation 4200
796 (Agilent). Double size selection to remove primer dimers and fragments exceeding
797 1 kb was performed using SPRIselect beads (Beckman Coulter). Another quality
798 control was performed with High Sensitivity DNA assay using the Fragment
799 Analyzer (Agilent). The 4nM libraries pool was sequenced with Illumina NextSeq
800 2000 (100 cycles, Pair-end 50 bp). Two biological replicates were performed per
801 tissue.

802 The data obtained in this work has been deposited in the GEO repository,
803 (accession number GSE231592).

804

805 *ATAC-Seq bioinformatic and statistical analysis*

806 Data bioinformatic analysis was performed on the Galaxy server [80]. Raw
807 sequencing fastq files for each library were assessed for quality, adapter content
808 and duplication rates with FastQC (v0.11.9)
809 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Adapters were
810 trimmed and reads with length < 20 bp were removed using cutadapt (v1.16.5)[81]
811 (3´ adapter sequence: CTGTCTCTTATACACATCT). The trimmed reads were
812 aligned to the mouse reference genome (GRCm38/mm10 Dec. 2011) using
813 Bowtie2 (v2.4.2)[82] [parameters paired-end options (-X/–maxins 1000, --fr, --
814 dovetail), --very-sensitive]. The aligned reads were filtered using BamTools Filter
815 (v2.4.1) [83] (parameters --isProperPair true, --mapQuality ≥ 30, and --
816 reference!=chrM) and the duplicate reads were removed using Picard
817 MarkDuplicates (v2.18.2.2) (http://broadinstitute.github.io/picard/). The resulting
818 BAM files were converted to BED and the reads that overlap to the blacklisted
819 regions listed in https://www.encodeproject.org/files/ENCFF547MET/ were
820 removed using Bedtools (v2.30.0) [84]. The filtered BED files were used as inputs for
821 peak calling. Peaks were called from individual replicates using MACS2 callpeak
822 (v2.1.1) [85] (parameters --format single-end BED, --nomodel, --extsize 200, --shift -
823 100, --qvalue 0.05).

824 BAM files were then converted to BigWig with "deeptools" [86] normalizing to 1x
825 effective genome size (2652783500 for GRCm38/mm10). BigWig files were used
826 for data visualization in IGV with the addition of the conservation scoring by phyloP
827 (phylogenetic p-values)[87] for 60 vertebrate genomes from the UCSC genome
828 browser (mm10.60way.phyloP60way.bw file downloaded from
829 http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phyloP60way/) and ChIP-seq
830 data from Hoxa13 (GSE81356) [56] and Gli3 (GSE133710) [57].

831 The raw count matrix was created by merging peaks across samples as in
832 [88,89]. DESeq2 object was built from the raw count matrix and blind dispersion
833 estimated with the "vst" function from the "DESeq2" package in R [90]. Top 2500 most
834 variable peaks estimated with the "rowVars" function in R were selected for the
835 principal component analysis (PCA). PCA was performed using the built-in
836 "prcomp" function in R. 500 top loadings contributing to PC1 and PC2 were
837 annotated using ChIPseeker package in R [91]. Hierarchical clustering was performed
838 on the Z-score normalized reads. Correlation distances between the peaks were

computed with “as.dist” function and clustering was performed using complete linkage method with “hclust” function. Clusters were defined using a threshold h = 0.6. For statistical analysis raw reads were analyzed using EdgeR [92]. ANOVA-like test was performed on all samples and pairwise comparisons between the samples were made by analyzing individual contrasts. Venn diagrams were created from the lists of selected differentially accessible peaks with “eulerr” package in R. Heat Maps were made from data frames containing log2 normalized counts across samples with the “pheatmap” package (https://CRAN.R-project.org/package=pheatmap).

Footprinting analysis of ATAC-Seq data

We further used the program RGT HINT-ATAC [73] and motif matching and enrichment to identify, plot 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). The BAM and BED files for the peaks called in the control GT and the 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[93] for finding the motif-predicted binding sites (mpbs) independently. Motif enrichment analysis was performed on the footprinting results for the peaks of pattern 1, using as a background the footprinting results for all the peaks called in control GT. Position frequency matrices for Lef1 (MA0768.2) and Gata3 (MA0037.4) binding motifs in *Mus Musculus* were obtained from Jaspar database.

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, histopathology facility for their help with histology procedures, 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.

References

1. Duarte, P., Brattig Correia, R., Nóvoa, A. & Mallo, M. Regulatory changes associated with the head to trunk developmental transition. *BMC Biol* **21**, 170 (2023).
2. 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).
3. Binagui-Casas, A., Dias, A., Guillot, C., Metzis, V. & Saunders, D. Building consensus in neuromesodermal research: Current advances and future biomedical perspectives. *Curr Opin Cell Biol* **73**, 133–140 (2021).
4. Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J. F. Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis. *Dev Cell* **17**, 365–376 (2009).
5. Wymeersch, F. J., Wilson, V. & Tsakiridis, A. Understanding axial progenitor biology in vivo and in vitro. *Development* **148**, dev180612 (2021).
6. Wilson, V., Olivera-Martinez, I. & Storey, K. G. Stem cells, signals and vertebrate body axis extension. *Development* **136**, 2133–2133 (2009).
7. Prummel, K. D., Nieuwenhuize, S. & Mosimann, C. The lateral plate mesoderm. *Development (Cambridge)* **147**, dev175059 (2020).
8. 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).
9. Yamada, G. *et al.* Molecular genetic cascades for external genitalia formation: An emerging organogenesis program. *Developmental Dynamics* **235**, 1738–1752 (2006).
10. Miyagawa, S. *et al.* Disruption of the temporally regulated cloaca endodermal β-catenin signaling causes anorectal malformations. *Cell Death Differ* **21**, 990–997 (2014).

11. 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).
12. Huang, Y. C., Chen, F. & Li, X. Clarification of mammalian cloacal morphogenesis using high-resolution episcopic microscopy. *Dev Biol* **409**, 106–113 (2016).
13. Tickle, C. How the embryo makes a limb: Determination, polarity and identity. *J Anat* **227**, 418–430 (2015).
14. Rallis, C. *et al.* Tbx5 is required for forelimb bud formation and continued outgrowth. *Development* **130**, 2741–2751 (2003).
15. Naiche, L. a & Papaioannou, V. E. Tbx4 is not required for hindlimb identity or post-bud hindlimb outgrowth. *Development* **134**, 93–103 (2007).
16. Naiche, L. A. & Papaioannou, V. E. Loss of Tbx4 blocks hindlimb development and affects vascularization and fusion of the allantois. *Development* **130**, 2681–2693 (2003).
17. Kondo, T., Zakany, J., Innis, J. W. & Duboule, D. Of fingers, toes and penises [1]. *Nature* **390**, 29 (1997).
18. Cohn, M. J. Development of the external genitalia: Conserved and divergent mechanisms of appendage patterning. *Developmental Dynamics* **240**, 1108–1115 (2011).
19. Lin, C. *et al.* Delineating a Conserved Genetic Cassette Promoting Outgrowth of Body Appendages. *PLoS Genet* **9**, e1003231 (2013).
20. Tschopp, P. *et al.* A relative shift in cloacal location repositions external genitalia in amniote evolution. *Nature* **516**, 391–394 (2014).
21. Lozovska, A. *et al.* Tgfbr1 regulates lateral plate mesoderm and endoderm reorganization during the trunk to tail transition. *BioRxiv* (2023) doi:10.1101/2023.08.22.554351.
22. Zakin, L., Reversade, B., Kuroda, H., Lyons, K. & De Robertis, E. Sirenomelia in Bmp7 and Tsg compound mutant mice: requirement for Bmp signaling in the development of ventral posterior mesoderm. *Development* **132**, 2489–2499 (2005).
23. Suzuki, K. *et al.* Reduced BMP Signaling Results in Hindlimb Fusion with Lethal Pelvic/Urogenital Organ Aplasia: A New Mouse Model of Sirenomelia. *PLoS One* **7**, e43453 (2012).
24. Suzuki, K. *et al.* Regulation of outgrowth and apoptosis for the terminal appendage: External genitalia: Development by concerted actions of BMP signaling. *Development* **130**, 6209–6220 (2003).
25. Kajioka, D. *et al.* Bmp4 is an essential growth factor for the initiation of genital tubercle (GT) outgrowth. *Congenit Anom (Kyoto)* **60**, 15–21 (2020).
26. 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).
27. Haraguchi, R. *et al.* Molecular analysis of external genitalia formation: The role of fibroblast growth factor (fgf) genes during genital tubercle formation. *Development* **127**, 2471–2479 (2000).
28. Seifert, A. W., Yamaguchi, T. & Cohn, M. J. Functional and phylogenetic analysis shows that Fgf8 is a marker of genital induction in mammals but is not required for external genital development. *Development* **136**, 2643–2651 (2009).

29. Seifert, A. W., Bouldin, C. M., Choi, K. S., Harfe, B. D. & Cohn, M. J. Multiphasic and tissue-specific roles of sonic hedgehog in cloacal septation and external genitalia development. *Development* **136**, 3949–3957 (2009).
30. Haraguchi, R. *et al.* Unique functions of sonic hedgehog signaling during external genitalia development. *Development* **128**, 4241–4250 (2001).
31. 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).
32. 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).
33. Warot, X., Fromental-Ramain, C., Fraulob, V., Chambon, P. & Dollé, P. Gene dosage-dependent effects of the Hoxa-13 and Hoxd-13 mutations on morphogenesis of the terminal parts of the digestive and urogenital tracts. *Development* **124**, 4781–4791 (1997).
34. 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).
35. Su, T. *et al.* LIM homeodomain transcription factor Isl1 affects urethral epithelium differentiation and apoptosis via Shh. *Cell Death Dis* **10**, 713 (2019).
36. Tickle, C. & Towers, M. Sonic hedgehog signaling in limb development. *Front Cell Dev Biol* **5**, 14 (2017).
37. Dias, A. *et al.* A Tgfbr1/Snai1-dependent developmental module at the core of vertebrate axial elongation. *Elife* **9**, e56615 (2020).
38. 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).
39. 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).
40. Matsubara, Y. *et al.* Anatomical integration of the sacral-hindlimb unit coordinated by GDF11 underlies variation in hindlimb positioning in tetrapods. *Nat Ecol Evol* **1**, 1392–1399 (2017).
41. Aires, R. *et al.* Oct4 Is a Key Regulator of Vertebrate Trunk Length Diversity. *Dev Cell* **38**, 262–274 (2016).
42. McPherron, A. C., Huynh, T. V & Lee, S.-J. Redundancy of myostatin and growth/differentiation factor 11 function. *BMC Dev Biol* **9**, 24 (2009).
43. Heldin, C.-H. & Moustakas, A. Signaling receptors for TGF-β family members. *Cold Spring Harb Perspect Biol* **8**, 1–33 (2016).
44. Goumans, M.-J. *et al.* Activin Receptor-like Kinase (ALK)1 Is an Antagonistic Mediator of Lateral TGF/ALK5 Signaling. *Mol Cell* **12**, 817–828 (2003).
45. Carvalho, R. L. C. *et al.* Compensatory signalling induced in the yolk sac vasculature by deletion of TGFβ receptors in mice. *J Cell Sci* **120**, 4269–4277 (2007).
46. Ramachandran, A. *et al.* TGF-β uses a novel mode of receptor activation to phosphorylate SMAD1/5 and induce epithelial-to-mesenchymal transition. *Elife* **7**, e31756 (2018).

1006 47. Wang, W. *et al.* The TGF β type I receptor TGF β RI functions as an inhibitor of BMP
1007 signaling in cartilage. *Proceedings of the National Academy of Sciences USA* **116**,
1008 15570–15579 (2019).
1009 48. Pineault, K. M., Novoa, A., Lozovska, A., Wellik, D. M. & Mallo, M. Two
1010 CRISPR/Cas9-mediated methods for targeting complex insertions, deletions, or
1011 replacements in mouse. *MethodsX* **6**, 2088–2100 (2019).
1012 49. Wieser, R., Wrana, J. L. & Massague, J. GS domain mutations that constitutively
1013 activate T beta R-I, the downstream signaling component in the TGF-beta receptor
1014 complex. *Embo J* **14**, 2199–2208 (1995).
1015 50. Dünker, N. & Krieglstein, K. Tgfβ2-/-Tgfβ3-/- double knockout mice display severe
1016 midline fusion defects and early embryonic lethality. *Anat Embryol (Berl)* **206**, 73–
1017 83 (2002).
1018 51. Yang, L. *et al.* Isl1 Cre reveals a common Bmp pathway in heart and limb
1019 development. *Development* **133**, 1575–1585 (2006).
1020 52. Min, H. *et al.* Fgf-10 is required for both limb and lung development and exhibits
1021 striking functional similarity to Drosophila branchless. *Genes Dev* **12**, 3156–3161
1022 (1998).
1023 53. Davis, C. A., Holmyard, D. P., Millen, K. J. & Joyner, A. L. Examining pattern
1024 formation in mouse, chicken and frog embryos with an En-specific antiserum.
1025 *Development* **111**, 287–298 (1991).
1026 54. Riddle, R. D. *et al.* Induction of the LIM Homeobox Gene Lmx7 by WNT7a
1027 Establishes Dorsoventral Pattern in the Vertebrate Limb. *Cell* **83**, 631–640 (1995).
1028 55. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and
1029 enables interrogation of frozen tissues. *Nat Methods* **14**, 959–962 (2017).
1030 56. Sheth, R. *et al.* Distal Limb Patterning Requires Modulation of cis-Regulatory
1031 Activities by HOX13. *Cell Rep* **17**, 2913–2926 (2016).
1032 57. Lex, R. K. *et al.* GLI transcriptional repression regulates tissue-specific enhancer
1033 activity in response to hedgehog signaling. *Elife* **9**, e50670 (2020).
1034 58. Malkmus, J. *et al.* Spatial regulation by multiple Gremlin1 enhancers provides digit
1035 development with cis-regulatory robustness and evolutionary plasticity. *Nat*
1036 *Commun* **12**, 5557 (2021).
1037 59. Zhu, J., Patel, R., Trofka, A., Harfe, B. D. & Mackem, S. Sonic hedgehog is not a limb
1038 morphogen but acts as a trigger to specify all digits in mice. *Dev Cell* **57**, 2048–2062
1039 (2022).
1040 60. Michos, O. *et al.* Gremlin-mediated BMP antagonism induces the epithelial-
1041 mesenchymal feedback signaling controlling metanephric kidney and limb
1042 organogenesis. *Development* **131**, 3401–3410 (2004).
1043 61. Nissim, S., Hasso, S. M., Fallon, J. F. & Tabin, C. J. Regulation of Gremlin expression
1044 in the posterior limb bud. *Dev Biol* **299**, 12–21 (2006).
1045 62. Farin, H. F. *et al.* Tbx2 Terminates Shh/Fgf Signaling in the Developing Mouse Limb
1046 Bud by Direct Repression of Gremlin1. *PLoS Genet* **9**, e1003467 (2013).
1047 63. Zuniga, A. *et al.* Mouse limb deformity mutations disrupt a global control region
1048 within the large regulatory landscape required for Gremlin expression. *Genes Dev*
1049 **18**, 1553–1564 (2004).
1050 64. Li, Q. *et al.* A Gli silencer is required for robust repression of gremlin in the
1051 vertebrate limb bud. *Development (Cambridge)* **141**, 1906–1914 (2014).

65. 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).
66. 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).
67. 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).
68. Sagai, T. *et al.* A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development* **136**, 1665–1674 (2009).
69. Jia, S. & Meng, A. TGF β family signaling and development. *Development* **148**, dev188490 (2021).
70. 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).
71. Coda, D. M. *et al.* A network of transcription factors governs the dynamics of NODAL/Activin transcriptional responses. *J Cell Sci* **135**, jjcs259972 (2022).
72. Xi, Q., He, W., Zhang, X. H. F., Le, H. Van & Massague, J. Genome-wide impact of the BRG1 SWI/SNF chromatin remodeler on the transforming growth factor β transcriptional program. *Journal of Biological Chemistry* **283**, 1146–1155 (2008).
73. Li, Z. *et al.* Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* **20**, 45 (2019).
74. Liang, L. *et al.* Complementary Alu sequences mediate enhancer–promoter selectivity. *Nature* **619**, 868–875 (2023).
75. Chiang, C. *et al.* Manifestation of the limb prepattern: Limb development in the absence of sonic hedgehog function. *Dev Biol* **236**, 421–435 (2001).
76. Hogan, B., Beddington, R., Constantini, F. & Lacy, E. *Manipulating the mouse embryo: a laboratory manual*. (Cold Spring Harbour Laboratory Press, 1994).
77. 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).
78. 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.
79. Lösel, P. D. *et al.* Introducing Biomedisa as an open-source online platform for biomedical image segmentation. *Nat Commun* **11**, 5577 (2020).
80. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res* **44**, W3–W10 (2016).
81. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* **17**, 10–12 (2011).
82. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
83. 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).

84. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
85. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728–1740 (2012).
86. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
87. 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).
88. 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).
89. Erarslan-Uysal, B. *et al.* Chromatin accessibility landscape of pediatric T-lymphoblastic leukemia and human T-cell precursors. *EMBO Mol Med* **12**, e12104 (2020).
90. 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).
91. Wang, Q. *et al.* Exploring Epigenomic Datasets by ChIPseeker. *Curr Protoc* **2**, e585 (2022).
92. 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).
93. 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).
94. 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).
95. 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).
96. Srivastava, D. *et al.* Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* **16**, 154–160 (1997).
97. Aires, R. *et al.* Tail Bud Progenitor Activity Relies on a Network Comprising Gdf11, Lin28, and Hox13 Genes. *Dev Cell* **48**, 383–395 (2019).
98. 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).
99. 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).
100. Chapman, D. L. *et al.* Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Developmental Dynamics* **206**, 379–390 (1996).

Figures

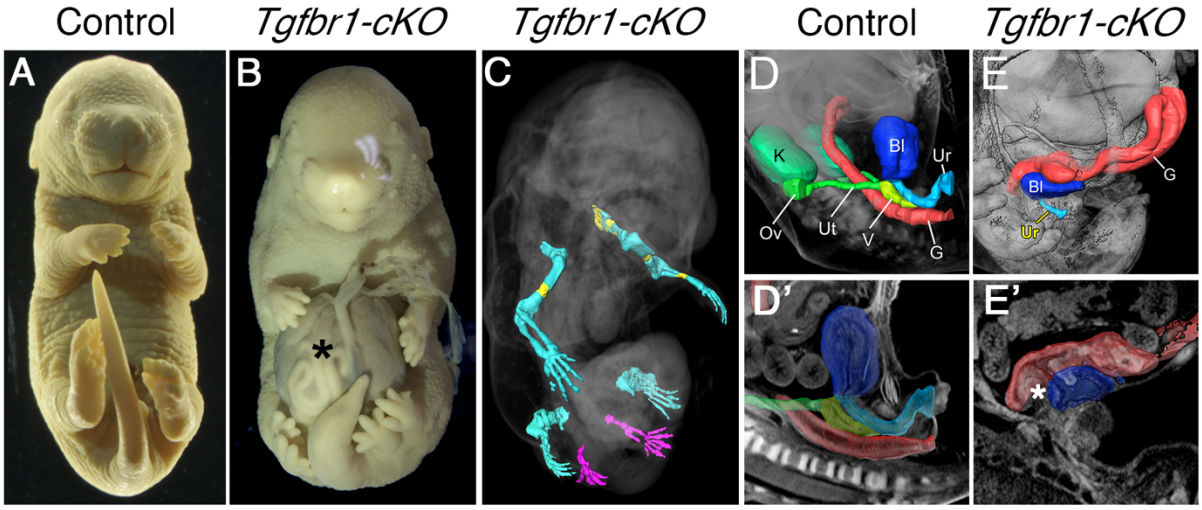


Figure 1. Malformations in the E16.5 *Tgfbr1-cKO* fetuses. (A, B) Images of fixed control (A) and *Tgfbr1-cKO* (B) fetuses. The mutant fetus shows the presence of an omphalocele (asterisk) and hindlimb duplication. (C) 3D reconstruction of the limb skeleton of a *Tgfbr1-cKO* fetus obtained by OPT and 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.

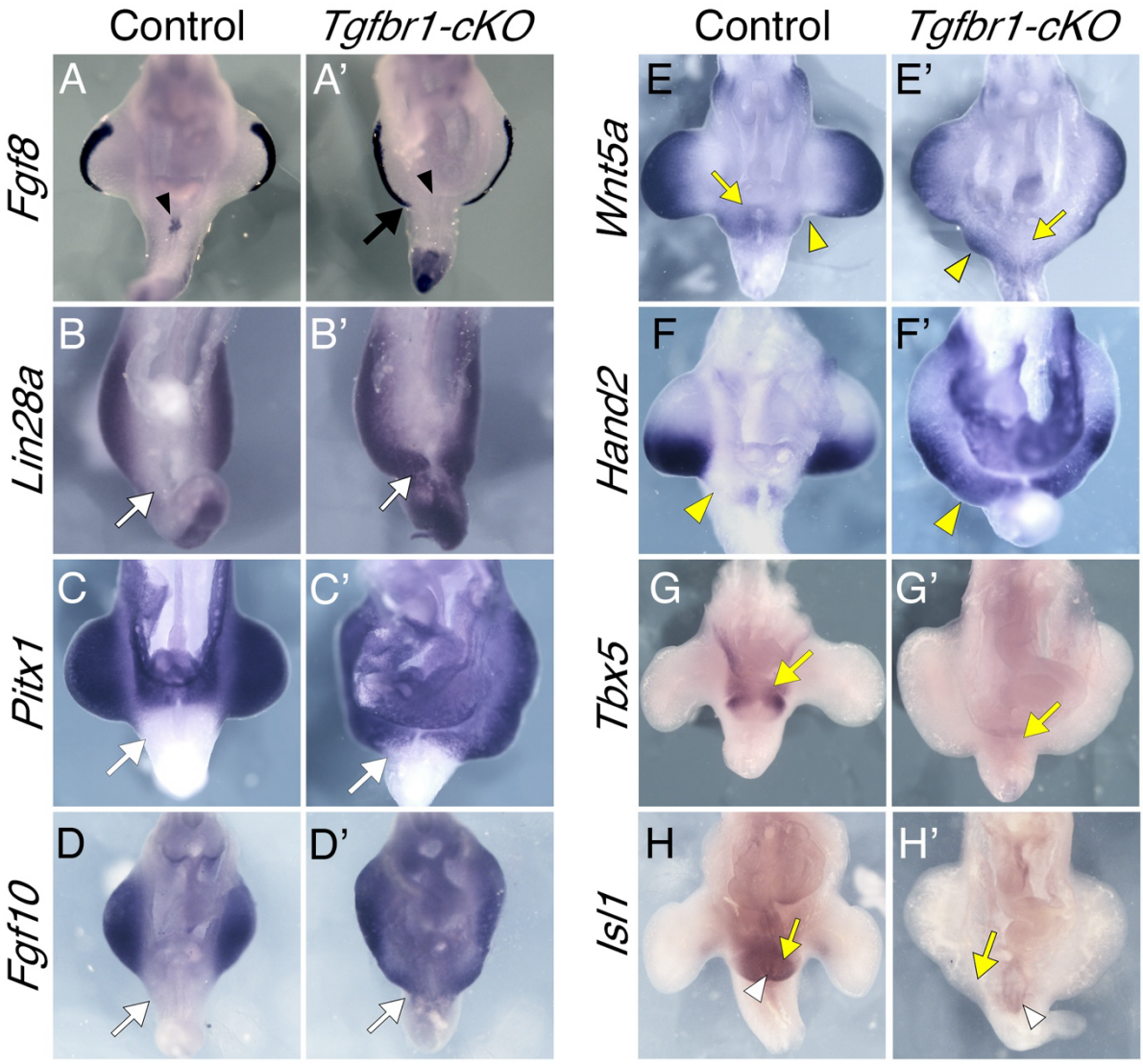


Figure 2. The pericloacal mesenchyme of the *Tgfbr1-cKO* embryos adopts hindlimb fate. Ventral views of the hindlimb/cloacal region of wild type (A-H) or *Tgfbr1-cKO* embryos (A'-H') stained for *Fgf8* (E10.5)(A, A'), *Lin28a* (E10.5) (B, B'), *Pitx1* (E10.5) (n=2; C, C'), *Fgf10* (E10.5) (D, D'), *Wnt5a* (E11.0) (E, E'), *Hand2* (E11.0) (F, F'), *Tbx5* (E11.5) (G, G') or *Isl1* (E11.5) (H, H'). *Fgf8* expression shows posterior and medial extension 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 *Tgfbr1-cKO* embryos (white arrows). *Wnt5a* and *Hand2* are expressed in the posterior prominence of the *Tgfbr1-cKO* hindlimbs, while forming clear separation of the hindlimb and GT domains in the control embryos (yellow arrowheads). Pericloacal expression of *Wnt5a*, *Tbx5* and *Isl1* is lost in the mutant embryos (yellow arrows). *Isl1* expression in cloacal endoderm is marked by white arrowhead. At least n=3 embryos were analyzed per probe, unless stated otherwise.

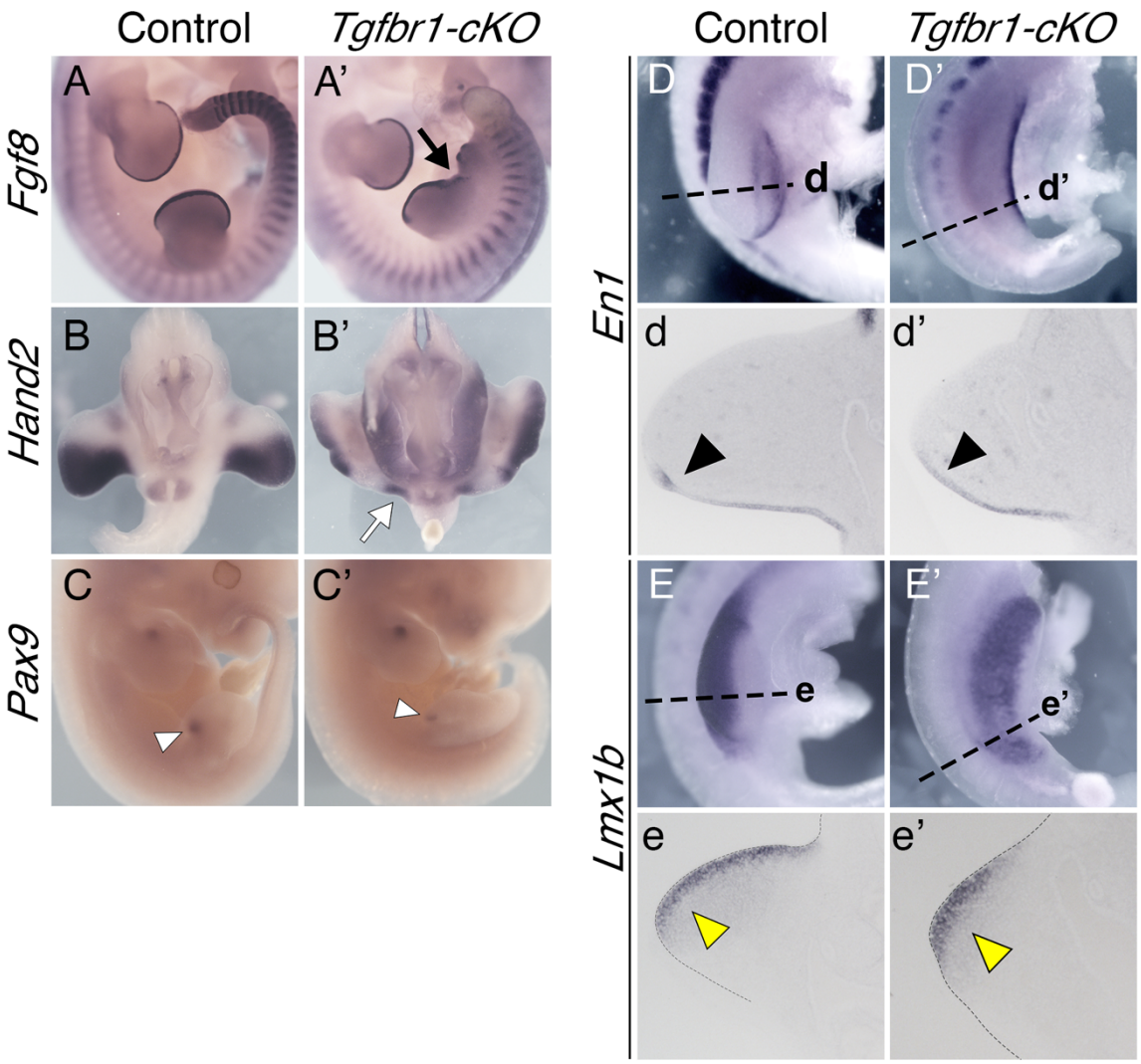
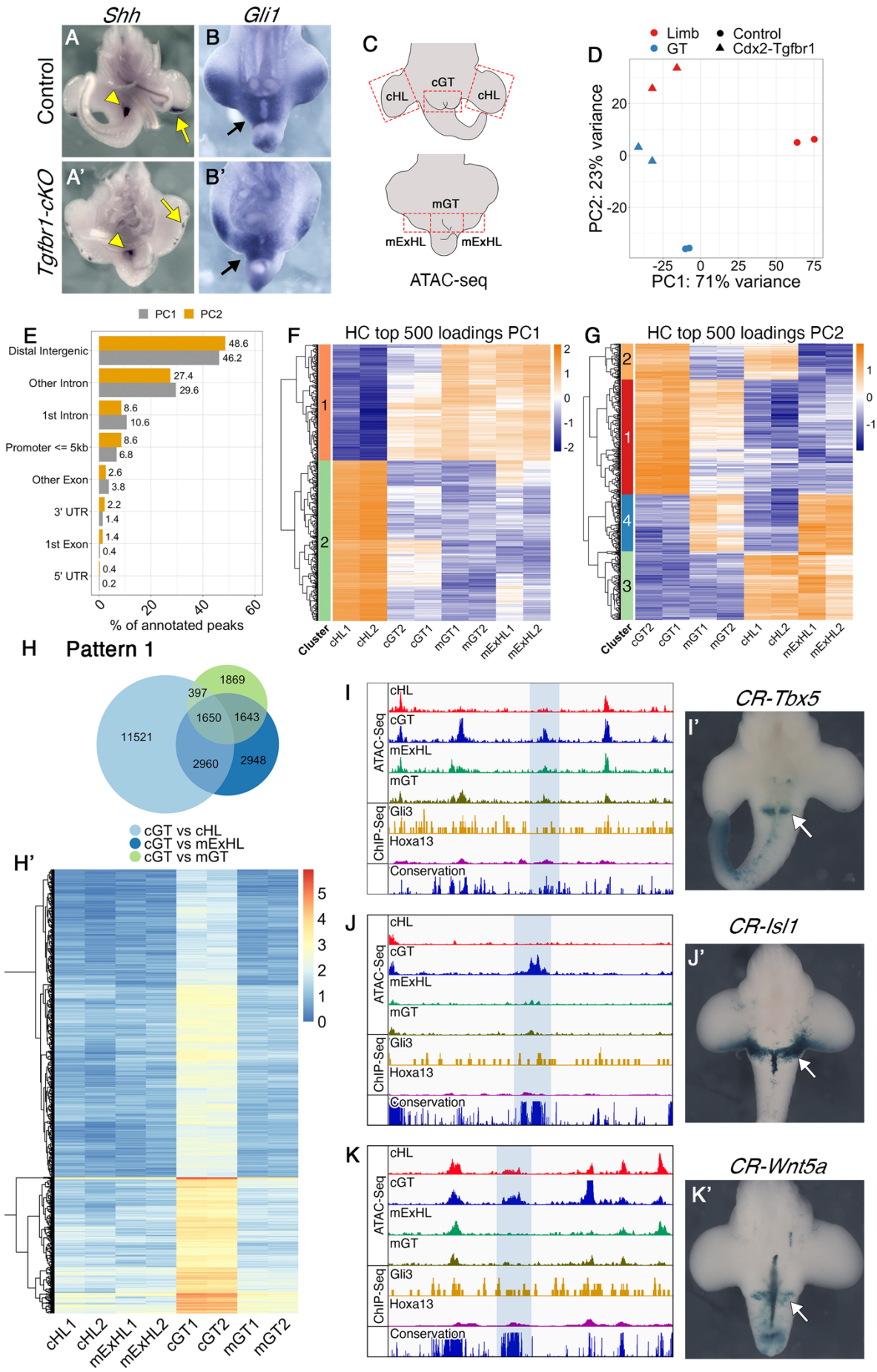


Figure 3. Limb patterning in the *Tgfbr1-cKO* embryos. (A, A') Expression of *Fgf8* in E11.5 control (A) and *Tgfbr1-cKO* (A') embryos. The black arrow shows separation between the two hindlimb prominences. (B, B') Expression of *Hand2* in E11.5 control (B) and *Tgfbr1-cKO* (B') embryos. White arrow in (B') shows a posterior expression domain of the prospective mutant extra hindlimb. (C, C') Anterior *Pax9* expression in E11.5 control (C) and *Tgfbr1-cKO* (C') hindlimbs. A single expression signal is observed in the hindlimb region of *Tgfbr1-cKO* embryo corresponding to the most anterior part of the structure (white arrowhead). (D-e') Dorsal-ventral polarity in the *Tgfbr1-cKO* mutant's hindlimb. Expression of the ventral limb marker *En1* in E10.5 control (D, d) and *Tgfbr1-cKO* (D', d') embryos. (D, D') show lateral views of the hindlimb region in whole mount embryos; (d, d') show transversal sections through the region indicated in D, D'. 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 (E, e) and *Tgfbr1-cKO* (E', e') embryos. (E, E') show lateral views of the hindlimb region in whole mounted embryos; (e, e') show transversal sections through the region indicated in E, E'. Yellow arrowheads show *Lmx1b* expression in the dorsal mesenchyme of the limb bud. n=3 embryos were analyzed per probe.


Shh
Gli1
Control
Tgfbr1-cKO
A
B
A'
B'
C
cGT
cHL
cHL
mGT
mExHL
mExHL
ATAC-seq
D
Limb
GT
Control
Cdx2-Tgfbr1
PC2: 23% variance
PC1: 71% variance
E
PC1
PC2
Distal Intergenic
Other Intron
1st Intron
Promoter <= 5kb
Other Exon
3' UTR
1st Exon
5' UTR
% of annotated peaks
F
HC top 500 loadings PC1
G
HC top 500 loadings PC2
Cluster
cHL1
cHL2
cGT2
cGT1
mGT1
mGT2
mExHL1
mExHL2
H
Pattern 1
11521
397
1869
1650
1643
2960
2948
cGT vs cHL
cGT vs mExHL
cGT vs mGT
H'
I
J
K
ATAC-Seq
ChIP-Seq
cHL
cGT
mExHL
mGT
Gli3
Hoxa13
Conservation
CR-Tbx5
I'
CR-Isl1
J'
CR-Wnt5a
K'

1194

Figure 4. 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) and *Tgfbr1-cKO* (A') embryos. *Gli1* expression in E10.5 control (B) and *Tgfbr1-cKO* (B') embryos (n=3 embryos were analyzed per probe). (C) Schematic representation of the tissue dissected for the ATAC-seq experiments in control (top) and mutant (bottom) embryos. (D) PCA plot of the ATAC-seq samples. (E) Frequencies at the annotated genomic regions of 500 top loadings from the PC1 (grey) and PC2 (yellow). (F, G) Hierarchical clustering (HC) of PC1 (F) and PC2 (G) 500 top loadings. Heatmaps show scaled normalized counts. (H, H') Chromatin regions following pattern 1. (H) 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]. (H') Heatmap showing log2 normalized counts across samples of the regions following pattern 1. (I, J, K) ATAC-seq tracks showing accessible chromatin in control GT, but not in the other samples in a region 93.7 kb upstream of the *Tbx5* TSS (I), 24.0 kb downstream of the *Isl1* TSS (J), and 91.7 kb upstream of the *Wnt5a* TSS (K). The regions highlighted with the blue shade are highly conserved among placental animals (bottom track). Two independent biological replicates were analyzed per tissue in the ATAC-seq experiments. (I', J', K') The highlighted elements drive β-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. None of the analyzed enhancers showed enrichment in Hoxa13 or Gli3 binding activity in the forelimbs, consistent with the GT specificity of those elements. cHL: control hindlimb; mExHL: mutant extra hindlimb; cGT: control GT; mGT: mutant GT; CR: conserved region.

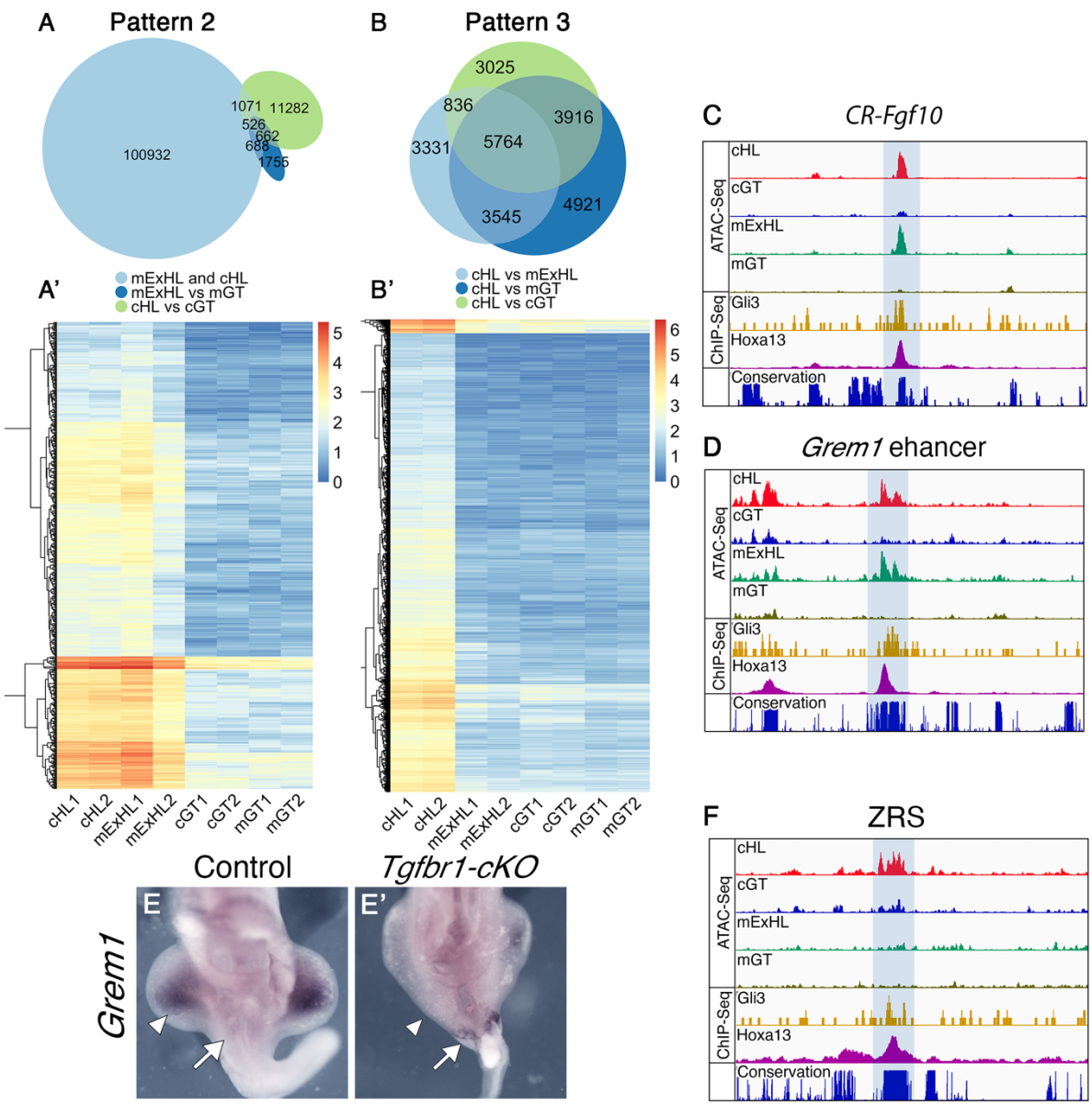


Figure 5. (A, A') Chromatin regions following pattern 2 and pattern 3. (A) Venn diagram showing interception between the regions more accessible in mutant extra limb than in mutant GT [FDR<0.001, logFC>1.5] and more accessible in control hindlimb than in control GT [FDR<0.001, logFC>1.5], but at the 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]. (B') Heatmap showing log2 normalized counts for the chromatin regions following pattern 3 across samples. (C, D) ATAC-Seq profiles showing *Fgf10* putative regulatory element (C) and *Grem1* enhancer (D) 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) (obtained from GSE81356 and GSE133710, respectively), indicating their binding to those elements in the forelimb bud. The lower track shows conservation in placental mammals. (E, E') *Grem1* expression in E10.5 wild type (E) or *Tgfbr1-cKO* (E') 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

1238 (white arrow). *Grem1* is expressed in the developing limb bud of the control embryo, but it
1239 is almost not detectable in the developing limb buds of the mutant embryo (white
1240 arrowheads) (n=3 embryos are analyzed per probe). (F) ATAC-seq data through the ZRS
1241 (highlighted with the blue shadow), showing that it is accessible in the control limb but not
1242 in any of the other samples. Two independent biological replicates were analyzed for all
1243 ATAC-seq experiments. CR: conserved region; cHL: control hindlimb; mExHL: mutant extra
1244 hindlimb; cGT: control GT; mGT: mutant GT.
1245

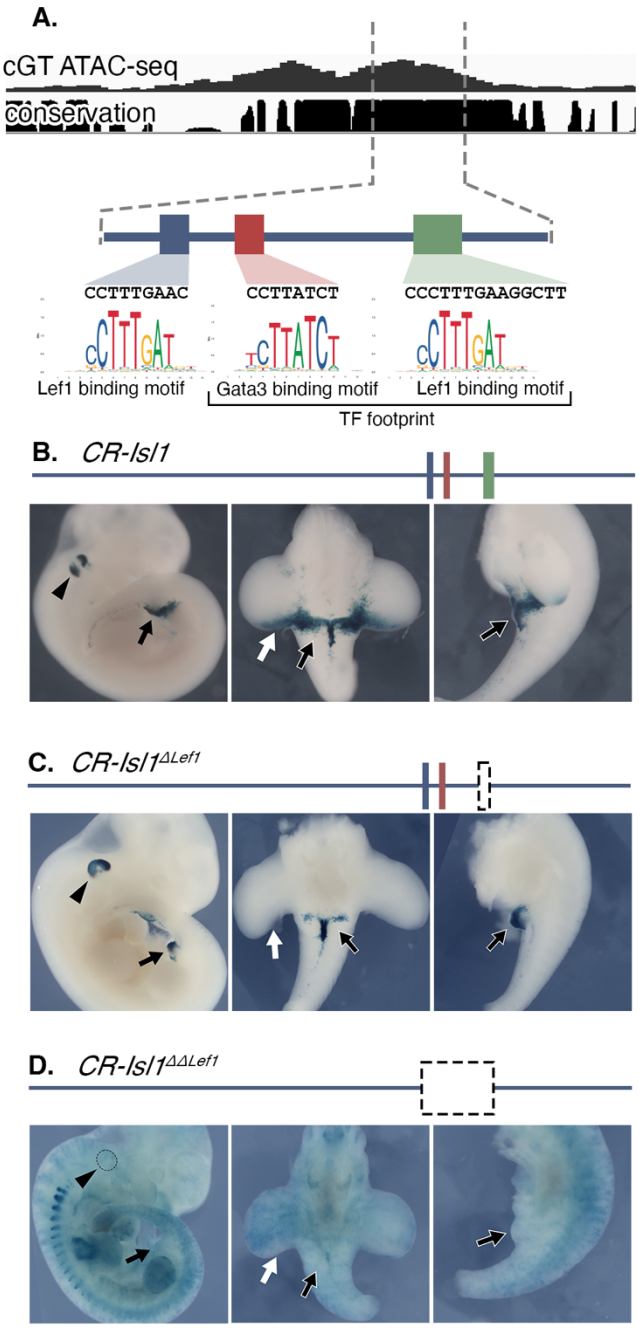
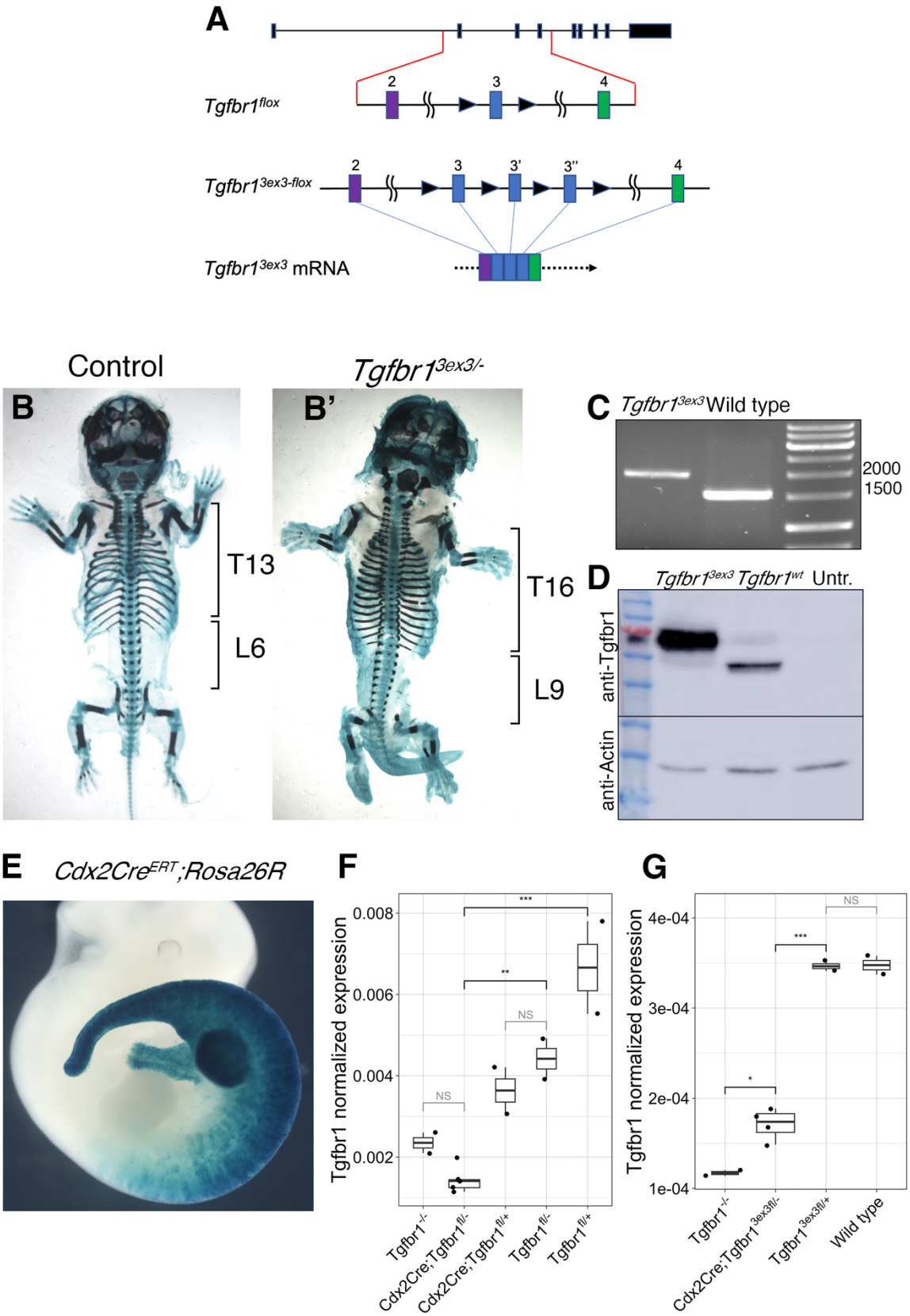
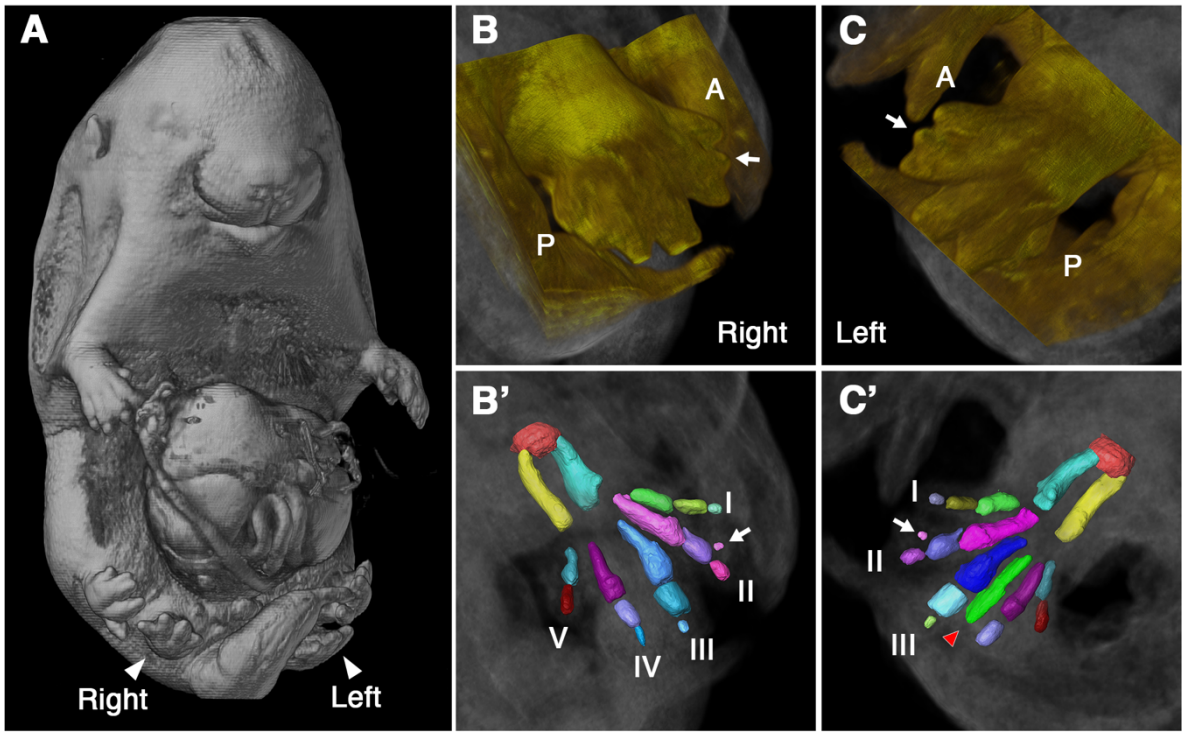


Figure 6. *CR-Isl1* is a potential Wnt-responsive element. (A) Schematic representation of the *CR-Isl1* region of Fig. 4J. The top panel shows the ATAC-seq signal in control GT and the placental conservation score. The footprint corresponded to a ~65 bp region containing a Gata3 binding motif (red), and Lef1 binding motif (green). An additional Lef1 binding motif (blue) was identified 21 bp upstream of the Gata3 motif, although it was not included within the footprint. The genomic sequences of these motifs are shown along with position frequency matrices (PFMs). (B-D) Activation of the β-galactosidase reporter by the *CR-Isl1* (B), *CR-Isl1*$^{\Delta Lef1}$ (C), and *CR-Isl1*$^{\Delta\Delta Lef1}$ (D) elements in E11.5 embryos. Black arrowheads show expression in otic vesicle; white arrow indicates expression in the posterior limb bud; black arrow shows expression in the GT.

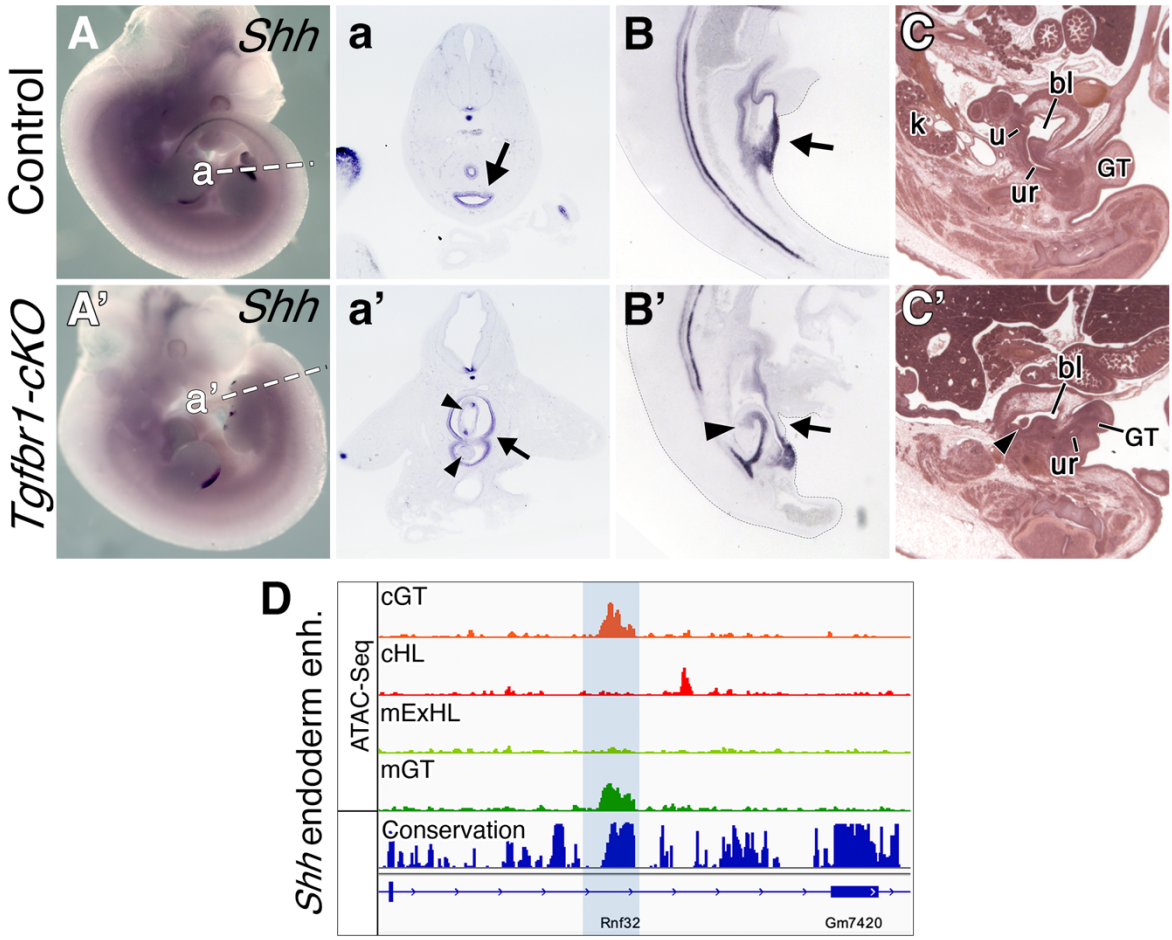


Supplementary figure 1. Generation on the *Tgfbr1-cKO*. (A) Scheme showing *Tgfbr1*^{flox} and *Tgfbr1*$^{3ex3\text{-}flox}$ alleles. Highlighted is the region containing exons 2, 3 and 4 (colored boxes) and the LoxP sites (black triangles). Also represented is the generation of a mRNA

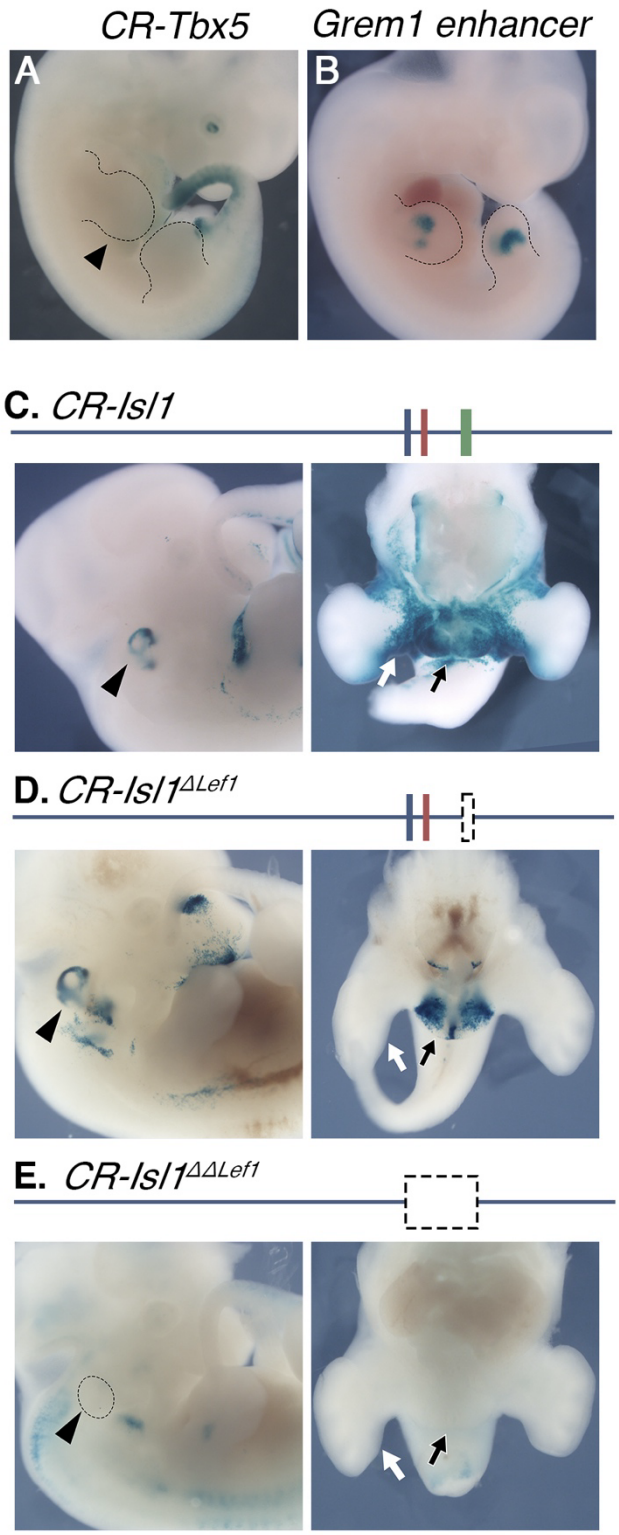
with a triplicated exon 3 after splicing of the transcript from the *Tgfbr1*$^{3ex3\text{-}flox}$ allele. (B, B') Skeletal staining of a E17.5 wild type (B) and *Tgfbr1*$^{3ex3\text{-}flox/-}$ fetus (B'), showing the presence of additional thoracic (T) and lumbar (L) vertebrae, resembling a *Gdf11* mutant phenotype. (C) RT-PCR fragment spanning the whole *Tgfbr1* coding region using RNA isolated from wild type (1,5 kb) and *Tgfbr1*$^{3ex3\text{-}flox/3ex3\text{-}flox}$ (2 kb) embryos. (D). Western blot analysis of protein extracts from HEK293T cells transfected with mammalian expression vectors carrying either the *Tgfbr1*^{3ex3} or the *Tgfbr1* coding regions, showing that the *Tgfbr1*^{3ex3} transcript produces a larger protein than its wild type counterpart. (E) Activation of β-galactosidase expression in a tamoxifen-treated E10.5 *Cdx2Cre*^{ERT}*;Rosa26R* embryo following the scheme used in this study. (F) RT-qPCR showing recombination efficiency in *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 functional allele. (G) RT-qPCR showing recombination efficiency in *Tgfbr1*$^{3ex3\text{-}flox/-}$*;Cdx2Cre*$^{ERT+/0}$ embryos. Normalized *Tgfbr1* expression level in *Tgfbr1*$^{3ex3\text{-}flox/-}$*;Cdx2Cre*$^{ERT+/0}$ embryos is significantly lower than in control embryos, although not reaching *Tgfbr1*$^{-/-}$ level, consistent with lower amount of tamoxifen given compared to the *Tgfbr1*^{flox} line. NS: nonsignificant, *p=0.018, **p<0.01, ***p<0.001.



Supplementary Figure 2. Skeletal phenotype of the *Tgfbr1-cKO* extra hindlimb. (A) Whole mount image of the OPT acquired E16.5 fetus. (B, B') Whole mount (B) and 3D reconstruction of the skeleton (B') of the right extra hindlimb. The autopod contained 5 digits, most of them with two phalanges, except for digit V, which only generated one phalanx and digit II, which contained two adjacent distal phalanges (white arrow in B and B'). (C, C') Whole mount (C) and 3D reconstruction of the skeleton (C') of the left extra hindlimb. The autopod contained 6 metatarsal bones. Anterior digits developed two phalanges, except for digit II, which generated two adjacent distal phalanges (white arrow in C and C'); the posterior digits developed one phalanx, except for one digit that did not develop any distal phalanges (red arrowhead in C').



Supplementary figure 3. Urogenital malformations in the *Tgfbr1-cKO.* (A-a') *Shh* expression in E11.5 control (A, a) and *Tgfbr1-cKO* (A', a') embryos. (a, a') show transversal sections through the cloacal region at the level indicated by the dashed line in A and A'. (B, B') Sagittal sections through E10.5 control (B) and *Tgfbr1-cKO* (B') embryos stained with a *Shh* probe. (C, C') Sagittal histological sections through bladder region of E16.5 control (C) and *Tgfbr1-cKO* (C') fetuses. Black arrows in a, a', B, B' indicate the cloaca; arrowheads in a', B', C' indicate intra-cloacal protrusions; bl: bladder; GT: genital tubercle; k: kidney; u: ureter; ur: urethra. (D) ATAC-Seq profiles through the region containing the enhancer regulating endodermal *Shh* expression (blue shadow). Bottom track shows conservation in placental mammals. cHL: control hindlimb; mExHL: mutant extra hindlimb; cGT: control GT; mGT: mutant GT.



Supplementary Figure 4. (A) The putative *Tbx5* GT enhancer shown in Fig. 4I' does not drive β-galactosidase expression in the forelimb (black arrowhead, forelimb outlined with the dashed line). (B) β-galactosidase expression driven by the *Grem1* enhancer in Fig. 5D using a transgenic reporter assay [n=2/8] (limbs outlined with the dashed line). (C-E) Activation of the β-galactosidase reporter by the *CR-Isl1* (C), *CR-Isl1*$^{\Delta Lef1}$ (D), and *CR-Isl1*$^{\Delta\Delta Lef1}$ (E) region in E12.0 embryos. Of note, the embryo shown in D is the one with the strongest reporter expression in the GT region of all *Isl1*$^{\Delta Lef1}$ transgenic embryos analyzed.

1316 Top panels show schematic representation of DNA region: blue rectangle: 5' Lef1 binding
1317 motif; red rectangle: Gata3 binding motif; green rectangle: 3' Lef1 binding motif within the
1318 TF footprint observed in the *Isl1*-related Pattern 1 element. Black arrowheads show
1319 expression in otic vesicle; white arrow shows expression in posterior limb bud; black arrow
1320 indicates expression in GT. CR: conserved region.
1321
1322
1323 **Supplementary movie 1.** OPT image from a E16.5 *Tgfbr1-cKO* fetus highlighting
1324 the limbs. The extra hindlimbs are shown in magenta.
1325
1326 **Supplementary movie 2.** OPT image from a E16.5 wild type fetus highlighting
1327 the caudal end of the intestine (red), the kidneys (cold green), the ovaries and
1328 uterus (bright green), the vagina (yellow), the urinary bladder (dark blue) and the
1329 urethra (light blue).
1330
1331 **Supplementary movie 3.** OPT image from a E16.5 *Tgfbr1-cKO* fetus highlighting
1332 the caudal end of the intestine (red), the urinary bladder (dark blue) and the
1333 urethra (light blue).
1334
1335
1336 **Supplementary Table 1.** Transcription factor footprint analysis of pattern 1
1337 elements. Column A: number of pattern 1 regions with at least 1 occurrence of the
1338 specified transcription factor; column B: number of pattern1 regions with no
1339 occurrence of the specified transcription factor; C: number of background regions
1340 with at least 1 occurrence of the specified transcription factor; D: number of
1341 background regions with no occurrence of the specified transcription factor.
1342
1343 **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-Isl1 - Left1 sites deletions	Primer1	GCGAATTCGGCTGAATGGAGGTCATTGGA
	Primer2	GCGAATTCTTCTGGCTTCAGTTCCTGAGAC
	Primer3	GCGAATTCCAGGTATGAAAAGCTGTGC
CR-Tbx5	Forward	GACTCGAGCTTTATGTATCTGAGCACACTG
	Reverse	GAAGATCTTCCTTCAACAAACCAT CCAC
CR-Wnt5a	Forward	CTGCTCGAGACTAGTACCTGCTGCTTATCAACCTAGTG
	Reverse	CAGCCATGGCTCCTACCTCACTGACTATCCAG
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		
En1	Davis and Joyner, 1988, ref [94]	
Fgf10	Whole coding region cloned into TOPO. Oligonucleotides listed above	
Fgf8	Crossley and Martin, 1995, ref [95]	
Grem1	PCR fragment with the whole coding region and T7 promoter at the 3' end. Oligonucleotides listed above	
Hand2	Srivastava et al, 1997, ref [96]	
Isl1	Jurberg et al 2013, ref [2]	
Lin28a	Aires et al., 2019, ref [97]	
Lmx1b	PCR fragment encompassing nucleotides 504 to 1202 of the mRNA cloned into pKS bluescript. Oligonucleotides listed above.	
Pitx1	Szeto et al, 1999, ref [98]	
Shh	Echelard et al, 1993, ref [99]	
Tbx5	Chapman et al, 1996, ref [100]	
Wnt5a	Yamaguchi et al, 1999, ref [32]	

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

Element	Coordinates
CR-Isl1	chr13:116,284,356-116,285,685
CR-Tbx5	chr5:119,701,323-119,702,815
CR-Wnt5a	chr14:28,412,522-28,413,803
CR-Fgf10	chr13:118,745,999-118,747,518
Grem1 enhancer	chr2: 113,580,982-113,582,008

1348