# A gene desert required for regulatory control of pleiotropic *Shox2* expression and embryonic survival

Samuel Abassah-Oppong[1,12,13], Brandon J. Mannion[2,11,12], Matteo Zoia[3,12], Raquel Rouco[4], Virginie Tissières[3,5,6], Cailyn H. Spurrell[2], Virginia Roland[3], Fabrice Darbellay[2,4], Anja Ljubojevic[1], Julie Gamart[3,6], Tabitha A. Festa -Daroux[1], Carly S. Sullivan[1], Eddie Rodríguez-Carballo[7,14], Yoko Fukuda-Yuzawa[2], Riana Hunter[2], Catherine S. Novak[2], Ingrid Plajzer-Frick[2], Stella Tran[2], Jennifer A. Akiyama[2], Diane E. Dickel[2], Javier Lopez-Rios[5], Iros Barozzi[2,8], Guillaume Andrey[4], Axel Visel[2,9,10], Len A. Pennacchio[2,9,11], John Cobb[1,*] and Marco Osterwalder[2,3,6*]

[1] Department of Biological Sciences, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, T2N 1N4, Canada.

[2] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[3] Department for BioMedical Research (DBMR), University of Bern, 3008 Bern, Switzerland.

[4] Department of Genetic Medicine and Development and iGE3, Faculty of Medicine, University of Geneva, 1211 Geneva, Switzerland.

[5] Centro Andaluz de Biología del Desarrollo (CABD), CSIC-Universidad Pablo de Olavide-Junta de Andalucía, 41013 Seville, Spain.

[6] Department of Cardiology, Bern University Hospital, 3010 Bern, Switzerland.

[7] Department of Genetics and Evolution, University of Geneva, 1211 Geneva, Switzerland.

[8] Center for Cancer Research, Medical University of Vienna, Vienna, Austria.

[9] US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[10] School of Natural Sciences, University of California, Merced, Merced, CA 95343, USA.

[11] Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA.

[12] These authors contributed equally.

[13] Present address: Philander Smith College, Biology Department, Little Rock, AR 72202.

[14] Present address: Department of Molecular Biology, University of Geneva, 1211 Geneva, Switzerland.

[*] Correspondence: jacobb@calgary.ca(J.C.), marco.osterwalder@unibe.ch (M.O.)

**Keywords**: *Shox2*, gene desert, enhancers, gene regulation, chromatin architecture, limb development, cardiac development, pacemaker, transcription factors.

## ABSTRACT

Gene deserts are defined as genomic regions devoid of protein coding genes and spanning more than 500 kilobases, collectively encompassing about 25% of the human genome. Approximately 30% of all gene deserts are enriched for conserved elements with *cis*-regulatory signatures. These are located predominantly near developmental transcription factors (TFs) but despite predicted critical functions, the transcriptional contributions and biological necessity of most gene deserts remain elusive. Here, we explore the *cis*-regulatory impact of a gene desert flanking the *Shox2* gene, a TF indispensable for proximal limb, craniofacial and cardiac pacemaker development. Using a functional genomics approach in mouse embryos we identify the gene desert as a hub for numerous *Shox2*-overlapping enhancers arranged in a globular chromatin domain with tissue-specific features. In accordance, using endogenous CRISPR deletion, we demonstrate that the gene desert interval is essential for *Shox2* transcriptional control in developing limbs, craniofacial compartments, and the heart. Phenotypically, gene desert ablation leads to pacemaker-related embryonic lethality due to *Shox2* depletion in the cardiac sinus venosus. We show that this role is partially mediated through a distal gene desert enhancer, providing evidence for intra-gene desert regulatory robustness. Finally, we uncover a multi-layered functional role of the gene desert by revealing an additional requirement for stylopod morphogenesis, mediated through an array of proximal limb enhancers (PLEs). In summary, our study establishes the *Shox2* gene desert as a fundamental genomic unit that controls pleiotropic gene expression through modular arrangement and coordinated dynamics of tissue-specific enhancers.

70 **INTRODUCTION**

71 Functional assessment of gene deserts, gene-free chromosomal segments larger than 500 kilobases
72 (kb), has posed considerable challenges since these large noncoding regions were shown to be a
73 prominent feature of the human genome more than 20 years ago[1]. Stable gene deserts (n=172 in the
74 human genome, ~30% of all gene deserts) share more than 2% genomic sequence conservation
75 between human and chicken, are enriched for putative enhancer elements and frequently located near
76 developmental genes, suggesting a critical role in embryonic development and organogenesis[2–4].
77 However, genomic deletion of an initially selected pair of gene deserts displayed mild effects on the
78 expression of nearby genes and absence of overt phenotypic alterations[5]. In contrast, gene deserts
79 centromeric and telomeric to the *HoxD* cluster were shown to harbor "regulatory archipelagos" i.e.,
80 multiple tissue-specific enhancers that collectively orchestrate spatiotemporal and colinear *HoxD*
81 *gene* expression in developing limbs and other embryonic compartments[6,7]. These antagonistic gene
82 deserts represent individual topologically associating domains (TADs) separated by the *HoxD* cluster
83 which acts as a dynamic and resilient CTCF-enriched boundary region[8,9]. Despite such critical roles,
84 the functional requirement of only few gene deserts have been studied in detail, including the
85 investigation of chromatin topology and functional enhancer landscapes in the TADs of other
86 developmental key regulators such as *Sox9, Shh* or *Fgf8*[10–12].

87 Self-associating TADs identified by 3D chromatin conformation capture are described as
88 primary higher-order chromatin structures that constrain *cis*-regulatory interactions to target genes
89 and facilitate long-range enhancer-promoter (E-P) contacts[13,14]. TADs are thought to emerge through
90 Cohesin-mediated chromatin loop extrusion and are delimited by association of CTCF to convergent
91 binding sites[15,16]. Re-distribution of E-P interactions can lead to pathogenic effects due to
92 perturbation of CTCF-bound TAD boundaries or re-configuration of TADs[10,17]. Therefore, functional
93 characterization of the 3D chromatin topology and transcriptional enhancer landscapes across gene
94 deserts is a prerequisite for understanding the developmental mechanisms underlying mammalian
95 embryogenesis and human syndromes[18]. Recent functional studies in mice have uncovered that

96 mRNA expression levels of developmental regulator genes frequently depend on additive
97 contributions of enhancers within TADs[19–22]. Hereby, the contribution of each implicated enhancer
98 to total gene dosage can vary, illustrating the complexity of transcriptional regulation through E-P
99 interactions[23]. In addition, nucleotide mutations affecting TF binding sites in enhancers can disturb
100 spatiotemporal gene expression patterns, with the potential to trigger phenotypic abnormalities such
101 as congenital malformations due to altered properties of developmental cell populations[24–26].

102 In the current study, we focused on the functional characterization of a stable gene desert
103 downstream (centromeric) of the mouse short stature homeobox 2 (*Shox2*) transcription factor (TF).
104 Tightly controlled *Shox2* expression is essential for accurate development of the stylopod (humerus
105 and femur), craniofacial compartments (maxillary-mandibular joint, secondary palate), the facial
106 motor nucleus and its associated facial nerves, and a subset of neurons of the dorsal root ganglia[27–
107 33]. In addition, *Shox2* in the cardiac sinus venosus (SV) is required for differentiation of progenitors
108 of the sinoatrial node (SAN), the dominant pacemaker population during embryogenesis and
109 adulthood[34–36]. *Shox2* inactivation disrupts Nkx2-5 antagonism in SAN pacemaker progenitors and
110 results in hypoplasia of the SAN and venous valves, leading to bradycardia and embryonic
111 lethality[34,35,37]. In accordance with this role, *SHOX2*-associated coding and non-coding variants in
112 humans were implicated with SAN dysfunction and atrial fibrillation[38–40]. Tbx5 and Isl1 cardiac TF
113 regulators were shown to act upstream of *Shox2* in SAN development[41–44] and *Isl1* is sufficient to
114 rescue *Shox2*-mediated bradycardia in zebrafish hearts[45].

115 The human SHOX gene located on the pseudo-autosomal region (PAR1) of the X and Y
116 chromosomes represents a paralog of SHOX2 (on chromosome 3), hence dividing Shox gene
117 function. *SHOX* is associated with defects and syndromes affecting skeletal, limb and craniofacial
118 morphogenesis[28,46,47]. Rodents have lost their *SHOX* gene during evolution along with other pseudo-
119 autosomal genes and mouse *Shox2* features an identical DNA-interacting homeodomain replaceable
120 by human *SHOX* in a mouse knock-in model[27,48]. Remarkably, while *Shox2*/*SHOX2* genes show
121 highly conserved locus architecture, the *SHOX* gene also features a downstream gene desert of similar

122 extension, containing neural (hindbrain) enhancers with overlapping activities[47]. Our previous
123 studies revealed that *Shox2* transcription in the developing mouse stylopod is partially controlled by
124 a pair of human-conserved limb enhancers termed hs741 and hs1262/LHB-A, the latter residing in
125 the gene desert[19,47,49]. However, the rather moderate loss of *Shox2* limb expression in absence of these
126 enhancers indicated increased complexity and potential redundancies in the underlying enhancer
127 landscape[19].

128 Here we identified the *Shox2* gene desert as a critical *cis*-regulatory domain encoding an array
129 of distal enhancers with specific subregional activities, predominantly in limb, craniofacial, neuronal,
130 and cardiac cell populations. We found that interaction of these enhancers with the *Shox2* promoter
131 is likely facilitated by a chromatin loop anchored downstream of the *Shox2* gene body and exhibiting
132 tissue-specific features. Genome editing further demonstrated essential pleiotropic functions of the
133 gene desert, including a requirement for craniofacial patterning, limb morphogenesis, and embryonic
134 viability through enhancer-mediated control of SAN progenitor specification. Our results identify the
135 *Shox2* gene desert as a dynamic enhancer hub ensuring pleiotropic and resilient *Shox2* expression as
136 an essential component of the gene regulatory networks (GRNs) orchestrating mammalian
137 development.

138 ## RESULTS

139 **Gene desert enhancers recapitulate patterns of pleiotropic *Shox2* expression.**

140 The gene encoding the *Shox2* transcriptional regulator is located in a 1 megabase (Mb) TAD
141 (chr3:66337001-67337000) and flanked by a stable gene desert spanning 675kb of downstream
142 (centromeric) genomic sequence (**Fig. 1A**). The *Shox2* TAD only contains one other protein coding
143 gene, *Rsrc1*, located adjacent to *Shox2* on the upstream (telomeric) side and known for roles in pre-
144 mRNA splicing and neuronal transcription[50,51] (**Fig. 1A**). Genes located beyond the TAD boundaries
145 show either near-ubiquitous (*Mlf1*) or *Shox2*-divergent (*Veph1*, *Ptx3*) expression signatures across
146 tissues and timepoints (**Fig. S1A**). While *Shox2* transcription is dynamically regulated in multiple
147 tissues including proximal limbs, craniofacial subregions, cranial nerve, brain, and the cardiac sinus
148 venosus (SV), only a limited number of *Shox2*-associated enhancer sequences have been previously
149 validated in mouse embryos[19,47,49,52] (**Fig. 1A, B, S1A**). These studies identified a handful of enhancer
150 elements in the *Shox2* TAD driving reporter activity almost exclusively in the mouse embryonic brain
151 (hs1251, hs1262) and limbs (hs741, hs1262, hs638) (Vista Enhancer Browser) (**Fig. 1A**). To predict
152 *Shox2* enhancers more systematically, and to estimate the number of developmental enhancers in the
153 gene desert, we established a map of stringent enhancer activities based on chromatin state profiles[53]
154 (ChromHMM) and H3K27 acetylation (H3K27ac) ChIP-seq peak calls across 66 embryonic and
155 perinatal tissue-stage combinations from ENCODE[54] (https://www.encode.project.org) (see
156 ***Methods***). After excluding promoter regions, this analysis identified 20 elements within the *Shox2*-
157 TAD and its border regions, each with robust enhancer marks in at least one of the tissues and
158 timepoints (E11.5-15.5) examined (**Fig. S1A and Table S1**). Remarkably, 17 of the 20 elements
159 mapping to the *Shox2* TAD or border regions were located within the downstream gene desert, with
160 the majority of H3K27ac signatures overlapping *Shox2* expression profiles across multiple tissues
161 and timepoints, indicating a role in regulation of pleiotropic *Shox2* expression (**Figs. 1B, S1A**). The
162 previously validated hs741 and hs1262 limb enhancers were not among these predictions as elevated
163 H3K27ac marks are present at these enhancers already at earlier stages[19,55].

164 To determine the *in vivo* activity patterns for each of the predicted gene desert enhancer (DE)
165 elements, we performed *LacZ* transgenic reporter analysis in mouse embryos at E11.5, a stage
166 characterized by wide-spread and functionally relevant *Shox2* expression in multiple tissues[19,47] (**Fig.**
167 **1B, C**). This analysis included the validation of 16 genomic elements (DE +329kb and +331kb were
168 part of a single reporter construct) and revealed reproducible enhancer activities in 12/16 cases (**Figs.**
169 **1B, S1B and Table S2**). Most of the individual enhancer activities localized to either craniofacial,
170 cranial nerve, mid-/hindbrain or limb subregions known to be dependent on *Shox2* expression and
171 function[27,28,30,31] (**Fig. 1C**). For example, DE9 (+475kb) and DE15 (+606kb), both exhibiting limb
172 and craniofacial H3K27ac marks, drove LacZ reporter expression exclusively in *Shox2*-overlapping
173 craniofacial domains in the medial nasal (MNP) and maxillary-mandibular (MXP, MDP) processes,
174 respectively (**Fig. 1C**). In line with DE15 activity, *Shox2* expression in the MXP-MDP junction is
175 known to be required for temporomandibular joint (TMJ) formation in jaw morphogenesis[28]. DE1, 5
176 and 12 instead showed activities predominantly in cranial nerve tissue, including the trigeminal
177 (TGn), facial (FGn) and jugular (JGn) ganglia, as well as the dorsal root ganglia (DRG) (**Fig. 1C**).
178 *Shox2* is expressed in all these neural crest-derived tissues, but a functional requirement has only
179 been demonstrated for FGn development and the mechanosensory neurons of the DRG[30,32]. While
180 no H3K27ac profiles for cranial nerve populations were available from ENCODE[56], both DE5 and
181 DE12 elements showed increased H3K27ac in craniofacial compartments at E11.5 (**Fig. 1B**), likely
182 reflecting the common neural-crest origin of a subset of these cell populations[57]. At mid-gestation,
183 *Shox2* is also expressed in the diencephalon (DiE), midbrain (MB) and hindbrain (HB), and is
184 specifically required for cerebellar development[31]. Gene desert enhancer assessment also identified
185 a set of brain enhancers (DE7, 14 and 16) overlapping *Shox2* domains in the DiE, MB and/or HB.
186 (**Fig. 1C**). Although H3K27ac marks were present in limbs at most predicted DEs, only two elements
187 (DE6 and DE10) drove LacZ reporter expression in the E11.5 limb mesenchyme in a sub-regionally
188 or limb type-restricted manner, respectively (**Fig. 1C**). However, despite elevated cardiac H3K27ac
189 in a subset of DEs, none of the validated elements drove reproducible LacZ reporter expression in

190 the heart at E11.5 (**Fig. 1B, 1C**). Taken together, our *in vivo* enhancer-reporter screen based on
191 systematic epigenomic profiling and transgenic reporter validation identified multiple DE elements
192 with *Shox2*-overlapping activities, pointing to a role of the gene desert as an enhancer hub directing
193 pleiotropic *Shox2* transcription.
194
195 **The *Shox2* gene desert shapes a chromatin loop with tissue-specific features.**
196 Recent studies have shown that sub-TAD interactions can be pre-formed or dynamic, and that 3D
197 chromatin topology can affect enhancer-promoter communication in distinct cell types or tissues[58–
198 60]. To explore the 3D chromatin topology across the *Shox2* TAD and flanking regions, we performed
199 region capture HiC (C-HiC) targeting a 3.5Mb interval in dissected E11.5 mouse embryonic
200 forelimbs, mandibles, and hearts, tissues known to be affected by *Shox2* loss-of-function (**Figs. 2A,**
201 **S2A**). C-HiC contact maps combined with analysis of insulation scores to infer inter-domain
202 boundaries revealed a tissue-invariant *Shox2*-containing TAD that matched the extension observed
203 in mESCs[61] (**Figs. 2A, S2A, Table S3**). C-HiC profiles further showed sub-TAD organization into
204 *Shox2*-flanking upstream (U-dom) and downstream (D-dom) domains as hallmarked by loop anchors
205 and insulation scores, with the D-dom spanning almost the entire gene desert (**Figs. 2A, S2A**). Virtual
206 4C (v4C) using a viewpoint centered on the *Shox2* transcriptional start site (TSS) further
207 demonstrated confinement of *Shox2*-interacting elements to U-dom and D-dom intervals, or TAD
208 boundary regions (**Fig. 2A, Table S3**). Remarkably, the most distal D-dom compartment spanning
209 ~170kb revealed dense chromatin contacts restricted to heart tissue and delimited by weak insulation
210 boundaries which co-localized with non-convergent CTCF sites (**Fig. 2A, B, S2A**). While this high-
211 density contact domain (HCD) contained the majority of the previously identified (non-cardiac) gene
212 desert enhancers (DE5-12), subtraction analysis further corroborated increased chromatin contacts
213 across the HCD and domain insulation specifically in cardiac tissue as opposed to limb or mandibular
214 tissue, indicating a potentially repressive function due to condensed chromatin state (**Fig. 2A-C,**
215 **S2B**). However, no region-specific accumulation of repressive histone marks (H3K27me3 or

216 H3K9me3) was observed in whole heart samples from ENCODE (**Fig. S3A, B**). Instead, V4C
217 subtraction analysis with defined viewpoints on positively validated DEs indicated that specifically
218 in heart tissue, enhancer elements outside the HCD (DE1, 15) were reduced in contacts with elements
219 inside (**Fig. S3C**). In turn, enhancer viewpoints inside the HCD (DE5, 9, 10) showed reduced contacts
220 with elements outside (**Fig. S3C**). Collectively, our results imply that *Shox2* is preferentially
221 regulated by upstream (U-dom) and downstream (D-dom) regulatory domains that contain distinct
222 sets of active tissue-specific enhancers. Hereby, the gene desert forms a topological chromatin
223 environment (D-dom) that in specific tissue context might enable certain enhancers to interact more
224 efficiently with the *Shox2* promoter.
225
226 **Control of pleiotropic *Shox2* dosage and embryonic survival by the gene desert.**
227 To explore the functional relevance of the gene desert as an interactive hub for *Shox2* enhancers in
228 mouse embryos, we used CRISPR/Cas9 in mouse zygotes to engineer an intra-TAD gene desert
229 deletion allele ($GD^{\Delta}$) (**Figs. 3A, S4A, B; Tables S4, S5**). F1 mice heterozygous for this allele ($GD^{\Delta/+}$)
230 were born at expected Mendelian ratios and showed no impaired viability and fertility. Following
231 intercross of $GD^{\Delta/+}$ heterozygotes we compared *Shox2* transcripts in $GD^{\Delta/\Delta}$ and wildtype (WT)
232 control embryos, with a focus on tissues marked by DE activities (**Figs. 1C, 3B-E**). Despite loss of
233 at least three enhancers with limb activities (hs1262, DE6, DE10), *Shox2* expression was still detected
234 in fore- and hindlimbs of $GD^{\Delta/\Delta}$ embryos, as determined by *in situ* hybridization (ISH) (**Fig. 3B**),
235 albeit at ~50% reduced levels as determined by RT-qPCR (**Fig. 3C, Table S6**). These results point
236 to a functional role of the gene desert in ensuring robust *Shox2* dosage during proximal limb
237 development[27]. *Shox2* expression in distinct craniofacial compartments was more severely affected
238 by the loss of the gene desert (**Fig. 3D, E**). Downregulation of *Shox2* transcripts was evident in the
239 MNP, anterior portion of the palatal shelves, and the proximal MXP-MDP domain of $GD^{\Delta/\Delta}$ embryos
240 at E10.5 and E11.5, compared to wild-type controls (**Fig. 3D**). Concordantly, and in contrast to *Rsrc1*
241 mRNA levels which remained normal, *Shox2* was depleted in the nasal process (NP) and MDP of

242 GD$^{\Delta/\Delta}$ embryos at E11.5 (**Fig. 3E**). Strikingly, these affected subregions corresponded to the activity
243 domains of the identified DE9 (MNP) and DE15 (MXP-MDP) gene desert enhancers indicating
244 essential craniofacial *Shox2* regulation (**Figs. 1C, 3F**). Taken together, these results demonstrate a
245 critical functional role of the gene desert in transcriptional regulation of *Shox2* during craniofacial
246 and proximal limb morphogenesis[27,28,33]. While our transgenic analysis also uncovered DEs with
247 activities in brain or cranial nerve regions (**Fig. 1C**), no overt reduction in spatial *Shox2* expression
248 was detected in corresponding subregions in GD$^{\Delta/\Delta}$ embryos (**Fig. 3D**). This is likely attributed to the
249 presence of *Shox2*-associated brain enhancers located in the U-dom (e.g., hs1413) or downstream of
250 the D-dom and the deleted gene desert interval (e.g., DE16) which show partially overlapping
251 activities.

252 Despite the lack of identification of any *in vivo* heart enhancers in the gene desert following
253 transgenic reporter analysis from epigenomic whole-heart predictions (**Fig. 1C**), spatial and
254 quantitative mRNA analysis in GD$^{\Delta/\Delta}$ embryos revealed absence of *Shox2* transcripts from the cardiac
255 sinus venosus (SV) that harbors the population of SAN pacemaker progenitors[62] (**Fig. 4A, B**). In
256 accordance with the essential role of *Shox2* in the differentiation of SAN progenitors and the related
257 lethality pattern in *Shox2*-deficient mouse embryos[34], cardiac *Shox2* depletion in GD$^{\Delta/\Delta}$ embryos
258 triggered arrested development and embryonic lethality at around E12 (n=5/5) (**Fig. S4C**).
259 Immunofluorescence further confirmed lack of Shox2 protein in the Hcn4-positive domain of SAN
260 pacemaker cells in the SV of GD$^{\Delta/\Delta}$ hearts compared to WT controls at E11.5 (**Fig. 4C, S4D**).
261 Together, these results demonstrated a requirement of the gene desert for embryonic viability directly
262 associated with transcriptional control of cardiac *Shox2*.

263

264 **Resilient gene desert enhancer architecture ensures robust cardiac *Shox2* expression.**

265 Abrogation of *Shox2* mRNA in the SV of GD$^{\Delta/\Delta}$ embryos implied the presence of enhancers with
266 cardiac activities, similar to the regulation of other TFs implicated in the differentiation of SAN
267 progenitor cells[44]. In agreement with our findings, a recent study[63] has reported that deletion of a

268 241kb interval within the gene desert (VS-250, mm10 chr3:66444310-66685547) is sufficient to
269 deplete *Shox2* in the SV. This resulted in a hypoplastic SAN and abnormally developed venous valve
270 primordia responsible for embryonic lethality[63]. We therefore concluded that loss of *Shox2* in hearts
271 of $GD^{\Delta/\Delta}$ embryos results from inactivation of one (or more) SV/SAN enhancer(s) in the VS-250
272 interval (**Fig. 4D**). While our epigenomic analysis from whole hearts identified multiple elements
273 with heart enhancer signatures (H3K27ac) (**Fig. 1B**), none was found to drive reproducible cardiac
274 activity in embryos at E11.5 by transgenic reporter analysis (**Fig. 1C**). To refine *Shox2*-associated
275 cardiac enhancer predictions we performed ATAC-seq from mouse embryonic hearts at E11.5 and
276 intersected the results with reprocessed open chromatin signatures from $HCN4^{+}$-GFP sorted SAN
277 pacemaker cells of mouse hearts at P0, available from two recent studies[44,64] (**Fig. 4D**, **Table S7**).
278 Intersection of peak calls within the VS-250 interval identified multiple sites with overlapping
279 accessible chromatin in embryonic hearts and perinatal SAN cells. While a subset of these candidate
280 SAN enhancer elements overlapped DEs validated for non-cardiac activities (DE 3, 4, 7-12), the
281 remaining ATAC-called elements (+319, +325, +389, +405, +417, +520) included yet
282 uncharacterized elements showing variable enrichment for Tbx5, Gata4 and/or Tead TFs which are
283 associated with SAN enhancer activation[44,65,66] (**Figs. 4D, S5A, Table S7**). To obtain complete
284 functional validation coverage, we subjected these new putative SV/SAN enhancer elements to LacZ
285 reporter transgenesis in mouse embryos (**Fig. 4D, Table S8**). This analysis identified a single element
286 located 325kb downstream of *Shox2* (+325) that was able to drive reproducible LacZ reporter
287 expression in the cardiac SV in a reproducible manner (**Fig. 4D, S5B**). To further define the core
288 region responsible for the SV-specific activity we divided the 4kb-spanning +325 module into two
289 elements: +325-A and +325-B (**Fig. 4E, Table S8**). These elements overlapped in a conserved block
290 of sequence (1.5kb) that showed an open chromatin peak in embryonic hearts at E11.5 and SAN cells
291 at P0, and also co-localized with Tbx5 enrichment at E12.5 (**Fig. 4E**). Both +325-A and +325-B
292 elements retained SV enhancer activity on their own in transgenic reporter assays, indicating that the
293 core sequence is responsible for SV activity (**Fig. 4E, S5B**). To identify cardiac TF interaction

294 partners in enhancers at the motif level, we then established a general framework based on a former
295 model of statistically significant matching motifs[67] and restricted to TFs expressed in the developing
296 heart at E11.5 (**Table S9**) (see ***Methods***). This approach identified a bi-directional Tbx5 motif in the
297 active core [*P*=1.69e-05 (+) and *P*=1.04e-05 (-)] of the +325 SV enhancer module which in addition
298 with ChIP-seq binding suggested direct interaction of Tbx5 (**Fig. 4E, S5C**). In contrast, no motifs or
299 binding of other established cardiac *Shox2* upstream regulators (e.g., Isl1) were identified in this core
300 sequence (**Fig. S5A, S5C**). In summary, our results identified the +325 module as a remote Tbx5-
301 interacting cardiac enhancer associated with transcriptional control of *Shox2* in the SV and thus likely
302 required for SAN progenitor differentiation[34,63].

303 The mouse +325 SV enhancer core module is conserved in the human genome where it is
304 located 268kb downstream (+268) of the TSS of the *SHOX2* orthologue. Taking advantage of fetal
305 left and right atrial (LA and RA) as well as left and right ventricular (LV and RV) tissue samples at
306 post conception week 17 (pcw17) available from the Human Developmental Biology Resource at
307 Newcastle University, we conducted H3K27ac ChIP-seq and RNA-seq to explore chamber-specific
308 SV enhancer activity during pre-natal human heart development (**Fig. 5A**). These experiments
309 uncovered an atrial-specific H3K27ac signature at the (+268) conserved enhancer module, matching
310 the transcriptional specificity of *SHOX2* distinct from the ubiquitous profile of *RSRC1* in human
311 hearts (**Fig. 5A**). This result indicating human-conserved activity prompted us to investigate the
312 developmental requirement of the SV enhancer *in vivo*. Therefore, we used CRISPR-Cas9 in mouse
313 zygotes (CRISPR-EZ)[68] to delete a 4.4kb region encompassing the +325 SV enhancer interval (SV-
314 Enh$^{\Delta}$) (**Figs. 5B, S5D, E; Tables S4, S5**). F1 mice heterozygous for the SV enhancer deletion (SV-
315 Enh$^{\Delta/+}$) were phenotypically normal and subsequently intercrossed to produce homozygous SV-
316 Enh$^{\Delta/\Delta}$ embryos. ISH analysis pointed to downregulation of *Shox2* transcripts in the SV region in SV-
317 Enh$^{\Delta/\Delta}$ embryos at E10.5 and qPCR analysis at the same stage demonstrated a ~60% reduction of
318 *Shox2* in hearts of SV-Enh$^{\Delta/\Delta}$ embryos compared to WT controls (**Fig. 5C**). Despite this reduction of
319 *Shox2* dosage in embryos, SV-Enh$^{\Delta/\Delta}$ mice were born at normal Mendelian frequency and showed

320 no overt phenotypic abnormalities during adulthood. Together, these results imply that multiple gene
321 desert enhancers are in control of *Shox2* expression in SAN progenitors and establish that in such a
322 system the +325 SV enhancer acts as a core module required for buffering of cardiac *Shox2* to protect
323 from dosage-reducing mutations.
324
325 **A gene desert limb enhancer repertoire promotes stylopod morphogenesis.**
326 Given another essential role of *Shox2* in proximal limb development we next addressed the
327 phenotypic requirement of the gene desert for skeletal limb morphogenesis. *Shox2* is essential for
328 stylopod formation and thus analysis of skeletal elements serves as an ideal readout for the study of
329 enhancer-related *Shox2* dosage reduction in the proximal limb[19,27]. Neither knockout of the hs1262
330 proximal limb enhancer[19] nor the identification of new limb enhancers (DE6, DE10) located in the
331 gene desert (**Fig. 1**) was sufficient to explain the ~50% *Shox2* reduction observed in proximal fore-
332 (FL) and hindlimbs (HL) of $GD^{\Delta/\Delta}$ embryos (**Fig. 3B, C**). To refine our epigenomic limb enhancer
333 predictions at the spatial level we reprocessed previously published ChIP-seq datasets from dissected
334 proximal and distal limbs at E12[55] which revealed multiple proximal-specific H3K27ac peaks (**Fig.**
335 **6A**). These included several elements not significantly enriched in H3K27ac maps from whole-mount
336 limb tissue (**Fig. 1B**). Interestingly, multiple elements marked by H3K27ac in proximal limbs also
337 showed H3K27me3 in distal limb mesenchyme reflecting compartment-specific bivalent epigenetic
338 regulation[55]. With the goal to identify the complement of H3K27ac-marked elements that interact
339 with the *Shox2* promoter we next performed circular chromosome conformation capture (4C-seq)
340 with a *Shox2* viewpoint from dissected proximal limbs at E12.5 (**Fig. 6B, Table S10**). Processing of
341 two replicates resulted in reproducible peaks which confirmed physical interaction between the
342 *Shox2* promoter and each of the *bona-fide* proximal limb enhancers characterized previously: hs741
343 located in the upstream domain (U-dom) and hs1262 located in the gene desert (D-dom)[19,47] (**Fig.**
344 **6B, C**). Other prominent 4C-seq peaks in the gene desert co-localized with either previously validated
345 enhancer elements with non-limb activities at E11.5 (DE1, 4, 6, 9, 15) or non-validated elements with

346 proximal limb-specific H3K27ac enrichment (+237kb and +568kb) (**Fig. 6B, C**). Open chromatin
347 and H3K27ac profiles further indicated that the *Shox2*-interacting DE4 (+407) element was unique
348 in its H3K27ac pattern (initiated at E11.5), while other proximal limb (candidate) enhancers showed
349 activity marks already at E10.5. Therefore, we decided to analyze the spatiotemporal enhancer
350 dynamics of newly identified (+237kb, +568kb) and temporally dynamic (+407) limb candidate
351 enhancer regions using stable transgenic LacZ reporter mouse lines. For comparison, we also
352 assessed the previously validated hs741 and hs1262 *Shox2* limb enhancers[19,47] (**Fig. 6C, S6A and**
353 **Table S11**). Remarkably, at E12.5, each element on its own was able to drive reporter expression in
354 the proximal fore- and hindlimb mesenchyme in a pattern overlapping *Shox2*, establishing a
355 complement of at least five proximal limb enhancers (PLEs) that contact *Shox2*, four of which reside
356 within the gene desert (PLE2-PLE5) (**Fig. 6C, S6A**). These activity patterns generally showed strong
357 reporter signal in the peripheral mesenchyme of the stylopod and zeugopod elements (**Fig. 6C, S6A**).
358 *Shox2* expression is progressively downregulated within the differentiating chondrocytes of the
359 proximal skeletal condensations of the limbs from E11.5, while its expression remains high in the
360 surrounding mesenchyme and perichondrium[49,69–71]. In accordance, activities of the newly
361 discovered elements (PLE3-5) remained excluded from the chondrogenic cores of the skeletal
362 condensations, consistent with a role in generating the *Shox2* expression pattern required for
363 stylopodial chondrocyte maturation and subsequent osteogenesis[11,27]. PLE3 (+237) was initiated in
364 the proximal limb mesenchyme at E11.5 with persistent signal until E13.5 and most closely
365 recapitulating the late *Shox2* expression pattern[27,49] (**Fig. S6A**). Instead, PLE4 (+407) drove reporter
366 activity already at E10.5 in a more widespread pattern leaking into distal forelimbs at later stages, in
367 line with H3K27ac enrichment in distal forelimbs (**Figs. 6A, S6A**). PLE5 (+568) was initiated only
368 at E12.5 and its activity remained restricted to the proximal-anterior (**Fig. S6A**). Together, these
369 diverse and partially overlapping enhancer activities pointed to dynamic interaction of *Shox2* gene
370 desert enhancers during limb development. In addition, to achieve insight into PLE configuration at
371 the chromatin level we performed 4C-seq with viewpoints at PLE2 and PLE4 which indicated the

372 formation of a complex involving PLE1, 3 and 4, but not PLE2 (**Fig. S6B-D**). These findings indicate
373 that PLE interactions might not necessarily be restricted to U-dom or D-dom sub-compartments for
374 *Shox2* regulation in the limb. Taken together, these results identify the gene desert as a multipartite
375 *Shox2* limb enhancer unit and indicate an instructive role in the transcriptional control of stylopod
376 morphogenesis.

377 To evaluate the functional and phenotypic contribution of the gene desert to stylopod
378 formation we combined our gene desert deletion allele with a *Prx1*-Cre conditional approach for
379 *Shox2* inactivation[27,72]. This enabled limb-specific conditional *Shox2* inactivation on one allele
380 ($Shox2^{\Delta c}$) paired with gene desert deletion on the other allele ($GD^{\Delta}$), allowing to bypass embryonic
381 lethality caused by loss of cardiac *Shox2* (**Figs. 6D, S4**). Remarkably, this abolishment of gene desert-
382 mediated *Shox2* regulation in limbs led to severe shortening of the stylopod with an approximate
383 60% reduction in humerus length and 80% decrease in femur extension in $GD^{\Delta}/Shox2^{\Delta c}$ newborn
384 mice (**Fig. S7A, B**). These skeletal abnormalities were in line with the stylopod phenotypes obtained
385 by limb-specific *Shox2* dosage reductions in previous studies[19,27]. Concordantly, micro-computed
386 tomography (μCT) from adult mouse limbs at P42 showed significant humerus length reduction of
387 approximately 40% and decreased femur length of about 50% (**Fig. 6E**). Our results thus demonstrate
388 an essential role of the gene desert in proximal limb morphogenesis and imply a significant functional
389 contribution of the PLE2-5 modules to spatiotemporal control of *Shox2* dosage in the limb.

390 In summary, our study identifies the *Shox2* gene desert as an essential and dynamic chromatin
391 unit that encodes an array of distributed tissue-specific enhancers that coordinately regulate stylopod
392 formation, craniofacial patterning, and SAN pacemaker dependent embryonic progression (**Fig. 7A-**
393 **C**). The arrangement of the enhancers appears modular but distributed in terms of tissue-specificities
394 (**Fig. 7A**). While craniofacial and neuronal gene desert enhancers are hallmarked by driving mostly
395 distinct subregional activities, limb enhancers (PLEs) show more overlapping activity domains,
396 pointing to potential redundant intra-gene desert interactions. Hereby, the detection of a high-density
397 contact domain (HCD) suggests that sub-TAD compartmentalization could further contribute to

398 modulation of subregional enhancer activities (**Fig. 7B**). Finally, the demonstrated phenotypic
399 requirement of the *Shox2* gene desert for multiple developmental processes underscores the
400 importance of functional studies focused on the non-coding genome for better mechanistic
401 understanding of congenital abnormalities (**Fig. 7C**).

402 **DISCUSSION**

403 There is now evidence that dismantling of duplicates of ancient genomic regulatory blocks (GRBs)
404 led to the emergence of gene deserts enriched in the neighborhood of regulatory genes such as TFs[73].
405 Functional assessment of TF gene deserts, including those in the *Hoxd* and *Sox9* loci, revealed that
406 distal long-range enhancers represent critical *cis*-regulatory modules that control subregional
407 expression domains through interaction with target gene promoters in a spatiotemporal
408 manner[6,7,74,75]. Gene deserts can thus be conceived as genomic units coordinating dynamic enhancer
409 activities in specific developmental processes, such as *HoxD*-dependent digit formation, and can be
410 also hi-jacked by evolutionary processes to enable phenotypic diversification[9,11,76]. Silencer modules,
411 insulating TAD boundaries and tethering elements (promoting long-range interactions) are involved
412 in restriction or modulation of E-P interactions in metazoan genomes and can further contribute to
413 gene desert functionality[77–79]. Recent studies also indicated that functional RNAs, such as lncRNAs
414 or circRNAs, represent elements with enhancer-modifying or distinct regulatory potential within
415 gene deserts[80]. Importantly, human disease-associated nucleotide variants in gene deserts are
416 frequently linked to enhancer function, contributing to the spectrum of enhanceropathies[81–83].
417 Furthermore, deletions, inversions and duplications can alter or re-distribute interaction of gene
418 desert enhancers with target gene promoters leading to congenital malformation or syndromes[10,17,18].
419 Despite these critical implications, the enhancer landscapes and related chromatin topology of most
420 gene deserts near developmental genes remain incompletely characterized at the functional level[84].
421 In the current study, we addressed the functional necessity and *cis*-regulatory architecture of a gene
422 desert flanking the *Shox2* transcriptional regulator, a critical determinant of embryogenesis and
423 essential for limb, craniofacial and SAN pacemaker morphogenesis[37,47,85]. We identify the *Shox2*
424 gene desert as a reservoir for highly subregional, tissue-specific enhancers underlying pleiotropic
425 *Shox2* dosage by demonstrating essential contributions to stylopod morphogenesis, craniofacial
426 patterning, and SV/SAN development. Our findings support a model in which gene deserts provide

427 a scaffold for preferential chromatin domains that generate enhancer-mediated cell type or tissue-
428 specific *cis*-regulatory output based on the integration of upstream signals.

429 Interpretation of gene desert function is dependent on accurate functional predictions of
430 enhancer activities embedded in the genomic interval. Our approach using ChromHMM-filtered
431 H3K27ac signatures from bulk tissues across a large range of embryonic stages (derived from
432 ENCODE) serves as a baseline for the mapping of tissue-specific enhancer activities. However, while
433 H3K27ac is known as the most specific canonical mark for active enhancers, it appears to not include
434 all enhancers[86–88]. For example, recent studies evaluating H3K27ac-based tissue-specific enhancer
435 predictions in mouse embryos revealed a substantial number of false-positives[54,89]. In turn, a
436 significant fraction (~14%) of validated *in vivo* enhancers were lacking enrichment of any canonical
437 enhancer marks (ATAC-seq, H3K4me1, H3K27ac)[89]. In line with these observations, our transgenic
438 reporter validation in many cases revealed more restricted or even distinct *in vivo* enhancer activities
439 than those predicted by epigenomic marks. Such discrepancies might be partially originating from
440 the use of bulk tissues or limited sensitivity of profiling techniques. In accordance, refinement of
441 enhancer predictions using region-specific open chromatin data in combination with chromatin
442 conformation capture (C-HiC, 4C-seq) enabled us to identify critical subregional cardiac and
443 proximal limb enhancers missed by the initial epigenomic prediction approach.

444 Genomic deletion analysis uncovered an important functional role of the gene desert for
445 pleiotropic expression and progression of embryonic development, the latter through direct control
446 of *Shox2* in SAN pacemaker progenitors. Consistent with our findings, a parallel study narrowed the
447 region essentially required for cardiac *Shox2* expression to a 241kb gene desert interval (termed VS-
448 250)[63]. Here, we have identified a human-conserved SV enhancer (+325) located within this essential
449 interval and specifically active in the SV/SAN region to maintain robust cardiac *Shox2* levels. These
450 results add to recent progress in uncovering SAN enhancers of cardiac pacemaker regulators,
451 including also *Isl1* or *Tbx3*[44,63]. Such findings not only shed light on the wiring of the GRNs driving
452 mammalian conduction system development but also offer the opportunity to identify mutational

453 targets linked to defects in the pacemaker system, such as arrhythmias[62]. Interestingly, removal of
454 the *Isl1* SAN enhancer (ISE) in mice, as for our +325 *Shox2* enhancer, led to reduced target gene
455 dosage but without subsequent embryonic or perinatal lethality[44]. These instances indicate that the
456 GRNs orchestrating SAN pacemaker development are buffered at the *cis*-regulatory level, which can
457 be enabled via partially redundant enhancer landscapes[19,90]. Similar to the binding profile of the +325
458 *Shox2* SV enhancer, a TF network involving Gata4, Tbx5 and Tead has been implicated in ISE
459 activation, confirming a key role of Tbx5 in the activation of SAN enhancers in working atrial
460 myocardium, while pacemaker-restricted identity may be established by repressive
461 mechanisms[43,44,62]. ISE activity was also correlated with abnormal SAN function in adult mice and
462 found to co-localize with resting heart rate SNPs, indicating potentially more sensitive GRN
463 architecture in humans[44]. Intriguingly, coding and non-coding variants in the human *SHOX2* locus
464 were recently associated with SAN dysfunction and atrial fibrillation, underscoring the value of
465 human-conserved SAN enhancer characterization for functional disease variant screening[38,40,91,92].

466 Arrangements of distributed enhancer landscapes conferring robust and cell type-specific
467 transcription emerged as a common feature of metazoan gene regulatory architecture[93–95]. Gene
468 deserts may thus not only function to promote robust expression boundaries and/or phenotypic
469 resilience, but also represent a platform enabling evolutionary plasticity[9,73]. The conventional model
470 of enhancer additivity based on individual small and stable regulatory contributions is likely
471 predominant in gene deserts[96]. In support, we uncovered at least four *Shox2*-associated gene desert
472 enhancers (PLE2-5) with overlapping activities in the proximal limb mesenchyme. Such regulatory
473 architecture resembles the multipartite enhancer landscapes in *Indian Hedgehog* (*Ihh*) or *Gremlin1*
474 loci, which as *Shox2* are involved in spatiotemporal coordination of proximal-distal limb identities
475 with chondrogenic cues[22,24]. Our study further reveals gene desert enhancers with seemingly unique
476 tissue specificities, such as the craniofacial DE9 and DE15 elements driving *Shox2*-overlapping
477 reporter expression in the nasal process and maxillary-mandibular region, respectively. DE15 may

478 be involved in jaw formation as *Shox2* inactivation in cranial neural crest cells in the maxilla-
479 mandibular junction leads to dysplasia and ankylosis of the TMJ in mice[28].

480 Our C-HiC experiments indicated that the repertoire of *Shox2* interacting elements (e.g.,
481 enhancers) is confined to the overarching TAD, without apparent cross-TAD boundary interactions[97].
482 The observed U-dom and D-dom assemblies (as evidenced by loop anchors) might reflect dynamic
483 loop structures to facilitate *Shox2* promoter scanning similar to the organization at *HoxA* and *HoxD*
484 loci that promotes nested and collinear gene expression[7,98]. C-HiC analysis also uncovered a high-
485 density contact domain (HCD) emerging only in heart tissue. The absence of convergent CTCF sites
486 flanking the HCD might reflect that a subset of contact domains form independently of cohesin-
487 mediated loop extrusion, for example based on self-aggregation of regions carrying identical
488 epigenetic marks or the emergence of globule structures resulting from phase separation[99–102].
489 Interestingly, the HCD genomic interval harbors several validated enhancers that were inactive in the
490 embryonic heart (DE5-12). An intriguing hypothesis raised by these observations is therefore that
491 HCDs could act to topologically sequester regulatory regions for modulation of target gene
492 interaction in a tissue-specific manner.

493 From a disease perspective, our findings also expand on former analyses demonstrating that
494 *Shox2* gene desert limb and hindbrain enhancer activities emerge within the similar-sized gene desert
495 flanking the human *SHOX*[47,103]. Pointing to functional homology with the mouse *Shox2* regulatory
496 region, disruption of enhancers within the gene desert downstream of *SHOX* has been associated with
497 Léri-Weill dyschondrosteosis (LWD) and idiopathic short stature (ISS) syndromes in a significant
498 fraction of cases[104]. Furthermore, *SHOX* haploinsufficiency is directly associated with the skeletal
499 abnormalities observed in LWD and Turner syndrome, the latter also involving craniofacial
500 abnormalities[105–107]. One study has also found a link between neurodevelopmental disorders and
501 microduplications at the *SHOX* locus, suggesting that such perturbations may alter neural
502 development or function[108]. Thus, considering the overlapping expression patterns and critical
503 functions of human *SHOX* and mouse *Shox2*, our results provide a blueprint for the investigation of

504 *SHOX* regulation in the hindbrain, thalamus, pharyngeal arches, and limbs[109,110]. It will be
505 particularly interesting to determine whether "orthologous" craniofacial, neural and/or limb
506 enhancers exist, and whether human *SHOX* enhancers share motif content or other enhancer grammar
507 characteristics with mouse *Shox2* enhancers. Indeed, in a recent example orthologous enhancer-like
508 sequence was identified 160kb downstream of human *SHOX* and 47kb downstream of mouse *Shox2*,
509 respectively, and drove overlapping activities in the hindbrain[47,103]. Such enhancers presumably
510 originate from a single ancestral *Shox* locus, preceding the duplication of *Shox* and *Shox2* paralogs
511 and are therefore considered evolutionary ancient. Within this context, future comparative studies
512 should include a search for deeply conserved orthologs of *SHOX* and *SHOX2* enhancers in basal
513 chordates such as amphioxus, which express their single *Shox2* gene in the developing hindbrain[111].
514 The recent identification of orthologous *Islet* gene enhancers in sponges and vertebrates demonstrate
515 the promise of such an approach[112]. Taken together, functional enhancer characterization along with
516 refined enhancer grammar and 3D interactions at the cell type level will likely be key to resolve the
517 regulatory complexity inherent to distributed enhancer landscapes and to understand how
518 transcriptional dynamics and morphological complexity are rooted in gene deserts.

519 # MATERIALS AND METHODS

520 **Ethics statement and approval of animal experimentation.**

521 All animal (mouse) experiments were performed in accordance with national laws and approved by
522 the national and local regulatory authorities. All animal work at Lawrence Berkeley National
523 Laboratory (LBNL, CA, USA) was reviewed and approved by the LBNL Animal Welfare
524 Committee. Knockout and transgenic mice were housed at the Animal Care Facility (the ACF) at
525 LBNL. Mice were monitored daily for food and water intake, and animals were inspected weekly by
526 the Chair of the Animal Welfare and Research Committee and the head of the animal facility in
527 consultation with the veterinary staff. The LBNL ACF is accredited by the American Association for
528 the Accreditation of Laboratory Animal Care International (AAALAC). Animal work at the
529 University of Calgary involving the production, housing and analysis of mouse lines depicted in
530 **Figures 6, S6 and S7**, was approved by the Life and Environmental Sciences Animal Care
531 Committee (LESACC). All experiments with mice were performed in accordance with Canadian
532 Council on Animal Care guidelines as approved by the University of Calgary LESACC, Protocol #
533 AC13-0053. Animal work in Switzerland involving SV enhancer KO mice (**Figs. 5 and S5**) was
534 approved by the regional commission on Animal Experimentation and the Cantonal Veterinary
535 Office of the city of Bern. The following developmental stages were used in this study: embryonic
536 day E10.5, E11.5, E12.5, E13.5 and newborn mice. Animals of both sexes were used in these
537 analyses. Sample size selection and randomization strategies were conducted as follows:

538 ***Transgenic mouse assays.*** Sample sizes were selected empirically based on our previous experience
539 of performing transgenic mouse assays for >3,000 total putative enhancers (VISTA Enhancer
540 Browser: https://enhancer.lbl.gov/). Mouse embryos were excluded from further analysis if they did
541 not encode the reporter transgene or if the developmental stage was not correct. All transgenic mice
542 were treated with identical experimental conditions. Randomization and experimenter blinding were
543 unnecessary and not performed.

544 ***Knockout mice.*** Sample sizes were selected empirically based on our previous studies[19,20]. All
545 phenotypic characterization of knockout mice employed a matched littermate selection strategy.
546 Analyzed *Shox2* gene desert and enhancer knockout embryos and mice described in this paper
547 resulted from crossing mice heterozygous for the respective deletion to allow for the comparison of
548 matched littermates of different genotypes. Embryonic littermates and samples from genetically
549 modified animals were dissected and processed blind to genotype.

550 **Transgenic reporter analysis in mouse embryos**

551 Transgenic reporter assays for validation of all elements except PLEs were performed at LBNL in
552 *Mus musculus* FVB strain mice and injection of *LacZ* reporter constructs was conducted as previously
553 described[19,113]. The related primer sequences and genomic coordinates are listed in **Tables S2 and**
554 **S8**. Predicted enhancer elements were PCR-amplified from mouse genomic DNA (Clontech) and
555 cloned into a Hsp68-*LacZ* expression vector for random integration[114]. For higher accuracy in
556 absence of position effects, the +325 SV enhancer element was analyzed in a *β*-globin-LacZ construct
557 for site-directed integration at the neutral *H11* locus[25,115]. PLE elements were PCR-amplified from
558 bacterial artificial chromosomes (**Table S11**) and then cloned into the *βlacz* plasmid containing a
559 minimal human *β*-globin promoter-*LacZ* cassette, as described[47]. Due to their large size, PLE3
560 (10,351 bp) and PLE5 (9,473 bp) were amplified with the proofreading polymerase in the
561 SequalPrep™ Long PCR Kit (Invitrogen). PLE transgenic mice and embryos were produced at the
562 University of Calgary Centre for Mouse Genomics by pronuclear injection of DNA constructs into
563 CD-1 single-cell stage embryos as described[116]. Male founder animals (or male F1 progeny produced
564 from transgenic females) were crossed to CD-1 females to produce transgenic embryos which were
565 stained with X-gal by standard techniques[113].

566 **CRISPR/Cas9 deletion mouse lines**

567 SgRNAs located 5' and 3' of the genomic sequence of interest were designed using CHOPCHOP[117]
568 (*Shox2* gene desert deletion) or CRISPOR[118] (http://crispor.tefor.net/) (SV enhancer deletion). Gene
569 desert deletion (582kb) was engineered using CRISPR/Cas9 genome editing in fertilized mouse

570 oocytes as described, with minor modifications[114]. Briefly, a mix containing Cas9 mRNA (final
571 concentration of 100 ng/ul) and two single guide RNAs (sgRNAs) (25 ng/ul each) was microinjected
572 into the cytoplasm of fertilized FVB strain oocytes. The SV enhancer deletion allele was engineered
573 using CRISPR-EZ[68] at the Center of Transgenic Models (CTM) of the University of Basel. Genomic
574 deletion coordinates and sgRNA sequences used for genome editing are listed in **Table S4**. Deletion
575 alleles in Founder (F0) mice and F1 offspring were identified using PCR and the exact deletion
576 breakpoints were verified by Sanger sequencing. Mice and embryos were PCR-genotyped using
577 primer pairs specific for the deleted genomic interval or wild-type counterpart (**Table S5**).

578 **ENCODE H3K27ac ChIP-seq and mRNA-seq analysis**

579 To establish a heatmap revealing putative enhancers and their temporal activities within the *Shox2*
580 TAD interval, a previously generated catalog of strong enhancers identified using ChromHMM[53]
581 across mouse development was used[54]. Briefly, calls across 66 different tissue-stage combinations
582 were merged and H3K27ac signals quantified as log2-transformed RPKM. Estimates of statistical
583 significance for these signals were associated to each region for each tissue-stage combination using
584 the corresponding H3K27ac ChIP-seq peak calls. These were downloaded from the ENCODE Data
585 Coordination Center (DCC) (http://www.encodeproject.org/, see **Table S1**, *sheet 3* for the complete
586 list of sample identifiers). To this purpose, short reads were aligned to the mm10 assembly of the
587 mouse genome using bowtie (ref), with the following parameters: *-a -m 1 -n 2 -l 32 -e 3001*. Peak
588 calling was performed using MACS v1.4[119], with the following arguments: *--gsize=mm --bw=300 -*
589 *-nomodel --shiftsize=100*. Experiment-matched input DNA was used as control. Evidence from two
590 biological replicates was combined using IDR (https://www.encodeproject.org/data-
591 standards/terms/). The *q*-value provided in the replicated peak calls was used to annotate each
592 putative enhancer region defined above. In case of regions overlapping more than one peak, the
593 lowest *q*-value was used. RNA-seq raw data was downloaded from the ENCODE DCC
594 (http://www.encodeproject.org/, see **Table S1**, *sheet 3* for the complete list of sample identifiers).

595 **Region Capture Hi-C (CHi-C)**

596 Embryonic forelimbs, mandibular processes, and hearts from wildtype FVB embryos at E11.5 were
597 micro-dissected in cold 1xPBS, pooled according to tissue type, and homogenized using a Dounce
598 tissue grinder. Cells were resuspended in 10% FCS (in PBS) and formaldehyde (37%) diluted to a
599 final 2% in a total volume of 1ml was added for fixation for 10 min, as previously described[60]. 1.25M
600 Glycine was used to quench fixation and pellets were snap-frozen in liquid nitrogen and stored at -
601 80C. Pellets were resuspended in fresh lysis buffer (10mM Tris, pH7.5, 10mM NaCl, 5mM MgCl2,
602 0.1 mM EGTA complemented with Protease Inhibitor) for nuclei isolation. Following 10min
603 incubation on ice, samples were washed with 1xPBS and frozen in liquid nitrogen. 3C-libraries were
604 prepared from thawed nuclei subjected to DpnII digestion (NEB, R0543M), re-ligated with T4 ligase
605 (Thermo Fisher Scientific) and de-crosslinking as described previously[60]. For 3C-library quality
606 control, 500ng of library sample along with digested and undigested control samples was assessed
607 using agarose gel electrophoresis (1% gel). Shearing on re-ligated products was performed using a
608 Covaris ultrasonicator (duty cycle: 10%, intensity 5, cycles per burst: 200, time: 2 cycles of 60s each).
609 Following adaptor ligation and amplification of sheared DNA fragments, libraries were hybridized
610 to custom-designed SureSelect beads (SureSelectXT Custom 0.5-2.9Mb library) and indexed
611 following Agilent's instructions. Multiplexed libraries were sequenced using 50bp paired-end
612 sequencing (HiSeq 4000 sequencer). CHi-C probes of the SureSelect library were designed to span
613 the *Shox2* genomic interval and adjacent TADs (mm10: chr3:65196079-68696078).

614 **CHi-C data processing and analysis**

615 CHi-C processing was performed using a previously published pipeline[26]. Briefly, sequenced reads
616 were mapped to the reference genome GRCm38/mm10 following the HiCUP pipeline[120] (v0.8.1) set
617 up with Bowtie2[121] (v2.4.5). Filtering and de-duplication was conducted using HiCUP (no size
618 selection, Nofill: 1, format: Sanger) and unique MAPQ ≥ 30 valid read pairs were obtained for FL,
619 MD and HT datasets (N=637163, N=577862 and N=592498, respectively). Binned contact maps
620 from valid read pairs were generated using Juicer command line tools[122] (v1.9.9) and raw .cool files

621 were generated with the hicConvertFormat tool (HiCExplorer v3.7.2) from native .hic out-puts
622 generated by Juicer. For normalization and diagonal filtering the Cooler matrix balancing tool[123]
623 (v0.8.11) was applied with the options '--mad-max 5 --min-nnz 10 --min-count 0 --ignore-diags 2 --
624 tol 1e-05 --max-iters 200 --cis-only'. Only the targeted genomic interval enriched in the capture step
625 (mm10: chr3:65196079-68696078) was selected for binning and balancing. Consequently, only read
626 pairs mapping to this interval were retained, shifted by the offset of 65,196,078 bp using custom
627 crhom.sizes files. Balanced maps were then exported at 5kb resolution with corrected coordinates
628 (transformed back to original values). Subtraction maps were directly generated from Cooler
629 balanced HiCmaps using hicCompareMatrices tool (HiCExplorer v3.7.2) with option '--operation
630 diff'. HiCExplorer[124] (v3.7.2) was used to determine normalized inter-domain insulation scores and
631 domain boundaries on Hi-C and subtraction maps using default parameters 'hicFindTADs -t 0.05 -d
632 0.01 -c fdr' computing p-values for minimal window length 50000. Hi-C maps and related graphs
633 were visualized from .cool files and bedgraph matrices, respectively, using pyGenomeTracks[125]
634 (*v.3.6*). GOTHiC[126] (v.1.32.0) was used to identify reliable and significance-based Hi-C interactions
635 from HiCUP validated read pairs (MAPQ10) with 'res=1000, restrictionFile, cistrans='all',
636 parallel=FALSE, cores=NULL' (R pipeline-template script, v.4.2.2) and a threshold of '-log(q-
637 value) > 1'.

638 **Virtual 4C (V4C)**

639 To determine target interactions of a defined element locally V4C profiles were generated as
640 described[60] from filtered unique read pairs (hicup.bam files) which also served as input for
641 computation of CHi-C maps (see above). Conditions for mapped read-pairs included MAPQ ≥ 30
642 and relative position of the two reads inside and outside the viewpoint, respectively. After
643 quantitation of reads outside of the viewpoint (per restriction fragment), read counts were distributed
644 into 3kb bins (with proportional distribution of read counts in case of overlap with more than one
645 bin). Following smoothing of each binned profile via averaging[60], peak profiles were generated using
646 custom Java code based on htsjdk v2.12.0 (https://samtools.github.io/htsjdk/). A 10kb viewpoint

647 containing the extended *Shox2* promoter region (chr3:66975788-66985788) was used for comparison
648 with Hi-C maps. The viewpoint and neighboring +/-5kb regions were excluded from computation of
649 the scaling factor. BigwigCompare tool (deepTools v3.5.1) was used to generate relative subtraction
650 Capture-C-like profiles.

651 **4C-seq from proximal forelimbs**

652 Per replicate, 10-12 proximal forelimbs from CD-1 embryos at E12.5 were dissected in PBS,
653 followed by 4C-seq tissue processing as described[127,128]. For tissue preparation, cells were dissociated
654 by incubating the pooled tissue in 250µl PBS supplemented with 10% fetal fetal calf serum (FCS)
655 and 1 mg/ml collagenase (Sigma) for 45 minutes at 37° C with shaking at 750 rpm. The solution was
656 passed through a cell strainer (Falcon) to obtain single cells which were fixed in 9.8 ml of 2%
657 formaldehyde in PBS/10% FCS for 10 minutes at room temperature and lysed. Libraries were
658 prepared by overnight digestion with NlaIII (New England Biolabs (NEB)) and ligation for 4.5 hours
659 with 100 units T4 DNA ligase (Promega, #M1794) under diluted conditions (7 ml), followed by de-
660 crosslinking overnight at 65°C after addition of 15ul of 20mg/ml proteinase K. After
661 phenol/chloroform extraction and ethanol precipitation the samples were digested overnight with the
662 secondary enzyme DpnII (NEB) followed by phenol/chloroform extraction, ethanol precipitation
663 purification and ligation for 4.5 hours in a 14 ml volume. The final ligation products were extracted
664 and precipitated as above followed by purification using Qiagen nucleotide removal columns. For
665 each viewpoint, libraries were prepared with 100 ng of template in each of 16 separate PCR reactions
666 using the Roche, Expand Long Template kit with primers incorporating Illumina adapters. Viewpoint
667 and primer details are presented in **Table S10**. PCR reactions for each viewpoint were pooled and
668 purified with the Qiagen PCR purification kit and sequenced with the Illumina HiSeq to generate
669 single 100bp reads. Demultiplexed reads were mapped and analyzed with the 4C-seq module of the
670 HTSstation pipeline as described[129]. Results are shown in UCSC browser format as normalized reads
671 per fragment after smoothing with an 11-fragment window and mapped to mm10 (**Fig. 6B**, **S6D**).

672 Raw and processed (bedgraph) sequence files are available under GEO accession number
673 GSE161194.

674 ### *In situ* hybridization (ISH) and quantitative real-time PCR (qPCR)

675 For assessment of spatial gene expression changes in mouse embryos, whole mount *in situ*
676 hybridization (ISH) using digoxigenin-labeled antisense riboprobes was performed as previously
677 described[130]. At least three independent embryos were analyzed for each genotype. Embryonic
678 tissues were imaged using a Leica MZ16 microscope coupled to a Leica DFC420 digital camera.
679 Brightness and contrast were adjusted uniformly using Photoshop CS5. For qPCR of samples
680 involving the gene desert deletion ($GD^{\Delta}$), isolation of RNA from micro-dissected embryonic tissues
681 was performed using the Ambion RNAqueous Total RNA Isolation Kit (Life Technologies)
682 according to the manufacturer's protocol. For qPCR of samples involving the SV enhancer deletion
683 ($SV\text{-}Enh^{\Delta}$), RNeasy Micro Kit (Qiagen) was used following the manufacturer's protocol. RNA was
684 reverse transcribed using SuperScript III (Life Technologies) with poly-dT priming according to
685 manufacturer instructions. qPCR was conducted on a LightCycler 480 (Roche) using KAPA SYBR
686 FAST qPCR Master Mix (Kapa Biosystems) for $GD^{\Delta}$ samples, and on a and on a ViiA 7 Real-Time
687 PCR System using PowerTrack SYBR Green Master Mix (Applied Biosystems) for $SV\text{-}Enh^{\Delta}$
688 samples, according to manufacturer instructions. The qPCR primers used (**Table S6**) were described
689 previously[19]. Relative gene expression levels were calculated using the $2^{-\Delta\Delta C_T}$ method, normalized to
690 the *Actb* housekeeping gene, and the mean of wild-type control samples was set to 1, as used
691 previously[19].

692 ### Immunofluorescence (IF)

693 IF was performed as previously described[19]. Briefly, mouse embryos at E11.5 were isolated in cold
694 PBS and fixed in 4% PFA for 2–3h. After incubation in a sucrose gradient and embedding in a 1:1
695 mixture of 30% sucrose and OCT compound, sagittal 10µm frozen tissue sections were obtained
696 using a cryostat. Selected cryo-sections were then incubated overnight with the following primary
697 antibodies: anti-Shox2 (1:300, Santa Cruz JK-6E, sc-81955), anti-SMA-Cy3 (1:250, Sigma, C6198),

698 anti-Hcn4 (1:500, Thermo Fisher, MA3-903) and anti-Nkx2.5 (1:500, Thermo Fisher, PA5-81452).
699 Goat-anti mouse, goat anti-rabbit and donkey anti-rat secondary antibodies conjugated to Alexa Fluor
700 488, 568, or 647 (1:1,000, Thermo Fisher Scientific) were used for detection. Hoechst 33258 (Sigma-
701 Aldrich) was utilized to counterstain nuclei. A Zeiss AxioImager fluorescence microscope in
702 combination with a Hamamatsu Orca-03 camera was used to acquire fluorescent images. Brightness
703 and contrast were adjusted uniformly using Photoshop CS5.

704 **Skeletal preparations**

705 Euthanized newborn mice were eviscerated, skinned and fixed in 1 % acetic acid in EtOH for 24
706 hours. Cartilage was stained overnight with 1 mg/mL Alcian blue 8GX (Sigma) in 20% acetic acid
707 in EtOH. After washing in EtOH for 12 hours and treatment with 1.5 % KOH for three hours, bones
708 were stained in 0.15 mg/mL Alizarin Red S (Sigma) in 0.5 % KOH for four hours, followed by de-
709 staining in 20 % glycerol, 0.5 % KOH.

710 **X-ray micro-computed tomography (µCT) of adult mouse skeletons**

711 Mice were euthanized at 6 weeks of age and whole-body µCT scans were generated using a Skyscan
712 1173 v1.6 µCT scanner (Bruker, Kontich, Belgium) at 80-85 kV and 56-62 µA with 45 µm
713 resolution[131] NRecon v1.7.4.2 (Bruker, Kontich, Belgium) was used to perform stack reconstructions,
714 and 3D landmarks were placed in MeshLab[132] (v2020.07) by one observer (CSS) blind to the
715 genotype identity of individual animals. To quantify the length of the stylopod bones, distances were
716 calculated between two landmarks placed at the proximal and distal ends of the humerus and femur
717 (the proximal epiphysis [PE] and olecranon fossa lateral [OFL] for the humerus, and the greater
718 trochanter [GT] and lateral inferior condyle [LIC] for the femur). To account for body size variability
719 between individuals, these measurements were normalized to the inter-landmark distance between
720 the proximal and distal ends of the third metatarsal. To assess intra-observer repeatability, CSS placed
721 the landmarks on scans of 12 mice (six GD$^{\Delta}$/*Shox2*$^{\Delta c}$, two GD$^{\Delta/+}$, two *Shox2*$^{\Delta c/+}$, and two WT) five
722 times each, with each session separated by at least 24 hours[133]. An absolute coefficient of variation

723 (CV) for each landmark was calculated and the average CV was 0.28% with a range of 0.14% -
724 0.42%.

725 **ATAC-seq**

726 ATAC-seq was performed as described[134] with minor modifications. Per replicate, pairs of wildtype
727 mouse embryonic hearts at E11.5 were micro-dissected in cold PBS and cell nuclei were dissociated
728 in Lysis buffer using a Dounce tissue grinder. Approx. 50'000 nuclei were then pelleted at 500 RCF
729 for 10 min at 4°C and resuspended in 50 µL transposition reaction mix containing 25 µL Nextera 2x
730 TD buffer and 2.5 µL TDE1 (Nextera Tn5 Transposase; Illumina) (cat. no. FC-121-1030) followed
731 by incubation for 30 minutes at 37°C with shaking. The reaction was purified using the Qiagen
732 MinElute PCR purification kit and amplified using defined PCR primers[135]. ATAC-seq libraries were
733 purified using the Qiagen MinElute PCR purification kit (ID: 28004), quantified by the Qubit
734 Fluorometer with the dsDNA HS Assay Kit (Life Technologies) and quality assessed using the
735 Agilent Bioanalyzer high sensitivity DNA analysis assay. Libraries were pooled and sequenced using
736 single end 50 bp reads on a HiSeq 4000 (Illumina).

737 **Mouse ATAC-seq and ChIP-seq data processing**

738 Analysis of mouse ATAC-seq and reprocessing of previously published ChIP-seq
739 datasets used in this study was performed using Adaptor trimming (trim_galore_v0.6.6) by
740 Cutadapt (https://cutadapt.readthedocs.io/), with default parameters '-j 1 -e 0.1 -q 20 -O 1' for single-
741 end, and '--paired -j 1 -e 0.1 -q 20 -O 1' for paired- end data (purging trimmed reads shorter than
742 20bp). For read mapping, Bowtie2[121] (version 2.4.2) was used with parameters '-q --no-unal -p 8
743 -X2000' (ATAC-seq) and '-q --no-unal -p 2' (ChIP-seq) for both single/paired-end samples.
744 Reads were aligned to the GRCm38/mm10 reference genome using pre-built Bowtie2 indexes from
745 the Illumina's iGenomes collection (http://bowtie-bio.sourceforge.net/bowtie2/). Duplicates and
746 low-quality reads (MAPQ = 255) for both single/paired-end samples were removed using
747 SAMtools (v1.12), with pipeline parameters 'markdup -r' and '-bh -q10', respectively[136]. ATAC-
748 seq peak calling was performed using MACSv2[119,137] (v2.1.0) with p-value < 0.01 and parameters '-

749 t -n -f BAM -g mm --nolambda --nomodel --shif 50 --extsize 100' for single-end, and '-t -n -f BAMPE
750 -g mm --nolambda --nomodel --shif 50 --extsize 100' for paired-end reads. For ChIP-seq peak calling,
751 '-t -c -n -f BAM -g mm' parameters were used instead. PyGenomeTracks[125] was used for
752 visualization of profiles and alignment with other datasets.

753 **Cardiac TF motif detection**

754 An enriched collection of position weight matrices (PWMs)[138] was limited to motifs of TFs expressed
755 in the developing heart at E11.5. After mapping of gene symbols to the equivalent identifiers in the
756 Ensembl103 release using the BiomaRt v2.5.0 package (R v4.1.2)[139], only those PWMs matching
757 TFs expressed in E11.5 hearts were selected for analysis[140] (ENCSR691OPQ). A mean FPKM ≥ 2
758 calculated across all RNA-seq replicates was used as threshold for significant expression. This
759 filtering resulted in a set of 576 mouse TFs. 1'376 corresponding PWMs were available for 282 of
760 these TFs[67] which were used for motif detection by FIMO (Find Individual Motif Occurrences)[67,141],
761 except for 14 that were omitted since in each case, since the match identified genome-wide was
762 included in a larger motif within the collection (**Table S9**). FIMO v5.3.0 with a standard p-value
763 cutoff of $10^{-4}$ and GC-content matched backgrounds were used for screening genomic sequence for
764 potential TF-binding sites. Motif conservation was computed using BWTOOL v1.0[142] based on the
765 average of individual nucleotide PhyloP (Placental) conservation scores provided by UCSC PHAST
766 package (http://hgdownload.cse.ucsc.edu/goldenpath/mm10/phyloP60way/).

767 **ChIP-seq and RNA-seq from human fetal hearts**

768 Aspects of this study involving human heart tissue samples were reviewed and approved by the
769 Human Subjects Committee at Lawrence Berkeley National Laboratory (Protocol Nos. 00023126
770 and 00022756). Fetal human heart embryonic samples at post conception week 17 (LV, RV, LA,
771 RA) were obtained from the Human Developmental Biology Resource at Newcastle University
772 (hdbr.org), in compliance with applicable state and federal laws and with fully informed consent.
773 Fetal samples were transported on dry ice and stored at -80C, as reported for the LV sample generated
774 and analyzed in a previous publication[143]. Fetal human RV, LA and RA tissue samples were

775 processed for ChIP-seq and RNA-seq analogous to the procedure published for the fetal LV
776 sample[143]. ChIP-seq libraries were prepared using the Illumina TruSeq library preparation kit, and
777 pooled and sequenced (50bp single end) using a HiSeq2000 (Illumina) and processed using a
778 previously published pipeline[19], with minor modifications. Briefly, ChIP-seq reads were obtained
779 following quality filtering and adaptor trimming using cutadapt_v1.1 with parameter '-m 25 -q 20'.
780 Bowtie[144] (v2.0.2.0) with parameter '-m 1 -v 2 -p 16' and MACS[119] (v1.4.2) with parameter '-mfold
781 = 10,30 -nomodel -p 0.0001' were used for read mapping (hg19) and peak calling, respectively.
782 Duplicates were removed with SAMtools[136]. RNA-seq libraries were prepared using the TruSeq
783 Stranded Total RNA with Ribo-Zero Human/Mouse/Rat kit (Illumina) according to manufacturer
784 instructions. An additional purification step was used to remove remaining high molecular weight
785 products, as published[143]. RNAseq libraries were pooled and sequenced via single end 50 bp reads
786 on a HiSeq 4000 (Illumina) and processed as previously published, with minor modifications[19].
787 Briefly, RNA-seq reads were preprocessed using quality filtering and adaptor trimming with
788 cutadapt_v1.1 ('-m 25 -q 25'). Tophat v2.0.6 was used to align RNA-seq reads to the mouse reference
789 genome (hg19) and the reads mapping to UCSC known genes were determined by HTSeq[145].
790 Normalized bigWig files were generated using bedtools (bedGraphToBigWig) and IGV browser was
791 used for visualization of profiles.

792 ## Data availability

793 Raw and processed files of the NGS datasets generated are available in the NCBI GEO database with
794 the accession codes GSE161194 (4C-seq) and GSE232887 (super-series including C-HiC, ATAC-
795 seq, ChIP-seq and RNA-seq data). Mouse and human genome coordinates used in this manuscript
796 are GRCm38/mm10 and GRCh37/hg19. All relevant transgenic *in vivo* enhancer data is made
797 available at the Vista Enhancer Browser (https://enhancer.lbl.gov). Correspondence and requests for
798 materials should be addressed to J.C. (jacobb@ucalgary.ca) or M.O. (marco.osterwalder@unibe.ch).

799 ## Competing interests

800 The authors declare no competing financial interests.

801 ## Acknowledgements

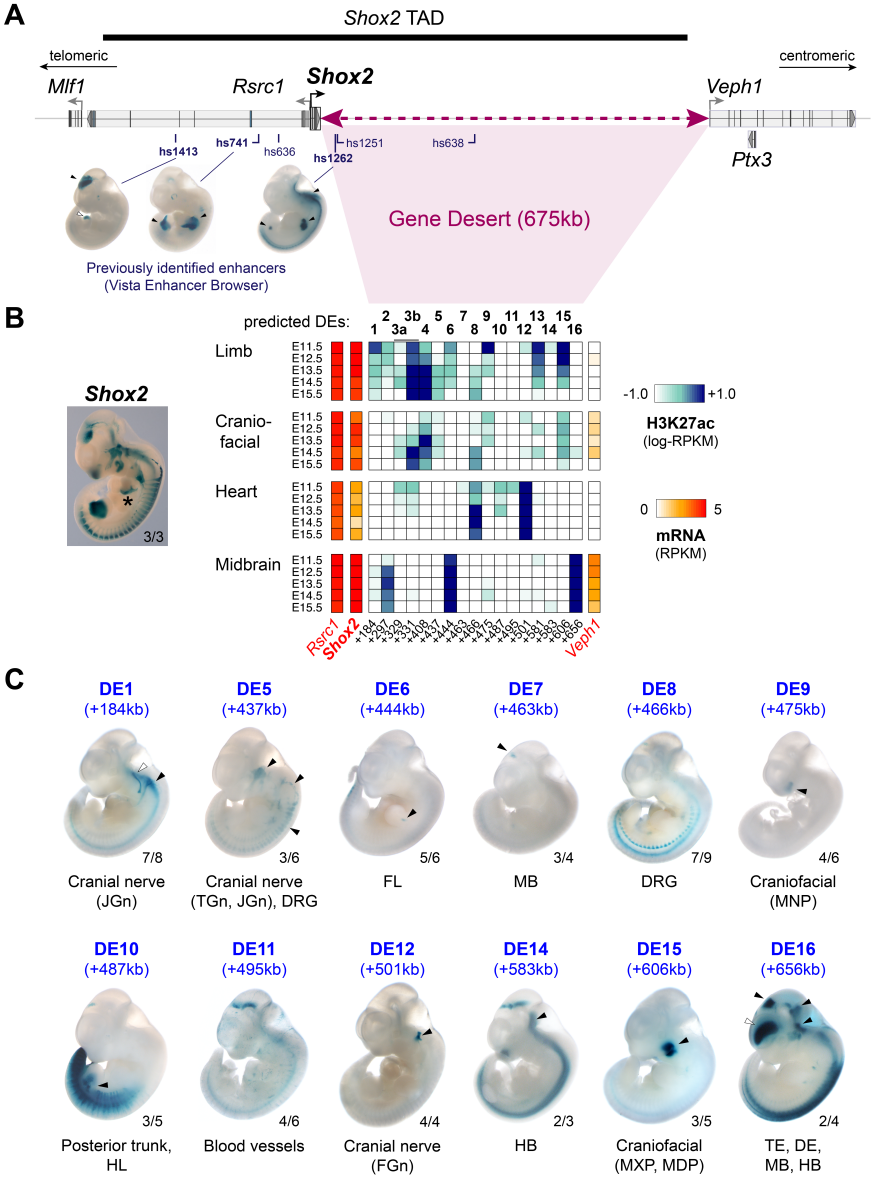
802 We thank L. Lopez-Delisle for sharing expertise on the use of pyGenomeTracks and Capture Hi-C
803 analysis, G. Kelman for preliminary ATAC-seq analysis and D. Duboule for hosting and supporting
804 4C-Seq experiments as well as training in his laboratory. We are grateful to P. Pelczar and the
805 members of the University of Basel Center for Transgenic Models (CTM) for generation of the mouse
806 enhancer deletion allele and thank C. Detotto and her team at the Central Animal Facilities (CAF) of
807 the University of Bern for excellent mouse care. We are grateful to M. Docquier and her team from
808 the iGE3 facility for preparation and sequencing of C-HiC libraries. We thank C. Fielding at the Clara
809 Christie Centre for Mouse Genomics for pronuclear injections conducted at the University of
810 Calgary, to J. Theodor and J. Anderson for use of the SkyScan 1173 uCT scanner and to C. Rolian
811 and C. Unger for help with morphometric analysis. We thank V. Rapp for cloning the +325 transgenic
812 reporter construct. We are grateful to the members of the L.A.P. and A.V. group for technical advice
813 and the members of the M.O. and J.C. labs for useful comments on the manuscript. This work was
814 supported by Swiss National Science Foundation (SNSF) grant PCEFP3_186993 (to M.O.), a
815 Discovery Grant (RGPIN/355731-2013) from the Natural Sciences and Engineering Research
816 Council of Canada (to J.C.) and National Institutes of Health grants R01HG003988, U54HG006997,

817 R24HL123879 and UM1HL098166 (to A.V. and L.A.P.). MO was also supported by grants of the
818 Swiss Heart Foundation (FF20110) and Novartis Foundation for Medical-Biological Research
819 (#21C183). J.L-R. is supported by the MICINN grants PID2020-113497GB-I00 and CEX2020-
820 001088-M (Unidad de Excelencia María de Maeztu institutional grant). G.A. is supported by Swiss
821 National Science Foundation Grant PP00P3_176802. F.D. is supported by a SNSF postdoc.mobility
822 fellowship (P400PB_194334). Research at the E.O. Lawrence Berkeley National Laboratory was
823 performed under Department of Energy Contract DE-AC02-05CH11231, University of California.

824 **Author contributions**

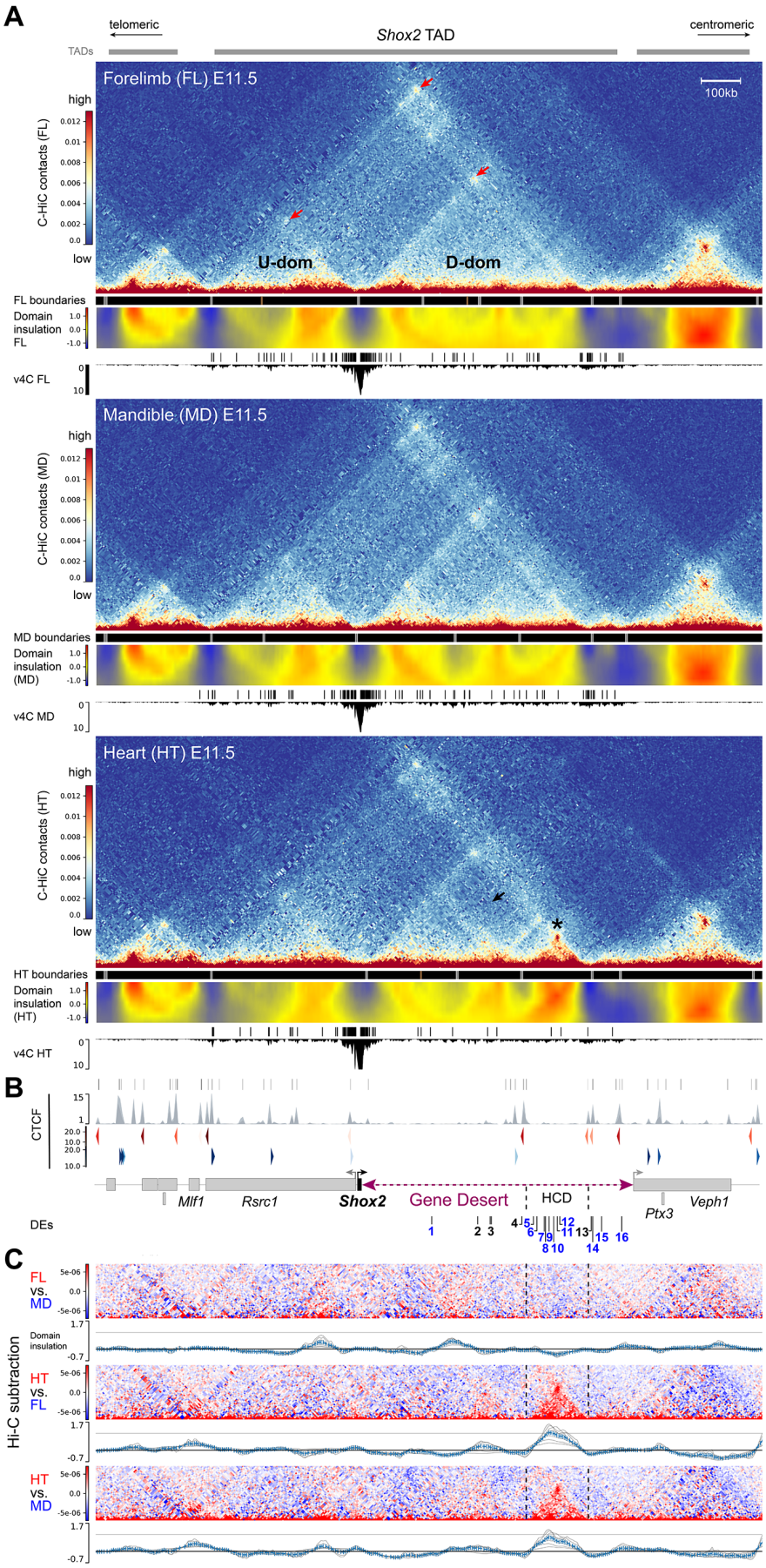
825 M.O. and J.C. conceived the study. S.A.-O., B.J.M, and M.Z. performed critical experimental (S.A.-
826 O., B.J.M.) and computational (M.Z.) analyses for the study. S.A.-O., B.J.M., J.C. and M.O. designed
827 and performed transgenic reporter and gene expression analyses. R.R. conducted experimental C-
828 HiC. V.T. and J.L.-R. executed the *in-situ* hybridization analysis. M.Z. performed C-HiC and ATAC-
829 seq/ChIP-seq data processing and analysis from all mouse datasets. I.B. set up the enhancer profiling
830 framework based on ENCODE data and ChromHMM. C.H.S. and B.J.M. conducted ChIP-seq and
831 RNA-seq from human heart tissues. Y. F.-Y. performed ChIP-seq and RNA-seq processing and
832 analysis of human heart datasets. S.A.-O., E.R-C., A.L., G.A. and J.C. performed 4C-seq experiments
833 and analysis. V.R. and J.G. conducted SV enhancer-deletion experiments. F.D., A.L., R.H., J.A.A.
834 performed additional experimental work related to transgenic reporter validation. T.A.F and C.S.S.
835 did skeletal phenotyping. C.S.N, I.P.-F. and S.T. performed pro-nuclear injections. G.A., D.E.D.,
836 A.V. and L.A.P. provided project funding and support. J.C. and M.O. provided project funding and
837 wrote the manuscript with input from the other authors.

838 # MAIN FIGURES AND FIGURE LEGENDS




839 **Figure 1. The *Shox2* gene desert constitutes a hub for tissue-specific enhancers**. (**A**) Genomic interval
840 containing the *Shox2* TAD[61] and previously identified *Shox2*-associated enhancer regions (see Vista Enhancer
841 browser). Vista IDs in bold mark enhancers with *Shox2*-overlapping and highly reproducible activities. Hs,
842 homo sapiens. (**B**) Heatmap showing H3K27 acetylation (ac) -predicted and ChromHMM-filtered putative
843 enhancers and their temporal signatures in tissues with dominant *Shox2* functions (see full matrix in **Fig. S1**).
844 Blue and red shades represent H3K27ac enrichment and mRNA expression levels, respectively. Distance to
845 *Shox2* TSS (+) is indicated in kb. Left: *Shox2* expression pattern (*Shox2*-LacZ/+) at E11.5[30]. *, indicates
846 removal of forelimb for better visibility of the heart. (**C**) Transgenic LacZ reporter validation of predicted
847 gene desert enhancers (DEs) in mouse embryos at E11.5. Arrowheads points to reproducible enhancer activity
848 with (black) or without (white) *Shox2* overlap. JGn, TGn, FGn: jugular, trigeminal, and facial ganglion,
849 respectively. DRG, dorsal root ganglia. FL, Forelimb. HL, Hindlimb. TE, Telencephalon. DiE, Diencephalon.

850 MB, Midbrain. HB, Hindbrain. MNP, medial nasal process. MXP, maxillary process. MDP, mandibular
851 process. Reproducibility numbers are indicated on the bottom right of each representative embryo shown
852 (reproducible tissue-specific staining vs. number of transgenic embryos with any LacZ signal).




853 **Figure 2. 3D chromatin architecture across the *Shox2* regulatory landscape in distinct tissues.** (**A**) C-
854 HiC analysis of the genomic region containing the *Shox2* TAD[61] in wildtype mouse embryonic forelimb (FL),

855 mandible (MD) and heart (HT) at E11.5 (see also **Fig. S2**). The chr3:65977711-67631930 (mm10) interval is
856 shown. Upper panels (for each tissue): Hi-C contact map revealing upstream (U-dom) and downstream (D-
857 dom) domains flanking the *Shox2* gene. Middle panels: Stronger (gray boxes, $p < 0.01$) and weaker (brown
858 boxes, $p > 0.01$, $< 0.05$) domain boundaries based on TAD separation score. A matrix showing normalized
859 inter-domain insulation score (blue = weak insulation, red = strong insulation) is plotted below. Bottom panels:
860 Virtual 4C (v4C) using a *Shox2*-centered viewpoint shows *Shox2* promoter interaction profiles in the different
861 tissues. *Shox2* contacting regions ($q < 0.1$) as determined by GOTHiC[126] are shown on top. Red arrows point
862 to chromatin domain anchors. Asterisk marks a high-density contact domain (HCD) observed only in heart
863 tissue (chr3:66402500-66572500). Black arrow indicates reduction of internal D-dom contacts between
864 elements inside the HCD and outside in the heart sample (see also **Fig. S2**). (**B**) Top: CTCF enrichment in
865 mESCs[61]. Bottom: CTCF motif orientation (red/blue) and strength (gradient). Protein coding genes (gene
866 bodies) are indicated below. DEs, predicted gene desert enhancers validated in **Fig. 1** (blue: tissue-specific
867 activity). (**C**) C-HiC subtraction to visualize tissue-specific contacts for each tissue comparison (red/blue).
868 Plots below display the corresponding subtracted inter-domain insulation scores. Dashed lines demarcate the
869 HCD borders.

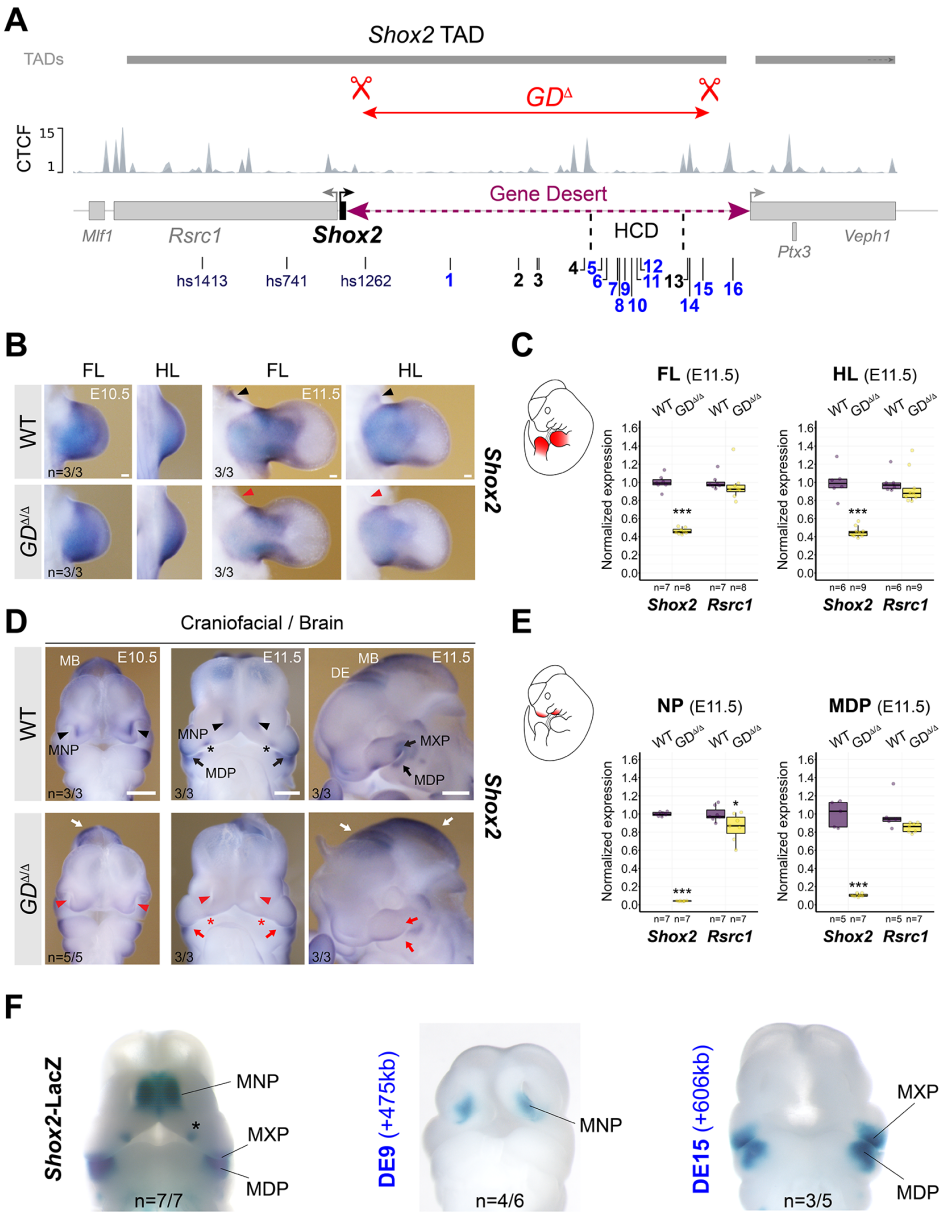
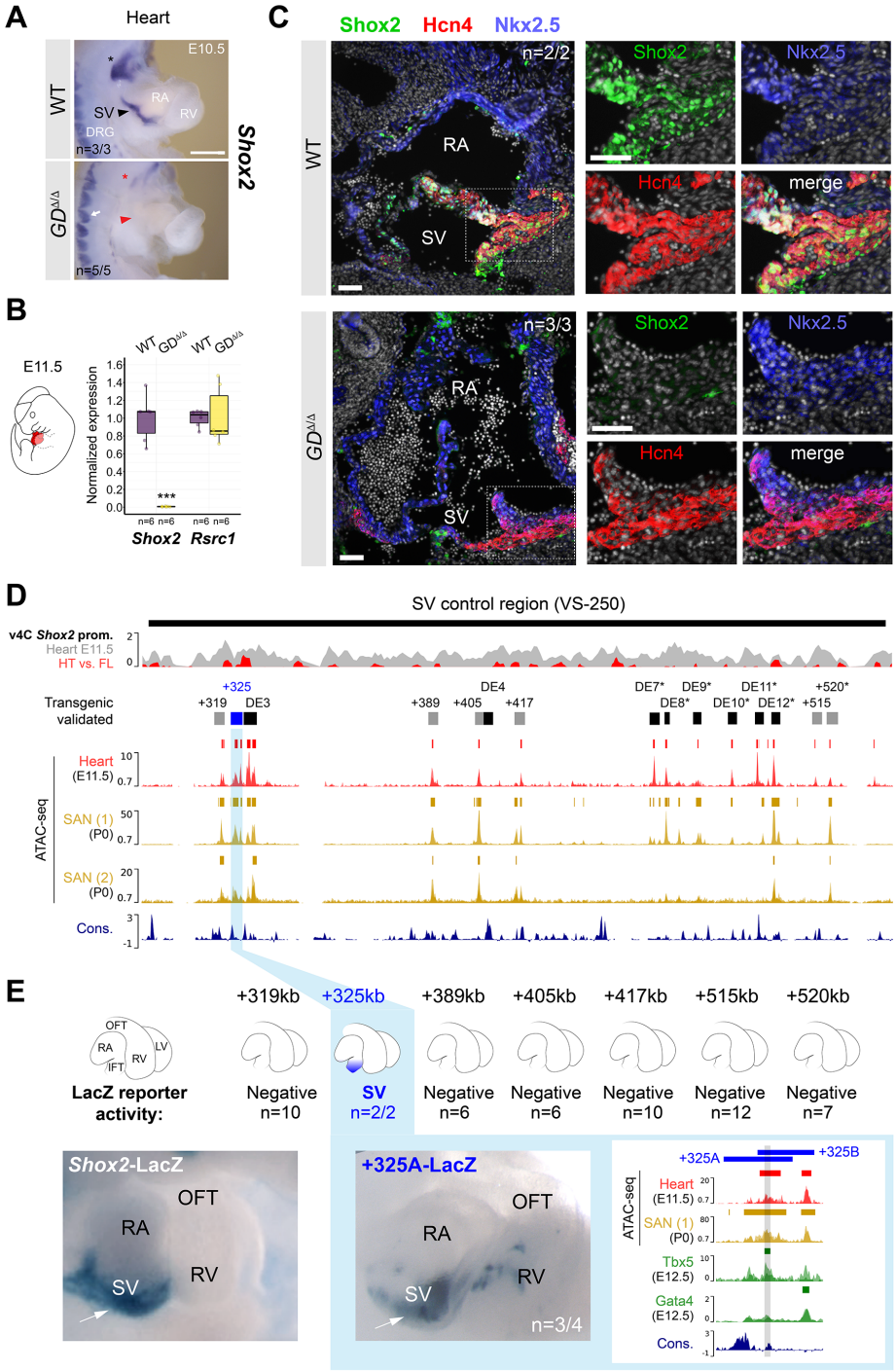



**Figure 3. Gene desert deletion reduces *Shox2* in limb and craniofacial compartments.** (**A**) CRISPR/Cas9-mediated deletion of the intra-TAD *Shox2* gene desert interval ($GD^{\Delta}$) (mm10, chr3:66365062-66947168). Vista (hs) and newly identified gene desert enhancers (1-16, active in blue) are displayed along with TAD interval and CTCF peaks from mESCs[61]. HCD, high-density contact domain (see **Fig. 2**). (**B**, **D**) ISH revealing spatial *Shox2* expression in fore- and hindlimb (FL/HL), craniofacial compartments, and brain in $GD^{\Delta/\Delta}$ embryos compared to wildtype (WT) controls at E10.5 and E11.5. Red arrowheads and red arrows point to regions with severely downregulated or reduced *Shox2* expression, respectively. Red asterisk demarcates *Shox2* loss in the anterior portion of the palatal shelves. White arrows indicate regions (diencephalon, DE and midbrain, MB) without overt changes in *Shox2* expression. Scale bars, 500 µm (B) and 100 µm (D). (**C, E**) Quantitative mRNA analysis (qPCR) in limb and craniofacial tissues of WT and $GD^{\Delta/\Delta}$ embryos. Box plots indicate interquartile range, median, maximum/minimum values (bars). Dots represent individual data points. ***, $P < 0.001$; *, $P < 0.05$ (two-tailed, unpaired t-test for qPCR). (**F**) DE9 and DE15 enhancer activities (**Fig.**

882 **1C**) overlap *Shox2* expression in medial nasal process (MNP) and maxillary-mandibular (MXP-MDP) regions,
883 respectively, in mouse embryos at E11.5. Asterisk marks anterior palatal shelf. “n” indicates number of
884 embryos per genotype or transgene analyzed, with similar results.




885 **Figure 4. The gene desert controls cardiac *Shox2* essential for embryonic viability.** (**A**) ISH revealing
886 absence of *Shox2* transcripts in the cardiac sinus venosus (SV) of GD$^{\Delta/\Delta}$ embryos at E10.5 (red arrowhead).
887 Red asterisk points to reduced *Shox2* in the nodose ganglion of the vagus nerve. White arrow indicates normal
888 *Shox2* expression in the dorsal root ganglia (DRG) of GD$^{\Delta/\Delta}$ embryos. Scale bar, 100 µm. (**B**) Quantitative
889 PCR (qPCR) revealing depletion of *Shox2* in GD$^{\Delta/\Delta}$ hearts compared to WT controls at E11.5. Box plots
890 indicate interquartile range, median, maximum/minimum values (bars). Dots represent individual data points.
891 ***$P < 0.001$ (two-tailed, unpaired t-test). (**C**) Co-localization of Shox2 (green), Hcn4 (red) and Nkx2-5 (blue)
892 in hearts of GD$^{\Delta/\Delta}$ and WT control embryos at E11.5. Shox2 is lost in the Hcn4-marked SAN pacemaker

893 myocardium in absence of the gene desert (dashed outline). Nuclei are shown in gray. Scale bars, 50μm. **(D)**
894 SAN enhancer candidate regions in the gene desert interval (VS-250) essential for *Shox2* expression in the
895 SV[63]. Top: Virtual 4C (v4C) *Shox2* promoter interaction signature in embryonic hearts (gray) overlapped with
896 heart (HT) versus forelimb (FL) subtraction profiles (red) (*see also* **Fig. S3A**). Below: ATAC-seq profiles and
897 peak-calls from embryonic hearts at E11.5 (this study) and SAN pacemaker cells from sorted *Hcn4*-GFP
898 mouse hearts at P0 (Fernandez-Perez *et al*. (1)[64] and Galang *et al*. (2)[44]). Desert enhancers (DEs) accessible in
899 SAN progenitors and additional predicted SAN enhancer elements are shown as black and gray bars,
900 respectively. Distance to the *Shox2* TSS in kb (+) is indicated. Cons, vertebrate conservation track by PhyloP.
901 (**E**) LacZ transgenesis identifies a 3.9kb genomic element 325kb downstream of the *Shox2* TSS (+325) able
902 to drive LacZ reporter expression in the SV at E11.5 (**Fig. S5**). 325A and 325B subregions each drive *Shox2*
903 overlapping SV activity (see **Fig. S5** for 325B). The conserved interval in the 325A/B overlapping region
904 shows a Tbx5 peak (gray bar) but no significant Gata4 enrichment in embryonic hearts at E12.5[65,66] (green
905 tracks). "n" denotes fraction of biological replicates with reproducible results. Single numbers represent total
906 number of transgenic embryos analyzed (without reproducible staining in the heart). RA, right atrium. RV,
907 right ventricle. OFT, outflow tract.

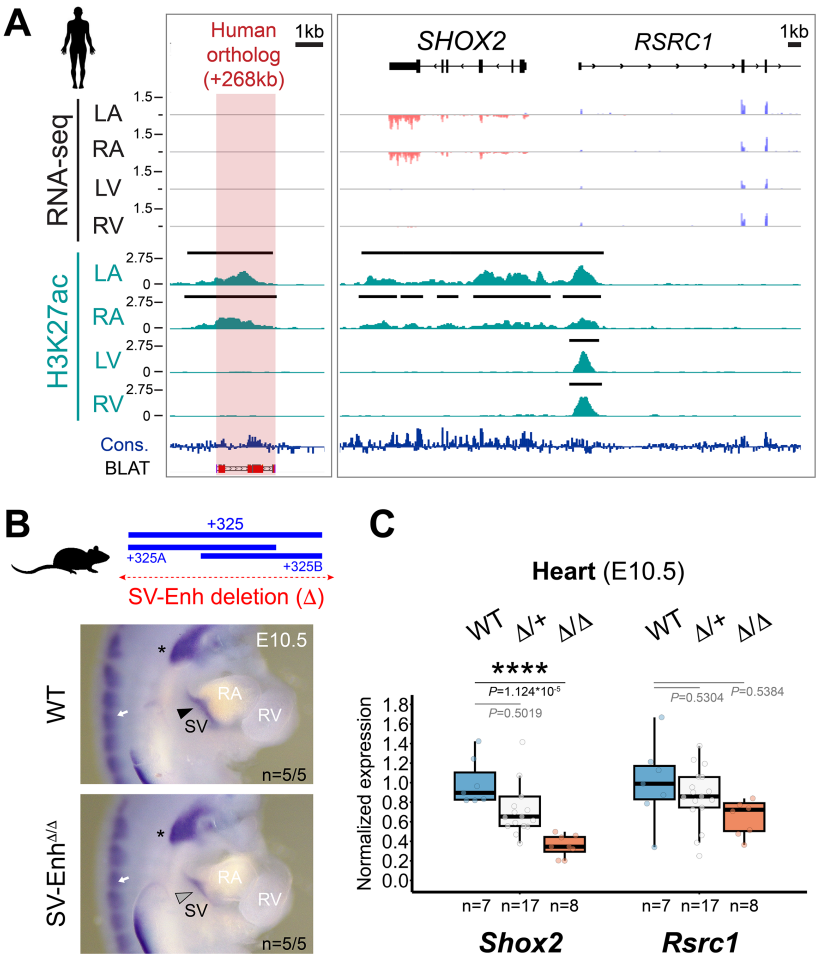
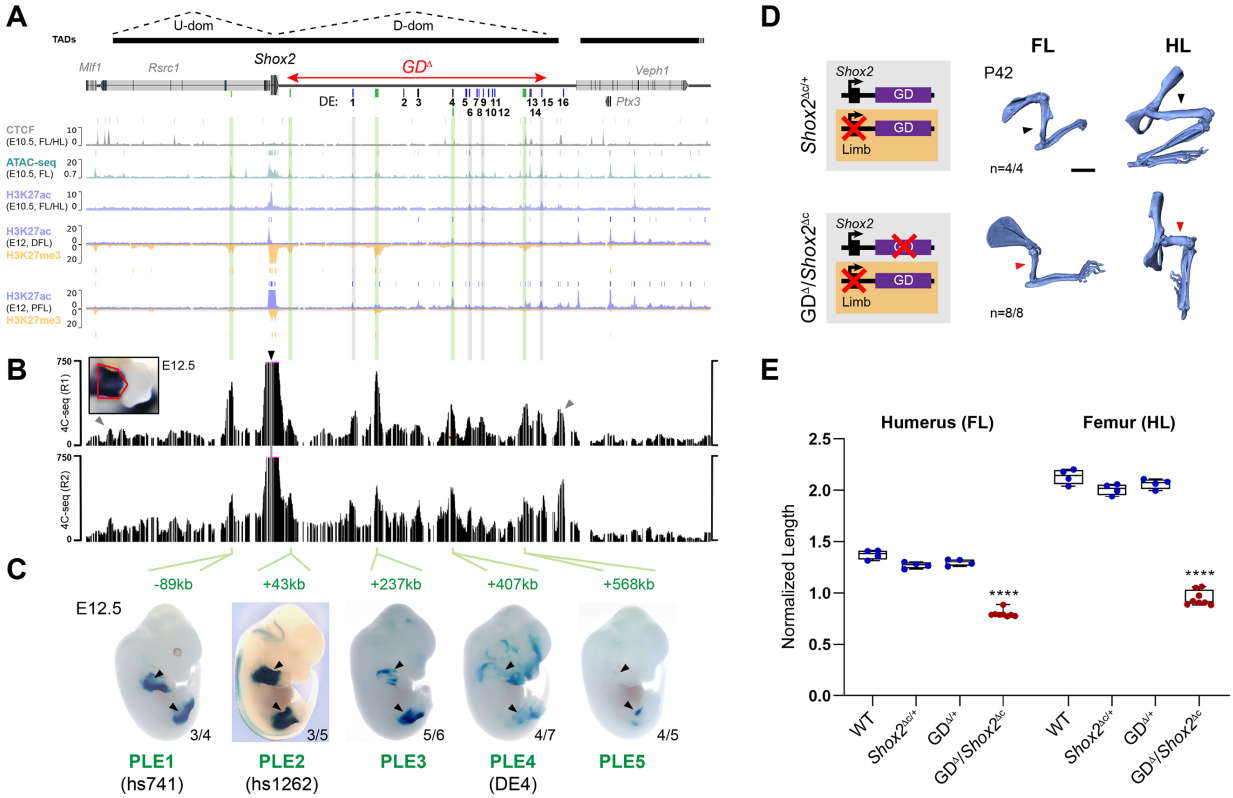



**Figure 5. Enhancer-mediated transcriptional robustness safeguards *Shox2* in the heart** (**A**) H3K27 acetylation ChIP-seq (H3K27ac) and RNA-seq profiles from human fetal heart compartments at PWC17 across the human orthologous sequence of the +325-mouse sinus venosus (SV) enhancer and the *SHOX2* interval. The left ventricle (LV) dataset has been previously published[143]. +268, distance to *SHOX2* TSS. Cons, mammalian conservation by PhyloP. (**B**) Top: Generation of a +325 SV enhancer deletion (4.4kb) allele in mice (SV-Enh$^{\Delta}$). Below: *Shox2* mRNA distribution (ISH) in SV-Enh$^{\Delta/\Delta}$ compared to WT mouse embryos at E10.5. (**C**) qPCR analysis of *Shox2* and *Rsrc1* mRNA levels in SV-Enh$^{\Delta/+}$ and SV-Enh$^{\Delta/\Delta}$ embryonic hearts at E10.5 compared to WT controls. Box plot indicates interquartile range, median, maximum/minimum values (bars) and individual biological replicates (n). *P*-values are shown, with **** P<0.0001 (two-tailed, unpaired t-test). Three outliers, two datapoints of *Shox2* Δ/+ replicates and one for *Rsrc1* (Δ/Δ), are outside of the scale shown. "n" indicates number of biological replicates analyzed, with similar results. LA, left atrium. RA, right atrium. RV, right ventricle.




920 **Figure 6. Stylopod morphogenesis is dependent on distributed proximal limb enhancers in the gene**
921 **desert**. (**A**) Re-processed ChIP-seq datasets from mouse embryonic limbs at E10.5 (CTCF, ATAC-seq,
922 H3K27ac) and E12 (H3K27ac, H3K27me3) showing epigenomic profiles at the *Shox2* locus[55,60,130]. Bars
923 above each track represent peak calls across replicates. DFL, distal forelimb. PFL, proximal forelimb. On top:
924 TAD extension in mESCs[61] (black bars) with desert enhancers identified in **Fig. 1** (DEs 1-16; blue indicates
925 validated activity). Red double arrow demarcates the deleted gene desert interval ($GD^{\Delta}$). (**B**) 4C-seq
926 interaction profiles from two independent biological replicates (R1, R2) of proximal limbs at E12.5 (red
927 outline). Black arrow indicates the 4C-seq viewpoint at the *Shox2* promoter. Gray arrowheads point to CTCF-
928 boundaries of the *Shox2*-TAD. Green lines (in **A**) indicate *Shox2*-interacting elements with putative proximal
929 limb activities. Gray lines mark 4C-seq peaks overlapping previously validated DEs without such activities
930 (**Fig. 1C**). (**C**) Identification of proximal limb enhancers (PLEs) through transgenic LacZ reporter assays in
931 mouse embryos at E12.5. Embryos shown are representatives from stable transgenic LacZ reporter lines (**Fig.**
932 **S6A**). Reproducibility numbers from original transgenic founders are listed for each element (bottom right).
933 (**D**) Left: Schematics illustrating gene desert inactivation ($GD^{\Delta}$) in the presence of reduced limb *Shox2* dosage
934 based on *Prx1*-Cre-mediated *Shox2* deletion ($Shox2^{\Delta c}$). Right: Micro-CT scans of fore (FL)- and hindlimb
935 (HL) skeletons of $GD^{\Delta}/Shox2^{\Delta c}$ and $Shox2^{\Delta c/+}$ control mice at postnatal day 42 (P42). Red arrowheads point
936 to severely reduced stylopods in $GD^{\Delta}/Shox2^{\Delta c}$ individuals compared to controls (black arrowheads). "n",
937 number of biological replicates with reproducible results. (**E**) Micro-CT stylopod quantification at P42 reveals
938 humerus and femur length reductions in $GD^{\Delta}/Shox2^{\Delta c}$ mice. ****, $P \leq 0.0001$ (ANOVA).

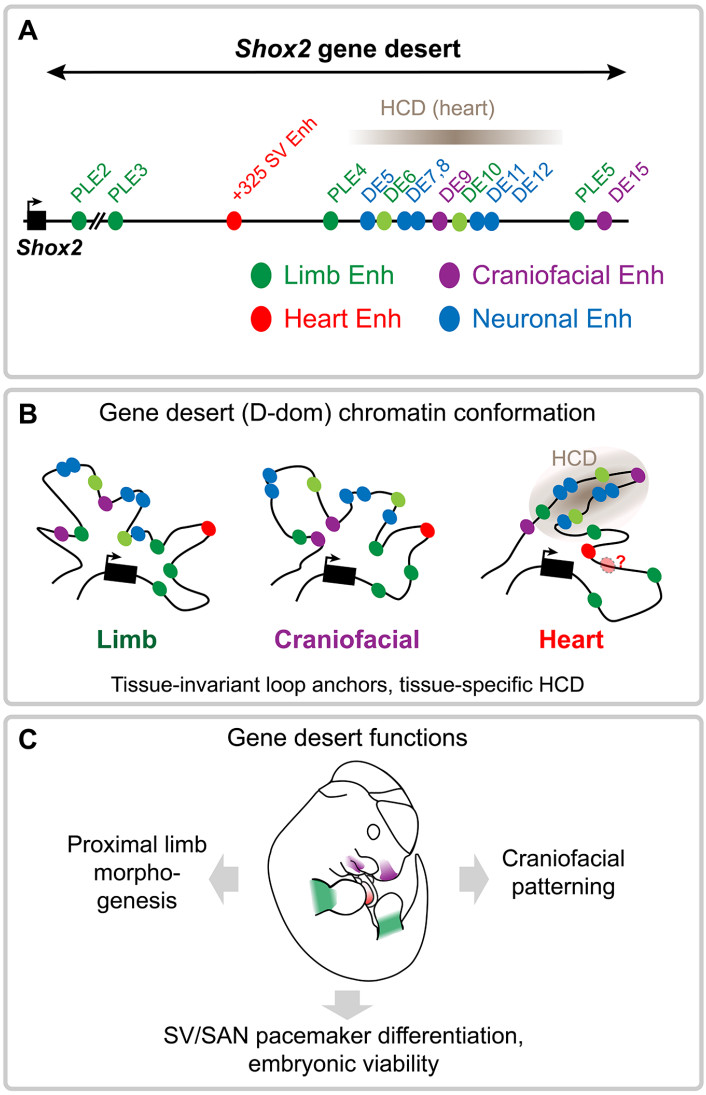



**Figure 7. Graphical summary.** (**A**) Identification of the *Shox2*-flanking gene desert as a reservoir for distributed transcriptional enhancers with activities in limb, craniofacial, cardiac, and neuronal cell populations. (**B**) The *Shox2* gene desert encodes distributed tissue-specific enhancers that are englobed in a dynamic chromatin domain (D-dom) with tissue-invariant loop anchors and a cardiac-specific high-density contact domain (HCD) that may influence enhancer activities. Additional gene desert enhancers are likely to participate in the regulation of cardiac *Shox2* in SAN progenitors. (**C**) Cumulative functions of gene desert enhancers orchestrate pleiotropic *Shox2* expression essential for proximal limb morphogenesis, craniofacial patterning, and cardiac pacemaker development.

947 **REFERENCES**


948 1. Craig Venter, J. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001).
949 2. Ovcharenko, I. *et al.* Evolution and functional classification of vertebrate gene deserts.
950 *Genome Res.* **15**, 137–145 (2005).
951 3. Nobrega, M. A., Ovcharenko, I., Afzal, V. & Rubin, E. M. Scanning human gene deserts for
952 long-range enhancers. *Science* **302**, 413 (2003).
953 4. Catarino, R. R. & Stark, A. Assessing sufficiency and necessity of enhancer activities for gene
954 expression and the mechanisms of transcription activation. *Genes Dev.* **32**, 202–223 (2018).
955 5. Nóbrega, M. A., Zhu, Y., Plajzer-Frick, I., Afzal, V. & Rubin, E. M. Megabase deletions of gene
956 deserts result in viable mice. *Nature* **431**, 984–988 (2004).
957 6. Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell*
958 **147**, 1132–1145 (2011).
959 7. Andrey, G. *et al.* A switch between topological domains underlies HoxD genes collinearity in
960 mouse limbs. *Science* **340**, 1234167 (2013).
961 8. Rodríguez-Carballo, E. *et al.* The HoxD cluster is a dynamic and resilient TAD boundary
962 controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* **31**, 2264–2281
963 (2017).
964 9. Darbellay, F. & Duboule, D. Topological Domains, Metagenes, and the Emergence of
965 Pleiotropic Regulations at Hox Loci. *Curr. Top. Dev. Biol.* **116**, 299–314 (2016).
966 10. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic
967 duplications. *Nature* **538**, 265–269 (2016).
968 11. Symmons, O. *et al.* The Shh Topological Domain Facilitates the Action of Remote Enhancers
969 by Reducing the Effects of Genomic Distances. *Dev. Cell* **39**, 529–543 (2016).
970 12. Marinić, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and
971 gene specificity of the Fgf8 regulatory landscape. *Dev. Cell* **24**, 530–542 (2013).
972 13. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression
973 control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
974 14. Furlong, E. E. M. & Levine, M. Developmental enhancers and chromosome topology. *Science*
975 **361**, 1341–1345 (2018).
976 15. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation
977 in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6456-65 (2015).
978 16. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**,
979 2038–2049 (2016).
980 17. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring
981 of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
982 18. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat.*
983 *Rev. Genet.* **19**, 453–467 (2018).

984 19. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian
985 development. *Nature* **554**, 239–243 (2018).
986 20. Dickel, D. E. *et al.* Ultraconserved Enhancers Are Required for Normal Development. *Cell*
987 **172**, 491-499.e15 (2018).
988 21. Hörnblad, A., Bastide, S., Langenfeld, K., Langa, F. & Spitz, F. Dissection of the Fgf8
989 regulatory landscape by in vivo CRISPR-editing reveals extensive intra- and inter-enhancer
990 redundancy. *Nat. Commun.* **12**, 439 (2021).
991 22. Will, A. J. *et al.* Composition and dosage of a multipartite enhancer cluster control
992 developmental expression of Ihh (Indian hedgehog). *Nat. Genet.* **49**, 1539–1545 (2017).
993 23. van Mierlo, G., Pushkarev, O., Kribelbauer, J. F. & Deplancke, B. Chromatin modules and
994 their implication in genomic organization and gene regulation. *Trends Genet.* (2022)
995 doi:10.1016/j.tig.2022.11.003.
996 24. Malkmus, J. *et al.* Spatial regulation by multiple Gremlin1 enhancers provides digit
997 development with cis-regulatory robustness and evolutionary plasticity. *Nat. Commun.* **12**,
998 5557 (2021).
999 25. Kvon, E. Z. *et al.* Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human
1000 Enhancer Variants. *Cell* **180**, 1262-1271.e15 (2020).
1001 26. Rouco, R. *et al.* Cell-specific alterations in Pitx1 regulatory landscape activation caused by the
1002 loss of a single enhancer. *Nat. Commun.* **12**, 7235 (2021).
1003 27. Cobb, J., Dierich, A., Huss-Garcia, Y. & Duboule, D. A mouse model for human short-stature
1004 syndromes identifies Shox2 as an upstream regulator of Runx2 during long-bone
1005 development. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4511–4515 (2006).
1006 28. Gu, S., Wei, N., Yu, L., Fei, J. & Chen, Y. Shox2-deficiency leads to dysplasia and ankylosis
1007 of the temporomandibular joint in mice. *Mech. Dev.* **125**, 729–742 (2008).
1008 29. Yu, L. *et al.* Shox2-deficient mice exhibit a rare type of incomplete clefting of the secondary
1009 palate. *Development* **132**, 4397–4406 (2005).
1010 30. Rosin, J. M., Kurrasch, D. M. & Cobb, J. Shox2 is required for the proper development of the
1011 facial motor nucleus and the establishment of the facial nerves. *BMC Neurosci.* **16**, 39 (2015).
1012 31. Rosin, J. M. *et al.* Mice lacking the transcription factor SHOX2 display impaired cerebellar
1013 development and deficits in motor coordination. *Dev. Biol.* **399**, 54–67 (2015).
1014 32. Scott, A. *et al.* Transcription factor short stature homeobox 2 is required for proper
1015 development of tropomyosin-related kinase B-expressing mechanosensory neurons. *J.*
1016 *Neurosci.* **31**, 6741–6749 (2011).
1017 33. Xu, J. *et al.* Shox2 regulates osteogenic differentiation and pattern formation during hard
1018 palate development in mice. *J. Biol. Chem.* **294**, 18294–18305 (2019).
1019 34. Blaschke, R. J. *et al.* Targeted mutation reveals essential functions of the homeodomain
1020 transcription factor Shox2 in sinoatrial and pacemaking development. *Circulation* **115**, 1830–
1021 1838 (2007).

1022 35. Espinoza-Lewis, R. A. *et al.* Shox2 is essential for the differentiation of cardiac pacemaker
1023 cells by repressing Nkx2-5. *Dev. Biol.* **327**, 376–385 (2009).
1024 36. van Eif, V. W. W. *et al.* Transcriptome analysis of mouse and human sinoatrial node cells
1025 reveals a conserved genetic program. *Development* **146**, (2019).
1026 37. Ye, W. *et al.* A common Shox2-Nkx2-5 antagonistic mechanism primes the pacemaker cell
1027 fate in the pulmonary vein myocardium and sinoatrial node. *Development* **142**, 2521–2532
1028 (2015).
1029 38. Hoffmann, S. *et al.* Coding and non-coding variants in the SHOX2 gene in patients with early-
1030 onset atrial fibrillation. *Basic Res. Cardiol.* **111**, 36 (2016).
1031 39. Hoffmann, S. *et al.* Functional Characterization of Rare Variants in the SHOX2 Gene
1032 Identified in Sinus Node Dysfunction and Atrial Fibrillation. *Front. Genet.* **10**, 648 (2019).
1033 40. Li, N. *et al.* A SHOX2 loss-of-function mutation underlying familial atrial fibrillation. *Int. J. Med.*
1034 *Sci.* **15**, 1564–1572 (2018).
1035 41. Mori, A. D. *et al.* Tbx5-dependent rheostatic control of cardiac gene expression and
1036 morphogenesis. *Dev. Biol.* **297**, 566–586 (2006).
1037 42. Vedantham, V., Galang, G., Evangelista, M., Deo, R. C. & Srivastava, D. RNA sequencing of
1038 mouse sinoatrial node reveals an upstream regulatory role for Islet-1 in cardiac pacemaker
1039 cells. *Circ. Res.* **116**, 797–803 (2015).
1040 43. Puskaric, S. *et al.* Shox2 mediates Tbx5 activity by regulating Bmp4 in the pacemaker region
1041 of the developing heart. *Hum. Mol. Genet.* **19**, 4625–4633 (2010).
1042 44. Galang, G. *et al.* ATAC-Seq Reveals an Isl1 Enhancer That Regulates Sinoatrial Node
1043 Development and Function. *Circ. Res.* **127**, 1502–1518 (2020).
1044 45. Hoffmann, S. *et al.* Islet1 is a direct transcriptional target of the homeodomain transcription
1045 factor Shox2 and rescues the Shox2-mediated bradycardia. *Basic Res. Cardiol.* **108**, 339
1046 (2013).
1047 46. Marchini, A., Ogata, T. & Rappold, G. A. A Track Record on SHOX: From Basic Research to
1048 Complex Models and Therapy. *Endocr. Rev.* **37**, 417–448 (2016).
1049 47. Rosin, J. M., Abassah-Oppong, S. & Cobb, J. Comparative transgenic analysis of enhancers
1050 from the human SHOX and mouse Shox2 genomic regions. *Hum. Mol. Genet.* **22**, 3063–3076
1051 (2013).
1052 48. Liu, H., Jiao, Z., Espinoza-Lewis, R. A., Chen, C. & Chen, Y. FUNCTIONAL REDUNDANCY
1053 BETWEEN HUMAN SHOX AND MOUSE SHOX2 IN THE REGULATION OF SINUS NODE
1054 FORMATION. *J. Am. Coll. Cardiol.* **57**, E53 (2011).
1055 49. Ye, W. *et al.* A unique stylopod patterning mechanism by Shox2-controlled osteogenesis.
1056 *Development* **143**, 2548–2560 (2016).
1057 50. Cazalla, D., Newton, K. & Cáceres, J. F. A novel SR-related protein is required for the second
1058 step of Pre-mRNA splicing. *Mol. Cell. Biol.* **25**, 2969–2980 (2005).

1059 51. Scala, M. *et al.* RSRC1 loss-of-function variants cause mild to moderate autosomal recessive
1060 intellectual disability. *Brain* **143**, e31 (2020).
1061 52. Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser--a
1062 database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88-92 (2007).
1063 53. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization.
1064 *Nat. Methods* **9**, 215–216 (2012).
1065 54. Gorkin, D. U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal development.
1066 *Nature* **583**, 744–751 (2020).
1067 55. Rodríguez-Carballo, E., Lopez-Delisle, L., Yakushiji-Kaminatsui, N., Ullate-Agote, A. &
1068 Duboule, D. Impact of genome architecture on the functional activation and repression of Hox
1069 regulatory landscapes. *BMC Biol.* **17**, 55 (2019).
1070 56. ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human
1071 and mouse genomes. *Nature* **583**, 699–710 (2020).
1072 57. Méndez-Maldonado, K., Vega-López, G. A., Aybar, M. J. & Velasco, I. Neurogenesis From
1073 Neural Crest Cells: Molecular Mechanisms in the Formation of Cranial Nerves and Ganglia.
1074 *Front Cell Dev Biol* **8**, 635 (2020).
1075 58. Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs
1076 reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
1077 59. Kragesteen, B. K. *et al.* Dynamic 3D chromatin architecture contributes to enhancer specificity
1078 and limb morphogenesis. *Nat. Genet.* **50**, 1463–1473 (2018).
1079 60. Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of Shh
1080 during limb development. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12390–12399 (2019).
1081 61. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**,
1082 557-572.e24 (2017).
1083 62. van Eif, V. W. W., Devalla, H. D., Boink, G. J. J. & Christoffels, V. M. Transcriptional
1084 regulation of the cardiac conduction system. *Nat. Rev. Cardiol.* **15**, 617–630 (2018).
1085 63. van Eif, V. W. W. *et al.* Genome-Wide Analysis Identifies an Essential Human TBX3
1086 Pacemaker Enhancer. *Circ. Res.* **127**, 1522–1535 (2020).
1087 64. Fernandez-Perez, A. *et al.* Hand2 Selectively Reorganizes Chromatin Accessibility to Induce
1088 Pacemaker-like Transcriptional Reprogramming. *Cell Rep.* **27**, 2354-2369.e7 (2019).
1089 65. Akerberg, B. N. *et al.* A reference map of murine cardiac transcription factor chromatin
1090 occupancy identifies dynamic and conserved enhancers. *Nat. Commun.* **10**, 4907 (2019).
1091 66. He, A. *et al.* Dynamic GATA4 enhancers shape the chromatin landscape central to heart
1092 development and disease. *Nat. Commun.* **5**, 4907 (2014).
1093 67. Monti, R. *et al.* Limb-Enhancer Genie: An accessible resource of accurate enhancer
1094 predictions in the developing limb. *PLoS Comput. Biol.* **13**, e1005720 (2017).

1095 68. Chen, S., Lee, B., Lee, A. Y.-F., Modzelewski, A. J. & He, L. Highly Efficient Mouse Genome
1096 Editing by CRISPR Ribonucleoprotein Electroporation of Zygotes. *J. Biol. Chem.* **291**, 14457–
1097 14467 (2016).
1098 69. Yu, L. *et al.* Shox2 is required for chondrocyte proliferation and maturation in proximal limb
1099 skeleton. *Dev. Biol.* **306**, 549–559 (2007).
1100 70. Bobick, B. E. & Cobb, J. Shox2 regulates progression through chondrogenesis in the mouse
1101 proximal limb. *J. Cell Sci.* **125**, 6071–6083 (2012).
1102 71. Neufeld, S. J., Wang, F. & Cobb, J. Genetic interactions between Shox2 and Hox genes
1103 during the regional growth and development of the mouse limb. *Genetics* **198**, 1117–1126
1104 (2014).
1105 72. Logan, M. *et al.* Expression of Cre Recombinase in the developing mouse limb bud driven by
1106 a Prxl enhancer. *Genesis* **33**, 77–80 (2002).
1107 73. Touceda-Suárez, M. *et al.* Ancient Genomic Regulatory Blocks Are a Source for Regulatory
1108 Gene Deserts in Vertebrates after Whole-Genome Duplications. *Mol. Biol. Evol.* **37**, 2857–
1109 2864 (2020).
1110 74. Long, H. K. *et al.* Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a
1111 Craniofacial Disorder. *Cell Stem Cell* **27**, 765-783.e14 (2020).
1112 75. de Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their
1113 regulatory landscapes. *Nature* **502**, 499–506 (2013).
1114 76. Lonfat, N., Montavon, T., Darbellay, F., Gitto, S. & Duboule, D. Convergent evolution of
1115 complex regulatory landscapes and pleiotropy at Hox loci. *Science* **346**, 1004–1006 (2014).
1116 77. Pang, B., van Weerd, J. H., Hamoen, F. L. & Snyder, M. P. Identification of non-coding
1117 silencer elements and their regulation of gene expression. *Nat. Rev. Mol. Cell Biol.* (2022)
1118 doi:10.1038/s41580-022-00549-9.
1119 78. Pachano, T., Haro, E. & Rada-Iglesias, A. Enhancer-gene specificity in development and
1120 disease. *Development* **149**, (2022).
1121 79. Batut, P. J. *et al.* Genome organization controls transcriptional dynamics during development.
1122 *Science* **375**, 566–570 (2022).
1123 80. Statello, L., Guo, C.-J., Chen, L.-L. & Huarte, M. Gene regulation by long non-coding RNAs
1124 and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118 (2021).
1125 81. Dickel, D. E. *et al.* Genome-wide compendium and functional assessment of in vivo heart
1126 enhancers. *Nat. Commun.* **7**, 12923 (2016).
1127 82. Claringbould, A. & Zaugg, J. B. Enhancers in disease: molecular basis and emerging
1128 treatment strategies. *Trends Mol. Med.* **27**, 1060–1073 (2021).
1129 83. van der Lee, R., Correard, S. & Wasserman, W. W. Deregulated Regulators: Disease-
1130 Causing cis Variants in Transcription Factor Genes. *Trends Genet.* **36**, 523–539 (2020).
1131 84. Corradin, O. & Scacheri, P. C. Enhancer variants: evaluating functions in common disease.
1132 *Genome Med.* **6**, 85 (2014).

1133 85. Sun, C., Zhang, T., Liu, C., Gu, S. & Chen, Y. Generation of Shox2-Cre allele for tissue
1134 specific manipulation of genes in the developing heart, palate, and limb. *Genesis* **51**, 515–522
1135 (2013).
1136 86. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental
1137 enhancers in humans. *Nature* **470**, 279–283 (2011).
1138 87. Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during
1139 mammalian development. *Cell* **155**, 1521–1531 (2013).
1140 88. Gasperini, M., Tome, J. M. & Shendure, J. Towards a comprehensive catalogue of validated
1141 and target-linked human enhancers. *Nat. Rev. Genet.* **21**, 292–310 (2020).
1142 89. Mannion, B. J. *et al.* Uncovering Hidden Enhancers Through Unbiased In Vivo Testing.
1143 *bioRxiv* 2022.05.29.493901 (2022) doi:10.1101/2022.05.29.493901.
1144 90. Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development
1145 and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).
1146 91. van Ouwerkerk, A. F. *et al.* Patient-Specific TBX5-G125R Variant Induces Profound
1147 Transcriptional Deregulation and Atrial Dysfunction. *Circulation* **145**, 606–619 (2022).
1148 92. Zhang, M. *et al.* Long-range Pitx2c enhancer-promoter interactions prevent predisposition to
1149 atrial fibrillation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22692–22698 (2019).
1150 93. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional
1151 enhancers. *Nature* **466**, 490–493 (2010).
1152 94. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster
1153 robustness of Drosophila gastrulation. *Curr. Biol.* **20**, 1562–1567 (2010).
1154 95. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory
1155 Networks. *Curr. Biol.* **26**, 38–51 (2016).
1156 96. Bolt, C. C. & Duboule, D. The regulatory landscapes of developmental genes. *Development*
1157 **147**, (2020).
1158 97. Cova, G. *et al.* Combinatorial effects on gene expression at the Lbx1/Fgf8 locus resolve split-
1159 hand/foot malformation type 3. *Nat. Commun.* **14**, 1475 (2023).
1160 98. Berlivet, S. *et al.* Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA
1161 genes in developing limbs. *PLoS Genet.* **9**, e1004018 (2013).
1162 99. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles
1163 of chromatin looping. *Cell* **159**, 1665–1680 (2014).
1164 100. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev.*
1165 *Genet.* **19**, 789–800 (2018).
1166 101. Conte, M. *et al.* Polymer physics indicates chromatin folding variability across single-cells
1167 results from state degeneracy in phase separation. *Nat. Commun.* **11**, 3289 (2020).
1168 102. Chang, L.-H., Ghosh, S. & Noordermeer, D. TADs and Their Borders: Free Movement or
1169 Building a Wall? *J. Mol. Biol.* **432**, 643–652 (2020).

1170 103. Skuplik, I. *et al.* Identification of a limb enhancer that is removed by pathogenic deletions
1171 downstream of the SHOX gene. *Sci. Rep.* **8**, 14292 (2018).
1172 104. Chen, J. *et al.* Enhancer deletions of the SHOX gene as a frequent cause of short stature: the
1173 essential role of a 250 kb downstream regulatory domain. *J. Med. Genet.* **46**, 834–839 (2009).
1174 105. Shears, D. J. *et al.* Mutation and deletion of the pseudoautosomal gene SHOX cause Leri-
1175 Weill dyschondrosteosis. *Nat. Genet.* **19**, 70–73 (1998).
1176 106. Rappold, G. A., Shanske, A. & Saenger, P. All shook up by SHOX deficiency. *The Journal of*
1177 *pediatrics* vol. 147 422–424 (2005).
1178 107. Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause
1179 growth failure in idiopathic short stature and Turner syndrome. *Nat. Genet.* **16**, 54–63 (1997).
1180 108. Tropeano, M. *et al.* Microduplications at the pseudoautosomal SHOX locus in autism
1181 spectrum disorders and related neurodevelopmental conditions. *J. Med. Genet.* **53**, 536–547
1182 (2016).
1183 109. Clement-Jones, M. *et al.* The short stature homeobox gene SHOX is involved in skeletal
1184 abnormalities in Turner syndrome. *Hum. Mol. Genet.* **9**, 695–702 (2000).
1185 110. Durand, C. *et al.* Alternative splicing and nonsense-mediated RNA decay contribute to the
1186 regulation of SHOX expression. *PLoS One* **6**, e18115 (2011).
1187 111. Jackman, W. R. & Kimmel, C. B. Coincident iterated gene expression in the amphioxus neural
1188 tube. *Evol. Dev.* **4**, 366–374 (2002).
1189 112. Wong, E. S. *et al.* Deep conservation of the enhancer regulatory code in animals. *Science*
1190 **370**, (2020).
1191 113. Kothary, R. *et al.* Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice.
1192 *Development* **105**, 707–714 (1989).
1193 114. Osterwalder, M. *et al.* Characterization of Mammalian In Vivo Enhancers Using Mouse
1194 Transgenesis and CRISPR Genome Editing. *Methods Mol. Biol.* **2403**, 147–186 (2022).
1195 115. Darbellay, F. *et al.* Chondrogenic Enhancer Landscape of Limb and Axial Skeleton
1196 Development. *bioRxiv* 2023.05.10.539849 (2023) doi:10.1101/2023.05.10.539849.
1197 116. Andras Nagy, M. G., Vintersten, K., and Behringer, R. Manipulating the Mouse Embryo: A
1198 Laboratory Manual, 3rd Edn. *Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press*
1199 (2003).
1200 117. Labun, K. *et al.* CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing.
1201 *Nucleic Acids Res.* **47**, W171–W174 (2019).
1202 118. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9
1203 genome editing experiments and screens. *Nucleic Acids Res.* **46**, W242–W245 (2018).
1204 119. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
1205 120. Wingett, S. *et al.* HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310
1206 (2015).

121. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
122. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
123. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
124. Wolff, J., Backofen, R. & Grüning, B. Loop detection using Hi-C data with HiCExplorer. *Gigascience* **11**, (2022).
125. Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 (2021).
126. Mifsud, B. *et al.* GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data. *PLoS One* **12**, e0174744 (2017).
127. Noordermeer, D. *et al.* The dynamic architecture of Hox gene clusters. *Science* **334**, 222–225 (2011).
128. Noordermeer, D. *et al.* Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *Elife* **3**, e02557 (2014).
129. David, F. P. A. *et al.* HTSstation: a web application and open-access libraries for high-throughput sequencing data analysis. *PLoS One* **9**, e85879 (2014).
130. Tissières, V. *et al.* Gene Regulatory and Expression Differences between Mouse and Pig Limb Buds Provide Insights into the Evolutionary Emergence of Artiodactyl Traits. *Cell Rep.* **31**, 107490 (2020).
131. Unger, C. M., Devine, J., Hallgrímsson, B. & Rolian, C. Selection for increased tibia length in mice alters skull shape through parallel changes in developmental mechanisms. *Elife* **10**, (2021).
132. Cignoni, P. *et al.* MeshLab: an Open-Source Mesh Processing Tool. *Sixth Eurographics Italian Chapter Conference* 129–136 (2008).
133. Cosman, M. N., Sparrow, L. M. & Rolian, C. Changes in shape and cross-sectional geometry in the tibia of mice selectively bred for increases in relative bone length. *J. Anat.* **228**, 940–951 (2016).
134. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9 (2015).
135. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
136. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
137. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).

1245 138. Diaferia, G. R. *et al.* Dissection of transcriptional and cis-regulatory control of differentiation in
1246 human pancreatic cancer. *EMBO J.* **35**, 595–617 (2016).
1247 139. Erwin, G. D. *et al.* Integrating diverse datasets improves developmental enhancer prediction.
1248 *PLoS Comput. Biol.* **10**, e1003677 (2014).
1249 140. Rahmanian, S. *et al.* Dynamics of microRNA expression during mouse prenatal development.
1250 *Genome Res.* **29**, 1900–1909 (2019).
1251 141. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif.
1252 *Bioinformatics* **27**, 1017–1018 (2011).
1253 142. Pohl, A. & Beato, M. bwtool: a tool for bigWig files. *Bioinformatics* **30**, 1618–1619 (2014).
1254 143. Spurrell, C. H. *et al.* Genome-wide fetalization of enhancer architecture in heart disease. *Cell*
1255 *Rep.* **40**, 111400 (2022).
1256 144. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient
1257 alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
1258 145. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput
1259 sequencing data. *Bioinformatics* **31**, 166–169 (2015).