Deletion of CTCF sites in the SHH locus alters enhancer-promoter interactions and leads to acheiropodia.

Aki Ushiki, Yichi Zhang, Chenling Xiong, Jingjing Zhao, Ilias
Georgakopoulos-Soares, Lauren Kane, Kirsty Jamieson, Michael J Bamshad,
Deborah A Nickerson, Yin Shen, et al.

▶ **To cite this version:**

HAL Id: hal-03344685
https://hal.univ-lille.fr/hal-03344685

Submitted on 15 Sep 2021

3

4   **Deletion of CTCF sites in the *SHH* locus alters enhancer-promoter interactions and leads to**
5   **acheiropodia**

6

7

8   Aki Ushiki[1,2], Yichi Zhang[1,2,3], Chenling Xiong[1,2], Jingjing Zhao[1,2], Ilias Georgakopoulos-Soares[1,2],
9   Lauren Kane[4], Kirsty Jamieson[2,5], Michael J. Bamshad[6,7,8], Deborah A. Nickerson[7,8], University of
10   Washington Center for Mendelian Genomics[#], Yin Shen[2,5], Laura A. Lettice[4], Elizabeth Lemos
11   Silveira-Lucas[9], Florence Petit[10], Nadav Ahituv[1,2]

12

13

14   [1] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco,
15   San Francisco, CA 94158, USA.
16   [2] Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94158,
17   USA.
18   [3] School of Pharmaceutical Sciences, Tsinghua University, Beijing, 100084, China
19   [4] MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh,
20   Crewe Road, Edinburgh EH4 2XU, UK
21   [5] Department of Neurology, University of California San Francisco, San Francisco, CA 94158, USA.
22   [6] Department of Pediatrics, University of Washington, Seattle, WA 98195, USA.
23   [7] Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA.
24   [8] Brotman-Baty Institute, Seattle, WA 98195, USA.
25   [9] Consultorio Genetica Clinica, Porto Alegre, 90440, Brazil
26   [10] CHU Lille, University of Lille, EA7364 RADEME, F-59000, Lille, France.
27   [#] Please see the end of the file for the consortium author list. Non-contributing members of the
28   consortium are listed within the Supplementary Information file.

29

30   Correspondence should be addressed to N.A. (email:nadav.ahituv@ucsf.edu).

31

32

33 **Abstract**

34 Acheiropodia, congenital limb truncation, is associated with homozygous deletions in

35 the *LMBR1* gene around ZRS, an enhancer regulating *SHH* during limb development. How these

36 deletions lead to this phenotype is unknown. Using whole-genome sequencing, we fine-mapped

37 the acheiropodia-associated region to 12 kb and show that it does not function as an enhancer.

38 CTCF and RAD21 ChIP-seq together with 4C-seq and DNA FISH identify three CTCF sites within

39 the acheiropodia-deleted region that mediate the interaction between the ZRS and

40 the *SHH* promoter. This interaction is substituted with other CTCF sites centromeric to the ZRS in

41 the disease state. Mouse knockouts of the orthologous 12 kb sequence have no apparent

42 abnormalities, showcasing the challenges in modelling CTCF alterations in animal models due to

43 inherent motif differences between species. Our results show that alterations in CTCF motifs can

44 lead to a Mendelian condition due to altered enhancer-promoter interactions.

45

46

Introduction

Acheiropodia (OMIM 200500) is a rare autosomal recessive disorder associated with bilateral congenital transverse defects of the upper and lower limbs including aplasia of the hands and feet[1]. Genetic analysis of five Brazilian families with acheiropodia, three of which were consanguineous, identified a homozygous deletion encompassing exon 4 of the limb development membrane protein 1 (*LMBR1*) gene to be associated with this phenotype[2]. The deletion was estimated to cover 4-6 kilo base (kb) on either side of this exon. However, no assays were done to fine map the deletion or functionally characterize how it could be causing acheiropodia.

While exon 4 of *LMBR1* was deleted in the individuals with acheiropodia, it is likely not the cause of this phenotype. LMBR1 is a membrane protein that is ubiquitously expressed[3] and a 35 kb deletion in mice that encompasses exons 1-3 of this gene did not lead to a limb phenotype[4]. *LMBR1* contains an enhancer within intron 5, named the zone of polarizing activity regulatory sequence (ZRS), that regulates the Sonic Hedgehog (*SHH*) gene during limb development. *SHH* encodes a ligand that plays a major role in the development of several tissues, including the limb[5]. In mice, *Shh* is expressed at the posterior part of the limb buds around embryonic day (E) 10-12[6,7,8] and plays a central role in digit patterning and limb outgrowth[9,10]. *Shh* homozygous knockout mice display early lethality with defective axial patterning and limb truncation reminding of acheiropodia[10]. In humans, heterozygous pathogenic variants in *SHH* are responsible for a large spectrum of central nervous system malformations without any limb malformation, of which the most severe is holoprosencephaly (OMIM 142945)[11]. Bi-allelic *SHH* disruption has not been described in humans. Mutations in the ZRS, located ~1 Mb away of *SHH*, cause non-syndromic limb malformations in humans, mice and many other species, consisting primarily of preaxial polydactyly due to ectopic *SHH* expression in the limb bud[12,13,14]. In addition, homozygous deletions encompassing the ZRS lead to acheiropodia in humans and mice[15,16]. Collectively, these results indicate that acheiropodia is likely caused by reduced *SHH* expression during limb development. However, the ZRS is completely intact in the Brazilian individuals with acheiropodia who are homozygous for the *LMBR1* exon 4 deletion, suggesting that other functional units associated with *SHH* limb expression may be disrupted by this deletion.

The architectural protein CCCTC-binding factor (CTCF) is known to play a central role in chromatin conformation[17]. It is involved in forming topologically associating domain (TAD), regions in the

79 genome that are on average ~880 kb in length and are defined as having more frequent

80 interactions within this domain than outside it[18,19]. In addition, CTCF is known to mediate long

81 range enhancer-promoter interactions[17]. CTCF-bound sites in a convergent orientation are thought

82 to halt chromatin loops that are progressively being extruded by the cohesin complex[20], facilitating

83 specific chromatin interaction. Previous studies in mice deleted individual and combinations of

84 CTCF sites in the *Shh* locus, some of them affecting interactions between the ZRS and *Shh*

85 promoter and leading to a reduction of up to ~52% of *Shh* expression in the limb, but none of which

86 led to an observable limb phenotype[4,21]. Interestingly, ectopic CTCF sites appeared in these CTCF

87 motif knockout mice likely supporting compensatory interactions[21].

88

89 We used whole-genome sequencing (WGS) to fine map the homozygous acheiropodia-associated

90 deletion in one of the probands from the Brazilian families, identifying a 12 kb deletion surrounding

91 *LMBR1* exon 4. Using a mouse transgenic enhancer assay, we show that this 12 kb sequence

92 does not have enhancer activity in the developing limb. Further analyses of this sequence using

93 CTCF and RAD21 ChIP-seq identified three CTCF sites in convergent orientation to *SHH* along

94 with RAD21 binding in this region. ChIP-seq analyses in the homozygous proband found an ectopic

95 CTCF site 27 kb centromeric to the ZRS. Consistent with these alterations of CTCF and RAD21

96 binding, interactions between the *SHH* promoter and the ZRS were found to be impaired in the

97 proband using 4C-seq and DNA fluorescence *in situ* hybridization (FISH). Finally, we generated a

98 mouse knockout of the orthologous 12 kb acheiropodia-associated region and did not find any limb

99 malformations, highlighting the differential chromatin interactions in this locus in mice compared to

100 humans. Combined, our results suggest that, in humans, CTCF sites adjacent to the ZRS are likely

101 needed as a scaffold to associate the *SHH* promoter to the ZRS and that this mechanism is

102 different in mice.

103

Results

Whole-genome sequencing identifies a 12 kb acheiropodia-associated deletion

We obtained genomic DNA and lymphoblastoid cell lines from a female proband with acheiropodia and her parents. The proband has terminal transverse hemimelia of the four limbs with truncation of both hands and feet[22]. Prior genetic testing identified a deletion overlapping exon 4 of the *LMBR1* gene with an estimate for the deletion's boundaries to be around 1.2–2.5 kb and 2.7–3.5 kb 5' and 3' of exon 4 respectively[2]. To identify the exact deletion coordinates and assess whether other pathogenic variants might explain the phenotype, we carried out WGS on the proband and her parents. Because of known consanguinity (**Fig. 1a**), we searched for regions of homozygosity in the proband, finding runs spanning a total of 302 mega base (Mb) within the genome (**Supplementary Table 1**). Previous genomic analyses of five consanguineous families with acheiropodia, including this family (Family 2 in[2]), found that all of them share a ~0.5 Mb region of homozygosity in the *LMBR1* gene locus. Based on these results and the known deletion of exon 4, we focused our analyses on this region, identifying a 4 Mb region of homozygosity from rs12719966 to rs1985369 (chr7:155,356,342- 159,326,530; hg38). No pathogenic or likely pathogenic variants were found in *SHH*. In the proband, we identified a 12,041 base pair (bp) homozygous deletion (chr7:156,816,030-156,828,070; hg38) that overlaps *LMBR1* exon 4 along with two base pairs (CA) that were inserted at the breakpoint (**Fig. 1b-d**, **Supplementary Fig. 1**). Both unaffected parents are heterozygous for this deletion (**Fig. 1b-c**, **Supplementary Fig. 1**). We reported this deletion in the Decipher database[23] (#411659) and did not identify any overlapping homozygous deletions in control databases[24]. To further validate our WGS results, we carried out both PCR analyses around the breakpoint (**Fig. 1c**) and Sanger sequencing of the breakpoint (**Fig. 1d**, **Supplementary Fig. 1**), the results of which corroborated our findings.

As homozygous deletions of the ZRS, which regulates *SHH* expression in the developing limb, were shown to lead to truncated limbs in mice and humans[15,16], we carried out detailed sequence analysis of this enhancer. WGS and Sanger sequencing analyses of the ZRS (chr7:156,790,916- 156,792,095; hg38) in the proband affected with acheiropodia did not reveal any rare variants in this enhancer. We did observe a homozygous single nucleotide polymorphism (SNP) rs10254391 (chr7:156,791,873; hg38) in the proband and that was heterozygous in both parents. As this SNP has a minor allele frequency of 0.26 in the global population, has been reported to be homozygous

136 in around 1,702 cases in GnomAD[25] and is thought to be benign based on the ClinVar database[26],

137 we concluded that it is not likely to be causative of this phenotype. Our results strongly suggest that

138 the acheiropodia in the proband is likely caused by the 12 kb homozygous deletion.

139

140 **The 12 kb deleted region does not function as a limb developmental enhancer**

141 To test whether this region functions as a developmental limb enhancer, we tested its ability to

142 drive limb expression in mouse embryos. We amplified this 12 kb sequence from a human BAC

143 (RP11-155D20), cloned it into the *Hsp68*-LacZ vector, that contains an *Hsp68* minimal promoter

144 followed by the LacZ reporter gene[27], and injected it into one-cell mouse embryos (**Fig. 2**).

145 Transgenic embryos were harvested at E11.5, a time point that is critical for *Shh* limb expression in

146 the zone of polarizing activity (ZPA[6,16]). We obtained six LacZ PCR positive embryos, three not

147 showing any LacZ expression whatsoever and three having inconsistent LacZ expression, none of

148 which have expression in the ZPA (**Fig. 2**). Previous studies have tested ZRS human

149 sequences/mutations in mice using this assay, finding LacZ expression in the ZPA[28,29]. We also

150 checked this 12 kb region for the presence of various histone modifications indicative of enhancer

151 activity from ENCODE[30] genomic data. Analysis of 18 different cell types found only a poised

152 enhancer mark, H3K4me1, in two of the cell lines, K562 and A549 (**Supplementary Fig. 2**).

153 Combined, these results strongly suggest that this 12 kb region does not function as an enhancer

154 in general and more specifically in the ZPA at E11.5.

155

156 **The 12 kb deletion leads to altered CTCF/RAD21 distribution**

157 We next analyzed the 12 kb deleted region for potential functional entities that could lead to the

158 acheiropodia phenotype. While it overlaps exon 4 of the *LMBR1* gene, mouse knockouts of this

159 gene do not have any apparent limb phenotype[4], the numerous mutations that were identified in it

160 in humans and mice are thought to lead to limb malformations due to altering ZRS copy number or

161 sequence[12] and an acheiropodia phenotype was observed in both *Shh* and ZRS homozygous

162 mouse knockouts[10,16] and homozygous ZRS deletion in humans[15]. We reasoned that the likely

163 cause of the acheiropodia in this proband is altered *SHH* expression during limb development.

164 Analysis of ENCODE[30] ChIP-seq datasets identified three CTCF-bound sites in this region, named

165 here as *LMBR1-SHH* CTCF (LSC) sites 3-5. These three CTCF-bound sites appear in numerous

166 ChIP-seq assays (LSC3: 118/191, LSC4: 97/191, LSC5: 139/191) from various human cell lines,

167 strongly suggesting that they are functional (**Supplementary Fig. 3**, **Supplementary Table 2**). As

168 CTCF motif orientation was shown to be important in determining the positioning of chromatin

169 looping[31], we next analyzed the orientation of these sites. We found that all three sites are in

170 convergent orientation to the *SHH* gene (**Fig. 3a**). We thus speculated that this 12 kb region may

171 function as a scaffolding region, enabling ZRS to interact with the *SHH* promoter.

172

173 To test whether this sequence functions as a scaffolding region, we carried out ChIP-seq for both

174 CTCF and RAD21, a member of the cohesin complex that along with CTCF is known to determine

175 chromatin looping[32]. ChIP-seq was done for both proteins using proband and wild type

176 lymphoblastoid cells. It is important to note that these cells were established using an Epstein Barr

177 virus which could affect our subsequent genomic studies. As a previous study[21] indicated that the

178 interaction between *Shh* and ZRS is "tissue-invariant", we reasoned that these cells could be used

179 for these analyses. We also checked the mRNA expression of *SHH* in the wild type and mutant

180 cells, observing overall low expression levels that were significantly higher in wild-type versus

181 proband cells (**Supplementary Fig.4**). In the wild type cells, we observed three CTCF ChIP-seq

182 peaks (LSC3-5) that have sites in convergent orientation to *SHH* and correspond to those found in

183 the ENCODE datasets (**Fig. 3a**-**b**, **Supplementary Fig. 5-6**). For RAD21, we also observed

184 binding in the 12 kb region, in particular at the LSC3 site (**Fig. 3a**-**b**, **Supplementary Fig. 5-6**). In

185 the proband's cells, we did not observe the CTCF and RAD21 peaks due to the 12 kb deletion.

186 Instead, we observed a novel RAD21 and CTCF peak in convergent orientation to *SHH* (LSC2)

187 near exon 6 of *LMBR1* that does not appear in wild type cells (**Fig. 3a**-**b**, **Supplementary Fig. 5-6**).

188

189 We next analyzed the CTCF motif scores of LSC1-5 to assess whether they could be associated

190 with the appearance of the novel CTCF binding (LSC2) observed in the proband's cell line. We

191 used the Find Individual Motif Occurrences (FIMO[33]) tool to assign motif scores for all five sites. We

192 extracted the CTCF motifs with a p-value threshold of 0.001 genome-wide and only picked motifs in

193 the *SHH-LMBR1* locus that overlapped CTCF peaks in our ChIP-seq. For LSC3-5, we observed

194 motif scores, determined by the weights at the corresponding position weight matrix summing up to

195 16, 5.2 and 17 respectively, while for LSC2 we obtained a score of 11 (**Fig. 3c**). This suggests that

196 with the loss of LSC3-5 due to the deletion, CTCF might bind to the LSC2 weaker binding affinity

197 motif instead of LSC3-5.

198

199 **The 12 kb deletion impairs the interaction between ZRS and the *SHH* promoter**

200   To examine whether the chromatin interaction between the ZRS and the *SHH* promoter are altered

201   due to the 12 kb deletion, we performed 4C-seq using the *SHH* promoter as a viewpoint. 4C-seq

202   was performed on both proband and wild type lymphoblastoid cell lines using standard methods[34]

203   (see Methods). In wild type cells, we observed that the *SHH* promoter strongly interacts with LSC1

204   and LSC3-5 (**Fig. 4a**, **Supplementary Fig. 5-6**). For the proband's cells, we did not observe

205   interactions with the ZRS and instead saw increased interactions between the *SHH* promoter and

206   LSC1 (**Fig. 4b**, **Supplementary Fig. 5-6**). Interestingly, in wild type cells we observed a weak

207   interaction with the ZRS compared to a much stronger interaction between LSC3-5 and the *SHH*

208   promoter (**Fig. 4b**, **Supplementary Fig. 5-6**). We also analyzed published CTCF Hi-ChIP data

209   from human GM12878 lymphoblastoid cells[35] and observed a much more robust interaction

210   between the *SHH* promoter and the 12 kb region compared to the ZRS (**Supplementary Fig.7**).

211

212   DNA FISH was also carried out on both proband and parental lymphoblastoid cell lines to

213   investigate chromosome conformation changes in an allele-specific manner using probes targeting

214   the *SHH* promoter, LSC1, LSC2 and LSC3-5 (**Fig. 5a**). To distinguish between the wild type and

215   mutant alleles in the parental cell lines, we used a plasmid containing the 12 kb acheiropodia-

216   associated region (**Fig. 5b**). LSC1 was found to be significantly closer to the *SHH* promoter on the

217   mutant allele compared to the wild type chromosome, suggestive of an increased interaction

218   between LSC1 and the *SHH* gene. We also observed that the novel LSC2 peak identified in the

219   proband's cell line is not found closer to the *SHH* promoter when comparing the wild type allele to

220   the mutant (**Fig. 5c**). However, we did observe a significant increase in the distance between the

221   *SHH* promoter and the region containing the 12 kb deletion that was specific to the mutant allele

222   consistent with the loss of interactions observed by 4C-seq (**Fig. 4**). These results further suggest

223   that the 12 kb acheiropodia-associated region functions as a scaffolding region between the ZRS

224   and the *SHH* promoter and its deletion impairs this interaction.

225

226   **Removal of the acheiropodia-associated region in mice does not lead to an observable**

227   **phenotype**

228   To further assess the function of this sequence in mice, we generated a mouse knockout of the

229   orthologous 12 kb acheiropodia-associated region. Using the liftOver tool in the UCSC Genome

230   Browser[36] (see methods) the human 12 kb acheiropodia-associated sequence was converted to its

231   orthologous mouse sequence (chr5:29,335,354-29,348,393; mm10). Of note, using FIMO[33] we

observed that mice have eight CTCF motifs in this orthologous region while humans have four, and only one of the eight overlapped mouse limb CTCF ChIP-seq data[37,38] (**Fig. 6a**). We also analyzed developing mouse embryonic limb (E10.5-E15.5) ChIP-seq datasets for various histone modifications (H3Kme1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3, H3K36me3) and ATAC-seq from ENCODE[30] and did not observe any peaks overlapping this region (**Supplementary Fig. 8**). Previous deletions of various CTCF sites in this region in mice did not show any apparent limb malformations[4,21]. However, these deletions did not cover the acheiropodia-associated region (**Fig. 6a**). We generated a knockout mouse which harbors the orthologous 12 kb acheiropodia-associated deletion along with additional sequence due to sgRNA selection constraints (chr5:29,334,962-29,348,393; mm10). Mouse knockouts were generated using the improved-Genome editing via Oviductal Nucleic Acids Delivery (*i*-GONAD[39]) technique. Founder mice and germ line transmission in F1 offspring with the desired deletion were validated by PCR, Sanger sequencing and Southern blot (**Supplementary Fig. 9**). We focused our subsequent phenotypic analyses on mouse line 517 that had a single nucleotide T insertion within the deleted region (**Supplementary Fig. 9a**).

To determine the functional effect of the deletion, we generated homozygous mice and phenotyped them using qRT-PCR, whole mount *in situ* hybridization (WISH) and alizarin red/alcian blue skeletal staining. Homozygous mice did not have any observable phenotype. qRT-PCR on E11.5 autopods from both forelimbs and hindlimb did not identify *Shh* expression changes between homozygous and wild type mice (**Fig. 6b**). WISH for *Shh* did not identify any changes in expression between homozygous and wild type E11.5 embryos (**Fig. 6c**). Finally, we checked the limb skeletal structure at E18.5 using alizarin red/alcian blue staining finding no apparent abnormalities in the homozygous embryos (**Fig. 6d**). These results highlight that mice are not an appropriate model to test the chromosomal interactions in humans for this region, likely due to the differences in CTCF site distribution and orientation. In addition, they also suggest that removal of *Lmbr1* exon 4 does not lead to a limb-associated phenotype in mice.

Discussion

We identified a 12 kb homozygous deletion that is associated with acheiropodia. We show that this 12 kb region does not have enhancer activity at mouse E11.5. Our CTCF and RAD21 ChIP-seq data indicate that this region has three CTCF binding sites along with RAD21 binding (**Fig. 3b**). Chromatin interaction analyses of this region suggests that it functions as a scaffolding region between the ZRS and the *SHH* promoter via three CTCF sites (LSC3-5). In the cells from the proband with acheiropodia, these sites are deleted and this interaction is substituted with another CTCF site (LSC1) centromeric to the ZRS. Due to this change in interaction, the ZRS does not interact with the *SHH* promoter (**Fig. 7**). Deletion of the orthologous region in mice did not lead to an observable phenotype, likely due to the inherent chromatin interaction and CTCF distribution differences between humans and mice in this region.

Our work suggests there are substantial differences in the regulation of chromosomal interactions linking ZRS to the *SHH* promoter between humans and mice. There are two previous reports that generated CTCF site-specific deletions around the ZRS in mouse[4,21]. Paliou *et al.* [21] deleted the i4[#] or i5[#] CTCF sites (**Fig. 6a**) individually or together and observed no major limb phenotype even though a 51% reduction of *Shh* expression was observed in E10.5 limb buds when both CTCF sites were deleted. Following the deletion of these two CTCF sites (i4[#] and i5[#]), ectopic CTCF sites also appeared, one within the ZRS (ZRS[#]) and the other near the transcription start site (TSS) of *Lmbr1* (termed here as i3[#]), both of which do not overlap our observed ectopic CTCF site, LSC2 (**Fig. 6a**). To characterize the function of the ectopic CTCF sites, three sites (i4[#], i5[#] and ZRS[#]) were deleted, leading to a depletion of all CTCF (including i3[#]) and RAD21 binding around the ZRS and significantly decreasing the interaction between *Shh* and ZRS. Although these triple deletions led to a 52% reduction of *Shh* expression in E10.5 limb buds, no limb abnormalities were observed. Williamson *et al.*[4] deleted three different CTCF sites individually around the *Shh* gene and ZRS (CTCF3*, 4*, and 5*; **Fig. 6a**). Mice homozygous for each deletion did not show an observable limb phenotype. They also generated a homozygous 35 kb deletion that contains CTCF4*, i3[#], 5* and the *Lmbr1* TSS and promoter and did not observe *Shh* gene expression changes in E11.5 limb buds measured by qRT-PCR and any apparent limb abnormalities. Of note, as previously mentioned, these results also suggest that *Lmbr1* itself is not necessary for limb development, as also observed in our 12 kb knockout mice. Combined, these CTCF mouse deletion studies,

10

including our study, imply that ZRS-*Shh* interactions are likely to be robust to individual or even triple CTCF perturbations. They also suggest that other CTCF sites, either those that were not tested in these studies or ones that appear ectopically following these manipulations, keep this interaction intact.

The phenotypic differences between human and mice are likely due to several factors including differences in CTCF location, motif score and orientation. In terms of location, humans have three CTCF ChIP-seq peaks (LSC3-5) in the deleted region; however, mouse has one CTCF peak in the orthologous 12 kb acheiropodia deleted sequence (i4[#]) and this region does not show strong evolutionary conservation between humans and mice (**Fig. 6a**). Our work suggests that these three human CTCF sites (LSC3-5) play a role as an anchor/scaffolding region for the interaction between ZRS and the *SHH* promoter. In mice, i4[#] and i5[#] likely play this role and their relative distal position between one another might be important for robustness. Analysis of previously published 4C-seq from E10.5 mouse limbs[21] showed interactions between the *Shh* promoter and i9[#], CTCF3*/i5[#] and ZRS[#] but not with the 12 kb acheiropodia-associated region (**Supplementary Fig. 10**). These CTCF sites could be working cooperatively to maintain the interaction between ZRS and the *Shh* promoter. For LSC1, where we observed increased interactions with the *SHH* promoter in the proband, CTCF ChIP-seq from ENCODE[30] has 161/191 assays showing CTCF-bound sites in this region while in ChIP-seq datasets from various mouse tissues/cells we only observed about half of the assays to have a peak in this region (**Supplementary Fig. 11**). Interestingly, for LSC2, while we only observed a CTCF ChIP-seq peak in the proband, likely due to compensation for the deletion of LSC3-5, in mice the homologous CTCF site, 3*/i5[#], shows a strong CTCF ChIP-seq peak in wild-type E10.5, E13.5 and E14.5 limbs (**Fig. 6a**) and in ENCODE[30] mouse datasets (**Supplementary Fig. 11**). Analyses of CTCF motif scores for LSC2 versus 3*/i5[#] shows a weaker score for LSC2 (**Supplementary Fig. 12a**), and this CTCF site also has a weak interaction with the *SHH* promoter (**Fig. 4**). Weaker CTCF sites might serve as a backup for failed enhancer-promoter interactions. Correspondingly, analyses of ZRS[#] and i3[#] ectopic CTCF sites which appeared in the i4[#] and i5[#] double deletion mice found them to have lower motif scores than i4[#] and i5[#] (**Supplementary Fig. 12b**).

In terms of orientation, all three CTCF sites in the 12 kb acheiropodia-deleted region (LSC3-5) are in convergent orientation to *SHH* along with LSC1 and LSC2, while the sites telomeric to LSC5,

324   LSC6 and LSC7, are in divergent orientation or both (**Fig 6a**). In mice, CTCF site 4* that is

325   homologous to LSC6 and 5* which is homologous to LSC7 are all in divergent orientation to *Shh*,

326   but i3[#] is in convergent orientation (**Fig 6a**). Carrying out a more global analysis of human and

327   mouse CTCF ChIP-seq peaks that compared human K562 to mouse CH12 cells, both

328   lymphoblasts, shows that only around 25% of the peaks overlap when converting their coordinates

329   to mouse or vice versa (**Supplementary Fig. 13**). These results are consistent with a recent report

330   that also analyzed the overlap of CTCF ChIP-seq peaks between these cells (K562 and CH12)

331   plus human GM12878 and mouse MEL cells[40]. This suggests that there are major differences

332   between human and mouse in terms of CTCF location. These differences could be due to various

333   selection pressures, proving more safeguards for enhancer-promoter interactions. It will also be

334   intriguing to test whether these changes in CTCF location and orientation could be involved in

335   phenotypic differences between species. Taken together, our results highlight that mouse is not a

336   useful model to assess the chromatin interactions in humans for this locus and that CTCF location,

337   orientation and number needs to be assessed between human and mice before using mice as an

338   animal model to dissect human nucleotide variation that affects CTCF binding.

339

340   The 12 kb acheiropodia-associated deleted region resides close to a topologically associated

341   domain (TAD) boundary that encompasses both *SHH* and *LMBR1*. Previous mouse genetic studies

342   have shown that TAD boundary alterations could alter chromatin interactions and lead to ectopic

343   gene expression[31,41]. While we cannot definitively exclude that this deletion is associated with TAD

344   boundary alterations, using the 3D Genome Browser[42], we have analyzed this TAD boundary in Hi-

345   C datasets from ten different human cell lines, finding that in all of them the boundary does not

346   overlap this 12 kb deleted region. We observed two different locations for this boundary that differ

347   between cell types. For five of the cell lines (HepG2, GM12878, NHEK, K562 and HMEC), this

348   boundary is thought to be located around the *LMBR1* TSS while for five other cell lines (H1-ESC,

349   G401, A549, epidermal keratinocyte and hippocampus) the boundary is estimated to be around the

350   transcription termination site of the DnaJ heat shock protein family (Hsp40) member B6 (*DNAJB6*)

351   gene (**Supplementary Fig. 14**). Human and mice TAD boundaries were shown to be relatively

352   conserved[43]. In mice, the *Shh-Lmbr1* TAD boundary resides around the *Lmbr1* TSS[44], similar to

353   what is observed in humans for five out of the ten cell lines. Symmons *et al.*[44] inverted a 450 kb

354   region (300 kb downstream and 150 kb upstream of *Lmbr1*) that contains this boundary. This led to

355   a complete loss of *Shh* expression in the ZPA and a limb truncation phenotype, similar to the ZRS

356 homozygous knockout. 4C-seq analysis revealed that the ZRS-*Shh* interaction was disrupted in

357 this inversion, further suggesting that altering this interaction can lead to an acheiropodia like

358 phenotype. In summary, while we cannot conclusively rule out that alteration of the TAD boundary

359 is responsible for this phenotype, our results strongly suggested that removal of these CTCF sites

360 in humans alters the interaction between ZRS and the *SHH* promoter, likely leading to the

361 acheiropodia phenotype.

362

363 CTCF plays a major role in enhancer-promoter interactions, facilitating transcriptional activity by

364 establishing chromatin loops between these elements[45,46]. However, only a small number of

365 genetic diseases where CTCF site-specific mutations lead to alterations of these enhancer-

366 promoter interactions have been reported. CTCF site-specific deletions were shown to be

367 associated with imprinting in the *IGF2/H19* locus, causing Beckwith-Wiedemann syndrome (BWS;

368 OMIM 130650). A 1.8 kb deletion that removes two CTCF sites in the normal imprinted and

369 silenced *IGF2* expression in the maternal allele was shown to lead to hypermethylation and biallelic

370 expression of *IGF2* and is thought to cause BWS[47]. Several reports have associated somatic

371 mutations in CTCF sites with various cancers[48,49]. Interestingly, analysis of somatic mutations from

372 the International Cancer Genome Consortium database[50] revealed that numerous mutations

373 overlap human stem cell CTCF loop anchors[51], suggesting that aberrant chromatin interactions

374 could be strongly associated with cancer. To our knowledge, this study is the first report a CTCF

375 mutation that is associated with a Mendelian condition. With WGS becoming more commonly used

376 in the clinic, it would be interesting to analyze disease-associated variants, in particular short indels,

377 for their overlap with CTCF motifs and chromatin interactions. In addition, our study shows that due

378 to differences in CTCF site location, motif sequence and orientation animal models may not be a

379 good proxy to analyze the effects of CTCF site variation. As more human genomes are sequenced

380 and the genomes of additional species become available, it will be important to consider the

381 phenotypic effects of nucleotide changes in CTCF sequences on disease and evolution.

382

Methods

Patient sample collection

The study was approved by the ethical committee of the University of California San Francisco, protocol number 10-03111, Comitê de Ética em Pesquisa da Prefeitura de Porto Alegre (Plataforma Brasil) protocol number 1.103.654 and the Brazilian Research Ethics Commission (CONEP) protocol number 223.811. Samples were obtained after receipt of informed consent. Genomic DNA was extracted from saliva using standard techniques. Blood samples were collected using standard techniques and used for the generation of lymphoblastoid cell lines. Clinical data were obtained from a physician examination and review of medical records.

Establishment of lymphoblastoid cell line and culture

Blood samples from the proband and parents were spun over Ficoll-Paque (Amersham Biosciences) gradients to enrich the sample for mononuclear cells. Epstein Bar virus (EBV)-transformed lymphoblastoid lines were generated from isolated peripheral blood lymphocytes. Briefly, cells were washed and resuspended in complete Iscove's modified Dulbecco's culture media supplemented with 10% v/v fetal bovine serum, antibiotics, and virus. The B95-8 EBV-infected marmoset cell line (ATCC, catalog no. CRL-1612) was used as the source for viral stocks. High molecular weight DNA was isolated from Ficoll-Paque enriched mononuclear cells using standard desalting procedures. Lymphoblastoid cells were maintained in RPMI1640 medium (Life Technologies, catalog no. 11875093) containing 15% fetal bovine serum (FBS) and penicillin-streptomycin.

Whole genome sequencing

Whole genome sequencing was performed at the University of Washington Center for Mendelian Genomics (University of Washington, Seattle). Initial quality control (QC) entailed DNA quantification, gender validation assay and molecular fingerprinting with a 63-SNP OpenArray assay derived from a custom exome SNP set. Following successful QC, at least 750 ng of genomic DNA was subjected to a series library construction steps utilizing the KAPA Hyper Prep kit (Roche), automated on the Perkin Elmer Janus platform. Libraries were validated using the Bio-Rad CFX384 Real-Time System and KAPA Library Quantification kit (Roche). Samples were sequenced on a HiSeq X using Illumina's HiSeq X Ten Reagent Kit (v2.5) to an average depth of 30X. Burrows-Wheeler Aligner[52], Genome Analysis ToolKit[53] and SeattleSeq Annotation server build 138

416   (https://snp.gs.washington.edu/SeattleSeqAnnotation138/) were used to generate BAM, vcf and

417   annotation files, respectively. Homozygosity mapping was performed with PLINK v1.07 software[54]

418   using the genotypes generated by the 63-SNP OpenArray assay. Structural variants were called

419   using Lumpy[55]. Alignments were visualized using the Integrative Genomics Viewer tool[56]. The

420   *LMBR1* deletion and ZRS variants were validated by PCR-Sanger sequencing (primers provided in

421   **Supplementary Table 3**).

422

Mouse transgenic enhancer assays

424   Mouse work was approved by the UCSF Institutional Animal Care and Use Committee (IACUC),

425   protocol number AN100466, and was conducted in accordance with AALAC and NIH guidelines.

426   The 12 kb acheiropodia associated region was amplified from a human BAC (RP11-155D20) by

427   PCR, cloned it into the *Hsp68*-LacZ vector[27] and sequence verified. All LacZ transgenic mice were

428   generated by Cyagen Biosciences using standard procedures[57], and harvested and stained for

429   LacZ expression at E11.5 as previously described[58]. Pictures were obtained using an M165FC

430   stereo microscope and a DFC500 12-megapixel camera (Leica).

431

ChIP-seq

433   Lymphoblastoid cells were plated on two different flasks and used for the experiment as

434   independent tubes considered as two technical replicates. Cells (1×10^7 cells) were fixed in

435   phosphate-buffered saline (PBS) with 0.96% formaldehyde for 8 minutes at room temperature.

436   Crosslinking was quenched with 125 mM Glycine. The cells were washed with PBS and

437   precipitated via centrifugation. The cell pellet was stored in -80°C until use. The pellet was lysed in

438   240 μL of Buffer B (LowCell# ChIP kit; Diagenode, catalog no. C01010072) and lysed chromatin

439   was sheared using a Covaris S2 sonicator to obtain on average 250 bp size fragments. ChIP was

440   performed using the LowCell# ChIP kit according to manufacturer's protocol with modifications. 120

441   μL of sheared chromatin was mixed with 880μL of Buffer A (LowCell# ChIP kit) supplemented with

442   complete protease inhibitor (Sigma-Aldrich, catalog no. 11873580001). 80μL of the solution was

443   saved as input control. To obtain magnetic bead-antibody complexes, 22μL of protein A-coated

444   paramagnetic beads (LowCell# ChIP kit) were washed twice with Buffer A (LowCell# ChIP kit) and

445   resuspended in 22 μL of Buffer A. 10μL of magnetic beads were mixed with 90μL of Buffer A

446   (LowCell# ChIP kit) and 6 μg antibody (final antibody concentration was 60 ng/μL in the binding

447   reaction). This mixture was gently agitated at 4°C for 2 hours. Antibody against CTCF (Active Motif,

448 catalog no. 61311) or RAD21 (Abcam, catalog no. ab992) was used for immunoprecipitation

449 respectively. The bead-antibody complex was precipitated with a magnet and the supernatant was

450 removed. 800 mL of shared chromatin was added to the bead-antibody complex and rotated at 4°C

451 overnight. The beads were then washed with Buffer A three times and Buffer C once. DNA was

452 purified using IPure kit v2 (Diagenode, catalog no. C03010015) according to the manufacturer's

453 protocol. Sequencing libraries were generated using the Accel-NGS 2S Plus DNA Library Kit (Swift

454 Biosciences, catalog no. 21024) according to manufacturer's protocol. Massively parallel

455 sequencing was performed on an Illumina HiSeq4000 with 50 bp single-end read. ChIP-seq was

456 done with two technical replicates. ChIP-seq data was analyzed following the ENCODE

457 transcription factor pipeline[59]. Both RAD21 and CTCF ChIP-seq raw reads were mapped against

458 the human genome (GRCh37; hg19) using bowtie2 (v2.2.6). Duplicate reads were marked using

459 Picard (v1.126) MarkDuplicates and multimapping, low quality, duplicated and non-properly paired

460 reads were removed. Library complexity measures and flagstats were generated for each BAM file.

461 BAM files were converted to tagAlign format and two subsampled pseudoreplicates were

462 generated for each sample with half the total reads. Reproducible peaks were identified using the

463 MASC2 (v2.1.1)[60] peak caller and the irreproducibility discovery rate (IDR (v2.0.4)) framework[59].

464 IDR analysis was performed using self-pseudoreplicates and the main samples to obtain self-

465 consistent sets of peaks. Final peak calls were filtered using the ENCODE blacklist[61] and an IDR of

466 2% with a signal value > 30. We combined replicates to obtain only highly reproducible peaks using

467 the IDR[59] and show them pooled in main figure and individual replicates in supplementary figures.

468 Differential enrichment analysis between the proband and wild type cells was performed by

469 DiffBind[62].

470

471 **CTCF motif analysis**

472 We used the position-weight matrix MA0139.1 from JASPAR[63] to scan for CTCF motifs. Genome-

473 wide CTCF motif identification was performed on the human (hg19) and mouse (mm9) genomes,

474 with FIMO[33], with a p-value threshold of 0.0001. For some CTCF peaks in the *SHH-LMBR1* locus

475 (LSC4, 6, and 7), when there was no CTCF motif that overlapped ChIP-seq peaks, we reduced the

476 p-value threshold to 0.001. CTCF orientation was determined by the strand in which the motif was

477 identified. For the human-mouse CTCF ChIP-seq comparisons, experiments from the cell lines

478 K562[30] and CH12[37] were analyzed. The UCSC Genome Broswer[36] liftOver tool was used with –

479 minMatch=0.01 to transfer the peak coordinates between the two species followed by BEDtools[64]

480 to intersect them and calculate the proportion of overlapping peaks.

481

482 **4C-seq**

483 Lymphoblastoid cells were plated on two different flasks and used for the experiment as

484 independent tubes deemed as two technical replicates. 4C-seq was performed using standard

485 procedures[34]. Briefly, 1×10^7 cells were fixed in PBS with 2% formaldehyde for 10 minutes at room

486 temperature. Crosslinking was quenched with 125 mM Glycine. The cells were precipitated via

487 centrifugation and resuspended in lysis buffer (50 mM Tris-HCl pH7.5, 150 mM NaCl, 5 mM EDTA,

488 0.5% NP-40, 1.15% Triton X-100, 1x complete proteinase inhibitors (Roche, catalog no.

489 11697498001)) and incubated for 10 minutes on ice. They were then precipitated via centrifugation

490 and washed with PBS. The cell pellet was stored in -80°C until use. The cell pellet was suspended

491 in *Dpn*II restriction enzyme buffer and treated with 0.3% SDS and 2.5% Triton X100 at 37°C for 1

492 hour, respectively. Chromatin was digested with 100 units of *Dpn*II (New England Biolabs, catalog

493 no. R0543) at 37°C for 3 hours. An additional 100 unit of *Dpn*II was added and the reaction was

494 incubated at 37°C overnight. After heat inactivation of the enzyme, 50 units of T4 DNA ligase

495 (Roche, catalog no. 10799009001) were applied for self-ligation of the digested chromatin and

496 placed for incubation at 16°C overnight. After purification of DNA using phenol-chloroform and

497 ethanol precipitation, DNA was digested with 50 units of *Nla*III (New England Biolabs, catalog no.

498 R0125) at 37°C overnight. Following heat inactivation of the enzyme at 65°C, 25 µg of DNA was

499 used for the second ligation reaction with 50 units of T4 DNA ligase at 16°C overnight. After

500 purification of DNA using phenol-chloroform and ethanol precipitation, the inverse PCR was

501 performed using NEBNext high-fidelity 2X PCR master mix (New England Biolabs, catalog no.

502 M0541). DNA was purified with AMPure XP beads (Beckman Coulter, catalog no. A63881). The

503 second round of PCR was performed using NEBNext high-fidelity 2X PCR master mix to attach

504 library adapters and index sequences. All PCR primer sequences are listed in **Supplementary**

505 **Table 3**. DNA was purified with the QIAquick PCR purification kit (Qiagen, catalog no. 28104).

506 Massively parallel sequencing was performed on an Illumina HiSeq4000 with 50 bp single-end

507 reads using a custom primer (**Supplementary Table 3**). 4C-seq was carried out using two

508 technical replicates. 4C-seq data was analyzed using the 4C-seq pipeline[65]. Briefly, 4C-seq raw

509 reads were trimmed to 50 bp with cutadapt 2.4. Valid 4C-seq reads containing 4C reading primer

510 were extracted from fastq file and parsed into raw.txt file aligned against the restriction-enzyme

511 digested genome GRCh37(hg19) using 4Cseqpipe version 0.7[65]. Raw files were translated into

512 final graphical depictions of contact profiles around viewpoints using 4Cseqpipe version 0.7[65].

513

514 **DNA FISH**

515 For DNA FISH, 0.5-1×10^6 lymphoblastoid cells were seeded on Poly-prep slides (Sigma) overnight.

516 They were then fixed in 4% paraformaldehyde for 10 minutes at room temperature and

517 permeabilized using 0.5% Triton X for 10 minutes[66]. Fosmid clones and plasmid were prepared

518 and labelled as previously described[67]. Cells were denatured for 30 minutes. For four-color FISH,

519 each slide was hybridized with between 80-100 ng of biotin-, digoxigenin- and directly labelled

520 probes, 18 µg of human Cot1 DNA (Invitrogen) and 5 µg salmon sperm DNA. Green496-dUTP

521 (Enzo Life Sciences) was used for direct labelling of fosmid probes. Washes and detection were as

522 previously described[67]. See **Supplementary Table 3** for Fosmid probe details.

523

524 Slides were imaged using a Photometrics Coolsnap HQ2CCD camera and a Zeiss AxioImager A1

525 fluorescence microscope with a Plan Apochromat 100×1.4NA objective, a Nikon Intensilight

526 Mercury based light source and either Chroma #89014ET (three-colour) or #89000ET (four-colour)

527 single excitation and emission filters (Chroma Technology Corp.) with the excitation and emission

528 filters installed in Prior motorized filter wheels. A piezo electrically driven objective mount (PIFOC

529 model P-721, Physik Instrumente) was used to control movement in the z dimension. Step size for

530 z stacks was set at 0.2 µm. Nikon Nis-Elements software was used to perform hardware control,

531 image capture and analysis. Images were deconvolved using a calculated point spread function

532 with the constrained iterative algorithm of Volocity (PerkinElmer). The quantitation module of

533 Volocity was used to calculate inter probe distances. To eliminate the possibility of measuring sister

534 chromatids, only alleles with single probe signals were analyzed.

535

536 **Generation of knockout mice**

537 Mouse work was approved by the UCSF IACUC, protocol number AN100466, and was conducted

538 in accordance with AALAC and NIH guidelines. The 12 kb acheiropodia associated sequence

539 (chr7:156,608,724-156,620,764; hg19) was converted to mouse sequence (chr5:29,335,354-

540 29,348,393; mm10) using the UCSC Genome Broswer[36] liftOver tool. Two gRNA were designed to

541 target the 5' and 3' ends of this region (**Supplementary Table 3**) using the gRNA design tool on

542 the Integrated DNA Technologies (IDT) website and selected based on low off-target and high on-

543   target scores. The acheiropodia deletion allele was generated using *i*-GONAD[39]. Briefly, after

544   reconstitution of two crRNA (IDT) and tracrRNA (IDT), these were mixed together (final

545   concentration 100 µM each) and incubated at 92°C for 2 minutes and left at room temperature for

546   10 minutes to prepare the crRNA/tracrRNA complex. The genome-editing mixture, (30 µM

547   crRNA/tracrRNA complex, 1 mg/ml Cas9 protein (IDT), Opti-MEM) was incubated at 37°C for 10

548   minutes. Estrus female FVB mice (Jackson Laboratory, catalog no. 001800) were mated to male

549   mice the night before. Presence of copulation plugs was confirmed by visual inspection the next

550   morning and the females having plugs were designed as Day 0.5 of gestation at noon and Day 0.7

551   of gestation at 16:00. Females on Day 0.7 were used for oviduct electroporation. Mice were

552   anesthetized using isoflurane, the ovary and oviducts were exposed by grasping the adipose tissue

553   surrounding the ovary. Approximately 1-2 µl of the genome-editing mixture was injected into the

554   oviduct lumen upstream of the ampulla using a micropipette. Immediately following injection, the

555   oviduct was covered with a piece of wet paper soaked in PBS and then grasped by tweezer-type

556   electrodes (Bulldog Bio, catalog no. CUY652P2.5 X4). The electroporation was performed using a

557   square-wave pulse generator BTXECM830 (BTX Genetronics Inc.). The electroporation conditions

558   used were 8 pulses of 50 V at 5 mseconds wave length. After electroporation, the oviducts were

559   placed in their original position, and the muscle layer incision was sutured using absorbable suture

560   chromic gut. The coat layer incision was closed by AutoClip kit (Fine Science Tools, catalog no.

561   12022-09). The animals were kept on a warming pad (37°C) during surgery and monitored for

562   anesthesia recovery following surgery.

563

564   **Sanger sequencing and Southern blot**

565   PCR-Sanger sequencing (primers provided in **Supplementary Table 3**) was preformed using

566   standard techniques[28]. For Southern blot analyses, genomic DNA were treated with *Bst*XI (New

567   England Biolabs, catalog no. R0113) and fractionated by agarose gel electrophoreses. Following

568   capillary transfer onto nylon membranes, blots were hybridized with Digoxigenin (DIG)-labeled

569   DNA probes (corresponding to chr5:29348565-29349037; mm10) amplified by the PCR DIG Probe

570   Synthesis Kit (Sigma-Aldrich, catalog no. 11636090910). The hybridized probe was

571   immunodetected with anti-digoxigenin Fab fragments conjugated to alkaline phosphatase (Sigma-

572   Aldrich, catalog no. 11093274910) and visualized with a CDP star (Sigma-Aldrich, catalog no.

573   11685627001) according to the manufacturer's protocol. Chemiluminescence was detected using

574   the FluorChem E (ProteinSimple, catalog no.92-14860-00).

RT-qPCR

Total RNA was collected from E11.5 limb buds or lymphoblastoid cells using TRIzol (Thermo

Fisher Scientific, catalog no. 15596026) and converted to cDNA using ReverTra Ace qPCR-RT

master mix with genomic DNA (gDNA) remover (Toyobo, catalog no. FSQ-301). qPCR was

performed using SsoFast EvaGreen supermix (Bio Rad, catalog no. 1725205). Primer sequences

used for qPCR are shown in **Supplementary Table 3.**

Whole-mount *in situ* hybridization

Mouse E11.5 embryos were fixed in 4% paraformaldehyde. A plasmid containing mouse *Shh*

cDNA (GenScript, catalog no. OMu22903D) was used as template for DIG-labeled probes. Mouse

whole-mount *in situ* hybridization was performed according to standard procedures[68].

Bone and cartilage staining

Embryos were harvested at E18.5 and limbs were dissected out and used for staining. Alcian

blue/Alizarin red staining was performed according to standard procedures for late-gestation stage

embryos[69].

Analysis of CTCF Hi-ChIP data

Analysis of the CTCF Hi-ChIP[35] data and figure generation were done using the HiCExplorer[70].

Data Availability

hg19 and hg38 human reference genome is available from NCBI GenBank assembly

"GCA_000001405.1 "[https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/]", and

"GCA_000001405.15 [https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/]", respectively).

The deleted sequence information is available from Decipher database "#411659

[https://decipher.sanger.ac.uk/patient/411659/overview/general]". ChIP-seq and 4C-seq data are

available from the Gene Expression Omnibus under accession number "GSE155324

[https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE155324]". ENCODE data are available

from the "UCSC genome browser [https://genome.ucsc.edu/]". Hi-C datasets are available from the

"3D Genome Browser [http://www.3dgenome.org/]". The CTCF motif was obtained from "JASPAR

[http://jaspar.genereg.net/]". All other relevant data supporting the key findings of this study are

607   available within the article and its Supplementary Information files or from the corresponding author

608   upon reasonable request. A Source Data file accompanies the manuscript. A reporting summary

609   for this Article is available as a Supplementary Information file.

610

References

1. Freire-Maia, A., Freire-Maia, N., Morton, N. E., Azevêdo, E. S. & Quelce-Salgado, A. Genetics of acheiropodia (the handless and footless families of Brazil). VI. Formal genetic analysis. *Am. J. Hum. Genet.* **27**, 521–7 (1975).

2. Ianakiev, P. *et al.* Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. *Am. J. Hum. Genet.* **68**, 38–45 (2001).

3. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).

4. Williamson, I. *et al.* Developmentally regulated Shh expression is robust to TAD perturbations. *Development* **146**, (2019).

5. Briscoe, J. & Thérond, P. P. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14**, 416–29 (2013).

6. Büscher, D., Bosse, B., Heymer, J. & Rüther, U. Evidence for genetic control of Sonic hedgehog by Gli3 in mouse limb development. *Mech. Dev.* **62**, 175–82 (1997).

7. Platt, K. A., Michaud, J. & Joyner, A. L. Expression of the mouse Gli and Ptc genes is adjacent to embryonic sources of hedgehog signals suggesting a conservation of pathways between flies and mice. *Mech. Dev.* **62**, 121–35 (1997).

8. Lewis, P. M. *et al.* Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**, 599–612 (2001).

9. Zeller, R., López-Ríos, J. & Zuniga, A. Vertebrate limb bud development: moving towards integrative analysis of organogenesis. *Nat. Rev. Genet.* **10**, 845–858 (2009).

10. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407–13 (1996).

11. Roessler, E. *et al.* Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat. Genet.* **14**, 357–360 (1996).

12. VanderMeer, J. E. & Ahituv, N. cis-regulatory mutations are a genetic cause of human limb malformations. *Dev. Dyn.* **240**, 920–30 (2011).

13. Hill, R. E., Heaney, S. J. H. & Lettice, L. A. Sonic hedgehog: restricted expression and limb dysmorphologies. *J. Anat.* **202**, 13–20 (2003).

14. Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat. Rev. Genet.* **18**, 245–258 (2017).

643 15. Shamseldin, H. E. *et al.* Novel copy number variants and major limb reduction malformation:
644   Report of three cases. *Am. J. Med. Genet. A* **170A**, 1245–1250 (2016).

645 16. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range
646   cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of
647   the mouse limb. *Development* **132**, 797–803 (2005).

648 17. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–211
649   (2009).

650 18. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression
651   control. *Nat. Rev. Genet.* **20**, 437–455 (2019).

652 19. Chang, L.-H., Ghosh, S. & Noordermeer, D. TADs and Their Borders: Free Movement or
653   Building a Wall? *J. Mol. Biol.* **432**, 643–652 (2020).

654 20. Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A. & Mirny, L. A. Emerging
655   Evidence of Chromosome Folding by Loop Extrusion. *Cold Spring Harb. Symp. Quant. Biol.*
656   **82**, 45–55 (2017).

657 21. Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of Shh
658   during limb development. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 12390–12399 (2019).

659 22. Silveira, E. L. & Freire-Maia, A. Acheiropodia: new cases from Brazil. *Clin. Genet.* **54**, 256–7
660   (1998).

661 23. Firth, H. V *et al.* DECIPHER: Database of Chromosomal Imbalance and Phenotype in
662   Humans Using Ensembl Resources. *Am. J. Hum. Genet.* **84**, 524–33 (2009).

663 24. MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L. & Scherer, S. W. The Database of
664   Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic*
665   *Acids Res.* **42**, D986-92 (2014).

666 25. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in
667   141,456 humans. *Nature* **581**, 434–443 (2020).

668 26. Landrum, M. J. *et al.* ClinVar: improving access to variant interpretations and supporting
669   evidence. *Nucleic Acids Res.* **46**, D1062–D1067 (2018).

670 27. Kothary, R. *et al.* A transgene containing lacZ inserted into the dystonia locus is expressed in
671   neural tube. *Nature* **335**, 435–7 (1988).

672 28. Laurell, T. *et al.* A novel 13 base pair insertion in the sonic hedgehog ZRS limb enhancer
673   (ZRS/LMBR1) causes preaxial polydactyly with triphalangeal thumb. *Hum. Mutat.* **33**, 1063–6
674   (2012).

675   29.   VanderMeer, J. E. *et al.* A novel ZRS mutation leads to preaxial polydactyly type 2 in a
676   heterozygous form and Werner mesomelic syndrome in a homozygous form. *Hum. Mutat.* **35**,
677   945–8 (2014).

678   30.   ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human
679   genome. *Nature* **489**, 57–74 (2012).

680   31.   Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and
681   Enhancer/Promoter Function. *Cell* **162**, 900–10 (2015).

682   32.   Merkenschlager, M. & Nora, E. P. CTCF and Cohesin in Genome Folding and Transcriptional
683   Gene Regulation. *Annu. Rev. Genomics Hum. Genet.* **17**, 17–43 (2016).

684   33.   Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif.
685   *Bioinformatics* **27**, 1017–8 (2011).

686   34.   Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from
687   beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**,
688   17–32 (2020).

689   35.   Mumbach, M. R. *et al.* HiChIRP reveals RNA-associated chromosome conformation. *Nat.*
690   *Methods* **16**, 489–492 (2019).

691   36.   Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

692   37.   Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature*
693   **515**, 355–64 (2014).

694   38.   Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs
695   reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).

696   39.   Gurumurthy, C. B. *et al.* Creation of CRISPR-based germline-genome-engineered mice
697   without ex vivo handling of zygotes by i-GONAD. *Nat. Protoc.* **14**, 2452–2482 (2019).

698   40.   Diehl, A. G., Ouyang, N. & Boyle, A. P. Transposable elements contribute to cell and
699   species-specific chromatin looping and gene regulation in mammalian genomes. *Nat.*
700   *Commun.* **11**, 1796 (2020).

701   41.   Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic
702   Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).

703   42.   Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome
704   organization and long-range chromatin interactions. *Genome Biol.* **19**, 151 (2018).

705   43.   Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of
706   chromatin interactions. *Nature* **485**, 376–80 (2012).

707   44.   Symmons, O. *et al.* The Shh Topological Domain Facilitates the Action of Remote Enhancers
708   by Reducing the Effects of Genomic Distances. *Dev. Cell* **39**, 529–543 (2016).

709   45.   Dowen, J. M. *et al.* Control of Cell Identity Genes Occurs in Insulated Neighborhoods in
710   Mammalian Chromosomes. *Cell* **159**, 374–387 (2014).

711   46.   Arzate-Mejía, R. G., Recillas-Targa, F. & Corces, V. G. Developing in 3D: the role of CTCF in
712   cell differentiation. *Development* **145**, dev137729 (2018).

713   47.   Sparago, A. *et al.* Microdeletions in the human H19 DMR result in loss of IGF2 imprinting and
714   Beckwith-Wiedemann syndrome. *Nat. Genet.* **36**, 958–960 (2004).

715   48.   Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat. Genet.*
716   **47**, 818–821 (2015).

717   49.   Norton, H. K. & Phillips-Cremins, J. E. Crossed wires: 3D genome misfolding in human
718   disease. *J. Cell Biol.* **216**, 3441–3452 (2017).

719   50.   Zhang, J. *et al.* International Cancer Genome Consortium Data Portal--a one-stop shop for
720   cancer genomics data. *Database (Oxford).* **2011**, bar026 (2011).

721   51.   Ji, X. *et al.* 3D Chromosome Regulatory Landscape of Human Pluripotent Cells. *Cell Stem*
722   *Cell* **18**, 262–275 (2016).

723   52.   Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
724   *Bioinformatics* **25**, 1754–60 (2009).

725   53.   McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
726   next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).

727   54.   Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
728   linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).

729   55.   Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for
730   structural variant discovery. *Genome Biol.* **15**, R84 (2014).

731   56.   Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–6 (2011).

732   57.   Pu, X., Young, A. P. & Kubisch, H. M. Production of Transgenic Mice by Pronuclear
733   Microinjection. in 17–41 (2019). doi:10.1007/978-1-4939-8831-0_2

734   58.   Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding
735   sequences. *Nature* **444**, 499–502 (2006).

736   59.   Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE
737   consortia. *Genome Res.* **22**, 1813–31 (2012).

738   60.   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

739   61.   Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of
740   Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).
741   62.   Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical
742   outcome in breast cancer. *Nature* **481**, 389–393 (2012).
743   63.   Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor
744   binding profiles. *Nucleic Acids Res.* **48**, D87–D92 (2020).
745   64.   Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
746   features. *Bioinformatics* **26**, 841–2 (2010).
747   65.   van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA
748   interactions. *Nat. Methods* **9**, 969–72 (2012).
749   66.   Eskeland, R. *et al.* Ring1B Compacts Chromatin Structure and Represses Gene Expression
750   Independent of Histone Ubiquitination. *Mol. Cell* **38**, 452–464 (2010).
751   67.   Morey, C., Da Silva, N. R., Perry, P. & Bickmore, W. A. Nuclear reorganisation and chromatin
752   decondensation are conserved, but distinct, mechanisms linked to Hox gene activation.
753   *Development* **134**, 909–919 (2007).
754   68.   Hargrave, M., Bowles, J. & Koopman, P. In situ hybridization of whole-mount embryos.
755   *Methods Mol. Biol.* **326**, 103–13 (2006).
756   69.   Rigueur, D. & Lyons, K. M. Whole-mount skeletal staining. *Methods Mol. Biol.* **1130**, 113–121
757   (2014).
758   70.   Wolff, J. *et al.* Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and
759   single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* **48**, W177–
760   W184 (2020).
761   71.   Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
762
763

764   **# University of Washington Center for Mendelian Genomics**

765   Michael J. Bamshad[6,7,8] & Deborah A. Nickerson[7,8]
766   Non-contributing members of the consortium are listed within the Supplementary Information file.

767

768

769

770

Author contributions

A.U., F.P. and N.A. conceived and designed the study. C.X. helped in lymphoblastoid generation. A.U. performed ChIP-seq, 4C-seq and generated knockout mice. L.K. and L.A.L. performed and analyzed DNA FISH. A.U. and Y.Z. analyzed the phenotype of knockout mice. Y.Z. constructed the LacZ plasmid. M.J.B., D.A.N. and the University of Washington Center for Mendelian Genomics carried out the whole-genome sequencing. F.P. analyzed whole-genome sequencing data. F.P. and C.X. carried out and analyzed Sanger sequencing. J.Z., I.G-S., K.J., and Y.S. performed computational analyses. A.U., F.P., and N.A. interpreted the data. A.U., F.P. and N.A. wrote the manuscript.

Competing Interests

803   There are no competing interests.

804

Figure Legends

Fig. 1| Fine-mapping of the acheiropodia-associated deletion. a, Pedigree of acheiropodia family with proband indicated via the arrow. Squares and circles represent males and females, respectively. **b**, WGS alignments showing a homozygous 12 kb deletion in the acheiropodia proband. The Y-axis is the read depth (number of reads for each nucleotide). The deletion appears in a heterozygous manner in both parents. BP: breakpoint; P: proband; M: mother; F: father. **c**, PCR amplification using three different primers pairs, whose location is indicated in **b**, further confirming the breakpoint in the proband (P) and mother (M) and father (F). PCR was performed several times using different primer sets to validate the deletion. **d**, Sanger sequencing of the acheiropodia patient showing the breakpoint sequence which also has a CA insertion.

Fig. 2 | Mouse transgenic enhancer assay for the 12 kb acheiropodia-associated sequence.
Schematic representation of the mouse transgenic enhancer assay (upper panel) showing the 12 kb acheiropodia associated sequence cloned upstream of *Hsp68* promoter-LacZ gene. Enhancer activity, as visualized by LacZ staining, was not observed in the ZPA for the six PCR positive E11.5 mouse embryos (lower panel).

Fig. 3 | CTCF and RAD21 distribution in the *LMBR1-SHH* locus. a, CTCF and RAD21 ChIP-seq enrichment in lymphoblastoid cells from wildtype (WT) and proband (Mut) at the *LMBR1-SHH* locus. GM12878 (lymphoblastoid cell line) TAD boundaries are shown in orange and gray horizontal bar. ZRS and the acheiropodia-associated deleted region are shown in orange and blue vertical lines respectively. CTCF orientations are shown as red triangles. The Y-axis is the signal p-value to reject the null hypothesis that the signal at that location is present in the control. **b**, Zoom in of the region around the *LMBR1* gene. **c**, CTCF motif from JASPAR[63] [http://jaspar.genereg.net/] and CTCF motif scores, as assigned by FIMO[33], overlapping CTCF peaks in the *LMBR1* locus.

Fig. 4 | Chromatin interactions with the *SHH* promoter. a, 4C contact profiles in lymphoblastoid cells from wildtype (WT) and proband (Mut) at the *LMBR1-SHH* locus. The viewpoint is depicted by a black arrowhead. The median and 20th and 80th percentiles of sliding 2-50 kb windows determine the main trend line. The color scale represents enrichment relative to the maximum medium value attainable at 12 kb resolution. CTCF and RAD21 ChIP-seq peaks are shown as

837   black and blue vertical line respectively. The ZRS and the acheiropodia-associated deleted region

838   are shown as orange and blue vertical lines respectively. CTCF orientations are shown as red

839   triangles. **b**, Zoom in of the region around the *LMBR1* gene.

840

841   **Figure 5 | DNA FISH showing the *SHH* promoter interaction with the acheiropodia-**

842   **associated region. a,** Schematic of the *LMBR1-SHH* locus showing the ZRS and the

843   acheiropodia-associated deleted region via orange and blue vertical lines respectively. CTCF

844   orientations are shown as red triangles and the locations to which the DNA FISH probes hybridize

845   to are depicted by blue bars. **b**, Images of representative nuclei from DNA FISH analysis of

846   parental and proband lymphoblastoid cells showing FISH signals for SHH, LSC2, LSC3-5 and 12kb

847   probes. Scale bars: 5 μm. **c**, Violin plots showing the distribution of interprobe distances (μm)

848   between SHH – LSC1, SHH – LSC2 and SHH – Deletion. The wild type allele was distinguished

849   from the mutant allele in the parental cell line using the 12kb probe. SHH – deletion is measured

850   from SHH to 12kb probe on the wild type allele and from SHH to LSC3-5 probe on the mutant allele.

851   The statistical significance between data sets was examined by a two-sided Mann–Whitney U-test,

852   ** = 0.004537 and **** = 2.083×10^{-11} (n= 75-150 alleles).

853

854   **Fig. 6 | Human and mouse genomic comparisons and phenotype of mice where the**

855   **orthologous region was deleted. a**, Comparison of the *LMBR1-SHH* locus between human and

856   mice. CTCF site deletions analyzed by Paliou *et al.*[21] are marked by purple lines and those

857   generated by [#], Williamson *et al.*[4] are denoted by yellow lines or gray rectangle, and marked by *.

858   CTCF motif orientation is shown via red triangles. The acheiropodia-associated deletion and its

859   mouse orthologous sequence are depicted by a blue rectangle. Mouse limb CTCF ChIP-seq data

860   from ENCODE[37], Andrey *et al.*[38], Paliou *et al.*[21] and human CTCF ChIP-seq data from this study

861   (WT= wild type; Mut= proband) are shown as black genomic tracks below the locus. The

862   conservation track is adopted from the Ensembl Genome Browser[71] with green lines indicating

863   conserved sequences between humans and mice. **b**, *Shh* gene expression levels dissected from

864   E11.5 mouse autopods from wild type (WT) and knockout (KO) mice as determined by qRT-PCR.

865   Each value represents the ratio of *Shh* gene expression to that of *β-Actin*, and values are mean ±

866   standard deviation. The expression value of WT group was arbitrarily set at 1.0. Each dot

867   represents one embryo and statistical differences were determined using a two-sided unpaired t

868   test (*P*= 0.7796, N.S., not significant). Source data are provided as a Source Data file. **c**, Whole-

869 mount *in situ* hybridization for *Shh* of wild type (WT) and knockout (KO) E11.5 mouse embryos.

870 Forelimbs (FL) and hindlimbs (HL) were dissected and shown in the lower panel. **d**, Wild type (WT)

871 and knockout (KO) E18.5 limb skeletal staining using alizarin red/alcian blue.

872

873 **Fig. 7 | Proposed model for the aberrant chromatin structure of the *LMBR1-SHH* locus in the**

874 **acheiropodia patient.** Model of chromatin structure in the *LMBR1-SHH* locus based on our ChIP-

875 seq and 4C-seq data. CTCF sites are shown as red triangles and the cohesin complex is shown as

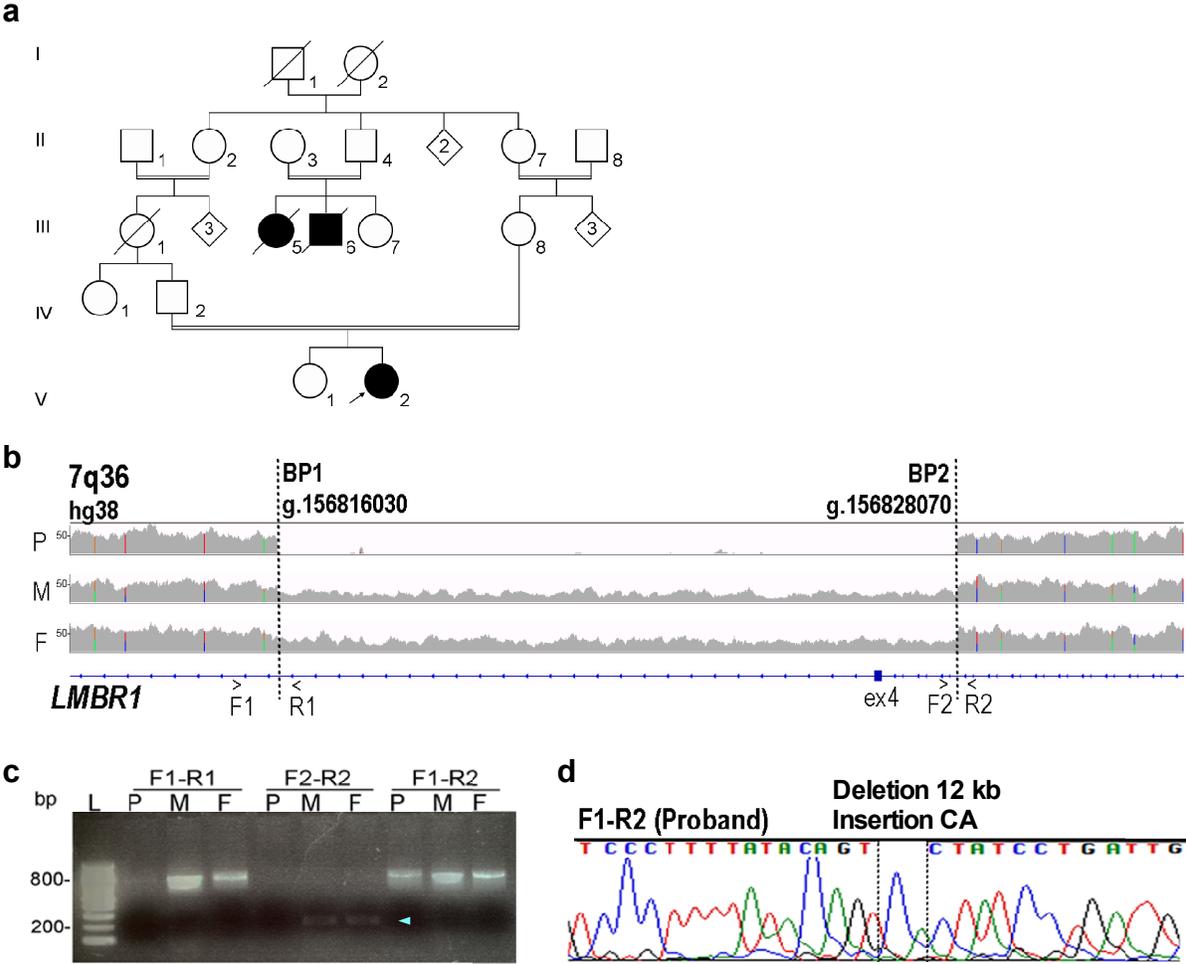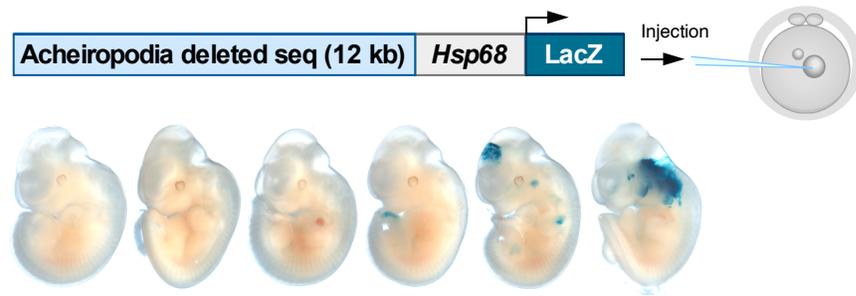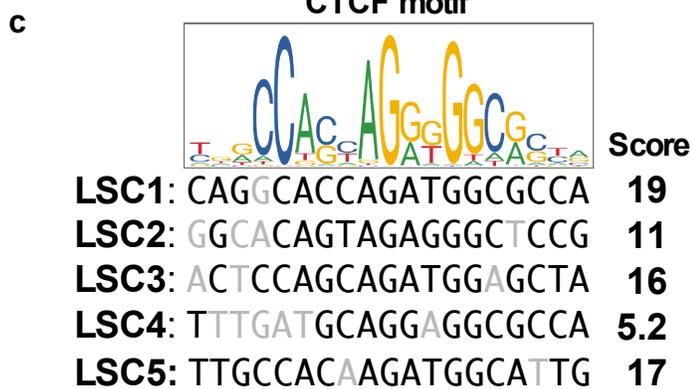876 a green ring. The ZRS is depicted as an orange oval.

877

878

Figure 1

a

b

7q36 hg38	**BP1** g.156816030
	BP2 g.156828070

P

M

F

LMBR1   F1   R1   ex4   F2   R2

c

bp   L   P M F   P M F   P M F
800-
200-

F1-R1   F2-R2   F1-R2

d

F1-R2 (Proband)   **Deletion 12 kb Insertion CA**

T C C C T T T T A T A C A G T   C T A T C C T G A T T G

Figure 2

Figure 3

a

500 kb

RBM33   SHH   RNF32   LMBR1   ZRS   Deletion   NOM1   MNX1

TAD

450   LSC1   LSC2   LSC3-5   6 7
CTCF orientation

CTCF WT   0

450
CTCF Mut   0

480
RAD21 WT   0

480
RAD21 Mut   0

b

100 kb

RNF32   LMBR1   ZRS   Deletion   NOM1

150   LSC1   LSC2   LSC3-5   LSC6 LSC7
CTCF WT   0

150
CTCF Mut   0

250
RAD21 WT   0

250
RAD21 Mut   0

c

CTCF motif

Score

LSC1: CAGGCACCAGATGGCGCCA   **19**
LSC2: GGCACAGTAGAGGGCTCCG   **11**
LSC3: ACTCCAGCAGATGGAGCTA   **16**
LSC4: TTTGATGCAGGAGGCGCCA   **5.2**
LSC5: TTGCCACAAGATGGCATTG   **17**

Figure 4

Figure 5

a

Scale
chr7: 155,500,000

500 kb

156,000,000

156,500,000

ZRS Deletion

157,000,000

EN2
CNPY1 RBM33
SHH

RNF32
LMBR1
MNX1

NOM1 UBE3C

LSC1

LSC2 3-5 6 7

CTCF orientation

Probes SHH

LSC1

LSC2 LSC3-5
12kb

b

SHH   Parental
LSC2
12kb   Mutant
         Wild type

SHH   Proband
LSC2

SHH   Proband
LSC3-5

c

Interprobe distances (μm)

****

**

□ Wild type
■ Mutant

SHH - LSC1   SHH - LSC2   SHH - deletion

Figure 6

Figure 7