Article

Non-coding deletions identify Maenli lncRNA as a limb-specific *En1* regulator

Lila Allou[1,2], Sara Balzano[3,4], Andreas Magg[1,2,5], Mathieu Quinodoz[3,6,7], Beryl Royer-Bertrand[4], Robert Schöpflin[1,2], Wing-Lee Chan[2,5], Carlos E. Speck-Martins[8], Daniel Rocha Carvalho[8], Luciano Farage[9], Charles Marques Lourenço[10], Regina Albuquerque[11], Srilakshmi Rajagopal[12], Sheela Nampoothiri[13], Belinda Campos-Xavier[4], Carole Chiesa[4], Florence Niel-Bütschi[4], Lars Wittler[14], Bernd Timmermann[15], Malte Spielmann[1,2,16], Michael I. Robson[1], Alessa Ringel[1], Verena Heinrich[17], Giulia Cova[1,2], Guillaume Andrey[1,18], Cesar A. Prada-Medina[1], Rosanna Pescini-Gobert[3], Sheila Unger[4], Luisa Bonafé[4], Phillip Grote[19], Carlo Rivolta[3,6,7,20], Stefan Mundlos[1,2,5] ✉ & Andrea Superti-Furga[4]

Long non-coding RNAs (lncRNAs) can be important components in gene-regulatory networks[1], but the exact nature and extent of their involvement in human Mendelian disease is largely unknown. Here we show that genetic ablation of a lncRNA locus on human chromosome 2 causes a severe congenital limb malformation. We identified homozygous 27–63-kilobase deletions located 300 kilobases upstream of the engrailed-1 gene (*EN1*) in patients with a complex limb malformation featuring mesomelic shortening, syndactyly and ventral nails (dorsal dimelia). Re-engineering of the human deletions in mice resulted in a complete loss of *En1* expression in the limb and a double dorsal-limb phenotype that recapitulates the human disease phenotype. Genome-wide transcriptome analysis in the developing mouse limb revealed a four-exon-long non-coding transcript within the deleted region, which we named Maenli. Functional dissection of the *Maenli* locus showed that its transcriptional activity is required for limb-specific *En1* activation in *cis*, thereby fine-tuning the gene-regulatory networks controlling dorso-ventral polarity in the developing limb bud. Its loss results in the *En1*-related dorsal ventral limb phenotype, a subset of the full *En1*-associated phenotype. Our findings demonstrate that mutations involving lncRNA loci can result in human Mendelian disease.

There has been enormous progress in exploring disease variants in the human genome. Yet, the interpretation of variants in the non-coding genome remains a challenge owing to the myriad mechanisms by which they can potentially cause disease. Besides disrupting *cis*-regulatory elements, non-coding variants may interfere with the function of non-coding transcripts. Indeed, a substantial fraction of the human genome is transcribed into RNA, although most transcripts lack protein-coding potential and are referred to as non-coding transcripts[2]. Characterization of a small number of these RNA molecules has revealed that they may have roles as regulators of gene expression through diverse modes of action[3]. However, the identification of functional non-coding transcript loci remains challenging. Thus, annotating non-coding transcript loci and unravelling their function will substantially improve our knowledge about gene regulation and

the identification and interpretation of non-coding genetic variants with respect to disease pathogenesis.

Non-coding deletions cause limb malformations

We identified 27–63-kb non-coding deletions of chromosome 2 in three unrelated individuals (patients 1–3) with a type of limb malformation that, to our knowledge, remains undescribed. Affected individuals had a severe shortening and deformation of the legs and feet, 3/4 syndactyly of the hands, as well as the presence of nails on the palmar side of fingers IV and V (Fig. 1a, Extended Data Fig. 1a, b, Supplementary Note 1). Radiographs showed normal femora but severely shortened tibiae, triangular fibulae and malformed or absent bones in the feet (Fig. 1a, Extended Data Fig. 1a, Supplementary Note 1). Exome sequencing did

Article

Fig. 1 | The human dorsal dimelia deletions result in limb-specific *En1* inactivation and a double dorsal-limb phenotype in mice. a, Photographs and X-rays of patients with dorsal dimelia showing the presence of ventral nails (1), osseous syndactyly (2), normal femurs (3), severely shortened and triangular tibiae (4), triangular fibulae (5) and malformed or absent bones in the feet (6). Brain magnetic resonance imaging (MRI) of patient 4 shows a complete absence of the cerebellum as well as dysplastic occipital lobes filling the posterior fossa (red circle). **b, c**, Overview of the human (**b**) and mouse (**c**) loci. Physical mappings of deletions in 2q14.2 (patients 1, 2 and 3) and a point mutation in *EN1* (patient 4) are shown in **b**. The detailed view of the region is derived from the UCSC Genome Browser (GRCh38/hg38 December 2013 assembly). Genes are represented in black. Deletions are depicted by horizontal blue lines. Vertical dashed black lines depict the location of the breakpoint deletions in patients 1–3. The minimal deleted region is 27 Kb (chr2:118561492–118589320). Vertebrate conservation is measured by 100-vertebrate PhyloP scores. Cen, centromeric; Tel, telomeric; dup, duplication; fs, frameshift; Ter, stop codon. **d**, Micro-computed tomography (μCT), ventral views and high-magnification views of wild-type and *Del(27)* forepaws illustrating the presence of circumferential nails (1), ventral nails formed opposite to the dorsal nails (2) and syndactyly (3). Scale bars, 1,000 μm, 2,000 μm and 500 μm, respectively. **e**, Whole-mount in situ hybridization (WISH) of wild-type and *Del(27)* embryos at E11.5 probed with *En1*. *Del(27)* embryos show a complete loss of *En1* expression in the limb (1) with detectable expression in the somites (2), and mid-hindbrain boundary (3). Scale bars, 1,000 μm (whole embryos) and 200 μm (high-magnification view of limbs).

not reveal any biallelic pathogenic variants related to the phenotype, but autozygosity analysis revealed a region of about 10 Mb on chromosome 2 where patients were homozygous by descent, raising the possibility of a structural variation undetected by exome sequencing. Indeed, comparative genomic hybridization (CGH) array and Sanger

sequencing of the genomic junctions revealed that patients 1 and 2 had identical homozygous deletions (27-kb) on chromosome 2 (Fig. 1b, Extended Data Fig. 1c). Patient 3 was homozygous for a 63-kb deletion and a 2.5-kb insertion in the same region (Fig. 1b, Extended Data Fig. 1c). We defined the region corresponding to the overlapping non-coding deletions as the minimal critical region (MCR) (Fig. 1b). The MCR is located ~300 kb upstream of engrailed-1 (*EN1*), a gene critical for dorsal–ventral patterning in the limb[4] (Fig. 1b). Thus, we hypothesized that the deletions caused altered expression of *EN1* that could explain the clinical features in these individuals. Furthermore, we identified an individual (patient 4) with a limb phenotype similar to those in patients 1–3 and, in addition, a severe neurological condition characterized by spasticity, seizures and complete aplasia of the cerebellum (Fig. 1a, Extended Data Fig. 1a, b, Supplementary Note 1). Sequencing of *EN1* exons in this individual revealed a biallelic loss-of-function mutation (NM_001426.3: c.317dupT; p.Ile107His*fs*Ter39) (Fig. 1b, Extended Data Fig. 1d). These data suggest that loss-of-function mutations in *EN1* result in a recessively inherited combined limb-brain malformation, as previously described for *En1*[−/−] mice[5]. The upstream deletions, however, cause a partial phenotype restricted to the limbs.

To investigate whether a homozygous deletion of the MCR is sufficient to drive the limb phenotypes observed in patients, we generated a deletion of the orthologous region in mouse embryonic stem cells (mouse ES cells) and produced homozygous mice (Fig. 1c). Mice with this deletion (hereafter *Del(27)* mice) were born in the expected Mendelian ratios but showed a strong limb malformation phenotype very similar to that for the previously described inactivation of the *En1* gene[5], consisting of dorsalization of ventral structures including ventral or circumferential claws, ventral hairs, loss of pads, ventral digits and syndactyly (Fig. 1d, Extended Data Fig. 2a–d). In situ hybridizations at mouse embryonic stage 11.5 (E11.5) showed that the deletion caused a specific loss of *En1* in the limb (Fig. 1e), where it is expressed in the ventral apical ectodermal ridge (AER) and ectoderm. Other expression domains, including the hindbrain, spinal cord and somites, remained unchanged (Fig. 1e, Extended Data Fig. 2e). The loss of *En1* in the limb resulted in an ectopic ventral expression of the dorsal marker genes *Lmx1b* and *Wnt7a*, as well as a proximoventral expansion of the AER marker *Fgf8* (Extended Data Fig. 3a), as previously described for *En1*[−/−] mice[4]. However, in contrast to complete *En1* inactivation[5], mice with the 27-kb deletion were viable and did not exhibit defects of the sternum or the brain (Extended Data Fig. 3b, c). Thus, the *Del(27)* mouse recapitulates the double dorsal-limb phenotype in humans. These data support the notion that the deletions identified in patients cause disease by specifically altering *EN1* expression in the limbs.

Identification of limb-expressed lncRNA Maenli

On the basis of our initial results, we hypothesized that the MCR contained DNA regulatory elements that control *EN1* expression in the limb. To test this hypothesis, we first performed capture Hi-C (cHi-C) in mouse embryonic day 9.5 (E9.5) limb buds to define the *En1* regulatory domain, as previously reported[6,7]. At this stage, *En1* begins to be expressed in the limb[8]. cHi-C revealed the MCR is located within a large topologically associated domain (TAD) of about 650 kb that includes *En1* on the centromeric end (Fig. 2a). We examined data from chromatin immunoprecipitation followed by sequencing (ChIP–seq) for CTCF and Rad21 in mouse limbs at E10.5[9] and found that the TAD is bordered by CTCF and Rad21 binding sites at either end that form a strong loop (Fig. 2a). These data indicate that the MCR is part of the *En1* regulatory landscape and is thus likely to contain regulatory information for *En1*. To determine the position of potential regulatory elements controlling *En1* in limb, we examined long-range chromatin interactions between the MCR and *En1* using the *En1* promoter as viewpoint. The virtual 4C profile derived from the cHi-C data showed intense interaction of the *En1* promoter with the other end of the TAD, but no prominent specific interaction with the MCR (Fig. 2a).
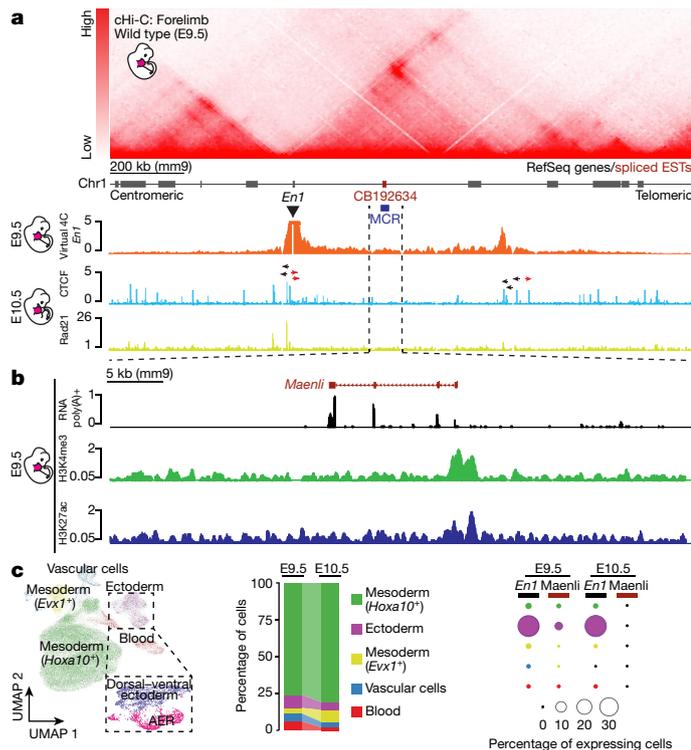
Fig. 2 | The *En1* TAD contains Maenli, a non-coding transcript. a, cHi-C map of the forelimb at E9.5 shows a large TAD encompassing the *En1* gene and the mouse region homologue of the human region deleted in patients with dorsal dimelia. The MCR encompasses the EST CB192634. Virtual 4C profile using the *En1* promoter as viewpoint (black triangle) shows no prominent interaction between *En1* and the MCR. ChIP–seq tracks for CTCF and Rad21 in E10.5 forelimb show that the *En1* TAD is bordered by CTCF and Rad21 binding sites. Orientation of the CTCF motif is indicated by the red and black arrowheads above the track. **b**, RNA-seq profile of E9.5 forelimb showing the presence of Maenli, a four-exon transcript located in the MCR. ChIP–seq tracks for H3K4me3 and H3K27ac in E9.5 forelimb show the presence of these active histone modifications at the *Maenli* TSS. **c**, Uniform manifold approximation and projection (UMAP) showing the five coloured clusters identified using 43,950 cells sampled from scRNA-seq of E9.5 and E10.5 mouse forelimbs. Ectodermal cell subtypes, AER and dorsal–ventral ectoderm cells are also shown. Alluvial plot shows the relative cell-type composition of each developmental stage. Dot plots of the percentage of cells expressing *En1* and Maenli in the different clusters identified at E9.5 and E10.5 developmental stages show that *En1* and Maenli are co-expressed in the ectoderm cell cluster of E9.5 mouse limbs.

We then searched for putative *cis*-regulatory elements and transcripts in the MCR using the Encyclopedia of DNA Elements (ENCODE) and the UCSC Genome Browser (http://genome.ucsc.edu/). We found that the MCR contains the expressed sequence tag (EST) CB192634, and exhibits the characteristics of a non-coding transcript (Fig. 2a). Thus, we hypothesized the potential involvement of a lncRNA locus in the control of *En1* expression and the mechanism of pathogenesis. To test this hypothesis, we performed poly(A)⁺ RNA sequencing (RNA-seq) and ChIP–seq for histone H3 lysine 4 trimethylation (H3K4me3) and histone H3 lysine 27 acetylation (H3K27ac) in E9.5 mouse limbs. The analyses revealed the presence of a four-exon transcript within the MCR (Fig. 2b), partially corresponding to the EST CB192634. Rapid amplification of cDNA ends (RACE) analysis of the mouse RNA revealed two alternatively spliced transcripts (Extended Data Fig. 4a–c). The major transcript (70%) was 1,016 nucleotides long and contained all four exons (Extended Data Fig. 4a, b, d). Its transcriptional start site (TSS) is located 251,108 bp downstream of the 3′-end of *En1* and

displays sevenfold lower expression than *En1* (Extended Data Fig. 4e). The PhyloCSF protein-coding evolutionary signature of the characterized transcript was indistinguishable from lncRNAs, indicating a lack of coding potential (Extended Data Fig. 4f). This is supported by the low conservation of its nucleotide sequence, which displays 40% similarity between mouse and human and low conservation across other species (Extended Data Figs. 4g, 5a, Supplementary Note 2). Moreover, its transcripts were detected in all fractions but enriched in the nucleus of fractionated cells from mouse limbs at E9.5 (Extended Data Fig. 5b). We named this transcript Maenli (short for master activator of engrailed-1 in the limb) (Fig. 2b).

The mouse *Maenli* locus encompasses around 11 kb and its TSS overlaps with H3K27ac and H3K4me3 marks specifically in the limb (Fig. 2b, Extended Data Fig. 5c). Quantitative real-time PCR (rtPCR) showed that Maenli is expressed in the limb at E9.5 but decreases thereafter, whereas *En1* expression persists (Extended Data Fig. 5d). The decrease in Maenli expression was consistent with the loss of H3K4me3 at the *Maenli* locus at E10.5 and E11.5 developmental stages (Extended Data Fig. 5e). To further define the cell population that expresses *Maenli* in the limb, we performed single-cell RNA-seq (scRNA-seq) from E9.5 and E10.5 mouse limbs (Fig. 2c). Bioinformatic analysis identified five major clusters corresponding to the major cell types of the limb: ectoderm, *Hoxa10*⁺ mesoderm, *Evx1*⁺ mesoderm, blood and vascular cells (Fig. 2c). Maenli RNA was present almost exclusively in the ectoderm cluster, overlapping with *En1* expression (Fig. 2c). To determine whether Maenli expression is restricted to the same domains as *En1* (AER and ventral ectoderm), we used the marker genes *En1*, *Fgf8* and *Wnt7a* as identifiers for ventral ectoderm, AER and dorsal ectoderm, respectively. We found Maenli to be enriched in the AER and the ventral ectoderm, with only a few Maenli⁺ cells in the dorsal ectoderm (Extended Data Fig. 5f). Of these *Wnt7a*⁺Maenli⁺ cells, the majority also expressed either *En1* or *Fgf8*. Thus, Maenli is not globally distributed throughout the limb; instead, its expression is restricted to the *En1*-expressing domains.

Maenli transcription in *cis* drives *En1* expression

To test whether *Maenli* is required for *En1* expression and limb development in vivo, we generated a deletion of a 1.4-kb DNA fragment encompassing the *Maenli* TSS (*Maenli*⁻/⁻). This deletion resulted in a more than 98% reduction in expression of Maenli and *En1* in the developing limb, and a double dorsal-limb phenotype in adult mice, similar to that observed in *Del(27)* mice (Fig. 3a, b, Extended Data Fig. 6a). Thus, the *Maenli* locus is causal for the limb malformations in mice resembling the human phenotype.

lncRNA loci have been described as being involved in the regulation of nearby genes via multiple independent mechanisms including: (1) lncRNA promoters acting as enhancers, (2) transcription and/or splicing-related processes, and (3) functions of the RNA products[3]. Thus, from the TSS deletion, it is not possible to distinguish whether enhancer activity at the *Maenli* TSS, transcription at the *Maenli* locus or the mature Maenli transcript itself drives *En1* activation. To distinguish between these possibilities, we first inserted an early triple polyadenylation termination sequence (pAS) 70 bp downstream of the *Maenli* TSS (*Maenli*^pAS/pAS) (Extended Data Fig. 6b). This insertion eliminated about 91% of Maenli expression while leaving the TSS sequence intact. This caused a reduction of approximately 90% in *En1* expression and, consequently, the double dorsal-limb phenotype (Fig. 3a, b, Extended Data Figs. 6a, 7a, b). By contrast, insertion of an inverted polyadenylation sequence (InvpAS) into the same position (*Maenli*^InvpAS/InvpAS) (Extended Data Fig. 6c) had no effect on Maenli and *En1* expression, and *Maenli*^InvpAS/InvpAS mice were indistinguishable from wild-type mice (Fig. 3a, b, Extended Data Fig. 6a). These data demonstrate that limb-specific *En1* expression is not driven by *Maenli* TSS enhancer activity, but rather by the process of *Maenli* transcription or production of the mature Maenli transcript itself.
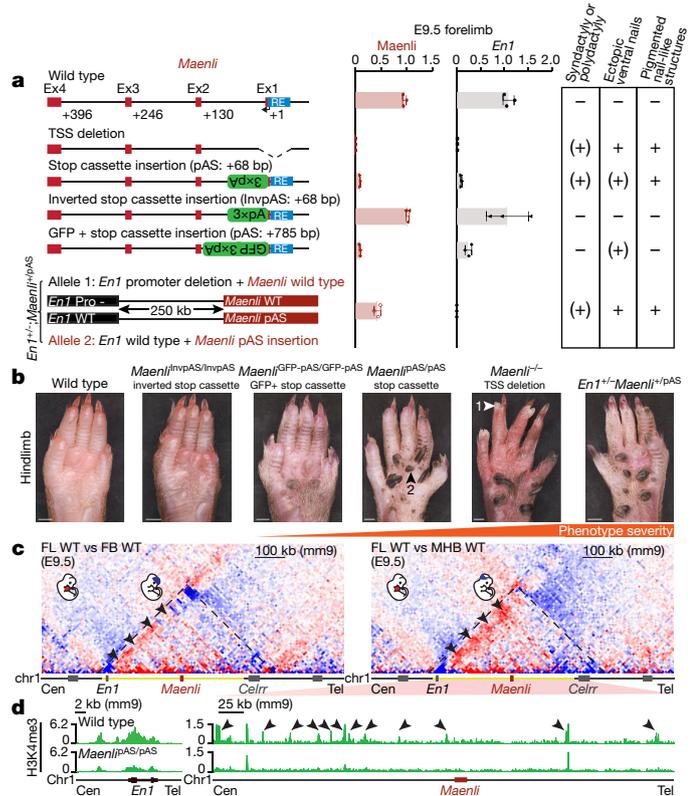
Fig. 3 | Maenli transcription in *cis* is required for the induction of *En1* expression in limb. a, Normalized rtPCR of Maenli and *En1* in E9.5 forelimb embryos shows a significant loss of expression of Maenli and *En1* upon deletion of the *Maenli* TSS (*Maenli*^−/−) (Maenli, $P = 7.43 \times 10^{-7}$; *En1*, $P = 4.43 \times 10^{-5}$) and insertion of a stop cassette in exon 1 of *Maenli* (*Maenli*^pAS/pAS) (Maenli, $P = 1.16 \times 10^{-6}$; *En1*, $P = 6.28 \times 10^{-5}$). By contrast, insertion of an inverted stop cassette in exon1 of *Maenli* (*Maenli*^InvpAS/InvpAS) had no significant effect on expression of Maenli and *En1* (Maenli, $P = 0.047$; *En1*, $P = 0.47$). Insertion of a GFP sequence lacking a stop codon and followed by a stop cassette (*Maenli*^GFP-pAS/GFP-pAS) resulted in an almost complete loss of Maenli expression ($P = 1.34 \times 10^{-6}$), whereas *En1* expression was decreased by about 70% ($P = 2.11 \times 10^{-4}$). *En1*^+/− *Maenli*^+/pAS double-heterozygous mice show a complete loss of *En1* expression in E9.5 limbs ($P = 9.50 \times 10^{-5}$). The main differential phenotypic features of the mutants at 7–8 weeks of age are shown on the right (−, no phenotype; +, fully penetrant; (+), not fully penetrant). Data were normalized to wild-type expression; one-tailed *t*-test; data are mean ± s.d.; $n = 3$ biologically independent wild-type and mutant limb pairs. Ex, exon; RE, regulatory element; TSS, transcription start site; 3×pA, triple polyadenylation termination sequence (stop cassette); GFP, green fluorescent protein; Del, deletion; Pro, promoter; FL, forelimb. **b**, Corresponding phenotypes of age-matched mice. The severity of the double dorsal-limb phenotype corresponds with the level of *En1* loss. 1, ectopic ventral nails; 2, pigmented nail-like structures. Scale bars, 2,000 μm. **c**, cHi-C subtraction maps of E9.5 wild-type forebrain (FB WT) and E9.5 wild-type midhindbrain boundary (MHB WT) from E9.5 wild-type forelimb (FL WT) (red indicates gain in limb and blue indicates loss in limb). Note the interaction stripe-like (indicated with black arrowheads), extending from *En1* through the entire limb regulatory domain (yellow line), including *Maenli* but excluding *Celrr*. The dashed triangle delimits the *En1* TAD. Cen, centromeric; Tel, telomeric. **d**, ChIP–seq tracks for H3K4me3 from wild-type, *Maenli*^pAS/pAS E11.5 mouse limbs. Note the loss of H3K4me3 at the *En1* promoter (left) and across the *En1* limb regulatory domain (indicated with black arrows).

To examine whether *En1* activation is mediated by a sequence-specific function of the Maenli transcript or the process of transcription, we first deleted each of the three downstream exons and three introns of *Maenli* (Extended Data Fig. 6d). None of these deletions significantly altered *En1* activation and the mutant mice were indistinguishable from wild-type mice (Extended Data Figs. 6a, 7c, d). These data suggest that *En1* activation does not require unique sequences or structures in the Maenli transcript itself. To test whether *En1* activation requires sequence-independent functions associated with transcription at the *Maenli* locus in vivo, we inserted a heterologous GFP sequence lacking a stop codon, followed by a pAS sequence 70 bp downstream of the *Maenli* TSS (*Maenli*^GFP-pAS/GFP-pAS) (Extended Data Fig. 6e). *Maenli*^GFP-pAS/GFP-pAS mouse embryos exhibited a more than 90% decrease in Maenli RNA (Fig. 3a, Extended Data Figs. 6a, 7b), similar to *Maenli*^pAS/pAS mutants. However, unlike *Maenli*^pAS/pAS mutants with about 90% lower *En1* expression and a severe phenotype, *Maenli*^GFP-pAS/GFP-pAS mutants showed about 70% decrease of *En1* expression and a milder phenotype (Fig. 3a, b, Extended Data Fig. 7b). Thus, while both mutants express similar levels of residual Maenli RNA, increasing the length of the transcribed region at the *Maenli* locus in *Maenli*^GFP-pAS/GFP-pAS mutants (Extended Data Fig. 7e) led to increased activation of *En1* expression, partially rescuing the limb phenotype. Consistent with these results, it has been shown that changes in the transcript length or amount of transcriptional activity at lncRNA loci influence the local regulation of gene expression[3]. These data support mediation of *En1* activation at least in part by sequence-independent functions associated with transcription at the *Maenli* locus. However, we do not exclude the possibility that exons could have redundant but sequence-dependent roles or that the RNA transcript itself exhibits sequence-non-specific functions.

To test whether *Maenli* functions in *cis* or in *trans* to regulate *En1*, we produced heterozygous *En1*^+/− mice (Extended Data Fig. 8) and crossed them with *Maenli*^+/pAS mice (Fig. 3a). *En1*^+/− *Maenli*^+/pAS mice recapitulated the *Maenli*^−/− phenotype, displaying the double dorsal-limb phenotype (Fig. 3b). *Maenli* expression was reduced to about 45% of wild-type levels in double heterozygotes and *En1* expression decreased by more than 99% (Fig. 3a). Thus, inactivation of *Maenli* on one allele cannot be rescued by Maenli transcripts generated from the sister allele. These data demonstrate that activation of *En1* in the limb is dependent on transcription of *Maenli* in *cis*.

Maenli establishes permissive chromatin

Several examples of *cis*-acting lncRNAs regulating gene expression from a distance have been described, among them CCAT1-L[10], *PVT1*[11] and ThymoD[12]. ThymoD, for example, activates *Bcl11b* by reconfiguring enhancer–promoter communication by recruiting CTCF binding and cohesin, releasing enhancers from the nuclear lamina, and depositing active epigenetic marks across the loop domain[12]. To investigate any effect of *Maenli* transcription on the position of the locus in regard to the nuclear lamina, we performed DNA adenine methyltransferase identification sequencing (DamID-seq) on a primary culture of cells isolated from E11.5 mouse limbs and mouse ES cells in which *En1* was active and inactive, respectively. The *En1* locus and its regulatory landscape were located to the nuclear interior in both active and inactive states, indicating that this mechanism of activation is not relevant for *En1* expression in the limb (Extended Data Fig. 9a). To investigate possible effects of *Maenli* transcription on recruitment of CTCF and/or cohesin, we performed ChIP–seq for CTCF and Rad21 from wild-type and *Maenli*^pAS/pAS mouse limbs at E11.5. We observed no major differences in CTCF or Rad21 binding (Extended Data Fig. 9b), indicating that the mechanism of action of *Maenli* is different from that of *ThymoD*.

To detect any differences in 3D chromatin folding between tissues expressing *En1* and those that do not, we performed cHi-C at E9.5 from limb and mid–hindbrain boundary (MHB), where *En1* is expressed independently of Maenli, and forebrain, where *En1* is not expressed. We detected no difference in the overall TAD configuration. However, forelimb-versus-MHB and forelimb-versus-forebrain subtraction cHi-C maps showed a limb-specific stripe-like interaction within the *En1* TAD that includes *En1* and *Maenli* but not the *Celrr* (cerebellum-expressed regulatory RNA) locus (Fig. 3c). Architectural stripes are chromatin configurations that have been described as originating at
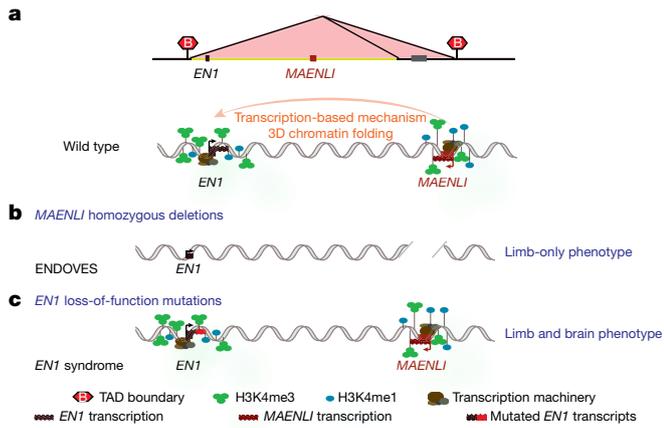
Fig. 4 | A model of MAENLI function in humans. MAENLI-mediated *EN1* expression is required for proper limb development; its loss causes *EN1*-related dorsoventral syndrome, a subset of the full *EN1* syndrome. **a**, The *EN1* TAD contains the *MAENLI* lncRNA locus located about 300 kb away from *EN1*. MAENLI transcription is required for *EN1* activation and normal limb development. **b**, Homozygous deletions involving the *MAENLI* locus result in a complete loss of *EN1* expression specifically in the limb (limb-only phenotype). **c**, Loss-of-function mutations in *EN1*, however, result in a combined limb and brain phenotype. The yellow line in a delimits the *EN1* limb regulatory domain within the *EN1* TAD.

strong-loop-extrusion CTCF anchor points that tether enhancers to their cognate promoters[13,14]. To investigate whether *Maenli* transcription has a role in initiating the observed chromatin conformation, we performed cHi-C from wild-type and *Maenli*[pAS/pAS] mouse limbs at E9.5. The forelimb wild-type-versus-mutant subtraction cHi-C map showed no major differences (Extended Data Fig. 9b), indicating that Maenli transcription has no major influence on the 3D conformation at the *En1* locus in limb. Moreover, the observed stripe-like structure does not appear to be induced by CTCF, but rather reflects increased activity across the locus, as no differences in CTCF binding were observed between *En1* expressing and non-expressing tissues (Extended Data Fig. 9b). Thus, although *Maenli* and the surrounding regulatory landscape are in greater proximity with *En1* in the limb, this seems to occur independently of Maenli or *En1* expression. Instead, a limb-specific 3D configuration brings the actively transcribing *Maenli* locus in close proximity to *En1* where it acts as its licensing factor.

Finally, to investigate possible effects of Maenli transcription on histone modification and chromatin accessibility at the *En1* locus, we performed ChIP–seq for H3K27ac, H3K4me1 and H3K4me3, as well as assay for transposase-accessible chromatin followed by sequencing (ATAC-seq) from wild-type and *Maenli*[pAS/pAS] mouse limbs at E11.5. These data revealed that Maenli transcription affected the deposition of active epigenetic marks at the *En1* and *Maenli* loci as well as the surrounding regulatory landscape (Fig. 3d, Extended Data Fig. 10a). Indeed, *Maenli*[pAS/pAS] and *Del(27)* mutants show a loss of H3K4me3 at the *En1* promoter, as well as at other peaks (overlapping conserved elements) normally established across the *En1* limb regulatory domain mainly at E10.5 and E11.5 developmental stages (Fig. 3d, Extended Data Figs. 5e, 10a). These data suggest that other putative enhancer elements might be involved in the regulation of *En1* expression in limb (Extended Data Fig. 10b). Of note, similar to HOTTIP lncRNA[15], our results show that *Maenli* is a key locus control element during embryonic limb development and is required for the deposition of H3K4me3 epigenetic mark. However, whereas HOTTIP binds to and targets MLL-1 complexes to the 5′ *HOXA* locus, yielding a broad domain of H3K4me3 and transcription activation[15], Maenli lncRNA does not seem to cooperate

with the chromatin-modifying MLL-1 complex. Indeed, we did not detect an enrichment of the MLL-1 catalytic subunit at the *En1* locus and its surrounding regulatory landscape in E11.5 mouse limb buds (Extended Data Fig. 10c). These data support the idea that HOTTIP and Maenli function by different mechanisms. However, we do not exclude the possibility of a 'HOTTIP-like' mechanism involving a different chromatin-modifying complex.

In sum, Maenli transcription in *cis* helps to establish a permissive chromatin environment at the *En1* locus and its surrounding regulatory landscape within a preformed 3D nuclear architecture. This would allow access to transcription factors required for *En1* activation and ventral ectodermal cell fate decision as well as the activation of putative enhancer elements that might be required for *En1* expression regulation.

Discussion

lncRNAs have been found to be differentially expressed in various types of cancer[16] and are dysregulated in other disease types[17–22]; yet it has not been shown that variants involving lncRNA loci can be sufficient to cause human Mendelian diseases. In this study, we provide compelling evidence that inactivation of Maenli lncRNA transcription results in an isolated dorsal dimelia limb phenotype that is a subset of the full *En1*-associated phenotype. We report non-coding deletions that disrupt the previously unannotated lncRNA locus as the molecular aetiology of a severe, recessively inherited congenital limb malformation (Fig. 4). In the era of whole-genome sequencing, the findings described here underscore the need for a systematic annotation and functional characterization of lncRNA loci to interpret and classify non-coding genetic variants. They highlight the importance of elucidating the complex diversity of lncRNA modes of action to assess their role in organ development and disease. Notably, our findings provide a conceptual framework for Mendelian diseases that extends beyond the involvement of coding genes and their regulatory sequences to include long non-coding transcripts acting on the genes themselves. This new paradigm may be the key to solving at least part of the substantial proportion of genetic diseases that remain unexplained.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03208-9.

1. Kopp, F. & Mendell, J. T. Functional classification and experimental dissection of long noncoding RNAs. *Cell* **172**, 393–407 (2018).
2. Hon, C. C. et al. An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature* **543**, 199–204 (2017).
3. Engreitz, J. M. et al. Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
4. Loomis, C. A. et al. The mouse *Engrailed-1* gene and ventral limb patterning. *Nature* **382**, 360–363 (1996).
5. Wurst, W., Auerbach, A. B. & Joyner, A. L. Multiple developmental defects in *Engrailed-1* mutant mice: an early mid-hindbrain deletion and patterning defects in forelimbs and sternum. *Development* **120**, 2065–2075 (1994).
6. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
7. Kragesteen, B. K. et al. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat. Genet.* **50**, 1463–1473 (2018).
8. Huettl, R. E. et al. Engrailed 1 mediates correct formation of limb innervation through two distinct mechanisms. *PLoS ONE* **10**, e0118505 (2015).
9. Andrey, G. et al. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
10. Xiang, J. F. et al. Human colorectal cancer-specific *CCAT1-L* lncRNA regulates long-range chromatin interactions at the *MYC* locus. *Cell Res.* **24**, 513–531 (2014).
11. Cho, S. W. et al. Promoter of lncRNA gene *PVT1* Is a tumor-suppressor DNA boundary element. *Cell* **173**, 1398–1412 (2018).
12. Isoda, T. et al. Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* **171**, 103–119 (2017).

Article

13. Vian, L. et al. The energetics and physiological impact of cohesin extrusion. *Cell* **175**, 292–294 (2018).
14. Kraft, K. et al. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat. Cell Biol*. **21**, 305–310 (2019).
15. Wang, K. C. et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
16. Gutschner, T. & Diederichs, S. The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol*. **9**, 703–719 (2012).
17. Ishii, N. et al. Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J. Hum. Genet*. **51**, 1087–1099 (2006).
18. Broadbent, H. M. et al. Susceptibility to coronary artery disease and diabetes is encoded by distinct, tightly linked SNPs in the *ANRIL* locus on chromosome 9p. *Hum. Mol. Genet*. **17**, 806–814 (2008).
19. Cabianca, D. S. et al. A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* **149**, 819–831 (2012).
20. van Dijk, M. et al. HELLP babies link a novel lincRNA to the trophoblast cell cycle. *J. Clin. Invest*. **122**, 4003–4011 (2012).
21. Maass, P. G. et al. A misplaced lncRNA causes brachydactyly in humans. *J. Clin. Invest*. **122**, 3990–4002 (2012).
22. Castellanos-Rubio, A. et al. A long noncoding RNA associated with susceptibility to celiac disease. *Science* **352**, 91–95 (2016).

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Subjects
Four individuals with congenital limb anomalies were observed at clinical genetic centres in Brazil and in India. Clinical and radiographic images were referred to us (A.S.-F. and S.U.) for diagnostic advice. Given the unique nature of their phenotypes, they were offered additional diagnostic investigations in accordance with regulations for studies on human subjects at the respective centres. Informed consent to participate to the study was obtained from all patients (or their legal guardian). Informed consent to publish identifiable pictures and clinical data was obtained from all patients (or their legal guardian). The study protocol was approved by the local ethics committees Commission cantonale d'éthique de la recherche sur l'être humain, Lausanne, Switzerland and Ethikkommission der Charité, Berlin, Germany, and followed relevant ethical regulations. Clinical details of the subjects are provided in Supplementary Note 1.

Exome sequencing and identification of variants
Molecular and cytogenetic studies on samples from patients were performed at the Lausanne laboratories. Total genomic DNA was extracted from peripheral blood leukocytes according to standard procedures. Exome sequencing was performed using Agilent SureSelect Human All Exon technology (Agilent Technologies). The captured regions were sequenced using Illumina HiSeq 2500 at the Genomic Technologies Facility in Lausanne, Switzerland. Bioinformatics analyses were performed as described previously[23]. In brief, raw reads were mapped to the human reference genome (hg19/GRCh37) using the Novoalign software (v.3.08.00, Novocraft Technologies). Next, Picard (v.2.14.0-SNAPSHOT) was used to remove duplicate reads and Genome Analysis Toolkit (GATK) (v.3.8)[24] was used to perform base quality score recalibration on both single-nucleotide variants and insertion–deletions. A VCF file with the variants was generated by HaplotypeCaller from GATK. DNA variants were then filtered based on quality, frequency in ExAC[25] (v.0.3.1), gnomAD[26] (v.2.1.1), 1000 Genomes[27], ESP (NHLBI Exome Variant Server, http://evs.gs.washington.edu/EVS), GME (GME Variome http://igm.ucsd.edu/gme/index.php), and ABraOM[28], and on predicted impact on protein sequence and messenger RNA (mRNA) splicing.

Homozygosity mapping and haplotype analysis
Large segments of homozygosity (>1 Mb) were computed from exome sequencing data with a tool developed in-house, AutoMap (unpublished), enabling the detection of shared regions of homozygosity in multiple probands. Sequences were then visualized to detect shared haplotypes.

Array CGH
Array CGH (CGHa) was performed using the Agilent SurePrint G3 Human CGH microarray 1X1M kit analysed on an Agilent DNA Microarray scanner and interpreted with the Agilent Genomics Workbench lite edition 6.5 software (all from Agilent Technologies). In patients 1 and 2, the region defined by probes A_16_P00453315 to A_16_P00453356 (included) was found to be homozygously deleted. In patient 3, the region defined by probes A_18_P21824892 to A_18_P13366110 (included) was homozygously deleted; this region is larger than, and comprises, the region deleted in patients 1 and 2.

Sanger sequencing
To determine the precise location of the rearrangement breakpoints in patients 1–3, a series of PCR amplifications using different pairs of primer combinations were performed. Sanger sequencing of the amplified PCR fragments was carried out using standard reagents and conditions. In patient 4, we analysed the entire coding sequence by direct sequencing of PCR products. Exons 1 and 2 of the *EN1* gene were PCR amplified and the coding sequence was screened by direct DNA sequencing. Sequences were analysed with the Seqscape 2.5 software (Applied Biosystems). Sequence variants are numbered starting from the first base of the ATG codon, numbering based on reference sequence NM_001426.3. Naming of variants follows the Human Genome Variation Society nomenclature using (https://variantvalidator.org/variantvalidator/). The sequences of all primers used in this study are listed in Supplementary Table 1.

Generation of structural variations using CRISPR–Cas9
Genome editing experiments were performed as previously described[29]. The size and position of the human structural variations (hg38) were converted to the mouse genome (mm9) using the UCSC liftOver tool. sgRNAs were designed within close proximity of the deletion breakpoints, using the platform at http://crispr.mit.edu/ to obtain candidate single guide RNA (sgRNA) sequences. To minimize off target effects, guide sequences were chosen to have a quality score above 95%. Complementary strands were annealed, phosphorylated, and cloned into the BbsI site of pX459 CRISPR–Cas vector (Addgene). The sequences of all sgRNAs used in this study are listed in Supplementary Table 1. G4 embryonic stem cells[30] (ESCs) (400,000) (129/Sv X C57BL/6 F1 hybrid) were seeded on CD1 mouse embryonic fibroblast (MEF) feeders and cultured under standard ESC culture conditions. The cells were transfected with 8 µg of each pX459-sgRNA construct using FUGENE HD reagent (Promega) under manufacturer conditions. After 12 h, cells were split, transferred into DR4 puromycin-resistant feeders, and selected with puromycin at a final concentration of 2 µg ml[-1] for 48 h. Clones were then grown for 5–6 more days, picked, and transferred onto 96-well plates on CD1 feeders. After 48 h of culture, plates were split in triplicates, two for freezing and one for growth and DNA collection. All clones were genotyped by PCR and quantitative PCR (qPCR) analyses. Positive clones were thawed and grown on CD1 feeders until they reached an average of four million cells. Three vials were frozen, and DNA was collected from the rest of the cells to confirm the genotyping results.

ESCs and feeder cells were tested for *Mycoplasma* contamination using the MycoAlert detection kit (Lonza) and MycoAlert Assay Control Set (Lonza). DR4 and CD1 feeder cell lines were directly produced from mouse embryos originating from DR4 and CD1 mice mattings, respectively.

Genotyping by PCR and sequencing
Primers were designed at a distance of 400–500 bp from each cutting site, on both sides of sgRNA targets. Each allele has thus a set of four primers: F1 (forward) and R1 (reverse) amplifying one targeted site and F2 (forward)/R2 (reverse) amplifying the other targeted site. Deletions, duplications and inversions were detected using the following set of primers: F1–R2, F2–R1, and F1–F2 and R1–R2, respectively. We genotyped each clone by running the PCR products on agarose gels and comparing PCR amplicon sizes. Clones showing evidence for the presence of the desired genomic rearrangement were further genotyped by Sanger sequencing. The sequences of all genotyping primers used in this study are listed in Supplementary Table 1.

Genotyping by quantitative PCR
qPCR was performed on genomic DNA, using the QuantStudio 7 Flex Real-Time PCR System (v.1.7.1) (Applied Biosystems). For the genotyping of copy-number variations, we designed four sets of primers. qPCR was carried out in a total volume of 20 µl containing 10 µl of Power Sybr Green Master Mix (Applied Biosystems), 0.4 µM of each primer and 10 ng genomic DNA. Thermal cycling conditions were 95 °C for 20 s, followed by 40 cycles at 95 °C for 3 s and 60 °C for 30 s. A non-coding region was selected as the control amplicon. Validation experiments

Article

demonstrated that amplification efficiency of the control and all target amplicons were approximately equal. All samples were run in triplicate. The dosage of each amplicon relative to the control amplicon and normalized to wild type DNA was determined using the $2^{-\Delta\Delta CT}$ method.

Transgenic mouse strains

Mice were generated from the corresponding ESC clone by diploid or tetraploid aggregation[31], after thawing a frozen ESC vial seeded on CD1 feeders and grown for 2 days. Female mice of the CD1 strain were used as foster mothers. Several founder animals for each mouse line were used for establishing line stock with variable intercrosses between single founder and C57BL/6 wild-type animals. Selection of animals for analysis and breeding was random. For all animal experiments, no choice of sample size was applied and investigators were not blinded. The data collection was performed according to the stage of each sample since mouse breeding and analysis required knowledge about the genotype at hand.

All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under license numbers G0248/13, G0243/18, G0346/13 and G0176/19. Routine bedding, food and water changes were performed. Mice were housed in a centrally controlled environment with a 12 h light:12 h dark cycle, temperature of 20–22.2 °C, and humidity of 30–50%. All animal experiments followed all relevant guidelines and regulations.

Whole-mount in situ hybridization

RNA expression in E11.5 mouse embryos (G4 and C57BL/6 backgrounds) was assessed by WISH using a digoxigenin-labelled antisense riboprobe transcribed from a cloned probe (PCR DIG Probe Synthesis Kit, Roche). Whole embryos were fixed overnight in 4% PFA/PBS. The embryos were then washed in PBS-Tween (PBST; 0.1% Tween), dehydrated for 10 min (min) each in 25%, 50% and 75% methanol/PBST, and finally stored at −20 °C in 100% methanol. The WISH protocol was as follows. Embryos were rehydrated on ice in reverse methanol/PBST steps, washed in PBST, bleached in 6% H_2O_2/PBST for 1 h and washed again in PBST. Embryos were then treated or not (embryos probed with *Lmx1b* were treated while embryos probed with *En1*, *Fgf8* and *Wnt7a* were not) in 10 µg ml^{-1} proteinase K/PBST for 3 min, incubated in glycine/PBST, washed in PBST, and finally refixed for 20 min with 4% PFA/PBS, 0.2% glutaraldehyde, and 0.1% Tween 20. After further washing steps with PBST, embryos were incubated at 68 °C in L1 buffer (50% deionized formamide, 5× saline sodium citrate, 1% SDS, 0.1% Tween 20 in diethyl pyrocarbonate, pH 4.5) for 10 min. Embryos were then incubated for 2 h at 68 °C in hybridization buffer 1 (L1 with 0.1% transfer RNA and 0.05% heparin). Afterwards, embryos were incubated overnight at 68 °C in hybridization buffer 2 (hybridization buffer 1 with 0.1% transfer RNA and 0.05% heparin and 1/500 digoxigenin probe). After overnight hybridization, removal of the unbound probe was done through a series of washing steps, 3 × 30 min each at 68 °C: L1; L2 (50% deionized formamide, 2 × saline sodium citrate pH 4.5, 0.1% Tween 20 in diethyl pyrocarbonate, pH 4.5); and L3 (2 × saline sodium citrate pH 4.5, 0.1% Tween 20 in diethyl pyrocarbonate, pH 4.5). Subsequently, embryos were treated for 1 h with RNase solution (0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween 20, 100 µg ml^{-1} RNase A in H_2O), followed by washing in Tris-buffered saline, 0.1% Tween 20 (TBST 1) (140 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl, 1% Tween 20, pH 7.5). Embryos were then blocked for 2 h at room temperature in blocking solution (TBST 1 with 2% fetal bovine serum and 0.2% bovine serum albumin (BSA)), followed by incubation at 4 °C overnight in blocking solution containing 1:5,000 anti-digoxigenin-alkaline phosphatase. After overnight incubation, removal of unbound antibody was done through a series of washing steps 8 × 30 min at room temperature with TBST 2 (TBST with 0.1% Tween 20 and 0.05% levamisole/tetramisole) and left overnight at 4 °C. Twenty-four hours later, staining of the embryos was initiated by washing at room temperature with alkaline phosphatase buffer

(0.02 M NaCl, 0.05 M MgCl$_2$, 0.1% Tween 20, 0.1 M Tris-HCl and 0.05% levamisole/tetramisole in H_2O) 3 × 20 min, followed by staining with BM Purple AP Substrate (Roche). At least three embryos were analysed from each mutant genotype. The stained embryos or their limb buds were imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera. The sequences of primers used to generate *En1*, *Lmx1b* and *Wnt7a* WISH probes for this study are listed in Supplementary Table 1. The *Fgf8* probe was previously used in Liska et al.[32].

RNA isolation and quantitative PCR with reverse transcription

To quantify RNA levels in wild type and mutant mice at E9.5 and E11.5 developmental stages (G4 background), limb buds were microdissected in cold PBS, immediately snap-frozen, and stored at −80 °C. Total RNAs were extracted using the RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. RNAs were treated with DNase I (Thermo Fisher Scientific) at 37 °C for 15 min followed by 10 min incubation at 70 °C for DNase I inactivation. Complementary DNAs (cDNAs) were generated using the Superscript III First-Strand Synthesis System (Thermo Fisher Scientific) whereby 1 µg of RNA was reverse-transcribed using oligo(dT)20 (except for the *GFP* quantification for which we used random hexamers). To quantify the relative abundance of transcripts, quantitative PCR with reverse transcription (RT–qPCR) analyses of 4–6 pairs of limb buds were done in technical triplicates using the Power Sybr Green Master Mix (Applied Biosystems) as described above. The fold change (FC) was determined using the $2^{-\Delta\Delta CT}$ method. The dosage of each amplicon was normalized to the endogenous control amplicon (*Rps9* cDNAs). The sequences of all RT–qPCR primers used in this study are listed in Supplementary Table 1.

Skeletal preparation

E18.5 fetuses and P2 pups (G4 background) were kept in H_2O for 1–2 h at room temperature and heat shocked at 65 °C for 1 min. The skin was taken off and the abdominal and thoracic viscera were removed using forceps. The fetuses were then fixed in 100% ethanol overnight. Afterwards, the cartilage was stained overnight using alcian blue staining solution (150 mg l^{-1} alcian blue 8GX in 80% ethanol and 20% acetic acid). Fetuses were then rinsed and postfixed in 100% ethanol overnight. After 24 h, initial clearing was done by incubating the fetuses for 20 min in 1% potassium hydroxide in H_2O, followed by alizarin red (50 mg l^{-1} alizarin red S in 0.2% potassium hydroxide) staining of bones overnight. Following this, rinsing and clearing was done for several days using low concentrations of potassium hydroxide. The stained embryos were dissected in 80% glycerol and limbs, sternums, and ribs were imaged using a ZEISS SteREO Discovery.V12 with cold light source CL9000 microscope and Leica DFC420 digital camera.

Phenotypic evaluation

Phenotypic analyses for 7–8 weeks mutant mouse lines were carried out for at least 3 animals per analysis. Penetrance of the phenotypes was investigated by analysing >30 offspring and considered fully penetrant if all animals were similarly affected.

Micro-computed tomography

Autopods of 10-week-old wild type and mutant mice (n = 3, G4 background) were wrapped in plastic film and scanned ex vivo using a SkyScan 1172 high-resolution micro-computed tomography system (Brucker microCT) at 5-µm resolution. 3D model reconstruction was done with the SkyScan image analysis software (computed tomography analyser and computed tomography volume) (Brucker microCT).

SureSelect design

The capture Hi-C SureSelect library probes were designed over the genomic interval (mm9, chr1: 119,650,000–124,400,000) using the SureDesign online tool from Agilent.

Capture Hi-C

cHi-C libraries were prepared from homozygous E9.5 forelimb buds, forebrain and MHB. cHi-C experiments were performed as triplicates (wild type samples; CD1 background) or duplicates (mutant samples; G4 background). Per biological replicate, 30 pairs E9.5 FL buds ($2-3 \times 10^6$ cells) were microdissected in PBS at room temperature. A single-cell suspension was obtained by incubating the tissue for 10 min at 37 °C in 1 ml Gibco trypsin-EDTA 0.05% (Thermo Fisher Scientific). Cells were resuspended in 5 ml 10% fetal bovine serum/PBS and fixed by adding 5 ml 4% formaldehyde (Sigma-Aldrich) at a final concentration of 2%. Cells were mixed for 10 min at room temperature. Fixation was quenched using 1.425 M glycine (Merck) on ice and immediately centrifuged at 2,100 rpm for 8 min. Supernatant was removed and the pellet resuspended in lysis buffer (final concentration of 10 mM Tris, pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 0.1 M EDTA, and 1 × cOmplete protease inhibitors (Sigma-Aldrich)) and incubated on ice for 10 min. Cells were then centrifuged at 2,900 rpm for 5 min at 4 °C, followed by removal of supernatant, snap-freezing, and storage at −80 °C. For the preparation of the 3C library, the pellet was resuspended in 60 µl 10× DpnII buffer (Thermo Fisher Scientific), and incubated with 15 µl 10% SDS for 1 h at 37 °C and 900 rpm 150 µl 10% Triton X-100 was then added and the pellet was incubated for 1 h at 37 °C and 900 rpm 600 µl of 1× DpnII buffer was added to the samples. A 10-µl aliquot was taken as undigested control and stored at −20 °C. The chromatin was digested using 40 µl 10 U µl⁻¹ DpnII for 4 h at 37 °C and 900 rpm; another 20 µl of DpnII enzyme was then added and samples were incubated overnight at 37 °C with rotation. After overnight incubation, samples were supplemented with 20 µl DpnII enzyme and incubated for four more hours at 37 °C with rotation. The DpnII restriction enzyme was inactivated at 65 °C for 20 min. Next, the digested chromatin was diluted and religated in 5.1 ml H_2O, 700 µl 10× ligation buffer (Thermo Fisher Scientific) and 1.67 µl 30 U µl⁻¹ T4 DNA ligase (Thermo Fisher Scientific). Ligation reactions were incubated at 16 °C overnight with rotation. A 100-µl aliquot was taken to test ligation efficiency and stored at −20 °C. The chimaeric chromatin products and test aliquots were de-cross-linked overnight by adding 30 µl and 5 µl 20 mg ml⁻¹ proteinase K, respectively, and incubated at 65 °C overnight. Following this, 30 µl or 5 µl 10 mg ml⁻¹ RNase A was added to the samples and aliquots, respectively, and incubated for 45 min at 37 °C. Chromatin was then precipitated by adding 1 volume phenol–chloroform to the samples and aliquots, vigorously shaking them, followed by centrifugation at 3,750 rpm at room temperature for 15 min. The upper phase containing the chromatin was transferred to a new tube and the volume was adjusted to 7 ml with H_2O. Samples were then supplemented with 1 ml 3 M sodium acetate, pH 5.6, 35 ml 100% ethanol, and 7 µl 20 mg ml⁻¹ glycogen and frozen overnight at −80 °C. The precipitated chromatin was isolated by centrifugation at 8,350g for 20 min at 4 °C. The chromatin pellet was washed with 10 ml 70% ethanol and further centrifuged at 3,300g for 15 min at 4 °C. Finally, the 3C library chromatin pellet was dried at room temperature and resuspended in 150 µl 10 mM Tris-HCl, pH 7.5 at 37 °C. To check the 3C library, the undigested, digested, and ligated aliquots were loaded on a 1% agarose gel. The 3C library was then sheared using a Covaris sonicator (duty cycle: 10%; intensity: 5; cycles per burst: 200; time: 6 cycles of 60 s each; set mode: frequency sweeping; temperature: 4–7 °C). Adaptors were added to the sheared DNA and amplified according to the manufacturer's instructions for Illumina sequencing (Agilent). The library was hybridized to the custom-designed SureSelect beads and indexed for sequencing (150 bp paired-end) following the manufacturer's instructions (Agilent).

Capture Hi-C processing

FASTQ files were processed with the HiCUP pipeline v.0.6.1[33] (no size selection, Nofill: 1, Format: Sanger) using Bowtie2 v.2.3.4.1[34] for mapping short reads to the reference genome mm9. After the mapping, filtering and deduplication steps of the HiCUP pipeline, replicates were merged by combining their bam files. Juicer tools v.1.7.6[35] was used to generate binned contact maps and to normalize maps by Knights and Ruiz (KR) matrix balancing[36]. For the generation of the contact maps, only the genomic region (chr1:120,200,001–124,400,000, mm9), part of the region enriched in the capturing step, was considered. Therefore, only read pairs mapping to the region of interest were kept and their coordinates were shifted towards the origin by the offset of this region. Afterwards, cHi-C maps were generated with the Juicer tools 'pre' command using a custom chrom.sizes file containing only the length of the genomic region of interest (4.2 Mb) and applying a minimum MAPQ of 30. After exporting KR-normalized cHi-C maps at 10-kb bin size, coordinates were shifted back to mm9 coordinates. Subtraction maps were computed from KR-normalized maps, which were scaled before subtraction across their sub-diagonals in a pairwise manner to account for differences in their signal decay pattern. Therefore, the values of each sub-diagonal of one map were divided by the sum of this sub-diagonal and multiplied afterwards by the average of these sums from both maps. Afterwards, the maps were scaled by 10^6/total sum. cHi-C maps were displayed as heat maps in which values above the 98.5th percentile were truncated to improve visualization.

Virtual 4C profile generation

In order to obtain more fine-grained interaction profile for the *En1* and *Maenli* promoters, a virtual 4C-like profile was generated from mapped, filtered and deduplicated cHi-C read pairs obtained from the HiCUP pipeline. A read-pair was considered in the profile, when it had a MAPQ ≥ 30 and one read mapped to the defined viewpoint region while the other one mapped outside of it. Out of these pairs, reads mapping outside of the viewpoint region were counted per restriction fragment and binned to a 1-kb grid. The count value of restriction fragments spanning more than one bin was distributed proportionally to the corresponding bins. Afterwards, the profile was smoothed by averaging over five bins and scaled by the factor 10^3/counts within the enriched region. For the computation of the scaling factor, the viewpoint region and a margin ±5 kb around it were excluded. The processing was performed with custom Java code using htsjdk v.2.12.0 (https://samtools.github.io/htsjdk/).

RNA sequencing

E9.5 forelimb buds were microdissected from wild type (CD1 background) and mutant embryos (G4 background) in cold PBS, immediately snap-frozen, and stored at −80 °C. RNA-seq experiments were performed in duplicates. Total RNAs were extracted using the RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. RNAs were treated with DNase I (Thermo Fisher Scientific) at 37 °C for 15 min followed by 10 min of incubation at 70 °C for DNase I inactivation. Samples were poly-A enriched and sequenced (150 pb paired-end) using Illumina technology following standard protocols.

RNA-seq data processing

Paired-end RNA-seq reads were mapped to a custom mouse reference genome (mm9), including *Maenli* coordinates (chr1:122,744,478–122,755,676) using the STAR mapper[37] (v.2.6.1d) (splice junctions based on RefSeq; options:–alignIntronMin 20 –alignIntronMax 500000 – outFilterMultimapNmax 5 – outFilterMismatchNmax 10 – outFilterMismatchNoverLmax 0.1). Reads per gene were counted as previously described[38], and used for differential expression analysis with the DEseq2 package[39].

scRNA-seq

scRNA-seq experiments were performed in single replicate. E9.5 forelimb buds of C57BL/6 and 129/Sv wild-type embryos as well as E10.5 forelimb buds of C57BL/6 wild-type embryos were microdissected in PBS at room temperature. A single-cell suspension was obtained by

Article

incubating the tissue for 10 min at 37 °C in 200 µl Gibco trypsin-EDTA 0.05% (Thermo Fisher Scientific) and 20 µl 5% BSA (Sigma-Aldrich). Digestion was then blocked by adding 400 µl of 5% BSA. Cells were then resuspended by pipetting and filtered using a 0.40-µm cell strainer (Falcon). Cells were washed once with 0.04% BSA with a centrifugation step performed for 5 min at 1,200 rpm then resuspended in 0.04% BSA. The cell concentration was determined using an automated cell counter (Bio-Rad) and cells were subjected to scRNA-seq (10x Genomics, Chromium Single Cell 3′ v2; one reaction per time point and per strain has been used) aiming for a target cell recovery of up to 10,000 sequenced cells per sequencing library (time point and strain). Single-cell libraries were generated according to the manufacturer's instructions with the following conditions: 8 PCR cycles were run during cDNA amplification and 12 PCR cycles were run during library generation and sample indexing to increase library complexity. Libraries were sequenced with a minimum of 230 million paired end reads.

scRNA-seq data processing
The Cell Ranger pipeline v.3 (10x Genomics Inc.) was used for each scRNA-seq sample to de-multiplex the raw base call files, generate the fastq files, and perform the alignment against a custom mouse reference genome mm10, including *Maenli* coordinates (ENSMUSG00000099336.2, chr1:120,847,901–120,859,099), to create the unique molecular identifier (UMI) count matrix. Only cells with a detected number of UMI counts among 500 and 40,000, a detected number of genes among 200 and 5,000, and less than 10% of mitochondrial UMI counts were considered for further analysis. In addition, we used scrublet[40] to identify potential duplets in our dataset. We filtered out cells with a scrublet score higher than 0.2. Each sample dataset was normalized independently using the SCT method[41] implemented in Seurat3[42] R package and then integrated the datasets using the Seurat3 IntegrateData function and considering the top 6,000 most variable genes. The first 20 principal components of this joint dataset were calculated and used for UMAP projection and cell clustering. To delimit the major forelimb bud cell types, we used the Louvain algorithm implemented in the Seurat3 function FindClusters with resolution equal 0.02. We identified marker genes for each of the main cell types using the receiver operating characteristic method of the Seurat3 FindAllMarkers function. We only considered genes expressed in at least 10% of each cell cluster and a $\log_2$ fold-change average higher than 0.25. Marker genes are available in Supplementary Table 2. For the sub-clustering of the ectoderm cells, we used a similar approach as described above. We normalized and integrated only ectodermal cells from all samples by considering the top 500 most variable genes among these cells. We clustered and UMAP embedded the ectodermal cells using the top 12 principal components. We used the marker gene *Fgf8* to distinguish between the AER cell cluster and the dorsal–ventral ectoderm cell cluster.

ChIP–seq
Chromatin immunoprecipitation was performed as previously described[43]. ChIP–seq experiments were performed in duplicates. In brief, 80–100 pairs of E9.5 forelimbs (CD1 background) or 10-15 pairs E11.5 forelimbs and hindlimbs (G4 background) were microdissected in PBS at room temperature. A single-cell suspension was obtained by incubating the tissue for 10 min at 37 °C in 1 ml Gibco trypsin-EDTA 0.05% (Thermo Fisher Scientific). Cells were resuspended in 15 ml 10% fetal bovine serum/PBS and fixed by adding 5 ml 4% formaldehyde (Sigma-Aldrich) at a final concentration of 1%. Cells were mixed for 10 min at 4 °C. Fixation was quenched using 1.425 M glycine (Merck) on ice and immediately centrifuged at 2,100 rpm for 8 min. Supernatant was removed and the pellet was immediately snap-frozen and stored at −80 °C. The extraction of nuclear lysate was performed as described in Lee et al.[44] and chromatin was sonicated with a Diagenode Bioruptor (45 cycles with 30 s pulse, 30 s pause, HI power). For ChIP, 25–35 µg of

chromatin was incubated with 6–8 µg of antibody (anti-H3K4me3, Millipore, cat. no. 07-473; anti-H3K4me1, Diagenode, cat. no. C15410037; anti-H3K27ac, Diagenode, cat. no. C15410174; anti-CTCF, Active Motif, cat. no. 61311; anti-Rad21, abcam, cat. no. ab992; anti-MLL1, Bethyl Laboratories, cat. no. A300-086A) overnight. Following this, blocked magnetic beads were added to the chromatin–antibody mixture and incubated overnight. Next, samples were washed 6 times with RIPA buffer and one time with TE buffer[44]. After elution, the preparation of the library for pulled-down DNA was performed as previously described[43]. ChIP–seq libraries were sequenced (75 bp single-end) using Illumina technology following standard protocols.

Previously published data used in this study include ChIP–seq for H3K4me3 from E10.5 forelimbs[9], ChIP–seq data for CTCF, Rad21, H3K4me1 and H3K4me3 from E13.5 midbrains[9], and ChIP–seq data for H3K27ac and H3K4me3 from E13.5 liver[9].

ATAC-seq
ATAC-seq experiments were performed in duplicates as previously described[45]. Tissues from E11.5 embryos (G4 background) were microdissected in PBS at room temperature. A single-cell suspension was obtained by incubating the tissue for 10 min at 37 °C in 1 ml Gibco trypsin-EDTA 0.05% (Thermo Fisher Scientific). The cells were resuspended in 10% fetal bovine serum/PBS and homogenized using a 0.40-µm cell strainer (Falcon). Following this, the cells were washed in cold D-PBS (Gibco) and lysed in fresh lysis buffer (10mM TrisHCl pH7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% (v/v) Igepal CA-630) for 10 min while being centrifuged. Supernatant was then discarded and cells were prepared for the transposition reaction using the Nextera Tn5 Transposase (Nextera kit, Illumina). After 30 min of incubation at 37 °C, the nuclei solution was purified using the MinElute PCR Purification kit (Qiagen). The transposed DNA was eluted in 10 µl of elution buffer and stored at −20 °C, if not immediately used. Barcoded adapters were added to the transposed DNA fragments by PCR using the NEBNext High Fidelity 2x PCR Master Mix (NEB). To avoid saturation in our PCR reaction, we initially performed 5 cycles. Then, an aliquot of 5 µl was used to perform a qPCR in order to determine the additional number of PCR cycles needed to reach the correct amount of DNA without over-amplification. Nextera qPCR primers were used for the amplification and determination of the additional number of PCR cycles. A second round of PCR amplification using the remaining 45 µl of the 1st PCR reaction was then performed (total PCR cycles ≤ 12). Finally, the samples were purified using the AMPure XP beads (Agencourt) and eluted in 20 µl. Samples concentration was measured using Qbit and the quality was assessed using the Bioanalyzer 2100 (Agilent). ATAC-seq libraries were sequenced (75 bp paired-end) using Illumina technology following standard protocols.

ChIP–seq and ATAC-seq data processing
Single-end ChIP–seq reads were mapped with bowtie (v2.2.6)[34] to mm9, filtered for mapping quality MAPQ≥10, and duplicates were removed using samtools rmdup (v.1.8). For generating coverage tracks, reads were extended to 300 bp (H3K27ac, H3K4me1 and H3K4me3) or 200 bp (CTCF and Rad21) and scaled to rpm (10^6 per number of unique reads) using bedtools genomecov v.2.27.1.

Paired-end ATAC-seq reads were trimmed to 30 bp to allow fragments with close-by transposition events (<50 bp) to map, this increases read coverage at nucleosome-free regions. Trimmed reads were mapped with Bowtie (v.2.2.6)[34], and duplicate fragments were removed using Picard RemoveDuplicates (https://broadinstitute.github.io/picard/). Coverage tracks (bigWig) were generated with deepTools (v.2.0)[46] for properly mapped read pairs (FLAG 0 × 2) with mapping quality ≥ 20.

RACE PCR
The sequences of *Maenli* RNA transcripts expressed in the mouse limb bud at E9.5 (CD1 background) were determined using 5′ and 3′ RACE

PCR using the commercially available 5′ RACE System for Rapid Amplification of cDNA Ends, version 2.0 (Invitrogen), and 3′ RACE System for Rapid Amplification of cDNA Ends (Invitrogen). RACE PCR was performed according to the manufacturer's recommended protocol, using primers specific to exon 4 of *Maenli* for the 5′ RACE and exon 2 of *Maenli* for the 3′ RACE. The sequences of these primers are listed in Supplementary Table 1.

Deletions of *Maenli* TSS, exons and introns, and *En1* promoter
We generated these alleles using CRISPR–Cas9 technology as described above for the human structural variation. We assessed the genotype of clonal cell lines as described above for the human structural variation. We note that intron knockouts, by design, do not affect the sequence of the spliced *Maenli* RNA. The sequences of all sgRNAs and primers used for genotyping are listed in Supplementary Table 1.

DamID-seq
DamID was performed as previously described[47]. One replicate was performed per experiment. In brief, mouse ES cells and cultured cells isolated from mouse hindlimb buds at E11.5 (HLCCs) (C57BL/6 background) were both separately transduced with lentiviruses encoding Dam–Lamin B1 and Dam only over 24 h. At 72 h post-transduction, genomic DNA was extracted from trypsinized mouse ES cells and HLCCs and processed into libraries for next generation sequencing (NGS). DamID-seq libraries were sequenced (75 bp paired-end).

DamID-seq data processing
Raw reads from DamID-seq experiments were mapped to the mouse reference genome mm9 using the alignment tool BWA mem (v.0.7.12)[48]. Mapped reads were normalized by reads per kilobase of transcript, per million mapped reads (RPKM) and the $\log_2$ fold change between the Dam–Lamin B1 transduced samples and the respective Dam-only-encoding samples was calculated for every DpnII restriction enzyme site (GATC).

Cellular fractionation
Chromatin-associated, soluble nuclear, and cytoplasmic fractions were isolated as previously described[49]. In brief, 80–100 pairs E9.5 forelimbs (CD1 background) were microdissected in PBS at room temperature. A single-cell suspension was obtained by incubating the tissue for 10 min at 37 °C in 1 ml Gibco trypsin-EDTA 0.05% (Thermo Fisher Scientific). Cells were then centrifuged at 500g for 5 min at 4 °C, resuspended in ice-cold 1× PBS, and aliquoted by centrifugation at 500g for 5 min at 4 °C. An aliquot of $10 × 10^6$ cells was then lysed in 380 µl ice-cold hypotonic lysis buffer (HLB) (10 mM Tris (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, 0.3% (vol/vol) NP-40, and 10% (vol/vol) glycerol) supplemented with SUPERase-In (Thermo Fisher Scientific), and incubated on ice for 10 min. Cells were then centrifuged at 1,000g for 3 min at 4 °C. The supernatant from this spin became the cytoplasmic fraction. 1 ml of RNA precipitation solution (RPS) (0.5 ml of 3 M sodium acetate (pH 5.5) mixed with 9.5 ml of ethanol) was added to the cytoplasmic fraction and stored at −20 °C for at least 1 h. The pellet was subsequently washed 3 times with 1 ml of ice-cold HLB by vortexing for 30 s and centrifuging at 200g for 2 min at 4 °C. The nuclei pellet was then resuspended in 380 µl of modified Wuarin–Schibler buffer (MWS) (10 mM Tris-HCl (pH 7.0), 4 mM EDTA, 0.3 M NaCl, 1 M urea, and 1% (vol/vol) NP-40) supplemented with SUPERase-In (Thermo Fisher Scientific), incubated on ice for 5 min, vortexed for 30 s, and incubated on ice for an additional 10 min. Next, nuclei were centrifuged at 1,000 g for 3 min at 4 °C. The supernatant from this spin became the nucleoplasmic fraction. 1 ml of RPS was added to the nucleoplasmic fraction and stored at −20 °C for at least 1 h. The pellet was subsequently washed 3 times with 1 ml of ice-cold MWS by vortexing for 30 s and centrifuging at 500g for 2 min at 4 °C. The pellet was then resuspended in 1 ml TRIzol and stored at −20 °C. Samples that have been incubated in RPS at −20 °C for at least 1 h were then vortexed for 30 s and centrifuged at 18,000g for 15 min at 4 °C. Pellets were subsequently washed by vortexing in ice-cold 70% (vol/vol) ethanol and centrifuging at 18,000g for 5 min at 4 °C. The supernatant was then discarded and the air-dry pellets were resuspended in 1 ml of TRIzol. Ten microlitres of 0.5 M EDTA were then added to all samples in TRIzol, heated to 65 °C with vortexing until the pellet was completely dissolved. Once samples reached room temperature, 200 µl of chloroform:isoamyl alcohol (1:24) was added to the samples, vortexed for 30 s, and centrifuged at 18,000g for 10 min at room temperature. Aqueous supernatants were then transferred to a new tube and 1 vol of isopropanol was added to each sample, incubated at −20 °C for at least 1 h, vortexed, and centrifuged at 18,000 g for 15 min at room temperature. RNA pellets were then washed with 70% (vol/vol) ethanol by vortexing and centrifuging at 18,000g for 5 min at room temperature. Air-dry pellets were then resuspended in an appropriate volume of RNase and DNase-free water. Purified RNA following DNase-I treatment (Thermo Fischer Scientific) was stored at −80 °C and used for downstream RT–qPCRs analyses as described above.

CRISPR-guided knock-in in mouse ES cells
For the insertion of a triple polyadenylation termination sequence (pAS), an inverted pAS (InvpAS) and a GFP sequence followed by a pAS sequence (GFPpAS) into exon 1 of *Maenli*; 3.4-kb and 4.2-kb constructs containing the respective sequences to be inserted were cloned into a targeting vector with asymmetric homology arms using the Gibson Assembly Cloning Kit (NEB) following the manufacturer's instructions. For knock-in of targeting constructs without selection marker, the targeting construct was transfected in combination with a pX459-sgRNA vector. Notably, the targeting construct did not either contain an intact protospacer adjacent motif site or guide sequence. Puromycin selection and clonal ESC line generation were performed as described above. Successfully targeted ESC lines were screened using PCR and validated for locus-specific integration after successful establishment of the ESC line. Successful clones were further genotyped by Sanger sequencing. The sequences of all sgRNAs and genotyping primers used in this study are listed in Supplementary Table 1. The polyA signal used to introduce the early transcription termination signal is an SV40 poly(A)[50].

Statistics and reproducibility
The $2^{-\Delta\Delta CT}$ method[51] was used to calculate the fold change between wild type and mutant samples. Using a one-tailed Student's *t*-test, we tested the ability of mutations to result in decreased *Maenli* and/or *En1* gene expression. The variance in mutant and wild-type samples was assumed equal.

The data presented in Extended Data Figs. 2a, 3b, c, 8a–d were reproducible in two independent experiments. The data presented in Fig. 1d (µCT), e, and Extended Data Figs. 2d, 3a, 4b, 7a were reproducible in three independent experiments.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
All datasets have been deposited in the Gene Expression Omnibus (GEO) database and are accessible under accession number GSE137335 (or published previously under GSE84795). As we do not have consents from patients to deposit the exome sequencing and CGH array data in an open access public repository, they are available only upon request from A.S.-F. (asuperti@unil.ch) and C.R. (carlo.rivolta@iob.ch). Data can be shared for research purposes with permission of the patient or his/her legal guardian. All the variants reported in this study have been deposited into ClinVar (NCBI) and are accessible under accession numbers SCV001450636, SCV001450638 and SCV001450639. The

Article

novel lncRNA (Maenli) transcript sequence identified and reported here has been deposited into GenBank (NCBI) and is accessible under MW294174. Previously published data used in this study include H3K4me3 ChIP–seq data from E10.5 forelimb, ChIP–seq data for CTCF, Rad21, H3K4me1 and H3K4me3 from E13.5 midbrain and ChIP–seq data for H3K27ac and H3K4me3 from E13.5 liver (GSE84793), data from the Exome Aggregation Consortium (ExAC v0.3.1; http://exac.broadinstitute.org/), the Genome Aggregation Database (gnomAD v2.1.1; https://gnomad.broadinstitute.org/), the 1000 Genomes Browser (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/), the Exome Sequencing Project (https://evs.gs.washington.edu/EVS/), the Greater Middle East Variome Project (http://igm.ucsd.edu/gme/), and the Brazilian genomic variants database (http://abraom.ib.usp.br/). Source data are provided with this paper.

Code availability

The AutoMap tool used for homozygosity mapping and haplotype analysis of exome sequencing data is publically available at https://github.com/mquinodo/AutoMap.

23. Royer-Bertrand, B. et al. Mutations in the heat-shock protein A9 (*HSPA9*) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci. Rep.* **5**, 17154 (2015).
24. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics*. **43**, 11.10.1–11.10.33 (2013).
25. Karczewski, K. J. et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. **45** (D1), D840–D845 (2017).
26. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
27. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
28. Naslavsky, M. S. et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum. Mutat*. **38**, 751–763 (2017).
29. Kraft, K. et al. Deletions, inversions, duplications: engineering of structural variants using CRISPR/Cas in mice. *Cell Rep*. **10**, 833–839 (2015).
30. George, S. H. et al. Developmental and adult phenotyping directly from mutant embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 4455–4460 (2007).
31. Artus, J. & Hadjantonakis, A. K. Generation of chimeras by aggregation of embryonic stem cells with diploid or tetraploid mouse embryos. *Methods Mol. Biol*. **693**, 37–56 (2011).
32. Liska, F. et al. Impairment of Sox9 expression in limb buds of rats homozygous for hypodactyly mutation. *Folia Biol*. **56**, 58–65 (2010).
33. Wingett, S. et al. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. **4**, 1310 (2015).
34. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
35. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. **3**, 95–98 (2016).
36. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
37. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
38. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
39. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550 (2014).
40. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst*. **8**, 281–291 (2019).
41. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. **20**, 296 (2019).
42. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol*. **36**, 411–420 (2018).
43. Ibrahim, D. M. et al. Distinct global shifts in genomic binding profiles of limb malformation-associated *HOXD13* mutations. *Genome Res*. **23**, 2091–2102 (2013).
44. Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc*. **1**, 729–748 (2006).
45. Paliou, C. et al. Preformed chromatin topology assists transcriptional robustness of *Shh* during limb development. *Proc. Natl Acad. Sci. USA* **116**, 12390–12399 (2019).
46. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. **44** (W1), W160–W165 (2016).
47. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat. Protoc*. **2**, 1467–1478 (2007).
48. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
49. Gagnon, K. T., Li, L., Janowski, B. A. & Corey, D. R. Analysis of nuclear RNA interference in human cells by subcellular fractionation and Argonaute loading. *Nat. Protoc*. **9**, 2045–2060 (2014).
50. Grote, P. et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206–214 (2013).
51. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408 (2001).

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03208-9.
Correspondence and requests for materials should be addressed to S.M.
Peer review information *Nature* thanks Annelien Bredenoord, Howard Chang, Moisés Mallo and Andrew Wilkie for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
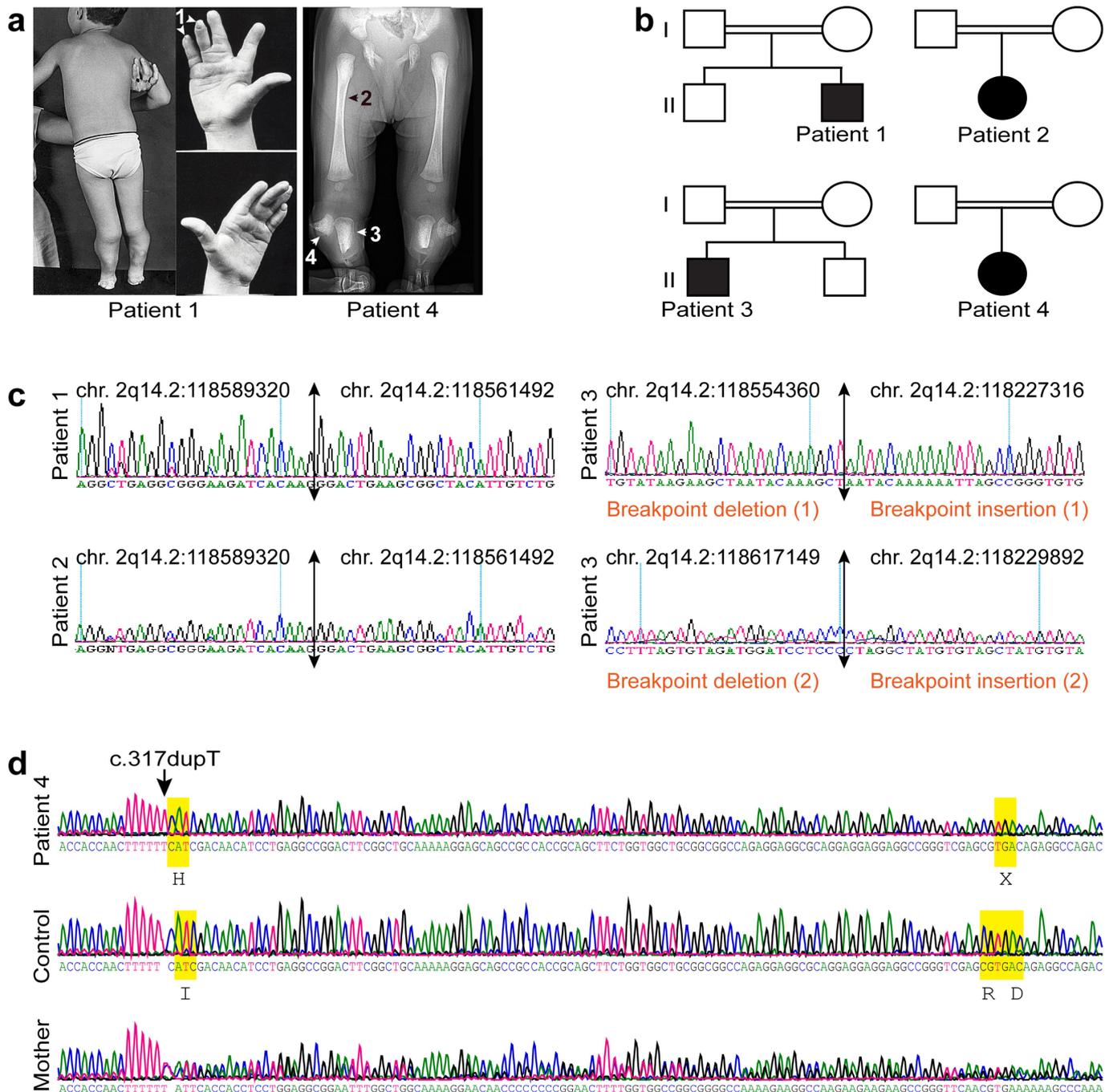
Extended Data Fig. 1 | Clinical and molecular features of the affected individuals. a, Photographs and X-ray of patients 1 and 4 showing the presence of ventral nails (1), normal femurs (2), severely shortened and triangular tibiae (3), triangular fibulae (4). **b**, Pedigrees of patients 1–4. **c**, Sequencing of ~3.0 and ~7.0 kb PCR fragments across the breakpoints revealing a deletion of 27,828 bp in patients 1 and 2, and a deletion of 62,789 bp with a 2,576 bp insertion at the breakpoint in patient 3, respectively. The genomic coordinates for chromosome 2 are based on the UCSC Genome Browser (GRCh38/hg38, December 2013 assembly). **d**, Conventional Sanger sequencing chromatograms show the homozygous mutation (c.317dupT) in patient 4. The presence of a monoallelic variant (c.317dupT) in the asymptomatic mother is also shown. The impact of the biallelic frameshift mutation (c.317dupT) on the amino acid sequence of the resulting EN1 protein is highlighted in yellow. dup, duplication; H, histidine; I, isoleucine; R, arginine; D, aspartic acid; X, stop codon.
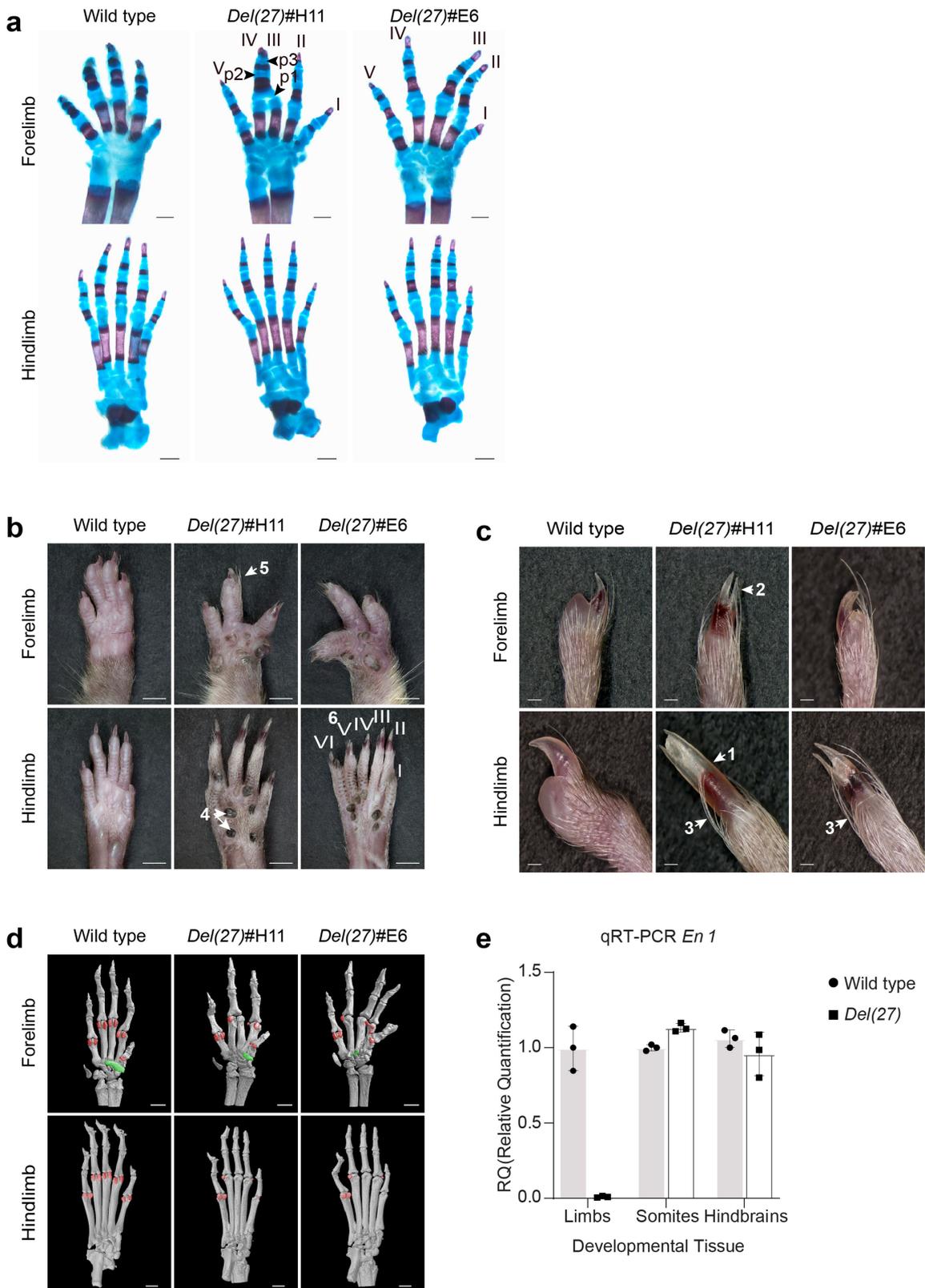
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | *Del(27)* mouse model exhibits a limb phenotype similar to *En1*[−/−] mice. a, Alcian blue (cartilage) and alizarin red (bone) stained limbs prepared from wild-type and *Del(27)* newborn mice. Fusion of digits III and IV at the level of phalanges 3 and 2 is observed for the *Del(27)*#H11 mutant. Digits II and III are displaced towards digit I (preaxial shift) for the *Del(27)*#E6 mutant. For both mutants, digits are flexed dorsally rather than ventrally. p, phalange. Scale bars, 500 μm. **b**, **c**, Ventral views (**b**) and high power digit views (**c**) of wild type and *Del(27)* adult fore- and hindpaws illustrating the presence of cylindrical nails (1), ventral nails formed opposite to the dorsal nails (2), normal dorsal and ectopic ventral hairs (3). The pigmented metatarsal pads (4) are elongated and hardened, resembling nails, and the fusion (5) and polydactyly of digits (6) are variable. Scale bars, 2,000 μm and 500 μm in **b** and **c**, respectively. **d**, μCT of wild type and *Del(27)* adult fore- and hindpaws show that the ventral skeletal elements, such as the falciform (green) and sesamoid bones (red), were lost or incompletely formed in *Del(27)* mutant paws. Scale bars, 1,000 μm. **e**, Normalized qRT–PCR of *En1* in limbs, somites, and hindbrains prepared from wild type and *Del(27)* E11.5 embryos show a complete loss of mRNA expression specifically in the limb buds (P value = 1.58×10^{-4}). No changes in *En1* mRNA levels were detectable in the somites and hindbrains (P value = 0,012; 0,16). Data were normalized to wild type expression. We used a one-tailed *t*-test to evaluate the significance of *En1* loss (n = 3 biologically independent wild type and mutant limb, somite, and hindbrain pairs; mean ± s.d.).
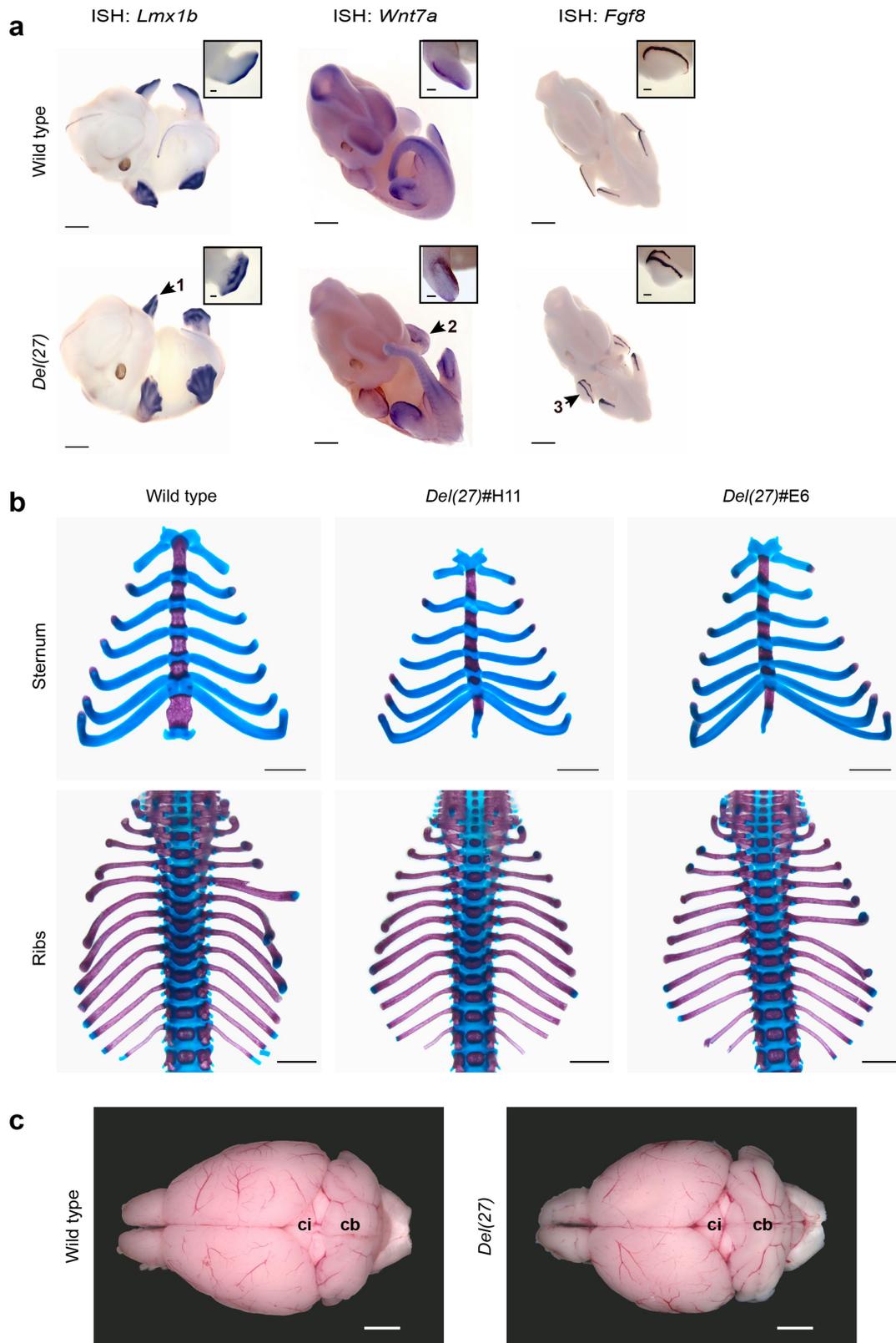
Extended Data Fig. 3 | *Del(27)* mouse model does not exhibit sternum, ribs, or brain defects. a, WISH of wild type and *Del(27)* embryos at E11.5 probed with *Fgf8*, *Lmx1b*, or *Wnt7a*. Limb mutants display ectopic ventral expression of the dorsalizing ge-nes *Lmx1b* (1) and *Wnt7a* (2) as well as aberrant proximoventral expansion of the *Fgf8* expression domain in the AER (3). Scale bars, 1,000 µm for the whole embryos and 200 µm for the high power view of the limb buds, respectively. **b**, Alcian blue- (cartilage) and alizarin red- (bone) stained sternum and ribs prepared from wild type and *Del(27)* newborn mice show that these structures are indistinguishable between wild type and mutant mice. Scale bars, 2,000 µm. **c**, Intact adult brains prepared from wild type and *Del(27)* mice show no brain abnormalities. cb, cerebellum; ci, colliculi. Scale bars, 2,000 µm.

a

11 Kb

Ex4   Ex3   Ex2 Ex1

5' GSP1   3' GSP1

Ex4   Ex3   Ex2 Ex1

Maenli Transcript 1 (1016bp)

Ex4   Ex2 Ex1

Maenli Transcript 2 (871bp)

b

1   2

1016 bp   1500 bp
871 bp   1000 bp
700 bp

c

TGGACAGATTCTCAATAATTC TATTTGAAGTCACTGGATCA
Exon2   Exon1

AGCAGAGCGACTTGCCATCA GAGGCTCCTGTCTGCCAAGA
Exon2   Exon3

AATGACTGGCAGGCATCACA GTTCCTTATTTCATGGGACA
Exon3   Exon4

GTTGTCCCATGAAATAAGGA CATGGCAAGTCGCTCTGCT
Exon4   Exon2

d

qRT-PCR *Maenli*

e

qRT-PCR *En1/Maenli*

f

1 Kb (mm10)

Maenli

PhyloCSF+ 1
PhyloCSF+ 2
PhyloCSF+ 3
PhyloCSF- 1
PhyloCSF- 2
PhyloCSF- 3

g

5 Kb (mm9)

Mammal Cons   2 1 0

Maenli

1 Kb (mm9)

Ex4   Ex3   Ex2   Ex1

GERP   3.08 0 -6.16

Extended Data Fig. 4 | See next page for caption.

Article

Extended Data Fig. 4 | *Maenli* has two alternatively spliced isoforms.
a, Diagram of the *Maenli* transcripts expressed in the mouse limb bud, determined using 5′ and 3′ RACE. Ex, exon; 5′GSP1, 5′ *Maenli*-specific primer from exon 4; 3′GSP1, 3′ *Maenli*-specific primer from exon 2. **b**, Full-length *Maenli* RT–PCR analysis. RT–PCR was carried out using cDNAs generated using total RNA isolated from E9.5 forelimb bud. Agarose gel electrophoresis profile shows the presence of the expected size of DNA fragments (1016 bp and 871 bp) corresponding to *Maenli*-transcript 1 and *Maenli*-transcript 2, respectively. 1, E9.5 forelimb bud cDNAs; 2, 1 kb plus DNA ladder. **c**, Sequencing of the subcloned RT–PCR products illustrates the exon skipping event involving exon 3 of *Maenli*. The double arrowheads indicate the exon/exon junctions.

d, Isoform-specific qRT–PCR of *Maenli* in E9.5 forelimb shows that the 4-exon *Maenli* transcript 1 is the major isoform ($n = 3$ biologically independent wild type forelimbs; mean ± s.d.). **e**, Normalized qRT–PCRs of *En1* and *Maenli* in E9.5 forelimb buds prepared from wild type embryos. *Maenli* expression is on average 7 times lower than *En1* mRNA level. Data were normalized to *Maenli* expression; $n = 3$ biologically independent wild type forelimbs; mean ± s.d. **f**, PhyloCSF tracks show a lack of protein coding potential for the *Maenli* transcript. **g**, Mammalian and vertebrate sequence conservation tracks show that the *Maenli* TSS region is highly conserved while the *Maenli* nucleotide sequence exhibits little conservation.
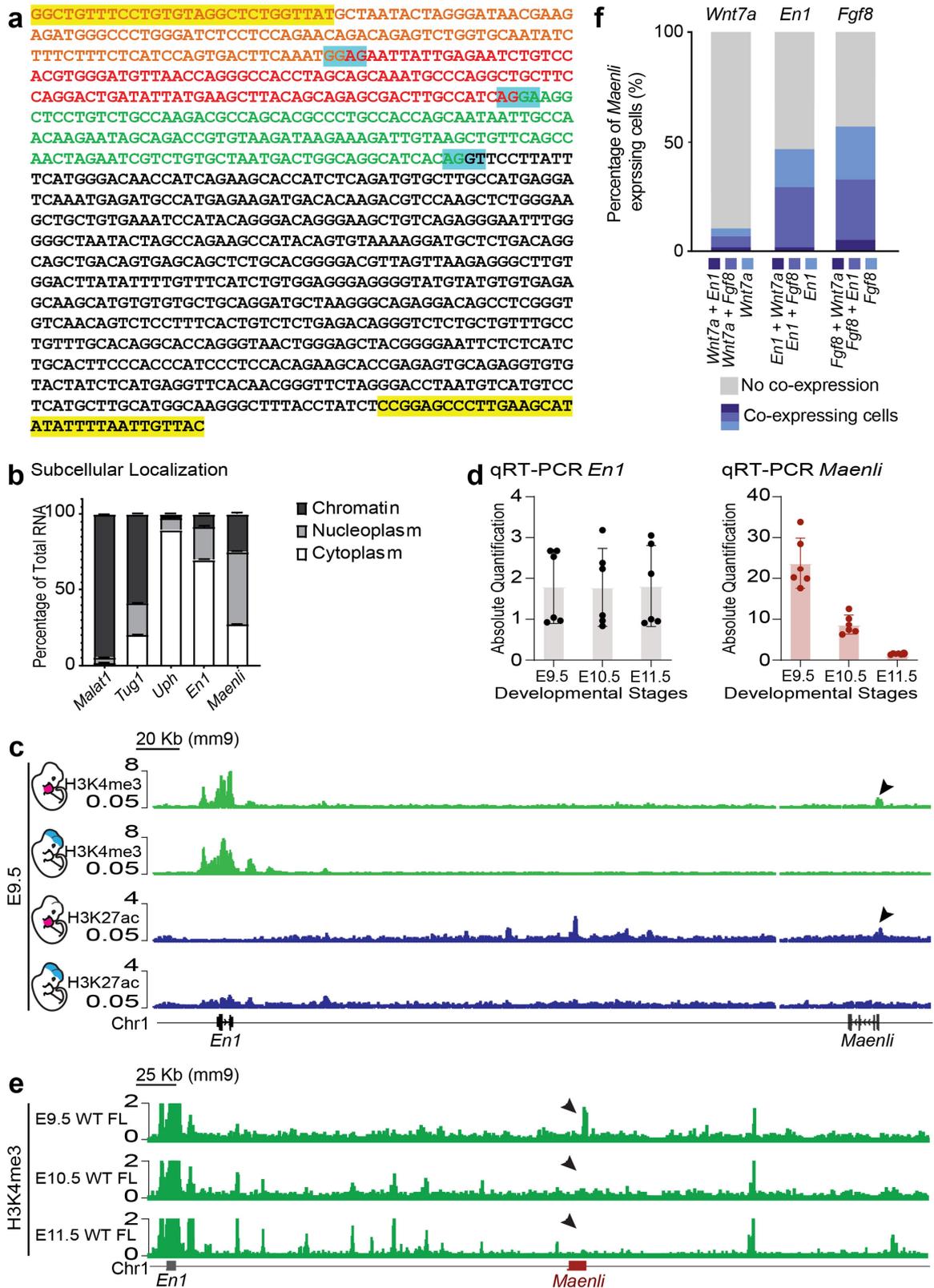
Extended Data Fig. 5 | See next page for caption.

Article

Extended Data Fig. 5 | *Maenli* is a nuclear-enriched lncRNA. a, Nucleotide sequence of the mouse *Maenli* transcript. Exons 1, 2, 3, and 4 are represented using different colours. The exon–exon junctions are highlighted in blue. The primers used for the amplification of the full *Maenli* cDNA are highlighted in yellow. **b**, Subcellular fractionation of *Malat1*, *Tug1*, *Uph*, *En1*, and *Maenli*. qPCR analysis of cDNAs derived from the cytoplasm, nucleoplasm, or chromatin fractions of cells obtained from E9.5 forelimb buds shows that *Maenli* is enriched in the nuclear fraction (75% of total RNA) (n = 3 biological replicates from 4 independent experiments; mean ± s.d.). **c**, ChIP–seq tracks for H3K4me3 and H3K27ac in E9.5 forelimb bud and brain show the presence of the active histone modifications at the *Maenli* TSS specifically in the developing limb bud, indicated with black arrowheads. No active histone modifications were detected at the *Maenli* TSS in the brain. **d**, Normalized qRT–PCRs of *En1* and *Maenli* in E9.5, E10.5, and E11.5 forelimb buds show a decrease of *Maenli* expression from E9.5 to E11.5, while *En1* expression is stable across the different developmental stages. Data were normalized to E11.5 *En1* or *Maenli* expression; n = 3 biologically independent wild type forelimbs; mean ± s.d. **e**, ChIP–seq tracks from wild type (WT) E9.5, E10.5, and E11.5 forelimb buds (FL) show the presence of the H3K4me3 active epigenetic mark at the *Maenli* TSS at E9.5 but not at later stages E10.5 and E11.5. WT, wild type; FL, forelimb. **f**, The percentage of ectodermal *Maenli*-expressing cells that express *Wnt7a* (dorsal ectoderm marker gene), *En1* (ventral ectoderm marker gene), and *Fgf8* (AER marker gene) is shown. *Maenli* is mostly co-expressed with *En1* and *Fgf8* and is thus enriched in the AER and ventral ectoderm, with only a few positive cells in the dorsal ectoderm. Of these *Maenli*/*Wnt7a* positive cells, the majority also express either *En1* or *Fgf8*.
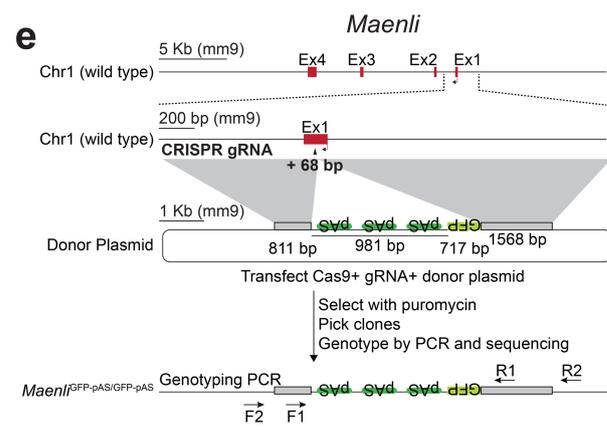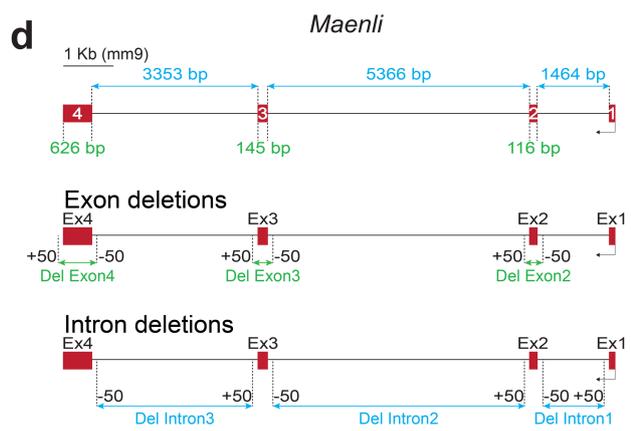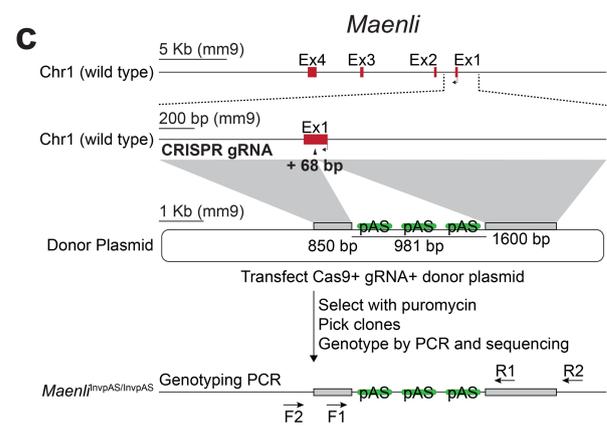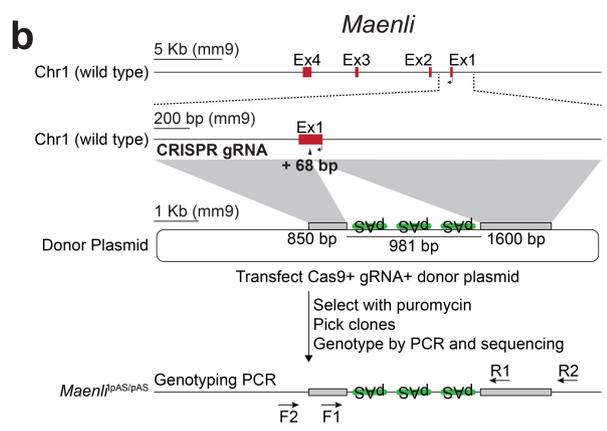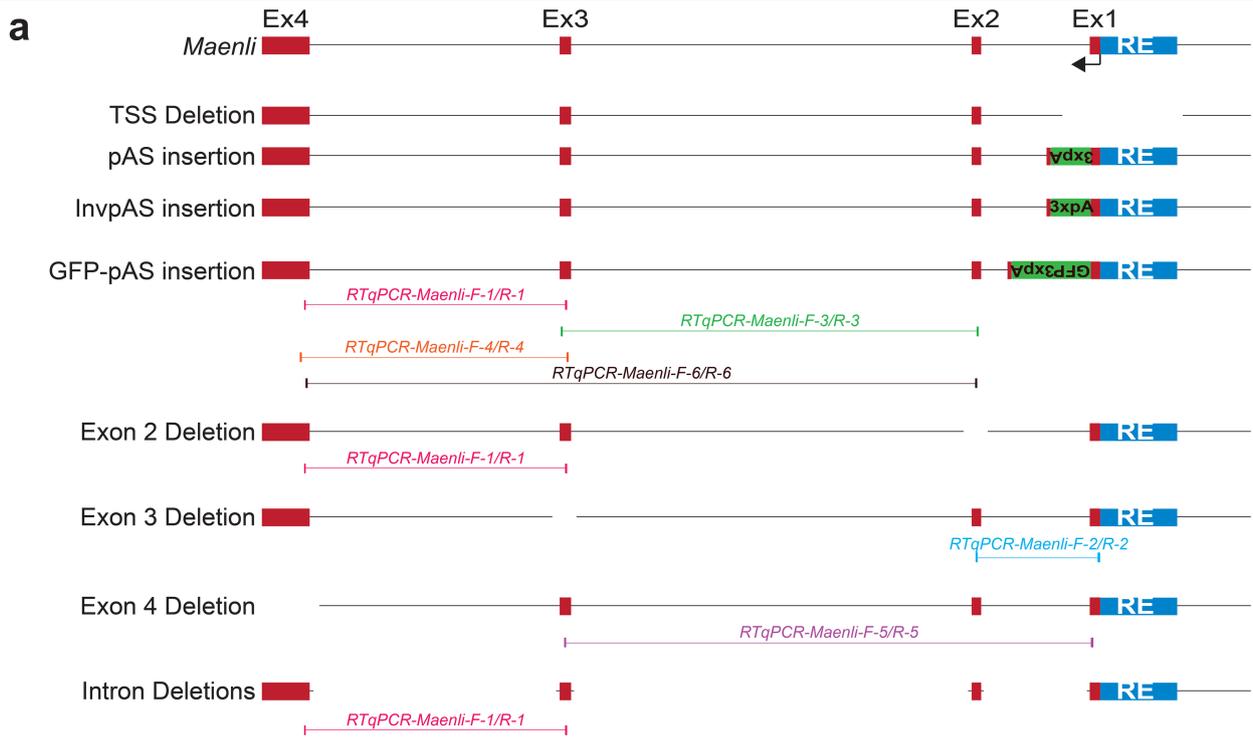
Extended Data Fig. 6 | See next page for caption.

Article

Extended Data Fig. 6 | Knock-ins and knockouts experimental design as well as quantification of *Maenli* expression primer design. a, A schematic of the *Maenli* locus showing the primer positions used for the quantification of *Maenli* expression in the different alleles included in Fig. 3a and Extended Data Fig. 7c. Six pairs of qRT–PCR primers were used for the quantification of *Maenli* expression. Depending on the genetic manipulation, different primer pairs were used for the quantification of *Maenli* expression, ensuring that the same sizes of qRT–PCR products were amplified in the wild type and in the mutant samples. One pair of qRT–PCR primers (qRT–PCR-*Rps9*-F/R) was used as an endogenous control. The fold change (FC) was determined using the $2^{-\Delta\Delta CT}$ method. The dosage of each amplicon was normalized to the endogenous control amplicon and relative to the corresponding wild type sample. As for every genetic manipulation, the qRT–PCR results of the mutant allele were relative (normalized) to the corresponding qRT–PCR results of the wild type allele, only one wild type sample is shown in Fig. 3a and Extended Data Fig. 7c

with an RQ value = 1 (100%) for simplicity. The sequences of the qRT–PCR primers used for the quantification of *Maenli* expression are listed in Supplementary Table 1. The exact position of the primers within the exons is indicated with a vertical dash. RE, regulatory element; Ex, exon; 3xpA, triple polyadenylation termination sequence; GFP, green fluorescent protein. **b**, **c**, **e**, Targeting strategy for the insertion of a triple polyadenylation termination sequence (pAS) (**b**), an inverted pAS (InvpAS) (**c**), and a GFP sequence followed by a pAS sequence (**e**) into exon 1 of *Maenli*. A schematic of the donor vector and CRISPR-mediated homologous recombination targeting strategy is shown. The primers F1, F2, R1, and R2 used for genotyping of the knock-in alleles by PCR correspond to the primers KI-PCRg-1-F, KI-PCRg-2-F, KI-PCRg-1-R, and KI-PCRg-2-R listed in Supplementary Table 1, respectively. Ex, exon; F, forward; R, reverse; GFP, green fluorescent protein. **d**, A schematic of the knockouts of *Maenli* exons and introns is shown. We note that the intron deletions, by design, do not affect the sequence of the spliced *Maenli* RNA. Del, deletion.
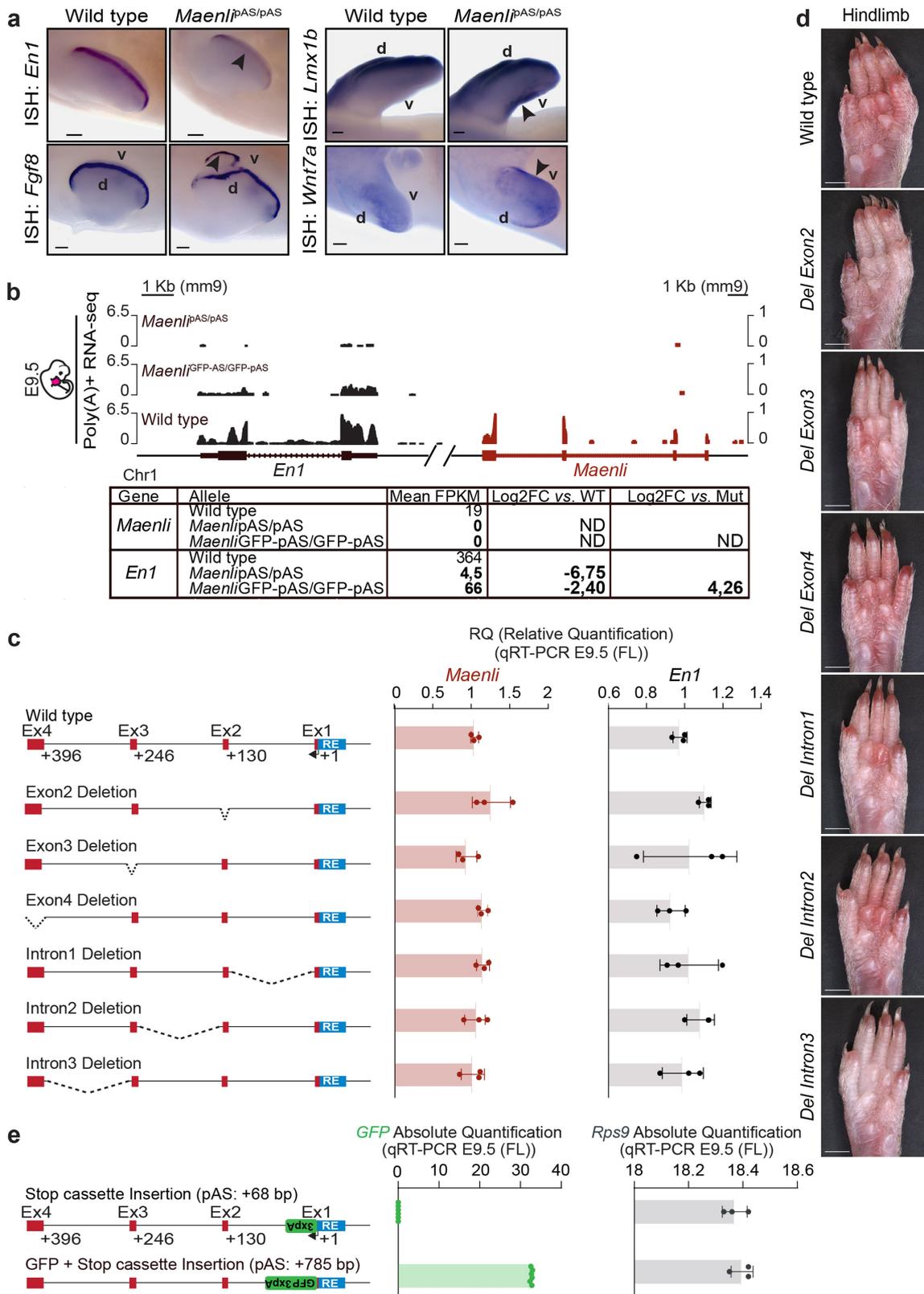
Extended Data Fig. 7 | See next page for caption.

Article

Extended Data Fig. 7 | *En1* activation is dependent on *Maenli* transcription.
a, WISH of wild type and *Maenli*[pAS/pAS] embryos at E11.5 probed with *En1*, *Fgf8*, *Lmx1b*, or *Wnt7a*. Limb mutants show an almost complete loss of *En1* expression (black arrowhead), ectopic ventral expression of the dorsalizing genes *Wnt7a* and *Lmx1b* (black arrows), as well as aberrant proximoventral expansion of the *Fgf8* expression domain in the AER (black arrow). d, dorsal; v, ventral. Scale bars, 200 µm. **b**, Poly(A)⁺ RNA-seq profiles of E9.5 forelimb buds show that the pAS insertion abolishes *Maenli* transcription in both *Maenli*[pAS/pAS] and *Maenli*[GFP-pAS/GFP-pAS] mutants. However, while *Maenli*[pAS/pAS] mutants exhibit an almost complete loss of *En1* expression, *Maenli*[GFP-pAS/GFP-pAS] mutants exhibit a less severe reduction in *En1* expression. A summary of the expression values for *En1* and *Maenli* is shown. FPKM, fragments per kilobase of exon per million fragments mapped; ND, not defined; Mut, *Maenli*[pAS/pAS]. **c**, Normalized qRT–PCRs of *Maenli* and *En1* in E9.5 forelimb embryos show that individual exon and intron deletions of *Maenli* had no significant effect on *Maenli* and *En1* expression (*P* value = 0,10; 0,28; 0,14, 0,06; 0,38; 0,40; 0,04; 0,36; 0,19; 0,31; 0,04; 0,41). Data were normalized to wild type *Maenli* or *En1* expression (*n* = 3 biologically independent wild type and mutant limb pairs; a one-tailed *t*-test was used to evaluate the significance of *Maenli* and *En1* changes; mean ± s.d.). **d**, Ventral views of adult hindpaws prepared from wild type and mutant mice carrying a deletion of each of the three downstream exons and three introns of *Maenli*. Note that the mutant hindpaws are indistinguishable from wild type. Del, deletion. Scale bars, 2,000 µm. **e**, Normalized qRT–PCRs of *GFP* and *Rps9* in E9.5 forelimb show that transcription of the orthologous GFP sequence is detected in *Maenli*[GFP-pAS/GFP-pAS] mutants but not in *Maenli*[pAS/pAS] mutants (cycle threshold undetermined). *Rps9* was used as an endogenous control to assess RNA quality and quantity. Data were normalized to *Maenli*[GFP-pAS/GFP-pAS] *GFP* and *Rps9* expression (*n* = 3 biologically independent mutant limb pairs; mean ± s.d.). Ex, exon; RE, regulatory element; GFP, green fluorescent protein; 3xpA, triple polyadenylation termination sequence (stop cassette); FL, forelimb.
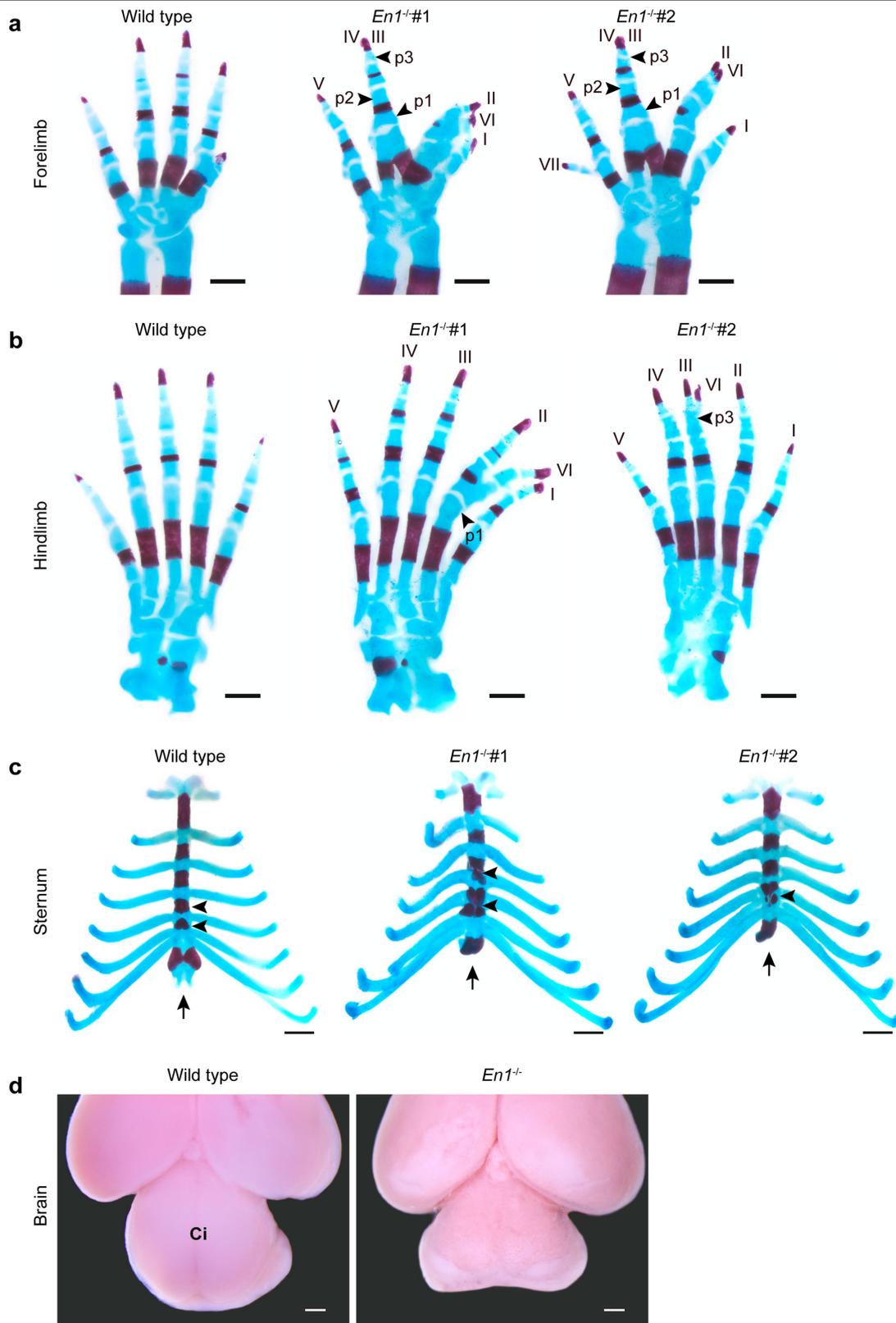
Extended Data Fig. 8 | Limb, brain, and sternum defects in *En1*^−/− mice produced by crossings of *En1*^+/− mice. a, Alcian blue- (cartilage) and alizarin red- (bone) stained forelimbs prepared from wild type and *En1*^−/− E18.5 embryos show the presence of syndactyly of the 3rd and 4th digits at the level of phalanges 3, 2, and 1 for both mutants, as well as the presence of polydactyly (VI) and fusion of digits II and VI. Note the presence of the postaxial polydactyly (VII) for *En1*^−/−#2. p, phalange. Scale bars, 2,000 µm. **b**, Alcian blue- (cartilage) and alizarin red- (bone) stained hindlimbs prepared from wild type and *En1*^−/− E18.5 embryos show the presence of ectopic ventral digits (VI) fused at the level of phalanges 1 and 3 for *En1*^−/−#1 and *En1*^−/−#2, respectively. p, phalange. Scale bars, 500 µm. **c**, Wild type and mutant sternums stained with alcian blue and alizarin red show truncated sternums (black arrows) of *En1*^−/− mutants and abnormal placement of the ribs. The sternebrae are reduced in size and intercostal ossification is reduced and irregular (black arrowheads). Scale bars, 1,000 µm. **d**, Intact brains prepared from wild type and *En1*^−/− E18.5 embryos show the absence of the colliculi in the mutants. ci, colliculi. Scale bars, 500 µm.
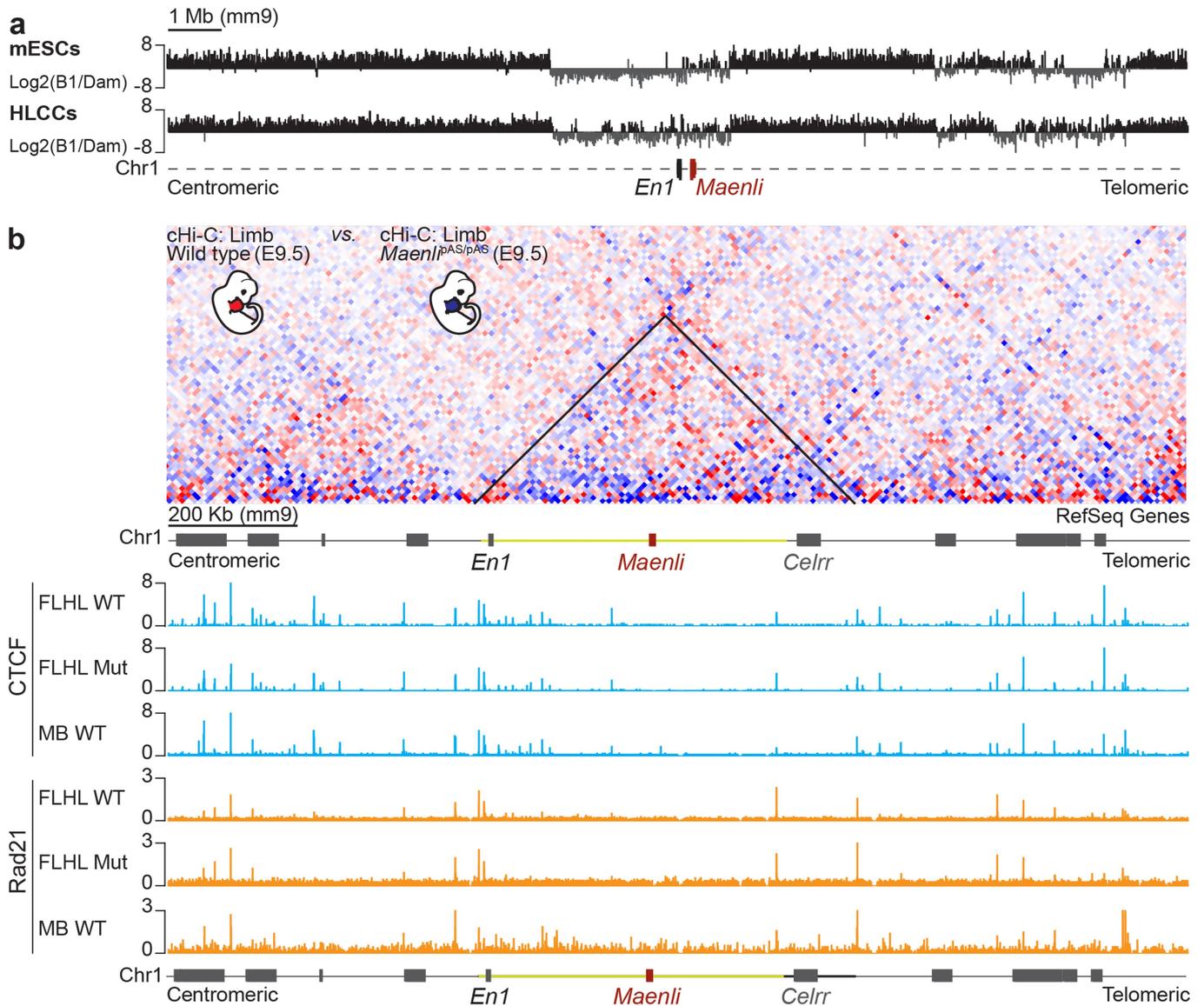
Extended Data Fig. 9 | *Maenli* expression does not seem to modulate neither 3D chromatin folding nor CTCF and cohesin occupancy. a, DamID-seq tracks from wild type mouse ES cells and in vitro culture of hindlimb buds from E11.5 embryos show that the *En1* locus and its regulatory landscape are located to the nuclear interior in both cell types. mESCs, mouse embryonic stem cells; HLCCs, primary culture of cells derived from E11.5 hindlimb buds. **b**, Subtraction map of E9.5 *Maenli*^pAS/pAS^ forelimb cHi-C signal from E9.5 wild type forelimb cHi-C signal shows no major differences between the wild type and mutant limb buds. ChIP–seq profiles for CTCF and Rad21 from wild type (WT) and *Maenli*^pAS/pAS^ (Mut) E11.5 limb buds (FLHL) are shown. No major differences in CTCF or Rad21 binding were observed between the wild type and mutant limb buds, but also between the wild type limb buds and a control sample (wild type midbrain (MB)) used for the detection of limb-specific peaks. FLHL, forelimb-hindlimb; MB, midbrain; Mut, *Maenli*^pAS/pAS^; WT, wild type. The black triangle delimits the *En1* TAD. The yellow segment line delimits the *En1* limb regulatory domain. The black segment line delimits the end of the *En1* TAD.
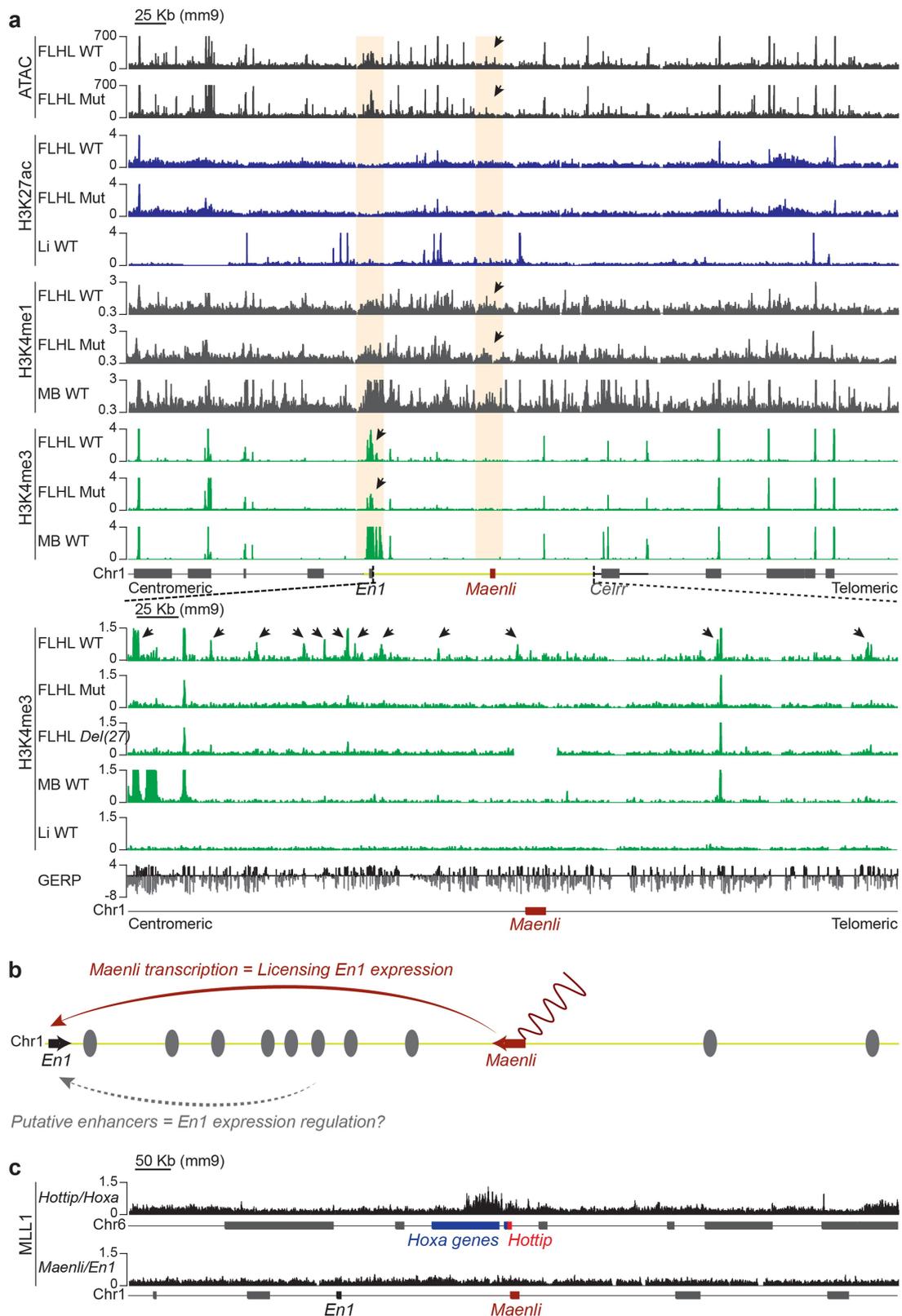
Extended Data Fig. 10 | *Maenli* transcription loss affects the deposition of H3K4me3 across the *En1* limb regulatory domain. a, ATAC reads and ChIP–seq profiles for H3K27ac, H3K4me1 and H3K4me3 from wild type (WT) and *Maenli*[pAS/pAS] (Mut) E11.5 limb buds (FLHL) are shown. Major changes in chromatin accessibility and histone marks are highlighted in orange and indicated with black arrows. A control sample (liver (Li) or midbrain (MB)) used for the detection of limb-specific peaks is shown when possible. A zoom in of the *En1* limb regulatory domain for H3K4me3 ChIP–seq tracks from wild type (WT), *Maenli*[pAS/pAS] (Mut) and *Del(27)* E11.5 limb buds (FLHL) show the loss of H3K4me3 across the *En1* limb regulatory domain. Control samples including the midbrain (MB) and liver (Li) used for the detection of limb-specific peaks are shown. FLHL, forelimb-hindlimb; Li, liver; MB, midbrain; Mut, *Maenli*[pAS/pAS]; WT, wild type. The yellow segment line delimits the *En1* limb regulatory domain. The black segment line delimits the end of the *En1* TAD. Vertebrate conservation is measured using the Genomic Evolutionary Rate Profiling (GERP) method. **b**, Genetic model depicting the regulation of *En1* expression in the developing limb bud. **c**, ChIP–seq tracks for MLL1 in E11.5 limb buds show that MLL1 is enriched at the *HOTTIP/HOXA* locus but not at the *Maenli/En1* locus.

Corresponding author(s): Stefan Mundlos

Last updated by author(s): Dec 22, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection No commercial, open source or custom software/code was used to collect the data in this study.

Data analysis
- For exome sequencing analyses: 1) raw reads were mapped to the human reference genome (hg19/GRCh37) using the Novoalign software (V3.08.00, Novocraft Technologies), Picard (version 2.14.0-SNAPSHOT) was used to remove duplicate reads, Genome Analysis Toolkit (GATK) (version 3.8) was used to perform base quality score recalibration on both single-nucleotide variants and insertion–deletions. A VCF file with the variants was generated by HaplotypeCaller from GATK. DNA variants were filtered based on quality, frequency in ExAC v0.3.1, gnomAD v2.1.1, 1000 Genomes, ESP (NHLBI Exome Variant Server, http://evs.gs.washington.edu/EVS), GME (GME Variome http://igm.ucsd.edu/gme/index.php), and ABraOM.
- Homozygosity mapping and haplotype analysis were computed from exome sequencing data with a tool developed in-house, AutoMap (Unpublished).
- Array CGH were analyzed on an Agilent DNA Microarray scanner and interpreted with the Agilent Genomics Workbench lite edition 6.5 software (all from Agilent Technologies, Santa Clara CA 95051, U.S.A).
- Sanger sequencing data were analyzed with the Seqscape 2.5 software (Applied Biosystems, Foster City, CA, USA).
- Naming of variants follows the Human Genome Variation Society nomenclature using (https://variantvalidator.org/variantvalidator/).
- The size and position of the human structural variations (hg38) were converted to the mouse genome (mm9) using the UCSC liftOver (toolhttps://genome.ucsc.edu/cgi-bin/hgLiftOver).
- sgRNAs were designed within close proximity of the deletion breakpoints, using the http://crispr.mit.edu/ platform to obtain candidate sgRNA sequences.
- qRT-PCRs results were analyzed using the QuantStudio 7 Flex Real-Time PCR Software (version 1.7.1) (Applied Biosystems).
- 3D model reconstruction of micro-computed tomography data was done with the SkyScan image analysis software (computed tomography analyzer and computed tomography volume) (Brucker microCT). The SkyScan 1172 scanner used to perform these experiments is equipped with this software.
- The capture Hi-C SureSelect library probes were designed using the SureDesign online tool from Agilent (https://earray.chem.agilent.com/

suredesign/).
-For the cHi-C data analysis, raw sequencing reads were processed with the HiCUP pipeline v0.6.1 (no size selection, Nofill: 1, Format: Sanger) using Bowtie2 v2.3.4.1 for mapping short reads to the reference genome mm9. Juicer tools v1.7.6 was used to generate binned contact maps and to normalize maps by Knights and Ruiz (KR) matrix balancing.
-cHi-C maps were generated with the Juicer tools v.1.7.6.
-Data processing for the generation of the virtual 4C profile was performed with custom Java code using htsjdk v.2.12.0 (https://samtools.github.io/htsjdk/).
-Paired-end RNA-seq reads were mapped to a custom mouse reference genome (mm9), including Maenli coordinates (chr1:122,744,478-122,755,676) using the STAR mapper (version 2.6.1d) (splice junctions based on RefSeq; options: --alignIntronMin 20 − alignIntronMax 500000 − outFilterMultimapNmax 5 − outFilterMismatchNmax 10 − outFilterMismatchNoverLmax 0.1).
-RNA-seq differential expression analysis was performed with the DEseq2 package.
-The Cell Ranger pipeline version 3 (10x Genomics Inc.) was used for each scRNA-seq sample to de-multiplex the raw base call files, generate the fastq files, and perform the alignment against a custom mouse reference genome mm10, including Maenli coordinates (ENSMUSG00000099336.2), to create the UMI count matrix.
-We used scrublet to identify potential duplets in our dataset. Scurblet is available as online tool and has been incorporated into SPRING (https://kleintools.hms.harvard.edu/tools/spring.html).
-Each sample dataset was normalized independently using the SCT method implemented in Seurat3 R package and then integrated the datasets using the Seurat3 IntegrateData function and considering the top 6000 most variable genes.
-To delimitate the major forelimb bud cell types, we used the Louvain algorithm implemented in the Seurat3 function FindClusters with resolution equal 0.02.
-We identified marker genes for each of the main cell types using the ROC method of the Seurat3 FindAllMarkers function.
-Single-end ChIP-seq reads were mapped with bowtie (v2.2.6) to mm9, duplicates were removed using samtools rmdup (v1.8). The generation of coverage tracks was performed using bedtools genomecov v2.27.1.
-Paired-end ATAC-seq reads were mapped with Bowtie (v2.2.6), duplicate fragments were removed using Picard RemoveDuplicates (https://broadinstitute.github.io/picard/), coverage tracks (bigWig) were generated with deepTools v2.0.
-Raw reads from DamID-seq experiments were mapped to the mouse reference genome mm9 using the alignment tool BWA mem (version 0.7.12).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All datasets have been deposited in the Gene Expression Omnibus (GEO) database and are accessible under GSE137335 (or published previously under GSE84795). The source data behind Fig. 3a and Extended Data Figs. 2e, 4d, 4e, 5b, 5d, 7c, 7e are available within the manuscript files. As we do not have consents from patients to deposit the exome sequencing and comparative genomic hybridization array (CGHa) data in an open access public repository, they are available only upon request from Dr. Andrea Superti-Furga (Division of Genetic Medicine, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland; asuperti@unil.ch) and Dr. Carlo Rivolta (Medical Genetics Unit, Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; carlo.rivolta@iob.ch). Data can be shared for research purposes with permission of the patient or his/her legal guardian. All the variants reported in this study have been deposited into ClinVar (NCBI) and are accessible under SCV001450636, SCV001450638, and SCV001450639. The novel lncRNA (Maenli) transcript sequence identified and reported here has been deposited into GenBank (NCBI) and is accessible under MW294174.
Previously published data used in this study include H3K4me3 ChIP-Seq data from E10.5 FL, ChIP-seq data for CTCF, Rad21, H3K4me1 and H3K4me3 from E13.5 Midbrain and ChIP-seq data for H3K27ac and H3K4me3 from E13.5 Liver (GSE84793), data from the Exome Aggregation Consortium (ExAC v0.3.1; http://exac.broadinstitute.org/), the Genome Aggregation Database (gnomAD v2.1.1; https://gnomad.broadinstitute.org/), the 1000 Genomes Browser (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/), the Exome Sequencing Project (ESP; https://evs.gs.washington.edu/EVS/), the Greater Middle East Variome Project (GME Variome; http://igm.ucsd.edu/gme/), and the Brazilian genomic variants database (ABraOM; http://abraom.ib.usp.br/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	WISH experiments were performed from 3 independent biological mouse embryos. qRT-PCRs experiments were performed using 3 independent biological replicates. Skeletal preparations were performed using 2 independent biological mouse embryos or pups. Phenotypic evaluation was performed using at least 3 animals per experiment. Micro-computed tomography was performed using 3 biological replicates. Capture Hi-C experiments were performed using 3 biological replicates per experiment. RNA-seq analyses were performed using 2 biological replicates. scRNA-seq experiments were performed from one biological replicate. ChIP-seq experiments were performed using 2 biological

replicates except for anti-MLL1 ChIP-seq experiments that have been performed from one biological replicate. ATAC-seq experiments were performed using 2 biological replicates. Cellular fractionation experiments were performed from 4 independent biological replicates. DamID-seq experiments were performed from one biological replicate.
No specific choice of sample sizes was made in this study. We considered as appropriate a sample size of minimum 2 per experiment for non quantitative ones and a sample size of minimun 3 per experiment for quantitative ones.
In our study, an additional biological replicate would be required for RNA-seq, ChIP-seq and ATAC-seq to be able to perform statistical analysis; as well as for scRNA-seq and DamID-seq to confirm the results.

Data exclusions	Samples/ animals were included/ excluded according to genotype. Genotyping was established and verified in appropriate control experiments.
Replication	All experiments were replicated 2 or 3 times and were everytime successfully reproducible.
Randomization	Several founder animals for each mouse line were used for establishing line stock with variable intercrosses between single founder and C57BL/6 wildtype animals. Selection of animals for analysis and breeding was random.
Blinding	Investigators were not blinded during experiments, the data collection was performed according to the stage of each sample since mouse breeding and analysis required knowledge about the genotype at hand.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The following specific antibodies were used for this study: anti-H3K4me3, Millipore, cat. No. 07-473; anti-H3K4me1, Diagenode, cat. No. C15410037; anti-H3K27ac, Diagenode, cat. No. C15410174; anti-CTCF, Active Motif, cat. No. 61311; anti-Rad21, abcam, cat.No. ab992; anti-MLL1, Bethyl Laboratories, cat. No. A300-086A.
Validation	-anti-H3K4me3: Application Chromatin immunoprecipitation was performed using the Magna ChIP™ HiSens kit (cat# 17-10460), 1 μL Anti-trimethyl-Histone H3 (Lys4) antibody (cat# 07-473) , 20 μL Protein A/G beads, and 1 million crosslinked HeLa cell chromatin. -anti-H3K4me1: Judson RN et al. Inhibition of methyltransferase Setd7 allows the in vitro expansion of myogenic stem cells with improved therapeutic potential. Cell Stem Cell. 2018 Feb 1; 22(2): 177–190.e7. ChIP; Mouse. -anti-H3K27ac: Application Chromatin immunoprecipitation was performed on sheared chromatin from 1 million HeLaS3 cells using 2 μg of the Diagenode antibody against H3K27ac (Cat. No. C15410174). -anti-CTCF: Despang A et al. Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. Nat Genet. 2019 Aug;51(8):1263-1271. ChIP; Mouse. -anti-Rad21: Amat R et al. Rapid reversible changes in compartments and local chromatin organization revealed by hyperosmotic shock. Genome Res. 2019. 29: 18-28. ChIP; Human. -anti-MLL1: Schmidt et al. CEBPA-mutated leukemia is sensitive to genetic and pharmacological targeting of the MLL1 complex. Leukemia. 2019 July;33(7):1608-1619. ChIP; Mouse.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	We used mouse embryonic stem cells (mESCs) from (129/Sv×C57BL/6J F1 hybrid) (G4) backgrounds. These cells were obtained from Dr. Anders Nagy (George et al., 2007). CD1 and DR4 feeder cell lines, produced from CD1 and DR4 transgenic embryos, were used to culture the G4 cells.
Authentication	Genetically modified ESCs were used to produce embryos using Tetraploid and Diploid aggregation. Genotyping confirmed the presence of the desired mutation. CD1 and DR4 feeder cell lines were directly produced from mouse embryos originating from DR4 and CD1 mice crosses, respectively.
Mycoplasma contamination	All the cell lines were tested and were negative for mycoplasma contamination.

Commonly misidentified lines (See [ICLAC](#) register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	Mice from CD1, C57BL/6J, or (129/Sv×C57BL/6J) F1 hybrid backgrounds were used in our study. Males and females from embryonic days E9.5, E10.5, and E11.5, postnatal stage P2, and adult animals (7-8 or 10 weeks) were used in our experiments. Routine bedding, food, and water changes were performed. Mice were housed in a centrally controlled environment with a 12-h light/12-h dark cycle, temperature of 20–22.2°C, and humidity of 30–50%. All animal experiments followed all relevant guidelines and regulations.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	All animal procedures were conducted as approved by the local authorities (LAGeSo Berlin) under the license numbers G0248/13 and G0346/13.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	There are in total 4 human research participants included in this study. Each patient is a direct descendent of consanguineous parents (4 families in total). Patient1: male, homozygous for a 27 kb deletion including the newly characterized Maenli lncRNA. At birth, anomalies in the upper and lower limbs were detected. Patient 2: female, homozygous for a 27 kb deletion including the newly characterized Maenli lncRNA. At birth, anomalies in extremities leading to reduction of lower limbs and congenital anomalies in both hands were detected. Patient 3: male, homozygous for a 63 kb deletion including the newly characterized Maenli lncRNA. He did not have any obvious dysmorphism but had predominant limb malformations involving all 4 extremities. Patient 4: female, homozygous for c.317dupT/p.Ile107HisfsTer39 variant in the EN1 gene. Abnormalities in the upper and lower limbs were noticed at birth. At age 2 months, she presented tonic focal seizures. Severe hypotonia was noticed. A diagnosis of neurogenic bladder was made at the age of 3 months. At age 3 years and 10 months, clinical examination showed failure to thrive (weight and stature below 3rd percentile), microcephaly, poor visual contact, generalized hypotonia, no speech, no social smile; presence of a gastrostomy (G)-tube in the abdomen and umbilical hernia.
Recruitment	The 4 patients with congenital limb anomalies were observed at clinical genetic centers in Brazil and in India. Given the unique nature of their phenotypes, they were offered additional diagnostic investigations in accordance with regulations for studies on human subjects at the respective centers.
Ethics oversight	The study has been performed in accordance with national and international regulations. Informed consent was obtained from each individual or their legal guardian to participate in the study, for genomic analysis, and to publish fotos and clinical details. The study was approved by the local ethics committees of the participating centers in Lausanne (Commission cantonale d'éthique de la recherche sur l'etre humain), Switzerland and Berlin (Ethikkommission der Charité), Germany.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Datasets are available through the Gene Expression Omnibus (GEO) under accession number GSE137335.
Files in database submission	Brain-E95-WT-Mm9-Rep1-1-H3K27ac-ChIP-L16552_R1_001.bw Brain-E95-WT-Mm9-Rep1-2-H3K27ac-ChIP-L16552_R1_001.bw Brain-E95-WT-Mm9-Rep1-H3K4me3-ChIP-L16554_R1_001.bw Brain-E95-WT-Mm9-Rep2-1-H3K4me3-ChIP-L16555_R1_001.bw Brain-E95-WT-Mm9-Rep2-1-H3K27ac-ChIP-L16553_R1_001.bw Brain-E95-WT-Mm9-Rep2-2-H3K4me3-ChIP-L16555_R1_001.bw Brain-E95-WT-Mm9-Rep2-2-H3K27ac-ChIP-L16553_R1_001.bw FL-E95-WT-Mm9-Rep1-1-H3K4me3-ChIP-L14203_R1_001.bw FL-E95-WT-Mm9-Rep1-1-H3K27ac-ChIP-L16550_R1_001.bw FL-E95-WT-Mm9-Rep1-2-H3K4me3-ChIP-L14203_R1_001.bw FL-E95-WT-Mm9-Rep1-2-H3K27ac-ChIP-L16550_R1_001.bw FL-E95-WT-Mm9-Rep2-1-H3K27ac-ChIP-L16551_R1_001.bw FL-E95-WT-Mm9-Rep2-2-H3K27ac-ChIP-L16551_R1_001.bw FL-E95-WT-Mm9-Rep2-H3K4me3-ChIP-L15854_R1_001.bw FLHL-E115-3xpA-Mm9-Rep1-1-H3K27ac-ChIP-L16622_R1_001.bw

FLHL-E115-3xpA-Mm9-Rep1-2-H3K27ac-ChIP-L16622_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep1-CTCF-ChIP-L17110_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep1-H3K4me1-ChIP-L15850_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep1-H3K4me3-ChIP-L15852_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep1-Rad21-ChIP-L17114_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep2-1-H3K27ac-ChIP-L16623_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep2-2-H3K27ac-ChIP-L16623_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep2-CTCF-ChIP-L17111_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep2-H3K4me1-ChIP-L15851_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep2-H3K4me3-ChIP-L15853_R1_001.bw
FLHL-E115-3xpA-Mm9-Rep2-Rad21-ChIP-L17115_R1_001.bw
FLHL-E115-Del27-Mm9-Rep1-H3K4me1-ChIP-L14442_R1_001.bw
FLHL-E115-Del27-Mm9-Rep1-H3K4me3-ChIP-L14447_R1_001.bw
FLHL-E115-Del27-Mm9-Rep1-H3K27ac-ChIP-L16620_R1_001.bw
FLHL-E115-Del27-Mm9-Rep2-1-H3K27ac-ChIP-L16621_R1_001.bw
FLHL-E115-Del27-Mm9-Rep2-2-H3K27ac-ChIP-L16621_R1_001.bw
FLHL-E115-Del27-Mm9-Rep2-H3K4me1-ChIP-L14446_R1_001.bw
FLHL-E115-Del27-Mm9-Rep2-H3K4me3-ChIP-L14890_R1_001.bw
FLHL-E115-WT-Mm9-Rep1-CTCFT-ChIP-L17108_R1_001.bw
FLHL-E115-WT-Mm9-Rep1-H3K27ac-ChIP-L16618_R1_001.bw
FLHL-E115-WT-Mm9-Rep1-Rad21-ChIP-L17112_R1_001.bw
FLHL-E115-WT-Mm9-Rep2-1-H3K27ac-ChIP-L16619_R1_001.bw
FLHL-E115-WT-Mm9-Rep2-2-H3K27ac-ChIP-L16619_R1_001.bw
FLHL-E115-WT-Mm9-Rep2-CTCF-ChIP-L17109_R1_001.bw
FLHL-E115-WT-Mm9-Rep2-Rad21-ChIP-L17113_R1_001.bw
Brain-E95-WT-Mm9-Rep1-1-H3K27ac-ChIP-L16552_R1_001.fastq.gz
Brain-E95-WT-Mm9-Rep1-2-H3K27ac-ChIP-L16552_R1_001.fastq.gz
Brain-E95-WT-Mm9-Rep1-H3K4me3-ChIP-L16554_R1_001.fastq.gz
Brain-E95-WT-Mm9-Rep2-1-H3K4me3-ChIP-L16555_R1_001.fastq.gz
Brain-E95-WT-Mm9-Rep2-1-H3K27ac-ChIP-L16553_R1_001.fastq.gz
Brain-E95-WT-Mm9-Rep2-2-H3K4me3-ChIP-L16555_R1_001.fastq.gz
Brain-E95-WT-Mm9-Rep2-2-H3K27ac-ChIP-L16553_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep1-1-H3K4me3-ChIP-L14203_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep1-1-H3K27ac-ChIP-L16550_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep1-2-H3K4me3-ChIP-L14203_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep1-2-H3K27ac-ChIP-L16550_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep2-1-H3K27ac-ChIP-L16551_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep2-2-H3K27ac-ChIP-L16551_R1_001.fastq.gz
FL-E95-WT-Mm9-Rep2-H3K4me3-ChIP-L15854_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep1-1-H3K27ac-ChIP-L16622_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep1-2-H3K27ac-ChIP-L16622_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep1-CTCF-ChIP-L17110_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep1-H3K4me1-ChIP-L15850_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep1-H3K4me3-ChIP-L15852_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep1-Rad21-ChIP-L17114_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep2-1-H3K27ac-ChIP-L16623_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep2-2-H3K27ac-ChIP-L16623_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep2-CTCF-ChIP-L17111_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep2-H3K4me1-ChIP-L15851_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep2-H3K4me3-ChIP-L15853_R1_001.fastq.gz
FLHL-E115-3xpA-Mm9-Rep2-Rad21-ChIP-L17115_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep1-H3K4me1-ChIP-L14442_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep1-H3K4me3-ChIP-L14447_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep1-H3K27ac-ChIP-L16620_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep2-1-H3K27ac-ChIP-L16621_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep2-2-H3K27ac-ChIP-L16621_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep2-H3K4me1-ChIP-L14446_R1_001.fastq.gz
FLHL-E115-Del27-Mm9-Rep2-H3K4me3-ChIP-L14890_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep1-CTCFT-ChIP-L17108_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep1-H3K27ac-ChIP-L16618_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep1-Rad21-ChIP-L17112_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep2-1-H3K27ac-ChIP-L16619_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep2-2-H3K27ac-ChIP-L16619_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep2-CTCF-ChIP-L17109_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep2-Rad21-ChIP-L17113_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep1-H3K4me1-ChIP-L14891_R1_001.bw
FLHL-E115-WT-Mm9-Rep2-H3K4me1-ChIP-L14892_R1_001.bw
FLHL-E115-WT-Mm9-Rep1-H3K4me3-ChIP-L14893_R1_001.bw
FLHL-E115-WT-Mm9-Rep2-H3K4me3-ChIP-L14894_R1_001.bw
FLHL-E115-WT-Mm9-Rep1-H3K4me1-ChIP-L14891_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep2-H3K4me1-ChIP-L14892_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep1-H3K4me3-ChIP-L14893_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep2-H3K4me3-ChIP-L14894_R1_001.fastq.gz
FLHL-E115-WT-Mm9-Rep1-MLL1-ChIP-L15994_R1_001.bw
FLHL-E115-WT-Mm9-Rep1-MLL1-ChIP-L15994_R1_001.fastq.gz

Genome browser session
(e.g. UCSC)

No longer applicable.

Methodology

Replicates

Except for anti-MLL1 immunoprecipitation, we performed 2 indepedent biological replicates for each immunoprecipitation. Each biological replicate is composed of a pull of different limb buds from at least three different embryos.

Sequencing depth

All samples were 75 bp single-end reads sequencing. The sequencing depth for each experiment was approximately 50 million total reads. We obtained between 15 to 30 million uniquely mapped reads for each biological sample.

Antibodies

The following specific antibodies were used for this study: anti-H3K4me3, Millipore, cat. No. 07-473; anti-H3K4me1, Diagenode, cat. No. C15410037; anti-H3K27ac, Diagenode, cat. No. C15410174; anti-CTCF, Active Motif, cat. No. 61311; anti-Rad21, abcam, cat.No. ab992; anti-MLL1, Bethyl Laboratories, cat. No. A300-086A

Peak calling parameters

No peak calling was performed in this study.

Data quality

Experiments were checked for similiar global enrichment folds between the different conditions using spp cross-correlation plots and also visually inspected for reproducibility by comparing binding profiles of inspected proteins at loci/regions that were not potentially affected by the inserted mutations.

Software

Single-end ChIP-seq reads were mapped with bowtie (v2.2.6) to mm9, filtered for mapping quality MAPQ≥10, and duplicates were removed using samtools rmdup (v1.8). For generating coverage tracks, reads were extended to 300bp (H3K27ac, H3K4me1, and H3K4me3) or 200bp (CTCF,Rad21, and MLL1) and scaled to r.p.m. (106 per number of unique reads) using bedtools genomecov v2.27.1.