Enhancer redundancy provides phenotypic robustness in mammalian development

Permalink
https://escholarship.org/uc/item/3z44x6vv

Journal
NATURE, 554(7691)

ISSN
0028-0836

Authors
Osterwalder, M
Barozzi, I
Tissieres, V
et al.

Publication Date
2018-02-08

DOI
10.1038/nature25461

Peer reviewed
Enhancer redundancy provides phenotypic robustness in mammalian development

Marco Osterwalder1, Iros Barozzi1, Virginie Tissières2,3, Yoko Fukuda–Yuzawa1, Brandon J. Mannion1, Sarah Y. Afzal1, Elizabeth A. Lee1, Yiwen Zhu1, Ingrid Plajzer–Frick1, Catherine S. Pickle1, Momoe Kato1, Tyler H. Garvin1, Quan T. Pham1, Anne N. Harrington1, Jennifer A. Akiyama1, Veena Afzal1, Javier Lopez–Rios2,3, Diane E. Dickel1, Axel Visel1,4,5 & Len A. Pennacchio1,4,6

Distant-acting tissue-specific enhancers, which regulate gene expression, vastly outnumber protein-coding genes in mammalian genomes, but the functional importance of this regulatory complexity remains unclear1,2. Here we show that the pervasive presence of multiple enhancers with similar activities near the same gene confers phenotypic robustness to loss-of-function mutations in individual enhancers. We used genome editing to create 23 mouse deletion lines and inter-crosses, including both single and combinatorial enhancer deletions at seven distinct loci required for limb development. Unexpectedly, none of the ten deletions of individual enhancers caused noticeable changes in limb morphology. By contrast, the removal of pairs of limb enhancers near the same gene resulted in discernible phenotypes, indicating that enhancers function redundantly in establishing normal morphology. In a genetic background sensitized by reduced baseline expression of the target gene, even single enhancer deletions caused limb abnormalities, suggesting that functional redundancy is conferred by additive effects of enhancers on gene expression levels. A genome-wide analysis integrating epigenomic and transcriptomic data from 29 developmental mouse tissues revealed that mammalian genes are very commonly associated with multiple enhancers that have similar spatiotemporal activity. Systematic exploration of three representative developmental structures (limb, brain and heart) uncovered more than one thousand cases in which five or more enhancers with redundant activity patterns were found near the same gene. Together, our data indicate that enhancer redundancy is a remarkably widespread feature of mammalian genomes that provides an effective regulatory buffer to prevent deleterious phenotypic consequences upon the loss of individual enhancers.

Enhancers are a principal class of cis-regulatory elements that orchestrate precise gene expression patterns, which are essential for numerous processes including embryonic development3. They are now routinely predicted by genome-wide chromatin profiling methods, which identify positions of open chromatin or enhancer-associated histone marks5. Enhancers predicted by these high-throughput approaches outnumber genes by approximately an order of magnitude1, raising the question of their functional significance. In particular, it remains unclear whether mammalian enhancers typically regulate numerous processes including embryonic development2,4–7, or if this regulatory complexity more commonly results in functional redundancy among enhancers associated with the same gene6,8–10.

Using the developing limb as a model for gene regulation during morphogenetic processes8,11,12, we investigated the functional importance of enhancers in vivo. We used CRISPR–Cas9 genome editing to

Figure 1 | Lack of limb morphological abnormalities in ten enhancer deletion lines. a, All selected enhancers are active in the limb mesenchyme (blue shading) at E11.5, are marked by epigenomic H3K27ac and DNase I hypersensitivity (DHS) at E11.5, and contain a conserved core sequence (Cons). Target gene expression and limb morphology were assessed following deletion of individual enhancers (Extended Data Fig. 1a–j). b, None of the individual enhancer deletions caused obvious defects in the structure of skeletal elements. Enhancers are identified by VISTA ID numbers. Enhancer activities (left, E11.5) and forelimb skeletons of enhancer knockout (KO) embryos (right, E18.5) are shown (see Extended Data Fig. 3 for wild-type controls). Predicted target gene and enhancer distance (+, downstream; −, upstream) from the transcriptional start site (TSS) are indicated. n represents independent biological replicates with similar results. Scale bars, 100 μm (E11.5), 1 mm (E18.5).
individually delete ten embryonic enhancers, each with strong evolutionary conservation and robust limb activity in transgenic mouse reporter assays\(^{13-17}\) (VISTA Enhancer Browser: https://enhancer.lbl.gov/) (Fig. 1a, Extended Data Fig. 1a–j and Supplementary Table 1). Each enhancer (identified by VISTA ID number) is located in the vicinity of a gene associated with human congenital limb malformations, and deletion of these genes in mice results in limb phenotypes ranging from polydactyly (Gli3) to complete loss of limbs (E11.5) overlaps spatial RNA expression of the associated target gene, suggesting that these enhancers are part of the regulatory architecture that controls the expression of these genes\(^{16-21}\) (Extended Data Fig. 2). Capture-C chromatin conformation data from embryonic limbs\(^{22}\) confirmed that at least six of these enhancers physically interacted with their predicted target genes (Extended Data Fig. 1k). This framework enabled us to investigate the functional contribution of each enhancer by comparing the potential limb skeletal abnormalities caused by enhancer loss to the phenotypes observed in gene knockout mice.

Unexpectedly, we did not detect any abnormalities in bone number, shape, length, position or mineralization in mice in which any of the ten single enhancers was deleted (Fig. 1b and Extended Data Fig. 3). Similarly, we observed neither significant differences in predicted target gene expression in embryonic limbs for nine out of ten individual enhancer deletions, nor obvious changes in local H3K27ac (acetylation of lysine 27 on histone H3) signatures outside the deleted enhancers (Extended Data Figs 2, 4). Together, these results suggest that a substantial proportion of limb enhancers, even if highly conserved in evolution, are not individually essential for normal limb morphogenesis.

One possible explanation for the lack of an obvious phenotype in individual limb enhancer knockout lines is that different enhancers associated with the same gene may have spatiotemporally redundant, rather than unique, activity. Our selected panel of enhancers (Fig. 1b and Extended Data Fig. 1a–j) included three enhancer pairs with overlapping limb activity domains and the same predicted target gene (mm1179–hs1586, hs741–hs1262, and hs1467–mm636; Extended Data Fig. 5a–d, g, j), such that both deletions occurred in cis. In two out of three cases, involving enhancer pairs near Gli3 and Shox2, homozygous DKO embryos showed phenotypic abnormalities affecting skeletal limb morphology (Fig. 2a–d and Extended Data Fig. 5a–c). Using iterative CRISPR–Cas9 genome editing, we generated double enhancer knockout (DKO) mice for each enhancer pair (Extended Data Fig. 5a–d, g, j), such that both deletions occurred in cis. In two out of three cases, involving enhancer pairs near Gli3 and Shox2, homozygous DKO embryos showed phenotypic abnormalities affecting skeletal limb morphology (Fig. 2a–d and Extended Data Fig. 5a–c). Using iterative CRISPR–Cas9 genome editing, we generated double enhancer knockout (DKO) mice for each enhancer pair (Extended Data Fig. 5a–c). Using iterative CRISPR–Cas9 genome editing, we generated double enhancer knockout (DKO) mice for each enhancer pair (Extended Data Fig. 5a–c).

To examine the degree of overlap between the activity patterns of phenotypically redundant enhancers at the cellular level, we generated transgenic mouse lines expressing fluorescent reporters under the
control of each of the Gli3 or Shox2 enhancers (mm1179–GFP, hs1586–mCherry, hs741–GFP and hs1262–mCherry). Using immunofluorescence on limb sections from double transgenic embryos, we tracked the activity of each of the four enhancers during limb development (Fig. 2e, f and Extended Data Fig. 6). Consistent with the preaxial polydactyly observed in Gli3 DKO embryos, limb progenitor cells marked by both Gli3 enhancers were observed at high density in the anterior limb mesenchyme (Fig. 2e and Extended Data Fig. 6c, d). In Shox2 double enhancer reporter embryos, a major accumulation of cells with dual Shox2 enhancer activities is present in a proximal limb mesenchymal cell population known to harbour stylopod progenitors12 (Fig. 2f). In conjunction with our deletion studies, these results illustrate the degree of functional overlap between pairs of enhancers near the same gene at the cellular level.

Considering the apparent contrast between the morphological redundancy of pairs of enhancers and the strong evolutionary conservation of each individual enhancer, we studied the phenotypic effect of single and combinatorial enhancer deletions in sensitized genetic backgrounds carrying heterozygous deletions of the presumptive target genes (Fig. 3). We used CRISPR–Cas9 to engineer Gli3 and Shox2 gene loss-of-function alleles, which recapitulated expected gene dosage reductions and previously published phenotypes (Extended Data Figs 7, 8). We then used these alleles to generate compound heterozygous mice harbouring one or more disrupted enhancers with a wild-type gene on one allele and a disrupted gene but wild-type enhancers on the other allele (Fig. 3). For Gli3, the absence of either enhancer (mm1179 or hs1586) in the presence of only one functional Gli3 allele resulted in a supernumerary anterior digit (Fig. 3a and Extended Data Fig. 8a), which is more severe than the terminally bifurcated thumb observed in Gli3 heterozygotes (Fig. 3a). Similarly, for Shox2 the removal of either neighbouring enhancer (hs1262 or hs741) in combination with compound heterozygous deletion of the Shox2 gene resulted in a more pronounced reduction in femur length than observed in Shox2 heterozygotes (Fig. 3b). For both pairs of enhancers, compound heterozygous mice carrying deletions of both enhancers on one allele and a deletion of the gene on the other allele showed even more severe phenotypes. In the case of Gli3, loss of both enhancers over a Gli3 null allele resulted in greatly reduced expression of Gli3 (Extended Data Fig. 7b, c) and severe pre-axial polydactyly in forelimbs, similar in severity to homozygous loss of the Gli3 gene23 (Fig. 3a and Extended Data Fig. 8a). Likewise, compound heterozygous deletion of enhancers hs741 and hs1262 over a Shox2 gene deletion strongly reduced Shox2 expression (Extended Data Fig. 7e, f) and resulted in a severe reduction in femur length and substantial shortening of the humerus (Fig. 3b and Extended Data Fig. 8b, c), consistent with the phenotypes that result from homozygous Shox2 gene loss18,25. Together, our data demonstrate that these developmental enhancers, although seemingly dispensable under non-sensitized conditions, show individual functional contributions to limb development under conditions of reduced genetic robustness.

The lack of phenotypic change upon deletion of individual enhancers, and the functional redundancy observed among enhancer pairs, raises the question of how commonly such redundancy occurs in mammalian gene regulatory landscapes. To explore this question systematically, we devised a genome-wide, correlation-based computational approach to estimate the number of enhancers that regulate each gene
Enhancers with redundant signatures are prevalent near developmental genes. a. Enhancer–gene assignments based on correlation of H3K27ac and mRNA profiles across a wide array of tissues (Extended Data Fig. 9a). Top, at an example locus encompassing Tbx3, Tbx5, and Lhx5, up to 25 enhancers are assigned to each of these three genes (blue, pink and brown boxes, Extended Data Fig. 9c). Genes showing fewer than five assigned enhancers are shown in grey. Bottom, heat maps showing meta-profiles of each gene’s expression profile across tissues (red shades), along with the cumulative activity profile of its assigned enhancers (blue shades). b. Distribution of the number of enhancers assigned to developmental transcription factors (TFs) with biased expression in limb ($P = 5 \times 10^{-25}$ versus housekeeping), forebrain ($P = 8 \times 10^{-15}$), and heart ($P = 3 \times 10^{-25}$) (two-sided Mann–Whitney tests). Box plots show median, interquartile values, range, and outliers (individual points). c. Complete spectrum of genes with at least one assigned enhancer, sorted by decreasing enhancer number. Limb-biased transcription factors are highlighted in green. d. Total number of enhancers (in all tissues analysed) assigned to each transcription factor in e, with the number of assigned enhancers predicted specifically in limb at E11.5 (dark green) or any other stage analysed (light green).

Figure 4 | Enhancers with redundant signatures are prevalent near developmental genes. a. Enhancer–gene assignments based on correlation of H3K27ac and mRNA profiles across a wide array of tissues (Extended Data Fig. 9a). Top, at an example locus encompassing Tbx3, Tbx5, and Lhx5, up to 25 enhancers are assigned to each of these three genes (blue, pink and brown boxes, Extended Data Fig. 9c). Genes showing fewer than five assigned enhancers are shown in grey. Bottom, heat maps showing meta-profiles of each gene’s expression profile across tissues (red shades), along with the cumulative activity profile of its assigned enhancers (blue shades). b. Distribution of the number of enhancers assigned to developmental transcription factors (TFs) with biased expression in limb ($P = 5 \times 10^{-25}$ versus housekeeping), forebrain ($P = 8 \times 10^{-15}$), and heart ($P = 3 \times 10^{-25}$) (two-sided Mann–Whitney tests). Box plots show median, interquartile values, range, and outliers (individual points). c. Complete spectrum of genes with at least one assigned enhancer, sorted by decreasing enhancer number. Limb-biased transcription factors are highlighted in green. d. Total number of enhancers (in all tissues analysed) assigned to each transcription factor in e, with the number of assigned enhancers predicted specifically in limb at E11.5 (dark green) or any other stage analysed (light green).

Our observations have implications for the interpretation of noncoding regulatory variants in relation to human phenotypes. Our findings suggest that many loss-of-function enhancer mutations will cause, at most, subtle phenotypes in humans. Thus, for many genetic loci, enhancer-associated disease phenotypes may be more likely to result from gain-of-function mutations that either expand enhancer activity or alter the positions of enhancers relative to genes.

Received 27 March; accepted 18 December 2017. Published online 31 January 2018.


Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by National Institutes of Health grants R01HG003988, U54HG006997, R24HL123879 and U111HL098166 (to A.V. and L.A.P.) and the University of Basel and the Novartis Foundation for Biomedical Research (to J.L.-R.). M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. We thank B. Ren for providing access to A.V. and L.A.P. and the University of Basel and the Novartis Foundation for Biomedical Research (to J.L.-R.). M.O. was supported by a Swiss National Science Foundation (SNSF) fellowship. We thank B. Ren for providing access to A.V. and L.A.P. and the University of Basel and the Novartis Foundation for Biomedical Research (to J.L.-R).
METHODS
Experimental design. All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory (LBNL) Animal Welfare Committee. All mice used in this study were housed at the Animal Care Facility (ACF) at LBNL. Mice were monitored daily for food and water intake, and animals were inspected weekly by the Chair of the Animal Welfare and Research Committee and the head of the animal facility in consultation with the veterinary staff. The LBNL ACF is accredited by the American Association for the Accreditation of Laboratory Animal Care (AAALAC). Transgenic mouse embryos were injected with optimized CRISPR–Cas9 reagents as described previously45 using CHOPCHOP45 (see Supplementary Table 1 for sgRNA sequences and coordinates). Knockout mice were engineered as described previously46 using the TRUSeq Stranded mRNA Sample Prep Kit (Illumina), following the manufacturer’s instructions, and purified, eluted, and quantified as described previously47. RNA-seq libraries were generated for 400–600 ng of total RNA with Illumina’s TruSeq Stranded mRNA Kit. RNA-seq data were analyzed using the TopHat2 and Cufflinks tools.

Transgenic mouse assay selection and randomization. Sample sizes were selected empirically on the basis of our previous experience of performing transgenic mouse assays for more than 2,000 total putative enhancers (VISTA Enhancer Browser: https://enhancer.lbl.gov/). Mouse embryos were excluded from further analysis if they did not contain the reporter transgene or if the developmental stage was not correct. All transgenic mice were treated with identical experimental conditions. Randomization and experimenter blinding were unnecessary and not performed.

Enhancer knockout selection and randomization. Sample sizes were selected empirically on the basis of our previous studies15. All phenotypic characterization of knockout mice used a matched littermate selection strategy. All phenotyped mice described in the paper resulted from crossing heterozygous enhancer deletion mice together to allow the comparison of matched littermates of different genotypes. Embryonic samples used for ChIP–seq hybridizations, RNA-seq, and skeletal preparations were dissected blinded to genotype.

In vivo transgenic reporter assays. Enhancer names in this study are the unique identifiers used in the VISTA Enhancer Browser (https://enhancer.lbl.gov/). Embryo enhancer activities (in situ hybridizations) were assayed using an hsp68-LacZ-based reporter vector similar to that described above, with the exception of a fluorescent reporter replacing LacZ. The enhancer–reporter combinations were generated as follows: mm1179–sGFP, hs1586–mCherry, hs741–sGFP and hs1262–mCherry. sGFP is a fusion of Sun1 and 2×sGFP as described42 and localizes to the nuclear membrane. Mice carrying the individual fluorescent reporter transgenes were then generated via pronuclear injection (using FVB strain zygotes), and stable lines were established from founders showing reproducible reporter activity in the embryonic limb.

Generation of enhancer knockout mice using CRISPR–Cas9. Mouse strains lacking limb enhancer(s) or harbouring gene loss-of-function alleles were generated using in vivo CRISPR–Cas9 editing, as previously described, with only minor modifications43,44. Parental lines used in this study were the VISTA enhancers (hs1586 at E10.5 and hs1262) were amplified from human genomic DNA and cloned into an hsp68-LacZ expression vector as previously described44. Transgenic mouse embryos were injected with sgRNAs targeting the secondary enhancer for deletion. hs1586 and hs1262 were located on chromosome 18 and 20 of the mouse genome, respectively. All phenotypic characterization of knockout mice used a matched littermate selection strategy. All phenotyped mice described in the paper resulted from crossing heterozygous enhancer deletion mice together to allow the comparison of matched littermates of different genotypes. Embryonic samples used for ChIP–seq hybridizations, RNA-seq, and skeletal preparations were dissected blinded to genotype.

In situ hybridization and skeletal preparations. To assess spatial changes in gene expression in mouse embryonic limbs, whole mount in situ hybridization using digoxigenin-labelled antisense riboprobes was carried out as previously described46. Forelimbs and hindlimbs from at least three independent embryos were analysed for each genotype (including wild-type littermate controls). Mouse embryonic skeletons at E18.5 were stained with Alcian blue and Alizarin red to differentiate cartilage (blue) and bone (red) using standard methods47. For comparison of limb skeletons from enhancer knockout embryos and wild-type littermates, general parameters such as bone number, shape, length, position or mineralization were assessed. Embryonic limbs and limb skeletons were imaged, and skeletal elements were measured, using a Leica MZ6 stereo-microscope coupled to a Leica DFC300FX or DFC420 digital camera. Brightness and contrast were adjusted uniformly across the sample set using Photoshop CS5. Measurements of the ossified portions of humerus and femur (stipodial elements) were normalized to those of the ulna and tibia (related zygomatic elements), respectively (as shown in Figs 2d, 3b and Extended Data Figs 5a, 8c).

Quantitative real-time PCR (qPCR) and RNA-seq. RNA was isolated from microdissected forelimbs or hindlimbs of mouse embryos at E11.5 using the Ambion RNeasy Total RNA Isolation Kit (Life Technologies) according to the manufacturer’s instructions. For qPCR, RNA was treated with RNase-free DNase (Promega) and reverse transcribed using SuperScript III (Life Technologies) with random hexamer or poly-dT priming according to the manufacturer’s instructions. qPCR was performed on a LightCycler 480 (Roche) using KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems) according to the manufacturer’s instructions. qPCR primers (listed in Supplementary Table 4) were designed in silico using Primer3 (http://primer3.wi.mit.edu/), and amplification span exon–exon junctions in order to prevent amplification of genomic DNA. Relative gene expression levels were calculated using the 2−ΔΔct method48, normalized to the Actb housekeeping gene, and the mean of wild-type control samples was set to 1.

For RNA-seq, RNA samples were treated with DNase (TURBO DNA-free Kit, Life Technologies), and RNA quality was verified using a 2100 Bioanalyzer (Agilent) with an RNA 6000 Nano Kit (Agilent). RNA-seq libraries were generated using the TruSeq Stranded mRNA Sample Prep Kit (Illumina), following the manufacturer’s instructions, and purified, eluted, and quantified as described previously47. RNA-seq libraries were pooled (0.5× per lane) and sequenced using single-end 50-bp reads on a HiSeq4000 (Illumina). Immunofluorescence. Mouse embryonic limbs at E10.5, E11.5 or E12.5 were dissected in cold PBS and fixed in 4% PFA for 2–3 h. Following incubation in a sucrose gradient and embedding in a 1:1 mixture of 30% sucrose and optimum cutting solution temperature, sagittal 10-μm frozen sections were cut using a cryostat. Cryosections were incubated overnight with the following primary antibodies: chicken anti-GFP (1:500, Thermo Fisher Scientific, A10262), rabbit anti-mCherry (1:1,000, Thermo Fisher Scientific, PA5-34979) and goat anti-sox-9 (1:500, R&D Systems, AF3075). Goat-anti chicken, goat anti-rabbit and donkey anti-goat secondary antibodies conjugated to Alexa Fluor 488, 568, 594 or 647 (1:1,000, Thermo Fisher Scientific) were used for detection. Hoechst 33258 (Sigma-Aldrich) was used to counterstain nuclei. Fluorescent images were acquired using a Zeiss Axiolab fluorescent microscope in combination with a Hamamatsu Orca-03 camera. Brightness and contrast were adjusted uniformly using Photoshop CS5.

ChIP–seq. For each of six single enhancer knockout lines, ChIP–seq to H3K27ac was performed using a protocol optimized for mouse embryonic tissues49. In brief, forelimbs and hindlimbs of wild-type embryos were dissected at E11.5, formaldehyde crosslinked, and sheared using a Diagenode Bioruptor. Sonication was performed on a rotator for 30 min at 4°C. Freshly rinsed Dynabeads (1:1 protein A:protein G mix) were then added to the antibody-treated chromatin, and immunoprecipitation was performed on a rotator for 60 min at 4°C. Libraries were prepared using the Illumina Truseq DNA sample prep kit following the manufacturer’s instructions with minor modifications. Library quality was assessed using a 2100 Bioanalyzer with the High Sensitivity DNA Kit (Agilent), and quantification was performed using a Qubit Fluorometer with the dsDNA HS Assay Kit (Life Technologies). ChIP–seq and input libraries were pooled and sequenced paired-end 50 bp using the TruSeq Stranded mRNA Sample Prep Kit (Illumina). RNA-seq data were analysed using a combination of the Ballgown and DESeq2 packages (https://www.bioconductor.org/packages/Ballgown/).
top 100 differentially expressed genes, sorted by false discovery rate (FDR), are listed in Supplementary Tables 5–7.

For read mapping and peak calling of ChIP–seq datasets, bowtie2 (version 0.12.8) with parameter ‘-m 1 -v 2’ and MACS2 (version 1.4.2) with parameter ‘-mfold = 10.30 -nomodel -p 0.0001’ were used, respectively. Biological replicates were combined using MSCP56, with the following parameters: ‘-r biological -s 1E-10 -W 1E-6 -n Highest -c 2’. The predicted enhancer intervals were assigned the best P-value (as defined by MACS2) among the overlapping peaks.

ENCODEnRNA-seq data analysis. Raw data were downloaded from the Data Coordination Center of the ENCODE project (https://www.encodeproject.org/, see Supplementary Table 8 for the complete list of sample identifiers). Short reads were aligned to the mm10 assembly of the mouse genome using bowtie2, with the following parameters: ‘-a -m 1 -n 2 -1 32 -e 3001. Peak calling was performed using MACS v1.4.5, with the following arguments: ‘-gsize = mm-bw = 300–nomodel -shiftsize = 100’. Experiment-matched input DNA was used as a control.

ENCODEnRNA-seq data analysis. Raw data were downloaded from the ENCODE Data Coordination Center (https://www.encodeproject.org/, see Supplementary Table 8 for the complete list of sample identifiers). Short reads were aligned to the mm10 assembly of the mouse genome using TopHat v2.0.8.57 and Gencode vM358 as the reference transcriptome. Cuffnorm v2.2.151 was run to quantify transcripts across conditions using the Gencode vM3 transcriptome as the reference and setting ‘library-norm-method to geometric. Only genes with a level of expression of at least one RPM (reads per kilobase of exons per million mapped reads) in at least one of the considered conditions were included in further analyses. Small and non-coding RNAs were excluded by retaining only those genes with a Gencode biotype58 supporting protein-coding functionality.

Classifying genes by tissue-biased patterns of expression. For each protein-coding gene in the mouse genome, the expression variability across the twenty-nine ENCODE RNA-seq experiments from multiple tissues and developmental time points was evaluated using two metrics: a measure of tissue-specificity (τ)59 ranging from 0 (consistent expression across all conditions) to 1 (expression in one single condition); and a measure of relative expression in a condition of interest (for example, limb at E11.5). Given a gene, the latter was defined as the difference between the percentile of expression of the gene in the given condition and the median percentile of expression across all the samples. A large positive number indicates a gene that is much more expressed in the condition of interest than the average.

Tissue-biased genes were defined as showing τ ≥ 0.7 and relative expression higher than the 95th percentile. Housekeeping genes were defined as having τ < 0.4 and relative expression between the 5th and 95th percentiles. The complete lists of genes assigned to each category are available in Supplementary Table 9.

Gene classification based on pre-specified functional categories. Tissue-biased developmental transcription factors (sometimes referred to as tissue-specific transcription factors) were defined as genes with biased expression in a given tissue (see previous section), associated with abnormal developmental phenotypes in the same tissue (terms extracted from the Mouse Genome Informatics (MGI) database60, listed in Supplementary Table 10) and annotated as a transcription factor under the terms GO:0003700 or GO:0005705 in the Gene Ontology (GO)61. Annotations were downloaded from GO and MGI on July 7, 2016.

Topologically associated domains. TAD coordinates26 estimated from mouse embryonic stem cell Hi-C data were downloaded from http://chromosome.sdsce.edu/mouse/hi-c/download.html. Coordinates were converted from mm9 to mm10 using liftOver62.

A statistical framework defining enhancer–promoter associations genome-wide. A list of putative enhancer regions was first defined as follows: after excluding any region annotated to the mitochondrial or any random chromosome, the BED coordinates of the H3K27ac peaks across the twenty-nine conditions (different combinations of tissue and developmental stage as defined by the ENCODE consortium, see ‘ENCODEnChIP–seq data analysis’ above) were merged using the mergeBed utility from BEDTools v.2.10.063. For a more robust signal estimation (see below), regions shorter than 500 bp were enlarged to 1 kb from their central coordinate. Promoters, defined as regions within 2.5 kb of the transcriptional start sites of genes annotated in Gencode vM358, were then excluded using subtractBed from BEDTools v.2.17.0. After that, any remaining region shorter than 1 kb was excluded. Uniquely aligned, de-duplicated reads were then used to quantify the H3K27ac signals at each region, for each one of the 29 conditions. These signals were measured using the coverageBed utility from BEDTools v.2.17.063, normalized to RPMK (according to the sequencing depth of each specific sample), and log2-transformed. The resulting list of 74,366 predicted enhancers and their corresponding H3K27ac signal quantifications, along with the mRNA expression measurements for the protein-coding genes (as defined in ‘Classifying genes by tissue-biased patterns of expression’), were used as input for the statistical framework described below. The main steps of the approach are also outlined in Extended Data Fig. 9b.

For each previously defined TAD in the mouse genome, we retrieved all of the enhancers predicted and the genes expressed in at least one of the twenty-nine conditions considered that fell within that TAD. Pairwise correlations between all possible enhancer–gene combinations within the TAD were then evaluated by calculating Spearman’s rank correlation coefficient (SCC) between the H3K27ac pattern of enrichment at the enhancer and the mRNA expression of the gene associated with it. Each putative enhancer was initially assigned to the gene showing the highest SCC value (in the very rare case of ties, all of the genes showing the same SCC value were assigned to the enhancer). After that, a null distribution of SCC values was estimated empirically, by pairing the enhancer with 1,000 randomly picked genes from the same chromosome. The z-score for the correlation coefficient was then calculated by subtracting the mean and dividing by the standard deviation estimated from the empirical null. The corresponding P-value was calculated using the pnorm function in R. Finally, only those putative enhancers showing a P-value < 0.05 and a SCC = 0.25 were retained, resulting in a set of 34,882 enhancers with an assigned target (Supplementary Table 11). Considering the entire, genome-wide set of pairwise associations, a P = 0.05 corresponds to a Benjamini–Hochberg corrected FDR of 0.087. This analysis resulted in the assignment of one or more putative enhancers to 9,365 protein-coding genes (Supplementary Table 12). To define a set of genes with many redundant enhancers, we considered enhancers as redundant only if they were associated with the same gene by correlation and showed a strong peak of H3K27ac in the same exact tissue under examination (for example, both enhancers are active in limb and linked to the Gli3 gene). Although this correlative approach may result in a subset of false-positive assignments for individual genes, it enables an approximation of both regulatory complexity and potential enhancer redundancy. For each condition, we found 1,276 genes that showed multiple assigned enhancers such that at least five of the enhancers were all active in the same tissue (limb, heart or brain). We then used a permutation scheme to directly evaluate the statistical robustness of this conclusion (that is, 1,276 genes with 5 or more redundant enhancers in either developing limbs, heart or forebrain), which considered increasingly higher correlation values between the activity of putative enhancers and expression of genes (Extended Data Fig. 10c–f). By re-shuffling the expression values of each gene across conditions (100 genome-wide permutations), we estimated the FDR of observing a gene with five or more enhancers attached to it, for increasingly lower correlation coefficients. Each permutation consisted of the same enhancers and genes, in which the H3K27ac values were left as in the actual data whereas the RNA expression values of the genes across the different samples were randomly reshuffled. For each genome-wide permuted matrix, the entire statistical approach described above was re-run and a map of enhancer–promoter associations was generated. For each value of Spearman’s correlation coefficient (0.25 to 0.75, with a 0.01 step) the number of genes showing five or more enhancers in the permuted data was calculated. The average across the 100 iterations was then computed and used for FDR estimation. This was calculated as the average number of genes showing five or more enhancers across the permuted data, over the number of genes derived from the actual data.

Statistical analysis. Statistical analyses are described in detail in the Methods sections above. Whenever a P value is reported in the text, the statistical test is also indicated. Unless specified otherwise, all the statistics were estimated and plots drawn using the statistical computing environment R (https://www.r-project.org) or GraphPad Prism 7 software.

Data availability. ChIP–seq and RNA-seq datasets are available in the NCBI GEO database with the accession code GSE93730. Additional data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Extended Data Figure 1 | CRISPR deletion of ten limb enhancers and regulatory interaction landscape of associated target genes.

a–j, Left, representative activity patterns of the selected enhancers in mouse embryos at E11.5 (VISTA enhancer browser)\textsuperscript{13} and the respective genomic enhancer regions tested in transgenic assays (Tg, blue bar), along with the regions deleted in enhancer knockout mice (Del, red bar). Corresponding H3K27 acetylation patterns (green) in wild-type mouse embryonic forelimbs at E11.5 (this study) are depicted with open chromatin (ENCODE DHS in forelimbs at E11.5, purple) and the Placental Mammal basewise conservation track by PhyloP (Cons, blue/red). Scale bars, 500 bp. VISTA enhancer IDs (mm and hs numbers) are indicated on the left, with the distance of the enhancer from the transcriptional start site of the predicted target gene in the mouse genome. Numbers at the bottom right of each embryo indicate the reproducibility of the enhancer reporter assay. Arrowheads mark additional activity domains (other than limb): hs1262 (hindbrain, reproducibility: 5/6, also shown previously\textsuperscript{17}), mm917 (dorsal root ganglion, 7/7) and hs1603 (nose, 7/7; and branchial arch, 5/7). Asterisk indicates potential craniofacial enhancer activity for mm636, which was observed in 3 of 9 embryos\textsuperscript{64}. Right, PCR validation strategy and results for enhancer knockout lines. Red scissors indicate CRISPR-mediated deletion breakpoints. PCR was used to detect the wild-type (+) and enhancer deletion (Δ) alleles. Below, Sanger sequencing traces show the deletion breakpoints (indicated by the dashed line) for the enhancer knockout alleles. PCR genotyping results are shown with amplicon sizes indicated on the left (enhancer deletion allele in red). Primers (Ctrl or Ctrl2) amplifying an unrelated genomic region were included as a PCR positive control. See Supplementary Table 3 for all primer sequences and related PCR product sizes. k, Top, Hi-C interaction heat maps of topologically associated chromatin domains (mouse embryonic stem cell TADs)\textsuperscript{26}. Bottom, selected enhancers (blue triangles) and their predicted target genes (TSS indicated as black bar). The Capture-C UCSC browser track (purple) illustrates three-dimensional chromatin interaction profiles from E11.5 embryonic limbs (3-kb window) using promoters of the predicted enhancer target genes as anchor points\textsuperscript{22}. H3K27ac enrichment (green) in wild-type forelimbs at E11.5 (this study) is shown below. Six of the ten enhancers selected for deletion analysis display local Capture-C enrichment (*), indicating physical interaction with the predicted target gene promoter at E10.5 or E11.5, based on the stringent statistical approach (95th percentile threshold) applied in the original study\textsuperscript{22}. Other genes present in the TAD are shown in grey.
Extended Data Figure 2 | No major differences in expression of predicted target genes in individual enhancer knockouts. a, Spatial enhancer activity domains (LacZ, see also Fig. 1b) are compared to mRNA expression domains (by *in situ* hybridization) of the predicted target genes in embryonic forelimbs and hindlimbs at E11.5. No significant changes in expression patterns were observed in enhancer knockouts compared to wild-type limbs, except in limbs lacking hs741, where a small subdomain of target gene expression was lost (red arrowhead marks loss of the posterior *Shox2* domain in the distal limb, compared with black arrowhead in wild type). Transcript distribution was reproduced in at least n = 3 independent biological replicates. b, Quantitative real-time PCR using limbs of homozygous null (KO, red dots) and wild-type (Wt, blue dots) embryos at E11.5 reveals lack of significantly downregulated transcript levels of predicted enhancer target genes in nine out of ten cases. Box plots indicate median, interquartile values, range and individual biological replicates. Outliers are shown as circled data points. **P = 0.0012, unpaired, two-tailed t-test. n.s., not significant. Scale bars, 100 μm.
Extended Data Figure 3 | Absence of obvious morphological abnormalities in limb enhancer knockouts. Side-by-side comparison of enhancer knockout limb skeletons and wild-type littermate controls at E18.5. Neither forelimbs (this figure) nor hindlimbs (data not shown) of the enhancer knockout lines revealed any obvious morphological differences in comparison to wild-type littermates. Cartilage is stained blue and bone dark red. The number of embryos with normal limb phenotypes over the total number of homozygous-null embryos examined is shown in the bottom left. n represents number of independent biological replicates with similar results. Scale bar, 1 mm.
Extended Data Figure 4 | Absence of compensatory enhancer signatures in limbs of enhancer knockout embryos. a, Layered ChIP–seq H3K27 acetylation (ac) profiles surrounding the deleted enhancers and from wild-type (blue, \( n = 4 \) independent biological replicates) and enhancer knockout embryos (orange, at least \( n = 2 \) biological replicates). For all samples, E11.5 forelimb was profiled. For display, replicates were merged using bigWigMerge (UCSC tools) and normalized. Red triangles indicate the positions of individual enhancer deletions. b, H3K27ac enrichments in targeted regions marked by red triangles in a, showing the absence of H3K27ac at the deletion site in individual enhancer knockout (orange) compared to wild-type (blue) samples. Blue bars indicate locations of enhancer sequences. Dashed red lines demarcate the regions deleted by CRISPR. Vertebrate basewise conservation track by PhyloP (Cons) is shown.
Extended Data Figure 5 | Transcriptional and phenotypic impact of dual enhancer deletions engineered by iterative CRISPR–Cas9 genome editing. a–c, Top, enhancer pairs with overlapping limb activities (LacZ), coinciding with domains of predicted target gene expression visualized by in situ hybridization (ISH). For Sox9 enhancers, black arrowheads indicate overlapping domains. Schematics, double enhancer deletion strategy to delete the three enhancer pairs with overlapping activity (see Methods). Grey numbers indicate enhancer distance (kb) from the TSS. Bottom, Sanger sequencing verification of the secondary enhancer deletion. Deletion breakpoint is marked by the dashed line. Grey horizontal bars indicate bases present in the primary deletions (single enhancer knockout lines, see Extended Data Fig. 1a–j).
d–f, RNA-seq confirmed significantly reduced Gli3 expression in hand plates of DKO embryos but not individual enhancer knockout embryos (compared to wild-type hand plates). f, Unaffected hindlimb morphology in mm1179/hs1586 DKO embryos. Red arrowhead points to digit 1 duplication in forelimbs (see also Fig. 2). g, Shox2 expression (in situ hybridization) in forelimbs and hindlimbs of hs741/hs1262 DKO embryos. The distal-posterior domain (arrowhead) is dependent on hs741 (Extended Data Fig. 2a). h, Reduced Shox2 expression in forelimbs and hindlimbs of hs741/hs1262 DKO embryos (qPCR). Expression of the nearby Rsrc1 gene was unchanged. i, Left, representative limb skeletons of wild-type and hs741/hs1262 DKO embryos. Hu, humerus; Ul, ulna; Fe, femur; Ti, tibia. Right, mild but significant reduction in humerus ossification length (double arrows) in hs741/hs1262 DKO limb skeletons. n.s., not significant.

*P = 1.66 × 10⁻⁷ (two-tailed, unpaired t-test). 

© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
Extended Data Figure 6 | Cellular resolution of redundant Gli3 enhancer activities at the onset of digit formation. **a, b**, Individual Gli3 enhancer activities as detected by immunofluorescence (mm1179, green; hs1586, red) in forelimbs of transgenic reporter embryos. Sox9 (grey) marks chondrogenic progenitors of the mesenchymal condensations forming digit primordia (digits 1–5, from anterior to posterior). 
**c, d**, Co-localization of mm1179 and hs1586 enhancer activities in hand plates of double enhancer transgenic embryos. Close-ups (right) show that the anterior mesenchyme (Fig. 2c) harbours many cells with dual enhancer activities (yellow). A fraction of double enhancer-positive cells carries the signature of Sox9 digit progenitors (white, bottom). n = 3 independent embryos per genotype were analysed, with similar results. Nuclei, detected via Hoechst staining, are blue. Scale bars, 100 μm (a, b); 50 μm (c, d).
Extended Data Figure 7 | Generation of Gli3 and Shox2 knockout alleles and characterization of enhancer deletions in a sensitized background.

a, d, Top, schematic showing CRISPR–Cas9-mediated deletions used to generate Gli3 and Shox2 loss-of-function alleles. Genotyping primers used to validate targeted deletion events are indicated. Bottom, Sanger sequencing confirmation of deletion event, with grey and red dashed lines indicating breakpoints. Right, PCR genotyping examples with the size of the product specific for the deletion allele depicted in red (primers listed in Supplementary Table 3).

b, e, In situ hybridization showing the gradual decrease in anterior Gli3 transcript in forelimbs of wild-type, Gli3Δ/+ and sensitized mm1179/hs1586 DKO (DKO/Gli3Δ) embryos. Arrowheads point to the domains where Shox2 expression is nearly abolished in enhancer DKO/Shox2Δ embryos.
f, qPCR revealing significantly downregulated Shox2 mRNA levels in hindlimbs of DKO/Shox2Δ compared to Shox2Δ/+ embryos. *P < 0.05; **P < 0.001; ***P < 0.001 (two-tailed, unpaired t-test). n.s., not significant. For in situ hybridization, transcript distribution was reproduced in at least n = 3 independent biological replicates. Scale bars, 100 μm.
Extended Data Figure 8 | See next page for caption.
Extended Data Figure 8 | Limb phenotypes of individual and combinatorial Gli3 and Shox2 enhancer knockouts in the presence of reduced target gene dosage. a, Skeletal phenotypes resulting from mm1179 and hs1586 enhancer deletions in combination with reduction to one copy of the Gli3 gene at E18.5. Genotypes are shown on the left with red crosses indicating elements deleted by CRISPR–Cas9. While forelimbs of Gli3Δ/+ embryos displayed bifurcated digit 1 terminal phalanges, hindlimbs showed an extra toe structure but without detectable cartilage template. Four out of seven mm1179Δ/Gli3Δ embryos displayed additional bifurcation of digit 2 of the right forelimb (a), which suggests that removal of mm1179 reduces Gli3 levels in the anterior forelimb more than deletion of hs1586. An almost complete anterior extra toe formed in hindlimbs of embryos with single or dual enhancer deletions in the sensitized background (black asterisks). Loss of both Gli3 copies resulted in anterior hindlimb polydactyly with altered digit identities (red asterisks). b, Allelic series depicting shortening of the stylopod (humerus and femur) in limb skeletons with individual or combined hs741 and hs1262 enhancer deletions in a Shox2 sensitized condition (see also Fig. 3b). Stylopod ossification length (double arrows) appears less reduced in forelimbs (humerus, Hu) than in hindlimbs (femur, Fe) of embryos lacking the activity of both enhancers (hs741Δ/, hs1262Δ/Shox2Δ). Tibia (Ti) and ulna (Ul) were normal in all genotypes examined. c, Humerus ossification length (normalized to ulna ossification length) is significantly reduced in embryos lacking either hs741 or hs1262 in the presence of only one copy of Shox2. In embryos lacking both enhancers in the sensitized background, significant shortening of humerus ossification is observed (compared to all other genotypes). n indicates the number of independent biological replicates with similar results. Box plots indicate median, interquartile values, range and individual biological replicates. ***P < 0.001; *P < 0.05 (two-tailed, unpaired t-test). Scale bars, 500 μm.
Extended Data Figure 9 | See next page for caption.
Extended Data Figure 9 | A correlative framework to define enhancer–promoter associations across the mouse genome. **a**, The TAD including the transcriptional regulators *Tbx3*, *Tbx5* and *Lhx5* illustrates the statistical framework to define enhancer–promoter associations genome-wide. For each predicted enhancer, correlation between its H3K27ac signal (blue arrowhead, blue-shaded heat map) with the mRNA expression profiles of every gene in the TAD (red-shaded heat map) across all available tissues and developmental stages was assessed. The enhancer was then assigned to the most highly correlated gene, *Tbx3* in the case of enhancer 3. **b**, Schematic depicting the underlying statistical framework used to determine genome-wide enhancer–promoter interactions (see Methods). **c**, Activity pattern for the enhancers assigned to *Tbx3*, *Tbx5* and *Lhx5*. Genomic coordinates are listed on the right. For each predicted enhancer–gene pair, Spearman’s correlation coefficient (SCC, *n* = 29) and the corresponding empirically estimated *P* value (from 1,000 random enhancer–gene pairings) are shown in Supplementary Table 11. **d**, Identifying genes with biased expression in embryonic limb, forebrain, or heart. Expression variability across 29 RNA-seq datasets from multiple tissues and developmental time points, measures of tissue specificity (Tau (τ), x-axis) and specific tissue-biased expression at E11.5 (y-axis) for each protein-coding gene were calculated (see Methods). Housekeeping genes were defined as displaying τ ≤ 0.4 and relative expression in the limb between the 5th and 95th percentiles. Tissue-biased genes were defined as showing τ ≥ 0.7 and relative expression higher than the 95th percentile. **d**, Distribution of enhancer numbers assigned to each gene, for the different gene categories. Genes with tissue-biased expression profiles were associated with a significantly higher number of enhancers than housekeeping genes. *P* = 4 × 10⁻¹²¹ (*n* = 553), *P* = 7 × 10⁻⁹⁷ (*n* = 626) and *P* = 6 × 10⁻⁸³ (*n* = 826) for limb, forebrain and heart biased genes, respectively (two-sided Mann–Whitney tests), *n* = 1,287 for housekeeping genes. Box plots indicate median, interquartile values and range. Outliers are shown as individual points.
Extended Data Figure 10 | See next page for caption.
Extended Data Figure 10 | Enhancer redundancy as a widespread feature of developmental genes and robustness to the choice of thresholds used in the correlative approach. a, b, Top, number of enhancers assigned to each gene through the correlative framework, with developmental transcription factors (TFs) showing biased expression in forebrain (a, blue dots) or heart (b, orange dots) indicated. Classification of tissue-biased developmental transcription factors is described in Methods. Genes with at least one assigned enhancer are displayed and sorted according to the number of assigned enhancers (left to right). Bottom, bar plot showing the total number of enhancers assigned to each of the transcription factors highlighted in the top panels. For each gene, a colour code shows the number of predicted enhancers assigned to that gene in the relevant tissue (a, heart; b, forebrain) at E11.5 (dark colour), in the relevant tissue at any other developmental stage included in the analysis (light colour), or in any other tissue (white). c, Estimated FDR (based on genome-wide permutations, see Methods) of observing a gene with five or more enhancers assigned to it, for increasingly larger correlation coefficients (0.25 to 0.75). The red solid line indicates an FDR of 0.05. The red arrow and the black dashed line highlight the lowest correlation coefficient (0.47, considering a step of 0.01) with an FDR ≤ 0.05 (FDR = 0.0495). d, Number of genes showing five or more enhancers assigned to them, for increasingly larger correlation coefficients (0.25 to 0.75). The total number of genes (SCC ≥ 0.25) along with the number of genes identified using the threshold set in c (SCC > = 0.47) is indicated (1,276 and 1,058, respectively; see Supplementary Tables 11, 12). e, Bubble plot showing the number of genes with five or more enhancers assigned to them, at increasingly higher correlation between enhancer and target gene expression (x-axis) and between enhancers assigned to the same gene (y-axis). f, Bubble plot displaying the fold-enrichment (linear) for developmental transcription factor genes among each set in c.
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. Sample size
   Describe how sample size was determined.
   Methods, section "Experimental Design" (pages 23 and 24).
   Specific information related to the different experimental approaches can be obtained in the individual methods sections.

2. Data exclusions
   Describe any data exclusions.
   Methods, section "Experimental Design" (pages 23 and 24).
   Specific information related to the different experimental approaches can be obtained in the individual methods sections.

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   Attempts at replication were successful and statistical parameters reproducibility numbers are indicated in figure panels/text as required.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Methods, section "Experimental Design" (pages 23 and 24).

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Methods, section "Experimental Design" (pages 23 and 24).
   Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

   - The exact sample size ($n$) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
   - A statement indicating how many times each experiment was replicated
   - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons
   - The test results (e.g. $P$ values) given as exact values whenever possible and with confidence intervals noted
   - A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
   - Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.
Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

Statistics were estimated and the plots were drawn using the statistical computing environment R (www.r-project.org). Bar plots and some of the box plots were drawn using the GraphPad Prism software (www.graphpad.com).

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Materials used are readily available from the authors or from standard commercial sources (see Methods for specific reagents).

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

For ChIP-seq the anti-H3K27ac antibody from Active Motif (cat no. 39133) was used:
It has been validated and used e.g. in the following study: e.g. Kuwahara, M., Ise, W., et al. (2016), 'Bach2-Batf interactions control Th2-type immune response by regulating the IL-4 amplification loop.', Nat Commun, 7, pp. 12596

For Immunofluorescence the following primary antibodies have been used (see Methods, page 33):
chicken 15 anti-GFP (1:500, Thermo Fisher Scientific, A10262), rabbit anti-mCherry (1:1,000, Thermo Fisher Scientific, PA5-34974) and goat anti-Sox9 (1:500, R&D Systems, AF3075).
These antibodies were validated in several publications and are all widely used for immuno-detection of the corresponding proteins in mouse tissues (see websites of vendors).

For in situ hybridization a standard protocol involving the Anti-Digoxigenin-AP antibody (Roche, 11093274910) was used (see Methods page 32 for reference, described in detail in Kvon et al., Cell 2016).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used in this study.

b. Describe the method of cell line authentication used.

No eukaryotic cell lines were used in this study.

c. Report whether the cell lines were tested for mycoplasma contamination.

No eukaryotic cell lines were used in this study.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

No eukaryotic cell lines were used in this study.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

See Methods, section "Experimental Design" (page 23). Experiments were performed in Mus musculus FVB strain mice. The following developmental stages were used in this study: embryonic day E10.5, E11.5, E12.5 and E18.5 mice. Animals of both sexes were used in the analysis and adult mice involved in breedings were up to 1 year of age.
Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study did not involve human research participants.
ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

- **Data deposition**

  1. For all ChIP-seq data:
     a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.
     b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

  2. Provide all necessary reviewer access links.
     *The entry may remain private before publication.*

  3. Provide a list of all files available in the database submission.

All ChIP-seq files (raw data and BED files) have been deposited in the GEO database (accession GSE93730) as stated in the Methods section of the manuscript. The secure token to access these files is the following: wlbwucztklnoz

<table>
<thead>
<tr>
<th>File ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSM2461145</td>
<td>Wt_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461146</td>
<td>Wt_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461153</td>
<td>Wt_replicate_03_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461154</td>
<td>Wt_replicate_04_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461149</td>
<td>Wt_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461150</td>
<td>Wt_replicate_02_input</td>
</tr>
<tr>
<td>GSM2461159</td>
<td>Wt_replicate_03_input</td>
</tr>
<tr>
<td>GSM2461160</td>
<td>Wt_replicate_04_input</td>
</tr>
<tr>
<td>GSM2461147</td>
<td>mm1179KO_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461148</td>
<td>mm1179KO_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461157</td>
<td>mm1179KO_replicate_03_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461158</td>
<td>mm1179KO_replicate_04_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461151</td>
<td>mm1179KO_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461152</td>
<td>mm1179KO_replicate_02_input</td>
</tr>
<tr>
<td>GSM2461163</td>
<td>mm1179KO_replicate_03_input</td>
</tr>
<tr>
<td>GSM2461164</td>
<td>mm1179KO_replicate_04_input</td>
</tr>
<tr>
<td>GSM2461155</td>
<td>hs2KQ_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461156</td>
<td>hs2KQ_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461161</td>
<td>hs2KQ_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461162</td>
<td>hs2KQ_replicate_02_input</td>
</tr>
<tr>
<td>GSM2461129</td>
<td>hs1586KO_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461137</td>
<td>hs1586KO_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461133</td>
<td>hs1586KO_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461141</td>
<td>hs1586KO_replicate_02_input</td>
</tr>
<tr>
<td>GSM2461130</td>
<td>hs1262KO_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461138</td>
<td>hs1262KO_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461134</td>
<td>hs1262KO_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461142</td>
<td>hs1262KO_replicate_02_input</td>
</tr>
<tr>
<td>GSM2461131</td>
<td>hs1467KO_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461139</td>
<td>hs1467KO_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461135</td>
<td>hs1467KO_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461143</td>
<td>hs1467KO_replicate_02_input</td>
</tr>
<tr>
<td>GSM2461132</td>
<td>hs1603KO_replicate_01_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461140</td>
<td>hs1603KO_replicate_02_ChIP_H3K27ac</td>
</tr>
<tr>
<td>GSM2461136</td>
<td>hs1603KO_replicate_01_input</td>
</tr>
<tr>
<td>GSM2461144</td>
<td>hs1603KO_replicate_02_input</td>
</tr>
</tbody>
</table>
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

<table>
<thead>
<tr>
<th>Methodological details</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Describe the experimental replicates.</td>
</tr>
<tr>
<td>6. Describe the sequencing depth for each experiment.</td>
</tr>
<tr>
<td>7. Describe the antibodies used for the ChIP-seq experiments.</td>
</tr>
<tr>
<td>8. Describe the peak calling parameters.</td>
</tr>
<tr>
<td>9. Describe the methods used to ensure data quality.</td>
</tr>
<tr>
<td>10. Describe the software used to collect and analyze the ChIP-seq data.</td>
</tr>
</tbody>
</table>