







# Modeling uniquely human gene regulatory function via targeted humanization of the mouse genome

Emily V. Dutrow <sup>1,8</sup>, Deena Emera<sup>1,9</sup>, Kristina Yim<sup>1</sup>, Severin Uebbing <sup>1</sup>, Acadia A. Kocher<sup>1</sup>, Martina Krenzer<sup>1,10</sup>, Timothy Nottoli <sup>2,3</sup>, Daniel B. Burkhardt<sup>1,11</sup>, Smita Krishnaswamy <sup>1,4</sup>, Angeliki Louvi <sup>5,6</sup> & James P. Noonan <sup>1,6,7</sup>✉

The evolution of uniquely human traits likely entailed changes in developmental gene regulation. Human Accelerated Regions (HARs), which include transcriptional enhancers harboring a significant excess of human-specific sequence changes, are leading candidates for driving gene regulatory modifications in human development. However, insight into whether HARs alter the level, distribution, and timing of endogenous gene expression remains limited. We examined the role of the HAR *HACNS1* (HAR2) in human evolution by interrogating its molecular functions in a genetically humanized mouse model. We find that *HACNS1* maintains its human-specific enhancer activity in the mouse embryo and modifies expression of *Gbx2*, which encodes a transcription factor, during limb development. Using single-cell RNA-sequencing, we demonstrate that *Gbx2* is upregulated in the limb chondrogenic mesenchyme of *HACNS1* homozygous embryos, supporting that *HACNS1* alters gene expression in cell types involved in skeletal patterning. Our findings illustrate that humanized mouse models provide mechanistic insight into how HARs modified gene expression in human evolution.

<sup>1</sup>Department of Genetics, Yale School of Medicine, New Haven, CT 06510, USA. <sup>2</sup>Department of Comparative Medicine, Yale School of Medicine, New Haven, CT 06510, USA. <sup>3</sup>Yale Genome Editing Center, Yale School of Medicine, New Haven, CT 06510, USA. <sup>4</sup>Department of Computer Science, Yale University, New Haven, CT 06520, USA. <sup>5</sup>Department of Neurosurgery, Yale School of Medicine, New Haven, CT 06510, USA. <sup>6</sup>Department of Neuroscience, Yale School of Medicine, New Haven, CT 06510, USA. <sup>7</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520, USA. <sup>8</sup>Present address: Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA. <sup>9</sup>Present address: Center for Reproductive Longevity and Equality, Buck Institute for Research on Aging, Novato, CA 94945, USA. <sup>10</sup>Present address: Neuroscience Research Training Program, Department of Psychiatry, Yale School of Medicine, New Haven, CT 06510, USA. <sup>11</sup>Present address: Cellarity, Cambridge, MA 02139, USA. ✉email: [james.noonan@yale.edu](mailto:james.noonan@yale.edu)

The evolution of uniquely human physical traits required human-specific genetic changes that altered development<sup>1,2</sup>. Discovering the locations of these changes in the genome and determining their biological impact is a major challenge. However, over the last decade comparative studies have begun to reveal potential genetic drivers underlying novel human biological features. These efforts have identified a prominent class of elements in the genome that are highly conserved across many species but show a significant excess of human-specific sequence changes<sup>3–7</sup>. These elements, collectively named Human Accelerated Regions (HARs), are prime candidates to encode novel human molecular functions. Many HARs act as transcriptional enhancers during embryonic development, particularly in structures showing human-specific morphological changes such as the brain and limb<sup>7–13</sup>. HARs have also been shown to exhibit human-specific changes in enhancer activity<sup>7,9,10,14–19</sup>. These findings suggest a critical contribution for HARs in human evolution and support the long-standing hypothesis that changes in developmental gene regulatory programs contribute to evolutionary innovation<sup>20,21</sup>.

Despite these advances, the role of HARs in altering regulatory function in vivo remains poorly understood. Here we employed a knock-in approach in mouse to characterize the effects of the HAR *HACNS1* on gene expression and regulation compared to its chimpanzee ortholog during embryonic development (Fig. 1A). A similar approach has been used to model the transcriptional and developmental effects of changes in enhancer activity in other mammalian lineages, notably bats<sup>22</sup>. We chose to model *HACNS1* (also known as HAR2 and 2xHAR.3), as it exhibits the strongest acceleration signature of any noncoding HAR yet identified (Fig. 1B). *HACNS1* was also the first HAR demonstrated to exhibit a human-specific gain in enhancer activity during development. In a mouse transgenic enhancer assay, *HACNS1* was shown to drive increased expression of a *LacZ* reporter gene in the embryonic mouse limb compared to its chimpanzee and rhesus macaque orthologs<sup>10</sup>. Furthermore, *HACNS1* exhibits increased levels of histone H3K27 acetylation (H3K27ac), which is correlated with enhancer activity, in human versus rhesus macaque and mouse embryonic limb<sup>11</sup>. Together, these findings suggest that *HACNS1* may have contributed to changes in limb development during human evolution.

To compare the functions of *HACNS1* and its chimpanzee and mouse orthologs in the same developmental system, we used homologous recombination to replace the endogenous mouse sequence with the human or chimpanzee counterpart. We found that *HACNS1* maintains its human-specific enhancer activity in the mouse embryo and alters the expression of the nearby transcription factor encoding gene *Gbx2* in limb chondrocytes, a cell type required for skeletal morphogenesis. Our findings support that HARs can direct changes in endogenous gene expression during development and illustrate the power of genetically humanized mouse models to provide insight into regulatory pathways and developmental mechanisms modified in human evolution.

## Results

**Generating a *HACNS1* knock-in mouse model.** We designed a targeting construct for homologous recombination including a 1.2 kb human sequence encompassing *HACNS1* that was previously shown to encode human-specific enhancer activity in transgenic mouse embryos<sup>10</sup>. We replaced the orthologous mouse locus using homology-directed repair in C57BL6/J-*A<sup>w</sup>-J/J* (B6 agouti) embryonic stem (ES) cells (Fig. 1B, Supplementary Figs. 1A, B, Supplementary Data 1; Methods). To provide a control that would enable us to distinguish bona fide human-specific functions of *HACNS1* from possible primate-rodent differences,

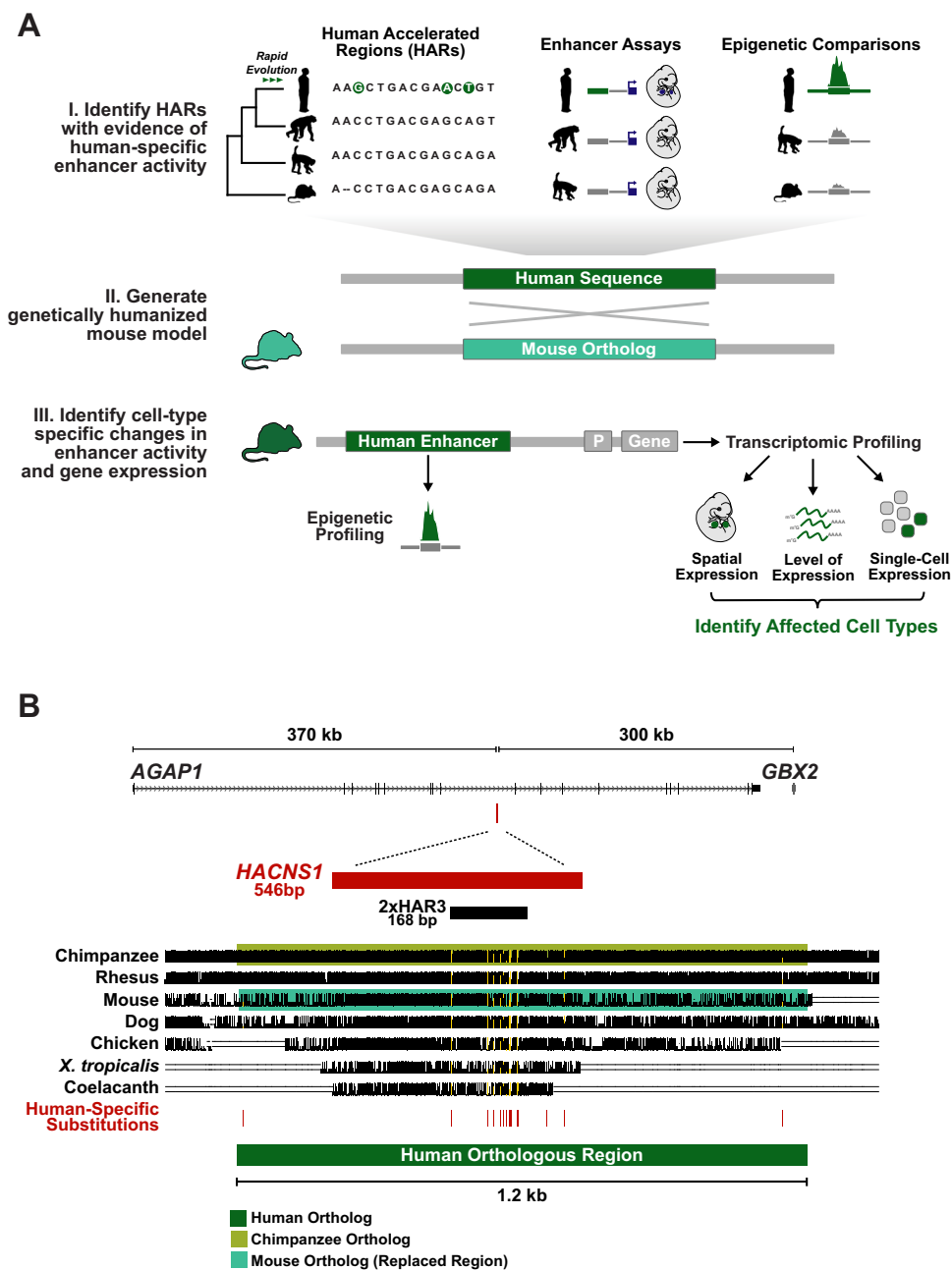
we also generated a mouse model for the orthologous chimpanzee sequence using the same approach (Supplementary Figs. 1A–C). The 1.2 kb chimpanzee sequence shows no evidence of evolutionary acceleration and includes 22 single nucleotide differences relative to the human sequence<sup>3</sup>; 15 of these are human-specific based on comparisons to other primate genomes (Methods, Supplementary Data 2). Previous studies indicate that multiple human-specific substitutions contribute to the gain of function in *HACNS1*<sup>10</sup>. Twelve of the 15 substitutions introduce one or more predicted transcription factor binding sites to the human sequence (Supplementary Data 2, Supplementary Fig. 1A). An extensive comparison of sequence conservation and divergence among the human, chimpanzee, and mouse sequences is provided in the Supplementary Note (Supplementary Materials).

In order to verify the integrity of the edited loci, we sequenced a 40 kb region encompassing the human or chimpanzee sequence replacement, the homology arms used for targeting, and flanking genomic regions in mice homozygous for either *HACNS1* or the chimpanzee ortholog (Supplementary Fig. 1C; Methods). We found no evidence of aberrant editing, sequence rearrangements, or other off-target mutations at either edited locus. We also verified that each homozygous line carried two copies of the human or chimpanzee sequence using quantitative real-time PCR (RT-qPCR) (Supplementary Fig. 1D; associated Source Data).

***HACNS1* maintains its human-specific enhancer activity in the mouse embryo.** We used chromatin immunoprecipitation (ChIP) to determine if *HACNS1* exhibits epigenetic signatures of increased enhancer activity in mice. We first performed epigenetic profiling in the developing mouse limb bud based on prior evidence that *HACNS1* drives increased reporter gene activity in transgenic enhancer assays and exhibits increased H3K27ac marking in the human embryonic limb<sup>10,11</sup>. We profiled both H3K27ac and H3K4 dimethylation (H3K4me2), which is also associated with enhancer activity, in embryonic day (E) 11.5 limb buds from embryos homozygous for *HACNS1*, embryos homozygous for the chimpanzee ortholog, and wild type embryos. We found a strong signature of H3K27ac marking at *HACNS1* in the limb buds of *HACNS1* homozygous embryos (Fig. 2, Supplementary Fig. 2). The chimpanzee and mouse sequences both showed significant but weaker H3K27ac enrichment relative to the human sequence, supporting the conclusion that *HACNS1* maintains its human-specific enhancer activity in the mouse genomic and developmental context.

We used DESeq2 implemented in HOMER (Methods) to identify genome-wide significant differences in H3K27ac and H3K4me2 levels in E11.5 limb buds from mice homozygous for *HACNS1* or the chimpanzee ortholog versus wild type<sup>23,24</sup>. We found that H3K27ac and H3K4me2 levels were significantly increased at the edited *HACNS1* locus compared to the endogenous mouse locus (Fig. 2, Supplementary Fig. 2A, B, Supplementary Data 3). In contrast, the level of H3K27ac at the edited chimpanzee locus was not significantly different than that at the endogenous locus (Fig. 2, Supplementary Fig. 2A, Supplementary Data 3). The levels of H3K4me2 were significantly increased at both the humanized and orthologous chimpanzee loci in each respective line compared to mouse (Fig. 2, Supplementary Fig. 2, Supplementary Data 3). As high levels of H3K4me2 coupled with low levels of H3K27ac are associated with weak enhancer activity<sup>25</sup>, it is likely that the chimpanzee sequence is not acting as a strong enhancer in the limb bud overall, a finding further supported by the gene expression analyses described in Fig. 3.

Previous transgenic mouse enhancer assays indicated that *HACNS1* drives increased reporter gene activity in the pharyngeal

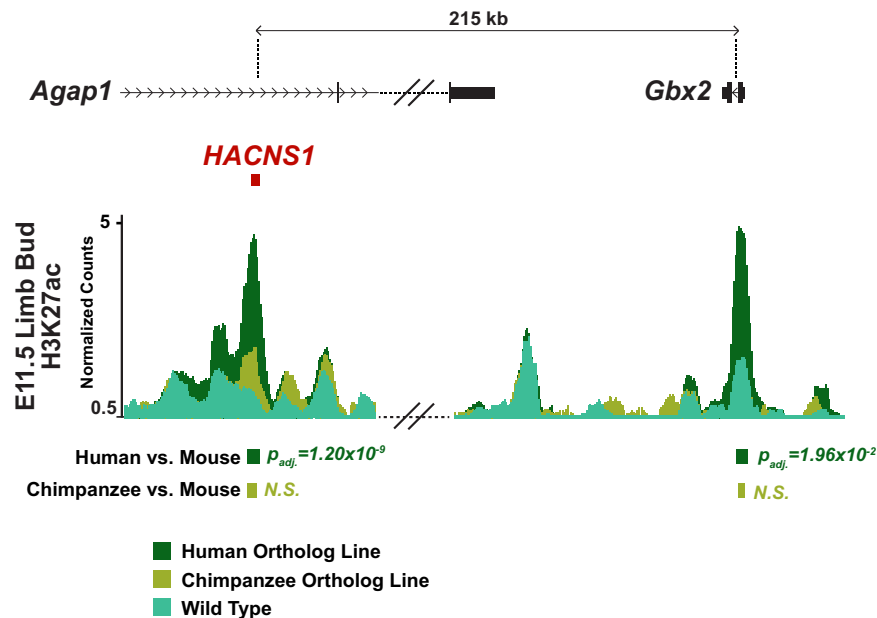


**Fig. 1** Generating a knock-in mouse model for the Human Accelerated Region *HACNS1*. **A** Schematic illustrating the generalized workflow we developed to characterize the gene regulatory functions of HARs with prior evidence of human-specific enhancer activity, which we applied to *HACNS1* in this study. **B** The location of *HACNS1* in the human genome (GRCh37/hg19) relative to the nearby genes *AGAP1* and *GBX2*. Below, alignment of the human sequence used to generate the *HACNS1* knock-in mouse with orthologous sequences from other vertebrate genomes, obtained from the UCSC hg19 100-way Multiz alignment (see Supplementary Data 1 for coordinates). The chimpanzee orthologous sequence used to generate the chimpanzee control line is highlighted in olive, and the mouse sequence replaced in each line is highlighted in teal. The location of each human-specific substitution is indicated by a red line, and the corresponding positions in the alignment are highlighted in yellow. The locations of *HACNS1* and 2xHAR3 are shown above the alignment<sup>3,6</sup>.

arch compared to the chimpanzee ortholog<sup>10</sup>. We therefore profiled H3K27ac and H3K4me2 in pharyngeal arch tissue from E11.5 embryos homozygous for either *HACNS1* or the chimpanzee ortholog. We detected reproducible, significant enrichment of H3K27ac in the pharyngeal arch at the humanized and orthologous chimpanzee loci compared to input controls, but the H3K27ac signal at neither the human nor the chimpanzee ortholog locus was significantly different compared to the mouse endogenous locus (Supplementary Fig. 2). We did, however, identify a significant gain of H3K4me2 signal in the pharyngeal arch at the humanized and orthologous chimpanzee

loci compared to the mouse locus (Supplementary Fig. 2, Supplementary Data 3).

In order to identify downstream epigenetic changes resulting from *HACNS1* activation in the limb bud, we searched for other genome-wide gains of H3K27ac and H3K4me2 at enhancers and promoters. We identified a significant gain of H3K27ac in *HACNS1* versus wild type limb buds at the promoter of the nearby gene *Gbx2* (Fig. 2, Supplementary Fig. 2, Supplementary Data 3). While significant H3K27ac enrichment was found in all three lines at the *Gbx2* promoter compared to input controls, H3K27ac levels were not significantly increased at *Gbx2* in limb



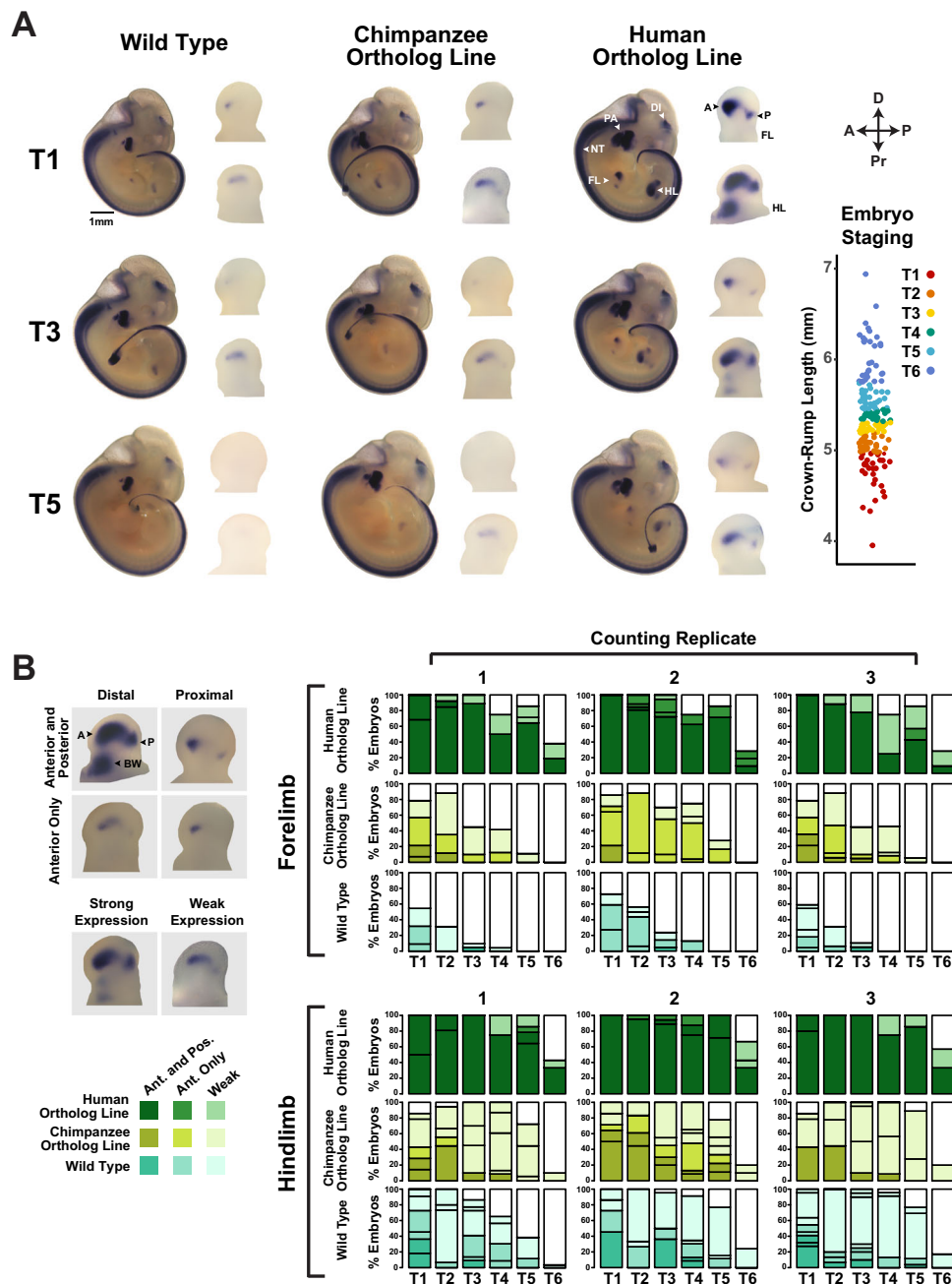
**Fig. 2 Epigenetic signatures of increased activity at *HACNS1* and the *Gbx2* promoter in the *HACNS1* homozygous mouse limb bud.** Epigenetic profiling in the *HACNS1* homozygous E11.5 limb bud compared to the chimpanzee ortholog line and wild type. The normalized H3K27ac signals are shown for the *HACNS1* line (in dark green), the chimpanzee ortholog line (in olive), and wild type (in teal) (see “Methods”). The location of the edited *HACNS1* locus in the human ortholog line relative to nearby genes is shown above the track. The double slanted lines indicate intervening H3K27ac signal data between the edited and wild type loci and *Gbx2* that were removed for clarity; see Supplementary Fig. 2 for complete views for each line as well as input signals. H3K27ac peak calls showing significant increases in signal between *HACNS1* homozygous and wild type, and the corresponding peak regions compared between the chimpanzee control line and wild type, are shown below the signal track. Litter-matched embryos were used for each comparison (see “Methods”). N.S. not significant. All peak calls for each line are shown in Supplementary Fig. 2. Adjusted *P* values were obtained using DESeq2 (implemented in HOMER) with a Wald test followed by Benjamini-Hochberg correction<sup>23,24</sup>.

buds with the chimpanzee ortholog compared to wild type, indicating the gain of activity is specific to *HACNS1*. H3K4me2 was also enriched at the promoter of *Gbx2* in all three lines compared to input controls (Supplementary Fig. 2B). After multiple testing correction, we did not identify any significant differentially marked regions outside of the *HACNS1-Gbx2* locus between either the *HACNS1* homozygous or the chimpanzee ortholog line compared to wild type for each chromatin mark in either tissue (Supplementary Data 3). Moreover, we did not detect a significant increase in either H3K27ac or H3K4me2 levels at the *Gbx2* promoter in the pharyngeal arch in *HACNS1* homozygous embryos (Supplementary Figs. 2C, D). This may be due to a lack of statistical power to detect small differences in histone modification levels given the number of replicates in the analysis, or our use of whole tissues to map histone modification profiles, which could obscure spatially restricted changes.

*Gbx2* encodes a transcription factor with a highly complex expression pattern and multiple functions in the developing embryo. GBX2 has been implicated in midbrain and hindbrain development<sup>26,27</sup>, guidance of thalamocortical projections<sup>28,29</sup>, ear development<sup>30</sup>, and pharyngeal arch patterning<sup>31</sup>. *Gbx2* is expressed in developing mouse limb at E10.5; however, its role in limb development remains undetermined as no limb phenotype has been reported in *Gbx2* knockout mice<sup>27</sup>. *HACNS1* and *GBX2* are located in the same topologically associated domain (TAD), and TADs have been shown to restrict enhancer interactions to genes within their boundaries<sup>32,33</sup>. In E11.5 limbs from *HACNS1* homozygous mice, the only significant increases in H3K27ac we detected in this TAD were at the *HACNS1* knock-in locus and at the endogenous *Gbx2* promoter (Supplementary Data 3). Together, these results suggest that *Gbx2* is a regulatory target of *HACNS1*, evoking the hypothesis that the gain of function in *HACNS1* might alter *Gbx2* expression in the limb.

***HACNS1* drives spatial and quantitative changes in *Gbx2* expression in the limb bud.** To visualize potential expression changes resulting from *HACNS1*-driven upregulation of the *Gbx2* promoter in *HACNS1* homozygous mouse embryos, we used in situ hybridization (ISH) (Fig. 3). We analyzed *Gbx2* expression in >50 E11.5 embryos per genotype (Fig. 3A, B; Methods and associated Source Data). In wild type embryos, we observed single foci of *Gbx2* expression in forelimb and hindlimb (Fig. 3A, left). In contrast, embryos homozygous for *HACNS1* showed substantially increased *Gbx2* expression in both forelimb and hindlimb (Fig. 3A, right). Embryos homozygous for the chimpanzee ortholog showed a weak increase in *Gbx2* expression compared to wild type (Fig. 3A, B). *Gbx2* expression in *HACNS1* homozygous embryos was increased in two distinct anterior and posterior regions in the forelimb and hindlimb bud, as well as an anterior proximal region in the latter. Overall, *Gbx2* expression in the limb bud was temporally dynamic in embryos of all genotypes. Embryos from the same litter vary in developmental age such that individual embryos collected at E11.5 range from E11 to E12. Therefore, we established a fine staging scheme to characterize changes in *Gbx2* expression within this short developmental interval. We assigned embryos to 6 temporally ordered groups (designated T1-T6, and ranging from ~40 to 48 somites, although we did not use somite counts to stage embryos; see Methods for more details) according to crown-rump length and used a blinded approach to qualitatively assess staining patterns (Fig. 3A, B and associated Source Data)<sup>10,34-36</sup>.

We identified differences in the distribution of *Gbx2* expression in the forelimb and hindlimb buds of *HACNS1* compared to both chimpanzee ortholog and wild type embryos across all 6 developmental time points (Fig. 3B). At the earliest time point (T1), we found that *Gbx2* was strongly expressed in distinct anterior-distal and posterior domains in *HACNS1* forelimb and



**Fig. 3** Spatial and temporal changes in *Gbx2* expression driven by *HACNS1* in *HACNS1* knock-in mouse embryos. **A** Spatial and temporal expression of *Gbx2* in *HACNS1* homozygous, chimpanzee ortholog line, and wild type E11-E12 embryos visualized by whole-mount in situ hybridization (ISH). Representative images are shown for each genotype at three fine-scale time points; see text and associated Source Data for details on staging. Magnified views of *Gbx2* expression in limb buds are shown to the right of each embryo. Annotations of anatomical structures and developmental axes are indicated at the top right: FL forelimb, HL hindlimb, DI diencephalon, NT neural tube, PA pharyngeal arch, A anterior, P posterior. The arrows at the top far right indicate the anterior-posterior (A-P) and proximal-distal (Pr-D) axes for the magnified limb buds. Bottom right: Crown-rump lengths for all embryos assayed for *Gbx2* mRNA expression by ISH. Each point indicates a single embryo. Colors denote each fine-scale time point (T1-T6). **B** *Left*: representative images of anterior, posterior, proximal, distal (top), and strong versus weak *Gbx2* staining patterns (bottom). Anterior (A), posterior (P), and body wall (BW) domains are denoted on top left limb bud. *Right*: *Gbx2* ISH staining pattern data across 6 developmental timepoints from each of three independent, blinded scorers (marked at top as counting replicates 1-3; see text and Fig. 3A for timepoint scheme and associated Source Data for annotations). The darkest shade for *HACNS1* homozygous (dark green), chimpanzee ortholog line (olive), and wild type (teal) represents percentage of forelimbs or hindlimbs showing strong anterior and posterior limb bud staining. Medium-dark shade, as shown in the legend on the left, denotes strong anterior staining only, while the lightest shade denotes weak staining in any domain. Independent biological specimens were analyzed for  $n = 139$  (wild type),  $n = 103$  (chimpanzee ortholog line), and  $n = 106$  (human ortholog line) forelimbs and  $n = 137$  (wild type),  $n = 103$  (chimpanzee ortholog line), and  $n = 102$  (human ortholog line) hindlimbs. For body wall and pharyngeal arch scoring data see Supplementary Fig. 3A, B and associated Source Data.

hindlimb buds (Fig. 3A, B). Robust expression of *Gbx2* in *HACNS1* limb buds persisted through the remaining time points (up to T6), though the size of the anterior and posterior domains decreased over time. Strong expression of *Gbx2* in *HACNS1* homozygous embryos persisted for a longer period of time in hindlimb than in forelimb, consistent with the delayed developmental maturation of the former<sup>37</sup>. In addition, *HACNS1* homozygous embryos showed a hindlimb-specific anterior-proximal expression domain adjacent to the body wall across all 6 time points (Fig. 3A, B, Supplementary Fig. 3A).

In contrast to the robust expression observed in limb buds of *HACNS1* homozygous mice, *Gbx2* expression in limb buds from both the chimpanzee ortholog and wild type lines was weak and mostly evident at early time points (Fig. 3). Chimpanzee ortholog line embryos and wild type embryos both showed weak distal *Gbx2* expression foci in early forelimb and hindlimb that were generally restricted to the anterior limb bud (Fig. 3A, B). Weak distal expression was primarily restricted to approximately T1-T2 in wild type forelimb but persisted until approximately T4 in a subset of embryos with the chimpanzee ortholog (Fig. 3A, B). Weak distal expression persisted in hindlimb through T5-T6 in both the chimpanzee line and wild type (Fig. 3B, bottom). These findings suggest that the chimpanzee ortholog line exhibits a modest increase in *Gbx2* expression compared to wild type, potentially due to primate-rodent sequence differences affecting enhancer activity that our experimental design was intended to control for (see Supplemental Note). However, the *HACNS1* knock-in line exhibits profound changes in *Gbx2* limb bud expression compared to both. Together, these findings suggest that *HACNS1* drives spatial and quantitative changes in *Gbx2* expression in the limb, as well as a temporal extension of expression compared to wild type.

In addition to the forelimb and hindlimb bud, *Gbx2* was also expressed in the neural tube, diencephalon, and pharyngeal arches of embryos homozygous for *HACNS1* or the chimpanzee ortholog, and in wild type embryos (Fig. 3A)<sup>27,29,31</sup>. Whereas *Gbx2* expression was primarily restricted to the first pharyngeal arch in embryos with the chimpanzee ortholog and in wild type embryos, it expanded dorsally into the second pharyngeal arch in *HACNS1* homozygous embryos during T1-T5 (Supplementary Fig. 3B). However, we chose to focus on limb, due to the absence of a significant gain in H3K27ac marking at the *HACNS1* knock-in locus or the *Gbx2* promoter in *HACNS1* homozygous versus wild type pharyngeal arch (Supplementary Fig. 2).

In order to quantify the gain of *Gbx2* expression in *HACNS1* knock-in mice, we used real-time quantitative reverse transcription PCR (RT-qPCR) in pooled forelimb and hindlimb buds from embryos homozygous for *HACNS1*, the chimpanzee ortholog, and the endogenous mouse locus at time points T1-T6. We found that *Gbx2* expression was increased in forelimb and hindlimb of *HACNS1* embryos versus both chimpanzee ortholog and wild type at all 6 time points (Supplementary Fig. 3C and associated Source Data). Although we detected an increase in *Gbx2* expression in forelimb and hindlimb of embryos with the chimpanzee ortholog versus wild type at early time points, this change was substantially weaker than that between *HACNS1* and wild type or *HACNS1* and the chimpanzee line. Consistent with our ISH results, *Gbx2* expression in *HACNS1* knock-in line forelimb and hindlimb was strongest at the earliest time points and persisted longer in hindlimb than in forelimb (Fig. S3C). While *Gbx2* expression declined over time in all three genotypes, it persisted longer in *HACNS1* knock-in line forelimb and hindlimb.

### ***HACNS1* drives increased *Gbx2* expression in limb chondrogenic mesenchymal cells.**

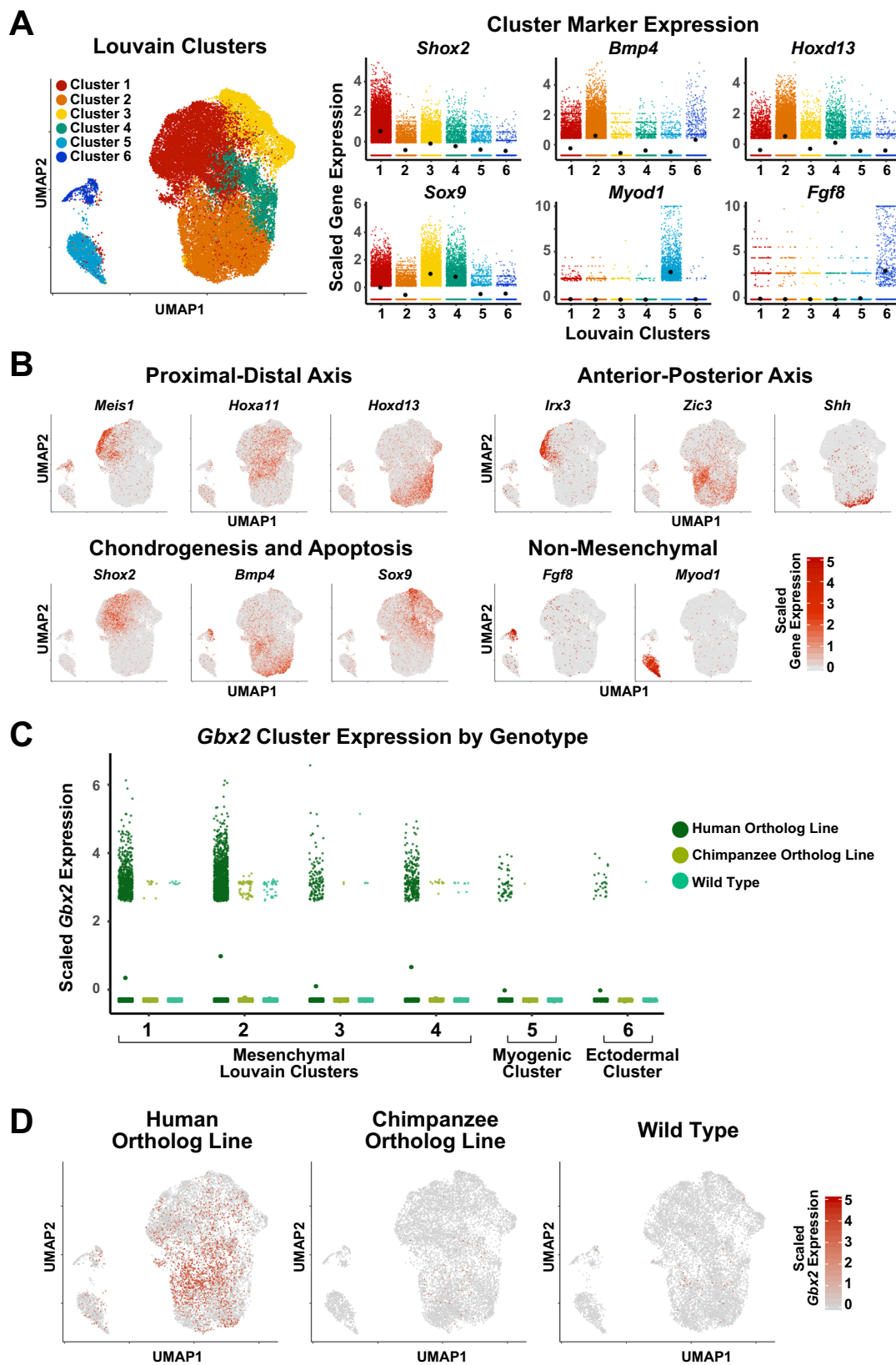
In order to identify the specific cell

types expressing *Gbx2* as well as genes that are co-expressed with *Gbx2* in the developing limb of all three genotypes, we performed single-cell RNA-sequencing (scRNA-seq) in E11.5 hindlimbs, which showed the most pronounced upregulation of *Gbx2* in spatial and quantitative expression analyses (Fig. 3, Supplementary Fig. 3). Using the 10x Genomics scRNA-seq platform for cell barcoding, library preparation, and sequencing, we obtained transcriptomes from ~10,000 cells per genotype. We used the Seurat toolkit for data preprocessing and library size normalization (Methods)<sup>38</sup>. During pre-processing, we removed endothelial and blood cells (*Cd34*-positive; *Pf4*-positive; *Hbb*-positive), as our analysis indicated that these developmentally and transcriptionally distinct cell types do not express *Gbx2* in any of our datasets<sup>39,40</sup>. After normalization of the filtered data using the SCTransform method in Seurat and integration of data from all samples into a single dataset using the Seurat v3 integration workflow (Methods), we performed clustering analysis on the integrated dataset to identify cell type categories present in all three genotypes<sup>38</sup>. To visualize similarities between cells, we used Uniform Manifold Approximation and Projection (UMAP), a dimensionality reduction method for data visualization, followed by the Louvain method for community detection to identify cell subtypes<sup>41,42</sup>.

This analysis revealed three distinct groups: (a) mesenchymal cell subtypes based on expression of the known markers *Sox9* (clusters 1, 3, 4), *Bmp4* (cluster 2), *Shox2* (cluster 1), and *Hoxd13* (clusters 1–4) (Fig. 4A); (b) non-mesenchymal cell types, including myogenic cells (cluster 5, *Myod*); and (c) ectodermal cells (cluster 6, *Fgf8*) (Fig. 4A)<sup>43–48</sup>. Furthermore, our analysis revealed finer separation of mesenchymal cells according to known limb patterning markers. We first examined the expression of known proximal-distal limb bud markers *Meis1*, *Hoxa11*, and *Hoxd13* (proximal, medial, and distal, respectively)<sup>49</sup>. Cells expressing each of these markers showed a distinct localization in the UMAP embedding (*Meis1* + cells in the top left, *Hoxa11* + cell in the center, and *Hoxd13* + cells in the lower right; Fig. 4B, upper left), suggesting our analysis recovered transcriptional and cell-type transitions along the proximal-distal patterning axis (Fig. 4B).

We also found that the first axis of the UMAP embedding clearly recapitulated known gene expression gradients along the anterior-posterior limb bud axis based on expression of the anterior-proximal marker *Irx3*, the anterior marker *Zic3*, and the posterior-proximal marker *Shh* (Fig. 4B, upper right)<sup>43,50–52</sup>. Using markers of chondrogenic (*Sox9*, *Shox2*) versus non-chondrogenic (*Bmp4*) mesenchyme, we found that the second UMAP axis followed the chondrogenic versus interdigital apoptotic fate gradient (Fig. 4B)<sup>43,45,48</sup>. We also found that the expression patterns of these markers were broadly conserved between genotypes, with each genotype showing comparable subsets of proximal, distal, anterior, posterior, chondrogenic, and non-chondrogenic cell types (Supplementary Figs. 4A, B). Collectively, our scRNA-seq analyses identified specific conserved cell types and spatial transcriptional gradients in the developing hindlimb bud across all three genotypes.

We then sought to define genotype-specific differences in *Gbx2* expression. In order to identify the cell types expressing *Gbx2* in *HACNS1* homozygous hindlimb buds, we examined the distribution of *Gbx2*-positive cells across cell clusters in all three genotypes. We found that *Gbx2* was upregulated in *HACNS1* homozygous hindlimbs versus chimpanzee ortholog and wild type hindlimbs, primarily in the mesenchymal cell clusters (clusters 1–4), consistent with the ISH and RT-qPCR expression analyses (Fig. 3, Supplementary Fig. 3C and Fig. 4C). In *HACNS1* homozygous hindlimbs, 24% of cells expressed *Gbx2*, 96% of which were mesenchymal cells, whereas less than 1% of cells in



chimpanzee ortholog and wild type hindlimbs expressed *Gbx2* (Methods; Supplementary Data 4). Greater than 98% of *Gbx2*-positive cells in the chimpanzee ortholog line and wild type hindlimb were mesenchymal, and the majority of *Gbx2*-positive cells in each of these lines were assigned to Louvain cluster 2 (70% and 68%, respectively; Supplementary Data 4). Only one

non-mesenchymal cell from chimpanzee ortholog hindlimb and one non-mesenchymal cell from wild type hindlimb was *Gbx2*-positive (Fig. 4C; Supplementary Data 4). UMAP embedding of cells revealed that *Gbx2*-positive cells in *HACNS1* homozygous hindlimb buds largely clustered within a distinct subset of mesenchymal cells belonging primarily to Louvain clusters 1, 2

**Fig. 4 Single-cell transcriptome analysis of E11.5 hindlimb bud in *HACNS1* knock-in, chimpanzee ortholog line, and wild type embryos.** **A** Left: UMAP embedding of *HACNS1* homozygous, chimpanzee ortholog line, and wild type cells. The colors indicate cell clusters identified by Louvain clustering. Right: Expression of known limb bud cell-type marker genes in each cluster. Black dots denote cluster mean expression. **B** UMAP embedding of hindlimb bud cells from *HACNS1* homozygous, chimpanzee ortholog line, and wild type, showing expression of proximal-distal, anterior-posterior, chondrogenesis-apoptosis, and non-mesenchymal markers. See text and Supplementary Fig. 4 for details. **C** Expression of *Gbx2* in each Louvain cluster, separated by genotype. Dots denote cluster mean expression. **D** UMAP embeddings illustrating cells expressing *Gbx2* (indicated in red) in *HACNS1* homozygous, chimpanzee ortholog line, and wild type cells. All gene expression data shown in plots and UMAP embeddings (**A–D**) were imputed using ALRA and centered and scaled using z-scores (see “Methods”)<sup>81</sup>.

and 4 (Fig. 4C, D; Supplementary Data 4). The genotype-specific differences in *Gbx2* expression were consistent between the imputed and unimputed data as well as across individual replicates (Supplementary Fig. 4C, D).

To identify genes whose expression is associated with *Gbx2*, we used k-Nearest-Neighbors Conditional-Density Resampled Estimate of Mutual Information (kNN-DREMI), which computes scores quantifying the strength of the relationship between two genes<sup>53,54</sup>. Using kNN-DREMI scores, we ranked each gene expressed in *HACNS1* homozygous limbs by the strength of its association with *Gbx2*. To determine if genes associated with *Gbx2* were enriched in particular functions, we then performed Gene Set Enrichment Analysis (GSEA) on this set of ranked genes. We found that *Gbx2* expression was associated with genes in several limb development-related ontologies, including “Cartilage Morphogenesis” (Kolmogorov–Smirnov (KS)  $P = 1.61 \times 10^{-3}$ ) and “Regulation of Chondrocyte Differentiation” (KS  $P = 5.84 \times 10^{-3}$ ); the latter overlapped considerably with “Collagen Fibril Organization” (KS  $P = 1.30 \times 10^{-4}$ ) (Fig. 5A, Supplementary Data 5). These results indicate that in the *HACNS1* homozygous hindlimb, *Gbx2* is co-regulated with genes expressed in condensing mesenchymal cells destined to become chondrocytes (Fig. 5A).

We also used Manifold Enhancement of Latent Dimensions (MELD) to quantify the differences in the transcriptional profiles of *HACNS1* homozygous limb buds compared to the chimpanzee ortholog line and wild type. MELD is an unsupervised learning algorithm and is therefore an orthogonal approach that is naïve to our identification of *Gbx2* as the target of *HACNS1*<sup>55</sup>. MELD uses graph signal processing to quantify the relative likelihood of observing each cell in each of multiple experimental conditions based on its transcriptional profile. In this case, MELD is used to quantify the relative likelihood (RL) of observing a cell in the *HACNS1* hindlimb versus the chimpanzee ortholog or wild type hindlimb. Rather than explicitly classifying genes as differentially expressed in one condition versus another, the RL value can be used to identify trends in gene expression across cells that are associated with the *HACNS1* knock-in condition<sup>55</sup>.

To identify overall gene expression patterns characteristic of *HACNS1* homozygous hindlimb bud cells, we used kNN-DREMI to associate gene expression with the *HACNS1* knock-in condition (*HACNS1* RL) calculated by MELD. We then used the resulting gene rankings to identify enriched biological functions via GSEA, as described above for *Gbx2* expression. We found that genes associated with both *HACNS1* RL and *Gbx2* expression converged on related biological processes. Performing GSEA using genes ranked by mutual information with *HACNS1* RL revealed significant enrichment of the “Chondrocyte Differentiation” ontology (KS  $P = 7.00 \times 10^{-5}$ ), along with four other categories also significantly enriched in the *Gbx2* expression analysis: “Hindlimb Morphogenesis” (KS  $P = 1.42 \times 10^{-3}$ ), “Cartilage Development Involved in Endochondral Bone Morphogenesis” (KS  $P = 1.11 \times 10^{-3}$ ), “Collagen Fibril Organization” (KS  $P = 1.40 \times 10^{-5}$ ), and “Positive Regulation of Phosphatidylinositol 3-kinase Signaling” (KS  $P = 4.40 \times 10^{-5}$ ), of which the last two are also implicated in chondrocyte differentiation (Fig. 5A, Supplementary Data 6)<sup>56,57</sup>. “Collagen Fibril Organization” is the

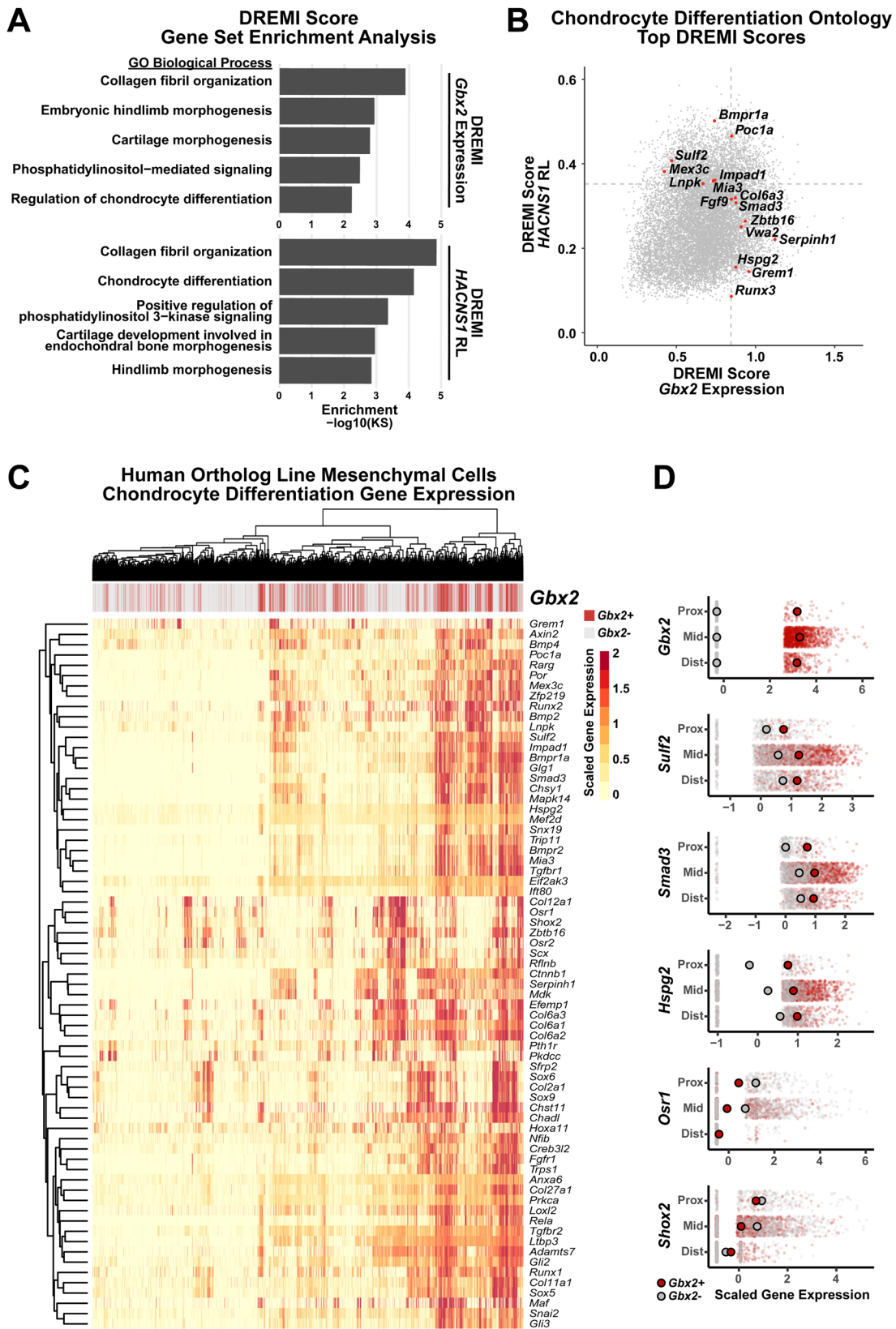
most significantly enriched GO term for genes associated with *HACNS1* RL and is the second most enriched for genes associated with *Gbx2* expression. The top GO term for genes associated with *Gbx2* expression, “Roof of Mouth Development”, (KS  $P = 1.10 \times 10^{-5}$ ), shares >20% of its genes with “Embryonic Hindlimb Morphogenesis” (KS  $P = 1.18 \times 10^{-3}$ )<sup>58</sup>. This illustrates that many genes involved in limb development are also implicated in craniofacial development, and likely accounts for why craniofacial development-related GO terms were also enriched in our analysis.

These findings led us to examine the expression patterns of chondrocyte differentiation-related genes in *HACNS1* homozygous mesenchymal cells belonging to Louvain clusters 1–4 (Fig. 5C). We clustered *HACNS1* homozygous mesenchymal cells by humanized RL and *Gbx2* expression and examined the expression of the “Chondrocyte Differentiation” ontology genes within *Gbx2*-positive cells (Fig. 5C). This clustering analysis revealed higher expression of positive regulators of chondrocyte differentiation (e.g., *Sox9*, *Col2a1*, *Bmp2*, and *Runx2*) specifically in *Gbx2*-positive versus *Gbx2*-negative humanized mesenchymal cells, supporting that *HACNS1*-driven upregulation of *Gbx2* occurs in chondrogenic cells (Fig. 5C)<sup>57,59–61</sup>. We also identified a subset of *Gbx2*-positive humanized cells that were also positive for *Bmp4*, which is expressed in the apoptotic interdigital domains<sup>62</sup>. These findings suggest that upregulation of *Gbx2* is associated with the interdependent pathways of digit condensation and interdigital cell fate specification required for digit morphogenesis.

The gain of a *Gbx2* anterior proximal hindlimb expression domain in *HACNS1* knock-in mice suggests that *Gbx2* upregulation may impact development of multiple segments in the limb (Fig. 3A, B). To explore this further, we characterized the distribution of *Gbx2* expression along the proximal-distal (PD) axis in the *HACNS1* homozygous hindlimb single-cell dataset. We defined three groups of mesenchymal cells along the (PD) axis based on expression of the patterning markers shown in Fig. 4: distal (*Hoxd13*+, *Hoxa11*–, *Meis1*–); proximal (*Hoxd13*–, *Hoxa11*–, *Meis1*–); and intermediate (all remaining *Hoxa11*+) cells. *Gbx2* was expressed in all three subtypes (Fig. 5D, top). We also examined expression of several “Chondrocyte Differentiation” ontology genes (selected from Fig. 5C) in *Gbx2*+ and *Gbx2*– cells in each subtype. Cells expressing *Gbx2* also expressed chondrocyte differentiation markers along the PD axis, with a subset of markers showing modestly increased expression in *Gbx2*+ cells (Fig. 5D, middle and bottom). Together, these results link *Gbx2* upregulation to chondrocyte differentiation in multiple developing regions of the hindlimb.

**Morphometric analysis of *HACNS1* knock-in mouse limbs.** To determine if *Gbx2* upregulation and downstream transcriptional changes in *HACNS1* limb buds affect digit formation or overall limb morphology, we performed morphometric analysis of skeletal preparations for embryos homozygous for *HACNS1*, the chimpanzee ortholog, and wild type embryos (Methods). We performed morphometric analysis at E18.5 in order to capture any major phenotypic effects of targeted humanization that





occurred by the end of embryonic skeletogenesis<sup>22</sup>. We did not detect gross morphological differences among genotypes; the three major limb segments (autopod, zeugopod, and stylopod) were present in both *HACNS1* skeletons and chimpanzee ortholog skeletons (Supplementary Fig. 6). We also examined digit length (normalized to body size based on the length of the ossified

humerus), and intradigital (phalange to metacarpal or metatarsal length) and interdigital ratios. Again, we found no significant differences in digit length or autopod proportions between genotypes (Supplementary Fig. 6, Supplementary Data 7–9; Source Data). Although these analyses failed to detect morphological differences, it is possible that subtle phenotypes do indeed exist in

**Fig. 5 Gbx2-positive mesenchymal cell expression of chondrocyte differentiation markers in *HACNS1* homozygous limb bud.** **A** Ontology enrichments of genes with expression associated with *Gbx2* expression (top) and the relative likelihood of the *HACNS1* knock-in condition (*HACNS1* RL, bottom) in *HACNS1* homozygous mesenchymal cells. The log-transformed Gene Set Enrichment Analysis Kolmogorov-Smirnov *P* value for each category is plotted on the x-axis. Ontologies shown are those overlapping in the *Gbx2* expression and *HACNS1* RL ontology enrichment lists. See also Supplementary Data 5 and 6. **B** *HACNS1* RL and *Gbx2* kNN-DREMI scores are plotted for all genes. Genes ranked in the top 20% of kNN-DREMI scores in the Chondrocyte Differentiation ontology (GO:0002062) for the union of the *HACNS1* RL and *Gbx2* kNN-DREMI analysis gene lists are colored in red and labeled. Dotted lines indicate the top 20% of values for each dataset. **C** Heatmap showing expression of genes belonging to the ontology “Chondrocyte Differentiation” (GO:0002062) in all *HACNS1* homozygous mesenchymal cells (Louvain clusters 1–4). Hierarchical clustering was used to determine the order of cells (in columns) and genes (in rows). The bar at the top of the heatmap shows *Gbx2*-positive and *Gbx2*-negative cells in red and gray, respectively. **D** Expression of selected genes in *Gbx2*-positive (red) versus *Gbx2*-negative (gray) mesenchymal cells belonging to Louvain clusters 1 and 2, partitioned by proximal-distal axis markers as follows: Proximal cells (Prox) are *Meis1*<sup>+</sup>, *Hoxd13*<sup>–</sup>, *Hoxa11*<sup>–</sup>; distal cells (Dist) are *Hoxd13*<sup>+</sup>, *Hoxa11*<sup>–</sup> and *Meis1*<sup>–</sup>; and intermediate cells (Mid) are all remaining *Hoxa11*<sup>+</sup> cells. Cells were randomly down-sampled to enable comparison of equal numbers of *Gbx2*-positive and *Gbx2*-negative cells. Larger red and gray dots respectively denote mean expression of each indicated gene in each group in *Gbx2*-positive and *Gbx2*-negative cells. All gene expression values shown in C and D were imputed using ALRA and centered and scaled using z-scores (see “Methods”)<sup>81</sup>.

the *HACNS1* homozygous limb, which may be revealed in future studies.

## Discussion

Understanding how uniquely human genetic changes altered developmental processes is essential to understanding the evolution of our species. Here we investigated the role of the Human Accelerated Region *HACNS1* in human limb evolution by directly interrogating its biological functions in a genetically humanized mouse model. This *in vivo* approach enabled us to identify spatial and temporal changes in gene expression driven by *HACNS1* and to characterize the specific cell types affected by these changes, providing insight into the developmental processes modified due to human-specific alterations in enhancer activity.

First, we determined that *HACNS1* is active in the mouse genomic context, recapitulating the significant level of H3K27ac marking previously observed in the developing human limb in the *trans*-regulatory environment of the developing mouse limb. Second, we found that *Gbx2* exhibited increased promoter activity in the *HACNS1* homozygous limb bud, strongly supporting that it is regulated by *HACNS1*, and then demonstrated that *HACNS1* drives robust changes in *Gbx2* expression in the forelimb and hindlimb bud. These findings support the long-standing hypothesis that discrete regulatory changes altering expression of pleiotropic developmental regulators in specific tissues contribute to the evolution of phenotypic differences—in the case of this study, molecular phenotypes—across species<sup>20,21</sup>. Third, by performing scRNA-seq, we identified the spectrum of distinct cell types that show upregulation of *Gbx2* in the developing limb, as defined by transcriptome signatures. This analysis not only established that *Gbx2* is expressed in mesenchymal cells in the limb bud, but also placed these cells in the developmental process of chondrogenesis. By both characterizing *Gbx2*-positive cells in the *HACNS1* homozygous limb bud and identifying overall expression trends associated with the *HACNS1* knock-in condition without reference to *Gbx2* expression, we implicated changes in *Gbx2* regulation in chondrocyte differentiation, a critical process in skeletal development.

We found that the human-specific gain of function in *HACNS1* drives quantitative, spatial, and temporal changes in *Gbx2* expression in the *HACNS1* knock-in mouse limb bud. One hypothesis consistent with these findings is that *HACNS1* acts to both spatially expand and prolong *Gbx2* expression and its potential effects on chondrocyte differentiation in the developing limb. *Gbx2* is expressed in wild type mouse limb at E10.5, and the mouse ortholog of *HACNS1* is marked by H3K27ac at this time point, suggesting it is contributing to *Gbx2* regulation<sup>27,63,64,65</sup>. However, the mouse ortholog is annotated as a poised enhancer in the E11 limb bud, consistent with the weak expression of *Gbx2*

we observe in E11–E12 wild type limb buds<sup>65</sup>. In contrast, *HACNS1* shows robust H3K27ac marking in the E11.5 limb, and *Gbx2* is strongly expressed in the E11 *HACNS1* limb bud. These results, coupled with our finding that *Gbx2* is primarily expressed in limb mesenchymal cells in all three genetic backgrounds we interrogated, suggests that the gain of function in *HACNS1* may be modifying an ancestral *Gbx2* regulatory program in the limb. The temporal extension of *Gbx2* expression we observed in *HACNS1* knock-in limb buds also raises the possibility that *HACNS1* contributes to human-specific heterochronic changes in this ancestral regulatory program. This is particularly intriguing given the protracted timing of human development compared to chimpanzee and mouse. Further insight into these questions will require deciphering the *Gbx2* regulatory network in both *HACNS1* knock-in and wild type limbs, including identifying downstream targets of GBX2 and upstream regulators of *HACNS1* and the orthologous mouse enhancer. We note that several of the transcription factors whose predicted binding sites are present only in *HACNS1* have been implicated in limb bud patterning and chondrogenesis at E11, including *Ets1* and *Gabpa*<sup>66</sup>, *Tfap2B*<sup>67</sup>, and *Runx2* (Supplementary Data 2)<sup>68</sup>. Additional longitudinal single-cell expression studies in the *HACNS1* knock-in limb may also reveal downstream transcriptional changes that could not be identified in the single time point we examine here.

The modest, but observable, differences in *Gbx2* expression in the chimpanzee ortholog line compared to wild type may be due to primate-rodent sequence changes that altered ancestral enhancer activity at the *HACNS1* locus, predating the human-specific changes we focused on in this study. Although *HACNS1* itself shows very high levels of sequence conservation with the chimpanzee and mouse orthologs, the flanking human and chimpanzee sequences we included in our knock-in models harbor multiple differences, including single nucleotide substitutions and gaps, relative to the endogenous mouse locus (Fig. 1B; Supplemental Note). These sequence changes may have contributed either to gain of function at the *HACNS1* locus during primate evolution, or to loss of function on the rodent lineage. Distinguishing between these mechanisms will require *in vivo* genetic studies of *HACNS1* orthologs from multiple primate, rodent, and outgroup species, and of primate and rodent ancestral orthologs inferred by ancestral sequence reconstruction.

Our morphological studies in the *HACNS1* mouse model also have several limitations that we note here. We did not identify major changes in skeletal morphology at E18.5 that were associated with targeted humanization, despite the large number of biological replicates we used in our analyses. However, this does not preclude subtle changes in limb length or other features in adult mice, or changes in soft tissues that would not be detected

using skeletal preparations. Furthermore, we did not exhaustively characterize other tissues, including the pharyngeal arch and diencephalon, in which *Gbx2* has known developmental functions<sup>26–29,31</sup>. Although we did not observe overt craniofacial phenotypes in *HACNS1* knock-in mice, the dorsal expansion of *Gbx2* expression into the second pharyngeal arch that we detected in this background may result in more subtle developmental effects. The second arch contributes to the development of multiple structures, including the temporal styloid process, the hyoid bone, facial muscles, and the facial nerve<sup>69,70</sup>. Understanding which of these structures might be altered in *HACNS1* humanized mice could be explored in future studies.

We have shown that HAR knock-in mouse models represent a viable and fruitful approach for studying gene regulatory mechanisms relevant for human evolution within the complete genomic, tissue-level, and developmental framework of a living organism. We note that the lack of an overt morphological phenotype in the *HACNS1* model, despite the strong effects on gene expression that we observed, suggests it may be difficult to anticipate the full range of phenotypes HARs might generate in model systems based on their degree of acceleration. Nevertheless, it is not entirely surprising that the *HACNS1*-driven molecular phenotypes we observed did not produce morphological phenotypes. Genetic changes in any one enhancer are unlikely to be sufficient to replicate human-specific morphological changes entirely in an experimental model. The evolution of uniquely human physical traits likely entailed modifications in the expression of many genes, potentially driven by multiple HARs and other human-specific genetic changes. Our study provides insight into how a single HAR alters gene regulation and expression at critical developmental time points, yielding an important entry point for understanding the larger developmental networks that changed during human limb evolution, of which *Gbx2* is a part. Knock-in mouse models offer the means to study additional HARs contributing to human-specific phenotypes at once, either through intercrossing mouse lines harboring edited unlinked loci or by iterative editing of one locus, allowing us to expand our understanding of human limb evolution. Our study thus establishes a framework for using genetically humanized mouse models to link sequence changes that arose on the human lineage to the unique traits that distinguish our species.

## Methods

**Mouse line generation and validation.** All animal work was performed in accordance with approved Yale IACUC protocols (#2019–11167 and #2020–07271). The *HACNS1* and chimpanzee ortholog lines were generated at the Yale Genome Editing Center using standard gene targeting techniques in mouse ES cells<sup>71</sup>. C57BL/6J-*A<sup>w</sup>-/-*J mouse ES cells, generated by the Yale Genome Editing Center from C57BL/6J-*A<sup>w</sup>-/-*J mice obtained from The Jackson Laboratory (RRID:IMSR\_JAX:000051), were edited by electroporation of a GFP cloning vector containing human (1241 bp) or chimpanzee (1240 bp) sequence flanked by C57BL/6J mouse sequence homology arms, floxed pPGKneo vector, and diphtheria toxin sequence (Supplementary Fig. 1A)<sup>72</sup>. The genomic coordinates of the human (hg19; [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.13/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/)), chimpanzee (panTro4; [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001515.6/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001515.6/)), and mouse (mm9; [https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001635.18/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.18/)) sequences used in the editing constructs, including the mouse homology arm sequences, are listed in Supplementary Data 1<sup>73</sup>. Positive clones were karyotyped and only clones exhibiting a normal mouse karyotype were used for blastocyst injection. Resulting G0 chimeras were backcrossed to wild type C57BL/6J (RRID:IMSR\_JAX:000664) and crossed with an actin-Cre C57BL/6J mouse line to remove the neo cassette. All mice used in our analysis were from F9 or later generations. Mice were maintained in a Yale Animal Resources Center (YARC) managed facility under a standard 12 h light/dark cycle and environmental monitoring according to YARC policies and procedures.

Genotyping primers specific to *HACNS1*, chimpanzee, and mouse orthologs are listed in Supplementary Data 10. Cloning primers listed in Supplementary Data 10 were used to amplify edited loci for cloning and Sanger sequencing for comparison to the hg19 or panTro4 sequence. Sanger sequencing data is available at [http://noonan.ycg.yale.edu/noonan\\_public/Dutrow\\_HACNS1/](http://noonan.ycg.yale.edu/noonan_public/Dutrow_HACNS1/). The sequence identity between the human (hg19, chr2:236773456–236774696) and chimpanzee alleles

(panTro4, chr2B:241105291–241106530) is 98.2% (22 substitutions total, of which 15 are fixed in humans). Human-specific substitutions were defined as fixed if the derived allele frequency in dbSNP (v153) was  $\geq 0.9999$  and if the ancestral sequence state was conserved between chimpanzee, rhesus macaque, orangutan, and marmoset. We provide a detailed analysis of sequence differences between the human, chimpanzee and mouse orthologs in the Supplementary Note (Supplementary Materials). *HACNS1*-*GBX2* locus TAD coordinates (hg19 chr2:236655261–237135261) are from H1 human ES cell Hi-C data; *HACNS1* and *GBX2* are present in the same TAD and *GBX2* is the only annotated protein-coding gene whose promoter is included in this TAD<sup>32</sup>.

Copy number verification qPCR was performed using genomic DNA from three F9 mice from each line using Power SYBR Green Mastermix (Thermo Fisher Scientific #4368577) and the StepOnePlus Real-Time PCR System (Applied Biosystems) with primers listed in Supplementary Data 10. All biological replicates of each genotype were run in triplicate and Ct values of each were normalized to a control region on a different chromosome (see Supplementary Data 10). Primary qPCR results are available as Source Data.

**Chromatin Immunoprecipitation, ChIP-qPCR and ChIP-seq.** Tissue for chromatin preparation was collected from E11.5 forelimb and hindlimb bud pairs or pharyngeal arch tissue from *HACNS1* and chimpanzee ortholog line heterozygous crosses to obtain pooled, litter matched limb bud or pharyngeal arch samples for all three genotypes (*HACNS1* homozygous, chimpanzee ortholog line, and wild type). Two biological replicates were used per genotype per tissue, each with tissue pooled from three embryos. Pooled tissue was crosslinked and sonicated as previously described<sup>74</sup>. Chromatin for each genotype, tissue, and replicate was used for H3K27ac or H3K4me2 immunoprecipitation with 7.5  $\mu$ g antibody and  $\sim$ 5  $\mu$ g tissue per ChIP assay using Active Motif #39133 (RRID: AB\_2561016) and Active Motif #39913 (RRID: AB\_2614976) as previously described<sup>74,75</sup>. ChIP-qPCR was performed using Power SYBR Green Mastermix (Thermo Fisher Scientific #4368577) with primers listed in Supplementary Data 11. Samples were sequenced (2  $\times$  100 bp) using standard Illumina protocols on an Illumina HiSeq 4000 (RRID: SCR\_016386). To control for batch effects, all samples of the same tissue type were multiplexed and sequenced on a single lane.

Reference genomes edited to replace the mouse ortholog of *HACNS1* with the human or chimpanzee sequence were built using Bowtie (v2.2.8; RRID: SCR\_005476)<sup>76</sup>. ChIP-seq raw reads were aligned to the mm9, mm9 with chimpanzee ortholog, or humanized mm9 reference genome using Bowtie with -sensitive and -no-unal settings. GC composition was assessed using fastQC and showed that GC content and bias were consistent across all experiments<sup>77,78</sup>. Tag directories for each experiment were generated using makeTagDirectory in HOMER with default settings and standard normalization to 10 million tags, and were used to generate bigwig files for visualization with makeUCSCfile<sup>23</sup>. All peaks were called with HOMER (v4.9.1 RRID: SCR\_010881) using default settings for -histone (IP vs input fold change = 4,  $p = 0.0001$ , peak size = 500, minDist = 1000)<sup>23</sup>. All differential peaks were called with DESeq2 implemented in HOMER using getDifferentialPeaksReplicates.pl with default settings (fold change cutoff = 2, FDR cutoff = 5%); briefly, reads from each comparison are pooled, with ChIP and inputs pooled separately, such that new peaks are called and used for quantitative comparison between genotypes<sup>23,24</sup>. The complete datasets of all peaks tested in differential analyses can be found at [http://noonan.ycg.yale.edu/noonan\\_public/Dutrow\\_HACNS1/](http://noonan.ycg.yale.edu/noonan_public/Dutrow_HACNS1/).

**RNA extraction and RT-qPCR.** E11–E12 embryos were collected from six *HACNS1* homozygous, chimpanzee ortholog line, or wild type litters generated by crossing homozygous animals for each line. All embryos within each genotype group were ordered based on stage (>70 total embryos) and were divided into six timepoint groups per genotype consisting of forelimb or hindlimb buds from 4–6 pooled embryos per time point per genotype per tissue. RNA was purified using the Qiagen miRNeasy Kit (#74106). Invitrogen Superscript III Reverse Transcription Kit (#18080-051) was used to prepare cDNA from each sample. qPCR with the resulting cDNA was performed using Power SYBR Green Mastermix (Thermo Fisher Scientific #4368577). All samples were analyzed in triplicate using primers listed in Supplementary Data 12 and Ct values of *Gbx2* were normalized to *Hprt1*. Primary RT-qPCR results are available as Source Data.

**Whole mount in situ hybridization.** E11–E12 mouse embryos were collected from *HACNS1* homozygous ( $n = 7$  litters), chimpanzee ortholog line ( $n = 8$  litters), and wild type ( $n = 12$  litters) homozygous crosses. Embryos were fixed and hybridized with the same preparation of antisense *Gbx2* mRNA probe under identical conditions as previously described<sup>78,79</sup>. The *Gbx2* probe used for hybridization contains the full mouse consensus CDS sequence (CCDS15150.1); NCBI CCDS Release 23; <https://www.ncbi.nlm.nih.gov/projects/CCDS/CcDSBrowse.cgi?REQUEST=ALLFIELDS&DATA=CCDS15150.1&ORGANISM=10090&BUILD=CURRENTTBUILDS> (NCBI CCDS Release 23 CCDS15150.1). The 178 embryos (55 from the *HACNS1* knock-in line, 52 from the chimpanzee ortholog line, and 71 from wild type) were divided into temporally-ordered sextiles within the E11–E12 window ( $\sim$ 40–48 somites, although we did not rely on somite counts for staging) based on measurement of crown-rump length for each individual embryo<sup>35</sup>. For the data shown in Fig. 3B, embryos were assessed for staining pattern by three individuals blinded to genotype under a stereo microscope (Leica S6D). For the data shown

in Supplementary Fig. 3A, B embryos were annotated by a single scorer blinded to genotype. The scoring scheme was based on previous studies, notably to assess whole-mount gene expression patterns as described in the VISTA Enhancer Browser (<http://enhancer.lbl.gov/>)<sup>10,34,36</sup>. Embryos were assigned to one of eleven categories of *Gbx2* expression pattern based on the anterior-posterior and proximal-distal localization of staining as well the intensity (strong versus weak) of staining: 1: anterior and posterior (AP); 2: anterior distal and posterior distal (APD); 3: distal (D); 4: anterior distal (AD); 5: anterior (A); 6: weak anterior and posterior (APL); 7: weak anterior (AL); 8: weak distal (DL); 9: weak anterior and posterior distal (APDL); 10: weak anterior distal (ADL); 11: no staining (N). Categories were merged for clarity in Fig. 3B in the following manner: categories 1–3: anterior and posterior; categories 4–5: anterior only; categories 6–10: weak staining. See Fig. 3B for representative images of staining patterns illustrating the scoring scheme used for qualitative assessment of expression. Representative images were taken using a Zeiss Stemi 2000-C stereomicroscope fitted with an AxioCam MRC5 digital camera and Zeiss AxioVision software. Images and associated annotations are available as Source Data.

### Single-cell RNA-sequencing

**Sample preparation.** Tissue for scRNA-seq was collected at E11.5 from two human ortholog line homozygous litters, two chimpanzee ortholog line homozygous litters, and two wild type litters. Embryos were staged as previously described in order to obtain samples from stage-matched T3 embryos from each genotype. Left hindlimb buds from three embryos per genotype per replicate were pooled. Following dissection, the tissue was immediately placed in CMFSG saline-glucose solution (1x Calcium-magnesium-free phosphate buffered saline from Thermo Fisher Scientific #21-040-CV with 0.1% glucose from Corning 45% Glucose #45001-116) on ice. Gibco TrypLE Express digestion solution was used for cellular dissociation (Thermo Fisher Scientific # 2605010). The dissociation reaction was stopped using 1xDMEM (ATCC 30-2002) with 10% heat-inactivated Fetal Bovine Serum (Sigma-Aldrich #F4135). The dissociated cells were filtered through a 40 µm strainer and harvested by centrifugation at 4 °C. Cells were washed and resuspended in 1x Calcium-magnesium-free phosphate buffered saline (Thermo Fisher Scientific #21-040-CV) with 0.04% BSA (Sigma-Aldrich #SRE0036). Cell number and viability were estimated on a Countess II Automated Cell Counter prior to library preparation of 10,000 cells (estimated cell recovery from 16,000 input cells) per sample using Chromium Single Cell 3' GEM, Library & Gel Bead Kit v3 (10X Genomics PN-1000075). Libraries were sequenced (2 × 75 bp) on an Illumina HiSeq 4000 (RRID: SCR\_016386). To control for batch effects, all samples were multiplexed across all lanes. Count matrices were produced from raw sequencing data using the Cell Ranger v3.0.2 package from 10X Genomics (RRID: SCR\_017344).

**Data filtering and preprocessing.** Matrices from the 10x Cell Ranger platform were filtered and preprocessed using Seurat v3.0.1 (RRID: SCR\_016341)<sup>38</sup>. Prior to the generation of Seurat objects, *Xist* gene counts were eliminated in order to avoid clustering by sex within mixed sample populations. Genes expressed in fewer than 5 cells per sample were removed. Cells with greater than 7.5% or 2% counts from mitochondrial genes or hemoglobin genes, respectively, were removed. Cells with total gene count (nGene) z-scores less than -1 (corresponding to ~700 or fewer detected genes) or greater than 4 (corresponding to ~6000 or greater detected genes) were removed, as were cells with total UMI count (nUMI) z-scores greater than 7 (corresponding to ~50,000 or greater detected UMIs; see Supplementary Fig. 5). One chimpanzee ortholog line replicate was removed during pre-processing due to high overall mitochondrial gene expression, indicative of low viability. Prior to data integration, expression values from each sample were normalized based on library size for pre-processing purposes only using the Seurat tool `NormalizeData`<sup>38</sup>. Louvain clustering as implemented in Seurat was performed for pre-processing purposes only using `FindVariableFeatures`, `ScaleData`, `RunPCA`, `FindNeighbors`, and `FindClusters` in order to remove endothelial cell clusters (*Cd34*-positive and *Pf4*-positive), clusters characterized by aberrant mitochondrial gene expression (low *mt-Co1*), and transcriptionally distinct clusters containing fewer than 30 cells per sample<sup>38,42</sup>. The numbers of cells remaining after pre-processing for each sample are listed in Supplementary Data 13.

**Data normalization and integration.** All subsequent normalization and integration steps after pre-processing were performed with raw counts for all cells retained after pre-processing (see Supplementary Data 13). Cell cycle scores were computed using `CellCycleScoring` in Seurat to regress out the difference between G2M and S phases, effectively preserving differences between cycling and non-cycling cells while reducing differences related to cell cycle amongst proliferating cells<sup>38</sup>. In addition to cell cycle scores, percent mitochondrial gene expression and nUMI values were regressed using `SCTransform` (SCT) in order to reduce the effects of sequencing depth and minor differences in mitochondrial DNA expression related to viability<sup>38,80</sup>. All SCT normalized datasets containing all genes from individual samples were integrated using `SelectIntegrationFeatures`, `PrepSCTIntegration`, `FindIntegrationAnchors`, and `IntegrateData`<sup>38,80</sup>.

Following integration, the combined dataset was randomly down-sampled to contain a maximum of 10,000 cells per genotype prior to embedding and clustering using `SubsetData` in Seurat<sup>38</sup>. PCA, UMAP, and Louvain clustering were

implemented in Seurat using `RunPCA`, `RunUMAP`, `FindNeighbors`, and `FindClusters`<sup>38,41</sup>. Percentages of cells belonging to each Louvain cluster are shown in Supplementary Data 13. Normalized data from all samples combined were used for imputation using ALRA with default settings for the purposes of data visualization as shown in Fig. 4A–D, Supplementary Fig. 4A, B, and Fig. 5C, D<sup>81</sup>. Marker gene expression was compared between ALRA-imputed and unimputed data to establish that imputation did not substantially distort marker gene expression patterns in our dataset (Supplementary Fig. 4, Supplementary Data 13). Data normalization and integration, UMAP embedding, and Louvain clustering were performed prior to imputation. The threshold for identifying *Gbx2*-positive cells was set as an imputed *Gbx2* expression value greater than 0.1. This threshold was also used for identifying percentages of marker gene-positive cells in unimputed and imputed data as shown in Supplementary Data 13. All gene expression scaling and centering for visualization purposes was performed on normalized imputed or unimputed data using the Seurat `ScaleData` function with default parameters (`scale.max = 10`)<sup>38</sup>.

**MELD, MAGIC, kNN-DREMI analyses.** Cells belonging to mesenchymal cell clusters (clusters 1–4, see Fig. 4A, C) from all genotypes were used for MELD, MAGIC, kNN-DREMI, and Gene Set Enrichment Analysis (GSEA). Scaled data matrices from the Seurat object integrated assay were loaded using `scprep` for MELD, MAGIC, and kNN-DREMI (<https://github.com/krishnaswamylab/scprep>). MELD and MAGIC both denoise scRNA-seq data using graphs to model cellular state space. The same graph signal was used for both MELD and MAGIC as calculated by `graphools` (1.5.2) with `n_pca = 20`, `decay = 40`, and `knn = 10`. MELD was run on one-hot vectors for each genotype independently using default parameters<sup>55</sup>. MAGIC was performed using the same graph signal as MELD<sup>54</sup>. We used the kNN-DREMI implementation provided in `scprep` and kNN-DREMI was run on MAGIC-imputed data<sup>53</sup>. kNN-DREMI analysis was used in order to identify genes with expression levels associated with either *Gbx2* expression in humanized hindlimb or cells with increased humanized RL as calculated using MELD. MAGIC was employed only for the purpose of generating denoised gene expression values for kNN-DREMI analysis of gene-gene relationships but was not used for data visualization, clustering, or sample-associated density estimation using MELD.

**Gene set enrichment analysis.** GSEA was performed using `topGO` v.2.34.0 (RRID: SCR\_014798) on all expressed genes that were ranked by *Gbx2*-DREMI or humanized RL-DREMI score from the aforementioned humanized mesenchymal cell kNN-DREMI analysis<sup>82</sup>. Significant nodes were identified using a Kolmogorov-Smirnov test and the algorithm = “elim” argument. Ontologies listed in Supplementary Data 5 and 6 are the top 30 nodes with fewer than 100 annotated genes (to remove non-specific categories) and at least one gene in the top 20% of DREMI scores. Heatmap hierarchical clustering was performed using `phemap` v1.0.12 (RRID: SCR\_016418)<sup>83</sup>.

**Skeletal staining.** E18.5 skeletons from two litters from each of *HACNS1* homozygous, chimpanzee ortholog line, and wild type homozygous crosses ( $n = 48$  embryos) were stained with Alcian Blue and Alizarin Red as previously described<sup>71</sup>. Skeletons were imaged under a stereo microscope (Leica S6D) and measured by a single scorer blinded to genotype using `ImageJ` 2.0.0. Bone and cartilage lengths of the forelimb and hindlimb pelvic girdle, stylopod, zeugopod, and autopod were measured blinded to genotype using `ImageJ` 2.0.0. Forelimb measurements include metacarpals 1–5 (cartilage), proximal phalanges 1–5 (cartilage), intermediate phalanges 2–5 (cartilage), distal phalanges 1–5 (cartilage), scapula (bone and cartilage), humerus (bone and cartilage), radius (bone and cartilage), and ulna (bone and cartilage). Hindlimb measurements include metatarsals 1–5 (cartilage), proximal phalanges 1–5 (cartilage), intermediate phalanges 2–5 (cartilage), distal phalanges 1–5 (cartilage), tibia (bone and cartilage), femur (bone and cartilage), pelvis (cartilage), ilium (bone), ischium (bone), pubis (bone), fibula (bone), calcaneum (cartilage), and talus (cartilage). Digit length was calculated as the sum of all metacarpal/metatarsal and phalanx segments. Raw measurements and digit length were normalized to the length of ossified humerus or femur for forelimb or hindlimb digits, respectively. Phalange to metacarpal ratio was calculated as the ratio of the sum of the phalange lengths of each digit to the corresponding metacarpal/metatarsal segment. Interdigital ratios were calculated using raw digit lengths. Raw measurements and images are available as Source Data.

**ANOVA analysis for gene expression and morphometry.** ANOVA analysis was performed with the `lme4` package in R (RRID: SCR\_015654) using default parameters to dissect the effects of genotype on limb segment length (morphometric data)<sup>84</sup>. We calculated the effects of genotype, litter, sex, forelimb versus hindlimb, digit number, and right versus left (RL) on normalized digit length, phalange to metacarpal ratio and interdigital ratio ( $Length\ Ratio \sim Genotype * (1 | Genotype/Litter) * Sex * Limb * Digit * (1 | RL) * (1 | Litter/Embryo) * (1 | Sex/Embryo) * (1 | Genotype/Embryo)$ ). Correction for multiple comparisons was performed using the Benjamini & Hochberg method<sup>85</sup>.

**Statistics and reproducibility.** All ChIP-seq findings were validated using ChIP-qPCR of both the sequenced samples as well as additional biological replicates. Specificity of H3K27ac and H3K4me2 antibodies was validated by the authors using dot blot analysis. Additional validation measures including dot blot analysis and ChIP-qPCR were performed by Active Motif (<https://www.activemotif.com/documents/tds/39133.pdf> and <https://www.activemotif.com/documents/tds/39913.pdf>). RT-qPCR results shown in Supplementary Fig. 3C were validated with additional biological and technical replicates. All attempts at replication were successful. No statistical methods were used to predetermine sample size for ChIP-seq and RT-qPCR analyses and no data were excluded from these analyses. All samples prepared for ChIP-seq, RT-qPCR, ISH, scRNA-seq, or morphometric analysis as shown in the final figures were treated identically and processed in parallel. No statistical methods were used to predetermine optimal sample sizes for morphometric and ISH analyses. Instead, morphometric studies and ISH analyses were done using large sample sizes: limb samples from 48 embryos for morphometry and over 100 embryos obtained from multiple litters for each genotype for ISH analyses. For ISH and morphometric analyses, no data were excluded from the analyses; missing data values indicate samples that could not be evaluated/measured due to damage to the specimen. One scRNA-seq replicate from the chimpanzee ortholog line was excluded based on high overall mitochondrial gene expression, indicative of low sample quality based on preestablished filtering metrics. Qualitative analysis of ISH results were performed using a blinded approach by randomizing embryo identification numbers prior to annotation as described in the Methods. Morphometric data was collected blinded to genotype by randomizing sample identification numbers. ChIP-seq, RT-qPCR, and scRNA-seq were performed without group allocation blinding as all biological and technical replicates were processed identically and in parallel and no qualitative analyses were required for these experiments. We did not consider the sex of embryonic samples as a variable in our studies.

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

### Data availability

The Gene Expression Omnibus accession number for the data reported in this paper is “GSE141471”. The SRA accession number for the raw ChIP-seq and scRNA-seq data is “SRP234725” and the BioProject accession number is “PRJNA593575”. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable request. Source data are provided with this paper.

Received: 28 April 2021; Accepted: 16 December 2021;

Published online: 13 January 2022

### References

- Reilly, S. K. & Noonan, J. P. Evolution of gene regulation in humans. *Annu. Rev. Genomics Hum. Genet.* **17**, 45–67 (2016).
- Carroll, S. B. Genetics and the making of Homo sapiens. *Nature* **422**, 849–857 (2003).
- Prabhakar, S., Noonan, J. P., Pääbo, S. & Rubin, E. M. Accelerated evolution of conserved noncoding sequences in humans. *Science* **314**, 786 (2006).
- Pollard, K. S. et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**, 167–172 (2006).
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
- Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
- Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130025 (2013).
- Reilly, S. K. et al. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015).
- Kamm, G. B., Pisciotto, F., Kligler, R. & Franchini, L. F. The developmental brain gene NPAS3 contains the largest number of accelerated regulatory sequences in the human genome. *Mol. Biol. Evol.* **30**, 1088–1102 (2013).
- Prabhakar, S. et al. Human-specific gain of function in a developmental enhancer. *Science* **321**, 1346–1350 (2008).
- Cotney, J. et al. The evolution of lineage-specific regulatory activities in the human embryonic limb. *Cell* **154**, 185–196 (2013).
- Won, H., Huang, J., Opland, C. K., Hartl, C. L. & Geschwind, D. H. Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat. Commun.* **10**, 2396 (2019).
- Richard, D. et al. Evolutionary selection and constraint on human knee chondrocyte regulation impacts osteoarthritis risk. *Cell* **181**, 362–381.e328 (2020).
- Boyd, J. L. et al. Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* **25**, 772–779 (2015).
- Ryu, H. et al. Massively parallel dissection of human accelerated regions in human and chimpanzee neural progenitors. *bioRxiv*, 256313, <https://doi.org/10.1101/256313> (2018).
- Uebbing, S. et al. Massively parallel discovery of human-specific substitutions that alter enhancer activity. *Proc. Natl. Acad. Sci. USA* **118**, e2007049118 (2021).
- Caporale, A. L., Gonda, C. M. & Franchini, L. F. Transcriptional enhancers in the FOXP2 locus underwent accelerated evolution in the human lineage. *Mol. Biol. Evol.* **36**, 2432–2450 (2019).
- Acosta, S. et al. A Human Accelerated Region participates in early human forebrain patterning and expansion. *bioRxiv*, 777235, <https://doi.org/10.1101/777235> (2019).
- Aldea, D. et al. Repeated mutation of a developmental enhancer contributed to human thermoregulatory evolution. *Proc. Natl. Acad. Sci.* **118**, e2021722118 (2021).
- Wray, G. A. The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).
- King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).
- Cretekos, C. J. et al. Regulatory divergence modifies limb length between mammals. *Genes Dev.* **22**, 141–151 (2008).
- Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
- 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).
- Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Li, J. Y., Lao, Z. & Joyner, A. L. New regulatory interactions and cellular responses in the isthmus organizer region revealed by altering Gbx2 expression. *Development* **132**, 1971–1981 (2005).
- Wassarman, K. M. et al. Specification of the anterior hindbrain and establishment of a normal mid/hindbrain organizer is dependent on Gbx2 gene function. *Development* **124**, 2923–2934 (1997).
- Chatterjee, M. et al. Gbx2 regulates thalamocortical axon guidance by modifying the LIM and Robo codes. *Development* **139**, 4633–4643 (2012).
- Miyashita-Lin, E. M., Hevner, R., Wassarman, K. M., Martinez, S. & Rubenstein, J. L. Early neocortical regionalization in the absence of thalamic innervation. *Science* **285**, 906–909 (1999).
- Lin, Z., Cantos, R., Patente, M. & Wu, D. K. Gbx2 is required for the morphogenesis of the mouse inner ear: a downstream candidate of hindbrain signaling. *Development* **132**, 2309–2318 (2005).
- Byrd, N. A. & Meyers, E. N. Loss of Gbx2 results in neural crest cell patterning and pharyngeal arch artery defects in the mouse embryo. *Dev. Biol.* **284**, 233–245 (2005).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
- Visel, A., Minovitsky, S., Dubchak, I. & Pennacchio, L. A. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* **35**, D88–D92 (2007).
- Kaufman, M. H. Atlas of Mouse Development (Academic Press, 1995).
- Pennacchio, L. A. et al. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).
- Martin, P. Tissue patterning in the developing mouse limb. *Int. J. Dev. Biol.* **34**, 323–336 (1990).
- Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e1821 (2019).
- Civin, C. I. et al. Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells. *J. Immunol.* **133**, 157–165 (1984).
- Deutsch, E., Johnson, S. A. & Seeger, W. H. Differentiation of certain platelet factors related to blood coagulation. *Circ. Res.* **3**, 110–115 (1955).
- McInnes, L., Healy, J., Saul, N. & Großberger, L. Umap: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech.: theory Exp.* **2008**, P10008 (2008).
- Clement-Jones, M. et al. The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome. *Hum. Mol. Genet.* **9**, 695–702 (2000).

44. Favier, B. & Dollé, P. Developmental functions of mammalian Hox genes. *Mol. Hum. Reprod.* **3**, 115–131 (1997).
45. Gañan, Y., Macías, D., Duterque-Coquilaud, M., Ros, M. A. & Hurlé, J. M. Role of TGF beta s and BMPs as signals controlling the position of the digits and the areas of interdigital cell death in the developing chick limb autopod. *Development* **122**, 2349–2357 (1996).
46. Martin, G. R. The roles of FGFs in the early development of vertebrate limbs. *Genes Dev.* **12**, 1571–1586 (1998).
47. Sassoon, D. et al. Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. *Nature* **341**, 303–307 (1989).
48. Wright, E. et al. The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. *Nat. Genet.* **9**, 15–20 (1995).
49. Tabin, C. & Wolpert, L. Rethinking the proximodistal axis of the vertebrate limb in the molecular era. *Genes Dev.* **21**, 1433–1442 (2007).
50. Li, D. et al. Formation of proximal and anterior limb skeleton requires early function of *Irx3* and *Irx5* and is negatively regulated by *Shh* signaling. *Dev. Cell* **29**, 233–240 (2014).
51. Quinn, M. E., Haaning, A. & Ware, S. M. Preaxial polydactyly caused by *Gli3* haploinsufficiency is rescued by *Zic3* loss of function in mice. *Hum. Mol. Genet.* **21**, 1888–1896 (2012).
52. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**, 1401–1416 (1993).
53. Krishnaswamy, S. et al. Systems biology. Conditional density-based analysis of T cell signaling in single-cell data. *Science* **346**, 1250689 (2014).
54. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729.e27 (2018).
55. Burkhardt, D. B. et al. Quantifying the effect of experimental perturbations at single-cell resolution. *Nat. Biotechnol.* **39**, 619–629 (2021).
56. Kita, K., Kimura, T., Nakamura, N., Yoshikawa, H. & Nakano, T. PI3K/Akt signaling as a key regulatory pathway for chondrocyte terminal differentiation. *Genes Cells* **13**, 839–850 (2008).
57. Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N. & de Crombrughe, B. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(III) collagen gene. *Mol. Cell Biol.* **17**, 2336–2346 (1997).
58. Binns, D. et al. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
59. Shea, C. M., Edgar, C. M., Einhorn, T. A. & Gerstenfeld, L. C. BMP treatment of C3H10T1/2 mesenchymal stem cells induces both chondrogenesis and osteogenesis. *J. Cell Biochem.* **90**, 1112–1127 (2003).
60. Takigawa, Y. et al. The transcription factor *Znf219* regulates chondrocyte differentiation by assembling a transcription factory with *Sox9*. *J. Cell Sci.* **123**, 3780–3788 (2010).
61. Yoshida, C. A. et al. *Runx2* and *Runx3* are essential for chondrocyte maturation, and *Runx2* regulates limb growth through induction of Indian hedgehog. *Genes Dev.* **18**, 952–963 (2004).
62. Francis, P. H., Richardson, M. K., Brickell, P. M. & Tickle, C. Bone morphogenetic proteins and a signalling pathway that controls patterning in the developing chick limb. *Development* **120**, 209–218 (1994).
63. Rhinn, M., Lun, K., Werner, M., Simeone, A. & Brand, M. Isolation and expression of the homeobox gene *Gbx1* during mouse development. *Dev. Dyn.* **229**, 334–339 (2004).
64. Waters, S. T., Wilson, C. P. & Lewandoski, M. Cloning and embryonic expression analysis of the mouse *Gbx1* gene. *Gene Expr. Patterns* **3**, 313–317 (2003).
65. Gorkin, D. U. et al. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020).
66. Lettice, L. A. et al. Opposing functions of the ETS factor family define *Shh* spatial expression in limb buds and underlie polydactyly. *Dev. Cell* **22**, 459–467 (2012).
67. Zhao, F., Bosserhoff, A. K., Buettner, R. & Moser, M. A heart-hand syndrome gene: *Tfap2b* plays a critical role in the development and remodeling of mouse ductus arteriosus and limb patterning. *PLoS One* **6**, e22908 (2011).
68. Kim, E. J. et al. *Ihh* and *Runx2/Runx3* signaling interact to coordinate early chondrogenesis: a mouse model. *PLoS One* **8**, e55296 (2013).
69. Cunningham, M. L. Craniofacial development, by G.H. Sperber. *Am. J. Med. Genet.* **103**, 91–91 (2001).
70. Rodríguez-Vázquez, J. F., Mérida-Velasco, J. R., Verdugo-López, S., Sánchez-Montesinos, I. & Mérida-Velasco, J. A. Morphogenesis of the second pharyngeal arch cartilage (Reichert's cartilage) in human embryos. *J. Anat.* **208**, 179–189 (2006).
71. Behringer, R., Gertsenstein, M., Nagy, K. V. & Nagy, A. Manipulating the mouse embryo: a laboratory manual. (Cold Spring Harbor Laboratory Press, 2014).
72. Petersen, M. C. et al. Insulin receptor Thr1160 phosphorylation mediates lipid-induced hepatic insulin resistance. *J. Clin. Invest* **126**, 4361–4371 (2016).
73. Rhead, B. et al. The UCSC Genome Browser database: update 2010. *Nucleic Acids Res* **38**, D613–D619 (2010).
74. Cotney, J. et al. Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. *Genome Res.* **22**, 1069–1080 (2012).
75. Cotney, J. L. & Noonan, J. P. Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. *Cold Spring Harb. Protoc.* **2015**, 419 (2015).
76. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
77. Andrews, S. FastQC: a quality control tool for high throughput sequence data [Online]. Available online at: [www.bioinformatics.babraham.ac.uk/projects/fastqc/](http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). (2010).
78. Louvi, A. & Wassef, M. Ectopic engrailed 1 expression in the dorsal midline causes cell death, abnormal differentiation of circumventricular organs and errors in axonal pathfinding. *Development* **127**, 4061–4071 (2000).
79. Louvi, A., Yoshida, M. & Grove, E. A. The derivatives of the Wnt3a lineage in the central nervous system. *J. Comp. Neurol.* **504**, 550–569 (2007).
80. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv*, 576827, <https://doi.org/10.1101/576827> (2019).
81. Linderman, G. C., Zhao, J. & Kluger, Y. Zero-preserving imputation of scRNA-seq data using low-rank approximation. *bioRxiv*, 397588, <https://doi.org/10.1101/397588> (2018).
82. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for gene ontology. R package version 2.34.0. (2018).
83. Kolde, R. pheatmap: Pretty Heatmaps. R package version 1.0.8. (2015).
84. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
85. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.: Ser. B (Methodol.)* **57**, 289–300 (1995).

## Acknowledgements

We thank Steven K. Reilly for contributing graphics shown in Fig. 1A. This work was supported by a grant from the National Institute of General Medical Sciences (NIGMS) (R01 GM094780, to J.P.N.), a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01 HD102030 to J.P.N.), funds from the Yale School of Medicine (to J.P.N.), and funds from the Yale Program on Neurogenetics (to A.L.). E.V.D. was supported in part by NIGMS training grant T32 GM007499. D.E. was supported by an NIH F32 Postdoctoral Fellowship (NIGMS) (F32 GM106628). A.K. was supported by an NSF Graduate Research Fellowship (DGE-1122492). K.Y. was supported by an NSF Graduate Research Fellowship (DGE-1752134). S.U. was supported by a Research Fellowship (352711928) from the Deutsche Forschungsgemeinschaft (DFG). M.K. was supported by a Research Fellowship (387495052) from the Deutsche Forschungsgemeinschaft (DFG). D.B. was supported by an NIH F31 Predoctoral Fellowship (NICHD) (F31 HD097958). S.K. was supported by NIH grants R01 GM135929 and R01 GM130847, and Chan-Zuckerberg Initiative grants 182702 and CZF2019-002440. This research program and related results were also made possible by the support of the NOMIS foundation (to J.P.N.).

## Author contributions

E.V.D. and J.P.N. conceived of and designed the study with input from D.E.; E.V.D. performed the mouse line validation, chromatin immunoprecipitation, RT-qPCR, single cell RNA-sequencing, and skeletal staining experiments; E.V.D. and A.L. carried out the in situ hybridization experiments with assistance from A.A.K. and M.K.; D.E. and T.N. initially generated the mouse lines used in the study. Data analysis and interpretation for the single-cell RNA-sequencing experiments was performed by E.V.D. with input from K.Y., D.B.B. and S.K. S.U. performed the ANOVA analyses. E.V.D., A.L. and J.P.N. wrote the manuscript with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27899-w>.

**Correspondence** and requests for materials should be addressed to James P. Noonan.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022