The Human Accelerated Region *HACNS1* modifies developmental gene expression in humanized mice

Emily V. Dutrow[1], Deena Emera[1†], Kristina Yim[1], Severin Uebbing[1], Acadia A. Kocher[1], Martina Krenzer[1], Timothy Nottoli[2,3], Daniel B. Burkhardt[1], Smita Krishnaswamy[1,4], Angeliki Louvi[5,6], James P. Noonan[1,6,7+]

1. Department of Genetics, Yale School of Medicine, New Haven, CT, 06510, USA.
2. Department of Comparative Medicine, Yale School of Medicine, New Haven, CT, 06510, USA.
3. Yale Genome Editing Center, Yale School of Medicine, New Haven, CT, 06510, USA.
4. Department of Computer Science, Yale University, New Haven, CT, 06510, USA.
5. Department of Neurosurgery, Yale School of Medicine, New Haven, CT, 06510, USA.
6. Department of Neuroscience, Yale School of Medicine, New Haven, CT, 06510, USA.
7. Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, 06520, USA.

†Present address: Center for Female Reproductive Longevity and Equality, Buck Institute for Research on Aging, Novato, CA 94945, USA

+Corresponding author: james.noonan@yale.edu

Abstract

Morphological innovations that arose during human evolution are ultimately encoded in genetic changes that altered development. Human Accelerated Regions (HARs), which include developmental enhancers that harbor a significant excess of human-specific sequence changes, are leading candidates for driving novel physical modifications in humans. Here we examine the role of the HAR *HACNS1* (also known as HAR2) in human limb evolution by directly interrogating its cellular and developmental functions in a humanized mouse model. *HACNS1* encodes an enhancer with human-specific activity in the developing limb in transgenic mouse reporter assays, and exhibits increased epigenetic signatures of enhancer activity in the human embryonic limb compared to its orthologs in rhesus macaque and mouse. Here we find that *HACNS1* maintains its human-specific enhancer activity compared to its chimpanzee ortholog in the mouse embryonic limb, and that it alters expression of the transcription factor gene *Gbx2* during limb development. Using single-cell RNA-sequencing, we demonstrate that *Gbx2* is upregulated in humanized limb bud chondrogenic mesenchyme, implicating *HACNS1*-mediated *Gbx2* expression in early skeletal patterning. Our findings establish that HARs direct changes in the level and distribution of gene expression during development, and illustrate how humanized mouse models provide insight into regulatory pathways modified in human evolution.

Introduction

The evolution of uniquely human physical traits, including changes in limb morphology that allow us to use tools and to walk upright or the expansion of the cerebral cortex, required human-specific genetic changes that altered development (Aiello and Dean, 1990; Carroll, 2003; Geschwind and Rakic, 2013; Reilly and Noonan, 2016). Discovering the locations of these changes and determining their biological impact is a major challenge. However, over the last decade comparative studies have begun to reveal potential genetic drivers underlying uniquely human morphological 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 (Capra et al., 2013; Lindblad-Toh et al., 2011; Pollard et al., 2006, 2010; Prabhakar et al., 2006). These elements, collectively named Human Accelerated Regions (HARs), are prime candidates to encode novel human molecular functions. Experimental studies revealed that many HARs act as transcriptional enhancers during embryonic development, particularly in structures showing human-specific morphological changes such as the brain and limb (Capra et al., 2013; Cotney et al., 2013; Kamm et al., 2013; Prabhakar et al., 2008; Reilly et al., 2015; Won et al., 2019). HARs have also been shown to exhibit human-specific changes in enhancer activity, both in transgenic mouse enhancer assays and in massively parallel reporter assays in cultured cells (Capra et al., 2013; Kamm et al., 2013; Prabhakar et al., 2008; Ryu et al., 2018; Uebbing et al., 2019). These findings suggest a critical contribution for HARs in human evolution and support the long-standing hypothesis that changes in gene regulatory programs during development contribute to evolutionary innovation (Britten and Davidson, 1971; King and Wilson, 1975; Wray, 2007).

Despite these advances, the role of HARs in altering developmental processes remains poorly understood. Reporter assays in transgenic mice and cultured cells have identified HARs that show differential enhancer activity compared to their chimpanzee orthologs, but cannot provide insight into developmental phenotypes. Similar approaches have been used to drive expression of HAR putative target genes in transgenic mice. For example, the potential impact of *HARE5* on expression of its putative target gene *FZD8* was assessed using transgenes in which either *HARE5* or its chimpanzee ortholog drove expression of a mouse *Fzd8* cDNA (Boyd et al., 2015). Mice with the *HARE5*::Fzd8 transgene showed modest increases in neuron number and neocortical size compared to the chimpanzee ortholog control line. While such approaches

contribute to our understanding of HAR regulatory activity, they do not address how HARs function in native genomic contexts.

Here we use a humanized mouse model approach to directly characterize the effect of human-specific sequence changes in HARs on gene expression and regulation during embryonic development. We chose to model *HACNS1* (also known as HAR2 or 2xHAR.3), which exhibits the strongest acceleration signature of any noncoding HAR yet identified, with 13 human-specific substitutions in a 546 base-pair interval (Fig. 1). *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 (Fig.1B; Prabhakar et al., 2008). Furthermore, *HACNS1* exhibits increased levels of the histone modification H3K27ac, which is correlated with enhancer activity, in the human embryonic limb compared to rhesus macaque and mouse (Fig. 1C; Cotney et al., 2013). Together, these findings suggest that *HACNS1* may have contributed to changes in limb development during human evolution.

We used homologous recombination to replace the mouse ortholog of *HACNS1* with the human or chimpanzee sequence, enabling us to compare the functions of all three orthologs in the same developmental system. We found that *HACNS1* maintains its human-specific activity in the mouse embryo, altering the expression of the nearby transcription factor *Gbx2* in the developing limb bud. Through single-cell RNA-sequencing, we showed that the gain of function in *HACNS1* alters gene expression in limb chondrocytes, suggesting it may affect skeletal morphogenesis. Our findings establish that HARs direct human-specific changes in gene expression during development, and illustrate how humanized mouse models provide insight into regulatory pathways and developmental mechanisms modified in human evolution.

Results

Design and generation of the *HACNS1* humanized mouse line

To generate a humanized *HACNS1* mouse model, we designed a targeting construct for homologous recombination that included a 1.2 kb human sequence encompassing *HACNS1*, previously shown to encode human-specific enhancer activity in transgenic mouse embryos (Prabhakar et al., 2008). We replaced the orthologous mouse locus using homology-directed repair in C57BL6/J-A^{w-J}/J (B6 agouti) embryonic stem (ES) cells (Fig. 1D, Fig. S1A,B; Materials and Methods). To provide a control that would enable us to identify *bona fide* human-specific functions of *HACNS1*, we used the same approach to generate a mouse model for the orthologous chimpanzee sequence. The 1.2 kb chimpanzee sequence includes 22 single nucleotide differences from the human allele, 15 of which are specific to the human lineage compared to other primates (See Materials and Methods).

In order to verify the integrity of the edited sequences, 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* (*HACNS1*$^{+/+}$), or the chimpanzee ortholog (Fig. S1C; Materials and Methods). We found no evidence of aberrant editing, sequence rearrangement, 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 (qRT-PCR) (Fig. S1D).
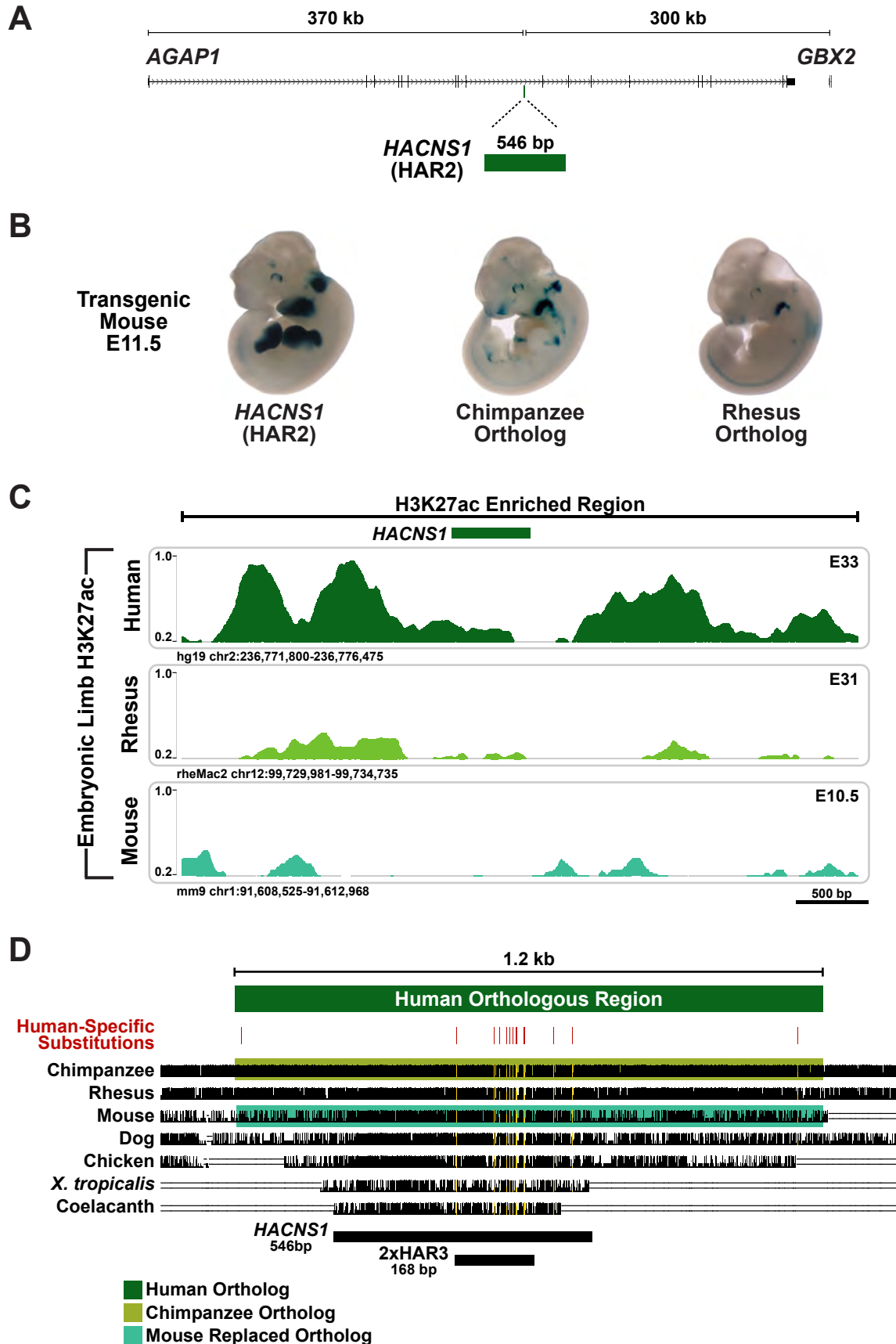
5

A

370 kb — 300 kb

AGAP1 *GBX2*

HACNS1 (HAR2) — 546 bp

B

Transgenic Mouse E11.5

HACNS1 (HAR2) — Chimpanzee Ortholog — Rhesus Ortholog

C

H3K27ac Enriched Region

HACNS1

Embryonic Limb H3K27ac

Human — E33
hg19 chr2:236,771,800-236,776,475

Rhesus — E31
rheMac2 chr12:99,729,981-99,734,735

Mouse — E10.5
mm9 chr1:91,608,525-91,612,968

500 bp

D

1.2 kb

Human Orthologous Region

Human-Specific Substitutions

Chimpanzee
Rhesus
Mouse
Dog
Chicken
X. tropicalis
Coelacanth

HACNS1 546bp
2xHAR3 168 bp

■ Human Ortholog
■ Chimpanzee Ortholog
■ Mouse Replaced Ortholog

6

Fig. 1. *HACNS1* **encodes an enhancer with human-specific activity in the developing limb.** (**A**) The location of *HACNS1* in the human genome (GRCh37/hg19), relative to the nearby genes *AGAP1* and *GBX2*. (**B**) *LacZ* reporter gene expression (in blue) driven by *HACNS1,* the chimpanzee ortholog, or the rhesus macaque ortholog in embryonic day 11.5 (E11.5) transgenic mouse embryos. Images are adapted from Prabhakar *et al.* 2008. (**C**) The normalized H3K27ac signal in the developing limb bud at *HACNS1* and the orthologous loci in rhesus macaque and mouse. Data were obtained from Cotney *et al.* 2013. The region of significant H3K27ac enrichment in human is shown by a black bar (top), and the location of *HACNS1* is shown as a green bar. (**D**) Alignment of the human sequence used to generate the *HACNS1* humanized mouse with orthologous sequences from other vertebrate genomes, obtained from the UCSC hg19 100-way Multiz alignment (see Table S1 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 shown by a red line, and the corresponding positions in the alignment are highlighted in yellow. The locations of *HACNS1* (Prabhakar et al., 2006) and 2xHAR3 (Lindblad-Toh et al., 2011) are shown below the alignment.

7

***HACNS1* exhibits human-specific enhancer activity in the humanized mouse embryo.**

We used chromatin immunoprecipitation (ChIP) to determine if *HACNS1* exhibits epigenetic signatures of increased enhancer activity in the humanized mouse model. We initially focused on the limb for epigenetic profiling, as the human-specific activity of *HACNS1* was previously demonstrated in mouse embryonic limb in transgenic reporter assays and by a gain of H3K27ac in human embryonic limb (Cotney et al., 2013; Prabhakar et al., 2008). We profiled H3K27ac and H3K4me2, which are both associated with enhancer activity, in embryonic day 11.5 (E11.5) limb buds from *HACNS1*[+/+] embryos, embryos homozygous for the chimpanzee ortholog, and wild type embryos. We found a strong signature of H3K27ac marking at *HACNS1* in *HACNS1*[+/+] limb (Fig. 2, S2). The chimpanzee and mouse sequences both showed significant but weaker H3K27ac enrichment than the human sequence, supporting the conclusion that *HACNS1* maintains its human-specific enhancer activity in the humanized mouse model.

To quantify the differences in epigenetic marking we observed, we used DESeq2 implemented in HOMER (Materials and Methods; Heinz et al., 2010; Love et al., 2014) to identify genome-wide significant changes in H3K27ac and H3K4me2 levels in limb bud of mice homozygous for *HACNS1* or the chimpanzee ortholog versus wild type. We found that H3K27ac and H3K4me2 levels at the edited *HACNS1* locus were significantly increased in *HACNS1*[+/+] limb bud compared to wild type (Fig. 2, S2, Table S4). Although the chimpanzee locus also exhibited increased H3K27ac marking over input controls, the level in the chimpanzee locus was not significantly differentially increased compared to wild type (Fig. S2A, Table S4). The levels of H3K4me2 were significantly increased at both the humanized and the chimpanzee orthologous loci in each respective line compared to wild type (Fig. 2, S2, Table S4). Increased levels of H3K4me2 coupled with low levels of H3K27ac are associated with weak enhancer activity, so it is likely that the chimpanzee sequence is not acting as a strong enhancer in the limb bud overall but may be active in other tissues (Ernst et al., 2011), a finding further supported by gene expression analyses described in Figure 3 below. However, our results suggest that the chimpanzee ortholog may show stronger enhancer activity than the mouse sequence, pointing to potential primate-rodent differences in addition to the strong human-specific gain of function in *HACNS1*.

8

Because previous transgenic mouse enhancer assays showed that *HACNS1* also drives increased reporter gene expression activity in the pharyngeal arch (Prabhakar et al., 2008), we profiled H3K27ac and H3K4me2 in pharyngeal arch tissue from $HACNS1^{+/+}$ embryos, embryos homozygous for the chimpanzee ortholog, and wild type embryos at E11.5. We detected reproducible, significant enrichment of H3K27ac in the pharyngeal arch at the humanized and orthologous chimpanzee locus compared to input controls, but H3K27ac marking at the humanized locus was not significantly different than that at the wild type locus (Fig. S2). We did, however, identify a significant gain of H3K4me2 signals in the pharyngeal arch at the humanized and orthologous chimpanzee locus compared to wild type (Fig. S2, Table S4).

In order to identify downstream epigenetic changes resulting from *HACNS1* activation in $HACNS1^{+/+}$ 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 $HACNS1^{+/+}$ limb bud versus wild type at the promoter of the nearby gene *Gbx2* (Fig. 2, S2, Table S4). 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 buds homozygous for 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.

Gbx2 encodes a transcription factor with multiple functions during development. GBX2 has been implicated in midbrain and hindbrain development (Li et al., 2005; Wassarman et al., 1997), guidance of thalamocortical projections (Chatterjee et al., 2012; Hevner et al., 2002; Miyashita-Lin et al., 1999), ear development (Lin et al., 2005; Miyazaki et al., 2006), and pharyngeal arch patterning (Byrd and Meyers, 2005). *Gbx2* was previously found to be expressed in developing mouse limb; however, its role in limb development remains undetermined given the lack of a reported limb phenotype in *Gbx2* knockout mice (Wassarman et al., 1997). Our findings suggest that *Gbx2* is a regulatory target of *HACNS1*, and that *HACNS1* gain of function may alter its expression in the humanized mouse limb.
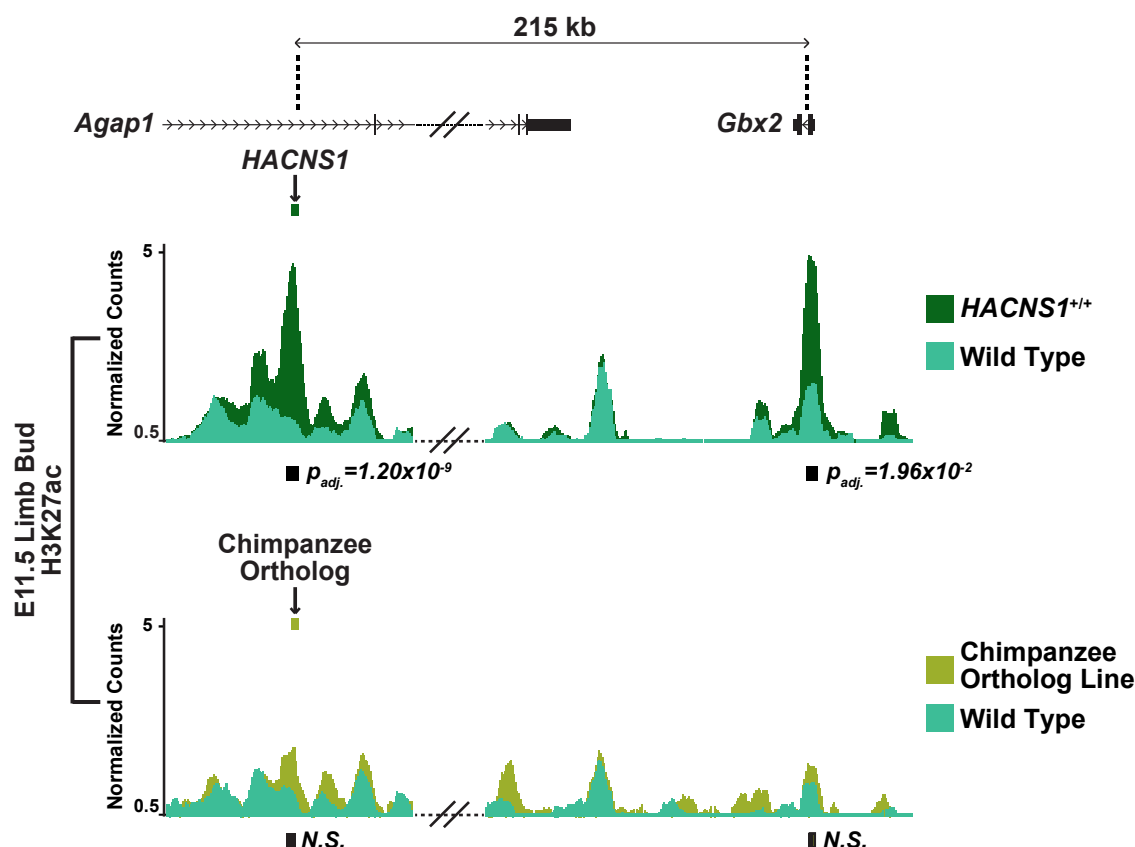
Fig. 2. *HACNS1* and the *Gbx2* promoter exhibit epigenetic signatures of increased activity in the humanized mouse limb bud. *Top:* Epigenetic profiling in humanized E11.5 limb bud compared to wild type. The normalized H3K27ac signals are shown for *HACNS1*[+/+] (in dark green) and litter-matched wild type (in teal). *Bottom:* Epigenetic profiling in chimpanzee ortholog line E11.5 limb bud compared to wild type. The normalized H3K27ac signals are shown for the chimpanzee ortholog line (in olive) and litter-matched wild type (in teal). The location of the edited locus in each line relative to nearby genes is shown above each 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 Fig. S2 for the complete view in each line. H3K27ac peak calls showing significant increases in signal between *HACNS1*[+/+] and wild type, and the corresponding peak regions compared between the chimpanzee control line and wild type are shown as black squares below each respective track (see Materials and Methods); *N.S.* = not significant. All peak calls for each line are shown in Fig. S2. Benjamini-Hochberg adjusted *P* values were obtained using DESeq2 implemented in HOMER (Heinz et al., 2010; Love et al., 2014).

10

HACNS1 drives spatial and quantitative changes in *Gbx2* expression in the limb bud

To identify potential expression changes resulting from *HACNS1*-driven activation of the *Gbx2* promoter in humanized mouse embryos, we used *in situ* hybridization (ISH). We found that $HACNS1^{+/+}$ E11.5 embryos showed increased *Gbx2* expression in both forelimb and hindlimb compared to embryos homozygous for the chimpanzee ortholog and wild type embryos (Fig. 3A). We observed an increase in *Gbx2* expression in $HACNS1^{+/+}$ embryos 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* limb bud expression was temporally dynamic, decreasing from E11 to E12. We therefore established a fine staging scheme to characterize changes in *Gbx2* expression within this short time interval. We used crown-rump length measurements to assign embryos to 6 temporally ordered groups (designated T1-T6, and ranging from approximately 36 to 43 somites; Tam, 1981) (Fig. 3A).

We identified substantial changes in the distribution of *Gbx2* expression in the forelimb and hindlimb buds of $HACNS1^{+/+}$ embryos compared to both homozygous chimpanzee ortholog embryos and wild type embryos across all 6 developmental stages (Fig. 3, S3). At the earliest time point, we found that *Gbx2* was strongly expressed in distinct anterior-distal and posterior domains in $HACNS1^{+/+}$ forelimb and hindlimb bud (Fig. 3, S3). Robust expression of *Gbx2* in $HACNS1^{+/+}$ limb buds persisted throughout the remaining timepoints, though the size of the anterior and posterior domains decreased over time. We also observed that strong expression of *Gbx2 in HACNS1$^{+/+}$* embryos persisted for a longer period of time in hindlimb bud than in forelimb, consistent with the delayed developmental maturation of the hindlimb compared to forelimb (Martin, 1990). In addition, $HACNS1^{+/+}$ embryos showed a hindlimb-specific anterior-proximal expression domain adjacent to the body wall across all 6 timepoints (Fig. S3B).

Compared to the robust expression in the humanized limb, *Gbx2* expression in both homozygous chimpanzee ortholog and wild type limb buds was weak and restricted to early time points (Fig. 3, S3A). Both homozygous chimpanzee ortholog embryos and wild type embryos showed a weak distal *Gbx2* expression pattern in early forelimb and hindlimb that was primarily restricted to the anterior portion of the limb bud (Fig. 3, S3A). While weak distal expression was restricted to approximately T1-T2 and T1-T4 in wild type forelimb and hindlimb, respectively, it persisted until approximately T3 in forelimb and T5 in hindlimb in homozygous chimpanzee ortholog embryos (Fig. 3, S3A). This finding further supports that the chimpanzee ortholog may

exhibit greater enhancer activity than the mouse ortholog, but substantially less activity than *HACNS1*.

In addition to the forelimb and hindlimb bud, *Gbx2* was also expressed in the neural tube, diencephalon, and pharyngeal arch of $HACNS1^{+/+}$embryos, homozygous chimpanzee ortholog embryos, and wild type embryos (Fig. 3A, S3C). Whereas *Gbx2* expression was primarily restricted to the first pharyngeal arch in chimpanzee ortholog line and wild type embryos, we observed a gain of a dorsal expansion of *Gbx2* expression into the second pharyngeal arch in $HACNS1^{+/+}$ embryos during T1-T5 (Fig. S3C). However, we chose to focus on limb due to a lack of comparative epigenetic profiling data in human pharyngeal arch and the absence of a significant gain in H3K27ac marking in $HACNS1^{+/+}$ versus wild type pharyngeal arch.

In order to quantify the gain of *Gbx2* expression in humanized limb bud, we used real-time quantitative reverse transcription PCR (qRT-PCR) in forelimb and hindlimb buds from $HACNS1^{+/+}$ embryos, homozygous chimpanzee ortholog embryos, and wild type embryos at each of the six developmental stages we defined above. We found that *Gbx2* showed increased expression in $HACNS1^{+/+}$ forelimb and hindlimb versus the chimpanzee ortholog line and wild type at all 6 timepoints (Fig. 3C). Although we detected an increase in *Gbx2* expression in the chimpanzee ortholog line versus wild type in early forelimb and hindlimb timepoints, this change was weaker than that between $HACNS1^{+/+}$ and wild type. Consistent with our ISH results, we found that *Gbx2* expression in humanized forelimb and hindlimb was strongest at the earliest timepoints and that expression persisted longer in hindlimb than forelimb (Fig. 3C). In order to determine the significance of the effects of genotype and timepoint on *Gbx2* expression, we used analysis of variance (ANOVA). We determined that genotype, timepoint, and genotype-timepoint interaction effects were significant, indicating that *Gbx2* expression was significantly increased in humanized mouse. While *Gbx2* expression declines over time in all three genotypes, it persists longer in humanized forelimb and hindlimb (Tables 1, 2).
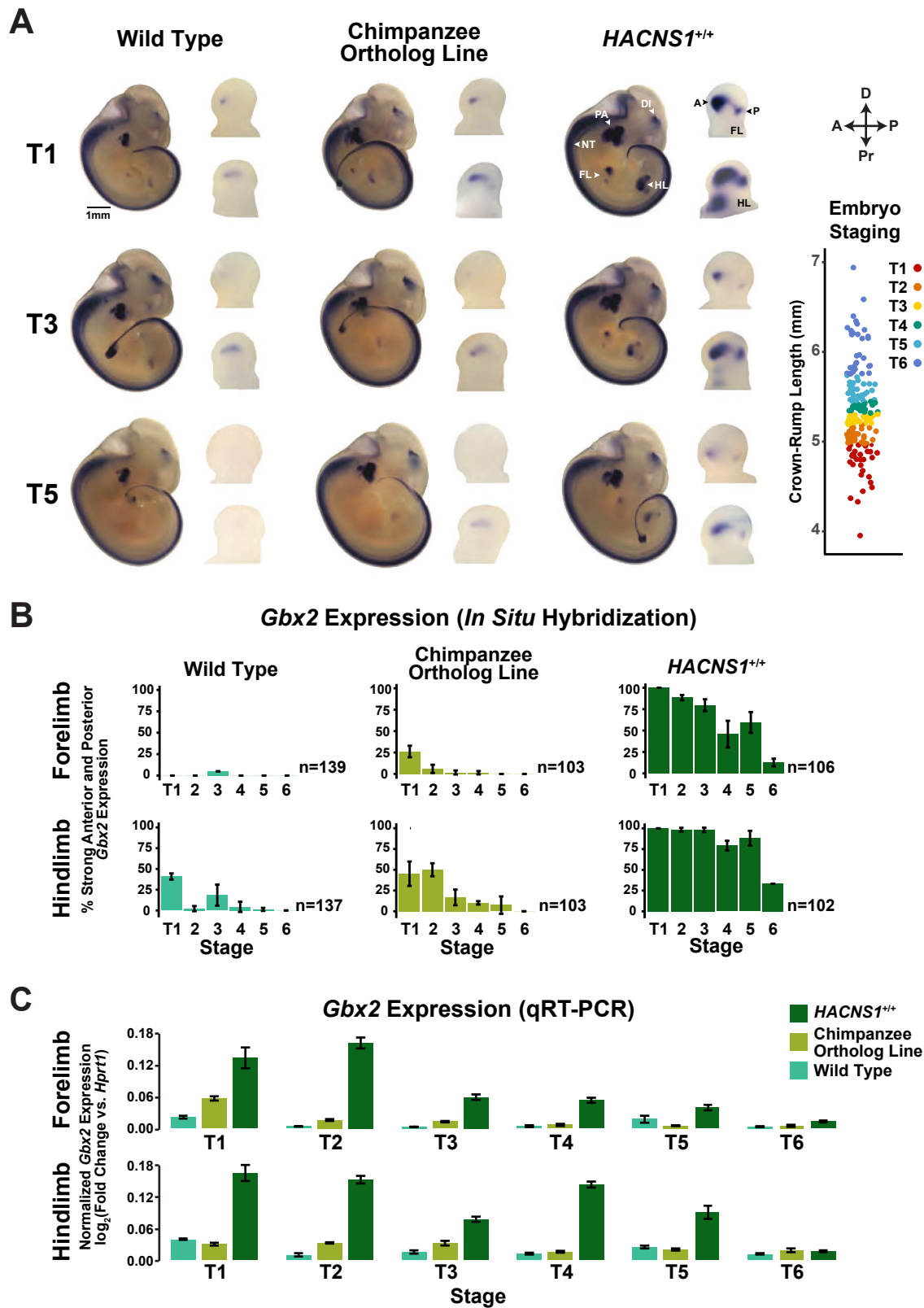
Fig. 3. *HACNS1* drives spatial and temporal changes in *Gbx2* expression in humanized mouse embryos. (A) Spatial and temporal expression of *Gbx2* in *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type E11-E12 embryos visualized by whole-mount *in situ* hybridization (ISH). Representative images are shown for each genotype at three fine-scale time points; see text and Fig. S3 for details on staging. Magnified images showing *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)** Percentage of embryos showing strong anterior and posterior expression of *Gbx2* at each time point in *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type forelimb (top) and hindlimb (bottom). Error bars denote standard deviation for results obtained from three independent, blinded analyses; for scoring scheme and full analysis see Fig. S3A. **(C)** Normalized *Gbx2* expression in *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type forelimb and hindlimb across all six timepoints measured using qRT-PCR. Error bars denote standard deviation across three technical replicates.

14

Table 1. Time Series qRT-PCR Forelimb ANOVA Analysis

Variable	*Df*	*Sums of Squares*	*Mean Squares*	*F*	*P$_{adj.}$*
Genotype	2	99.56	49.78	73.85	6.66E-15
Timepoint	1	41.15	41.15	61.05	6.33E-10
Genotype:Timepoint	2	7.65	3.82	5.67	6.14E-03
Residuals	48				

Table 2. Time Series qRT-PCR Hindlimb ANOVA Analysis

Variable	*Df*	*Sums of Squares*	*Mean Squares*	*F*	*P$_{adj.}$*
Genotype	2	59.45	29.72	66.62	4.26E-14
Timepoint	1	13.18	13.18	29.55	2.70E-06
Genotype:Timepoint	2	4.20	2.10	4.71	1.36E-02
Residuals	48				

15

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

Using bulk tissue analyses, we found that *HACNS1* drives increased *Gbx2* expression in distinct anterior and posterior domains in the humanized limb bud. In order to identify the precise cell types expressing *Gbx2* as well as genes potentially coregulated with *Gbx2,* we performed single-cell RNA-sequencing (scRNA-seq) in E11.5 hindlimb bud from $HACNS1^{+/+}$ embryos, homozygous chimpanzee ortholog embryos, and wild type embryos. Using the 10x Genomics scRNA-seq platform for cell barcoding, library preparation, and sequencing, we obtained transcriptomes from approximately 10,000 cells per genotype. We used the Seurat toolkit for data preprocessing and library size normalization (see Materials and Methods; Stuart et al., 2019). During pre-processing, we removed endothelial and blood cells (*Cd34*-positive, Civin et al., 1984; *Pf4*-positive, Deutsch et al., 1955; *Hbb*-positive, Bunn and Forget, 1986), as preliminary analysis indicated that these transcriptionally and developmentally distinct cell types do not express *Gbx2* in any of our datasets. 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 (see Materials and Methods), we performed a clustering analysis on the integrated dataset to identify cell type categories present in all three genotypes (Stuart et al., 2019). In order to visualize similarities between cells, we used Uniform Manifold Approximation and Projection for Dimensional Reduction (UMAP), a dimensionality reduction method for data visualization (McInnes et al., 2018).

Though the UMAP embedding reveals a continuum of cell types, we used the Louvain method for community detection to obtain a finer granularity of subtyping (Blondel et al., 2008). This analysis revealed three distinct groups: clusters 1-4 comprised mesenchymal cell subtypes based on expression of the known marker genes *Sox9* (clusters 1, 3, 4; Wright et al., 1995), *Bmp4* (cluster 2*;* Gañan et al., 1996)*, Shox2* (cluster 1*;* Clement-Jones, 2000), and *Hoxd13* (clusters 1-4; Favier, 1997) (Fig. 4A). The remaining two groups comprised non-mesenchymal cell types, including myogenic cells (cluster 5, *Myod1*; Sassoon et al., 1989) and ectodermal cells (cluster 6, *Fgf8*; Martin, 1998) (Fig. 4A). Furthermore, the UMAP embedding revealed finer separation of mesenchymal cells according to known limb geographical patterning markers. We first examined the expression of known proximal-distal limb bud markers *Meis1, Hoxa11*, and *Hoxd13* (proximal, medial, and distal, respectively; Tabin and Wolpert, 2007). 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 recapitulates known gene expression gradients along the anterior-posterior limb bud axis based on expression of the anterior-proximal marker *Irx3* (Li et al., 2014)*,* the anterior marker *Zic3* (Quinn et al., 2012)*,* and the posterior-proximal marker *Shh* (Riddle et al., 1993) (Fig. 4B, upper right). Using markers of chondrogenic (*Sox9, Shox2*; Wright et al., 1995; Clement-Jones, 2000) versus non-chondrogenic (*Bmp4*; Gañan et al., 1996) mesenchyme, we found that the second UMAP axis follows the chondrogenic versus interdigital apoptotic fate gradient (Fig. 4B). We found that the expression patterns of the aforementioned markers were broadly conserved between genotypes, with each genotype showing comparable subsets of proximal, distal, anterior, posterior, chondrogenic, and non-chondrogenic cell types (Fig. S4). Collectively, these results establish that our scRNA-seq analysis 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 humanized limb bud, we examined the distribution of *Gbx2*-positive cells across cell clusters in hindlimb buds from all three genotypes. We found that *Gbx2* was upregulated in *HACNS1*[+/+] hindlimb versus homozygous chimpanzee ortholog and wild type hindlimbs, primarily in the mesenchymal cell clusters (clusters 1-4), consistent with our ISH and qRT-PCR expression analyses (Fig. 4C and Fig. 3). In *HACNS1*[+/+] hindlimb, 24% of cells expressed *Gbx2*, 96% of which were mesenchymal cells, whereas less than 1% of homozygous chimpanzee ortholog or wild type cells expressed *Gbx2* (see Materials and Methods; Table S10). UMAP embedding of these cells reveals that *Gbx2*-positive cells in *HACNS1*[+/+] hindlimb bud largely cluster within a distinct subset of mesenchymal cells belonging primarily to Louvain clusters 1, 2 and 4 (Fig. 4C, D; Table S10).
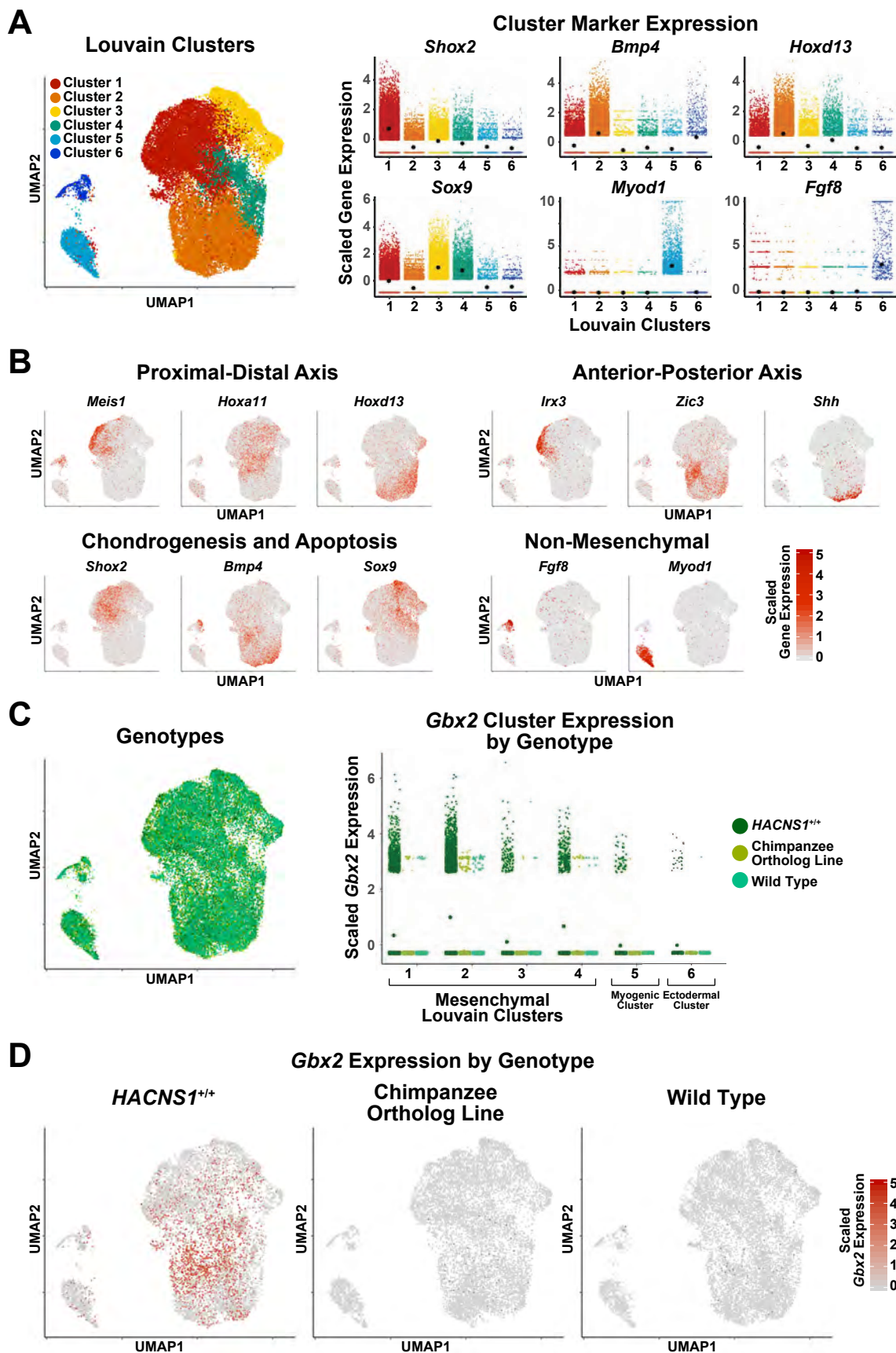
17

Fig. 4. Single-cell transcriptome analysis of the E11.5 hindlimb bud in *HACNS1*$^{+/+}$**, chimpanzee ortholog line, and wild type mice. (A)** *Left:* UMAP embedding of *HACNS1*$^{+/+}$, 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*$^{+/+}$, chimpanzee ortholog line, and wild type, showing expression of proximal-distal, anterior-posterior, chondrogenesis-apoptosis, and non-mesenchymal markers. See text and Fig. S4 for details. **(C)** *Left:* UMAP embedding showing *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type cells colored by genotype. *Right:* Expression of *Gbx2* in each Louvain cluster, separated by genotype. Black dots denote cluster mean expression. **(D)** UMAP embeddings illustrating cells expressing *Gbx2* (indicated in red) in *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type cells. All gene expression data shown in violin plots and UMAP embeddings (A-D) were imputed using ALRA (Linderman et al., 2018) and centered and scaled using z-scores.

To find 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 (van Dijk et al., 2018; Krishnaswamy et al., 2014). Using kNN-DREMI scores, we ranked each gene expressed in *HACNS1*$^{+/+}$ limb by the strength of its association with *Gbx2*. To determine if genes associated with *Gbx2* were enriched in particular functions, we performed Gene Set Enrichment Analysis (GSEA) on all expressed genes ranked by strength of association with *Gbx2*. We found that *Gbx2* expression is associated with genes in several limb development-related ontologies, including "Cartilage Morphogenesis" (Kolmogorov-Smirnov (KS) P=1.61x10^{-3}), and "Regulation of Chondrocyte Differentiation" (KS P=5.84x10^{-3}); the latter overlaps considerably with "Collagen Fibril Organization" (KS P=1.30x10^{-4}) (Fig. 5A, Table S11). These results indicate that in humanized mesenchyme, *Gbx2* is co-regulated with genes expressed in condensing mesenchymal cells destined to become chondrocytes (Fig. 5A).

We also used an orthogonal approach that is naïve to our identification of *Gbx2* as the target of *HACNS1* in order to characterize cells with altered transcriptional profiles in humanized limb bud. We used Manifold Enhancement of Latent Dimensions (MELD), a method that uses graph signal processing to identify gene expression trends in the data corresponding to genotype (Burkhardt et al., 2019), in order to distinguish cells that are most prototypical of an experimental condition - in our case, *HACNS1*$^{+/+}$ hindlimb cells versus hindlimb cells from the chimpanzee ortholog line and wild type. The output of MELD is an "Enhanced Experimental Signal," which assigns a score for each cell that reflects the degree to which that cell is altered by the experimental condition - in this case, humanization (Burkhardt et al., 2019). In order to identify gene expression patterns characteristic of humanized hindlimb bud cells, we used kNN-DREMI to associate gene expression with EES scores generated by MELD. We then used the resulting gene rankings to identify enriched biological functions via GSEA, as previously described for *Gbx2* expression. We found that genes associated with both EES scores and increased *Gbx2* expression converged on related biological processes. Performing GSEA using genes ranked by mutual information with EES score revealed significant enrichment of the "Chondrocyte Differentiation" ontology (KS P=7.00x10^{-5}), along with four other categories also significantly enriched in the *Gbx2* expression analysis: "Hindlimb Morphogenesis" (KS P=1.42x10^{-3}), "Cartilage Development Involved in Endochondral Bone Morphogenesis" (KS

P=1.11x10^{-3}), "Collagen Fibril Organization" (KS P=1.40x10^{-5}), and "Positive Regulation of Phosphatidylinositol 3-kinase Signaling" (KS P=4.40x10^{-5}). This last category, along with collagen expression, is implicated in chondrocyte differentiation (Kita et al., 2008; Lefebvre et al., 1997; Fig. 5A, Table S12)

The convergence of both the *Gbx2*-centric and *Gbx2*-naïve approaches to identifying downstream effects of humanization on overlapping sets of genes implicated in chondrocyte differentiation (as shown in Fig. 5B) led us to examine the expression patterns of chondrocyte differentiation-related genes in humanized mesenchymal cells belonging to Louvain clusters 1-4 (Fig. 5C). We clustered *HACNS1$^{+/+}$* mesenchymal cells by EES score and *Gbx2* expression and examined the expression of the "Chondrocyte Differentiation" ontology genes within *Gbx2*-positive cells in order to visualize the expression of genes in the ontology compared to *Gbx2* (Fig. 5C). This analysis identified *Gbx2*-positive humanized cells that showed higher expression of positive regulators of chondrocyte differentiation (e.g. *Sox9*, *Col2a1*, *Bmp2*, and *Runx2*) compared to other mesenchymal cells, indicating that *HACNS1*-driven upregulation of *Gbx2* occurs in chondrogenic cells destined for digit formation (Fig. 5C; Lefebvre et al., 1997; Shea et al., 2003; Takigawa et al., 2010; Yoshida et al., 2004). We also identified a subset of *Gbx2*-positive humanized cells that showed expression of *Bmp4*, which is expressed in the apoptotic interdigital domains (Francis et al., 1994). These findings suggest that upregulation of *Gbx2* impacts the interdependent pathways of both digit condensation and interdigital cell fate specification required for digit morphogenesis.
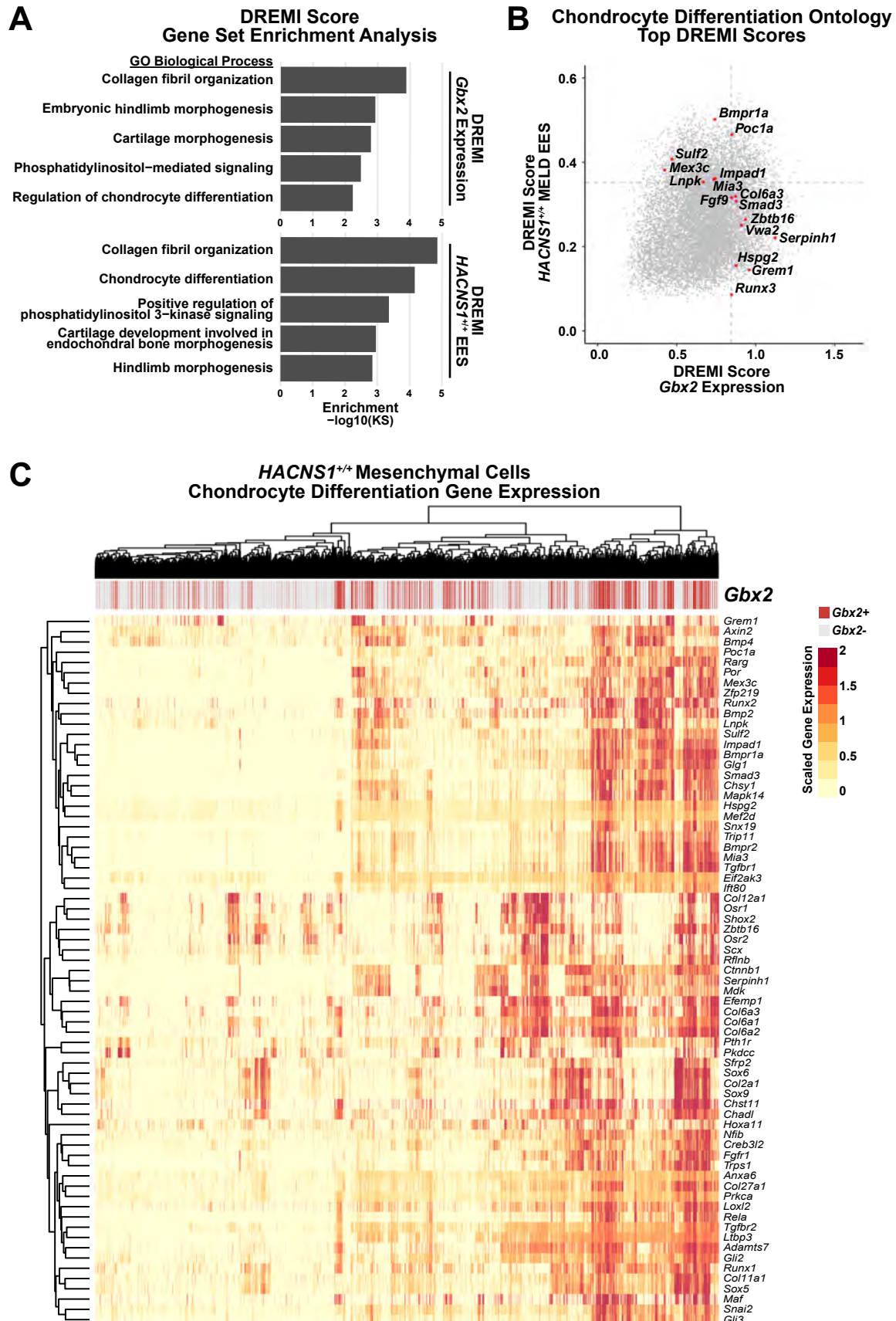
A DREMI Score
Gene Set Enrichment Analysis

B Chondrocyte Differentiation Ontology
Top DREMI Scores

C *HACNS1*⁺/⁺ Mesenchymal Cells
Chondrocyte Differentiation Gene Expression

Fig. 5. *Gbx2*-positive mesenchymal cells in *HACNS1*$^{+/+}$ limb bud express markers of chondrocyte differentiation. (A) Ontology enrichments of genes with expression associated with *Gbx2* expression (top) and the humanized state (EES score, bottom) in *HACNS1*$^{+/+}$ mesenchyme. 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 EES score ontology enrichments lists. See also tables S11 and S12. **(B)** EES and *Gbx2* kNN-DREMI scores 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 EES and *Gbx2* kNN-DREMI analysis gene lists are colored in red and labeled. **(C)** Heatmap showing expression of genes belonging to the ontology "Chondrocyte Differentiation" (GO:0002062) in all *HACNS1*$^{+/+}$ 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. The gene expression values shown are imputed using ALRA (Linderman et al., 2018) and centered and scaled using z-scores.

Humanized mouse limbs show no overt morphological changes

In order 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 at E18.5 for *HACNS1*$^{+/+}$ embryos, homozygous chimpanzee ortholog embryos, and wild type embryos (Materials and Methods). We found no gross morphological differences among genotypes. The three major limb segments (autopod, zeugopod, and stylopod) were present in *HACNS1*$^{+/+}$ and homozygous chimpanzee ortholog skeletons (Fig. 6, S5). To assess fine morphological differences, we examined digit length (normalized to body size based on ossified humerus length), and intradigital (phalange to metacarpal or metatarsal length) and interdigital ratios. Again, we found no significant differences in digit length or proportions between *HACNS1*$^{+/+}$, homozygous chimpanzee ortholog and wild type autopods (Fig. 6, S5, Tables S6-S8).

Fig. 6. Limb skeletal morphology is overtly similar in *HACNS1^{+/+}*, chimpanzee ortholog line, and wild type mice. Representative images of E18.5 forelimbs and hindlimbs of the indicated genotypes stained with Alizarin Red (violet; bone) and Alcian Blue (blue; cartilage). Forelimb and hindlimb autopod, zeugopod, and stylopod are shown on the left, and numbers indicate digit identity. High magnification images of forelimb and hindlimb autopods are shown on the right.

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 humanized mouse model. This *in vivo* approach enabled us to identify the spatial and temporal changes in developmental 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 its human-specific function in the *trans*-regulatory environment of the developing limb. Second, we found that *Gbx2* showed increased promoter activity in the *HACNS1*$^{+/+}$ limb bud, strongly supporting that it is regulated by *HACNS1*, and then demonstrated that *HACNS1* drives strong and highly localized 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 morphological differences across species (Britten and Davidson, 1971; King and Wilson, 1975; Wray, 2007). Third, we identified the specific cell types that show upregulation of *Gbx2* in the developing limb. Our single-cell transcriptome analyses captured the spectrum of embryonic limb cell types, enabling us to not only identify the specific subset of mesenchymal cells that express *Gbx2*, but also to place those cells in the developmental process of chondrogenesis. By both characterizing *Gbx2*-positive humanized cells and identifying humanized-prototypical cells without reference to *Gbx2* expression, we implicated changes in *Gbx2* regulation in chondrocyte differentiation, a process that serves as the basis for digit formation.

Our finding that *HACNS1*-driven *Gbx2* upregulation is implicated in chondrocyte development indicates a possible role for *HACNS1* in the modification of forelimb and hindlimb digit proportions on the human lineage. A primary distinguishing trait of the human forelimb is the greater relative length of the thumb versus the other digits (Aiello and Dean, 1990). In humans, this is due in part to the shortening of digits 2-5, which correlates with opposability and dexterity of the hand (Almécija et al., 2015). Differences between human and great ape feet include shorter digits with a non-opposable hallux, longer heel bones, and unique transverse and longitudinal arches defined by relative metatarsal-tarsal joint orientation in humans (Holowka

26

and Lieberman, 2018; Holowka et al., 2017). The changes in limb development that confer human-specific morphological outcomes therefore involve discrete modifications to the proportions and positioning of ancestral skeletal structures.

Our findings indicating a role for the *HACNS1-Gbx2* pathway in modifying digit chondrogenesis during human limb evolution serve as the foundation for identifying the broader network of developmental changes that drove the emergence of human-specific limb traits. Relative phalange shortening in the forelimb and hindlimb are implicated in human limb evolution, and the similarity in *Gbx2* expression patterns in humanized forelimb and hindlimb bud suggests that *Gbx2* upregulation could contribute to parallel morphological changes by influencing early chondrocyte differentiation and digit formation in both the upper and lower limbs (Aiello and Dean, 1990; Almécija et al., 2015). While the cellular mechanisms controlling relative segment growth within digits remain unclear, studies in vertebrates with differing limb morphologies have shown that patterns of developmental gene expression tend to diverge during the pre- and post-patterning phases of limb bud development (Saxena et al., 2017). The timing of *HACNS1*-driven upregulation of *Gbx2* at the nexus of the pre- and post-patterning phases in the humanized mouse limb is intriguing, given the plasticity of cell fate specification during this time period (Hiscock et al., 2017; Saxena et al., 2017). A shift in the chondrogenic trajectory favoring chondrocyte differentiation over proliferation in the presumptive phalanges, as suggested in this study, could therefore contribute to intradigital ratio modifications that evolved on the human lineage.

We have shown that humanized mouse models provide a viable and fruitful path for studying gene regulatory mechanisms in human evolution within the complete genomic, tissue-level, and developmental framework of a living mammalian system. The effect of genetic changes in any one enhancer is likely to be insufficient to replicate human-specific morphological changes entirely in an experimental model. However, changes in expression of single genes, particularly pleiotropic regulators such as *Gbx2*, can have large molecular effects on developmental pathways. The *HACNS1* humanized mouse model provides an entry point for understanding the larger developmental networks that changed during human limb evolution, of which *Gbx2* is a part. Identification of changes in chondrogenesis downstream of *HACNS1,* in conjunction with our understanding of regulatory activity in the developing human limb, can be used to identify related factors modifying human chondrogenesis. Using humanized mouse

models, additional molecular drivers contributing to human-specific phenotypes may be studied at once, either through intercrossing mouse lines with unlinked loci or by iterative editing of one locus, allowing us to expand our understanding of human limb evolution. Furthermore, humanized mouse models provide the means to interrogate the larger developmental networks involving additional HARs in order to understand their contributions to human-specific developmental features. Our study thus provides a novel framework for using humanized mouse models to link individual sequences that arose on the human lineage since species divergence to the unique traits that distinguish our species.
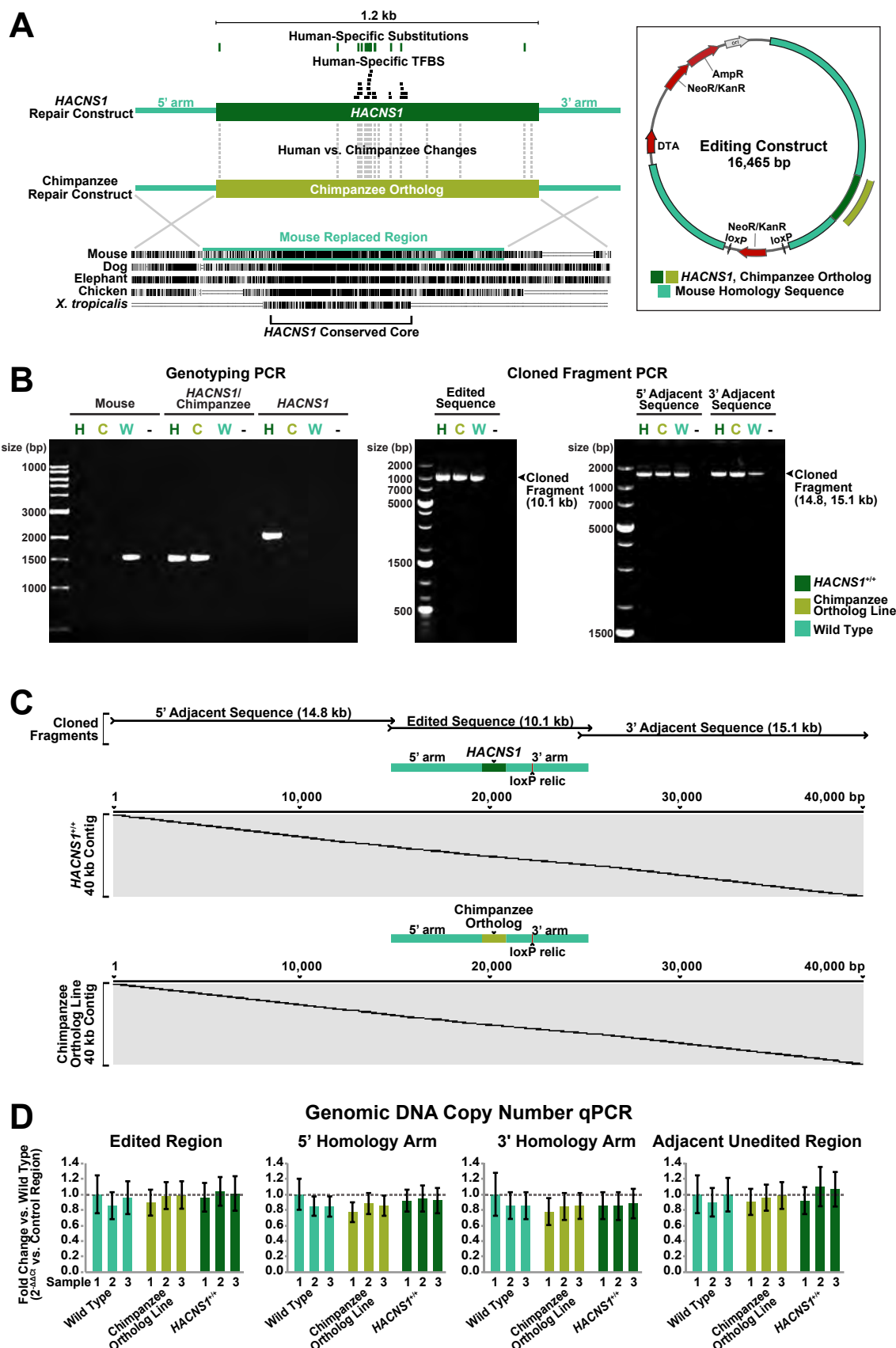
Fig. S1. Development and validation of the *HACNS1* and chimpanzee ortholog mouse models. (A) *Left: HACNS1* and chimpanzee ortholog line editing constructs are shown with the orthologous replaced region in mouse genome aligned to other vertebrate species derived from the 100-way Multiz alignment in the UCSC hg19 assembly. Non-polymorphic, fixed human-specific substitutions are shown above the human construct and all human versus chimpanzee sequence differences are shown below. The transcription factor binding sites (TFBS) unique to *HACNS1* versus the chimpanzee ortholog shown are based on predictions of JASPAR core mammal motifs in the human and chimpanzee genomes as described in Uebbing *et al.* 2019 (see also Table S13). *Right:* Embryonic stem cell editing construct showing antibiotic resistance (NeoR/KanR), diphtheria toxin (DTA), mouse homology, and location of human or chimpanzee sequences. **(B)** *Left:* PCR products generated with primers specific to the mouse, both *HACNS1* and chimpanzee, and *HACNS1* only orthologs were used for genotyping of $HACNS1^{+/+}$ (labeled as H), chimpanzee ortholog line (labeled as C), and wild type (labeled as W) mice from the F9 or later generation. *Middle:* PCR products were generated using primers outside the homology arms for Sanger sequencing of the edited locus. *Right:* PCR products were generated using primers anchored in the 5' homology arm and 14.8 kb upstream (5' adjacent sequence) and primers anchored in 3' homology arm and 15.1 kb downstream (3' adjacent sequence) for Sanger sequencing of the regions surrounding the editing locus. **(C)** Sanger sequencing contigs of cloned PCR products from (B), spanning the 40 kb region surrounding edited locus for $HACNS1^{+/+}$ (top) and the chimpanzee ortholog line (bottom). **(D)** $HACNS1^{+/+}$, chimpanzee ortholog line, and wild type genomic DNA qPCR for primers specific to the edited region, 5' homology arm, 3' homology arm, and adjacent unedited region. All Ct values are normalized to a region on chromosome 5. Three biological replicate samples are shown per genotype. Error bars denote standard deviation between technical replicates.
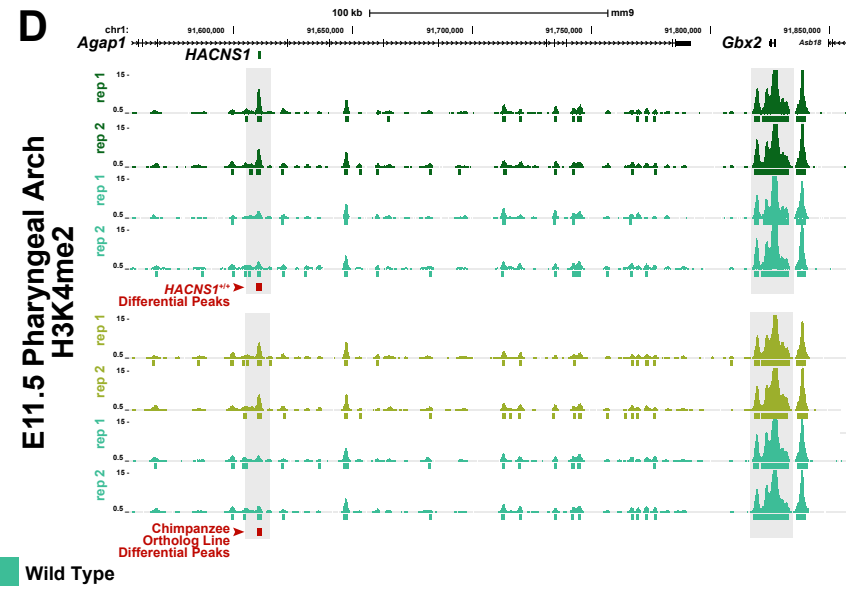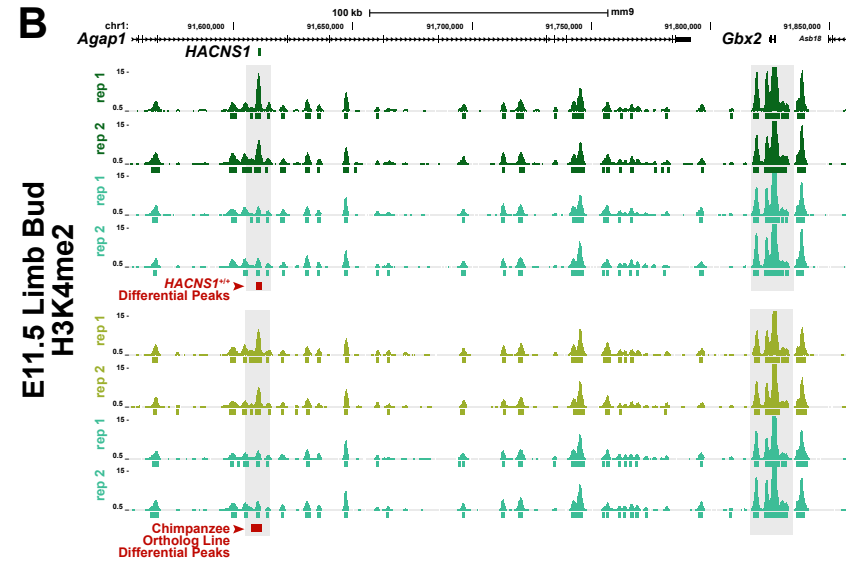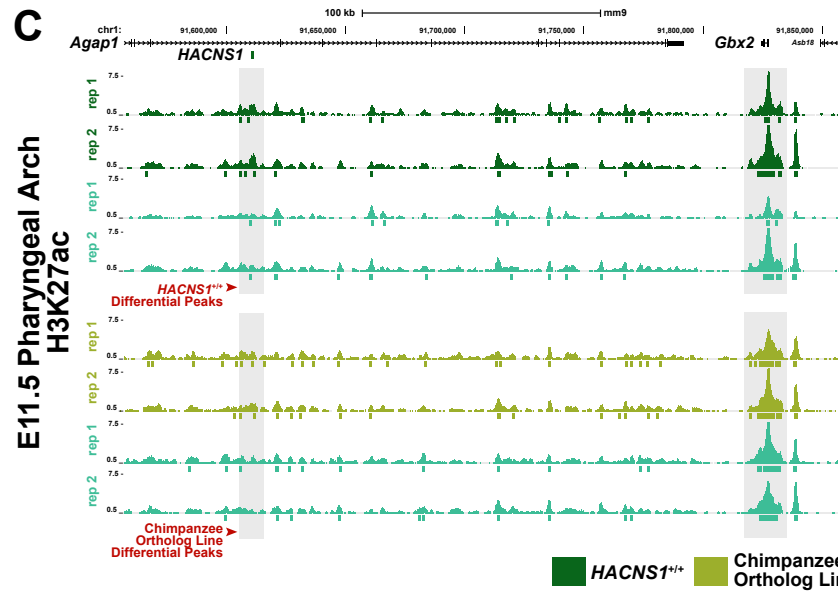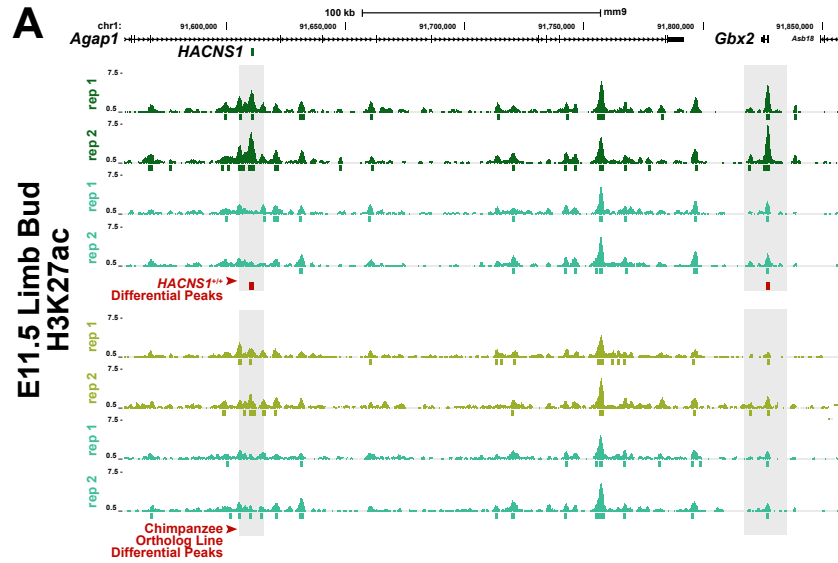
Fig. S2. H3K37ac and H3K4me2 ChIP-seq analyses in limb bud and pharyngeal arch. Normalized H3K27ac (left) and H3K4me2 (right) epigenetic signals in the region spanning the full *HACNS1-Gbx2* locus for two biological replicates per genotype for E11.5 limb bud (A, B) and pharyngeal arch (C, D). The location of the edited *HACNS1* locus relative to nearby genes is shown above each track group with a green bar directly below the corresponding UCSC mm9 genome track. *HACNS1* and *Gbx2* loci are highlighted in gray. All significant peaks are represented by genotype-specific colored bars below the signal tracks for *HACNS1*[+/+] (in dark green), chimpanzee ortholog line (in olive), and litter-matched wild type (in teal). Peak calls showing significant signal increases between *HACNS1*[+/+] and litter-matched wild type, or chimpanzee ortholog line and litter-matched wild type, are shown as red squares below each track group. Detailed differential peak information is available in Table S4.
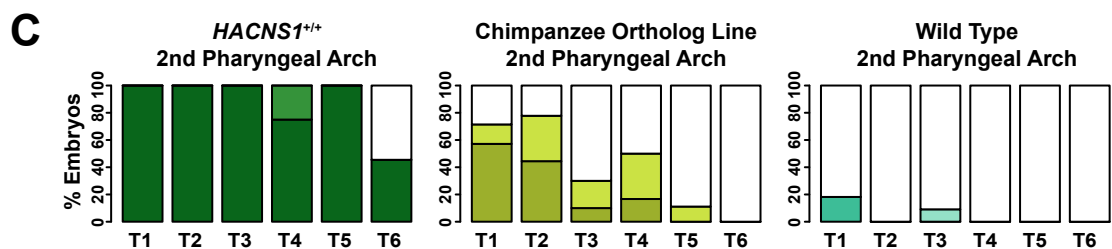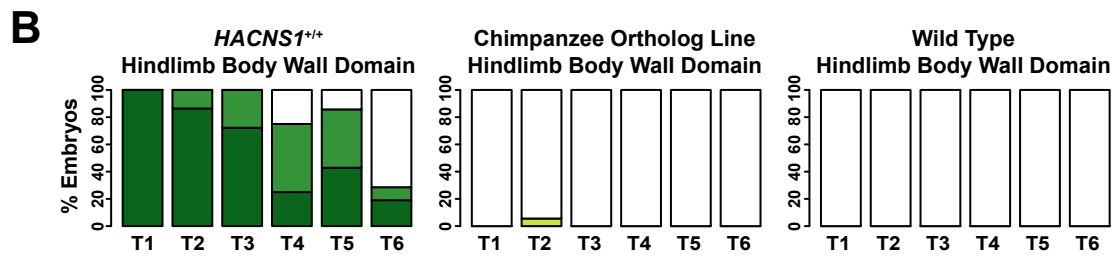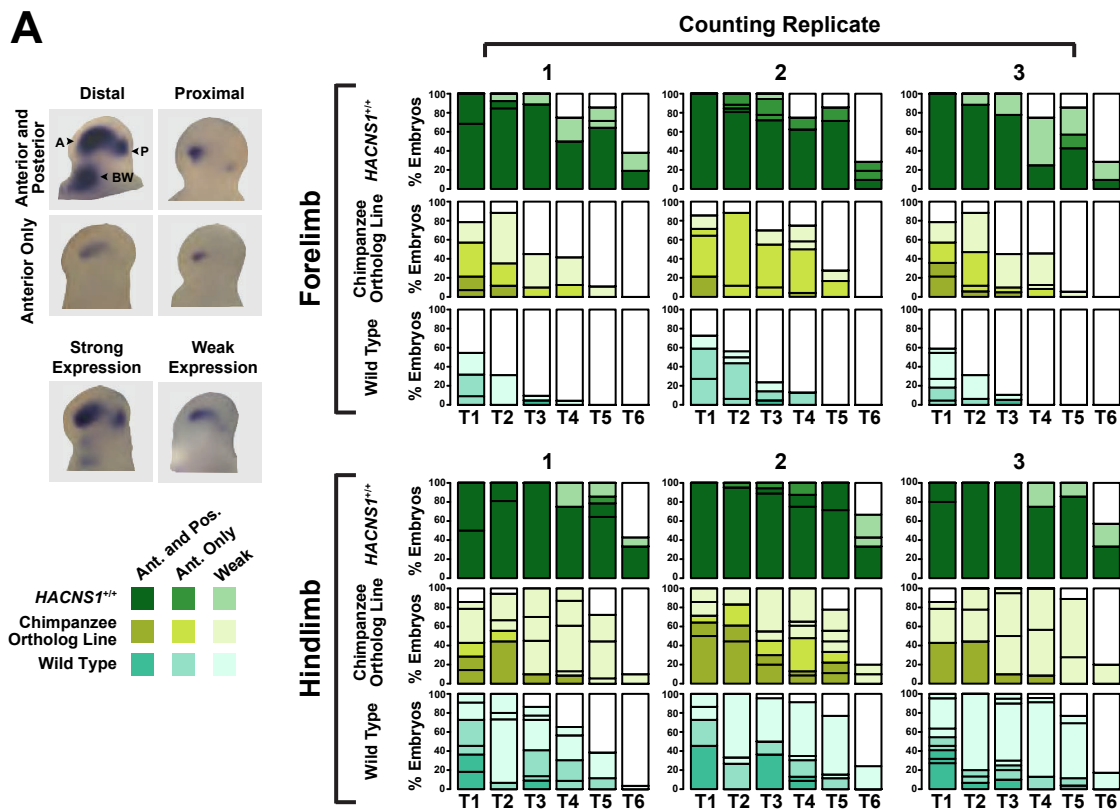
Fig. S3. Qualitative analysis of *Gbx2* expression patterns. (A) *Left:* representative images of anterior, posterior, proximal, distal (top), and strong versus weak staining patterns (bottom). Anterior (A), posterior (P), and body wall (BW) domains are denoted on top left limb bud. *Right: Gbx2* ISH 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). The darkest shade for $HACNS1^{+/+}$ (dark green), chimpanzee ortholog line (olive), and wild type (teal) represents percentage of embryos showing strong anterior and posterior limb bud staining. Medium-dark shade, as shown in legend on the left, denotes strong anterior staining only, while the lightest shade denotes weak staining in any domain. **(B)** Counting data for presence or absence of hindlimb bud body wall domain are shown for each genotype. Strong versus weak staining is denoted by darker versus lighter shade, as in (A). **(C)** Counting data for presence or absence of 2nd pharyngeal arch staining domain are shown for each genotype group. Strong versus weak staining is denoted by darker versus lighter shade, as in (A).
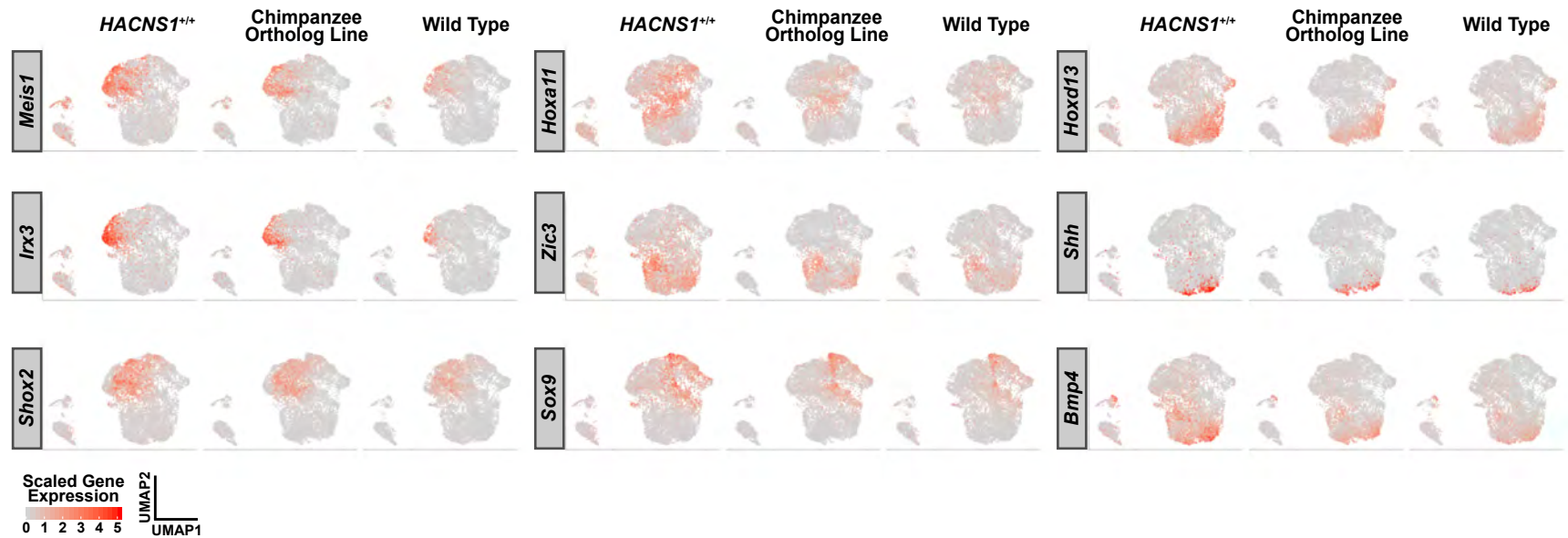
Fig. S4. Limb bud patterning marker expression in *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type E11.5 hindlimb bud. UMAP embedding of hindlimb bud cells from *HACNS1*$^{+/+}$, chimpanzee ortholog line, and wild type mice, showing conserved expression of representative proximal-distal (top row), anterior-posterior (middle row), and chondrogenesis-apoptosis marker genes (bottom row). All gene expression data are imputed using ALRA (Linderman et al., 2018) and centered and scaled using z-scores. See text and Fig. 4B for details on marker genes.
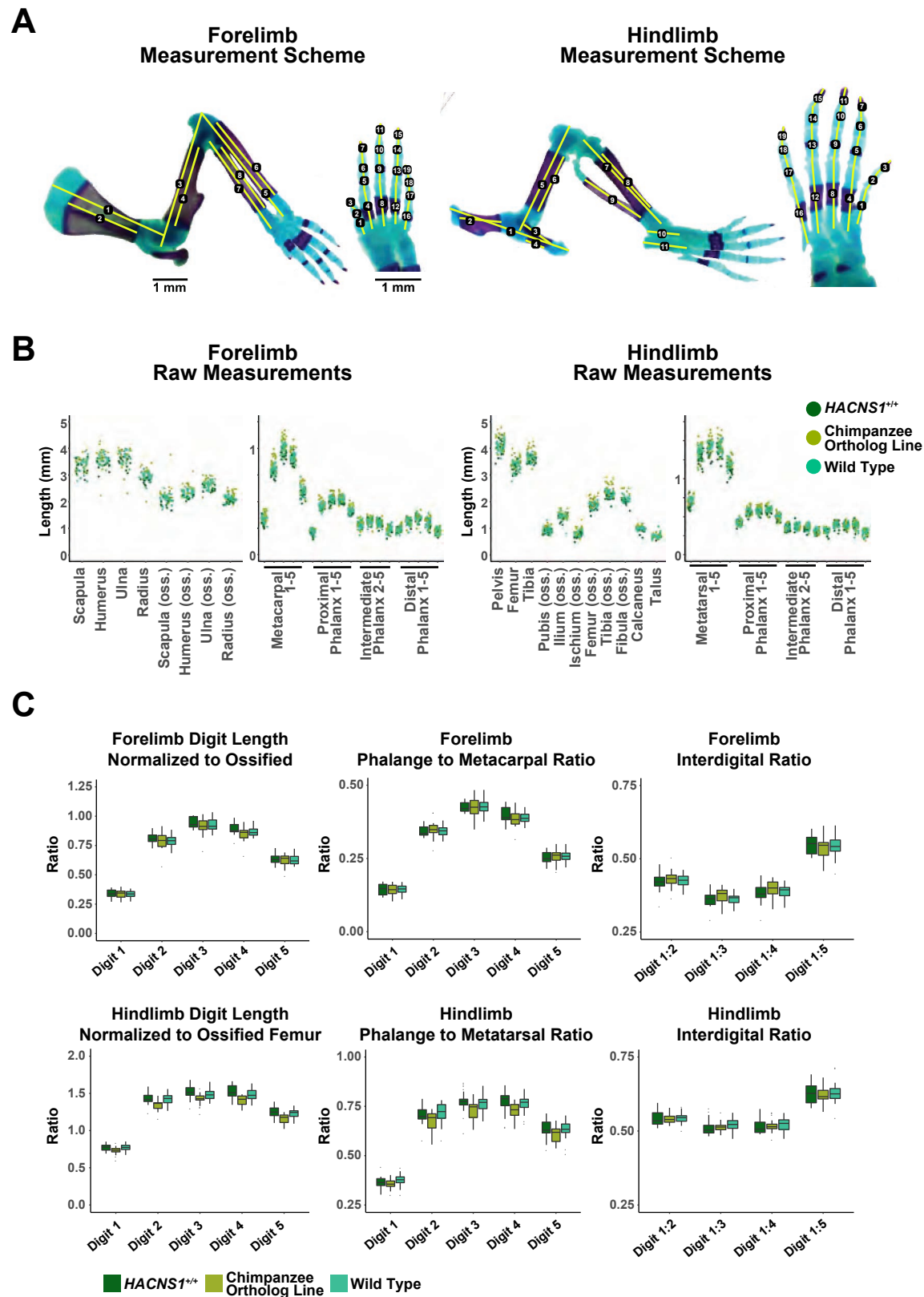
A Forelimb Measurement Scheme / Hindlimb Measurement Scheme

B Forelimb Raw Measurements / Hindlimb Raw Measurements

Legend: *HACNS1*⁺/⁺ ; Chimpanzee Ortholog Line ; Wild Type

C
- Forelimb Digit Length Normalized to Ossified
- Forelimb Phalange to Metacarpal Ratio
- Forelimb Interdigital Ratio
- Hindlimb Digit Length Normalized to Ossified Femur
- Hindlimb Phalange to Metatarsal Ratio
- Hindlimb Interdigital Ratio

Legend: *HACNS1*⁺/⁺ ; Chimpanzee Ortholog Line ; Wild Type

Fig. S5. Morphometric analyses of *HACNS1*[+/+], chimpanzee ortholog line, and wild type skeleton. (A) Measurement scheme for E18.5 skeleton morphometric analysis is shown with yellow lines denoting measured segments for forelimb zeugopod and stylopod cartilage and/or bone: scapula (1,2), humerus (3,4), ulna (5,6), radius (7,8); hindlimb zeugopod and stylopod cartilage and/or bone: pelvis (1), ilium (2), pubis (3), ischium (4), femur (5,6), tibia (7,8), fibula (9), talus (10), calcaneus (11), and metacarpal/metatarsal and phalange autopod segments (digit 1: 1-3; digit 2: 4-7; digit 3: 8-11; digit 4: 12-15; digit 5:16-19). **(B)** Raw data for all measured segments are plotted and colored by genotype. **(C)** Normalized digit length, phalange to metacarpal/ metatarsal ratio, and interdigital ratio for forelimb and hindlimb digits are shown by genotype. Digit length is calculated as sum of all metacarpal/metatarsal and phalange segments. Forelimb and hindlimb digit lengths are normalized to ossified humerus and femur length of the same sample digit length, respectively. For ANOVA analysis of morphometric data see Tables S6-S8.

37

Materials and Methods

Mouse Line Generation and Validation

The *HACNS1* and chimpanzee ortholog lines were generated at the Yale Genome Editing Center using standard techniques in modified ES cells (Behringer et al., 2014). C57BL6/J-A^{w-J}/J mouse ES cells (first described in Petersen et al., 2016) were edited by electroporation of a GFP cloning vector containing human (1,241 bp) or chimpanzee (1,240 bp) sequence flanked by C57BL/6J mouse sequence homology arms, floxed pPGKneo vector, and diphtheria toxin sequence (Fig. S1A). The genomic coordinates of the human (hg19) and mouse (mm9) sequences used in the editing constructs, including the mouse homology arm sequences, are listed in Table S1 (Rhead et al., 2009). G0 chimeras were backcrossed to wild type C57BL/6J and crossed with an actin-Cre C57BL/6J mouse line to remove vector sequence. All mice used in our analysis were from F9 or later generations. All animal work was performed in accordance with approved Yale IACUC protocols.

Genotyping primers specific to *HACNS1*, chimpanzee, and mouse orthologs are listed in Table S2. Cloning primers listed in Table S2 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 noonan.ycga.yale.edu/noonan_public/Dutrow_HACNS1/. Human-specific substitutions were defined as fixed in the human state if the derived allele frequency in dbSNP (v153) was >= 0.9999 and if the ancestral sequence state was conserved between chimpanzee, rhesus macaque, orangutan, and marmoset.

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 Table S2. 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 Table S2).

38

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*$^{+/+}$, 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 (Cotney et al., 2012). Chromatin for each genotype, tissue, and replicate was used for H3K27ac or H3K4me2 immunoprecipitation using Active Motif #39133 and Active Motif #39913 antibodies as previously described (Cotney and Noonan, 2015; Cotney et al., 2012). ChIP-qPCR was performed using Power SYBR Green Mastermix (Thermo Fisher Scientific #4368577) with primers listed in Table S3. Samples were sequenced (2×100 bp) using standard Illumina protocols on an Illumina HiSeq 4000. 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) (Langmead and Salzberg, 2012). 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 (Langmead and Salzberg, 2012). All peaks were called with HOMER (v4.9.1) using default settings for --histone (IP vs input fold change=4, p=0.0001, peak size=500, minDist=1000) (Heinz et al., 2010). 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 (Heinz et al., 2010; Love et al., 2014).

RNA Extraction and qRT-PCR

E11-E12 embryos were collected from six $HACNS1^{+/+}$, 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 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 Table S5 and Ct values of *Gbx2* were normalized to *Hprt1*.

Whole mount *In Situ* Hybridization

E11-E12 mouse embryos were collected from $HACNS1^{+/+}$ (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 (Louvi and Wassef, 2000; Louvi et al., 2007). The *Gbx2* probe used for hybridization contains the full mouse consensus CDS sequence (CCDS15150.1; NCBI CCDS Release 23). The 178 embryos were divided into sextiles based on crown-rump length and assessed for staining pattern by three individuals blinded to genotype under a stereo microscope (Leica S6D). Representative images were taken using a Zeiss Stemi stereomicroscope.

Single-Cell RNA-Sequencing

Sample Preparation

Tissue for scRNA-seq was collected from two each of $HACNS1^{+/+}$, chimpanzee ortholog line, and wild type litters from E11.5 homozygous crosses. 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 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. 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.

Filtering and Preprocessing

Matrices from the 10x Cell Ranger platform were filtered and preprocessed using Seurat v3.0.1 (Stuart et al., 2019). 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 or greater than 4 were removed, as well as those with total UMI count (nUMI) z-scores greater than 7. 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 (Stuart et al., 2019). 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 (Blondel et al., 2008;

Stuart et al., 2019). The numbers of cells remaining after pre-processing for each sample are listed in Table S9.

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 Table S9). Cell cycle scores were computed using CellCycleScoring in Seurat to regress out the difference between G2M and S phases in order to preserve only differences between cycling and non-cycling cells and regress out differences related to cell cycle amongst proliferating cells (Stuart et al., 2019). In addition to cell cycle scores, percent mitochondrial gene expression and nUMI values were regressed using SCTransform (SCT) in order to remove the effects of sequencing depth and minor differences in mitochondrial DNA expression related to viability (Hafemeister and Satija, 2019). All SCT normalized datasets containing all genes from individual samples were integrated using SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData (Hafemeister and Satija, 2019; Stuart et al., 2019).

Following integration, the combined dataset was randomly down-sampled to contain 10,000 cells per genotype prior to embedding and clustering using SubsetData in Seurat (Stuart et al., 2019). PCA and UMAP were implemented in Seurat using RunPCA, RunUMAP, FindNeighbors, and FindClusters (McInnes et al., 2018; Stuart et al., 2019). Normalized data from all samples combined was used for imputation using ALRA with default settings (Linderman et al., 2018). The threshold for identifying *Gbx2*-positive cells was set as an imputed *Gbx2* expression value greater than 0.1.

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). Imputation was performed with MAGIC using gt.Graph(data, n_pca=20, decay=40, knn=10),

magic.MAGIC(knn=G.knn, decay=G.decay), and magic_op.fit_transform(data, graph=G) (van Dijk et al., 2018). MELD was run on one-hot vectors for each genotype independently with the same graph signal as MAGIC using meld_op.fit_transform(RES.toarray(), G) (Burkhardt et al., 2019). kNN-DREMI was run on MAGIC-imputed data, and with MELD EES values corresponding to the humanized condition for humanized cells, with scprep using scprep.stats.knnDREMI (Krishnaswamy et al., 2014).

Gene Set Enrichment Analysis

GSEA was performed using topGO v.2.34.0 on all expressed genes that were ranked by *Gbx2*-DREMI or EES-DREMI score from the aforementioned humanized mesenchymal cell kNN-DREMI analysis (Alexa and Rahnenfuhrer, 2018). Significant nodes were identified using the Kolmogorov–Smirnov test and *elim* algorithm. Annotated genes were marked as significant if ranked in the top 20% of kNN-DREMI scores. Ontologies listed in Tables S11 and S12 are the top 30 nodes with fewer than 100 annotated genes (to remove non-specific categories) and at least one significant gene. Heatmap hierarchical clustering was performed using pheatmap v1.0.12 (Kolde, 2015).

Skeletal Staining

E18.5 skeletons from two litters from each of $HACNS1^{+/+}$, chimpanzee ortholog line, and wild type homozygous crosses (n=48 embryos) were stained with Alcian Blue and Alizarin Red as previously described (Behringer et al., 2014). Skeletons were imaged under a stereo microscope (Leica S6D). 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

43

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 are available at noonan.ycga.yale.edu/noonan_public/Dutrow_HACNS1/.

ANOVA Analysis for Gene Expression and Morphometry

ANOVA analysis was performed with the lme4 package in R using default parameters to dissect the effects of genotype on *Gbx2* expression (qRT-PCR data) and limb segment length (morphometric data) (Bates et al., 2015). For the qRT-PCR dataset, we employed an additive model of ΔCt to calculate the main effects of genotype and timepoint and a genotype:timepoint interaction term *(Ct_{Gbx2} − Ct_{Hprt1} ~ Genotype * Timepoint)*. For the morphometric datasets, 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 ~ Genotype * (1|Genotype/Litter) * Sex * Limb * Digit * (1|RL) * (1|Litter/Embryo) * (1|Sex/Embryo) * (1|Genotype/Embryo))*. For both datasets, multiple comparisons adjustment was performed using the Benjamini & Hochberg method (Benjamini and Hochberg, 1995).

Supplemental Tables

Table S1. Genomic Sequence Coordinates of Editing Construct Template DNA

Table S2. Oligonucleotides for Genotyping, Cloning, and Copy Number Analysis

Table S3. Oligonucleotides Used for ChIP-qPCR

Table S4. ChIP-seq Significant Replicating Differential Peaks

Table S5. Oligonucleotides Used for qRT-PCR

Table S6. Normalized Digit Length ANOVA

Table S7. Phalange to Metacarpal/Metatarsal Ratio ANOVA

Table S8. Interdigital Ratio ANOVA

Table S9. scRNA-seq Sample Summary

Table S10. scRNA-seq *Gbx2* Expression Summary

Table S11. *Gbx2* kNN-DREMI GSEA Results

Table S12. EES kNN-DREMI GSEA Results

Table S13. Predicted Transcription Factor Binding Site Changes

Data Availability

The Gene Expression Omnibus accession number for the data reported in this paper is GSE141471.

Acknowledgements

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, qRT-PCR, 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 experiment was performed by E.V.D. with input from K.Y. and D.B.B.. S.U. performed the ANOVA analyses. E.V.D. and J.P.N. wrote the manuscript with input from all authors.

References

Aiello, L., and Dean, C. (1990). An Introduction to Human Evolutionary Anatomy (Academic Press).

Alexa, A., and Rahnenführer, J. (2018). topGO: Enrichment Analysis for Gene Ontology. R package version 2.34.0.

Almécija, S., Smaers, J.B., and Jungers, W.L. (2015). The evolution of human and ape hand proportions. Nat. Commun. *6*, 7717.

Bates, D., Mächler, M., Bolker, B.M., and Walker, S.C. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. *67*, 1-48.

Behringer, R., Gertsenstein, M., Vintersten Nagy, K., and Nagy, A. (2014). Manipulating the Mouse Embryo: a Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B *57*, 289–300.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *2008*, P10008.

Boyd, J.L., Skove, S.L., Rouanet, J.P., Pilaz, L.J., Bepler, T., Gordân, R., Wray, G.A., and Silver, D.L. (2015). Human-chimpanzee differences in a *FZD8* enhancer alter cell-cycle dynamics in the developing neocortex. Curr. Biol. *25*, 772–779.

Britten, R.J., and Davidson, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. *46*, 111–138.

Bunn, H., and Forget, B. (1986). Hemoglobin: Molecular, Genetic and Clinical Aspects (Philadelphia: W.B. Saunders Co.).

Burkhardt, D.B., Stanley, J.S., Perdigoto, A.L., Gigante, S.A., Herold, K.C., Wolf, G., Giraldez, A., Dijk, D. van, and Krishnaswamy, S. (2019). Enhancing experimental signals in single-cell RNA-sequencing data using graph signal processing. bioRxiv, https://doi.org/10.1101/532846.

Byrd, N.A., and Meyers, E.N. (2005). Loss of *Gbx2* results in neural crest cell patterning and pharyngeal arch artery defects in the mouse embryo. Dev. Biol. *284*, 233–245.

Capra, J.A., Erwin, G.D., Mckinsey, G., Rubenstein, J.L.R., and Pollard, K.S. (2013). Many human accelerated regions are developmental enhancers. Philos. Trans. R. Soc. B Biol. Sci. *368*, 20130025.

Carroll, S.B. (2003). Genetics and the making of *Homo sapiens*. Nature *422*, 849–857.

Chatterjee, M., Li, K., Chen, L., Maisano, X., Guo, Q., Gan, L., and Li, J.Y.H. (2012). *Gbx2* regulates thalamocortical axon guidance by modifying the LIM and Robo codes. Dev. *139*, 4633–4643.

Civin, C.I., Strauss, L.C., Brovall, C., Fackler, M.J., Schwartz, J.F., and Shaper, J.H. (1984). 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.

Clement-Jones, M. (2000). The short stature homeobox gene *SHOX* is involved in skeletal abnormalities in Turner syndrome. Hum. Mol. Genet. *9*, 695–702.

Cotney, J.L., and Noonan, J.P. (2015). Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing. Cold Spring Harb. Protoc. *2015*, 191–199.

Cotney, J., Leng, J., Oh, S., DeMare, L.E., Reilly, S.K., Gerstein, M.B., and Noonan, J.P. (2012). Chromatin state signatures associated with tissue-specific gene expression and enhancer activity in the embryonic limb. Genome Res. *22*, 1069–1080.

Cotney, J., Leng, J., Yin, J., Reilly, S.K., Demare, L.E., Emera, D., Ayoub, A.E., Rakic, P., and Noonan, J.P. (2013). The evolution of lineage-specific regulatory activities in the human embryonic limb. Cell *154*, 185–196.

Deutsch, E., Johnson, S.A., and Seegers, W.H. (1955). Differentiation of certain platelet factors related to blood coagulation. Circ. Res. *3*, 110–115.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell *174*, 716-729.e27.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Favier, B. (1997). Developmental functions of mammalian Hox genes. Mol. Hum. Reprod. *3*, 115–131.

Francis, P.H., Richardson, M.K., Brickell, P.M., and Tickle, C. (1994). Bone morphogenetic proteins and a signalling pathway that controls patterning in the developing chick limb.

Development *120*, 209–218.

Gañan, Y., Macias, D., Duterque-Coquillaud, M., Ros, M.A., and Hurle, J.M. (1996). Role of TGFβs 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.

Geschwind, D.H., and Rakic, P. (2013). Cortical evolution: judge the brain by its cover. Neuron *80*, 633–647.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. bioRxiv, https://doi.org/10.1101/576827.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hevner, R.F., Miyashita-Lin, E., and Rubenstein, J.L.R. (2002). Cortical and thalamic axon pathfinding defects in *Tbr1*, *Gbx2*, and *Pax6* mutant mice: evidence that cortical and thalamic axons interact and guide each other. J. Comp. Neurol. *447*, 8–17.

Hiscock, T.W., Tschopp, P., and Tabin, C.J. (2017). On the formation of digits and joints during limb development. Dev. Cell *41*, 459–465.

Holowka, N.B., and Lieberman, D.E. (2018). Rethinking the evolution of the human foot: insights from experimental research. J. Exp. Biol. *221*, jeb174425.

Holowka, N.B., O'Neill, M.C., Thompson, N.E., and Demes, B. (2017). Chimpanzee and human midfoot motion during bipedal walking and the evolution of the longitudinal arch of the foot. J. Hum. Evol. *104*, 23–31.

Kamm, G.B., Pisciottano, F., Kliger, R., and Franchini, L.F. (2013). The developmental brain gene *NPAS3* contains the largest number of accelerated regulatory sequences in the human genome. Mol. Biol. Evol. *30*, 1088-1102.

King, M.C., and Wilson, A.C. (1975). Evolution at two levels in humans and chimpanzees. Science *188*, 107–116.

Kita, K., Kimura, T., Nakamura, N., Yoshikawa, H., and Nakano, T. (2008). PI3K/Akt signaling as a key regulatory pathway for chondrocyte terminal differentiation. Genes to Cells *13*, 839–850.

Kolde, R. (2019). pheatmap : Pretty Heatmaps. R package version 1.0.12.

Krishnaswamy, S., Spitzer, M.H., Mingueneau, M., Bendall, S.C., Litvin, O., Stone, E., Pe'er, D., and Nolan, G.P. (2014). Conditional density-based analysis of T cell signaling in single-cell data. Science *346*, 1250689.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lefebvre, V., Huang, W., Harley, V.R., Goodfellow, P.N., and de Crombrugghe, B. (1997). SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. Mol. Cell. Biol. *17*, 2336–2346.

Li, D., Sakuma, R., Vakili, N.A., Mo, R., Puviindran, V., Deimling, S., Zhang, X., Hopyan, S., and Hui, C. chung (2014). 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.

Li, J.Y.H., Lao, Z., and Joyner, A.L. (2005). New regulatory interactions and cellular responses in the isthmic organizer region revealed by altering Gbx2 expression. Development *132*, 1971–1981.

Lin, Z., Cantos, R., Patente, M., and Wu, D.K. (2005). Gbx2 is required for the morphogenesis of the mouse inner ear: A downstream candidate of hindbrain signaling. Development *132*, 2309–2318.

Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E., et al. (2011). A high-resolution map of human evolutionary constraint using 29 mammals. Nature *478*, 476-482.

Linderman, G.C., Zhao, J., and Kluger, Y. (2018). Zero-preserving imputation of scRNA-seq data using low-rank approximation. BioRxiv, https://doi.org/10.1101/397588.

Louvi, A., and Wassef, M. (2000). 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.

Louvi, A., Yoshida, M., and Grove, E.A. (2007). The derivatives of the Wnt3a lineage in the central nervous system. J. Comp. Neurol. *504*, 550–569.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*.

Martin, G.R. (1998). The roles of FGFs in the early development of vertebrate limbs. Genes Dev. *12*, 1571–1586.

Martin, P. (1990). Tissue patterning in the developing mouse limb. Int. J. Dev. Biol. *34*, 323–336.

McInnes, L., Healy, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. ArXiv e-prints 1802.03426.

Miyashita-Lin, E.M., Hevner, R., Wassarman, K.M., Martinez, S., and Rubenstein, J.L.R. (1999). Early neocortical regionalization in the absence of thalamic innervation. Science *285*, 906–909.

Miyazaki, H., Kobayashi, T., Nakamura, H., and Funahashi, J.I. (2006). Role of Gbx2 and Otx2 in the formation of cochlear ganglion and endolymphatic duct. Dev. Growth Differ. *48*, 429–438.

Petersen, M.C., Madiraju, A.K., Gassaway, B.M., Marcel, M., Nasiri, A.R., Butrico, G., Marcucci, M.J., Zhang, D., Abulizi, A., Zhang, X.M., et al. (2016). Insulin receptor Thr1160

phosphorylation mediates lipid-induced hepatic insulin resistance. J. Clin. Invest. *126*, 4361–4371.

Pollard, K.S., Salama, S.R., Lambert, N., Lambot, M.-A., Coppens, S., Pedersen, J.S., Katzman, S., King, B., Onodera, C., Siepel, A., et al. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature *443*, 167.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110-121.

Prabhakar, S., Noonan, J.P., Pääbo, S., and Rubin, E.M. (2006). Accelerated evolution of conserved noncoding sequences in humans. Science *314*, 786.

Prabhakar, S., Visel, A., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Morrison, H., FitzPatrick, D.R., Afzal, V., et al. (2008). Human-specific gain of function in a developmental enhancer. Science *321*, 1346–1350.

Quinn, M.E., Haaning, A., and Ware, S.M. (2012). Preaxial polydactyly caused by Gli3 haploinsufficiency is rescued by Zic3 loss of function in mice. Hum. Mol. Genet. *21*, 1888–1896.

Reilly, S.K., and Noonan, J.P. (2016). Evolution of Gene Regulation in Humans. Annu. Rev. Genomics Hum. Genet. *17*, 45-67.

Reilly, S.K., Yin, J., Ayoub, A.E., Emera, D., Leng, J., Cotney, J., Sarro, R., Rakic, P., and Noonan, J.P. (2015). Evolutionary changes in promoter and enhancer activity during human corticogenesis. Science *347*, 1155–1159.

Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J., et al. (2009). The UCSC genome browser database: Update 2010. Nucleic Acids Res. *38*.

Riddle, R.D., Johnson, R.L., Laufer, E., and Tabin, C. (1993). Sonic hedgehog mediates the polarizing activity of the ZPA. Cell *75*, 1401–1416.

Ryu, H., Inoue, F., Whalen, S., Williams, A., Kircher, M., Martin, B., Alvarado, B., Samee, M.A.H., Keough, K., Thomas, S., et al. (2018). Massively parallel dissection of human accelerated regions in human and chimpanzee neural progenitors. BioRxiv, https://doi.org/10.1101/256313.

Sassoon, D., Lyons, G., Wright, W.E., Lin, V., Lassar, A., Weintraub, H., and Buckingham, M. (1989). Expression of two myogenic regulatory factors myogenin and MyoDl during mouse embryogenesis. Nature *341*, 303–307.

Saxena, A., Towers, M., and Cooper, K.L. (2017). The origins, scaling and loss of tetrapod digits. Philos. Trans. R. Soc. B Biol. Sci. *372*.

Shea, C.M., Edgar, C.M., Einhorn, T.A., and Gerstenfeld, L.C. (2003). BMP treatment of C3H10T1/2 mesenchymal stem cells induces both chondrogenesis and osteogenesis. J. Cell. Biochem. *90*, 1112–1127.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y.,

Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888-1902.e21.

Tabin, C., and Wolpert, L. (2007). Rethinking the proximodistal axis of the vertebrate limb in the molecular era. Genes Dev. *21*, 1433–1442.

Takigawa, Y., Hata, K., Muramatsu, S., Amano, K., Ono, K., Wakabayashi, M., Matsuda, A., Takada, K., Nishimura, R., and Yoneda, T. (2010). The transcription factor Znf219 regulates chondrocyte differentiation by assembling a transcription factory with Sox9. J. Cell Sci. *123*, 3780–3788.

Tam, P.P.L. (1981). The control of somitogenesis in mouse embryos. J. Embryol. Exp. Morphol. *65*, 103–128.

Uebbing, S., Gockley, J., Reilly, S.K., Kocher, A.A., Geller, E., Gandotra, N., Scharfe, C., Cotney, J., and Noonan, J.P. (2019). Massively parallel discovery of human-specific substitutions that alter neurodevelopmental enhancer activity. BioRxiv, https://doi.org/10.1101/865519.

Wassarman, K.M., Lewandoski, M., Campbell, K., Joyner, A.L., Rubenstein, J.L.R., Martinez, S., and Martin, G.R. (1997). Specification of the anterior hindbrain and establishment of a normal mid/hindbrain organizer is dependent on Gbx2 gene function. Development *124*, 2923–2934.

Won, H., Huang, J., Opland, C.K., Hartl, C.L., and Geschwind, D.H. (2019). Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. Nat. Commun. *10*.

Wray, G.A. (2007). The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. *8*, 206–216.

Wright, E., Hargrave, M.R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A., and Koopman, P. (1995). The Sry-related gene Sox9 is expressed during chondrogenesis in mouse embryos. Nat. Genet. *9*, 15–20.

Yoshida, C.A., Yamamoto, H., Fujita, T., Furuichi, T., Ito, K., Inoue, K.I., Yamana, K., Zanma, A., Takada, K., Ito, Y., et al. (2004). Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. Genes Dev. *18*, 952–963.