A spatio-temporally constrained gene regulatory network directed by PBX1/2 acquires limb patterning specificity via HAND2

Marta Losa
University of California, San Francisco
Iros Barozzi
Medical University Vienna https://orcid.org/0000-0003-0690-3473
Marco Osterwalder
University of Bern
Peyman Zarrineh
University of Manchester
Jean Denis Benazet
University of California, San Francisco
Brandon Chacon
University of California, San Francisco
Ausra Girdziusaite
University of Basel
Angela Morabito
University of Basel
Jianjian Zhu
National Cancer Institute
Susan Mackem
https://orcid.org/0000-0002-9095-9914
Terence Capellini
Harvard University https://orcid.org/0000-0003-3842-8478
Nicoletta Bobola
University of Manchester
Diane Dickel
Lawrence Berkeley National Laboratory https://orcid.org/0000-0001-5497-6824
Aimée Zuniga
University of Basel https://orcid.org/0000-0002-9953-3637
Axel Visel
Lawrence Berkeley National Laboratory https://orcid.org/0000-0002-4130-7784
Rolf Zeller

University of Basel https://orcid.org/0000-0002-3186-7403
Licia Selleri (Licia.Selleri@ucsf.edu)
University of California, San Francisco

Article

Keywords:

Posted Date: March 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1460337/v1

A spatio-temporally constrained gene regulatory network directed by PBX1/2 acquires limb patterning specificity via HAND2

Marta Losa[^*], Iros Barozzi[%*], Marco Osterwalder[#xy], Peyman Zarrineh[{}], Jean Denis Benazet[^], Brandon Chacon[^], Ausra Girdziusaite[+], Angela Morabito[+], Jianjian Zhu[$], Susan Mackem[$], Terence D. Capellini[&], Nicoletta Bobola[{}], Diane Dickel[#], Aimee Zuniga[+], Axel Visel[#()z], Rolf Zeller[+[]] and Licia Selleri[^[]@]

[^]Program in Craniofacial Biology, Institute for Human Genetics, Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, Department of Orofacial Sciences and Department of Anatomy, University of California at San Francisco, San Francisco, California, USA; [%]Center for Cancer Research, Medical University of Vienna, Vienna, Austria; [#]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA; [x]Department for Biomedical Research (DBMR), University of Bern, Bern, Switzerland; [y]Department of Cardiology, Bern University Hospital, Bern, Switzerland; [{}]School of Medical Sciences, University of Manchester, Manchester, United Kingdom; [+]Developmental Genetics, Department Biomedicine, University of Basel, Basel, Switzerland; [$]Cancer and Developmental Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD, USA; [&]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA; Broad Institute of Harvard and MIT, Cambridge, MA, USA; [()]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; [z]School of Natural Sciences, University of California, Merced, Merced, CA 95343, USA.

[*] Equal contributions

[[]] Co-senior authors

[@] Corresponding Author

ABSTRACT

A lingering question in developmental biology has centered on how PBX TALE transcription factors, viewed primarily as HOX co-factors, can confer functional specificity to spatially-restricted HOX proteins despite being ubiquitously distributed in vertebrate embryos. Here, using the murine hindlimb as a model, we investigate the elusive mechanisms whereby PBX homeoproteins themselves attain tissue-specific developmental functions. We first demonstrate that mesenchymal-specific loss of PBX1/2 or the transcriptional regulator HAND2 generates similar limb phenotypes. By combining tissue-specific and temporally-controlled mutagenesis with multi-omics approaches, we then reconstruct a GRN at organismal-level resolution that is collaboratively directed by PBX1/2 and HAND2 within subsets of posterior-proximal hindlimb mesenchymal cells. Genome-wide profiling of PBX1 binding across multiple embryonic tissues further reveals that HAND2 selects a subset of PBX-bound regions to impart limb patterning functionality. Our research elucidates fundamental principles by which promiscuous transcription factors cooperate with select cofactors to instruct distinct developmental programs.

INTRODUCTION

Through the years, the limb has provided an excellent model for understanding the genetic principles that control development, evolution, and congenital disease[1,2]. Numerous genetic studies in the mouse have led to significant insights into the genes and molecular pathways that direct limb bud patterning and morphogenesis. Our research has established that the gene family encoding PBX1/2/3 homeodomain transcription factors of the TALE superclass are essential regulators of limb development[3,4]. PBX homeoproteins execute hierarchical, overlapping, and iterative functions during limb development; they are essential in determining limb bud positioning and formation, limb axes establishment, as well as patterning and morphogenesis of most skeletal elements of the limb and girdle[4–7]. In particular, PBX homeoproteins control effectors of limb and girdle development, including genes encoding the SHH morphogen in the limb posterior mesenchyme[6] and transcriptional regulators such as ALX1 in pre-scapular domains[8].

PBX TALE transcription factors have long been regarded primarily as cofactors that promote the DNA-binding specificity of HOX proteins, which are notoriously known to have very low binding specificity[3]. However, it remains challenging to envision how PBX transcription factors with widespread distribution in the vertebrate embryo can confer functional specificity to HOX proteins that display domain-restricted localization. We previously demonstrated that PBX homeoproteins do not act solely as cofactors for HOX proteins in limb buds[3], but indeed execute upstream control of 5' *HoxA/D* gene expression[6]. In particular, homozygous loss of *Pbx1* (*Pbx1*$^{-/-}$) causes malformations of girdles and proximal stylopod skeletal structures[9], while homozygous inactivation of *Pbx2* or *Pbx3* does not yield limb skeletal phenotypes[10,11]. In contrast, compound constitutive loss-of-function of *Pbx1/2* results in multiple developmental abnormalities, including striking distal limb defects with loss of posterior digits[6], and exacerbates the proximal limb phenotypes reported in *Pbx1*$^{-/-}$ embryos[9]. Given that the mouse hindlimb bud lacks detectable *Pbx3* expression[12] and that *Pbx4*, the last known member of the *Pbx* family, is not expressed during limb bud development[13], compound loss of *Pbx1/2* in the developing hindlimb achieves a PBX-null state. Accordingly, *Pbx1/2* mutant hindlimbs exhibit more pronounced phenotypes than those observed in forelimbs, which express high levels of *Pbx3*[12]. However, compound constitutive homozygous deletion of *Pbx1* and *Pbx2* results in embryonic lethality by gestational day (E)10.5[6] due to striking cardiovascular defects[14], which prevents a mechanistic understanding of their key functions as transcriptional regulators of limb bud development.

The establishment of *cis*-regulatory landscapes and gene regulatory networks (GRNs) that control limb patterning and outgrowth has provided an additional layer to our understanding of this archetypical

developmental process[15]. For example, dissection of the chromatin organization and regulation of the murine *Hox* gene clusters[16,17] has contributed mechanistic insight into the critical roles of *Hox* genes in setting up regional identities along the body axis of bilaterians and during limb bud patterning[18]. Specifically, expression of all four *Hox9* paralogs is required for anterior-posterior (AP) polarization of the limb bud and initiation of SHH signaling[19]. Of note, our studies have shown that establishment of *Shh* expression in the limb zone of polarizing activity (ZPA) requires PBX1/2[6] and the bHLH transcription factor HAND2[20]. In particular, *Hand2* is part of a GRN that antagonizes *Gli3* expression to establish AP asymmetry and pentadactyly[20,21]. Establishment of the posterior domain of *Hand2* expression is not only dependent on its mutual antagonistic interaction with *Gli3* and early-wave *Hox* genes, but also on HOX-interacting MEIS1/2 transcription factors that bind to a *Hand2* limb enhancer[22]. Identification of HAND2 targets has uncovered a GRN of regulators that control compartmentalization of the limb bud mesenchyme, emphasizing an evolutionary conserved role of HAND2 in orchestrating early limb development upstream of SHH[23]. Lastly, the dissection of the *Gremlin1* (*Grem1*) *cis*-regulatory landscape, which directs the spatio-temporal expression of *Grem1* in limb buds, has provided insight into the robustness and evolutionary plasticity of distal limb and digit development[24].

While it was established that both PBX1/2 and HAND2 independently activate posterior *Shh* expression by directly interacting with its distal limb enhancer (called ZRS)[6,20,25], it is unknown whether their functions converge on regulating SHH signaling. If this is the case, potential genetic interactions between PBX1/2 and HAND2 and GRNs jointly controlled by these factors remain elusive. Here, we combined tissue-specific and temporally controlled gene inactivation in mouse embryonic hindlimb buds with a variety of transgenic and multi-omics approaches using dissected limb buds to reconstruct a multi-layered GRN of limb regulators that is collaboratively directed by PBX1/2 and HAND2. We further established that the genetic interaction of *Pbx1* with *Hand2* in limb mesenchyme is required for normal hindlimb bud development and that PBX1 acts also as an upstream regulator of *Hand2* expression. Lastly, by comparing PBX genome-wide occupancy *in vivo* across multiple embryonic tissues, we uncovered that PBX transcription factors indiscriminately occupy a vast pool of common genomic loci during development. However, PBX binding gains restrained functionality via cooperative interactions with HAND2, a critical regulator of limb AP asymmetry and pentadactyly, thus promoting early limb patterning specificity. Together, these studies enabled the reconstruction of a spatio-temporally constrained GRN at superior organismal- and tissue-level resolution. Broadly, this research establishes how promiscuous transcription factors attain context-specific developmental functions via cooperative interactions with select cofactors that enable tissue specificity during vertebrate organogenesis.

RESULTS

Requirement of PBX1 and PBX2 in limb bud mesenchyme during the onset of limb bud patterning and outgrowth

Given the dominant role of PBX1 among PBX family members during hindlimb development[6], we examined its spatio-temporal distribution from formation of the hindlimb bud (E9.0) to digit development (E12.0). Following its expression in the lateral plate mesoderm[6] (E8.0), PBX1 was detected in most limb mesenchymal progenitors and in the apical ectodermal ridge (AER) during the onset of hindlimb bud development (E9.0-10.5, Fig.1A-C'), while it was restricted to the proximal-most mesenchyme by E11.0 until E12.0 (Fig. S1A-C). To circumvent early embryonic lethality of *Pbx1/2* compound constitutive null embryos and to decipher tissue-specific *Pbx* functions during limb development, we conditionally inactivated *Pbx1*[26] on a *Pbx2*-deficient background[10]. This genetic approach enabled us to obtain a *Pbx*-null state in the hindlimb bud, where *Pbx3* is not expressed at detectable levels[12]. AER-specific deletion of *Pbx1* using a *Msx2Cre* deleter line[27,28] did not cause any limb skeletal abnormalities (Fig. S1D-G, Supplementary Table 1) and was verified by immunohistochemistry (Fig. S1H-I'). These genetic experiments demonstrate that PBX1 is dispensable for AER formation and maintenance.

To investigate PBX functions during early limb bud development, the *Hoxb6Cre* deleter line expressing the *Cre* recombinase under the control of the *Hoxb6* promoter[29] was used for conditional gene inactivation in the lateral plate mesoderm, posterior forelimb and entire hindlimb bud mesenchyme from limb initiation onward. Mesenchyme-specific deletion of *Pbx1* on a *Pbx2*-deficient background resulted in drastic limb defects (Fig. 1D-H) that phenocopy the limb skeletal abnormalities of constitutive compound *Pbx1/2* mutants[6]. Although the phenotypes of $Pbx1^{f/f};Pbx2^{+/-};Hoxb6^{Cre/+}$ (hereafter $Pbx1cKO^{Mes};Pbx2^{+/-}$) (Fig. 1E,F) and $Pbx1^{f/f};Pbx2^{-/-};Hoxb6^{Cre/+}$ (hereafter $Pbx1cKO^{Mes};Pbx2^{-/-}$) hindlimbs (Fig. 1G,H) displayed variable penetrance, the pelvic girdle was dysmorphic and the proximal as well as distal elements were malformed, hypoplastic, or absent in all mutant embryos at E14.5 (Supplementary Table 1). As expected, phenotypes were more severe when both *Pbx1* and *Pbx2* were inactivated, varying from hypoplastic autopodia (Fig. 1G') to complete autopod agenesis (Fig. 1H'). To assess the temporal requirement of *Pbx1* during hindlimb bud development, we used an inducible *Hoxb6Cre* ($Hoxb6^{CreERT}$) line[30]. Tamoxifen injections at E8.5, E9.5, E10.0 and E10.5, respectively, showed that *Pbx1/2* are required in the mesenchyme for hindlimb patterning prior to E9.5 (Fig. 1I,J and Fig. S1J), as their inactivation at or before this time-point resulted in hindlimb abnormalities that mimicked those observed using the *Hoxb6Cre* allele. In contrast, hindlimbs developed normally when tamoxifen was administered at or after E10.0. Altogether, concomitant inactivation of *Pbx1/2* with different *Cre* deleter lines (Fig. 1D-J, S1D-G and S1J) demonstrates that PBX1/2 are dispensable in the AER and are

required in the hindlimb mesenchyme at least until E9.5, i.e. during hindlimb bud initiation and early patterning.

Early mesenchymal-specific deletion of *Pbx1/Pbx2* or *Hand2* results in similar skeletal, cellular, and molecular alterations

The hindlimb skeletal phenotypes of E14.5 *Pbx1cKO^{Mes};Pbx2$^{-/-}$* mouse embryos are similar to those reported in embryos with mesenchymal loss of *Hand2* (hereafter *Hand2cKO^{Mes}*)[20,23] (Fig. 1K,L) and constitutive loss of *Shh* (*Shh$^{-/-}$*)[31]. Given the striking similarities of the distal phenotypes in these mutant limbs, we sought to investigate the cellular and molecular alterations underlying the respective skeletal defects. Limb bud mesenchymal apoptosis[32] was not increased in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* hindlimb buds in comparison to controls at early stages (34 somites, Fig. 1M). However, later in development, apoptosis of mesenchymal cells was significantly increased in the distal-anterior hindlimb bud mesenchyme of *Pbx1cKO^{Mes};Pbx2$^{-/-}$* embryos (43 and 49 somites, Fig. 1M). These findings were consistent with increased mesenchymal apoptosis in *Shh*- and *Hand2*-loss-of-function limb buds[20,31,33]. Interestingly, *Pbx1/2*-deficient hindlimb buds expressed negligible levels of *Shh*[6]. RNA *in situ* hybridization corroborated complete loss of *Shh* also in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* hindlimbs (Fig. 1N,O), as was the case in *Hand2cKO^{Mes}* hindlimb buds[20,23]. These results demonstrate that the striking similarities observed in the distal limb skeletal phenotypes of *Pbx1cKO^{Mes};Pbx2$^{-/-}$* and *Hand2cKO^{Mes}* embryos are underpinned by similar cellular and molecular alterations.

Given these similarities, we asked whether *Pbx1/2* and *Hand2* converge in orchestrating a shared GRN essential for limb bud patterning and morphogenesis beyond their independent requirements in *Shh* regulation. First, to assess a potential epistatic hierarchy, we examined the spatial expression of *Hand2* transcripts in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* hindlimb buds. Compared to controls, *Hand2* transcript levels were reduced in the proximal-most mesenchyme of *Pbx1cKO^{Mes};Pbx2$^{-/-}$* developing hindlimbs (Fig. 1P,Q), as reported for *Pbx1/2* constitutive mutants[6]. In contrast, *Pbx1* spatial expression was not altered in *Hand2cKO^{Mes}* hindlimb buds (Fig. 1R,S). Altogether, these findings corroborate previous reports that both PBX1/2 and HAND2 are essential to establish *Shh* expression in the posterior limb bud mesenchyme. They further suggest that *Pbx1* contributes to the regulation of *Hand2* expression, likely within a GRN that orchestrates the onset of limb bud patterning and outgrowth.

***Pbx1/2* and *Hand2* are co-expressed in a subset of posterior-proximal hindlimb bud mesenchymal cells**

To investigate the genetic hierarchy, individual contributions, and convergence of PBX and HAND2 in

regulating downstream limb GRNs, we combined mouse genetics with genome-wide transcriptome and epigenome profiling of dissected hindlimb buds (Fig. 2A). We isolated hindlimb buds from wild-type mouse embryos at E10.5 (36-38 somites) and performed single cell (sc) RNAseq analysis[34] to disentangle the cellular heterogeneity of early hindlimb development and assess potential co-expression of *Pbx1/2* and *Hand2* transcripts at single-cell resolution (Fig. 2B-F and Fig. S2,S3). Unsupervised clustering and limb bud marker gene analysis led to the identification of six main cell populations within E10.5 hindlimbs, including mesenchymal cells (*Prrx1*+ and *Meis2*+); epithelial cells and AER (*Epcam*+ and *Wnt6*+); erythrocytes (*Hba-a2*+ and *Klf1*+); endothelial cells (*Pecam1*+ and *Emcn*+); phagocytic cells (*Fcer1g*+ and *Spi1*+); and pluripotent progenitor cells (*Sox2*+ and *Pou5f1*+) (Fig. 2B,C, Fig. S2A,B, and S3A-F)[35,36]. *Pbx1* and *Pbx2* were detected in both the mesenchymal and epithelial cell populations, although their expression was predominant in mesenchymal cells, with *Pbx1* being overall more abundant than *Pbx2* (Fig. 2D and Fig. S2C). In contrast, *Hand2* expression was restricted to a subset of posterior mesenchymal cells (Fig. 2E)[20,23]. Additional analyses confirmed that *Hand2* is co-expressed with *Pbx1* and *Pbx2* in a fraction of mesenchymal cells (orange and red, Fig. 2F). These mesenchymal subpopulations were characterized by high levels of *Tbx3* and *Isl1* and low levels of *Lhx9* and *Lhx2* transcripts, which identify them as posterior-proximal hindlimb bud mesenchyme (Fig. S2B and S3G). In agreement, these subpopulations lacked anterior-proximal markers such as *Irx3* and *Irx5* (Fig. S3G). The expression of *HoxA* and *HoxD* genes further confirmed the spatial assignment of these cells to the posterior-proximal mesenchyme (Fig. S3G). In addition, immunofluorescence (IF) analysis of hindlimb bud sections revealed co-localization of PBX1 and HAND2 proteins in the posterior mesenchyme (Fig. 2G-J''). PBX1 was present in most mesenchymal cell nuclei, although at lower levels in the distal hindlimb mesenchyme, while HAND2 was restricted to the posterior-proximal mesenchyme. These results show that nuclear PBX1 and HAND2 transcription factors co-localize in the same posterior-proximal mesenchymal cells in hindlimb buds.

PBX and HAND2 regulate a GRN that controls early hindlimb bud development

To identify unique and shared transcriptional targets of PBX and HAND2 and their epigenetic landscapes in mouse hindlimb buds, we used chromatin immunoprecipitation followed by sequencing (ChIPseq)[37,38], in combination with ATACseq (genome-wide assay for transposase-accessible chromatin using sequencing)[39] in E10.5 hindlimb buds (middle panel, Fig. 2A). From all datasets obtained, only the peaks detected reproducibly across two replicates were selected for further analysis (Supplementary Table 2; see Methods). Replicated peaks for PBX1 and HAND2 were merged into one set of 32,691 putative regulatory elements: 6,157 bound by both transcription factors; 4,536 bound only by HAND2; and 21,998 bound only by PBX1 (Figure 3A,B and S4A-C). The majority of HAND2 peaks

were located distal to the transcription start site (TSS)[23] of annotated genes, while PBX1 peaks were distributed between promoters and TSS-distal regions (Fig. 3A,B). Regions co-bound by both PBX1 and HAND2 predominantly associated with distal elements, including both intergenic and intragenic regions (Fig. 3A). Remarkably, sites co-bound by PBX1 and HAND2 were located in regions of higher chromatin accessibility (Fig. 3C; *p*-value < 2.2e-16, Kruskal-Wallis Test) and had greater enrichment values for transcription factor-binding (Fig. 3D,E; *p*-value < 2.2e-16, Mann-Whitney Tests) than sites bound by only one of the two factors. Genomic Regions Enrichment Annotation (GREAT)[40] analysis revealed that the peaks bound by both PBX1 and HAND2 were mostly associated with genes known to regulate limb bud and/or skeletal development (Fig. 3F,G), while this was not the case for regions bound only by one transcription factor (Fig. S4D-G).

De novo motif discovery (using HOMER)[41] determined whether both PBX1 and HAND2, or only one of them, interacted directly with DNA or as part of transcriptional complexes (Fig. 3H). The HAND2 (annotated by similarity to HAND1::TCF3) and HOX-PBX (TGATTNTT; annotated to HOXC9) motifs were identified as the top two most-enriched motifs in shared peaks, whether they were centered on the PBX1 peak or the HAND2 peak (Fig. 3H; p-value < 0.05). Interestingly, the HOX-PBX motif was still the second most enriched sequence motif in peaks bound by HAND2 even in the absence of PBX1 binding. In contrast, PBX1 bound on its own with other TALE factors (MEIS/PREP) at promoter regions and distal elements (Fig. 3H; *p*-value < 0.05). Of note, the majority of genes associated with PBX1-specific peaks function in developmental processes other than limb morphogenesis (Fig. S4F). Moreover, motif analysis using a large set of published binding preferences[41] revealed significant associations of the PBX1-only peaks with binding of other homeobox transcription factors (HOX, MEIS, PDX2, LHX2) at TSS-distal sites (Fig. S4H; *p*-value < 1e-5; HOMER). These results point to the convergence of PBX1 and HAND2 on *cis*-regulatory modules within a GRN in the posterior-proximal limb mesenchyme that orchestrates limb and digit patterning (asterisks, Fig. 3F,G).

Next, we integrated our sets of PBX1 and HAND2 replicated peaks with the H3K27ac (associated with active enhancer marks) and H3K27me3 (associated with repressive enhancer marks) ChIPseqs[42] and the ATACseq profiles (denoting open chromatin) that we generated from E10.5 hindlimb buds. In addition, we intersected these data sets with published CTCF binding profiles (Fig. 3I)[43]. Our analyses revealed that: 1) Intergenic, intragenic and promoter regions co-bound by PBX1 and HAND2 associate with accessible chromatin and H3K27ac enrichment (Fig. 3I',I''); 2) regions bound by PBX1 only are significantly more accessible and more enriched with H3K27ac than regions bound by HAND2 only (Fig. 3I and Fig. S5A-D); 3) no significant associations are present with genomic regions marked by

H3K27me3 repressive marks, with the exception of a fraction of promoter regions bound by PBX1 (red arrow, Fig. S5D), which points to potential bivalent promoters; and 4) as expected, CTCF binding is not associated with any of the groups analyzed. Together, these analyses indicate that PBX1 and HAND2 interact with a common set of candidate *cis*-regulatory modules (CRMs) in accessible and active chromatin within genomic landscapes of genes that regulate limb development.

PBX1 directly regulates *Hand2* expression via specific CRMs active in the hindlimb bud mesenchyme

In light of the genetic evidence indicating that *Pbx1/2* regulate *Hand2* expression in the proximal-posterior limb bud mesenchyme (Fig. 1P,Q and Fig. 2D-F), we set out to identify the complement of limb enhancers critical for this process. To this end, we screened the *Hand2* TAD[44] by intersecting our PBX1, HAND2, H3K27ac ChIPseq and ATACseq datasets (Fig. 4A) (see Methods). This analysis uncovered 17 PBX1-bound putative CRMs with predicted limb enhancer activity, including 3 previously identified candidate limb enhancers (mm1687, mm1688, and mm1689) as validated by transgenic *LacZ* reporter assays at E11.5[45] (Fig. 4B,C,E, Supplementary Table 3). A similar strategy was used to analyze the remaining putative PBX1-bound limb enhancer elements (n=14) in mouse embryos at E10.5 or E11.5. These assays revealed the presence of 7 additional tissue-specific enhancers in the *Hand2* TAD with activity in limb buds and/or other embryonic compartments (Fig. 4D,F,G and Fig. S6; Supplementary Table 3). Of these, 3 showed reproducible limb activities in *Hand2* expressing domains, expanding the number of known *bona fide* embryonic limb enhancers in the *Hand2* locus to 6 in total, all of which were active at E10.5 and E11.5 (Fig. 4B-G and Fig. S6). However, only 2 of these enhancers, located 337kb (mm1689) and 226kb (mm1828) upstream of the *Hand2* TSS, respectively, displayed activity restricted to the posterior hindlimb bud mesenchyme at E10.5 (Fig. 4E,F; dashed rectangles) overlapping the endogenous *Hand2* expression domain. Interestingly, it was recently reported that deletion of mm1689 in the mouse results in loss of *Hand2* expression in limb buds[22]. To demonstrate that the activity of this CRM is regulated by PBX, we mutagenized all PBX and PBX-HOX binding sites within this element (Fig. 4H and Supplementary Methods), disrupting the conserved core of the binding motifs[3,46]. We then assayed the mutagenized (MUT) and wild-type (WT) versions of this enhancer using a CRISPR/Cas9-mediated site-specific transgenic mouse assay termed enSERT[47]. Activity of the mm1689 enhancer was completely lost with high reproducibility in its MUT version (n=9) compared to WT (n=2) (Fig 4H). Together, our results indicate that in early hindlimb buds the activity of CRMs in the *Hand2* genomic landscape critically depends on intact PBX binding sites.

PBX and HAND2 collaboratively control transcriptional regulators of limb patterning

To identify GRNs co-regulated by PBX1/2 and HAND2, we used *Pbx1cKO^{Mes};Pbx2$^{-/-}$* and *Hand2cKO^{Mes}* hindlimb buds at E10.5 for bulk RNAseq analysis in comparison to littermate controls (Fig. S7A,B). Statistical analyses identified 1,489 differentially expressed genes (DEGs) in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* and 375 DEGs in *Hand2cKO^{Mes}* mutant hindlimb buds. Intersection of both transcriptomic datasets, considering only genes that were robustly detected in both sets of experiments, identified 46 DEGs significantly upregulated in hindlimb buds from both mutant mouse lines, defining them as genes repressed by both PBX1/2 and HAND2 (Fig. 5A, left, Fig. S7B and Supplementary Table 4). GO analyses of the PBX1-HAND2 repressed transcriptional targets revealed a significant association with 'developmental processes' and 'transcription factors' categories (Fig. S7C). In addition, 37 genes were significantly downregulated in both types of mutant hindlimb buds compared to littermate controls (Fig. 5A, left, Fig. S7B,C, and Supplementary Table 4), identifying target genes positively regulated by both PBX1/2 and HAND2 transcription factors. Lastly, 31 DEGs were altered in a discordant manner in the two mutant mouse lines, *i.e.* one transcript was upregulated in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* and downregulated in *Hand2cKO^{Mes}* mutant hindlimb buds, or *vice versa* (Fig. 5A, left, and Fig. S7B,C). Of note, the GO term 'transcription factors' was significantly enriched in all DEG groups, suggesting that PBX1/2 and HAND2 act as master regulators of a downstream GRN comprising transcription factors with critical roles in limb development (Fig. S7C).

Upon signal transduction, many transcription factors translocate from the cytoplasm into the nucleus where they interact with other transcription factor complexes and/or bind directly to DNA[42]. Therefore, we investigated whether the PBX1/2-HAND2 target genes are co-expressed with *Pbx1/2* and *Hand2* in the same hindlimb bud cells by analyzing the mesenchymal clusters in our scRNAseq datasets (based on the statistical threshold 'area under the curve' >=0.55 to determine the extent to which each target gene was co-expressed with *Pbx1/2* and/or *Hand2*; see Methods). Remarkably, seven transcription factors with critical functions in limb and/or skeletal development, namely *Msx1*[48], *Alx3*[49], *Lhx9*[50], *Prrx1*[51], *Zfhx4*[52], *Ets2*[53] and *Snai1*[54,55] were above this statistical threshold (right panel, red asterisks, Fig. 5A). Genes up-regulated in both *Pbx1/2* and *Hand2* mutants were highly expressed in subcluster 7 (e.g. *Col1a2*), or in subclusters 1 and 5 (e.g. *Prrx1, Msx1*). In contrast, *Ets2, Polr3d*, and *Snai1*, which were downregulated in both mutants, were expressed at higher levels in subclusters 1 and 5. In addition, we used our bulk and sc RNAseq datasets, together with an available computational compendium (Dorothea[56]; see Methods) to evaluate whether PBX1/2-HAND2 target genes are co-expressed with *Pbx1/2* and *Hand2* in the same cell subpopulations. This analysis detected the largest number of PBX1/2 and HAND2 target genes in cells co-expressing *Pbx1, Pbx2* and *Hand2* within the mesenchymal sub-clusters 7 and 12 (Fig. 5B,C and Fig. S7D). Both promoter and intergenic regions of

these target genes interacted with both PBX1 and HAND2 in accessible and active chromatin regions (Fig. 5D). Together, these analyses establish that the co-regulated PBX1/2-HAND2 transcriptional targets required for limb development are: 1) significantly dysregulated in hindlimb buds with mesenchymal-specific loss of *Pbx1/2* and *Hand2;* 2) bound by both PBX1 and HAND2; 3) associated with open and active chromatin; and 4) co-expressed in posterior hindlimb mesenchymal cells at E10.5.

We then examined whether previously defined topological associating domains (TADs)[44,57] comprising the identified DEGs exhibited a higher number of putative regulatory elements bound by PBX1 and/or HAND2 than TADs comprising genes whose expression was not altered in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* or *Hand2cKO^{Mes}* mutant hindlimbs (Fig. S8). This analysis indicated that distal elements within TADs of DEGs identified in either of the two mutants showed a trend toward higher regulatory scores (based on both the total number and individual strength of the sites contacted by PBX1 and/or HAND2 within the TAD; see Methods) than genes whose expression was not altered in *Pbx1cKO^{Mes};Pbx2$^{-/-}$* and/or *Hand2cKO^{Mes}* hindlimb buds. Moreover, the genes within TADs with the highest number and strength of distal regions bound by PBX1 and/or HAND2 were generally upregulated or regulated in a discordant manner in both mutant hindlimb buds (red arrows, Fig. S8). These results indicate that: 1) TADs containing higher numbers of DEGs comprise higher numbers of genomic regions bound by PBX1 and/or HAND2; and 2) genes that are repressed by these two transcriptional regulators are embedded within complex *cis*-regulatory landscapes.

Genetic interaction between *Pbx1* and *Hand2* is essential for proximo-distal hindlimb and digit patterning

To validate the developmental impact of the novel PBX1/2-HAND2-dependent GRN reconstructed by multi-omics approaches, we assessed its functional importance for normal mouse hindlimb bud patterning and skeletal development by a classical genetic interaction experiment. We generated compound mutant embryos lacking *Pbx1* and one allele of *Hand2* in the hindlimb bud mesenchyme using the *Hoxb6Cre* deleter line (Fig. 6A-G). The hindlimb skeletal morphology was normal in both heterozygous *Pbx1* and *Hand2* embryos (*Pbx1$^{f/+}$;Hoxb6$^{Cre/+}$* and *Hand2$^{+/-}$;Hoxb6^{Cre}*) compared to WT controls. In contrast, pelvic malformations and a shorter femur (n=4/4) were detected in *Pbx1$^{f/f}$;Hoxb6$^{Cre/+}$* single homozygous mutant embryos. Compound *Pbx1$^{f/f}$;Hand2$^{+/-}$;Hoxb6$^{Cre/+}$* embryos lacking both copies of *Pbx1* and one copy of *Hand2* in the hindlimb bud mesenchyme revealed a striking genetic interaction in hindlimb skeletal and digit patterning. In particular, *Pbx1$^{f/f}$;Hand2$^{+/-}$;Hoxb6$^{Cre/+}$* hindlimb skeletons exhibited a conspicuously dysmorphic pelvis and shorter femur and severe fibular defects, including agenesis, in combination with variable reduction or loss of posterior

digits (Fig. 6D-F'). Despite the different penetrance of the autopod phenotype in compound *Pbx1*$^{f/f}$;*Hand2*$^{+/-}$;*Hoxb6*$^{Cre/+}$ mutants (Fig. 6D'-F'), no autopod defects were observed in either *Pbx1*$^{f/f}$;*Hoxb6*$^{Cre/+}$ or *Hand2*$^{+/-}$;*Hoxb6*$^{Cre/+}$ hindlimbs. These results establish that genetic interaction of *Pbx1* and *Hand2* is required for normal hindlimb bud patterning.

HAND2 "selects" a subset of PBX-bound CRMs to confer limb bud patterning specificity

As *Pbx* genes are broadly expressed during embryogenesis and fulfill essential pleiotropic roles during organogenesis[46], we next asked whether context-specific PBX functions are achieved by tissue-specific PBX binding or by cooperativity with different transcription factors that confer tissue-specificity. To this end, we compared PBX genome-wide binding profiles across multiple tissues of the developing embryo. Given the essential roles of PBX proteins in upper lip/primary palate morphogenesis that we reported[58,59], we generated additional PBX1 ChIPseq datasets from the murine midface (MF) at E10.5 and E11.5. In addition, as PBX drives a second branchial arch (BA2)-specific transcriptional program together with MEIS and HOXA2[37], the available PBX1 ChIPseq dataset from E11.5 BA2 was also included in this study. Intersection of the three datasets (PBX1 genome-wide occupancy in the hindlimb bud, MF, and BA2), using only peaks present in both ChIPseq replicates (overlap of at least 1 nt), identified more than 10,000 PBX1-bound genomic regions in each of the tissues analyzed, a large fraction of which was shared across all three embryonic tissues analyzed (Fig. 7A; see Supplementary Table 5 for precise numbers and percentages). The fraction of peaks shared across these datasets is very similar in size to the binding overlap expected across biological replicates of the same genome-wide binding experiment (>50%)[60]. This points to highly promiscuous PBX binding in different embryonic tissues. GREAT analysis of the PBX binding profiles in hindlimb bud and BA2 revealed significant associations with diverse developmental processes. GO terms associated with limb development (Fig. 7B,C) were included, but they were not the top enriched categories in the hindlimb (or in BA2) dataset. Next, we intersected PBX-bound peaks with HAND2 or HOXA2 binding[61], and we observed that HAND2 and HOXA2 each occupied only a small subset of PBX peaks. Focusing on the shared HAND2-PBX1 ChIPseq peaks in hindlimb buds revealed that "limb development" and "limb morphogenesis" were among the top three developmental processes (Fig 7D,E). Moreover, HAND2 showed preferential binding to PBX1-bound peaks in the hindlimb bud (48%) compared to BA2 (13%) or MF (18%) tissues (Fig 7D; Supplementary Table 5). In contrast, HOXA2 showed preferential binding to PBX1-bound peaks within BA2 (70%) compared to hindlimb buds (43%) (Fig 7F; Supplementary Table 5). Comparative analysis of HOXA2 binding between BA2 and MF tissues was not conducted, as the MF is a so-called "*Hox*-less" domain[62]. Notably, HAND2- and HOXA2-bound peaks showed only a minimal overlap (12%) and are exclusive (Fig. 7G; Supplementary Table 5). For example, PBX1 binds

to similar CRMs in the genomic landscape of *Snail1* and *Zfp503* target genes in hindlimb bud and BA2. However, PBX1 binds the majority of co-bound peaks at both loci with distinct co-factors, HAND2 or HOXA2, in the hindlimb and BA2, respectively (Fig. 7H,I). These data indicate that HAND2 and HOXA2 select for distinct PBX binding events in their domain of expression (hindlimb and BA2, respectively), conferring tissue-specificity to PBX function.

Lastly, comparing the hindlimb bud to BA2 tissue, we quantified PBX1 binding enrichment of peaks that also overlap with HAND2 or HOXA2 binding. We found that PBX1-HAND2 co-bound peaks in the hindlimb bud are significantly more enriched relative to PBX1 peaks not overlapping HAND2 or HOXA2 binding ($p < 2.2e-16$) (Fig. 7J). Similarly, co-binding with HOXA2 significantly increased PBX1 binding in BA2 ($p < 2.2e-16$) (Fig. 7J). In summary, these results demonstrate that promiscuous PBX transcription factors occupy a vast pool of shared genomic regions during embryonic development, but acquire restrained tissue specificity via cooperation with distinct transcriptional regulators, such as HAND2 in the developing hindlimb.

DISCUSSION

Disturbance of the gene expression programs that specify cell fates and control tissue patterning, as well as organ morphogenesis, causes congenital malformations. As transcription factors rarely work in isolation, the regulation of gene expression is often mediated by collaborative interactions of transcription factor within complexes[63]. Despite decades of research, it remains unclear how spatio-temporal interactions among transcription factors and select cofactors achieve functional developmental specificity and how transcription factor complexes regulate their target genes through multiple CRMs such as enhancers and repressors. Our previous studies reported that, despite broad expression of *Pbx* genes during embryonic development, their encoded transcription factors control distinct target and effector genes in different tissues and organs, including the limb bud[4,14,46,64–68]. Our research highlighted the essential roles of *Pbx1/2* during limb development[4,46], but did not identify the cell type(s) that require PBX function or the time window during which limb bud development depends on PBX. Beyond the phenotypic characterization and identification of target genes, our previous studies did not provide insight into how PBX factors attain functional specificity during development despite their widespread distribution in the embryo. Here, the tissue-specific mutagenesis, genome-wide and single-cell experimental strategies employed in the mouse allowed us to address these critical issues and provided new insight into broader questions related to transcription factor promiscuity and specificity during vertebrate development.

We previously reported that inactivation of both *Pbx1* and *Pbx2* results in more severe phenotypes in mouse hindlimbs than forelimbs[5,6], as a consequence of functional redundancy among *Pbx* family members, as *Pbx3* is not expressed in hindlimbs. Thus, in this study we chose the mouse hindlimb bud to gain a mechanistic understanding of how TALE PBX homeoproteins control specific transcriptional programs *in vivo* in a context-dependent manner. The combination of tissue and temporally-controlled gene inactivation in the mouse with multi-omics approaches using dissected hindlimb buds established that *Pbx1/2* are dispensable for AER formation and function, but essential in the mesenchyme for normal hindlimb bud initiation and early patterning. Thereafter, PBX1/2 become dispensable in the hindlimb mesenchyme, pointing to their crucial roles in mesodermal progenitors during the onset of limb bud development. Consistent with essential PBX1/2 functions in mesoderm patterning, we reported that they are required to control *Polycomb* and *Hox* gene expression in the paraxial mesoderm, which gives rise to the axial skeleton[5]. Of note, a recent study showed that induction of the paraxial mesoderm relies on a BRACHYURY-TALE-HOX gene code[69]. In this context, TALE-HOX transcription factors establish a chromatin landscape permissive for recruitment of the WNT-effector LEF1, which in turn

unlocks WNT-mediated transcriptional programs that drive paraxial mesodermal fates. Together, these studies establish that PBX1/2 homeoproteins, in cooperation with tissue-specific cofactors, execute essential roles in the specification and patterning of both axial and limb bud mesodermal lineages.

Limb bud mesenchyme-specific deletion of either *Pbx1/2* or *Hand2* causes similar limb skeletal defects and comparable cellular and molecular alterations. In addition, the phenotypes of mouse embryos lacking either *Pbx1/2* or *Hand2* in the limb mesenchyme phenocopy the morphological and molecular defects of embryos lacking *Shh*[31] or its limb ZRS enhancer[25] resulting in loss of posterior digits. We and others reported that activation of *Shh* localized expression in the posterior limb bud mesoderm is regulated by interactions of transcriptional regulators, including HOX, PBX, HAND2, and ETS with the ZRS, whereas TWIST1, ETV, and GATA factors prevent anterior ectopic activation of *Shh* expression[6,20,53,70–73]. Here, we demonstrate that *Pbx1/2* and *Hand2* cooperate to regulate a GRN that controls the initiation of hindlimb bud patterning. Lastly, we establish that PBX1 regulates *Hand2* expression via the control of select CRMs active within the posterior hindlimb mesenchyme. Indeed, mutations of all PBX and PBX-HOX binding sites for the mm1689 CRM demonstrated that PBX is essential for its activity. Of note, it was recently reported that deletion of this enhancer, which comprises also MEIS and HOXD13 binding sites, results in loss of *Hand2* expression in murine limb buds[22], strongly arguing that PBX is part of the transcriptional complex that is required for *Hand2* activation. Together, these findings provide the first mechanistic evidence that explains how the loss of *Pbx1/2* or *Hand2* results in similar skeletal, cellular, and molecular alterations during limb bud development.

Transcription factors that act within the same GRN must be co-expressed within the same cells, as they translocate into the nucleus to regulate common downstream target gene expression[42]. A recent study based on synthetic GRN modelling reported that widespread co-expression emerges in regulatory networks and is a direct indicator of active co-regulation within a given cellular context[74]. Combination of scRNASeq and immunofluorescence on hindlimb buds provides compelling evidence that both *Pbx1/2* and *Hand2* transcripts are co-expressed within restricted mesenchymal subpopulations in the posterior-proximal hindlimb bud, where PBX1 and HAND2 proteins also co-localize in nuclei. The spatio-temporally constrained PBX1/2-HAND2-dependent GRN comprises target genes that encode critical limb developmental regulators, within well-defined epigenetic landscapes, and achieves superior organismal- and tissue-level resolution within a confined hindlimb bud domain (Fig. 8).

It was recently reported that *Meis1/2* genes, which are broadly expressed in the mouse embryo and encode TALE homeodomain proteins that can dimerize with PBX[3,46], have also essential roles during

early limb bud development. Their genetic inactivation in the mouse results in distal abnormalities including loss of posterior digits[22,75], similar to *Pbx1/2*-deficient mice. Furthermore, it was shown that MEIS and TBX transcription factors control a limb-specific GRN by co-regulating enhancers associated with genes essential for limb bud initiation such as *Fgf10*[22]. While this study showed that expression of *Hand2* and *Shh* is altered in *Meis1/2*-deficient hindlimbs, it remains to be determined whether MEIS proteins are part of the same PBX1/2-HAND2-directed GRN, or of a larger shared GRN. Our *de novo* motif analysis provides evidence that PBX1 can bind DNA without HAND2, but likely together with other TALE factors such as MEIS1/PREP1 in hindlimb buds (see Fig. 3H, third column), in agreement with ChIPseq experiments conducted on whole embryos[76]. Therefore, we envisage a scenario whereby, together with PBX1/2 and HAND2, MEIS proteins are part of the same transcriptional complex with other homeodomain transcription factors (see Fig. S4H) that participate in the regulation of shared target genes with essential functions in hindlimb bud patterning. Supporting this view, it was reported that MEIS and PBX largely occupy the same genomic regions in the branchial arches[37,61]. Of note, transcription factor cooperativity does not only rely on direct protein-protein interaction but can also be indirect. In this case, the transcription factors do not interact but stabilize each other by binding to the same genomic regions, thus competing with nucleosomes[63,77,78].

Several studies have attributed 'pioneer factor' functions to PBX proteins[79–82]. For example, it was proposed that during early development PBX may penetrate repressive chromatin to mark specific genes for transcriptional activation as part of multimeric complexes[83,84]. Furthermore, it was shown that during early zebrafish embryogenesis TALE regulators access promoters, thereby facilitating chromatin accessibility of transcriptionally inactive genes, and subsequently HOX proteins are required to initiate transcription[79]. In the mouse, it was reported that during neural differentiation PBX1 binds the promoter/proximal enhancer of doublecortin (*Dcx*) in undifferentiated neural progenitors, when chromatin is still compacted by histone H1 and long before *Dcx* is expressed[81]. Our study generates new genome-wide evidence in support of the putative pioneer factor functions of PBX proteins: 1) PBX1 without HAND2 binds to regions that are significantly more accessible and more enriched with H3K27ac than regions bound by HAND2 without PBX1 (see Fig. 3I); 2) there is a significant association between repressed promoter elements marked by H3K27me3 and PBX1 binding without HAND2 (see Fig. S5D); and 3) in different embryonic tissues, including hindlimb bud, MF, and BA2, DNA regions bound by PBX1 are largely overlapping (see Fig. 7). Nonetheless, in order to unequivocally assign pioneer factor roles to PBX proteins it will be critical to establish that PBX-marked genes are not already primed for transcriptional activation by pre-existing histone modifications[85,86] prior to PBX binding. As PBX1/2 are required in the limb mesenchyme during hindlimb bud initiation, our study

supports the notion that these TALE regulators have critical functions in multipotent progenitor cell populations[69]. Of note, while in the mouse embryo the onset of *Hox* gene expression occurs only during gastrulation[87], *Pbx* genes are expressed earlier, namely in the blastocyst and morula onward[88]. Furthermore, *Pbx* gene transcripts are already detected in both mouse and human embryonic stem cells[89,90], consistent with putative pioneer factor functions.

PBX homeodomain proteins of the TALE superclass[3] are characterized by the insertion of a three-amino-acid loop in the homeodomain, a highly conserved DNA binding moiety shared by hundreds of transcription factors[91,92], which typically is not sufficient to direct transcription factors to their functional targets. Our study shows that PBX proteins bind indiscriminately to a vast number of CRMs shared across diverse embryonic tissues. Thus, binding of PBX TALE proteins alone seems insufficient for context-dependent functional activity, whereas PBX1 binding acquires tissue-specific functions via combinatorial interactions with different transcription factors, such as HAND2, which confers limb bud specificity (see Fig. 8). A similar model defining how TALE factors can activate target genes in different contexts was proposed for MEIS[37,61]. Based on our findings (see Fig. 3 and Fig. 7), one can envisage a scenario whereby increased accessibility reflects higher DNA binding affinity of PBX in the presence of HAND2, and ultimately longer residence time on chromatin. Of note, the duration of transcription factor binding to DNA -or dwell time- positively correlates with downstream transcriptional output[93,94].

PBX transcription factors have long been considered to serve primarily as cofactors for HOX proteins, and their heterodimerization with PBX has been classically proposed as a mechanism by which HOX proteins acquire DNA-binding selectivity and specificity[95–99]. However, it always remained difficult to envisage how transcription factors such as PBX homeoproteins, encoded by genes with widespread expression in the vertebrate embryo, can confer functional specificity to HOX proteins, that display domain-restricted localization[46]. Challenging this widely accepted model, we reported that PBX transcriptional regulators can hierarchically control *Hox* gene expression in limb buds[6] and also function in "*Hox*-less" embryonic domains, such as the developing head[58,59,68]. Using the limb bud as a model system, this study provides novel evidence that interaction with select cofactors, such as HAND2 in hindlimbs, restrains PBX proteins directing them to execute specific developmental functions in distinct embryonic tissues. These findings have important implications for our understanding of how promiscuous transcription factors achieve developmental specificity, shedding new light on previously unknown mechanisms underlying vertebrate tissue patterning and organogenesis.

MATERIALS AND METHODS

Mice

Mutant alleles used in this study were previously described and the conditions for genotyping were reported: conditional *Pbx1*[26], *Pbx2*[10], *Hand2*^{3xFLAG} allele[23], constitutive and conditional *Hand2*[20], *Hoxb6Cre*[29], *Hoxb6CRE-ERT*[30], *Msx2Cre*[27,28]. Experiments on *Pbx1*/*Pbx2* mice were performed following Weill Cornell Medical College and UCSF IACUC guidelines and experimental procedures concerning housing, husbandry, and welfare. Animal studies conducted in Switzerland involving *Hand2* and *Pbx*/*Hand2* compound mutant mice and embryos were approved by the Regional Commission on Animal Experimentation and the Cantonal Veterinary Office of the city of Basel. Animal work at Lawrence Berkeley National Laboratory (LBNL) was reviewed and approved by the LBNL Animal Welfare Committee. Stages of all embryos analyzed are indicated in all figures and figure legends. Wild-type (Swiss Webster) mice were purchased from Charles River Laboratories and were time-mated to obtain embryos for microdissections. Mouse embryos were collected from pregnant females via cesarean section at the described timepoint following observation of a vaginal plug. Noon the day of the plug was considered E0.5.

Skeletal preparations

Skeletal preparations were performed as previously described[6,8,20].

Whole-mount ISH

Whole-mount *in situ* hybridization was performed on E11.5 embryos as previously described[5,6,20]. At least three embryos for each genotype were analyzed by whole-mount *in situ* hybridization.

Immunofluorescence

Embryos were harvested at specified stages, fixed overnight at 4°C in 4% PFA in PBS, rinsed twice in PBS, and cryoprotected in 30% sucrose overnight at 4°C. Subsequently, they were embedded in OCT compound and cryosectioned at 12µm per section. Slides were blocked for 1h with 10% fetal bovine serum (FBS)/PBS and incubated overnight in 0.1% bovine serum albumin (BSA) with primary antibodies (Ab). Primary Abs used for immunofluorescence were: PBX1 (Cell Signalling, #4342, 1:200), FLAG M2 (Sigma, F1804, 1:500). Primary Ab binding was detected by AlexaFluor-conjugated secondary Abs (Invitrogen, 1:400-1:1000). Nuclei were stained with DAPI (Sigma). Fluorescence imaging was performed using a Leica SP5 confocal microscope.

scRNAseq

Microdissected hindlimb tissues were collected from 10 embryos at E10.5 (37-40 somites) in cold PBS. Tissues were dissociated to single cells using an enzymatic cocktail of Liberase and DNase I for 10 minutes at 37 degrees. Cells were passed through a 45 µm strainer to remove clumps and ensure a single cell suspension. Dead cells were eliminated with a 'Dead Cell Removal Kit' with magnetic beads (MACS, Milteny Biotech). A single cell suspension of pooled live hindlimb cells were loaded into one well for single-cell capture using the Chromium Single Cell 3' Reagent Kit V2 (10X Genomics). Library preparation for each sample was also performed using the Chromium Single Cell 3' Reagent Kit V2, and each sample was given a unique i7 index. Libraries were pooled and subjected to sequencing in an Illumina NovaSeq sequencer. Additional details in Suppl. Methods.

Bulk RNAseq

Hindlimb buds from mutant and littermate controls were dissected individually at the indicated gestational days and snap frozen in dry ice or liquid nitrogen. After genotyping, each biological replicate consisted of hindlimb buds from one individual mutant or littermate embryo, with a total of 3 biological replicates per genotype. RNA was extracted using the RNeasy Plus Micro kit (Qiagen, #74034) and RNA quantification performed using the Qubit RNA HS Assay Kit (Invitrogen, #Q32852). Quality control of input RNA was performed using the RNA 6000 Pico kit (Agilent, #5067–1513) on a 2100 Bioanalyzer (Agilent) or the Fragment Analyzer (Advanced Analytical) High Sensitivity RNA kit. All RNA samples for library preparation had RIN>9. RNA sequencing libraries were prepared from 100 ng of input RNA using the non-directional kit NEBNext Ultra™ II RNA Library Prep Kit for Illumina (NEB, #E7775) with the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, #E7490) to capture polyA RNAs. Library size and quality was checked using an Agilent 2100 Bioanalyzer with the High Sensitivity DNA kit (Agilent, #5067–4626) or the Fragment Analyzer CRISPR discovery kit. Concentration of the libraries was determined with the QuBit dsDNA HS Assay kit (Invitrogen, #Q32854). *Pbx1/2* libraries were sequenced in an Illumina HiSeq 4000 to generate 50 base pair single-end reads. *Hand2* libraries were sequenced using NextSeq500 to generate 75 single-end reads. Additional details in Suppl. Methods.

ChIPseq

Embryonic hindlimb buds and embryonic midfaces (the latter dissected as described[59]) were isolated from wild-type Swiss Webster E10.5 and E11.5 mouse embryos, and immediately crosslinked for 10 min in 1% formaldehyde (Electron Microscopy Sciences, #15710). ChIP assays were performed as reported[37,38,59]. The crosslinked material was sonicated to 200-500 bp DNA fragments with a Diagenode Bioruptor or Covaris S220 sonicator. Each ChIPseq assay was performed from pooling 60-80

embryonic hindlimbs, by O/N incubation with specific antibodies (5 µg) at 4°C followed by 30 min incubation with Dynabeads protein A (PBX1, H3K27ac and H3K27me3 ChIPseq) or 60 min incubation with Dynabeads protein G (HAND2^{3XF} ChIPseq). IP and input DNA were purified using the MicroChIP DiaPure kit (Diagenode, #C03040001). Antibodies used were: PBX1 (Cell Signaling, #4243S), FLAG for HAND2 (Sigma, F1804), H3K27ac (Abcam, ab4729) and H3K27me3 (Millipore, #07-449). Each chromatin immunoprecipitation assay was conducted in two independent biological replicates. Following ChIP, DNA libraries were constructed using the MicroPlex Library Preparation Kit v2 (Diagenode, C05010012). Libraries were sequenced in an Illumina HiSeq 4000 to generate 50 base pair single-end reads. Additional details in Suppl. Methods.

ATACseq

We used the Assay for Transposase Accessible Chromatin (ATACseq) protocol, as originally described[39] with minor modifications. About 75 000 single cells from a pair of mouse hindlimb buds were used and n=3 biological replicates were analyzed. ATACseq reads were processed as described for ChIPseq. Peak calling was also performed with MACS v1.4, using the following parameters: --gsize=mm --bw=150 --nomodel --nolambda --shiftsize=75. Evidence from the three replicates was combined using the same approach described for ChIPseq. ATACseq reads were processed as described for ChIPseq. Peak calling was also performed with MACS v1.4, using the following parameters: --gsize=mm --bw=150 --nomodel --nolambda --shiftsize=75. Evidence from the three replicates was combined using the approach described for ChIPseq.

***Hand2* candidate enhancer identification and selection for analysis of *LacZ* activity**

We intersected our ChIPseq and ATACseq datasets from E10.5 hindlimb buds to identify all regulatory elements within the *Hand2* TAD[44] . We selected all PBX1-bound replicated peaks showing at least a 15-fold-enrichment, excluding promoter regions. To identify the complement of transcriptional enhancers controlling *Hand2* expression in the limb mesenchyme, we intersected evolutionarily conserved elements with PBX1-interacting regions enriched for H3K27ac marks and ATACseq peaks.

Transgenic mouse reporter assays

Transgenic mouse assays at LBNL were performed in *Mus musculus* FVB strain mice. Animals of both sexes were used in these analyses. Sample size selection and randomization strategies were conducted as follows: sample sizes were selected empirically based on previous experience in transgenic mouse assays for >3,000 total putative enhancers (VISTA Enhancer Browser: https://enhancer.lbl.gov/). Mouse embryos were excluded from further analysis if they did not encode

the reporter transgene or if the developmental stage was not correct. All transgenic mice were treated with identical experimental conditions. Randomization and experimenter blinding were unnecessary and not performed. For validation of *in vivo* limb enhancer activities conventional transgenic mouse *LacZ* reporter assays involving an hsp68 minimal promoter (*Hsp68-LacZ*) were performed as previously described[100,101]. For comparison of wildtype and mutagenized enhancer versions (mm1689) enSERT was used for site-directed insertion of transgenic constructs at the H11 safe-harbor locus[47]. Hereby, Cas9 and sgRNAs were co-injected into the pronucleus of FVB single cell-stage mouse embryos (E0.5) with the targeting vector encoding a candidate enhancer element upstream of the *Shh*-promoter-*LacZ* reporter cassette (*Shh-LacZ*)[47]. Related genomic enhancer coordinates are listed in Supplementary Table 3. Predicted enhancer elements were PCR-amplified from mouse genomic DNA (Clontech) and cloned into the respective *LacZ* expression vector[101]. Embryos were excluded from further analysis if they did not contain a reporter transgene in tandem. CD-1 females served as pseudo-pregnant recipients for embryo transfer to produce transgenic embryos which were collected at E10.5 or E11.5 and stained with X-gal using standard techniques[101].

Data availability

ChIPseq, ATACseq, scRNAseq and RNAseq datasets have been deposited in the NCBI GEO database under the identifier GSE197859 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE197859; Token: edqfkysonzitpif). Additional data supporting the findings of this study are available from the corresponding author upon request.

Acknowledgements

We are grateful to Mr. R. Aho for all artwork; Drs. D. Penkov and F. Blasi for input on ChIPseq protocols and hospitality in their laboratory; A. Baur and O. Romashkina for expert technical assistance; and A. Offinger and her team for mouse care. We thank members of the Selleri laboratory and Drs. G. Panagiotakos, J. Bush, R. Blelloch, D. Lim, Y. Shen, and D. Wagner for useful discussions. Work funded by the National Institutes of Health (grants R01HD043997 to LS; R01DE028599, UM1HG009421, and R01HG003988 to AV); by the University of California, San Francisco Chancellor's Funds to LS; by the Swiss National Science Foundation grant no. 310030_184734 to RZ and AZ and grant PCEFP3_186993 to MO; and by core funding from the University of Basel to RZ and AZ. Research at the E.O. Lawrence Berkeley National Laboratory performed under US Department of Energy contract DE-AC02-05CH11231, University of California. ML was recipient of postdoctoral fellowship from the American Association for Anatomy and IB was funded by Imperial College Research Fellowship and by Medical University of Vienna.

687

688 **Author contributions**: LS and RZ designed the study; AZ designed the genetic interaction experiment;
689 ML, MO, JDB, BC, AG, AM and TC designed and performed experiments under the supervision of LS,
690 RZ, AZ, AV, DD and NB; IB, PZ, and NB performed bioinformatic analysis of all data sets. JZ and SM
691 provided an unpublished mouse line and intellectual input. Based on the draft written by ML and LS, the
692 manuscript was critically reviewed by RZ, AZ, IB, MO, TC, SM, NB and AV. All authors read the final
693 version of the manuscript.

REFERENCES

1. Royle, S., Tabin, C. & Young, J. Limb positioning and initiation: An evolutionary context of pattern and formation. *Developmental dynamics : an official publication of the American Association of Anatomists* (2021) doi:10.1002/DVDY.308.
2. Zuniga, A. & Zeller, R. Dynamic and self-regulatory interactions among gene regulatory networks control vertebrate limb bud morphogenesis. *Current topics in developmental biology* **139**, 61–88 (2020).
3. Moens, C. B. & Selleri, L. Hox cofactors in vertebrate development. *Developmental Biology* (2006) doi:10.1016/j.ydbio.2005.10.032.
4. Capellini, T. D., Zappavigna, V. & Selleri, L. Pbx homeodomain proteins: TALEnted regulators of limb patterning and outgrowth. *Developmental Dynamics* **240**, 1063–1086 (2011).
5. Capellini, T. D. *et al.* Pbx1/Pbx2 govern axial skeletal development by controlling Polycomb and Hox in mesoderm and Pax1/Pax9 in sclerotome. *Developmental Biology* **321**, 500–514 (2008).
6. Capellini, T. D. *et al.* Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development* **133**, 2263–2273 (2006).
7. Capellini, T. D. *et al.* Control of pelvic girdle development by genes of the Pbx family and Emx2. *Developmental Dynamics* **240**, 1173–1189 (2011).
8. Capellini, T. *et al.* Scapula development is governed by genetic interactions of Pbx1 with its family members and with Emx2 via their cooperative control of Alx1. *Development* **137**, 2559–2569 (2010).
9. Selleri, L. *et al.* Requirement for Pbx1 in skeletal patterning and programming chondrocyte proliferation and differentiation. *Development* **128**, 3543–3557 (2001).
10. Selleri, L. *et al.* The TALE Homeodomain Protein Pbx2 Is Not Essential for Development and Long-Term Survival. *Molecular and Cellular Biology* **24**, 5324–5331 (2004).
11. Rhee, J. W. *et al.* Pbx3 Deficiency Results in Central Hypoventilation. *The American Journal of Pathology* **165**, 1343 (2004).
12. di Giacomo, G. *et al.* Spatio-temporal expression of Pbx3 during mouse organogenesis. *Gene expression patterns : GEP* **6**, 747–757 (2006).
13. Wagner, K., Mincheva, A., Korn, B., Lichter, P. & Pöpperl, H. Pbx4, a new Pbx family member on mouse chromosome 8, is expressed during spermatogenesis. *Mechanisms of Development* **103**, 127–131 (2001).
14. Stankunas, K. *et al.* Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circulation research* **103**, 702–709 (2008).
15. Bolt, C. C. & Duboule, D. The regulatory landscapes of developmental genes. *Development* **147**, (2020).
16. Amândio, A. R. *et al.* Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse HoxD cluster. *Genes and Development* **35**, 1490–1509 (2021).
17. Fabre, P. J. *et al.* Heterogeneous combinatorial expression of Hoxd genes in single cells during limb development. *BMC Biology* **16**, 1–15 (2018).
18. Darbellay, F. & Duboule, D. Topological Domains, Metagenes, and the Emergence of Pleiotropic Regulations at Hox Loci. *Current topics in developmental biology* **116**, 299–314 (2016).
19. Xu, B. & Wellik, D. M. Axial Hox9 activity establishes the posterior field in the developing forelimb. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 4888–4891 (2011).
20. Galli, A. *et al.* Distinct Roles of Hand2 in Initiating Polarity and Posterior Shh Expression during the Onset of Mouse Limb Bud Development. *PLoS Genetics* **6**, e1000901 (2010).
21. te Welscher, P., Fernandez-Teran, M., Ros, M. A. & Zeller, R. Mutual genetic antagonism involving GLI3 and dHAND prepatterns the vertebrate limb bud mesenchyme prior to SHH signaling. *Genes & Development* **16**, 421–426 (2002).
22. Delgado, I. *et al.* Control of mouse limb initiation and antero-posterior patterning by Meis transcription factors. *Nature Communications 2021 12:1* **12**, 1–13 (2021).
23. Osterwalder, M. *et al.* HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Developmental cell* **31**, 345–357 (2014).
24. Malkmus, J. *et al.* Spatial regulation by multiple Gremlin1 enhancers provides digit development with cis-regulatory robustness and evolutionary plasticity. *Nature communications* **12**, (2021).
25. Sagai, T., Hosoya, M., Mizushina, Y., Tamura, M. & Shiroishi, T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**, 797–803 (2005).

26. Koss, M. *et al.* Congenital Asplenia in Mice and Humans with Mutations in a Pbx/Nkx2-5/p15 Module. *Developmental Cell* **22**, 913–926 (2012).
27. Liu, Y. *et al.* Regulation of the Msx2 homeobox gene during mouse embryogenesis: a transgene with 439 bp of 5' flanking sequence is expressed exclusively in the apical ectodermal ridge of the developing limb. *Mechanisms of development* **48**, 187–197 (1994).
28. Sun, X. *et al.* Conditional inactivation of Fgf4 reveals complexity of signalling during limb bud development. *Nature Genetics 2000 25:1* **25**, 83–86 (2000).
29. Lowe, L. A., Yamada, S. & Kuehn, M. R. HoxB6-Cre Transgenic Mice Express Cre Recombinase in Extra-Embryonic Mesoderm, in Lateral Plate and Limb Mesoderm and at the Midbrain/Hindbrain Junction. *Genesis* 118–120 (2000) doi:10.1002/(SICI)1526-968X(200002)26:2.
30. Nguyen, M., Zhu, J., Nakamura, E., Bao, X. & Mackem, S. Tamoxifen-dependent, inducible Hoxb6CreERT recombinase function in lateral plate and limb mesoderm, CNS isthmic organizer, posterior trunk neural crest, hindgut, and tailbud. *Developmental dynamics : an official publication of the American Association of Anatomists* **238**, 467–474 (2009).
31. Chiang, C. *et al.* Manifestation of the Limb Prepattern: Limb Development in the Absence of Sonic Hedgehog Function. *Developmental Biology* **236**, 421–435 (2001).
32. Fernández-Terán, M. A., Hinchliffe, J. R. & Ros, M. A. Birth and death of cells in limb development: A mapping study. *Developmental Dynamics* **235**, 2521–2537 (2006).
33. Charite, J., McFadden, D. G. & Olson, E. N. The bHLH transcription factor dHAND controls Sonic hedgehog expression and establishment of the zone of polarizing activity during limb development. *Development* **127**, 2461–2470 (2000).
34. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature 2019 566:7745* **566**, 496–502 (2019).
35. Kelly, N. H., Huynh, N. P. T. & Guilak, F. Single cell RNA-sequencing reveals cellular heterogeneity and trajectories of lineage specification during murine embryonic limb development. *Matrix Biology* **89**, 1–10 (2020).
36. Desanlis, I., Paul, R. & Kmita, M. Transcriptional Trajectories in Mouse Limb Buds Reveal the Transition from Anterior-Posterior to Proximal-Distal Patterning at Early Limb Bud Stage. *Journal of Developmental Biology 2020, Vol. 8, Page 31* **8**, 31 (2020).
37. Amin, S. *et al.* Hoxa2 Selectively Enhances Meis Binding to Change a Branchial Arch Ground State. *Developmental Cell* **32**, 265–277 (2015).
38. Losa, M. *et al.* A tissue-specific, Gata6-driven transcriptional program instructs remodeling of the mature arterial tree. *eLife* **6**, (2017).
39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods 2013 10:12* **10**, 1213–1218 (2013).
40. McLean, C. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495–501 (2010).
41. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* **38**, 576–589 (2010).
42. Lee, T. I. & Young, R. A. Transcriptional Regulation and its Misregulation in Disease. *Cell* **152**, 1237 (2013).
43. DeMare, L. E. *et al.* The genomic landscape of cohesin-associated chromatin interactions. *Genome research* **23**, 1224–1234 (2013).
44. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557 (2017).
45. Monti, R. *et al.* Limb-Enhancer Genie: An accessible resource of accurate enhancer predictions in the developing limb. *PLoS Computational Biology* **13**, e1005720 (2017).
46. Selleri, L., Zappavigna, V. & Ferretti, E. 'Building a perfect body': Control of vertebrate organogenesis by PBX-dependent regulatory networks. *Genes and Development* vol. 33 258–275 (2019).
47. Kvon, E. Z. *et al.* Comprehensive In Vivo Interrogation Reveals Phenotypic Impact of Human Enhancer Variants. *Cell* **180**, (2020).
48. Bensoussan-Trigano, V., Lallemand, Y., saint Cloment, C. & Robert, B. Msx1 and Msx2 in limb mesenchyme modulate digit number and identity. *Developmental Dynamics* **240**, 1190–1202 (2011).
49. Kuijper, S. *et al.* Genetics of shoulder girdle formation: roles of Tbx15 and aristaless-like genes. *Development* **132**, 1601–1610 (2005).

50. Tzchori, I. *et al.* LIM homeobox transcription factors integrate signaling events that control three-dimensional limb patterning and growth. *Development* **136**, 1375–1385 (2009).
51. Berge, D. ten, Brouwer, A., Korving, J., Martin, J. F. & Meijlink, F. Prx1 and Prx2 in skeletogenesis: roles in the craniofacial region, inner ear and limbs. *Development* **125**, 3831–3842 (1998).
52. Nakamura, E. *et al.* Zfhx4 regulates endochondral ossification as the transcriptional platform of Osterix in mice. *Communications Biology 2021 4:1* **4**, 1–11 (2021).
53. Lettice, L. A. *et al.* Opposing Functions of the ETS Factor Family Define Shh Spatial Expression in Limb Buds and Underlie Polydactyly. *Developmental Cell* **22**, 459 (2012).
54. Chen, Y. & Gridley, T. Compensatory regulation of the Snai1 and Snai2 genes during chondrogenesis. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **28**, 1412 (2013).
55. de Frutos, C. A. *et al.* Snail1 is a transcriptional effector of FGFR3 signaling during chondrogenesis and achondroplasias. *Developmental cell* **13**, 872–883 (2007).
56. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome research* **29**, 1363–1375 (2019).
57. Dixon, J. R. *et al.* Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **485**, 376 (2012).
58. Ferretti, E. *et al.* A Conserved Pbx-Wnt-p63-Irf6 Regulatory Module Controls Face Morphogenesis by Promoting Epithelial Apoptosis. *Developmental Cell* **21**, 627–641 (2011).
59. Losa, M. *et al.* Face morphogenesis is promoted by Pbx-dependent EMT via regulation of Snail1 during frontonasal prominence fusion. *Development* **145**, (2018).
60. Bardet, A. F., He, Q., Zeitlinger, J. & Stark, A. A computational pipeline for comparative ChIP-seq analyses. *Nature Protocols 2012 7:1* **7**, 45–61 (2011).
61. Bridoux, L. *et al.* HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation. *PLOS Genetics* **16**, e1009162 (2020).
62. Creuzet, S., Couly, G. & le Douarin, N. M. Patterning the neural crest derivatives during development of the vertebrate head: insights from avian studies. *Journal of anatomy* **207**, 447–459 (2005).
63. Morgunova, E. & Taipale, J. Structural perspective of cooperative transcription factor binding. *Current Opinion in Structural Biology* **47**, 1–8 (2017).
64. Hurtado, R. *et al.* Pbx1-dependent control of VMC differentiation kinetics underlies gross renal vascular patterning. *Development (Cambridge)* **142**, 2653–2664 (2015).
65. Brendolan, A. *et al.* A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. *Development* **132**, 3113–3126 (2005).
66. Golonzhka, O. *et al.* Pbx Regulates Patterning of the Cerebral Cortex in Progenitors and Postmitotic Neurons. *Neuron* **88**, 1192–1207 (2015).
67. McCulley, D. J. *et al.* PBX transcription factors drive pulmonary vascular adaptation to birth. *Journal of Clinical Investigation* **128**, 655–667 (2018).
68. Welsh, I. C. *et al.* Pbx loss in cranial neural crest, unlike in epithelium, results in cleft palate only and a broader midface. *Journal of Anatomy* (2018) doi:10.1111/joa.12821.
69. Mariani, L. *et al.* A TALE/HOX code unlocks WNT signalling response towards paraxial mesoderm. *Nature communications* **12**, (2021).
70. Kmita, M. *et al.* Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature 2005 435:7045* **435**, 1113–1116 (2005).
71. Kozhemyakina, E., Ionescu, A. & Lassar, A. B. GATA6 Is a Crucial Regulator of Shh in the Limb Bud. *PLOS Genetics* **10**, e1004072 (2014).
72. Lettice, L. A. *et al.* Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* **141**, 1715–1725 (2014).
73. Mao, J., McGlinn, E., Huang, P., Tabin, C. J. & McMahon, A. P. Fgf-Dependent Etv4/5 Activity Is Required for Posterior Restriction of Sonic hedgehog and Promoting Outgrowth of the Vertebrate Limb. *Developmental Cell* **16**, 600–606 (2009).
74. Yin, W., Mendoza, L., Monzon-Sandoval, J., Urrutia, A. O. & Gutierrez, H. Emergence of co-expression in gene regulatory networks. *PLOS ONE* **16**, e0247671 (2021).
75. Delgado, I. *et al.* Proximo-distal positional information encoded by an Fgf-regulated gradient of homeodomain transcription factors in the vertebrate limb. *Science Advances* **6**, eaaz0742 (2020).

76. Penkov, D. *et al.* Analysis of the DNA-Binding Profile and Function of TALE Homeoproteins Reveals Their Specialization and Specific Interactions with Hox Genes/Proteins. *Cell Reports* **3**, 1321–1333 (2013).
77. Mirny, L. A. Nucleosome-mediated cooperativity between transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22534–22539 (2010).
78. Moyle-Heyrman, G., Tims, H. S. & Widom, J. Structural constraints in collaborative competition of transcription factors against the nucleosome. *Journal of molecular biology* **412**, 634–646 (2011).
79. Choe, S. K., Ladam, F. & Sagerström, C. G. TALE factors poise promoters for activation by Hox proteins. *Developmental cell* **28**, 203–211 (2014).
80. Choe, S. K., Lu, P., Nakamura, M., Lee, J. & Sagerström, C. G. Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis. *Developmental cell* **17**, 561–567 (2009).
81. Grebbin, B. M. *et al.* PBX1 is required for adult subventricular zone neurogenesis. *Development (Cambridge)* **143**, 2281–2291 (2016).
82. Grebbin, B. M. & Schulte, D. PBX1 as Pioneer Factor: A Case Still Open. *Frontiers in Cell and Developmental Biology* **5**, (2017).
83. Berkes, C. A. *et al.* Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Molecular cell* **14**, 465–477 (2004).
84. Sagerström, C. G. PbX marks the spot. *Developmental cell* **6**, 737–738 (2004).
85. Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes & development* **28**, 2679–2692 (2014).
86. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nature Genetics 2017 50:2* **50**, 250–258 (2018).
87. Deschamps, J. & Duboule, D. Embryonic timing, axial stem cells, chromatin dynamics, and the Hox clock. *Genes & development* **31**, 1406–1416 (2017).
88. Sonnet, W., Rezsöhazy, R. & Donnay, I. Characterization of TALE genes expression during the first lineage segregation in mammalian embryos. *Developmental dynamics : an official publication of the American Association of Anatomists* **241**, 1827–1839 (2012).
89. Chan, K. K. K. *et al.* KLF4 and PBX1 directly regulate NANOG expression in human embryonic stem cells. *Stem cells* **27**, 2114–2125 (2009).
90. Jürgens, A. S. *et al.* PBX1 is dispensable for neural commitment of RA-treated murine ES cells. *In vitro cellular & developmental biology. Animal* **45**, 252–263 (2009).
91. Bürglin, T. R. & Affolter, M. Homeodomain proteins: an update. *Chromosoma 2015 125:3* **125**, 497–521 (2015).
92. Bobola, N. & Merabet, S. Homeodomain proteins in action: similar DNA binding preferences, highly variable connectivity. *Current opinion in genetics & development* **43**, 1–8 (2017).
93. Donovan, B. T. *et al.* Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. *The EMBO Journal* **38**, e100809 (2019).
94. Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G. & Lieb, J. D. Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature 2012 484:7393* **484**, 251–255 (2012).
95. Pöpperl, H. *et al.* Segmental expression of Hoxb-1 is controlled by a highly conserved autoregulatory loop dependent upon exd/pbx. *Cell* **81**, 1031–1042 (1995).
96. de Kumar, B. *et al.* HOXA1 and TALE proteins display cross-regulatory interactions and form a combinatorial binding code on HOXA1 targets. *Genome research* **27**, 1501–1512 (2017).
97. Parker, H. J. *et al.* A Hox-TALE regulatory circuit for neural crest patterning is conserved across vertebrates. *Nature communications* **10**, (2019).
98. Ryoo, H. D. & Mann, R. S. The control of trunk Hox specificity and activity by Extradenticle. *Genes & development* **13**, 1704–1716 (1999).
99. Mann, R. S., Lelli, K. M. & Joshi, R. Hox specificity unique roles for cofactors and collaborators. *Current topics in developmental biology* **88**, 63–101 (2009).
100. Kothary, R. *et al.* Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**, 707–714 (1989).
101. Osterwalder, M. *et al.* Characterization of Mammalian In Vivo Enhancers Using Mouse Transgenesis and CRISPR Genome Editing. in *Methods in Molecular Biology* vol. 2403 147–186 (Humana, New York, NY, 2022).

Figure 1

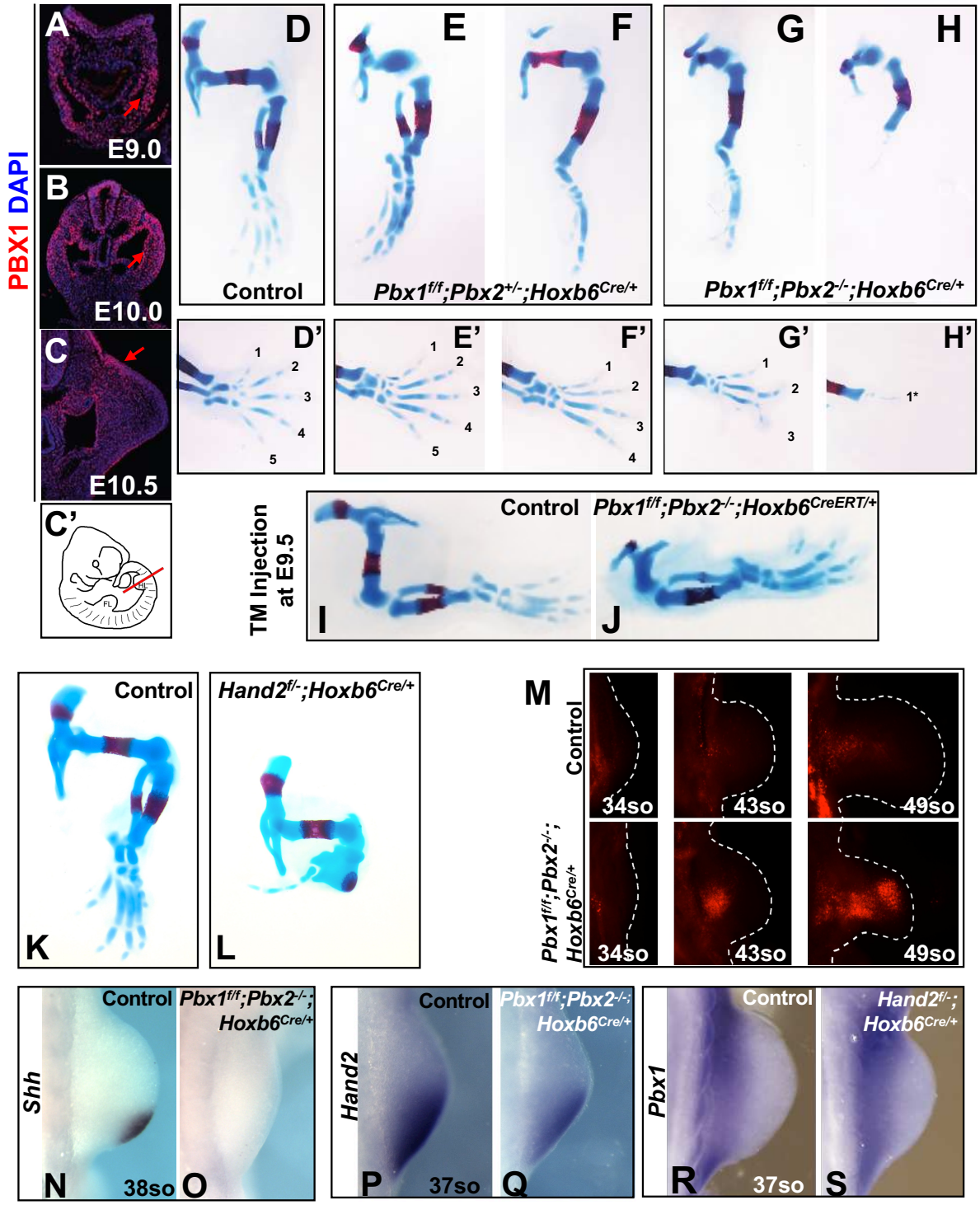


Figure 1. PBX1/2 are required in the mesenchyme for hindlimb patterning before E9.5

(A-C') IF showing PBX1 protein (red) localization in developing murine hindlimbs from E9.0 to E10.5. DAPI labels nuclei (blue); C' shows plane of section through hindlimb bud in E10.5 mouse embryo. (D-

H) E14.5 skeletal preparations of mutant hindlimbs with tissue-specific deletion of *Pbx1* on a *Pbx2*-deficient background using a *Hoxb6Cre* deleter line (referred in text as *Pbx1cKO^{Mes};Pbx2$^{+/-}$* and *Pbx1cKO^{Mes};Pbx2$^{-/-}$*); cartilage and bone visualized by Alcian Blue and Alizarin Red staining, respectively. Digits shown from anterior (digit 1) to posterior (digit 5). (I,J) Representative images of *Pbx1$^{f/f}$;Pbx2$^{-/-}$;Hoxb6$^{CreERT/+}$* hindlimb skeletal phenotype compared to *Pbx1$^{f/f}$;Pbx2$^{-/-}$* Control. Ratio of tibia/femur length in mutant embryos *versus* controls quantified in Suppl. Figure 1 (Panel J). (K,L) E14.5 skeletal preparations of mutant hindlimbs with tissue-specific deletion of *Hand2* using the *Hoxb6Cre* deleter (*Hand2$^{f/-}$;Hoxb6$^{Cre/+}$*). (M) Detection of apoptotic cells by LysoTracker Red. *Pbx1$^{f/f}$;Pbx2$^{-/-}$;HoxB6$^{Cre/+}$* hindlimb buds and respective controls at E10.25 (34 somites), E11.0 (43 somites), and E11.5 (48 somites). (N-S) Whole mount *in situ* hybridization (WISH) of *Shh* (N,O) and *Hand2* (P,Q) in E10.75 *Pbx1cKO^{Mes};Pbx2$^{-/-}$* hindlimbs *versus* controls. (R,S) WISH of *Pbx1* in E10.75 *Hand2cKO^{Mes}* hindlimb *versus* controls. In all panels, hindlimb buds oriented with anterior to the top and posterior to the bottom.

Figure 2

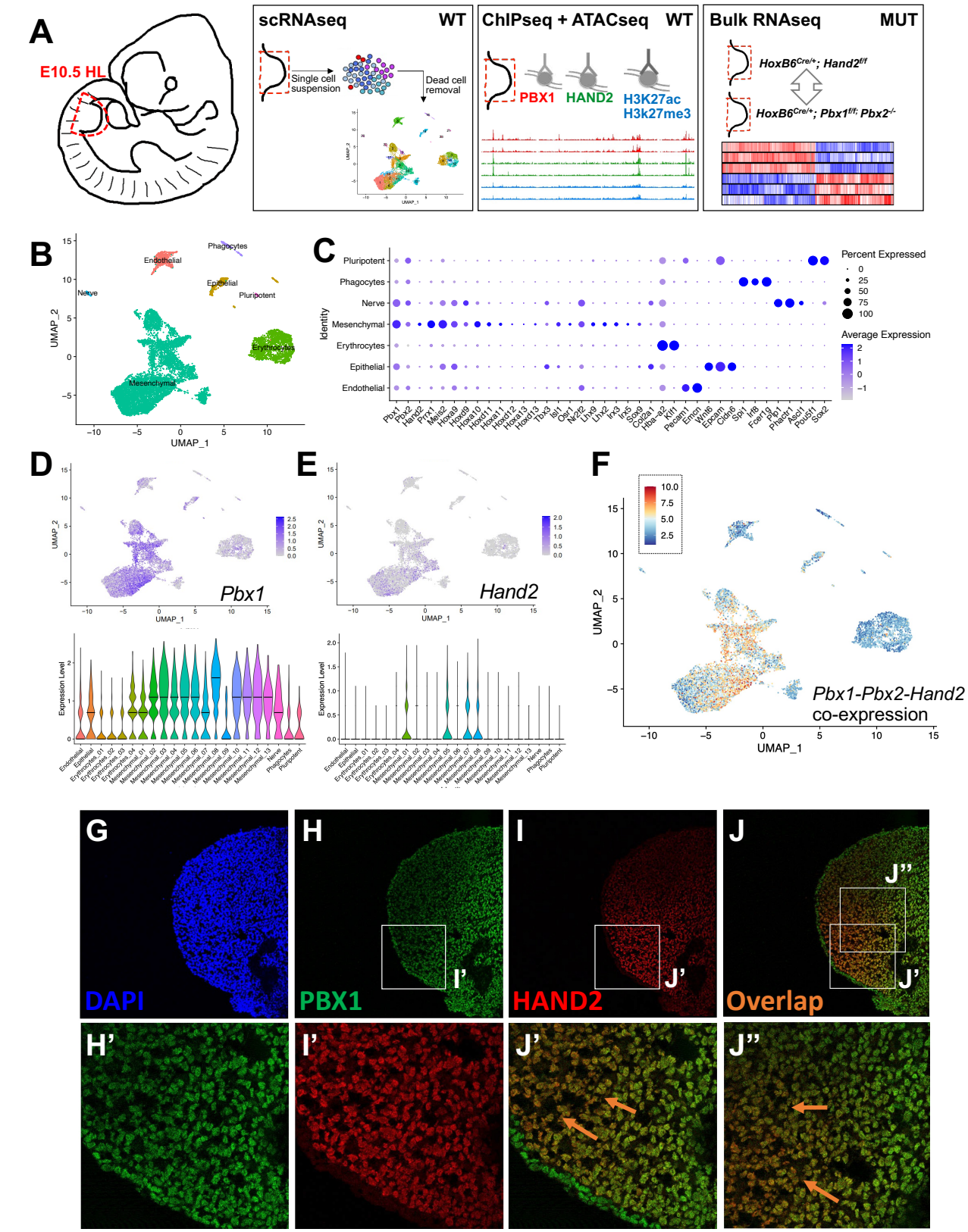

A
E10.5 HL
scRNAseq
WT
Single cell suspension
Dead cell removal
ChIPseq + ATACseq
WT
PBX1
HAND2
H3K27ac
H3K27me3
Bulk RNAseq
MUT
HoxB6Cre/+; Hand2f/f
HoxB6Cre/+; Pbx1f/f; Pbx2-/-
B
Phagocytes
Endothelial
Epithelial
Nerve
Pluripotent
Erythrocytes
Mesenchymal
UMAP_1
UMAP_2
C
Pluripotent
Phagocytes
Nerve
Mesenchymal
Erythrocytes
Epithelial
Endothelial
Identity
Percent Expressed
Average Expression
D
Pbx1
E
Hand2
Expression Level
F
Pbx1-Pbx2-Hand2
co-expression
G
DAPI
H
PBX1
I
HAND2
J
Overlap
H'
I'
J'
J''

Figure 2. Identification of hindlimb bud mesenchymal subpopulations co-expressing *Pbx1/2-Hand2*.

(A) Workflow of the genome-wide approaches performed in parallel in this study. (B) Uniform manifold approximation and projection (UMAP) representation of high-quality 9,859 hindlimb cells at E10.5-10.75, defined by scRNAseq. (C) Dot plot of top differentially expressed markers for each embryonic hindlimb population. Size of each dot represents proportion of cells within a given population that expresses the gene; intensity of color indicates average level of gene expression in indicated population. (D,E) UMAPs (upper) and violin plots (lower) showing normalized expression patterns for *Pbx1* and *Hand2*. Y axes in violin plots are at saturation and differ in scale. (F) UMAP highlighting co-expression of *Pbx1-Pbx2-Hand2.* (G-J') IF of PBX1 (green) and HAND2 (red) proteins in posterior-proximal hindlimb mesenchyme. Higher magnification shows co-localization (orange) in H', I' J' and J'' insets (orange arrows; double positive cells). DAPI (blue) labels nuclei.

Figure 3

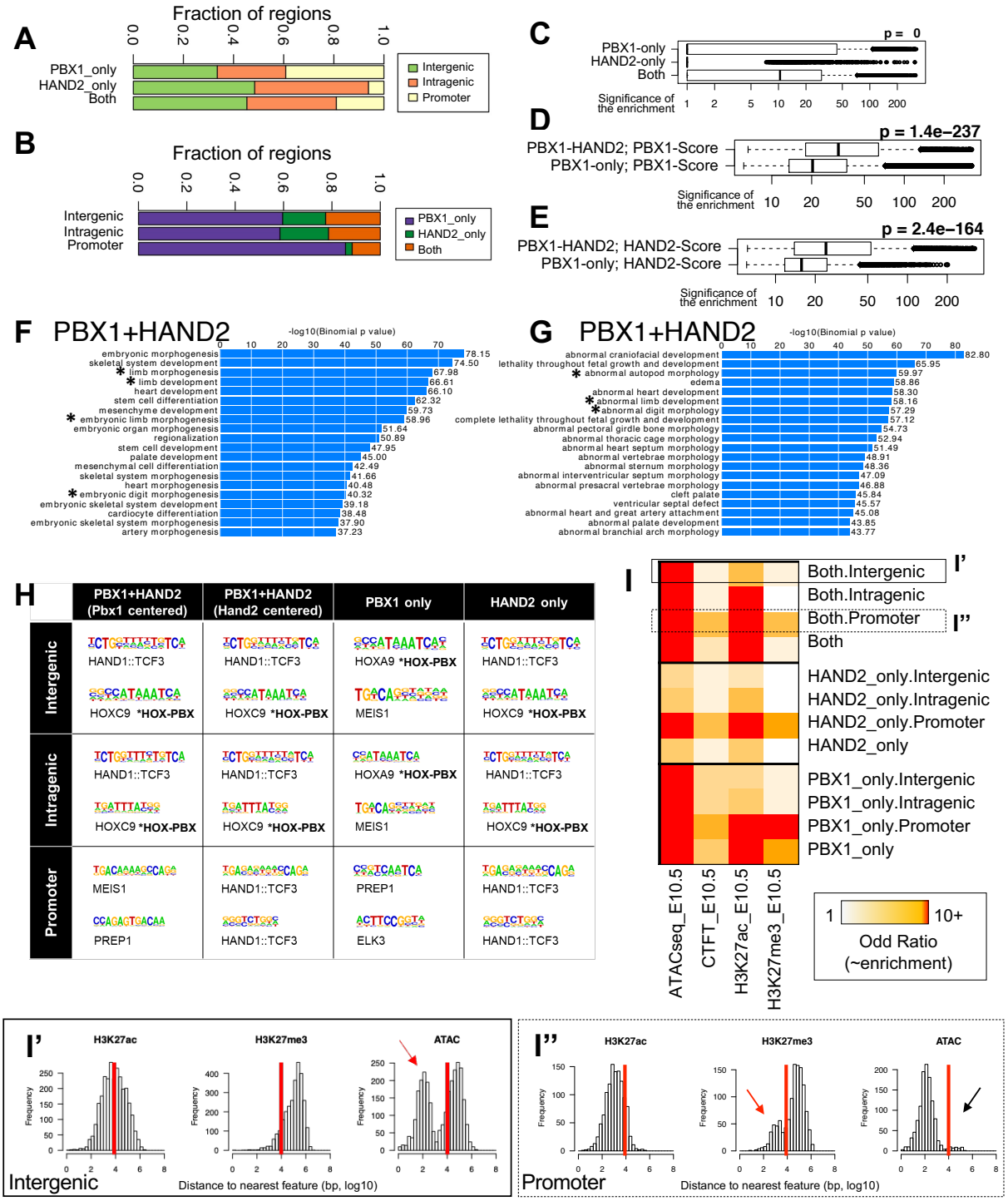
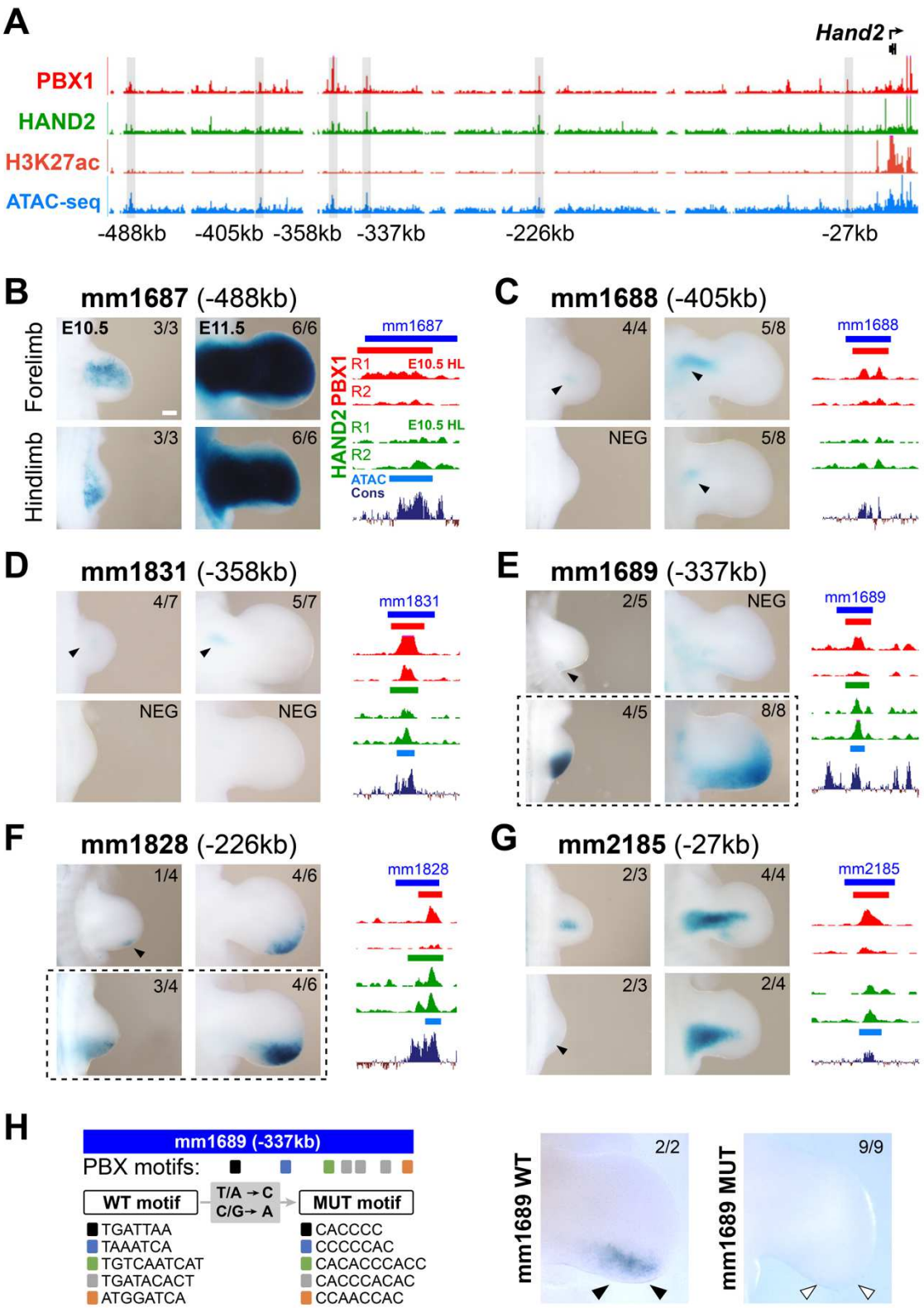


Figure 3. PBX1 and HAND2 control a shared gene regulatory network that drives early hindlimb development.

(A) Distribution of the proportion of PBX1-only, HAND2-only or PBX1-HAND2 co-bound peaks relative to gene intergenic, promoter, or intragenic regions. (B) Similar to (A); showing proportion of intergenic, promoter, or intragenic regions relative to PBX1-only, HAND2-only, or PBX1-HAND2 co-bound peaks. (C) Assessment of chromatin accessibility for PBX1-only and HAND2-only bound peaks *versus* PBX1-HAND2 co-bound peaks. X axis shows -log10(p-value) of ATACseq enrichment, on logarithmic scale. (D-E) Assessment of enrichment values for PBX1-only and HAND2-only bound peaks *versus* PBX1-HAND2 co-bound peaks. X axis shows the -log10(*p*-value) of ChIPseq enrichment, on logarithmic scale. (F,G) Top over-represented biological processes (F) and mouse phenotypes (G) associated with PBX1-HAND2 co-bound peaks. X axes show -log10 of binomial (uncorrected) *p*-values. Limb-related categories highlighted by asterisk*. (H) Top two most significantly enriched motifs identified by *de novo* motif enrichment analysis of peaks co-bound by PBX1-HAND2, PBX1-only, or HAND2-only (*q*-value <= 0.05). Identity of the most similar annotated motif indicated below each logo. (I) Heatmap summarizing enrichment (as odds ratio of observed *versus* expected fraction of peaks) of indicated chromatin features (columns) at peaks co-bound by PBX1-HAND2 (both), PBX1-only, or HAND2-only (rows). (I',I'') Distribution of H3K27ac/H3K27me3 or ATACseq peaks relative to distance to nearest peak bound by both PBX1-HAND2 in intergenic sites (I') or promoters (I'').

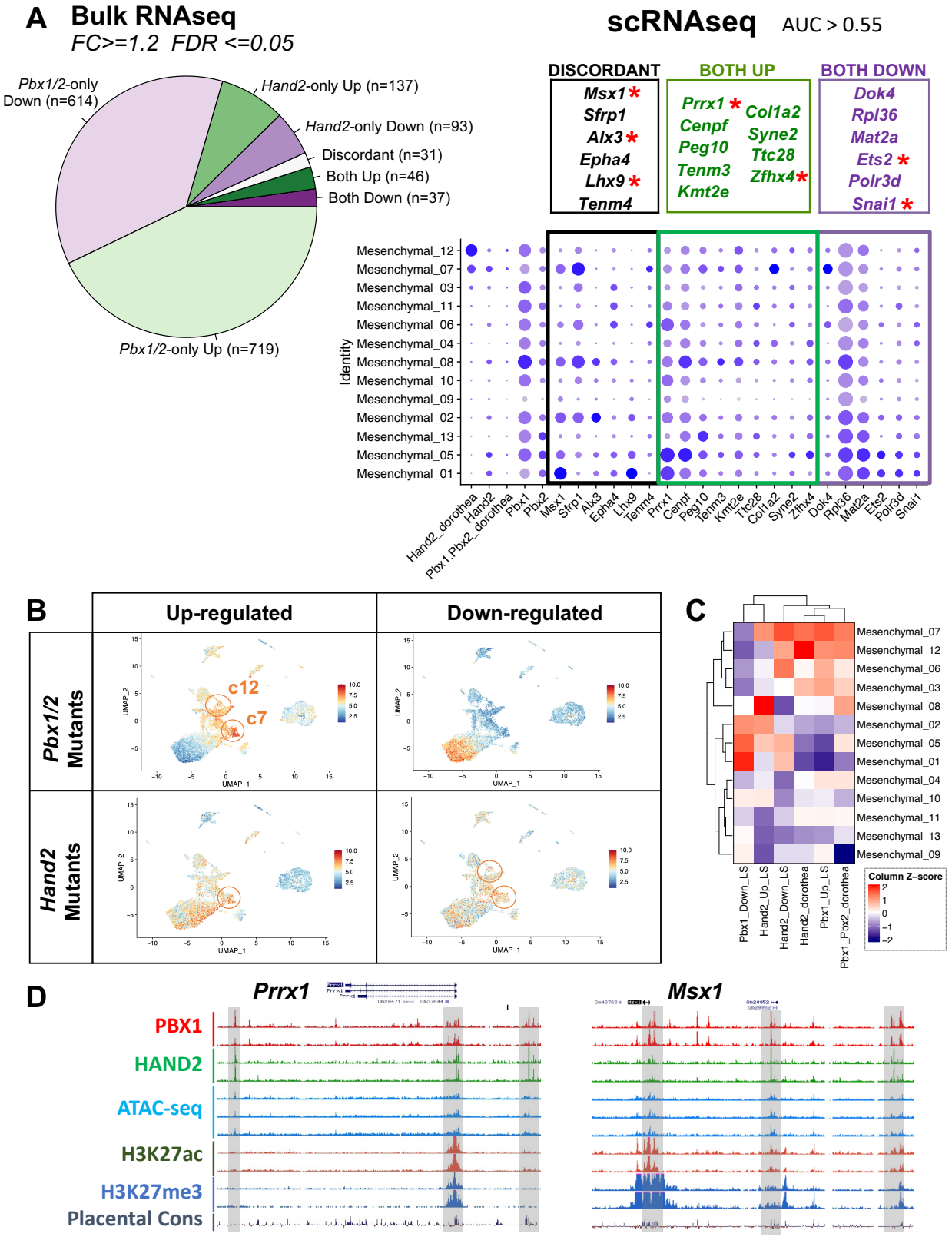
Figure 4


A
Hand2
PBX1
HAND2
H3K27ac
ATAC-seq
-488kb
-405kb
-358kb
-337kb
-226kb
-27kb
B
mm1687 (-488kb)
Forelimb
Hindlimb
E10.5
3/3
E11.5
6/6
3/3
6/6
mm1687
R1
E10.5 HL
R2
HAND2
PBX1
R1
E10.5 HL
R2
ATAC
Cons
C
mm1688 (-405kb)
4/4
5/8
NEG
5/8
mm1688
D
mm1831 (-358kb)
4/7
5/7
NEG
NEG
mm1831
E
mm1689 (-337kb)
2/5
NEG
4/5
8/8
mm1689
F
mm1828 (-226kb)
1/4
4/6
3/4
4/6
mm1828
G
mm2185 (-27kb)
2/3
4/4
2/3
2/4
mm2185
H
mm1689 (-337kb)
PBX motifs:
WT motif
T/A → C
C/G → A
MUT motif
TGATTAA
TAAATCA
TGTCAATCAT
TGATACACT
ATGGATCA
CACCCC
CCCCCAC
CACACCCACC
CACCCACAC
CCAACCAC
mm1689 WT
2/2
mm1689 MUT
9/9

964

Figure 4. PBX1 directly regulates the enhancer activities of specific CRMs in the *Hand2* TAD. (A) Overview of the genomic signatures of identified candidate limb enhancer elements (grey lines) within the *Hand2* TAD. *Hand2* transcriptional start site (TSS) marked by black arrow (top right). Enhancers listed based on their relative distance to *Hand2* TSS (bottom). (B-G) Analysis of transcriptional enhancer activity in E10.5 and E11.5 forelimb (top panels) and hindlimb (bottom panels) buds using mouse *LacZ* transgenic reporter assays. Left panels: Enhancer activity domains in embryonic forelimbs and hindlimb visualized by *LacZ* staining (blue). Weaker or more restricted domains highlighted by black arrowheads. Numbers at the top right of each panel indicate reproducibility in *LacZ* reporter assay (independent biological replicates with similar results). Right panels: UCSC genome browser tracks depicting replicated PBX1 (red) and HAND2 (green) ChIPseq peaks in E10.5 hindlimb buds for each element tested (blue). ATACseq peak calls (azure) and placental conservation (Cons, dark purple). The 6 enhancers display *bona fide* hindlimb-specific activity in at least one of the two time-points analyzed. Corresponding Vista enhancer IDs (mm: mus musculus) indicated. (H) Analysis of site-directed transcriptional enhancer activity of wild-type (WT) and mutant (MUT) mm1689 enhancer variants in E11.5 hindlimb buds using enSERT. Left Panel: mutagenesis strategy for enhancer mm1689, in which all PBX and PBX-HOX binding sites (WT motif) have been mutagenized (MUT motif). Right Panel: Activity of mm1689 is lost in MUT version. Numbers of embryos with reproducible *LacZ* staining (or lack thereof) in posterior limb bud mesenchyme (arrowheads) over total number of transgenic embryos analyzed (top right). Only transgenic embryos that carried at least two copies of the reporter transgene at the H11 locus were included in our analysis[47] (see also Methods).

Figure 5

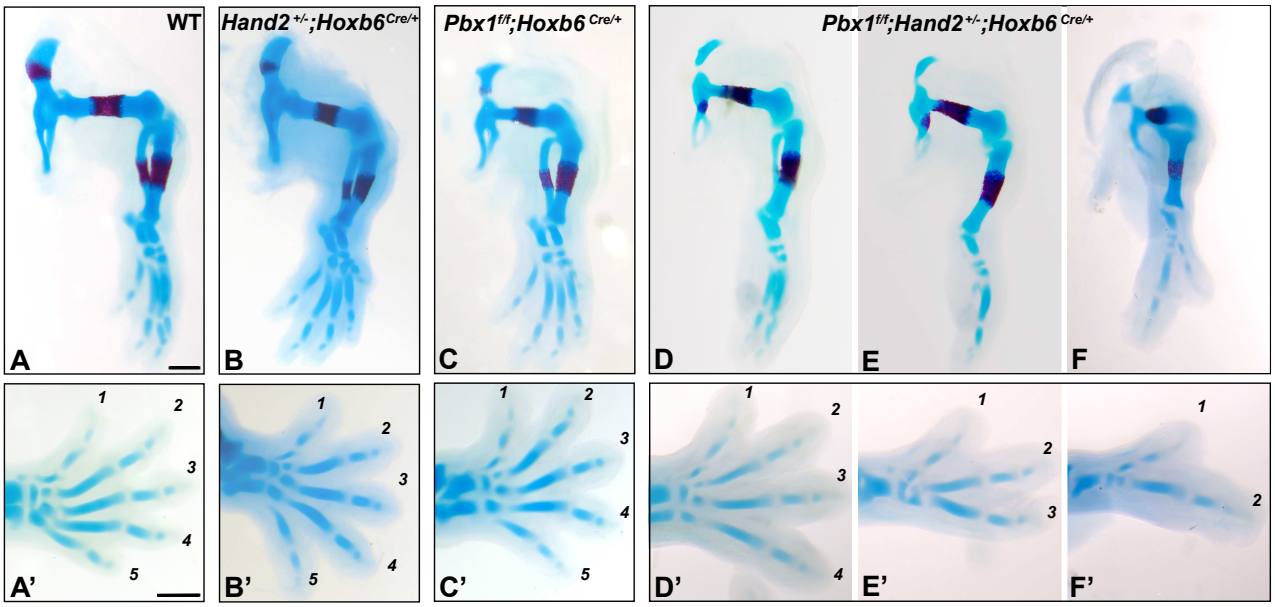

A
Bulk RNAseq
FC>=1.2 FDR <=0.05
Pbx1/2-only Down (n=614)
Hand2-only Up (n=137)
Hand2-only Down (n=93)
Discordant (n=31)
Both Up (n=46)
Both Down (n=37)
Pbx1/2-only Up (n=719)
scRNAseq
AUC > 0.55
DISCORDANT
BOTH UP
BOTH DOWN
Msx1 *
Sfrp1
Alx3 *
Epha4
Lhx9 *
Tenm4
Prrx1 *
Cenpf
Peg10
Tenm3
Kmt2e
Col1a2
Syne2
Ttc28
Zfhx4 *
Dok4
Rpl36
Mat2a
Ets2 *
Polr3d
Snai1 *
Identity
Mesenchymal_12
Mesenchymal_07
Mesenchymal_03
Mesenchymal_11
Mesenchymal_06
Mesenchymal_04
Mesenchymal_08
Mesenchymal_10
Mesenchymal_09
Mesenchymal_02
Mesenchymal_13
Mesenchymal_05
Mesenchymal_01
Hand2_dorothea
Hand2
Pbx1.Pbx2_dorothea
Pbx1
Pbx2
Msx1
Sfrp1
Alx3
Epha4
Lhx9
Tenm4
Prrx1
Cenpf
Peg10
Tenm3
Kmt2e
Ttc28
Col1a2
Syne2
Zfhx4
Dok4
Rpl36
Mat2a
Ets2
Polr3d
Snai1
B
Up-regulated
Down-regulated
Pbx1/2 Mutants
Hand2 Mutants
c12
c7
UMAP_1
UMAP_2
C
Mesenchymal_07
Mesenchymal_12
Mesenchymal_06
Mesenchymal_03
Mesenchymal_08
Mesenchymal_02
Mesenchymal_05
Mesenchymal_01
Mesenchymal_04
Mesenchymal_10
Mesenchymal_11
Mesenchymal_13
Mesenchymal_09
Pbx1_Down_LS
Hand2_Up_LS
Hand2_Down_LS
Hand2_dorothea
Pbx1_Up_LS
Pbx1_Pbx2_dorothea
Column Z-score
D
Prrx1
Msx1
PBX1
HAND2
ATAC-seq
H3K27ac
H3K27me3
Placental Cons

985

Figure 5. Target genes co-regulated by PBX1/2 and HAND2 are essential for patterning the hindlimb bud.

(A) Bulk RNAseq integration of datasets from E10.5 hindlimb buds of *Pbx1cKO^{Mes}*;*Pbx2$^{-/-}$* and *Hand2cKO^{Mes}* mutants *versus* respective controls. (Left) Pie chart indicating proportion of DEGs in both *Pbx1/2* and *Hand2* mutant hindlimb buds (Both) *versus* littermate controls, or in hindlimb buds from only one of the mutant genotypes (*Pbx1/2*-only and *Hand2*-only) *versus* littermate controls. Number of gene transcripts in each category indicated in brackets. (Right) Intersection of bulk RNAseq with scRNAseq. (Upper Panel) Within boxes, DEGs from both *Pbx1/2*-deficient and *Hand2*-deficient hindlimb buds and also co-expressed in the same cells as *Pbx1/2* and *Hand2* (AUC > 0.55). Transcription factors with known essential roles in limb patterning/morphogenesis highlighted by red asterisk. (Lower Panel) Dot plot of identified target genes showing their expression in the mesenchymal clusters identified by scRNAseq. Dot size represents proportion of cells within a given population that expresses the gene; color intensity indicates average expression level. (B) UMAP in which each cell is color-coded according to the expression of *Pbx1/2 and Hand2* target genes inferred based on RNAseq datasets from *Pbx1/2*- and *Hand2*-mutant hindlimbs. (C) Heatmap showing relative (z-score) median expression of different subsets of *Pbx1/2 and Hand2* target genes in scRNAseq mesenchymal subclusters. (D) UCSC genome browser tracks encompassing *Prrx1* and *Msx1* loci. Tracks: PBX1 and HAND2 ChIPseq, red and green, respectively; ATACseq, aqua; H3K27ac and H3K27me3, brown and blue, respectively; placental conservation, dark purple. Grey rectangles highlight predicted CRMs bound by PBX1 and HAND2.

Figure 6



G

GENOTYPE	SAMPLE NUMBER	DIGIT NUMBER	TIBIA/FIBULA PHENOTYPE	FEMUR PHENOTYPE	PELVIS
Control (Heterozygous or WT)	n=8	Normal	Normal	Normal	Normal
Hand2$^{+/-}$;Hoxb6$^{Cre/+}$ or Hand2$^{f/-}$	n=4	Normal	Normal	Normal	Normal
Pbx1$^{f/f}$;Hoxb6$^{Cre/+}$	n=4	Normal	Normal	Shorter femur (n=4)	Small and malformed pelvis (shorter ilium and foramen) (n=4)
Hand2$^{f/-}$;Hoxb6$^{Cre/+}$ (See Fig. 1L)	n=4	1 digit (n=3); 4 digits (n=1)	Tibia/Fibula fused and malformed (n=3)	Shorter femur (worse than *Pbx1Δc/Δc*) (n=4)	Small pelvis (n=4)
Pbx1$^{f/f}$;Hand2$^{+/-}$;Hoxb6$^{Cre/+}$	n=7	2 digits (n=1); 3 digits (n=2); 4 digits (n=2); 5 digits (n=2)	Fibula (defective or lost) (n=6)	Shorter femur (similar to *Pbx1Δc/Δc* or worse) (n=6)	Small and malformed pelvis (shorter ilium and foramen) (n=6)

Figure 6. *In vivo* genetic interaction of *Pbx1* and *Hand2* directs patterning of posterior hindlimb skeletal elements.

(A-F') Hindlimb skeletons at E14.5 (blue: cartilage; red: bone). Digits shown from anterior (digit 1) to posterior (digit 5). (A, A') Representative example of control embryo (Wild-type, WT) (n=8). (B, B') *Hand2$^{+/-}$;Hoxb6$^{Cre/+}$* hindlimbs with normal morphology. (C, C') *Pbx1$^{f/f}$; Hoxb6$^{Cre/+}$* hindlimbs with pelvic malformations and shorter femur (n=4/4). (D-F,D'-F') *Pbx1$^{f/f}$;Hand2$^{+/-}$;Hoxb6$^{Cre/+}$* hindlimbs with malformed pelvis, shorter or truncated femur, defects or loss of fibula, hypoplastic or absent tarsals and metatarsal, and variable loss of posterior digits (n=6/7). Autopod distal-most abnormalities (in tarsal, metatarsal, and phalanges) not observed in either *Pbx1$^{f/f}$;Hoxb6$^{Cre/+}$* or *Hand2$^{+/-}$;Hoxb6$^{Cre/+}$* mutants. Scale bar: 500μm. (G) Table listing total number of embryos analyzed and number of embryos with hindlimb skeletal defects shown in panels A-F.

Figure 7

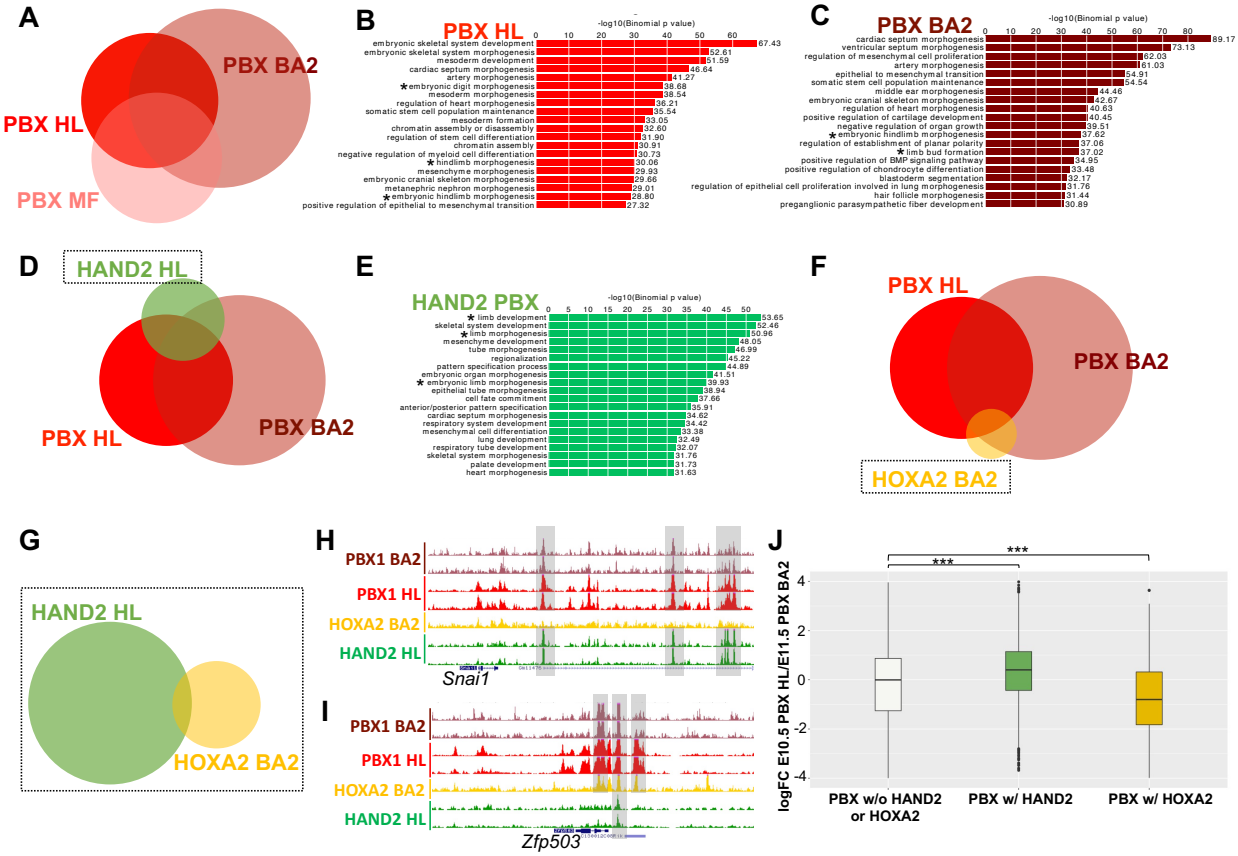


Figure 7. HAND2 selects different subsets of PBX1 peaks to confer early limb patterning functions to PBX1.

(A-C) Analysis of genome-wide PBX1 (PBX) binding in E10.5 hindlimb (HL), E11.5 second branchial arch (BA2) and E11.5 midface (MF). Overlap of PBX-bound peaks in hindlimb, BA2 and MF shown in (A). Venn diagram highlighting large overlap between PBX-bound regions across different embryonic tissues (see percentages, Supplementary Table 5). Bar plots showing enriched GO terms associated with genes involved in developmental processes (GREAT) of PBX-bound regions in HL (B) and BA2 (C). Limb development and morphogenesis GO categories (*). Size of bars corresponds to binomial raw (uncorrected) p-values (X axis values). (D) Venn diagram of PBX-bound regions in HL and BA2, with HAND2-bound regions in BA2. (E) Venn diagram of PBX-bound regions in HL and BA2, with HAND2-bound regions in BA2. (F) Overlap of PBX-bound regions in HL and BA2, with HOXA2-bound regions in BA2. (G) Venn diagram of regions bound by HOXA2 in BA2 with those bound by HAND2 in HL. (H,I) UCSC genome browser tracks (mm10) showing the generated ChIPseq datasets from PBX1 BA2 (brown); PBX1 HL (red); HOXA2 BA2 (yellow); and HAND2 HL (green) at the *Snai1* (H) and *Zfp503* (I) loci. Shared peaks highlighted by grey rectangles. (J) Boxplots of ratio (log2Fold Change;

1035 logFC) of normalized PBX1 ChIPseq signals between HL and BA2 at indicated subsets of PBX-bound
1036 regions.

Figure 8

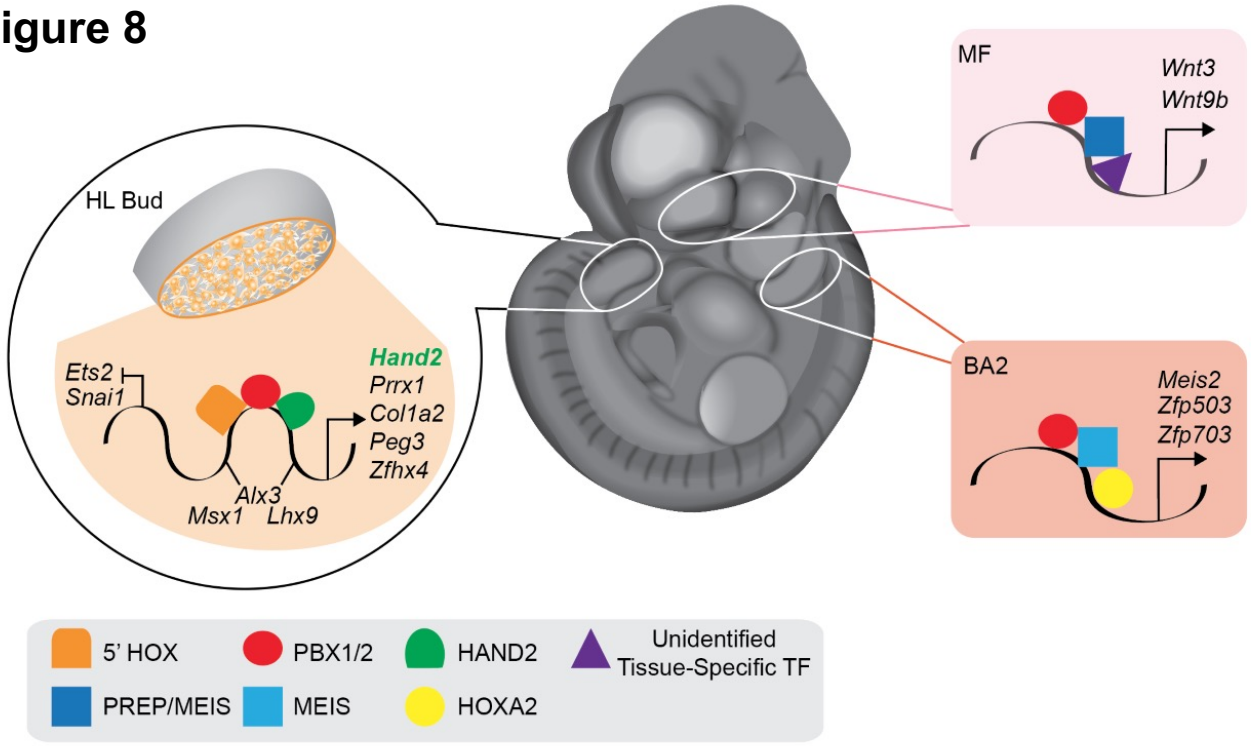


Figure 8. Promiscuous genome-wide PBX binding acquires tissue-specific developmental functions via interactions with distinct cofactors.

Model depicting how PBX homeoproteins must rely on DNA binding partnerships with different cofactors to execute context-dependent developmental functions. In the midface (MF), a "*Hox*-less" domain, PBX1/2 together with PREP/MEIS bind regulatory elements of *Wnt3* and *Wnt9b* (pink-shaded area) to direct their expression during upper lip/primary palate fusion[58]. In the second branchial arch (BA2), PBX1/2 activate a BA2-specific transcriptional program by binding the *Zfp503* and *Zfp703* loci with HOXA2 and MEIS[37] (ochre-shaded area). In restricted subset of posterior hindlimb bud mesenchymal cells (orange shading reflects co-expression of *Pbx1/2*, red, and *Hand2,* green), a PBX1/2-directed GRN acquires early limb patterning functions via interaction with HAND2. This spatio-temporally constrained GRN (orange-shaded area) activates known limb transcriptional regulators, such as *Prrx1* and *Hand2* itself, and represses other limb transcription factors, such as *Snai1* and *Ets2*. In contrast, PBX1/2 and HAND2 regulate other target genes (*Alx3, Msx1, Lhx9*) in a discordant manner.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- MergedSupplMaterials03142022NatureCommun.pdf
- SupplementaryTable2.xlsx
- SupplementaryTable4.xlsx