Title Page

Title: Dysregulation in Spatiotemporal Expression of *HMX1* Coordinated by Multifaceted Enhancers Drives an Auricular Disorder

Authors: Xiaopeng Xu[1,2,*], Qi Chen[3,*], Qingpei Huang[1,*], Timothy C. Cox[4,*], Hao Zhu[2], Jintian Hu[3], Xi Han[1], Ziqiu Meng[2], Bingqing Wang[3], Zhiying Liao[1], Wenxin Xu[1,5], Baichuan Xiao[2], Ruirui Lang[2], Jiqiang Liu[2], Qiang Li[6], Qingguo Zhang[3], Stylianos E. Antonarakis[7,8,9], Jiao Zhang[10,#], Xiaoying Fan[1,#], Huisheng Liu[1,#], Yong-Biao Zhang[2,11,#]

Affiliations:

[1] School of Biomedical Engineering, GMU-GIBH Joint School of Life Sciences, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou Medical University, Guangzhou National Laboratory, Bioland Laboratory, Guangzhou, 510320, China

[2] School of Bioengineering Medicine, Beihang University, 100191, China

[3] Department of Ear Reconstruction, Plastic Surgery Hospital, Chinese Academy of Medical Sciences, Beijing 100144, China

[4] Departments of Oral & Craniofacial Sciences, School of Dentistry, and Pediatrics, School of Medicine, University of Missouri-Kansas City, Kansas City, USA

[5] Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei, 230000, China

[6] Department of Plastic Surgery, Affiliated Hospital of Xuzhou Medical University, Xuzhou, 221000, China

[7] Department of Genetic Medicine and Development, University of Geneva Medical Faculty, Geneva 1211, Switzerland

[8] Medigenome, Swiss Institute of Genomic Medicine, 1207 Geneva, Switzerland

[9] iGE3 Institute of Genetics and Genomes in Geneva, Geneva, Switzerland

[10] Shandong collaborative innovation research institute of traditional Chinese medicine industry

[11] Key Laboratory of Big Data-Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology

Correspondence:

Yong-Biao Zhang, Address: 37 Xueyuan Road, Haidian District, Beijing, 100191, China, Tel: +86 13522471691, Email: zhangyongbiao@gmail.com

Huisheng Liu, Address: 9th Xingdaohuan North Road, Guangzhou International Bio Island, Haizhu District, Guangzhou, Guangdong, 510320, China, Email: liu_huisheng@gzlab.ac.cn

Xiaoying Fan, Address: 9th Xingdaohuan North Road, Guangzhou International Bio Island, Haizhu District, Guangzhou, Guangdong, 510320, China, Email: fan_xiaoying@gzlab.ac.cn

Jiao Zhang, Address: Minghu International Cell Medicine Industrial Park, Jingshi Road, Lixia District, Jinan, 250000, China, Email: joycezhang1978@hotmail.com

*The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint first authors

Abstract

Enhancers, through the combinatorial action of transcription factors (TFs), dictate both the spatial specificity and the levels of gene expression, and their aberrations can result in diseases. The association between *HMX1* and its downstream enhancer with ear deformities has been previously documented. However, the pathogenic variations and molecular mechanisms underlying the bilateral constricted ear (BCE) malformations remain unclear. This study identifies a copy number variation (CNV) encompassing three enhancers that induces BCE. These enhancers, collectively termed the positional identity hierarchical enhancer cluster (PI-HEC), co-regulate spatiotemporal expression of the *HMX1* gene in a coordinated and synergistic mode, each displaying unique activity-location-structure characteristics. A thorough exploration of this regulatory locus reveals that specific motif clusters within the PI-HEC variably modulate its activity and specificity, with the high mobility group (HMG) box combined with Coordinator and homeodomain (HD)-TFs notably influencing them respectively. Our findings based on various types of mouse models, reveal that both aberrant *Hmx1* expression in the fibroblasts of the basal pinna, originating from neural crest cells, and ectopic expression in the distal pinna structures contribute to the abnormal development of the outer ear, including the cartilage, muscle, and epidermis tissues. This study deepens our understanding of mammalian ear morphogenesis and sheds light on the complexity of gene expression regulation by enhancers and specific sequence motifs.

Introduction

Investigating the intricate morphological changes during mammalian embryogenesis remains a pivotal focus in developmental biology, a discipline that has recently seen significant progress [1]. The mammalian ear, a complex structure originating from cranial neural crest cells (CNCCs) and all three germ layers, exemplifies the complexity of developmental biology. It showcases a unique interplay of developmental processes and evolutionary adaptation[2,3]. The outer ear, essential for capturing sound waves and implicated in conductive hearing loss when dysfunctional, is less explored than its middle and inner counterparts[3]. This part, comprising the tragus and helix, is not only important for auditory function but also demonstrates considerable morphological diversity throughout evolution as well as in many types of diseases[4-7]. Unraveling the molecular mechanisms underlying this diversity is an ongoing field of investigation.

Bilateral constricted ear (BCE) often presents as an autosomal dominant trait and primarily manifests as curling of the upper portion of the external ear, affecting structures such as the helix, scapha, and antihelix. Comparative genetic studies across species, including sheep, cows, mice, and rats, have identified a correlation between BCE-like phenotypes and a non-coding regulatory region adjacent to the *Hmx1* gene that contains an evolutionarily conserved region (ECR)[8,9]. In humans, phenotypic presentations similar to those in concha-type microtia have been linked to genomic duplications encompassing the ECR near *HMX1*[10]. Notably, in rats, a 5.7-kb deletion that includes this ECR significantly diminishes *Hmx1* expression, leading to a dumbo ear phenotype, highlighting a critical regulatory role[11]. Furthermore, in human concha-type microtia, several genomic duplications exist that extend beyond the ECR, suggesting a complex and multifaceted genetic mechanism underlying outer ear development. The spectrum of pathogenic mutations involving the

ECR, including both duplications and deletions, emphasizes the complex interplay of gene expression in determining ear morphology.

Elucidating the function of cis-regulatory elements in specific cellular and temporal settings is key to unraveling the complex mechanisms of ear development and regulatory variations that influence the external ear phenotype[12,13]. Recent research underscores CNCCs as vital in craniofacial development, including the outer ear[14-16]. CNCCs positional identity is determined by a complex set of signaling factors that govern a distinct hierarchy of gene expression profiles. For example, positional identity along the anterior-posterior axis is influenced by bivalent gene expression such as *Hoxa2* in the second pharyngeal arch and *Hand2* in the mandibular arch[17]. Along the proximo-distal axis, gene expression of *Dlx* homeobox family members, for example, governs patterning within the mandibular arch. Patterning along all axes is crucial for proper craniofacial morphogenesis[18]. This highlights the importance of precise CNCCs positional identity[19]. For instance, ectopic *Hoxa2* expression in the pharyngeal arch 1 (PA1) results in mirror-image pinna duplication[7], while the loss of *Dlx5* and *Dlx6* in the mandibular arch results in the transformation of the lower jaw into a mirror image of the upper jaw components[20]. Single-cell RNA sequencing (scRNA-seq) studies of the mandibular arch further revealed a wide range of heterogeneous cellular domains (intra-arch identity), with disturbances leading to specific mandible and tooth defects[21]. Yet, the exact regulatory mechanisms of spatial gene expression in the craniofacial region, including *HMX1*, and within individual arches, are not fully understood. This knowledge gap limits our understanding of the genetic and epigenetic dynamics in human craniofacial morphogenesis.

This investigation presents a comprehensive analysis of human pedigrees with BCE, employing gene chips and next-generation sequencing to pinpoint a pathogenic locus, designated as

the BCE locus on chromosome 4. This locus encompasses an enhancer cluster, comprising three synergistically acting enhancers identified in human craniofacial tissues and *in vitro*-derived human CNCCs (hCNCCs). Transgenic LacZ mouse experiments revealed that the spatial specificity of the *Hmx1* gene expression in pharyngeal arch 2 (PA2) during ear development is governed by a hierarchical interplay among these enhancers, forming what we term the Positional Identity Hierarchical Enhancer Cluster (PI-HEC). In-depth examination of the dominant enhancer in PI-HEC unraveled a complex network of transcription factors (TFs) modulating enhancer specificity and function. Gene-editing in mouse models elucidated the link between pinna development and altered *Hmx1* expression, exposing a dysregulated molecular network affecting pinna structures such as cartilage, muscle, and epidermis. Our findings significantly advance the understanding of the genetics and regulatory mechanisms behind human outer ear development, providing novel molecular insights into the etiology of pinna development and the presentation of BCE.

Results

Copy number duplications detected in BCE

We explored the genetic basis of BCE, a condition noted in seven Chinese families showing autosomal dominant inheritance, using classical linkage analysis with genechip genotyping (**Fig. 1a and 1b**). Linkage analysis on pedigrees 1, 3, and 6 yielded a significant combined parametric LOD score of 7.8 (**Extended Data Fig. 1a**), with individual scores of 2.7, 1.8, and 3.3 respectively (**Extended Data Fig. 1b**). This led to the mapping of the susceptible locus to a 10.2 cM region on chromosome 4 (**Extended Data Fig. 1b**).

For fine mapping of this locus, target-capture sequencing was performed on 32 individuals across the three pedigrees. This detailed analysis revealed copy number duplications co-segregating in all sequenced patients from the three pedigrees (**Extended Data Fig. 1c**). The specific genomic ranges of these duplications were chr4:8691119-8795306, chr4:8679801-8750713, and chr4:8668379-8739905 (hg19) for each respective pedigree. Further investigations using chip-based methods and whole-genome sequencing in the remaining four pedigrees confirmed that probands in each of these pedigrees also carried copy number duplications in this non-coding region (**Fig. 1c**). Ultimately, our findings converged on a minimum genomic region of overlap of chr4:8691119-8726565 (termed the BCE locus), consistently present across all seven BCE pedigrees. Hi-C data from human embryo head tissue of CS17 stage[22], revealed the TAD structure around the BCE locus, suggesting that *HMX1* and *CPZ* are potential target genes (**Fig. 1c**).

The BCE locus, containing a cluster of enhancers, directly interacts with the *HMX1* promoter

Previous studies have identified a conserved enhancer within the BCE locus that participates in the regulation of *Hmx1* gene expression across different mammals[8,11]. However, understanding of the

full range of cis-regulatory elements at this locus, including their context-dependent characteristics and their role in controlling pinna development, remains incomplete. Given the spatiotemporal specificity of cis-regulatory elements, it is imperative to study them within embryonic contexts that are closely associated with pinna development. Studies in mice indicated that CNCCs are important for pinna organogenesis[7], and both murine studies[11] and single-cell sequencing data of human embryos[23] also indicated that *HMX1* is predominantly expressed in the craniofacial mesenchyme, including the PA1/2 and PA3/4 regions (**Extended Data Fig. 2a**). These data underscore the necessity of concentrating our research on CNCCs and craniofacial mesenchyme to thoroughly investigate cis-regulatory elements within the BCE locus and their roles in pinna development (**Extended Data Fig. 2a**).

To delineate the cis-regulatory elements at the BCE locus, we interrogated publicly available ChIP-seq data from both *in vivo* embryonic human head tissues (CS13-CS17)[24] and *in vitro*-derived hCNCCs[25]. This approach revealed a cluster of potential enhancers (termed EC1, EC2, and EC3) within the BCE locus across these datasets: EC1 and EC2 are both active in two datasets, EC3 was defined as an active enhancer *in vivo* but possesses weak histone signals *in vitro* (**Fig. 2a**). Consistent with findings in the human datasets, analogous sequences in publicly available mouse CNCCs (mCNCCs) data[17] were also identified as an enhancer cluster, notably in PA2 (**Extended Data Fig. 2b**). Comparative analyses employing Promoter Capture Hi-C (PC-HiC) revealed that the enhancer cluster region actively interacts with the *HMX1* promoter in hCNCCs, but not in human Embryonic Stem Cells (hESCs) (**Fig. 2b and Supplementary Table 1**). This observation suggests that the activity of the enhancer cluster is temporally regulated and cell-type specific, favoring *HMX1* and excluding interaction with the adjacent gene *CPZ*. Publicly available mouse PC-HiC data[26] further

159 revealed that the BCE locus interacts with the *Hmx1* promoter in PA2 tissue of E10.5 and pinna
160 structures of E12.5 and E14.5 (**Extended Data Fig. 2c**). The collective evidence suggests a
161 conserved role of the enhancer cluster in modulating *HMX1* expression in CNCCs, necessitating
162 further clarification of how enhancers within this complex coordinate the spatiotemporal pattern of
163 *HMX1* expression.

164 **Different intrinsic activities of the three enhancers at the BCE locus during hCNCCs**
165 **differentiation**

166 To investigate the spatiotemporal modulating patterns of the individual candidate enhancers hEC1,
167 hEC2, and hEC3 within the enhancer cluster, we further conducted a detailed analysis of epigenetic
168 datasets of two developmental stages of *in vitro* derived hCNCCs[25,27]: early (D11), representing
169 cells migrating out of the neural rosettes, and late (P4), representing post-migratory mesenchymal
170 cells (**Fig. 3a**). Active enhancer markers, P300 and H3K27ac, revealed distinct profiles: hEC1
171 displayed significant activity in late hCNCCs; hEC2 was divided into two regions, hEC2.1 and
172 hEC2.2, based on P300 signals; hEC2.1 showed strong H3K27ac activity at both stages, whereas
173 hEC2.2 exhibited weaker H3K27ac activity, indicating a less active state at both stages; hEC3,
174 despite showing strong accessibility in late hCNCCs and interaction with the *HMX1* promoter (**Fig.**
175 **2b**), lacked H3K27ac signals (**Fig. 3a**).

176 To evaluate the capacity of the candidate enhancers to drive reporter gene expression, we
177 adapted a well-established *in-vitro* differentiation system[28], further incorporating retinoic acid (RA)
178 at the late hCNCCs stage to induce transformation into PA (pharyngeal arch)-like hCNCCs
179 characteristic (**Fig. 3b and Extended Data Fig. 3**). *HMX1* gene expression initially rose
180 approximately 4-fold at the early hCNCCs stage, and subsequently surged to about 140-fold at the

late hCNCCs stage and 110-fold at the PA-like hCNCCs stage (**Fig. 3c**). In luciferase assays, none of the enhancers showed activity in H9 cells, consistent with the inaccessibility of these enhancers in hESCs (**Fig. 3d and Fig. 2b**). However, in PA-like hCNCCs, EC1 was a strong transcriptional activator (**Fig. 3e**), while hEC2.1 and hEC2 demonstrated weaker activity than hEC1 (**Fig. 3e**). Notably, hEC2.2 and hEC3 remained inactive in all tested contexts, consistent with their weak H3K27ac signals (**Fig. 3d and Fig.3e**). These findings suggest that hEC1 is a strong enhancer, hEC2 is a weak enhancer and hEC3 is primed for activation but it is not currently active or it may serve as a structural element to facilitate the contact between the BCE locus enhancer cluster and the *HMX1* promoter[29].

In a further experiment, we created knockout (KO) hESC lines with simultaneous deletions of hEC1 and hEC2 (**Fig. 3f, upper panel, and Extended Data Fig. 4a**). Both wild-type and KO hESC lines exhibited similar differentiation capacities into hCNCCs, and the expression levels of two marker genes, *SOX9* and *TWIST1*, were similar in both lines (**Extended Data Figs. 4b and 4c**). In the differentiation of hCNCCs to PA-like cells, *HMX1* expression remained similar between wild-type and KO cells at the H9 state, but diverged noticeably at the early hCNCCs stage, becoming more pronounced in the late and PA-like hCNCCs stages (**Fig. 3f, bottom panel**). In summary, our results establish the relationship between *HMX1* expression and enhancer activities in the BCE locus in different stages of hCNCCs development.

Hmx1 gene patterning along pinna development is related to epistatic interactions of the three enhancers in the BCE locus

We employed the whole-mount *in-situ* hybridization (WISH, E9.5) and transgenic Hmx1-P2A-EGFP mouse reporter line (E11.5 and E14.5) to elucidate the spatiotemporal dynamics of *Hmx1*

expression during embryonic ear development. Specifically, we focused on three developmental stages: embryonic day 9.5 (E9.5), representing early CNCCs development stage in PAs; E11.5, reflecting CNCC-derived mesenchyme development stage in PAs; and E14.5, corresponding to the initial stages of pinna development (**Fig. 3g and Extended Data Fig. 4j**). We observed the initiation of *Hmx1* expression at E9.5 in the superior region of PA1. By E11.5, this expression expanded to the eye, craniofacial mesenchyme, part of PA2, and the dorsal root ganglia (DRG). By E14.5, *Hmx1* predominantly localized to the lower part of the outer ear (lp), the basal pinna zone (bp), craniofacial mesenchyme (cm) and the proximal region of the mandibular prominence (pmp) (**Fig. 3g**).

In the LacZ assays assessing enhancer activity during mouse development, distinct patterns emerged for each enhancer. For hEC1, no activity was observed at E9.5. However, by E11.5, strong signals appeared in the proximal region of PA2, expanding to nearly the entire pinna, excluding the upper distal region, by E14.5 (**Fig. 3h and Extended Data Fig. 4d**). In contrast, hEC2.1 displayed only weak activity in the basal PA2 region and head at E9.5, with no detectable activity at later timepoints (**Fig. 3i and Extended Data Fig. 4e**). For hEC2, weak activity was noted in the distal PA1 and PA2 regions, head, and tail at E9.5. This activity persisted weakly in the distal PA2 region and hindlimbs at E11.5, and in the proximal region of the mandibular prominence, head, and hindlimbs at E14.5 (**Fig. 3j and Extended Data Fig. 4f**). hEC3 demonstrated two distinct patterns: one exhibited partial activity in proximal PA2 and a small area below PA2 (1/5 of samples), while the other showed activity in the limbs (4/5 of samples) (**Fig. 3l and Extended Data Fig. 4h**). These results were somewhat similar to those using cellular models; thus, hEC2 appears to be active in early CNCCs development in PAs, whereas hEC1 activity is initiated in CNCC-derived mesenchyme stage in PAs, with both maintaining activity through craniofacial development.

To enable a detailed comparison between the spatial specificity of the *Hmx1* gene expression at E11.5 and the LacZ staining patterns of three enhancers, we divided the *Hmx1* expression around the PA2 zone into three distinct regions: α (proximal PA2 region), β (distal PA2 region), and γ (caudal to PA2) (**Fig. 3g, right panel**). Our analysis elucidated the hierarchical relationship among the three enhancers: hEC1 exerts a predominant influence on regions α and, to a lesser extent, γ; hEC2 is primarily associated with region β; and hEC3 appears to affect region γ, albeit with an inconsistent pattern. (**Fig. 3h, j, l**). This led us to hypothesize that hEC1, while driving significant transcriptional activity, may not fully convey positional information. In contrast, hEC2 and hEC3, despite their weaker transcriptional activities, could provide the additional positional information required for precise *Hmx1* localization within the PA2 zone (intra-arch identity). To investigate this particular gene patterning regulation, we constructed two additional vectors: hEC1+hEC2 and hEC1+hEC2+hEC3, aiming to study the epistatic interactions among the enhancer cluster in the BCE locus (**Fig. 3k and 3m**). Notably, the pattern exhibited by the hEC1+hEC2 combination was distinct, showing reduced overall activity at E11.5 and a more defined spatial activity in the frontonasal prominence (FNP) and limbs. Crucially, it modulated the spatial position within PA2 (**Fig. 3k and Extended Data Fig. 4g**), encompassing most *Hmx1* expression regions, including α, β, and part of γ (**Fig. 3k**). The hEC1+hEC2+hEC3 combination further extended this pattern, closely mirroring the *Hmx1* spatial expression across regions α, β, and γ (**Fig. 3m and Extended Data Fig. 4i**). At E14.5, the staining pattern of hEC1+hEC2+hEC3 closely matched the spatial *Hmx1* expression around the pinna region, especially in the lower pinna part adjacent to the proximal mandibular region (**Fig. 3m**), correlating with areas of strong *Hmx1* expression (**Fig. 3g**).

These results indicate that spatial specificity of *Hmx1* expression around PA2 is finely tuned by the coordinated and synergistic enhancers within an enhancer cluster, which we term the Position Identity Hierarchical Enhancer Cluster (PI-HEC): hEC1 is responsible for most transcriptional output, while hEC2 and hEC3 provide the supplementary positional information for the transcriptional output.

Regulatory dynamics of motif clusters within human EC1 responsible for activity and specificity

Although the three enhancers within the PI-HEC coordinately and synergistically regulate the spatiotemporal expression characteristics of *Hmx1*, we also found that hEC1 possesses the dual attributes of dictating the primary activity and proximal regional identity. Additionally, building on our previous research with the orthologous mouse sequence (mEC1) of hEC1, which highlighted a crucial 32 bp sequence regulated by the Hox-Pbx-Meis complex[11], we extended our analysis to hEC1. To delve deeper into the regulatory dynamics, we targeted five motif clusters (D1, D2, D3, D4, and D5 of hEC1), identified by TF-binding motif analysis using FIMO (**Fig. 4a and Supplementary Table 3**). The impacts of deleting these individual clusters were evaluated using both *in vitro* (PA-like hCNCCs) and *in vivo* (LacZ assay) methods.

In vitro luciferase assays showed that deletions of D2, D3, D4, and D5 significantly diminished hEC1 activity, with D4 almost completely eliminating this activity (**Fig. 4b, top panel**). The *in vivo* LacZ assay in mouse embryos revealed similar trends but also some notable differences (**Fig. 4b, bottom panel**). The D1 deletion, while not affecting activity in the cell line, resulted in almost complete loss of hEC1 activity in PA2 and other embryonic regions. Deletions of D2 and D5 led to substantial reductions in staining, similar to the luciferase results (**Fig. 4b, bottom panel and**

Extended Data Fig. 5a). Interestingly, the D3 deletion, despite reducing activity *in vitro*, maintained visible LacZ staining across the embryo, with a spatial shift in the PA2 region from ventral to dorsal and ectopic PA1 staining. The D4 deletion eliminated staining throughout the embryo, in accordance with the cell line luciferase results.

These observations emphasize the complex regulatory mechanisms that control both the activity and specificity of the hEC1 enhancer. Concurrently, the distinct relationships of the five motif clusters for the activity and spatial specificity of gene expression provide a valuable model for elucidating the regulatory rules governing enhancers in pinna development[13].

hEC1 activity and specificity is coordinately controlled by multiple TFs

To investigate the proteins that bind to hEC1, we performed *in vitro* DNA pull-down assays using E11.5 PA2 tissue nuclear proteins and biotinylated hEC1 (**Extended Data Fig. 5b**). Proteomic analysis via LS/MS (Liquid Chromatography-Mass Spectrometry) highlighted homeodomain TFs (HD-TFs) such as Meis1, Pbx1/3, Dlx6 as the predominant hEC1-binding family of proteins (**Extended Data Fig. 5c and Supplementary Table 4**). Considering the lower sensitivity of this *in vitro* experiment and the prominence of a 'Coordinator' motif, a 17-bp long DNA sequence encompassing basic helix-loop-helix (bHLH)- and HD-binding motifs in hCNCCs[28,30], we further examined our motif prediction near the HD-TFs, and found that a pattern containing HD, high mobility group (HMG) box and Coordinator (predicted with TWIST1 motif) sequences are present in the D1 (DLX1-HD) and D3-D4 region (MEIS1/PBX1-HD) (**Fig. 4c and Supplementary Table 3**). Therefore, the discrepancy between the luciferase assay in the cell line and the transgenic LacZ regarding the effect of D1 deletion on hEC1 activity may result from the difference in the *DLX* gene family expression between these two systems (**Extended Data Fig. 3e**).

Here, we focus on the conserved D3-D4 region, containing three high-affinity binding sites (*P* value ≤ 1E-04) including MEIS1/PBX1, TCF7L1/2 and Coordinator and one low-affinity binding site (*P* value > 1E-04) of ALX1/4, which are contained within the larger 76 bp duplications associated with auricular malformations in sheep and cow microtia (**Fig. 4c**). To validate the TF-binding events, we performed CUT&RUN assays in *in vitro* derived hCNCCs targeting TCF7L2 and reanalyzed one publicly available TWIST1 CUT&RUN dataset in hCNCCs[30], and confirmed their binding capacity (**Fig. 4d and Extended Data Fig. 5d and 5e**). We then further applied enhanced chromatin occupancy (EChO) analysis, which identifies TF interaction foci[31], to explore TCF7L2 binding patterns in the hEC1 region. One high-scoring motif was found in the D4 region, consistent with the prediction result (**Fig. 4e and Supplementary Table 5**). The binding capacity of human MEIS1 protein in hEC1 was also validated using the EMSA method (**Fig. 4f**).

We further constructed two deletion regions, D3-1 (including HD motifs) and D4-1 (including HMG box and Coordinator motif) (**Fig. 4c**), with luciferase assay results demonstrating that removal of D4-1 dramatically suppressed the activity of hEC1 (**Fig. 4g**). *Meis1*, *Tcf7l2*, *Twist1/Tcf4*, that interacted with the D4-1 sequence, were all confirmed to be expressed in the PA2 region using public mCNCCs bulk RNA-seq and WISH datasets (**Fig. 4h and i**). Collectively, these findings suggest that the activity and specificity of hEC1 are modulated by a combination of different families of TFs. The underlying regulatory code resides within the characteristics of the enhancer sequences offers a complex process for pinna development.

Impact of spatial *Hmx1* expression on pinna development: insights from transgenic mouse models mimicking human BCE anomalies.

To investigate human BCE development mechanisms, we created a transgenic mouse model, *mEC1*$^{dup/dup}$, containing four copies of mEC1 sequences (**Fig. 5a, 5b and Extended Data Fig. 6a**). LacZ reporter activity for mEC1 in the pinna region was similar to that of hEC1 but less intense (**Fig. 3i, 5c and Extended Data Fig. 6b**). *mEC1*$^{dup/dup}$ mice showed significant pinna development abnormalities compared to wild-type (WT) mice (**Fig.5d, 5e, 5g**). The helix of the pinna was notably affected, with a reduced area (**Fig. 5e, 5f**), and the pinna angle changed, suggesting altered cartilage or muscle formation (**Fig. 5g, 5h**). Older transgenic mice exhibited progressive eye dysgenesis, though less consistent than ear malformations (**Extended Data Fig. 6c**). WISH showed spatially expanded *Hmx1* expression in the PA2 region at E10.5 and in the upper pinna region at E14.5 in the *mEC1*$^{dup/dup}$ mice (**Fig. 5i, 5j**), well correlating with the activity of mEC1 (**Fig. 5c**). These findings underscore the impact of enhancer copy number variation in developmental morphology, especially in pinna formation.

To explore the interplay between EC1 and *Hmx1* expression in CNCCs in ear development, we used multiple mouse models disrupting *Hmx1* function in various ways: *Wnt1::Cre;Hmx1*$^{fl/fl}$ (conditional *Hmx1* knockout in mCNCCs), *mEC1*$^{del/del}$, and *dumbo* (mutation in exon 1 of *Hmx1* resulting in truncated protein after residue Gln65[32]) (**Fig. 5k and Extended Data Fig. 6d**). The pinna morphology in these models exhibits similar low set, laterally protruding characteristics (**Fig. 5k**), suggesting that mEC1 is crucial for proper *Hmx1* expression in mCNCCs and normal pinna development. Finally, performing micro-CT analysis on all mouse models revealed that the normal development of the paroccipital process, a conical prominence of bone adjacent to the outer ear and serving as an attachment point for certain neck muscles, was inhibited in the *dumbo*,

Wnt1::Cre;Hmx1$^{fl/fl}$ and *mEC1*$^{del/del}$ mice but not in *mEC1*$^{dup/dup}$ (**Fig. 5l**), which may be related to the upright position (that is laterally protruding characteristics) of the external ear.

EC1 duplication results in ectopic expression of *Hmx1*, leading to disrupted cell differentiation in ear development

To gain greater insight into the mechanisms underlying abnormal pinna development in the *mEC1*$^{dup/dup}$ mouse model, we performed bulk RNA sequencing at the E14.5 embryonic stage in the whole pinna prominence (**Fig. 6a**). The duplication of the mEC1 led to overexpression of *Hmx1*, but the majority of differentially expressed genes (DEGs) were downregulated. This suggests that increased expression of *Hmx1* plays a critical role in suppressing numerous genes during pinna development (**Fig. 6a and Supplementary Table 7**). Gene Ontology (GO) enrichment analysis of these downregulated DEGs showed their primary involvement in pathways essential to ear development, such as epidermis development, extracellular collagen organization, and myofibril assembly (**Fig. 6b and Supplementary Table 8**). In addition, two cartilage development-related genes *Comp* and *Arhgap36* were also downregulated, although they did not rank high in the GO enrichment analysis (**Fig. 6a and 6b**). Masson's trichrome staining at E14.5 revealed delayed development of muscle and epidermis in the outer ear of *mEC1*$^{dup/dup}$ mice compared to that in wild-type mice, consistent with the bulk RNA-seq results (**Fig. 6c**). However, *Hmx1* is typically expressed in the craniofacial mesenchyme, which is not a source for muscle and epidermis in the pinna[33]. Thus, in conjunction with the WISH results (**Fig. 5j**), we hypothesize that the defects observed in the outer ear are due to ectopic expression of *Hmx1* in other pinna zones.

To definitively characterize the cell types expressing the *Hmx1* gene within the outer ear, and to further elucidate the molecular mechanisms involved in BCE, single-cell lineage tracing was

354 employed in Hmx1-P2A-GFP mice at pivotal developmental stages: E10.5, E12.5, and E14.5
355 (**Extended Data Fig. 7a**). Primarily, the *Hmx1*$^{+}$ cells were identified as fibroblasts, with a small
356 fraction defined as other cell types including neurons, myogenic cells, and chondrocytes also
357 detected (**Fig. 6d and Extended Data Fig. 7b**). The fibroblasts were further categorized into seven
358 subtypes based on distinct marker genes. These include skeletal muscle-related fibro (*Ogn*, *Mest*)[34],
359 two fibro-adipogenic progenitors (*Enpp2*$^{+}$, *Osr1*$^{+}$)[35,36], dermal fibro progenitors (upper: *Crabp1*,
360 *Twist2*; lower: *Malat1*, *Meg3*)[37], miotic CNCC-derived mesenchymal cells (*Ube2c*, *Top2a*)[38], and
361 perimysial cells (*Dcn*, *Lum*)[34] (**Fig. 6e and Extended Data Fig. 7c**). In the developmental trajectory
362 analysis with the cluster of CNCC-derived mesenchymal cells (C6) as the root, we demonstrated
363 that it has the potential to differentiate into the other six fibroblast-related cell clusters (C1, C2, C3,
364 C4, C5, C7) (**Fig. 6f**), suggesting an important role of the *Hmx1* gene on pinna fibroblast
365 development. We further noticed that genes that were substantially downregulated (logFC > 1) in
366 *mEC1*$^{dup/dup}$ mice were predominantly not expressed in *Hmx1*$^{+}$ cells of WT mice, substantiating our
367 earlier hypothesis that ectopic *Hmx1* expression suppresses cell fate differentiation (**Fig. 6g**).
368 To provide conclusive evidence for our hypothesis that defects in the outer ear are attributable
369 to aberrant *Hmx1* expression, *in situ* section RNA hybridization (*Hmx1*, *Dmkn*) was performed on
370 both upper and lower sections of the pinna in wild-type and *mEC1*$^{dup/dup}$ mice. The results
371 demonstrated upregulated expression of *Hmx1* gene in the basal pinna zone, and more notably, an
372 expansion of *Hmx1* expression from the basal pinna zone to the distal tip of the pinna, coupled with
373 a downregulation of *Dmkn* expression in these areas (**Fig. 6h and 6i**), which may result from a non-
374 cell autonomous effect of *Hmx1* expression on the tissue adjacent to the epidermis. Taken together,

375 our data provide strong evidence for the critical role of *Hmx1* in the observed abnormalities in the

376 outer ear.

377

Discussion

This study examines the genomes of seven human pedigrees, whose members present with constricted ears. All affected individuals showed a duplication of a genomic region termed the BCE locus near the *HMX1* gene. We present compelling evidence that the duplication of regulatory elements in the BCE locus erroneously modulates the spatiotemporal expression characteristics of *HMX1*, leading to the ear developmental abnormality.

The evolutionarily conserved region (ECR) downstream of *Hmx1* has been implicated in eye and external ear malformations across multiple species[8,9,11], and linked to human isolated bilateral concha-type microtia[10]. However, the explicit pathogenic mechanism in BCE patients remained unclear. This uncertainty was due to observations of larger genomic duplications extending beyond the single ECR element in previously reported pedigrees[10] and in our seven newly analyzed pedigrees. We have identified a minimal critical region, the 'BCE locus', encompassing three distinct cis-regulatory elements. We demonstrate that this locus is the causative factor for bilateral constricted ear by employing transgenic cell and mouse models. In many CNV-related developmental abnormalities, distinct cis-regulatory elements are organized as a cluster in the mutation interval[12,25,27]. The deletion or duplication of these elements induces morphological variations, stemming from differential gene expression[12,39]. Our mouse models reveal that the morphology of the pinna is particularly sensitive to changes in *Hmx1* gene expression due to variation in the copy numbers of enhancer sequences. Therefore, the pinna morphology variation among pedigrees may also be caused by a different number of cis-regulatory elements within the associated CNV or distinct interaction patterns between these elements and the *Hmx1* promoter,

399 which may require the study of patient-specific hiPSCs[27] for further analysis. Alternatively, this
400 variation may reflect the contribution of other family-specific genetics variants.

401 Enhancers activated within specific developmental windows are crucial for normal craniofacial
402 development, as evidenced by the study of long-range enhancer clusters in hCNCCs in the PRS
403 locus[25] and inter-TAD interactions controlling *Hoxa* function[22,26]. Our results also demonstrate that
404 the enhancers in the BCE locus are activated at different developmental stages of hCNCCs: hEC2
405 may be activated at early hCNCCs stage, and hEC1 is activated at late- and PA-like hCNCCs stage
406 (**Fig. 7a**). The anterior-posterior positional identity of gene expression in CNCCs is pivotal for
407 craniofacial development[17]. Likewise, optimizing the proximo-distal positional identity within the
408 same pharyngeal arch is equally crucial[18]. In this study, we elucidate the mechanism of *Hmx1* gene
409 patterning related to both the enhancer and motif cluster sequences. We have delineated an intra-
410 arch positional identity enhancer cluster, named PI-HEC. Redundancy within clusters of enhancers,
411 or within super enhancer (SE) regions, is a prevalent phenomenon that has been observed in multiple
412 genomic loci[40], providing phenotypic robustness[41,42], buffering gene expression against mutations[43]
413 and transcriptional noise[44]. Similar additive effects of a multipartite enhancer cluster have also been
414 reported in the *Ihh* locus[12]. Besides these two modes, synergy (super-additive) of different enhancers
415 is also observed in the EC1.45 of the PRS locus[25], the *Fgf5* locus[45], and regulating gap gene-
416 expression patterns in *Drosophila* embryos[46] and coordinating cell fate determination[47]. Compared
417 with aforementioned interaction modes of multipartite enhancers, the PI-HEC presents a distinct
418 scenario in which they possess both coordinated and synergistic properties: hEC1 serves as the
419 primary transcriptional activity driver in the proximal region, hEC2 enhances transcription weakly
420 but contributes distal positional information, and hEC3 alone is a weak, unstable enhancer but

421 compensates caudal positional information and becomes stable when combined with hEC1 and
422 hEC2 (**Fig. 7b**). This collaborative mode is somewhat similar to the interaction between
423 predominant sites and supportive sites studied in the estrogen receptor alpha binding sites (ERBS)[48]
424 and recently reported non-classical enhancers (facilitators) in the mouse *α-globin* cluster[49]. However,
425 this study present, for the first time, the synergistic and coordinated effects of multipartite regulatory
426 elements on the spatial regulation of gene expression in a disease-related locus. The underlying
427 mechanism for this pattern may be explained by the spatial proximity of these enhancers. In this
428 case, TFs or coactivators, responsible for driving transcription, recruited to hEC1 can spread to
429 hEC2 and hEC3[50,51]. However, this possibility needs to be further investigated in its original three-
430 dimensional genomic context[52]. Furthermore, whether this new mode of interplay between
431 multipartite enhancers is a common phenomenon in the regulation of intra-arch gene patterning in
432 CNCCs also requires further exploration, but it potentially signifies a paradigm shift in our
433 understanding of multipartite enhancers or SE dynamics[53].

434 The molecular basis of gene patterning is predominantly shaped by the interaction between
435 tissue-specific TFs and enhancer structural organization[13,51,54-56]. While enhancers have been
436 extensively studied for their role in governing spatial gene expression at embryonic level[12,25,57-60]
437 and across branchial arches[61], the precise mechanisms dictating gene expression within intra-arch
438 regions remain unclear. Our investigation into the spatial specificity of *Hmx1* expression highlights
439 that gene patterning within specific sub-structures, such as the PA2 zone, is regulated by the highly
440 restricted activity of enhancers. We have uncovered the vital role of TALE-type HD TFs, HMG box-
441 containing TFs and Coordinator-related TFs (TWIST1) in their interaction with hEC1, a key element
442 for its functional integrity. TALE-type HD TFs are critical for the formation of the PA2 ground

state[62], and our study also suggests a functional involvement in maintaining the proximal spatial position of hEC1. TCF7L2, a TF newly identified in this study as binding to hEC1, and the Coordinator motif in the D4 region, are both crucial for the activity of hEC1 (**Fig. 7c**). TCF7L2 is a Wnt signaling effector, modulating the neural crest gene expression in a position-dependent manner[63]; whereas the Coordinator motif, which can be bound by many bHLH and HD TFs, are important for shaping facial morphology[30]. Our study on the hEC1 structural arrangement further revealed that a novel coordinator pattern, consisting of HD, HMG-box and Coordinator motifs (in the D1 and D3-D4 regions), is crucial for enhancer activity and specificity. Whether these new regulatory rules exist in other craniofacial-specific enhancers remain to be investigated.

Duplication of genomic sequences, particularly those containing enhancers, can lead to neo-TAD formation at the *SOX9* locus and *KCNJ2* misexpression[64], precipitate ectopic gene expression[12], resulting in limb malformations, and even contribute to lung adenocarcinoma[65]. Specifically, the duplication of hEC1 also disrupts the spatial specificity of *Hmx1* expression, contributing to aberrations in the proximal helix and scapha in mice. These regions mainly originate from neural crest-derived mesenchymal cells[33,66,67], which aligns with our scRNA-seq data and underscores the pivotal role of *Hmx1* in the normal development of CNCC-derived fibroblasts in the outer ear (**Fig. 7d**). CNCC-specific knockout of *Hmx1* in mice leads to a low-set, protruding ear phenotype, consistent with findings in mEC1 deletion mice and 'dumbo' mice[32]. This suggests that appropriate *Hmx1* expression in CNCCs is essential for pinna organogenesis[7], further substantiating the validity of employing CNCCs as a cellular model for the investigation of BCE pathogenesis. Notably, the ear malformations observed in most human BCE patients within our study pedigrees are more severe than those in our *mEC1*$^{dup/dup}$ mouse model. This discrepancy may arise from a broader expansion

of *Hmx1* expression due to the extent of the duplication of the PI-HEC region, rather than solely the hEC1 duplication[12]. Despite these differences, the phenotypic hallmarks of the *mEC1*$^{dup/dup}$ mouse, including reduced ear area and a wrinkled helical rim, do grossly mimic the ear phenotype observed in patients.

In summary, our results reveal that CNCC-specific and precise spatial specificity of *Hmx1* expression is vital for the development of fibroblasts in the outer ear, with both loss and gain of *Hmx1* expression contributing to pinna malformation. The gene patterning of *Hmx1* across PA2 is intricately modulated by hEC1, hEC2, and hEC3, with hEC1's dominance as an enhancer being further refined by coordinator motif-transcription factor binding events. Our study reveals the complex interplay of gene regulatory activity in precise developmental space and time, and contributes to the understanding of the regulatory rules of gene expression.

Methods

Ethics statement. All procedures involving human samples were performed with the approval of the Ethics Committees of the School of Biological Science and Medicine Engineering at Beihang University (BM20210057) and the Plastic Surgery Hospital of the Chinese Academy of Medical Sciences (2017-07). The study has been registered and approved by China's Ministry of Science and Technology (project 2023-CJ0849). Written informed consent for participation in genetic and biological research was obtained from all subjects or their legal guardians. Consent for the publication of photographs was given after a separate explanation and request. All animal experiments were performed under the approval of the Animal Care Committee at Beihang University (BM20210057).

Samples and recruitment. We recruited members of seven Chinese families (from five separate provinces) in which bilateral constricted ear was segregating in a Mendelian fashion. The phenotype in all pedigrees segregated in a manner consistent with autosomal dominant inheritance. All affected subjects presented with a bilateral constricted ear phenotype, as shown in both ears of the probands in Fig. 1B. This phenotype is characterized by a malformation of the external ear, notably reduced size of the helix and absence of the triangular fossa, scaphoid fossa, and antihelix, giving a shell-like appearance. Contrastingly, the crus helicis, concha cavity, tragus, and earlobe maintained their normal structure. No accompanying facial abnormalities were observed.

Genotyping, target-capture sequencing, and whole-genome sequencing. We extracted DNA from venous blood samples using the TIANamp DNA Midi Kit (Tiangen Biotech, Beijing, China). Genotyping was performed on 40 samples using the Human Omni-Zhonghua chips (Illumina, CA, USA) for pedigrees 1, 2, 3, 4, 6, and 7. Target-capture sequencing was conducted on 32 samples from pedigrees 1, 3, and 7 using the Agilent SureSelect Kit (Agilent Technologies, Santa Clara, CA, USA). Whole-genome sequencing was carried out for the proband of Pedigree 5 using the DNBSEQ-T7 platform (Huada, Shenzhen, China).

For the chip data, we utilized the Genotyping and CNVPartition modules of GenomeStudio v2011.1 for genotype and CNV calling, achieving genotype call rates above 99.6%. The target-capture sequencing involved designing capture oligos for a 5.9Mb region encompassing the susceptible locus using the SureDesign tool (Agilent Technologies, Santa Clara, CA, USA). The DNA library was prepared following the manufacturer's specifications, and sequencing was executed on the Illumina HiSeq X10 system (Illumina, San Diego, CA, USA), yielding an average depth of 196X per sample.

Quality control for next-generation sequencing data was conducted using FastQC v1.1.0 and Cutadapt v1.15. Post quality control reads were mapped to the human reference genome (hg38) using BWA v0.7.16a[68]. PCR duplications were removed with Picard v2.27.0, and BAM file sorting and indexing were performed using Samtools v1.18. SNP and small indel calling were done using the HaplotypeCaller model of GATK v4. For identifying causal variants, we applied criteria including cosegregation with affected and unaffected individuals in each pedigree, minor allele frequency below 0.05 in dbSNP138, ExAC project, and the 1000 Genomes database, and minimum coverage depth of 20 with an alternative allele depth above 5.

Causal mutation screening. Our screening for causal mutations began with linkage analysis to pinpoint the suspected genomic locus for bilateral isolated constricted ear. We adhered to the quality control (QC) procedures outlined by Mao *et al.* for pruning SNPs in strong linkage disequilibrium, removing SNPs with Mendelian errors, and verifying sex specifications[15]. Parametric multi-point linkage analysis was conducted on Pedigrees 1, 3, and 7, both individually and collectively, using MERLIN v.1.1.2[69]. For consanguinity analysis, we used the same QC-approved data and performed pairwise IBD estimation for founders from the seven pedigrees using Plink v1.9[70]. A PI-HAT value greater than 0.1 between paired founders indicated a genetic relationship.

After examining all mutations within the coding region of the target area, we assessed copy number variations using differential coverage depth. We utilized the 'getSegmentReadCountsFromBAM' function in cn.mops for read

coverage of genomic segments[71]. To detect CNVs, the read count matrix was normalized for each base relative to the mean coverage of the target-capture regions. We excluded segments with an average read count below 90 to minimize noise from untargetable regions. This threshold was based on the theoretical depth in cases where a deletion was homozygous in 22 sequenced cases and heterozygous in 10 controls, given a mean depth of 196 for target regions. We then calculated and displayed the ratio of the average read count of cases to controls for each family. Furthermore, to illustrate the CNV of each case at the target-capture region, we calculated and plotted the case-to-control read count ratio for each family (Supplementary Fig. 1C). The expected ratios for CNV=3, CNV=2, and CNV=1 were approximately 1.5, 1, and 0.67, respectively.

Culture of human embryonic stem cells (hESCs). Female H9 (WA09) hESCs were cultured on six-well plates pre-coated with hESC-qualified Matrigel matrix (Corning, catalog no. 354277). The cells were maintained in mTeSR media (Stem Cell Technologies, catalog no. 85850), which was replenished daily. For routine passaging every 4-6 days, hESC were diluted at ratios between 1:6 and 1:50 using 0.5 mM EDTA-PBS (Thermo Fisher Scientific, catalog no. 15575020).

Generation of CRISPR/Cas9 genome-edited cell lines. For the deletion of specific regions in H9 hESCs via CRISPR/Cas9, six guide RNA (gRNA) sequences targeting the hEC1 and hEC2 regions (approximately 9 kb) were designed based on sequences from the Ensembl Genome Browser (Extended Data Fig. 4a). The gRNA sequences were as follows: g2 - CCAGACCCCTATGGAGCACGGGG, g4 - GCGGATCCAAGCTCACCTCAGG, g5 - CGATCTTAGTGCCTTCACCGCGG, g6 - GGGGTATTCTGCTGGCCGTAGGG, g1 - AGGAGTGTAGGAGCCGATGGTGG, and g3 - TCTGCTGTGACTCACGGTTGAGG. Confluent hESC colonies were dissociated into a single-cell suspension, and 3.0×10^6 cells were electroporated using the P3 Primary Cell 4D-Nucleofector L kit and 4D-Nucleofector system (Lonza Group Ltd., Basel, Switzerland), following the manufacturer's protocol (1320v, 30ms, 2 pulses). Post-transfection, the cells were plated in 6-well plates using

mTeSR media supplemented with 10 µM Y-27632 (Stem Cell Technologies, catalog no. 72304) for 48 hours. Cells with high viability were then selected in media containing puromycin (1 µg/ml) for 2-3 days. Drug-resistant wells were subsequently processed for single-cell isolation using Accutase (Sigma-Aldrich, catalog no. A6964), diluted, and re-plated into a 96-well plate. Pool cleavage efficiency was assessed using the EZ-editor Genotype Analysis System (UBIGENE). After two weeks, colonies were picked, split, and screened for enhancer deletion using specifically designed primers (Extended Data Fig.4a and Supplementary Table 2). Positive clones with confirmed homozygous deletions were expanded and prepared for cryopreservation for downstream analysis.

hCNCCs and PA-like hCNCCs derivation and culture. hESCs were differentiated into cranial neural crest cells (hCNCCs) following a previously described protocol with minor modifications[28]. Initially, confluent hESC colonies were treated with 2 mg/mL collagenase (Gibco, catalog no. 17104019) for 30-60 minutes until colony edges began to lift. These detached colonies were then cultured in neural crest cell induction (NCCI) medium, forming clusters of 100-200 cells in ultra-low attachment petri dishes. The NCCI medium comprised a 1:1 mixture of DMEM F-12 medium (Thermo Fisher Scientific, no. 10565018) and Neurobasal medium (Thermo Fisher Scientific, no. 21103049), supplemented with 0.5× N2 (Thermo Fisher Scientific, no. 17502048), 0.5× B27 (Thermo Fisher Scientific, no. 17504044), 20 ng/mL bFGF (Peprotech, catalog no. AF-100-18B), 20 ng/mL EGF (Peprotech, catalog no. AF-100-15), and 5 µg/mL bovine insulin (Gemini Bio, catalog no. 700-112P). After about 24 hours, embryoid bodies (EBs) formed and were maintained with bi-daily changes of NCCI medium. On day 4, EBs were transferred to new tissue culture treated dishes for approximately three days to initiate sphere attachment. Following attachment, media was changed daily, allowing cranial neural crest cells to migrate from the neural rosettes over about four days. Neuroepithelial spheres were then manually removed, and the remaining neural crest cells were harvested and passaged onto fibronectin (7.5 µg/mL)-coated plates (Millipore, catalog no. FC010-100MG) in early hCNCCs medium. This medium included a 1:1 DMEM F-12/Neurobasal mixture, 0.5× N2 supplement, 0.5× B27 supplement,

20 ng/mL bFGF, 20 ng/mL EGF, 1 mg/mL BSA (Gemini Bio, catalog no. 700–104 P), and 1× antibiotic-antimycotic (Gibco, catalog no. 15240062). After two additional passages, the medium was changed to late hCNCCs medium by supplementing the early hCNCCs medium with 3 μM CHIR 99021 (Selleckchem, catalog no. S2924) and 50 pg/mL BMP2 (Peprotech, catalog no. AF-120-02) to enhance cell proliferation and reduce migration[28]. Following two more passages, late-stage hCNCCs were induced to transition into pharyngeal arch (PA)-like cell states by adding 100 nM retinoic acid (RA) (Sigma-Aldrich, catalog no. R2625) to the late hCNCCs medium for additional two passages.

Immunofluorescence analysis. hCNCCs were cultured on fibronectin-coated (7.5 μg/mL) coverslips placed in 6-well plates. Upon reaching confluency, cells were fixed with 4% paraformaldehyde (PFA) for 10 minutes at room temperature and permeabilized with 0.1% Triton X-100 in PBS for 15 minutes. The cells were then incubated with primary antibodies targeting p75 (Abcam, catalog no. ab245134), NR2F1 (Abcam, catalog no. ab181137), AP2 (Santa Cruz Biotechnology, catalog no. sc-12726), and SOX9 (Millipore, catalog no. AB5535), each diluted in blocking buffer, and left overnight. After washing thrice for 5 minutes each with PBS, the cells were incubated for 1 hour in the dark with anti-mouse Alexa Fluor 488 antibody (Thermo Fisher Scientific, catalog no. A28175) and anti-rabbit Alexa Fluor 594 antibody (Thermo Fisher Scientific, catalog no. A-11012). Following another series of three 5-minute washes in PBS, the coverslips were mounted on slides using DAPI Fluoromount-G (SouthernBiotech, catalog no. 0100-20). Imaging was performed on a Carl Zeiss LSM 980 microscope, and the acquired images were processed using ImageJ software.

RNA extraction, cDNA preparation, and quantitative real-time PCR. For each of the three independent hCNCCs differentiation experiments, approximately 2×10^6 cells per well were washed with cold PBS and lysed using Buffer RLT (Qiagen, catalog no. 79216). RNA extraction was performed according to the manufacturer's protocol (RNeasy Mini Kit, Qiagen, catalog no. 74104). Subsequently, 1-2 μg of total RNA from each sample was reverse transcribed to cDNA using RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific, catalog no. EP0452). The

resulting cDNA samples were diluted, and qRT-PCR was conducted on a Bio-Rad CFX96 real-time PCR machine using SYBR Green PCR Master Mix (Thermo Fisher Scientific, catalog no. 4309155), following the manufacturer's instructions. The delta-delta Ct method was employed to calculate the relative expression levels.

Promoter-Capture Hi-C in late hCNCCs. The preparation of Promoter Capture Hi-C libraries was conducted following established protocols[72,73]. Cells were first fixed with 2% formaldehyde for 10 minutes at ambient temperature, followed by quenching with 0.2M glycine for 5 minutes. Post-fixation, cells were lysed, and endogenous nucleases were inactivated with 0.3% SDS. Chromatin was digested using 100U of HindIII (New England Biolabs, catalog no. R0104L), labeled with Biotin-14-dCTP (Invitrogen, catalog no. 19518018), and then ligated using 50U of T4 DNA ligase (New England Biolabs, catalog no. M0202L). After reversing the cross-links, DNA was isolated using the QIAamp DNA Mini Kit (Qiagen, catalog no. 51306), in strict accordance with the manufacturer's instructions. Subsequent steps involved fragmenting the DNA to 300-600 bp, end-repair, A-tailing, and adapter ligation using SureSelect adaptors. Biotinylated fragments were isolated through streptavidin-affinity pull-down and PCR amplified. Promoter-specific Capture Hi-C was performed using the SureSelect XT Library Prep Kit ILM (Agilent Technologies), which included a custom-designed biotinylated RNA bait library and paired-end blockers. Post-capture PCR enrichment was followed by final purification of the libraries with AMPure XP Beads (Beckman Coulter, catalog no. A63881). Library quality was evaluated using Bioanalyzer profiles (Agilent Technologies), and high-throughput sequencing was performed on the Illumina NovaSeq 6000 platform.

Luciferase assay. The luciferase assay was conducted as previously described[25]. In the case of hESCs, small clusters of 4-10 cells each were passaged into a 24-well plate with ROCK inhibitor (Y27632), and transfection was performed the following day. For pharyngeal arch-like hCNCCs, cells were transfected two hours post-seeding at approximately 160,000 cells per well in 24-well plates. Each assay consisted of four technical replicates. The transfection mix included 0.5 μg of pGL3-enhancer plasmids with firefly luciferase gene sequences driven by an SV40 promoter

(Promega, catalog no. E1761), 0.02 μg of pRL-TK Renilla luciferase (Promega, catalog no. E2261) as an internal control, and 1.5 μL of FuGENE HD (Promega, catalog no. E2691), all diluted in Opti-MEM I Reduced-Serum Medium (Gibco, catalog no. 31985062). The enhancer sequences were cloned into the cut site between KpnI and XhoI, and relevant plasmid construction primers are listed in Supplementary Table 2.

Twenty-four hours post-transfection, cells were harvested for measurement using the Dual-Luciferase Assay System (Promega, catalog no. E4550), following the manufacturer's instructions, on a CLARIOstar Plus machine (BMG LABTECH). Two or three independent biological replicates were conducted, and the relative firefly luciferase activity was calculated by normalizing the firefly signal values to the renilla signal values across the reaction period.

Enhancer-reporter assays in mouse embryos. Human enhancer sequences (hEC1-chr4: 8700174-8700974; hEC2.1-chr4: 8705075-8706475; hEC2-chr4: 8705075-8706475; hEC3-chr4: 8725074-8725674, hg38 version) were synthesized or PCR-amplified from H9 cell line genomic DNA. Combinations of these sequences, including (hEC1+hEC2) and (hEC1+hEC2+hEC3), along with deletion tiling variants of hEC1 (ΔD1-ΔD5), were assembled using NEBulider HiFi DNA Assembly Mix (New England Biolabs, catalog no. E2621L). These constructs were then cloned into PCR4-Shh:LacZ-H11 plasmids (Addgene plasmids, #139098), placing them upstream of a minimal *Shh* promoter and LacZ reporter gene. Similarly, the orthologous mouse enhancer, mEC1 (chr5: 35623357-35624110, mm39 version), was amplified from C57BL/6J mouse genomic DNA and cloned using the same method. The detailed primers used are in the Supplementary Table 2.

CRISPR/Cas9 microinjection protocol was employed to site-specifically integrate these enhancer sequences into the H11 locus of the FVB mouse strain, which served as surrogate mothers. F_0 embryos were harvested at developmental stages E9.5, E11.5, and E14.5. Only embryos carrying the transgenic reporter at the H11 locus were processed further for LacZ staining. Genotyping primers for each construct are listed in Supplementary Table 2. Whole mouse embryo LacZ staining was performed as previously established protocols[74]. For each enhancer and

630 developmental timepoint, a minimum of three independent positive embryos were analyzed. Imaging was conducted
631 using a Carl Zeiss Axio Zoom V16 microscope.

632 **Enhancer DNA pull-down assay.** Biotin-labeled enhancer DNA sequences were prepared via PCR and verified
633 through 1% agarose gel electrophoresis. Approximately 50 mouse embryos at E11.5 were microdissected to isolate
634 PA2 tissues, which were then washed with cold PBS and homogenized using a Dounce homogenizer. Nuclear
635 proteins were extracted using NE-PER nuclear and cytoplasmic extraction reagents (Thermo Fisher Scientific,
636 catalog no. 78833), following the manufacturer's protocol. Protein concentrations were quantified using a BCA
637 protein assay kits (Thermo Fisher Scientific, catalog no. 23225).

638 The DNA probe and nuclear proteins were pre-mixed and incubated with Bioeast Mag-SA (Streptavidin, SA,
639 Bioeast, catalog no. M2800S) for 1 hour at 4°C. The magnetic beads were then isolated using a magnetic stand and
640 washed thrice with cold PBS. The eluted proteins were analyzed by SDS-PAGE and silver staining to confirm
641 binding. Finally, protein identification was performed using liquid chromatography-mass spectrometry (LC-MS).

642 **CUT&RUN experiment in hCNCCs.** The CUT&RUN assay was conducted in hCNCCs using the
643 ChIC/CUT&RUN Kit (Epicypher, catalog no. 14-1048), incorporating several modifications from previously
644 reported methods. Approximately 500,000 cells were harvested using Accutase and subsequently incubated with
645 activated ConA Beads for 10 minutes at room temperature. Primary antibodies were then added to each 50 µL
646 reaction: IgG (Epicypher, catalog no. 13-0042k, 0.5 µg), H3K27ac (Active Motif, catalog no. 39133, 0.5 µg),
647 TCF7L2 (Cell Signaling, catalog no. 2569S, 1 µg), and the mixture was incubated overnight on a nutator at 4°C.

648 On the second day, pAG-MNase and cold calcium chloride were used to activate the cleavage of target DNA.
649 E. coli Spike-in DNA (0.1 ng) was also added for normalization in downstream analyses. DNA purification involved
650 phenol/chloroform extraction and ethanol precipitation. Library preparation for histone proteins was performed
651 using the CUT&RUN Library Prep Kit (Epicypher, catalog no. 14-1001). For TFs, modified procedures, as

previously reported (https:// doi.org/10.17504/protocols.io.bagaibse), were employed to enhance the preservation of small DNA fragments. Sequencing was carried out using paired-end reads (2 × 150 bp) on the Illumina NovaSeq 6000 platform.

Expression of recombinant proteins and electrophoretic mobility shift assays (EMSA). The coding sequences for MEIS1 protein was cloned into the pET-His6-MBP-TEV-LIC vector between BamHI and XhoI sites, incorporating a tobacco etch virus (TEV) protease cleavage site (ENLYFQ|G) between the maltose-binding protein (MBP) and the MEIS1 protein. BL21 (DE3) competent cells were transformed with this construct. Positive bacterial cultures were grown at 37°C until the optical density at 600 nm (OD600) reached 0.5 and then induced with 0.2 mM isopropyl β-d-1-thiogalactopyranoside (IPTG). Bacterial cells were lysed by sonication in buffer A (50 mM Tris-HCl, 500 mM NaCl, 10 mM 2-mercaptoethanol, 10% (v/v) glycerol, pH 7.5) and centrifuged to collect the supernatant. This supernatant was then applied to a Ni-NTA Superflow Cartridge (Sangon, catalog no. C600792), washed with buffer A supplemented with 50 mM imidazole, and eluted with an imidazole gradient from 100 mM to 500 mM in buffer A. The eluted protein was treated with TEV protease and dialyzed overnight in buffer A at 4°C. The protein mixture was re-applied to a Ni-NTA Superflow Cartridge and eluted with buffer B (50 mM Tris-HCl, 1 M NaCl, 10 mM 2-mercaptoethanol, 10% (v/v) glycerol, pH 8.0) containing 50 mM imidazole. Finally, the MBP-free protein was further purified using size-exclusion chromatography (Superdex 200 Increase 10/300 GL; Cytiva, catalog no. 28990944) and concentrated to approximately 1 mg/mL in buffer C (50 mM Tris, 200 mM NaCl, 10 mM 2-mercaptoethanol, 10% (v/v) glycerol, pH 7.5).

EMSA was performed using the LightShift Chemiluminescent EMSA kit (Thermo Fisher Scientific, catalog no. 20148) with the purified TF protein. Biotin-labeled probe containing predicted MEIS1-binding DNA sequence was synthesized and incubated with the MEIS1 following the manufacturer's protocol. The resultant gels were visualized using a Tanon 5200 system.

Generation of genome-edited mouse models via CRISPR/Cas9. This study utilized four distinct mouse models constructed using the CRISPR/Cas9 technique: mEC1 deletion ($mEC1^{del/+}$), mEC1 duplication ($mEC1^{dup/+}$), $Hmx1^{fl/+}$, $Hmx1^{wt/EGFP}$. gRNAs were designed using CHOPCHOP to target specific regions: $mEC1^{del/+}$: AGGACAGTCTCCAAGTCCGG-TGG; $mEC1^{dup/+}$: TGTGACGTGATCGACTCCAT-CGG; $Hmx1^{fl/+}$: gRNA1-CCCAGATTCAGGGCGTACAA-GGG, gRNA2- ACTGACTTGTTCCTACCTAC-AGG; $Hmx1^{wt/EGFP}$: gRNA1-ATGCCGGGGCTAGTGTGAGC-CGG, gRNA2-GCGCCGGCTCACACTAGCCC-CGG. A combination of sgRNAs, Cas9 protein, and donor vectors were co-injected into fertilized mouse embryos at the single-cell stage. F0 founder mice were identified via PCR and confirmed by Sanger sequencing. These founders were bred with wild-type mice for germline transmission testing and F1 heterozygous mouse generation. Different homozygous genotypes were subsequently produced by intercrossing heterozygous mice.

The generation of *Wnt1-Cre*, *ROSA26-CAG-LSL-tdTOMATO* and *dumbo* mouse lines has been previously described[11]. $Wnt1::Cre;Hmx1^{fl/fl}$ mice were generated by crossing $Wnt1::Cre;Hmx1^{fl/+}$ male mice with $Hmx1^{fl/fl}$ female mice, and the offspring were validated using PCR. $Wnt1::Cre;Rosa26^{tdTomato/+}::Hmx1^{EGFP/wt}$ bicolor fluorescent mice were generated by initially crossing $Hmx1^{EGFP/wt}$ male mice with $Rosa26^{tdTomato/+}$ female mice to produce $Rosa26^{tdTomato/tdTomato}::Hmx1^{EGFP/EGFP}$ mice, which were then crossed with *Wnt1-cre* mice to yield the final model. All mice were maintained on a C57BL/6J genetic background, housed in a Specific Pathogen Free (SPF) environment, and handled according to ethical standards. These mouse lines are available upon request from the corresponding author.

Sample preparation, RNA isolation and bulk RNA-seq library preparation. Forming pinna prominences from E14.5 embryos (both wild type and $mEC1^{dup/dup}$) were micro-dissected, washed with cold 1× PBS, and homogenized using a Tissue-Tearor. Total RNA extraction followed the same protocol as for cell RNA extraction. RNA-seq libraries were then prepared using the Illumina TruSeq RNA Library Prep Kit v2 (Illumina, catalog no. RS-122-

2001), adhering to the manufacturer's instructions. For each genotype, five independent biological replicates were processed.

Sample preparation, RNA isolation and single cell RNA-seq library preparation. Single cell RNA-seq samples were prepared as previously described[26]. Briefly, micro-dissected forming pinna prominences from *Hmx1*$^{wt/EGFP}$ embryos were enzymatically dissociated using 0.5% trypsin/1× EDTA for 10 minutes at 37 °C (for E10.5 and E12.5 stages) and papain digestion mix for 7 minutes (for E14.5 stage). The treated tissues were then rinsed in ice-cold 1× DMEM, filtered, and GFP^{+} cells were enriched using FACS (Sony, MA900). These collected cells were resuspended and diluted in 1× DPBS (Cytiva, catalog no. SH30028.FS) to a concentration of 1×10^{6} cells per ml. Gene expression libraries were prepared using GEXSCOPE Single Cell RNA Library Kits v.2 (Singleron, catalog no. SD-4180022). The single-cell suspensions were loaded onto the GEXSCOPE microchip, which facilitates automated single cell capture, cell lysis, cellular mRNA capture, and molecular labeling via the Singleron Matrix instrument. The barcoded cDNA was amplified and utilized for constructing single cell NGS libraries. The quality and quantity of the resulting libraries were assessed using Qubit 4.0 (Thermo Fisher Scientific) and Qseq100 (Bioptic). For gene expression analysis, samples were sequenced on an Illumina NovaSeq platform to a depth of 500 million reads per channel.

Whole-mount *in situ* hybridization (WISH). The WISH procedure was carried out on whole-mount embryos according to previously established protocols[26]. Initially, mouse embryos were carefully dissected to remove extraembryonic membranes and then immersed in cold 1× PBS. For fixation, embryos were submerged in 4% paraformaldehyde in 1× PBS and gently rocked for 6 hours or overnight at 4°C in a sealed vial. Post-fixation, embryos were thoroughly washed three times with PBST (1× PBS with 0.1% Tween-20), dehydrated through a graded series of methanol-PBST solutions, and finally placed in 100% methanol for storage at -20°C until they were processed further.

On the day of hybridization, the embryos were first rehydrated and then washed twice for 5 minutes each in PBST. They were incubated for 1 hour in 6% hydrogen peroxide/PBST and subsequently washed again with PBST at 4°C. Next, the embryos were treated with 10 µg/mL proteinase K (Roche, catalog no. 3115887001) in PBST for a duration of 10 to 30 minutes, varying according to different embryonic stages, at room temperature. The proteinase K reaction was halted by adding 2 mg/mL glycine in PBST. For pre-hybridization, embryos were incubated for 1 hour at 68°C in hybridization buffer, which comprised 50% deionized formamide, 5× SSC buffer, 1% SDS, 100 µg/mL yeast tRNA, and 50 µg/mL heparin. After this, digoxigenin-labeled probes (concentration ranging from 0.5 to 1 µg/mL) were added, and the embryos were incubated overnight at 68°C.

On the following day, embryos underwent a series of extensive washes. They were initially washed three times for 30 minutes each in Wash 1 (50% regular formamide, 5× SSC buffer, 1% SDS). This was followed by a 10-minute wash in a 1:1 mixture of Wash 1 and Wash 2 (0.5M NaCl, 10mM Tris HCl pH 7.5, 0.1% Tween-20), and then two 30-minute washes in Wash 2 containing 100 µg/mL RNaseA and 100 units/mL RNase T1 at 37°C. Further, embryos were washed three times for 30 minutes each in Wash 3 (50% regular formamide, 2× SSC buffer) at 65°C, and then three times for 10 minutes each in TBST (140mM NaCl, 2.5 mM NaCl, 25 mM Tris pH7.5, 0.1% Tween-20) at room temperature. The embryos were then blocked for 2.5 hours at room temperature in block buffer (1% blocking reagent in TBST) and incubated overnight at 4°C with anti-Digoxigenin-AP-conjugated antibody (dilution range: 1:1000 to 1:5000, Roche, catalog no. 11093274910) diluted in the same buffer.

On the third day, the embryos were subjected to five 10-minute washes, followed by five 60-minute washes in TBST, and then left to wash overnight at 4°C. Lastly, the embryos were washed twice for 20 minutes each in NTMT buffer (100mM NaCl, 100mM Tris pH 9.5, 50mM MgCl2, 0.1% Tween-20) at room temperature. The color reaction was developed using NBT/BCIP (Roche, catalog no. 11681451001) in NTMT buffer for 0.5 to 2 hours in the dark.

738 The reaction was stopped with three 10-minute washes in PBST, and the embryos were subsequently photographed
739 using a Carl Zeiss Axio Zoom V16 microscope, following the manufacturer's instructions.

740 **Frozen section in situ hybridizations.** Mouse embryos (wild type and *mEC1*$^{dup/dup}$) at E14.5 were dissected and
741 washed with cold 1× PBS before being fixed in 4% paraformaldehyde (PFA). The embryos underwent a dehydration
742 process using a series of sucrose/PBS solutions (10%, 20%, 30%). This was followed by an equal volume mixture
743 of OCT compound, with rotation for 30 minutes at room temperature. Embryos were then embedded in OCT on dry
744 ice and stored at -80°C until required. Cryostat sections of 20 µm thickness were prepared in a horizontal orientation
745 using a Leica CM1860. Before hybridization, frozen sections were thawed to room temperature and allowed to air
746 dry for at least one hour. They were then re-fixed using 4% PFA/PBS for 15 minutes, followed by three 5-minute
747 washes in 1× PBS. Acetylation was carried out in an acetic anhydride solution (1.35% (v/v) triethanolamine, 0.175%
748 (v/v) HCl, 0.25% (v/v) acetic anhydride) for 10 minutes at room temperature, and the sections were subsequently
749 washed three times for 5 minutes each in 1× PBS. Prehybridization was carried out at room temperature for 2 hours.
750 Digoxigenin (dig)-labeled probes were diluted to 1 µg/mL in hybridization buffer (50% deionized formamide, 5×
751 SSC buffer, 1% SDS, 100 µg/mL yeast tRNA, 50 µg/mL heparin), then heated to 80°C for 5 minutes. Approximately
752 400 µL of hybridization buffer was applied to each slide, covered with parafilm, and hybridized at 70°C for 16-18
753 hours. On the second day, the parafilm was removed by immersing slides in 5× SSC. This was followed by a series
754 of washes: 30 minutes in 5× SSC/50% formamide, 20 minutes in 2× SSC, and twice for 20 minutes each in 0.2×
755 SSC at 65°C. Subsequently, the sections were washed three times for 5 minutes each in MABT buffer at room
756 temperature. Blocking was performed for 1-2 hours in 2% Blocking reagent/MABT buffer. Anti-Dig-AP diluted at
757 1:2000 in blocking buffer was applied to the sections, which were then incubated overnight at 4°C. On day 3, the
758 sections were washed eight times for 5 minutes each in MABT, twice for 10 minutes each in NTMT buffer, followed
759 by application of BM purple (Roche, catalog no. 11442074001) in the dark for 2-6 hours, depending on the signal

development. The reaction was terminated by immersing the slides in MABT for two 10-minute washes, then post-fixing in 4% PFA/PBS for 30 minutes, and washing three times for 5 minutes each in 1× PBS. The sections were mounted with VectaMount AQ aqueous mounting medium (Vector Laboratories, catalog no. H-5501-60). Imaging was performed using a LEICA Aperio Versa 8 microscope, and images were processed with ImageScope x64 software (version 12.4.6.5003).

Micro-computed tomography (Micro-CT) analysis of ear structures.

Mice were euthanized and whole heads imaged using a SkyScan 1276 micro-computed tomograph (Bruker, Belgium). Scanning was performed at 18-micron resolution using the following parameters: 55 kV, 200 uA, 0.5 mm Al filter; 0.4° rotation step over 360°, with 3-frame averaging. All raw scan data were reconstructed into multiplanar slice data using NRecon V1.7.3.1 software (Bruker). Reconstructed data were then rendered in 3D with consistent thresholding parameters using Drishti V3.0 Volume Exploration software (GitHub.com/nci/Drishti)[75] for gross visual assessment of the craniofacial skeleton. Representative rendered images from animals of each genotype were captured and processed using Photoshop (Adobe Creative Cloud).

Computational analysis

RNA-seq data and GO enrichment analysis. RNA sequencing reads were initially processed using fastp for trimming of the paired-end data, with quality assessment performed via FASTQC (version 0.12.1). Subsequently, the reads were aligned to the mouse reference genome (GRCm39) using the STAR aligner (version v2.7.11a). Quantification of the alignment files was carried out using featureCounts (version 2.0.6), incorporating the gene model parameter. Differential gene expression analysis was conducted utilizing the limma (version 3.56.2), Glimma (version 2.10.0), and edgeR (version 3.42.4) packages. Significantly downregulated genes, defined by criteria of an FDR ≤ 0.05, log fold change (logFC) < 0, and log counts per million (logCPM) ≥ 1, were selected for Gene Ontology

(GO) enrichment analysis. This analysis was performed using the clusterProfiler package (version 4.8.3), with results visualized via the treeplot function.

ChIP-seq and CUT&RUN data analysis. Publicly available ChIP-seq datasets for this study were obtained from the GEO database, including datasets for d11hCNCC: GSE28874[76]; P4hCNCC: GSE145327[25] and GSE70751[28]; mNCC: GSE89435[17]; Pinna: GSE211900[26]. Sequencing reads were trimmed using fastp and aligned to either the mouse (GRCm39) or human (GRCh38) reference genome using bowtie2 (version 2.4.0). Duplicate reads (marked by Picard toolkits), non-unique, and low mapping quality reads were filtered out using sambamba. BigWig files, with signals normalized using the RPGC method, were generated for visualization using deepTools2.0 (version 3.4.2). Peak calling against the INPUT sample was performed using Macs2 software (version 2.2.7.1). Additionally, human embryonic craniofacial epigenetic data from stages CS13, CS14, CS15, and CS17 were accessed from https://cotney.research.uchc.edu/craniofacial/[24].

CUT&RUN sequencing data were analyzed employing the CUT&RUNTools 2.0 pipeline, as previously reported[77]. For histone peak calling, the fragment size filter was deactivated, while for TFs, fragments larger than 120 bp were filtered out. Normalized bigwig files were generated using an alignment scale factor based on spike-in reads from *E. coli* K12 strain MG1655. Two replicates for independent biological experiments were generated for each antibody (IgG, H3K27ac, and TCF7L2).

ATAC-seq data analysis. ATAC-seq datasets used in this study were sourced from existing databases (d11hCNCC: GSE108517[27]; hESC and P4hCNCC: GSE145327[25]; mNCC: GSE89436[17]; Pinna: GSE211899[26]). The adapter sequences were trimmed using the bbduk.sh script and the reads were subsequently aligned to the reference genome using bowtie2. Low-quality reads were filtered out following the methodology described in the ChIP-seq data analysis section. The alignmentSieve function with the –ATACshift parameter was employed to adjust for the

802 transposase binding pattern. For visualization purposes, bigwig files were generated utilizing the bamCoverage
803 function in deepTools 2.0 (version 3.5.2).

804 **Hi-C and PCHi-C data analysis.** For the capture Hi-C data of human cranial neural crest cells (hCNCCs), a genome
805 digest file for HindIII was created using hicup_digester. The HiCUP software (version 0.9.2) was then employed to
806 generate alignment files. To prepare files for the Chicago R package (version 1.30.0), scripts such as bam2chicago.sh
807 and makeDesignFiles.py from chicagoTools were utilized. The Chicago package produced washU_text format
808 results, showcasing interactions between two genomic positions. Interactions with a score ≥5 were considered high-
809 confidence and visualized as arcs using pygenometrack (version 3.9). Other PCHi-C datasets in this study (hESC:
810 GSE8682[78]; Pinna: GSE211901[26]) were downloaded in washU_txt format and similarly visualized with
811 pygenometrack without additional processing.

812 For the CS17 craniofacial Hi-C dataset (GSM5944420[24]), the hic file was obtained for further analysis. We
813 firstly converted hic format to cool format using hicConverFormat function in HiCExplorer (version 3.7.2) with –
814 *resolutions 100000*, then hicFindTADs was applied for calling TAD regions with parameters: -
815 *thresholdComparisons 0.05, -delta 0.01, -correctForMultipleTesting fdr*.

816 **scRNA-seq analysis.** The scRNA-seq reads were processed using the CeleScope pipeline, details of which are
817 available at https://github.com/singleron-RD/CeleScope. This processing generated a 10x Genomics-like raw
818 feature-barcode matrix suitable for downstream filtering and data integration using Seurat (version 4.3.0). For cell
819 filtering, we assumed that both the log10-transformed gene counts and UMI counts per cell for each sample followed
820 a normal distribution. We calculated the mean ± 1.96 times the standard deviation, defining the 95% confidence
821 interval, to determine the lower and upper thresholds. These thresholds were then converted back to the original
822 scale (as the power of 10 of the log values) to establish criteria for excluding low-quality cells. Furthermore, cells
823 exhibiting a mitochondrial gene expression ratio exceeding 5% were also categorized as low quality.

Data integration was accomplished using the CCA2-based pipeline of Seurat package. The SelectIntegrationFeatures function was utilized to identify variable features across datasets for integration. Subsequently, the FindIntegrationAnchors function was employed to identify anchor pairs between datasets, followed by the IntegrateData function to create an integrated assay, which included the newly integrated expression profiles. Based on the results of integration, we performed PCA dimensionality reduction using RunPCA on the variable features employed for integration. This was followed by UMAP dimensionality reduction using RunUMAP with parameters *-dims 1:20*, *-n.neighbors 50*. For unsupervised clustering, we executed the FindNeighbors function using default parameters, and FindClusters with *-resolution 0.4*.

PAGA developmental trajectory analysis. The integrated expression profiles were imported into SCANPY (version 1.9.3) for the construction of the fibroblast differentiation trajectory employing the Partition-based Graph Abstraction (PAGA) algorithm. For this analysis, a total of 4,994 cells, equally proportioned from each sample and cell type, were subsetted for trajectory construction. The scanpy.tl.paga function was utilized to construct a PAGA graph, which facilitated the visualization of cell population structures. This visualization was achieved using sc.tl.draw_graph, producing a force-directed graph drawing. The root cell for the trajectory was designated as a cell from the Mitotic CNCC-derived mesenchymal cell population. Diffusion pseudotime was calculated using scanpy.tl.diffmap and scanpy.tl.dpt functions. Additionally, the cellrank.kernels.PseudotimeKernel function from the CellRank package (version 2.0.2) was employed to construct the cellrank pseudotime kernel. This kernel was crucial for computing the cell-cell transition matrix. The pseudotime results were ultimately visualized as embeddings on the force-directed graph, providing insights into the developmental trajectory of fibroblast differentiation.

Motif analysis. Motif prediction on hEC1 sequence was performed using FIMO software in the MEME Suite (version 5.5.3). The motifs used in the prediction process were downloaded from HOCOMOCO database (v12) containing a collection of 1443 motifs, and the threshold *P*-value of 1E-4 was used to identify high-affinity binding

sites. After the preliminary results, we further used the single cell expression datasets from human embryos (GSE157329) to filter TFs that are not expressed (TFs must be expressed in cells expressing the *HMX1* gene in no less than 10% of the cells) with the *HMX1* gene. Additionally, through the ALX1/4 prediction *P*-value is larger than 1E-4, we also present it in the Figure 3 considering the very close distance to the very high MEIS1/PBX1 binding site and potentially influence. Finally, we further confirm the binding pattern of HD, HMG-box, and Coordinator in the D3-D4 region using sequence motif location tool (MoLoTool).

EChO analysis. The Enhanced Chromatin Occupancy (EChO) analysis was conducted following previously established protocols[31,63]. The process began with the generation of two bed files using bedtools: one representing the region file (TCF7L2 occupied peaks on hEC1 extended by ±500 bp) and the other being the fragment file. Detailed bedtools procedures for EChO can be found at https://github.com/FredHutch/EChO. Utilizing the region file, the foci mode of EChO was implemented to identify single base-pair foci as the centers of matrices. Additionally, the matrix mode was executed over a 400 bp window for each focus. This approach generated a matrix of base pair-resolution EChO fragment size values, facilitating subsequent visualization. For the analysis of high and low motif scores, TCF7L2 matrix profiles (including MA0523.1 and MA0523.2,) were retrieved from the JASPAR database. These profiles were scanned across a 40 bp window surrounding each focus. Motifs with scores exceeding 80% were categorized as high-scoring, while those above 70% were deemed low-scoring motifs.

Acknowledgments

We would like to thank the patients and their families for their participation in the study. In addition, we would like to thank Weiqiang Li, Peng Xiang (from Sun Yat-sen University) and Sahin Naqvi (from Stanford University) for helping in neural crest cell culture. We also would like to thank Yanxiao Zhang (from Westlake University) for advice on our LacZ experiments. This study was supported by grants from the National Natural Science Foundation of China (82202047 to X.X., 82171844 and 81970898 to Y-B.Z., and 81671933 to J.H), China Postdoctoral Science Foundation (2022M722228 to X.X.), and the Special Research Fund for Plastic Surgery Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College (YS202041 to Q.C.), the Key Research and Development Program of Guangzhou (202206060002 to X.F.), the Guangdong Provincial Pearl River Talents Program (2021QN02Y747 to X.F.) and a Stowers Family Endowment to T.C.C. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest statement

The authors have declared that no conflict of interest exists.

Author contributions

Y-B.Z., X.X., and H.L. designed and supervised the study. Q.C., Q.L., J.H., B.W., Q.Z., and Y-B.Z. recruited case samples. X.X., X.F., Q.H., Z.M., H.Z., Z.L., J.Z., J.L., and Q.C. planned and conducted laboratory experiments. X.X, X.H., B.X., W.X., R.L., Z.M., and T.C.C. analyzed data.

X.X., Y-B.Z., J.Z., X.F., T.C.C, S.E.A, and H.L. drafted and revised the manuscript. All authors have reviewed and contributed to the manuscript.

Data availability

All raw sequencing data and processed data in this study are currently being uploaded and will be made publicly accessible via the Gene Expression Omnibus (GEO) following the publication of the article. In addition, we used following public sequencing datasets: ChIP-seq datasets: GSE28874 (d11hCNCC), GSE145327 (P4hCNCC), GSE70751(P4hCNCC), GSE89435 (mCNCC), GSE211900 (Pinna). ATAC-seq datasets: GSE108517 (d11hCNCC), GSE145327 (hESC and P4hCNCC), GSE89436 (mCNCC), GSE211899 (Pinna). HiC dataset: GSM5944420 (CS17 head tissue). scRNA-seq dataset: GSE157329 (single-cell transcriptomes of 4- to 6-week human embryos). Additionally, human embryonic craniofacial epigenetic data from stages CS13, CS14, CS15, and CS17 were accessed from https://cotney.research.uchc.edu/craniofacial/.

Code availability

Computational procedures were conducted utilizing Python and R software, harnessing publicly available packages as delineated in the Methods and Reporting Summary sections. The analyses of data presented in this study did not rely on any unpublished custom computer code, mathematical algorithms, or software.

References

1 Atchley, W. R. & Hall, B. K. A model for development and evolution of complex morphological structures. *Biol Rev Camb Philos Soc* **66**, 101-157 (1991). https://doi.org/10.1111/j.1469-185x.1991.tb01138.x

2 Fuchs, J. C. & Tucker, A. S. Development and Integration of the Ear. *Curr Top Dev Biol* **115**, 213-232 (2015). https://doi.org/10.1016/bs.ctdb.2015.07.007

3 Anthwal, N. & Thompson, H. The development of the mammalian outer and middle ear. *J Anat* **228**, 217-232 (2016). https://doi.org/10.1111/joa.12344

4 Alasti, F. *et al.* A mutation in HOXA2 is responsible for autosomal-recessive microtia in an Iranian family. *Am J Hum Genet* **82**, 982-991 (2008). https://doi.org/10.1016/j.ajhg.2008.02.015

5 Zhang, Y., Fons, J. M., Hajihosseini, M. K., Zhang, T. & Tucker, A. S. An Essential Requirement for Fgf10 in Pinna Extension Sheds Light on Auricle Defects in LADD Syndrome. *Front Cell Dev Biol* **8**, 609643 (2020). https://doi.org/10.3389/fcell.2020.609643

6 Lausch, E. *et al.* TBX15 mutations cause craniofacial dysmorphism, hypoplasia of scapula and pelvis, and short stature in Cousin syndrome. *Am J Hum Genet* **83**, 649-655 (2008). https://doi.org/10.1016/j.ajhg.2008.10.011

7 Minoux, M. *et al.* Mouse Hoxa2 mutations provide a model for microtia and auricle duplication. *Development* **140**, 4386-4397 (2013). https://doi.org/10.1242/dev.098046

8 Quina, L. A. *et al.* Deletion of a conserved regulatory element required for Hmx1 expression in craniofacial mesenchyme in the dumbo rat: a newly identified cause of congenital ear malformation. *Dis Model Mech* **5**, 812-822 (2012). https://doi.org/10.1242/dmm.009910

9 Koch, C. T., Bruggmann, R., Tetens, J. & Drögemüller, C. A non-coding genomic duplication at the HMX1 locus is associated with crop ears in highland cattle. *PLoS ONE* **8**, e77841 (2013).

10 Si, N. *et al.* Duplications involving the long range HMX1 enhancer are associated with human isolated bilateral concha-type microtia. *J Transl Med* **18**, 244 (2020). https://doi.org/10.1186/s12967-020-02409-6

11 Rosin, J. M. *et al.* A distal 594 bp ECR specifies Hmx1 expression in pinna and lateral facial morphogenesis and is regulated by the Hox-Pbx-Meis complex. *Development* **143**, 2582-2592 (2016). https://doi.org/10.1242/dev.133736

12 Will, A. J. *et al.* Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat Genet* **49**, 1539-1545 (2017). https://doi.org/10.1038/ng.3939

13 Jindal, G. A. & Farley, E. K. Enhancer grammar in development, evolution, and disease: dependencies and interplay. *Dev Cell* **56**, 575-587 (2021). https://doi.org/10.1016/j.devcel.2021.02.016

14 Xu, X. *et al.* Novel risk factors for craniofacial microsomia and assessment of their utility in clinic diagnosis. *Hum. Mol. Genet.* (2021). https://doi.org/10.1093/hmg/ddab055

15 Mao, K. *et al.* FOXI3 pathogenic variants cause one form of craniofacial microsomia. *Nat Commun* **14**, 2026 (2023). https://doi.org/10.1038/s41467-023-37703-6

16 Zhang, Y. B. *et al.* Genome-wide association study identifies multiple susceptibility loci for craniofacial microsomia. *Nat Commun* **7**, 10605 (2016). https://doi.org/10.1038/ncomms10605

17 Minoux, M. *et al.* Gene bivalency at Polycomb domains regulates cranial neural crest positional identity. *Science* **355** (2017). https://doi.org/10.1126/science.aal2913

18 Selleri, L. & Rijli, F. M. Shaping faces: genetic and epigenetic control of craniofacial morphogenesis. *Nat Rev Genet* **24**, 610-626 (2023). https://doi.org/10.1038/s41576-023-00594-w

19 Minoux, M. & Rijli, F. M. Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development* **137**, 2605-2621 (2010). https://doi.org/10.1242/dev.040048

20 Depew, M. J., Lufkin, T. & Rubenstein, J. L. Specification of jaw subdivisions by Dlx genes. *Science* **298**, 381-385 (2002). https://doi.org/10.1126/science.1075703

21 Yuan, Y. *et al.* Spatiotemporal cellular movement and fate decisions during first pharyngeal arch morphogenesis. *Sci Adv* **6** (2020). https://doi.org/10.1126/sciadv.abb0119

22 Wilderman, A. *et al.* A distant global control region is essential for normal expression of anterior HOXA genes during mouse and human craniofacial development. *Nat Commun* **15**, 136 (2024). https://doi.org/10.1038/s41467-023-44506-2

23 Xu, Y. *et al.* A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat Cell Biol* **25**, 604-615 (2023). https://doi.org/10.1038/s41556-023-01108-w

24 Wilderman, A., VanOudenhove, J., Kron, J., Noonan, J. P. & Cotney, J. High-Resolution Epigenomic Atlas of Human Embryonic Craniofacial Development. *Cell Rep* **23**, 1581-1597 (2018). https://doi.org/10.1016/j.celrep.2018.03.129

25 Long, H. K. *et al.* Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a Craniofacial Disorder. *Cell Stem Cell* **27**, 765-783 e714 (2020). https://doi.org/10.1016/j.stem.2020.09.001

26 Kessler, S. *et al.* A multiple super-enhancer region establishes inter-TAD interactions and controls Hoxa function in cranial neural crest. *Nat Commun* **14**, 3242 (2023). https://doi.org/10.1038/s41467-023-38953-0

27 Laugsch, M. *et al.* Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell* **24**, 736-752 e712 (2019). https://doi.org/10.1016/j.stem.2019.03.004

28 Prescott, S. L. *et al.* Enhancer divergence and cis-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68-83 (2015). https://doi.org/10.1016/j.cell.2015.08.036

29 Chen, L. F. *et al.* Structural elements promote architectural stripe formation and facilitate ultra-long-range gene regulation at a human disease locus. *Mol Cell* (2023). https://doi.org/10.1016/j.molcel.2023.03.009

30 Kim, S. *et al.* DNA-guided transcription factor cooperativity shapes face and limb mesenchyme. *Cell* **187**, 692-711 e626 (2024). https://doi.org/10.1016/j.cell.2023.12.032

31 Meers, M. P., Janssens, D. H. & Henikoff, S. Pioneer Factor-Nucleosome Binding Events during Differentiation Are Motif Encoded. *Mol Cell* **75**, 562-575 e565 (2019). https://doi.org/10.1016/j.molcel.2019.05.025

32 Munroe, R. J. *et al.* Mouse H6 Homeobox 1 (Hmx1) mutations cause cranial abnormalities and reduced body mass. *BMC Dev Biol* **9**, 27 (2009). https://doi.org/10.1186/1471-213X-9-27

33 Allen, R. S., Biswas, S. K. & Seifert, A. W. Neural crest cells give rise to non-myogenic mesenchymal tissue in the adult murid ear pinna. *bioRxiv* (2023). https://doi.org/10.1101/2023.08.06.552195

34 Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat Commun* **11**, 3953 (2020). https://doi.org/10.1038/s41467-020-17740-1

35 Vallecillo-Garcia, P. *et al.* Odd skipped-related 1 identifies a population of embryonic fibro-adipogenic progenitors regulating myogenesis during limb development. *Nat Commun* **8**, 1218 (2017). https://doi.org/10.1038/s41467-017-01120-3

36 Contreras, O. & Harvey, R. P. Single-cell transcriptome dynamics of the autotaxin-lysophosphatidic acid axis during muscle regeneration reveal proliferative effects in mesenchymal fibro-adipogenic progenitors. *Front Cell Dev Biol* **11**, 1017660 (2023). https://doi.org/10.3389/fcell.2023.1017660

37 Phan, Q. M. *et al.* Lineage Commitment of Dermal Fibroblast Progenitors is Mediated by Chromatin De-repression. *bioRxiv* (2023). https://doi.org/10.1101/2023.03.07.531478

38 Han, X. *et al.* Runx2-Twist1 interaction coordinates cranial neural crest guidance of soft palate myogenesis. *Elife* **10** (2021). https://doi.org/10.7554/eLife.62387

39 Kemmler, C. L. *et al.* Conserved enhancers control notochord expression of vertebrate Brachyury. *Nat Commun* **14**, 6594 (2023). https://doi.org/10.1038/s41467-023-42151-3

40 Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. *Nat Rev Genet* **22**, 324-336 (2021). https://doi.org/10.1038/s41576-020-00311-x

41 Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**, 490-493 (2010). https://doi.org/10.1038/nature09158

42 Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239-243 (2018). https://doi.org/10.1038/nature25461

43 Wang, X. & Goldstein, D. B. Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease. *Am J Hum Genet* **106**, 215-233 (2020). https://doi.org/10.1016/j.ajhg.2020.01.012

44 Waymack, R., Fletcher, A., Enciso, G. & Wunderlich, Z. Shadow enhancers can suppress input transcription factor noise through distinct regulatory logic. *Elife* **9** (2020). https://doi.org/10.7554/eLife.59351

45 Thomas, H. F. *et al.* Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Mol Cell* **81**, 969-982 e913 (2021). https://doi.org/10.1016/j.molcel.2020.12.047

46 Perry, M. W., Boettiger, A. N. & Levine, M. Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo. *Proc Natl Acad Sci U S A* **108**, 13570-13575 (2011). https://doi.org/10.1073/pnas.1109873108

47 Choi, J. *et al.* Evidence for additive and synergistic action of mammalian enhancers during cell fate determination. *Elife* **10** (2021). https://doi.org/10.7554/eLife.65381

48 Carleton, J. B., Berrett, K. C. & Gertz, J. Multiplex Enhancer Interference Reveals Collaborative Control of Gene Regulation by Estrogen Receptor alpha-Bound Enhancers. *Cell Syst* **5**, 333-344 e335 (2017). https://doi.org/10.1016/j.cels.2017.08.011

49 Blayney, J. W. *et al.* Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* **186**, 5826-5839 e5818 (2023). https://doi.org/10.1016/j.cell.2023.11.030

50 Vockley, C. M. *et al.* Direct GR Binding Sites Potentiate Clusters of TF Binding across the Human Genome. *Cell* **166**, 1269-1281 e1219 (2016). https://doi.org/10.1016/j.cell.2016.07.049

51 Kim, S. & Wysocka, J. Deciphering the multi-scale, quantitative cis-regulatory code. *Mol Cell* **83**, 373-392 (2023). https://doi.org/10.1016/j.molcel.2022.12.032

52 Rajderkar, S. S. *et al.* Dynamic enhancer landscapes in human craniofacial development. *Nat Commun* **15**, 2030 (2024). https://doi.org/10.1038/s41467-024-46396-4

53 Zaugg, J. B. *et al.* Current challenges in understanding the role of enhancers in disease. *Nat Struct Mol Biol* **29**, 1148-1158 (2022). https://doi.org/10.1038/s41594-022-00896-3

54 Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613-626 (2012). https://doi.org/10.1038/nrg3207

55 Taskiran, II *et al.* Cell-type-directed design of synthetic enhancers. *Nature* (2023). https://doi.org/10.1038/s41586-023-06936-2

56 Wong, E. S. *et al.* Deep conservation of the enhancer regulatory code in animals. *Science* **370** (2020). https://doi.org/10.1126/science.aax8137

57 Marinic, M., Aktas, T., Ruf, S. & Spitz, F. An integrated holo-enhancer unit defines tissue and gene specificity of the Fgf8 regulatory landscape. *Dev Cell* **24**, 530-542 (2013). https://doi.org/10.1016/j.devcel.2013.01.025

58 Montavon, T. *et al.* A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**, 1132-1145 (2011). https://doi.org/10.1016/j.cell.2011.10.023

59 Dermitzakis, E. T. *et al.* Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**, 578-582 (2002). https://doi.org/10.1038/nature01251

60 Dermitzakis, E. T. *et al.* Evolutionary discrimination of mammalian conserved non-genic sequences (CNGs). *Science* **302**, 1033-1035 (2003). https://doi.org/10.1126/science.1087047

61 Attanasio, C. *et al.* Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* **342**, 1241006 (2013). https://doi.org/10.1126/science.1241006

62 Amin, S. *et al.* Hoxa2 selectively enhances Meis binding to change a branchial arch ground state. *Dev Cell* **32**, 265-277 (2015). https://doi.org/10.1016/j.devcel.2014.12.024

63 Azambuja, A. P. & Simoes-Costa, M. The connectome of neural crest enhancers reveals regulatory features of signaling systems. *Dev Cell* **56**, 1268-1282 e1266 (2021). https://doi.org/10.1016/j.devcel.2021.03.024

64 Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269 (2016). https://doi.org/10.1038/nature19800

65 Zhang, X. *et al.* Identification of focally amplified lineage-specific super-enhancers in human epithelial cancers. *Nat Genet* **48**, 176-182 (2016). https://doi.org/10.1038/ng.3470

66 Ma, S. *et al.* Developmentally regulated expression of integrin alpha-6 distinguishes neural crest derivatives in the skin. *Front Cell Dev Biol* **11**, 1140554 (2023). https://doi.org/10.3389/fcell.2023.1140554

67 Quiat, D. *et al.* An ancient founder mutation located between ROBO1 and ROBO2 is responsible for increased microtia risk in Amerindigenous populations. *Proc Natl Acad Sci U S A* **119**, e2203928119 (2022). https://doi.org/10.1073/pnas.2203928119

68 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009). https://doi.org/btp324 [pii]
10.1093/bioinformatics/btp324

69 Abecasis, G. R., Cherny, S. S., Cookson, W. O. & Cardon, L. R. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97-101 (2002). https://doi.org/10.1038/ng786

70 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). https://doi.org/10.1186/s13742-015-0047-8

71 Klambauer, G. *et al.* cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. *Nucleic Acids Res.* **40**, e69 (2012). https://doi.org/10.1093/nar/gks003

72 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014). https://doi.org/10.1016/j.cell.2014.11.021

73 Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598-606 (2015). https://doi.org/10.1038/ng.3286

74 Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642 e611 (2016). https://doi.org/10.1016/j.cell.2016.09.028

75 Stock, S. R. & Limaye, A. in *Developments in X-Ray Tomography VIII* (2012).

76 Rada-Iglesias, A. *et al.* Epigenomic annotation of enhancers predicts transcriptional regulators of human neural crest. *Cell Stem Cell* **11**, 633-648 (2012). https://doi.org/10.1016/j.stem.2012.07.006

77 Yu, F., Sankaran, V. G. & Yuan, G. C. CUT&RUNTools 2.0: a pipeline for single-cell and bulk-level CUT&RUN and CUT&Tag data analysis. *Bioinformatics* **38**, 252-254 (2021). https://doi.org/10.1093/bioinformatics/btab507

78 Freire-Pritchett, P. *et al.* Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *Elife* **6** (2017). https://doi.org/10.7554/eLife.21926

79 Cox, T. C., Camci, E. D., Vora, S., Luquetti, D. V. & Turner, E. E. The genetics of auricular development and malformation: new findings in model systems driving future directions for microtia research. *Eur J Med Genet* **57**, 394-401 (2014). https://doi.org/10.1016/j.ejmg.2014.05.003

80 Posey, K. L., Coustry, F. & Hecht, J. T. Cartilage oligomeric matrix protein: COMPopathies and beyond. *Matrix Biol* **71-72**, 161-173 (2018). https://doi.org/10.1016/j.matbio.2018.02.023

81 Melo, U. S. *et al.* Enhancer hijacking at the ARHGAP36 locus is associated with connective tissue to bone transformation. *Nat Commun* **14**, 2034 (2023). https://doi.org/10.1038/s41467-023-37585-8

Figure Legend

Fig. 1 | Duplication of non-coding region downstream of *HMX1* as a pathogenic factor in Bilateral Constricted Ears (BCEs).

a, Pedigrees of seven families with BCEs. Each individual's pedigree identity is indicated below their symbol. Gold arrows denote the proband in each family. Squares represent males and circles females; black shading indicates affected individuals, while gray shading denotes unknown phenotype. Triangles, stars, and rectangles symbolize chip analysis, target-capture sequencing, and whole-genome sequencing, respectively.

b, Representative images of affected ears from the proband patients and their families, in each pedigree.

c, Copy number duplication in each family, located in the intergenic region downstream of the *HMX1* gene, within the same TAD. The top panel displays Hi-C data from human embryo CS17 head tissue[22], with the black triangle marking the TAD region (TAD including BCE locus was marked red). In the middle panel, the H3K36me3 signal, a marker frequently observed at TAD boundaries, is depicted across this region. The lower panel specifies the duplication regions for each family, highlighting the minimum overlapping genomic area (chr4: 8691119-8728565, termed BCE locus) and the genes located within the TAD. Two protein-coding genes including *CPZ* and *HMX1* are highlighted.

Fig. 2 | Identification and characterization of candidate enhancers at the BCE locus.

a, Epigenomic profiling pinpoints enhancer elements at the BCE locus. Integrated epigenomic analysis of the BCE locus, combining *in vivo* data from human embryo craniofacial tissue at stages CS13-CS17 (including ear progenitor regions, upper part)[24] with *in vitro* data from hESCs-derived human Cranial Neural Crest Cells (hCNCCs, lower part)[28]. CNV region shared across seven pedigrees is outlined by the golden rectangle. The chromatin state is represented through color coding, as defined on the Cotney lab craniofacial epigenome website (https://cotney.research.uchc.edu/craniofacial/). The epigenetic marks H3K4me3, H3K27me3, H3K27ac, and the presence of the transcriptional coactivator p300, in addition to H3K4me1 and CTCF signals in hCNCCs, are displayed. Candidate enhancers EC1, EC2, and EC3 are identified based on both datasets and are indicated by black rectangles.

b, Differential chromatin interactions link CE enhancers to *HMX1* promoter. Promoter Capture Hi-C (PC-HiC) analysis demonstrates the differential physical interactions between the CE locus and the *HMX1* promoter in hESCs compared to hCNCCs. The line color denotes the interaction strength, and the TAD boundary identified in Fig. 1 is depicted. ATAC-seq data and PC-HiC data for hESCs are sourced from GSE145327[25] and GSE86821[78], respectively. Panels a and b share the same physical genomic coordinates.

Fig. 3 | Spatiotemporal regulation of *HMX1* expression by multipartite enhancers

a, Epigenetic profiling at the BCE locus in early (D11) and late hCNCCs (P4).

b, Schematic of hESCs differentiation to PA-like hCNCCs: The process involves transitioning H9 cell line hESCs to early hCNCCs using NCC induction (NCCI) medium, followed by BMP2/ChIR treatment to reach late hCNCCs stage, and finally achieving PA-like hCNCCs with retinoic acid (RA) introduction.

c, *HMX1* expression analysis across four cell states, presented as mean ± S.E.M. (n = 6).
d-e, Luciferase assays evaluating candidate enhancers at the BCE locus in hESC (**d**), and PA-like hCNCC (**e**), including SV40 enhancer (positive control) and empty vector (negative control). Results from three independent experiments, each with four technical replicates (n = 12), are shown.
f, CRISPR/Cas9 strategy to delete two active enhancers (hEC1 and hEC2, ~9 kb), and gene expression comparison across four cell states in WT and knockout cell lines (n = 3).
g, Spatiotemporal *Hmx1* expression from E9.5 to E14.5: Left panel – WISH image and fluorescent images of Hmx1-P2A-EGFP transgenic mouse reporter line (**Extended Data Fig. 4j**); Right panel - *Hmx1* expression predominantly in the PA2 zone and pinna region, and the spatial location of *Hmx1* expression at E11.5 is manually divided into α (proximal), β (distal), γ (caudal to PA2) positions according to corresponding PA2 zones. cm, craniofacial mesenchyme; bp, basal pinna; lp, lower part of pinna; dp, distal pinna; pmp, the proximal region of the mandibular prominence.
h-m, LacZ assays during pinna development for various enhancers (hEC1, hEC2.1, hEC2, hEC1+hEC2, hEC3, and hEC1+hEC2+hEC3). The orange number in the bottom right corner of each embryo indicates the count of positive LacZ staining embryos, with the black number denoting the total count of transgenic embryos examined (**Extended Data Fig. 4d-i**). The black star on the E9.5 embryo of hEC2.1 marks weak activity at the basal of PA2. Two clearly visible staining positions in the upper distal part of pinna and lower pinna part adjacent to the proximal mandibular region on hEC1+hEC2+hEC3 E14.5 embryo compared with other enhancer constructs were marked as red triangle.

Fig. 4 | The grammar of the hEC1 enhancer is shaped by multifactorial transcription factors
a, Motif analysis on the hEC1 sequence using FIMO is shown. The color intensity represents the affinity of the predicted TFBSs, with the previously identified core 32bp marked[11]. Five clusters of motifs, including D1 (162 bp), D2 (141 bp), D3 (182 bp), D4 (178 bp), and D5 (182 bp), are sequentially deleted.
b, hEC1 activity is evaluated using *in vitro* (luciferase assay in PA-like hCNCCs line, upper panel) and *in vivo* (LacZ staining, bottom panel) strategies, across three independent experiments with two technical replicates each (n = 6). The red number in the bottom right corner of the embryo represents the number of positive LacZ staining embryos, see also **Extended Data Fig. 5a**.
c, Dissection of motif structure across a core region linking D3 with D4. Two 76 bp duplication regions previously reported in sheep and cow are shown in the top panel, respectively. The middle panel displays DNA alignment across human, mouse, sheep and cow using T-coffee software. Four TFBSs were highlighted and the underlie marked the core motif sequence. The bottom panel shows motifs from HOCOMOCO V12 database. Two deletions including ΔD3-1 and ΔD4-1 used in **Fig. 5g** are shown.
d, CUT&RUN assay of H3K27ac, TWIST1 and TCF7L2 across the BCE locus.
e, Local minimal DNA protection by TF-DNA binding (EChO) reveals multiple TCF7L2-binding sites across the hEC1 sequence. Purple and blue lines represent foci positions associated with high- and low-score TCF7L2 motifs, respectively.
f, An EMSA assay reveals the binding of human MEIS1 protein to the predicted motif binding site in c.
g, A luciferase assay demonstrates that removal of ΔD4-1 containing HMG-box and Coordinator dramatically diminish the activity of hEC1 in hCNCCs (n = 4).

h, A complex heatmap shows TFs expression across four craniofacial regional zones: frontonasal process (FNP), maxilla (Mx), mandible (Md), and PA2 at E10.5 and E11.5[17]. Gene expression is scaled and clustering is performed.

i, WISH images of four critical TFs including *Meis1*, *Tcf7l2*, *Twist1* and *Tcf4*. Images are from Emage database (*https://www.emouseatlas.org/emage/home.php*).

Fig. 5 | mEC1 Knock-in in mouse recapitulates human constricted ear phenotypes, and mEC1-dependent *Hmx1* expression in mCNCCs is important for ear development.

a, A CRISPR/Cas9 genome editing strategy is used to insert mEC1.

b, Sanger sequencing of PCR product in *mEC1*$^{dup/dup}$ using two primers flanking mEC1.

c, The expression pattern of mEC1 at E14.5, as shown using enSERT technology, is presented (n = 7; see also **Extended Data Fig. 6b**).

d, Illustration of mouse pinna structure adapted from *Cox et al.,*[79], highlighting components such as the tragus, anti-tragus, scapha, helix, and anti-helix.

e, Side view comparison of wild type and homozygous mEC1 knock-in mouse models, with the red rectangle highlighting the altered outer ear phenotype. Scale bars represent 1 cm.

f, Statistical analysis of the inferior pinna area, marked by the red dotted box in e, is shown. Data are presented as mean ± S.E.M. (n = 15).

g, Top view comparison of wild type and homozygous mEC1 knock-in mouse models, with sections within the red rectangle enlarged in the bottom panel. Notably, the helix rim is wrinkled in the transgenic mouse model, as indicated by the red angle. Scale bars measure 1 cm.

h, Statistical analysis is conducted on the angle representing the extent of helix folding. Data are shown as mean ± SEM (n = 15).

i, j, Whole-mount in situ hybridization (WISH) showing *Hmx1* expression in WT and *mEC1*$^{dup/dup}$ embryos at E10.5 (i) and E14.5 (j). The white arrow indicates expanded *Hmx1* expression in the middle PA2 region (i) and in the upper part of the outer ear (j). Scale bars are 50 μm (i) and 1000 μm (j), respectively.

k, Loss of *Hmx1* expression through three different strategies results in a similar low-set and protruding ear phenotype. These strategies include *Hmx1*$^{fl/fl}$ (homozygous flox allele), *Wnt1::Cre;Hmx1*$^{fl/fl}$ (conditional deletion of *Hmx1* in neural crest cells), and *mEC1*$^{del/del}$ (homozygous deletion of mEC1). The 'dumbo' mutation, representing a missense mutation in exon 1 of Hmx1, results in a truncated protein. Scale bars measure 1000 μm.

l, Micro-CT analysis of ear-related structure in different mouse models at eight weeks, nine months and one year. The gold arrow points to the mastoid process.

Fig. 6 | *Hmx1* misexpression in mouse leads to widespread abnormal development in pinna structures including cartilage, fibroblasts, muscle, and epidermis

a, RNA-seq analysis compares the whole pinna structure between wild-type and E14.5 *mEC1*$^{dup/dup}$ mice. Each dot represents a single gene. Differentially expressed genes (DEGs) are defined as FDR ≤ 0.05, log2 CPM ≥ 1, and |log2 FC| ≥ 0. Up- and down-regulated genes are shown as gold and blue dots, respectively. DEGs with |log2 FC| ≥ 1 are highlighted and labeled, previous reported cartilage development-related genes including *Comp*[80] and *Arhgap36*[81] were marked as red.

b, Gene Ontology (GO) enrichment analysis of 284 downregulated genes is performed. The top 30 GO terms are displayed and grouped into four clusters.

c, Masson's trichrome staining at E14.5 (upper and lower pinna parts) is shown for WT and *mEC1*$^{dup/dup}$. Scale bars measure 200 μm.

d, UMAP dimensional reduction visualizes *Hmx1*$^{+}$ expressed cells combined across E10.5, E12.5, and E14.5.

e, A dot plot displays two marker genes for each cluster identified within the fibroblasts. The cell clusters of 1-7 is same as in d.

f, Developmental trajectory analysis from E10.5 CNCC mesenchymal cells (cluster 6) to the fibroblasts defined in (e) is illustrated. The cell clusters of 1-7 is same as in d.

g, The expression of cartilage, fibroblast, muscle, and epidermis development-related downregulated DEGs across nine cell clusters (including 7 fibroblast, chondrocyte, and myogenic cell clusters) is shown. The heatmap on the left panel depicts the fold change of each gene.

h, i, *In situ* hybridization staining for *Hmx1* and *Dmkn* expression in wild-type and *mEC1*$^{dup/dup}$ embryos at upper (**h**) and lower (**i**) regions of the pinna.

Fig. 7 | Model

a, Stage-specific enhancer activities in hCNCCs in the BCE locus. EC2 may be initiated at the early hCNCCs stage, driving weak expression of *HMX1*; EC1 drives most HMX1 transcriptional output at the PA-like hCNCCs stage, and EC3 may serve as a structural element.

b, Coordinated and synergistic PI-HEC patterning in PA2 and pinna zones. EC2.1, EC2, and EC3, represented by varying color intensities, are weak enhancers.

c, A motif pattern consisting of TALE-type HDs (HOXA2/MEIS1/PBX1), HMG box (TCF7L1/2), and Coordinator (TWIST1/TCF4) endows the dominant EC1 with specificity and activity.

d, A comprehensive disease model for BCE. Based on scRNA-seq and *mEC1*$^{dup/dup}$ mouse model, we demonstrated that *Hmx1* is mainly responsible for CNCCs-derived fibroblasts development in pinna structure, and duplication of EC1 causes aberrant expression of *Hmx1* in fibroblasts and ectopic expression in the other zones of the pinna, disrupting the gene network in cartilage, muscle, epidermis, and leading to outer ear malformation.

Extended Data Fig. 1 | Identification and fine mapping of the BCE Locus in three BCE pedigrees.

a, Linkage analysis using whole-genome chip-based data in three BCE families revealed a pronounced signal on chromosome 4, marked by a LOD score of 7.8. This highlights a shared chromosomal region associated with the BCE phenotype.

b, A detailed representation of the chromosome 4 linkage signal, including individual LOD scores for each family. This demonstrates the distinct yet overlapping linkage signals among these families.

c, Fine mapping of the linkage signals identified copy number duplications through target capture sequencing. For each family, one healthy individual was randomly selected as a control. This panel compares the sequencing coverage depth ratio between the remaining family members and this healthy control. Red lines illustrate the read depth ratio in the linked chromosomal region for patients compared to the control, whereas blue lines depict the ratio between healthy family members and the control.

Extended Data Fig. 2 | Single-Cell RNA sequencing and epigenetic landscaping of the BCE locus in human and mouse craniofacial development.

a, UMAP clustering and *HMX1* expression profiling in early human embryonic craniofacial regions. scRNA-seq analysis showcases the UMAP clustering of cell types from the human early organogenesis atlas, with cell type annotations as per Xu, Y. *et al.* (2023)[23]. The central feature plot visualizes the expression levels of the *HMX1* gene across different cell clusters, while the right panel provides a detailed view of *HMX1* expression within craniofacial mesenchyme cells specifically in the pharyngeal arch (PA) structures.

b, Conservation of epigenetic regulatory elements in mouse E8.5 NCC progenitors and E10.5 Md, PA2 NCCs. This panel illustrates the epigenetic landscape at the BCE locus in mouse neural crest cell progenitors at E8.5 and E10.5 using public available datasets[17]. Highlighted in yellow are the regions orthologous to human enhancers and the *Hmx1* promoter.

c, Dynamics of the BCE locus epigenetic landscape from mouse pharyngeal arch 2 (PA2) to pinna development. Here, the evolution of the epigenetic context is traced from E10.5 PA2 to E14.5 pinna tissues using public available datasets[26]. Blue and red arcs indicate chromatin interactions from two Promoter Capture Hi-C (PC-HiC) dataset replicates. ATAC-seq and CTCF tracks are also shown.

Extended Data Fig. 3 | *In vitro* differentiation system from H9 cells to PA-like hCNCCs state

a, Representative images depicting the progression from H9 cells to the late hCNCCs state.

b, Immunofluorescence analysis using markers p75, NR2F1, AP2, and SOX9 to verify cell identity during the differentiation process.

c, Quantitative real-time PCR (qPCR) is used to further evaluate the efficiency of hCNCCs differentiation. Results are presented as mean ± SEM (n = 3).

d, Representative images illustrating the transition from late hCNCCs to PA-like hCNCCs state.

e, Gene expression analysis of pharyngeal arch 2-specific TFs, including *HOXA2*, *MEIS1*, *PBX1*, *TBX15*, and *DLX6*. Data are expressed as mean ± SEM (n = 4).

Extended Data Fig. 4 | Spatiotemporal regulation of *Hmx1* expression by multipartite enhancers

a, A CRISPR/Cas9 strategy is employed to delete the hEC1+hEC2 region (~9 kb). Six gRNA sequences are designed along with corresponding primer pairs to validate knockout efficiency, with gel analysis images provided.

b, Representative images illustrate early hCNCCs migration in both wild-type (WT) and knockout (KO) cell lines. In the right panel, FACS analysis of CD271 (p75) in early hCNCCs is shown for both WT and KO cell lines.

c, Gene expression analysis of *SOX9* and *TWIST1*, two important TFs in hCNCCs, compares WT and KO. Data are presented as mean ± SEM (n = 4).

d-i, LacZ assays are conducted to track pinna development for various enhancers: hEC1 (d), hEC2.1 (e), hEC2 (f), hEC1+hEC2 (g), hEC3 (h), and hEC1+hEC2+hEC3 (i). Results showing all positive LacZ staining embryos for each enhancer are included.

j, The construction of the Hmx1-P2A-EGFP transgenic reporter line is described, with Sanger sequencing results of this mouse model presented.

Extended Data Fig. 5 | HD, HMG box and Coordinator TFs coordinately regulate hEC1 specificity and activity.

a, Display all positive LacZ staining embryos for ΔD1, ΔD2, ΔD3, ΔD4, and ΔD5. Arrows point to faint staining signals.

b, A schematic process of DNA pull-down using PA2 tissue from E11.5 mouse embryos and biotinylated hEC1 is presented.

c, Gene enrichment analysis of proteins identified in the DNA pull-down assay is conducted.. For fifteen DNA-binding proteins, InterPro enrichment is further performed on the DAVID website to detect TF classification.

d, A heatmap is presented, showing IgG, H3K27ac, and TCF7L2 binding profiles across ±2 kb from TCF7L2 peak summits (4560). Results from two independent biological replicates are shown.

e, *De novo* TCF7L2 motif analysis is performed using the MEME toolkit. The top-ranked motif is similar to TF7L2_HUMAN.H11MO.O.A; the eighth-ranked motif is similar to TWIST1_HUMAN.H11MO.O.A.

Extended Data Fig. 6 | Pinna malformation is driven by either the expansion or the loss of spatial *Hmx1* expression

a, A schematic diagram of the transgenic mEC1$^{dup/dup}$ mouse model is presented. A primer pair, P1 and P2, flanking the mEC1 element, is designed for genotyping: the wildtype allele measures 2374 bp and the transgenic allele 3487 bp. The right panel displays gel electrophoresis results for wildtype (WT), heterozygote (Het), and homozygote (Homo) in the white rectangle.

b, This figure Depicts all positive LacZ staining in mouse embryos for mEC1 at E14.5. The upper panel shows the whole embryo, while the bottom panel presents zoomed-in images of pinna staining. Black arrows and stars indicate distinct staining patterns compared to hEC1 LacZ results, particularly in the frontonasal prominence and fore- and hind-limbs.

c, Ear and eye malformations are compared between wildtype and *mEC1*$^{dup/dup}$ mice. Ear malformation in *mEC1*$^{dup/dup}$ is clearly visible, as indicated by the white arrow. Eye malformation, likely representing cataract, is marked with a red rectangle in older mice.

d, A schematic diagram of the transgenic *wnt1::Cre;Hmx1*$^{fl/fl}$ mouse model is shown. Primers F1, R1 are used to identify WT and *Hmx1*$^{fl/fl}$; and F2, R1 for genotyping *Wnt1::Cre;Hmx1*$^{fl/fl}$.

Extended Data Fig. 7 | ScRNA-seq reveals the critical role of *Hmx1* on the ear fibroblast development.

a, This figure depicts Lineage tracing of *Hmx1*$^{+}$ cells from E10.5 to E14.5 using the Hmx1-P2A-EGFP mouse reporter line. Fluorescence images for each timepoint are shown, with tissues micro-dissected for dissociation marked by white rectangles. *GFP*$^{+}$ cells are collected using FACS.

b, UMAP dimensional reduction visualization of *Hmx1*$^{+}$ expressing cells for each timepoint is presented.

c, Marker genes used for cell cluster annotation are identified.

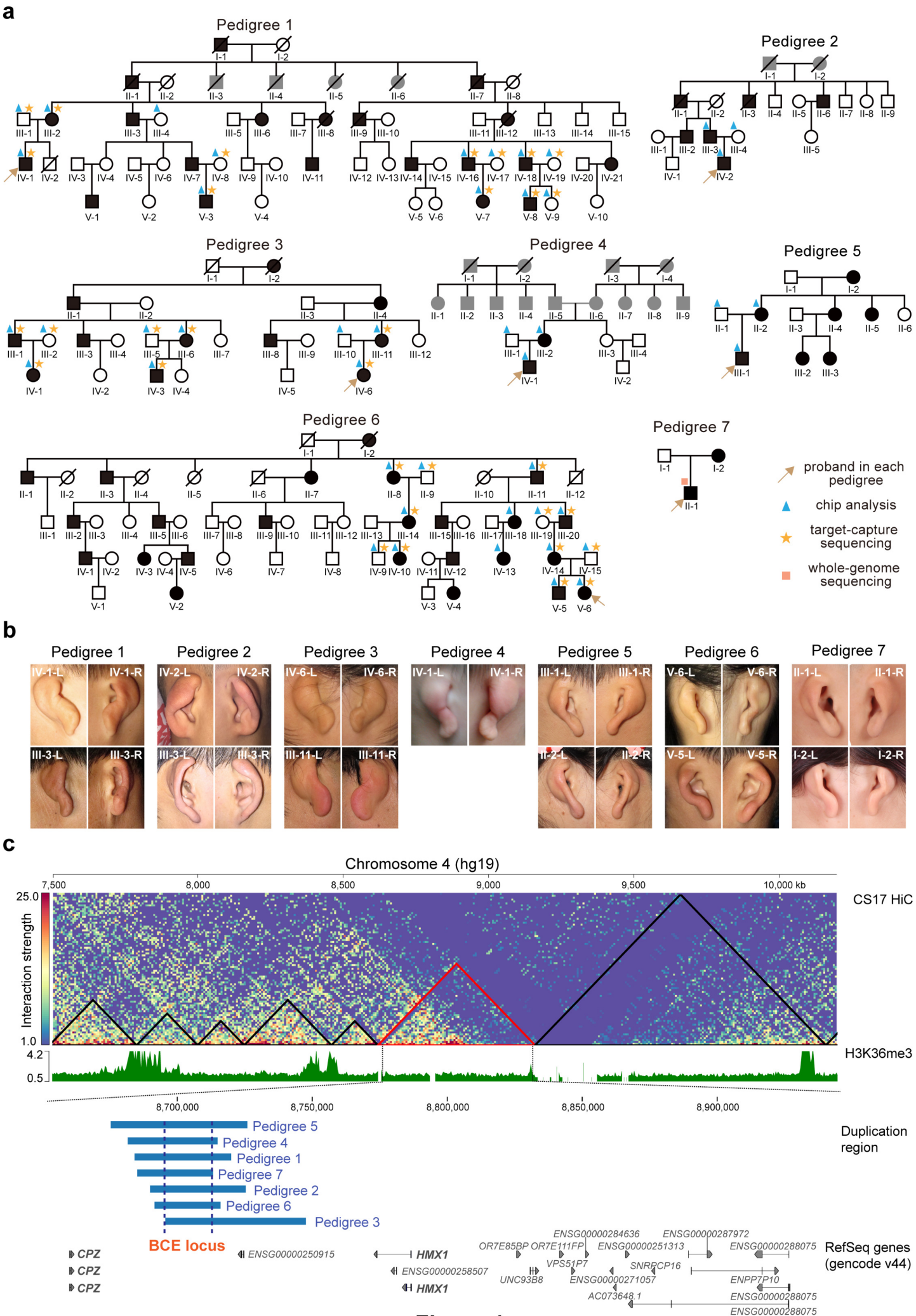

a
Pedigree 1
Pedigree 2
Pedigree 3
Pedigree 4
Pedigree 5
Pedigree 6
Pedigree 7
proband in each pedigree
chip analysis
target-capture sequencing
whole-genome sequencing
b
Pedigree 1
Pedigree 2
Pedigree 3
Pedigree 4
Pedigree 5
Pedigree 6
Pedigree 7
IV-1-L
IV-1-R
III-3-L
III-3-R
IV-2-L
IV-2-R
III-3-L
III-3-R
IV-6-L
IV-6-R
III-11-L
III-11-R
IV-1-L
IV-1-R
III-1-L
III-1-R
II-2-L
II-2-R
V-6-L
V-6-R
V-5-L
V-5-R
II-1-L
II-1-R
I-2-L
I-2-R
c
Chromosome 4 (hg19)
7,500
8,000
8,500
9,000
9,500
10,000 kb
25.0
1.0
Interaction strength
CS17 HiC
4.2
0.5
H3K36me3
8,700,000
8,750,000
8,800,000
8,850,000
8,900,000
Pedigree 5
Pedigree 4
Pedigree 1
Pedigree 7
Pedigree 2
Pedigree 6
Pedigree 3
Duplication region
BCE locus
CPZ
CPZ
CPZ
ENSG00000250915
HMX1
ENSG00000258507
HMX1
ENSG00000284636
ENSG00000287972
OR7E85BP
OR7E111FP
ENSG00000251313
ENSG00000288075
VPS51P7
SNRPCP16
UNC93B8
ENSG00000271057
ENPP7P10
AC073648.1
ENSG00000288075
ENSG00000288075
RefSeq genes (gencode v44)

Figure 1

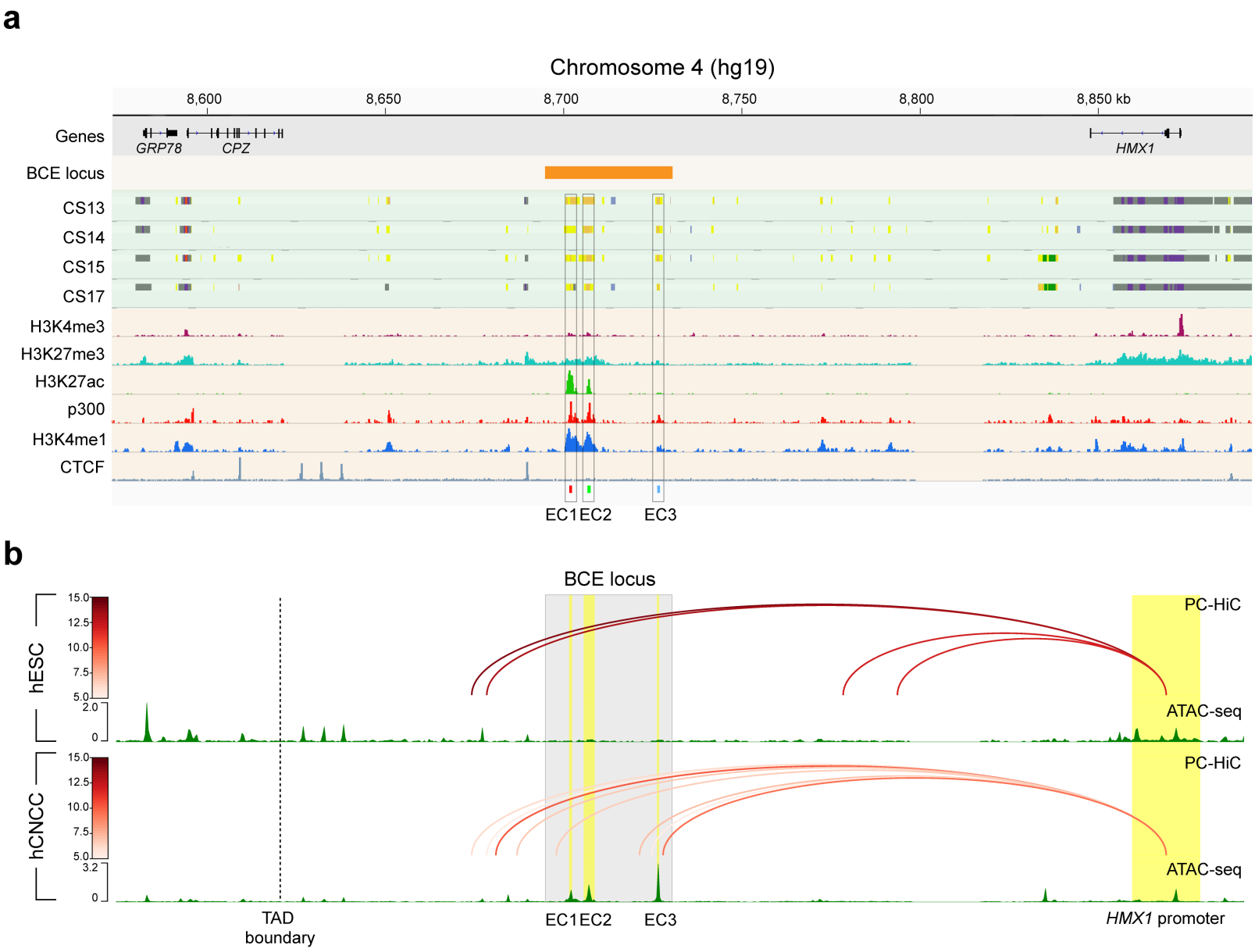

a
Chromosome 4 (hg19)
8,600
8,650
8,700
8,750
8,800
8,850 kb
Genes
GRP78
CPZ
HMX1
BCE locus
CS13
CS14
CS15
CS17
H3K4me3
H3K27me3
H3K27ac
p300
H3K4me1
CTCF
EC1
EC2
EC3
b
BCE locus
hESC
hCNCC
15.0
12.5
10.0
7.5
5.0
2.0
0
3.2
PC-HiC
ATAC-seq
TAD
boundary
HMX1 promoter

Figure 2

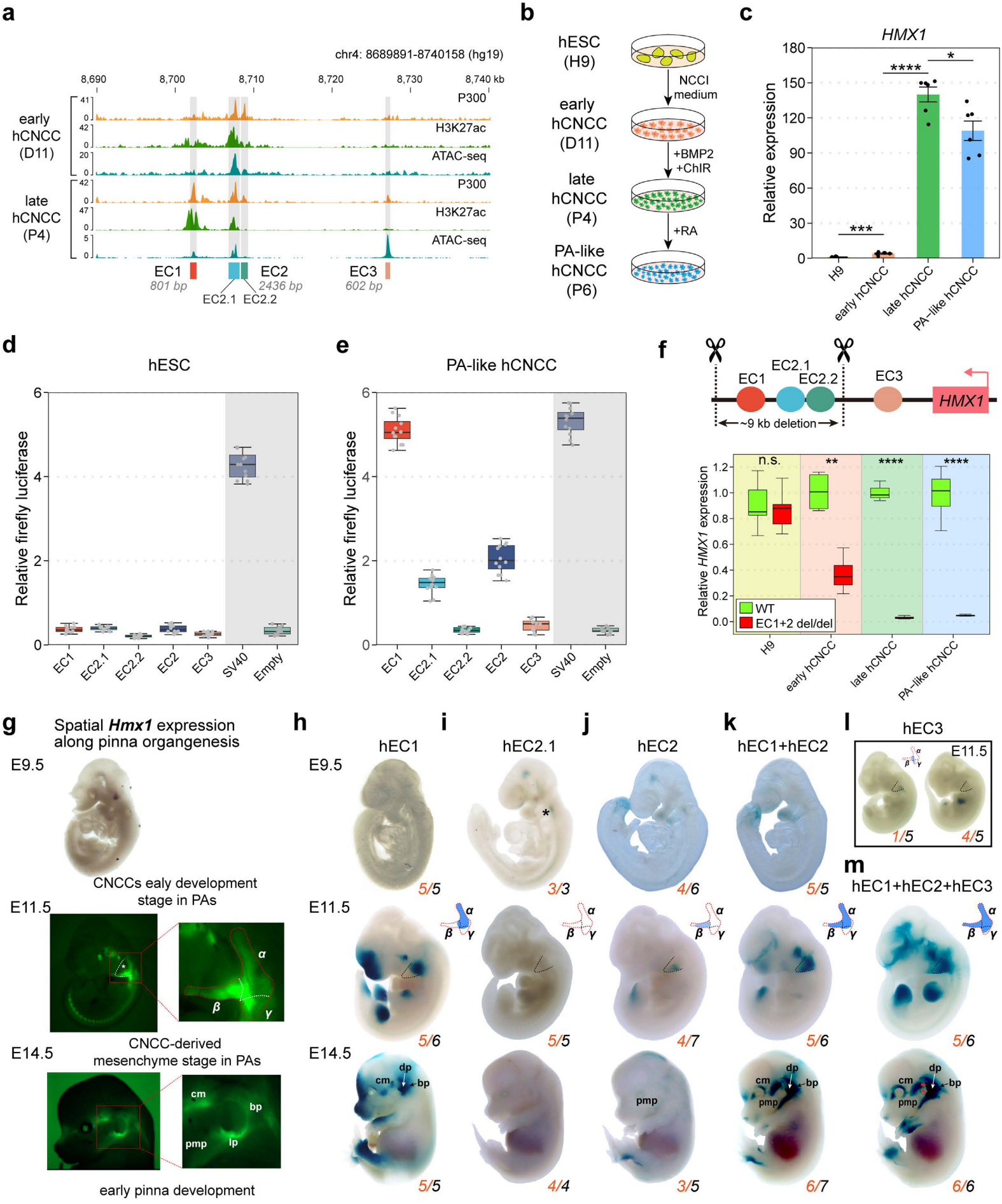

a
chr4: 8689891-8740158 (hg19)
8,690
8,700
8,710
8,720
8,730
8,740 kb
early hCNCC (D11)
late hCNCC (P4)
P300
H3K27ac
ATAC-seq
EC1
801 bp
EC2
2436 bp
EC3
602 bp
EC2.1
EC2.2
b
hESC (H9)
NCCI medium
early hCNCC (D11)
+BMP2 +ChIR
late hCNCC (P4)
+RA
PA-like hCNCC (P6)
c
HMX1
Relative expression
***
****
*
H9
early hCNCC
late hCNCC
PA-like hCNCC
d
hESC
Relative firefly luciferase
EC1
EC2.1
EC2.2
EC2
EC3
SV40
Empty
e
PA-like hCNCC
f
EC1
EC2.1
EC2.2
EC3
HMX1
~9 kb deletion
Relative HMX1 expression
n.s.
**
****
WT
EC1+2 del/del
g
Spatial Hmx1 expression along pinna organegenesis
E9.5
CNCCs ealy development stage in PAs
E11.5
CNCC-derived mesenchyme stage in PAs
E14.5
early pinna development
α
β
γ
cm
bp
pmp
lp
h
hEC1
i
hEC2.1
j
hEC2
k
hEC1+hEC2
l
hEC3
E11.5
1/5
4/5
m
hEC1+hEC2+hEC3
dp
5/5
3/3
4/6
5/5
5/6
5/5
4/7
5/6
5/6
4/4
3/5
6/7
6/6

Figure 3

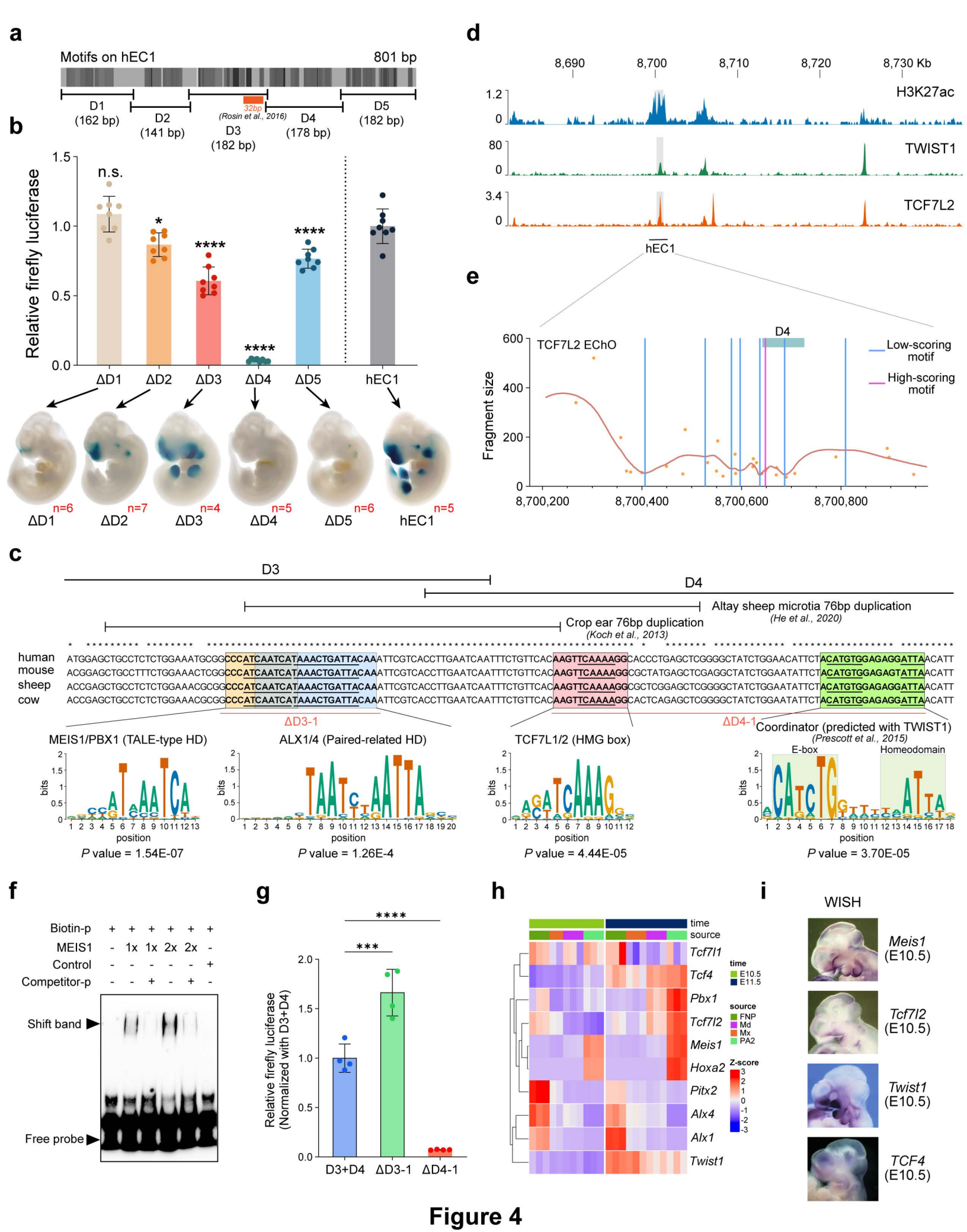

a
Motifs on hEC1
801 bp
D1 (162 bp)
D2 (141 bp)
32bp (Rosin et al., 2016)
D3 (182 bp)
D4 (178 bp)
D5 (182 bp)
b
Relative firefly luciferase
n.s.
*
****
****
****
ΔD1
ΔD2
ΔD3
ΔD4
ΔD5
hEC1
n=6
n=7
n=4
n=5
n=6
n=5
c
D3
D4
Altay sheep microtia 76bp duplication (He et al., 2020)
Crop ear 76bp duplication (Koch et al., 2013)
human
mouse
sheep
cow
ΔD3-1
ΔD4-1
MEIS1/PBX1 (TALE-type HD)
ALX1/4 (Paired-related HD)
TCF7L1/2 (HMG box)
Coordinator (predicted with TWIST1) (Prescott et al., 2015)
E-box
Homeodomain
bits
position
P value = 1.54E-07
P value = 1.26E-4
P value = 4.44E-05
P value = 3.70E-05
d
8,690
8,700
8,710
8,720
8,730 Kb
H3K27ac
TWIST1
TCF7L2
hEC1
e
D4
TCF7L2 EChO
Low-scoring motif
High-scoring motif
Fragment size
8,700,200
8,700,400
8,700,600
8,700,800
f
Biotin-p
MEIS1
Control
Competitor-p
Shift band
Free probe
g
Relative firefly luciferase (Normalized with D3+D4)
****
***
D3+D4
ΔD3-1
ΔD4-1
h
time
source
Tcf7l1
Tcf4
Pbx1
Tcf7l2
Meis1
Hoxa2
Pitx2
Alx4
Alx1
Twist1
E10.5
E11.5
FNP
Md
Mx
PA2
Z-score
i
WISH
Meis1 (E10.5)
Tcf7l2 (E10.5)
Twist1 (E10.5)
TCF4 (E10.5)

Figure 4

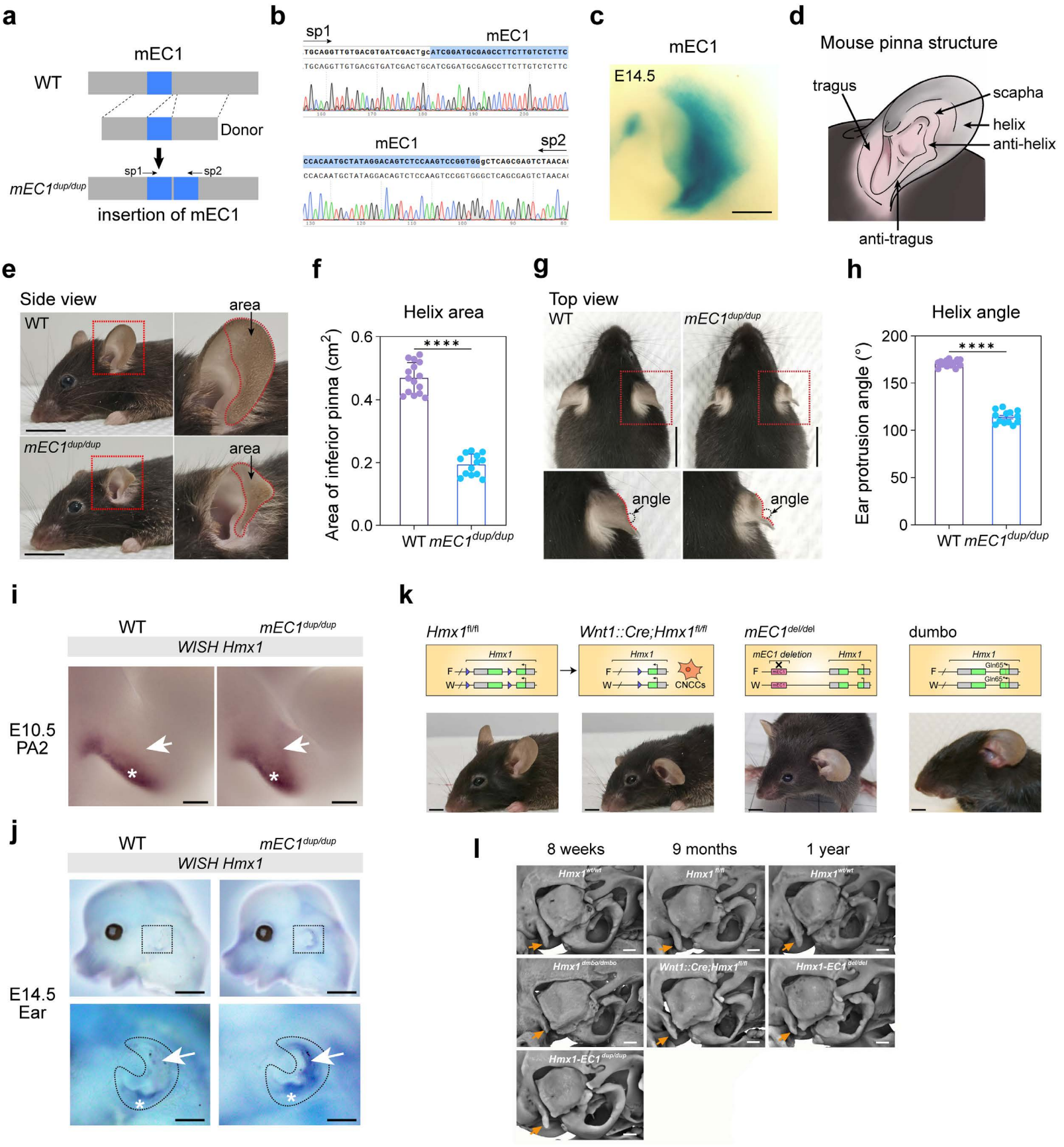

a
mEC1
WT
Donor
sp1
sp2
mEC1dup/dup
insertion of mEC1
b
sp1
mEC1
TGCAGGTTGTGACGTGATCGACTgcATCGGATGCGAGCCTTCTTGTCTCTTC
TGCAGGTTGTGACGTGATCGACTGCATCGGATGCGAGCCTTCTTGTCTCTTC
mEC1
sp2
CCACAATGCTATAGGACAGTCTCCAAGTCCGGTGGgCTCAGCGAGTCTAACA
CCACAATGCTATAGGACAGTCTCCAAGTCCGGTGGGCTCAGCGAGTCTAACA
c
mEC1
E14.5
d
Mouse pinna structure
tragus
scapha
helix
anti-helix
anti-tragus
e
Side view
WT
area
mEC1dup/dup
area
f
Helix area
Area of inferior pinna (cm2)
0.6
0.4
0.2
0.0
****
WT mEC1dup/dup
g
Top view
WT
mEC1dup/dup
angle
angle
h
Helix angle
Ear protrusion angle (°)
200
150
100
50
0
****
WT mEC1dup/dup
i
WT
mEC1dup/dup
WISH Hmx1
E10.5
PA2
k
Hmx1fl/fl
Wnt1::Cre;Hmx1fl/fl
mEC1del/del
dumbo
Hmx1
F
W
CNCCs
mEC1 deletion
Gln65*
j
WT
mEC1dup/dup
WISH Hmx1
E14.5
Ear
l
8 weeks
9 months
1 year
Hmx1wt/wt
Hmx1fl/fl
Hmx1wt/wt
Hmx1dmbo/dmbo
Wnt1::Cre;Hmx1fl/fl
Hmx1-EC1del/del
Hmx1-EC1dup/dup

Figure 5

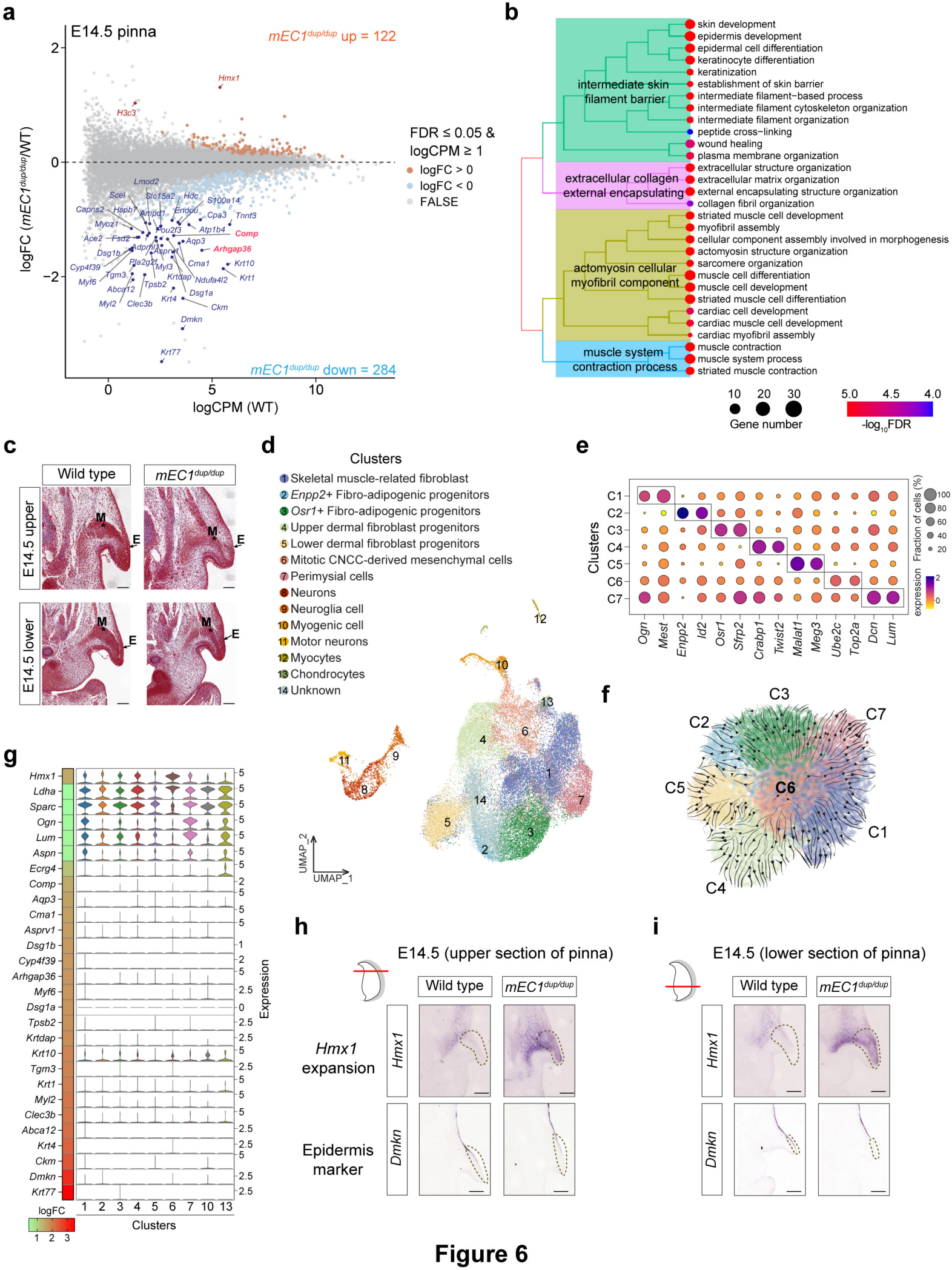

a
E14.5 pinna
mEC1dup/dup up = 122
mEC1dup/dup down = 284
FDR ≤ 0.05 & logCPM ≥ 1
logFC > 0
logFC < 0
FALSE
logFC (mEC1dup/dup/WT)
logCPM (WT)
Hmx1
H3c3
Lmod2
Scel
Slc15a2
Hdc
S100a14
Capns2
Hspb7
Arnpd1
Endou
Cpa3
Tnnt3
Myoz1
Pou2f3
Atp1b4
Comp
Ace2
Fsd2
Aqp3
Arhgap36
Dsg1b
Asprv1
Pla2g2f
Myl3
Cma1
Krt10
Cyp4f39
Tgm3
Krtdap
Ndufa4l2
Krt1
Myf6
Abca12
Tpsb2
Dsg1a
Myl2
Clec3b
Krt4
Ckm
Dmkn
Krt77
b
intermediate skin filament barrier
extracellular collagen external encapsulating
actomyosin cellular myofibril component
muscle system contraction process
skin development
epidermis development
epidermal cell differentiation
keratinocyte differentiation
keratinization
establishment of skin barrier
intermediate filament−based process
intermediate filament cytoskeleton organization
intermediate filament organization
peptide cross−linking
wound healing
plasma membrane organization
extracellular structure organization
extracellular matrix organization
external encapsulating structure organization
collagen fibril organization
striated muscle cell development
myofibril assembly
cellular component assembly involved in morphogenesis
actomyosin structure organization
sarcomere organization
muscle cell differentiation
muscle cell development
striated muscle cell differentiation
cardiac cell development
cardiac muscle cell development
cardiac myofibril assembly
muscle contraction
muscle system process
striated muscle contraction
Gene number
-log10FDR
c
Wild type
mEC1dup/dup
E14.5 upper
E14.5 lower
M
E
d
Clusters
1 Skeletal muscle-related fibroblast
2 Enpp2+ Fibro-adipogenic progenitors
3 Osr1+ Fibro-adipogenic progenitors
4 Upper dermal fibroblast progenitors
5 Lower dermal fibroblast progenitors
6 Mitotic CNCC-derived mesenchymal cells
7 Perimysial cells
8 Neurons
9 Neuroglia cell
10 Myogenic cell
11 Motor neurons
12 Myocytes
13 Chondrocytes
14 Unknown
UMAP_1
UMAP_2
e
Clusters
Fraction of cells (%)
expression
Ogn
Mest
Enpp2
Id2
Osr1
Sfrp2
Crabp1
Twist2
Malat1
Meg3
Ube2c
Top2a
Dcn
Lum
f
C1
C2
C3
C4
C5
C6
C7
g
Expression
logFC
Clusters
Hmx1
Ldha
Sparc
Ogn
Lum
Aspn
Ecrg4
Comp
Aqp3
Cma1
Asprv1
Dsg1b
Cyp4f39
Arhgap36
Myf6
Dsg1a
Tpsb2
Krtdap
Krt10
Tgm3
Krt1
Myl2
Clec3b
Abca12
Krt4
Ckm
Dmkn
Krt77
h
E14.5 (upper section of pinna)
Wild type
mEC1dup/dup
Hmx1 expansion
Epidermis marker
Hmx1
Dmkn
i
E14.5 (lower section of pinna)
Wild type
mEC1dup/dup
Hmx1
Dmkn

Figure 6

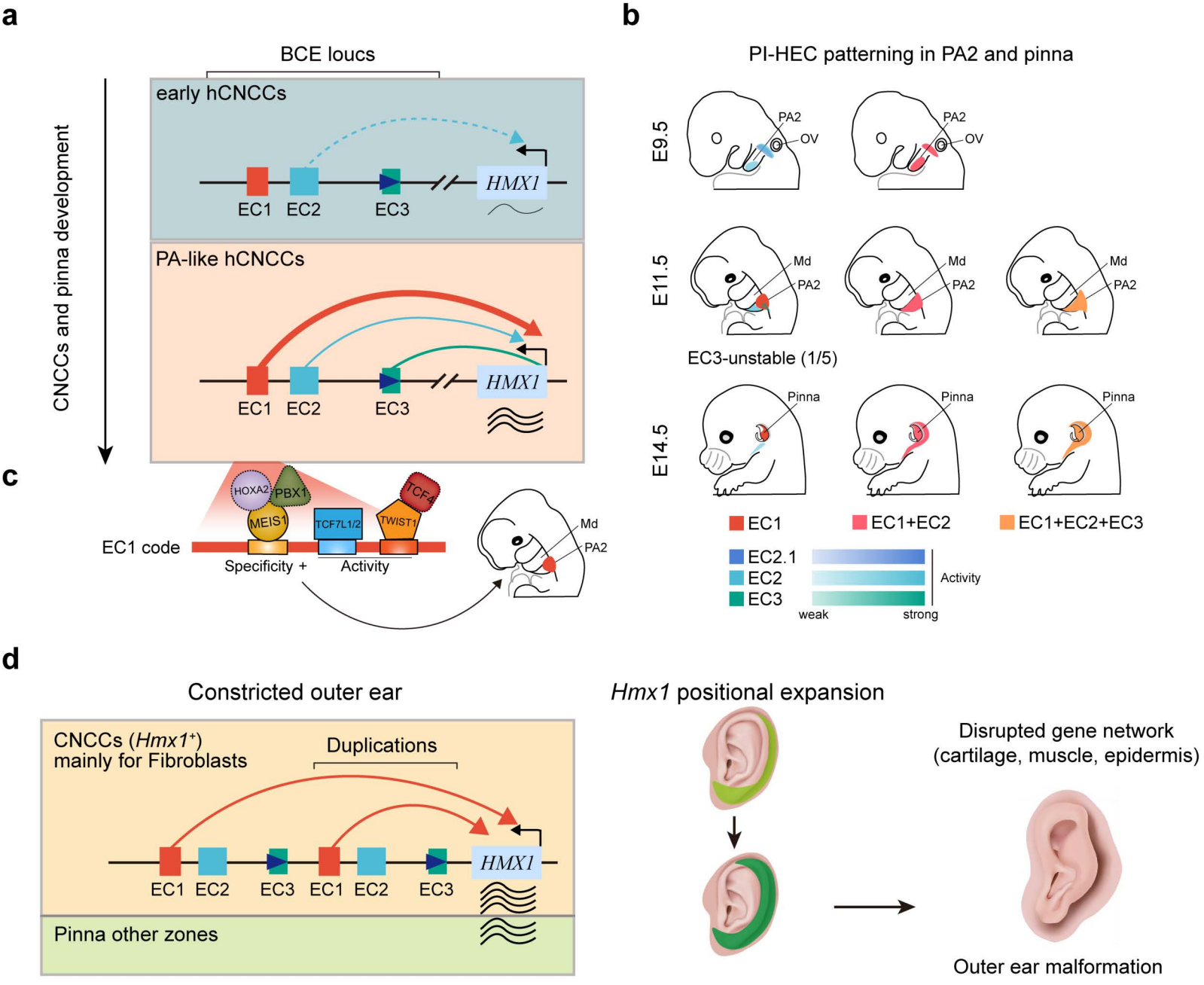
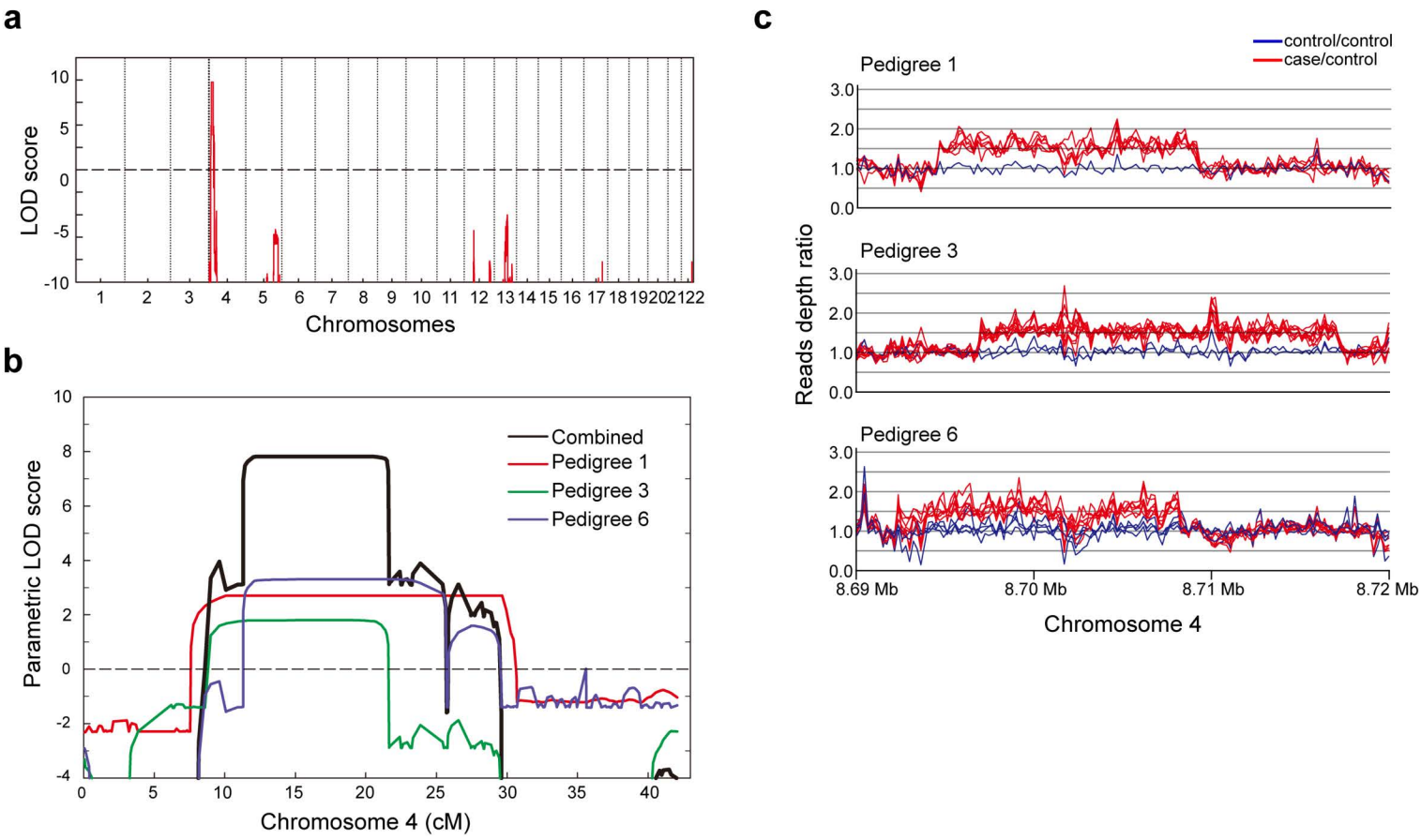
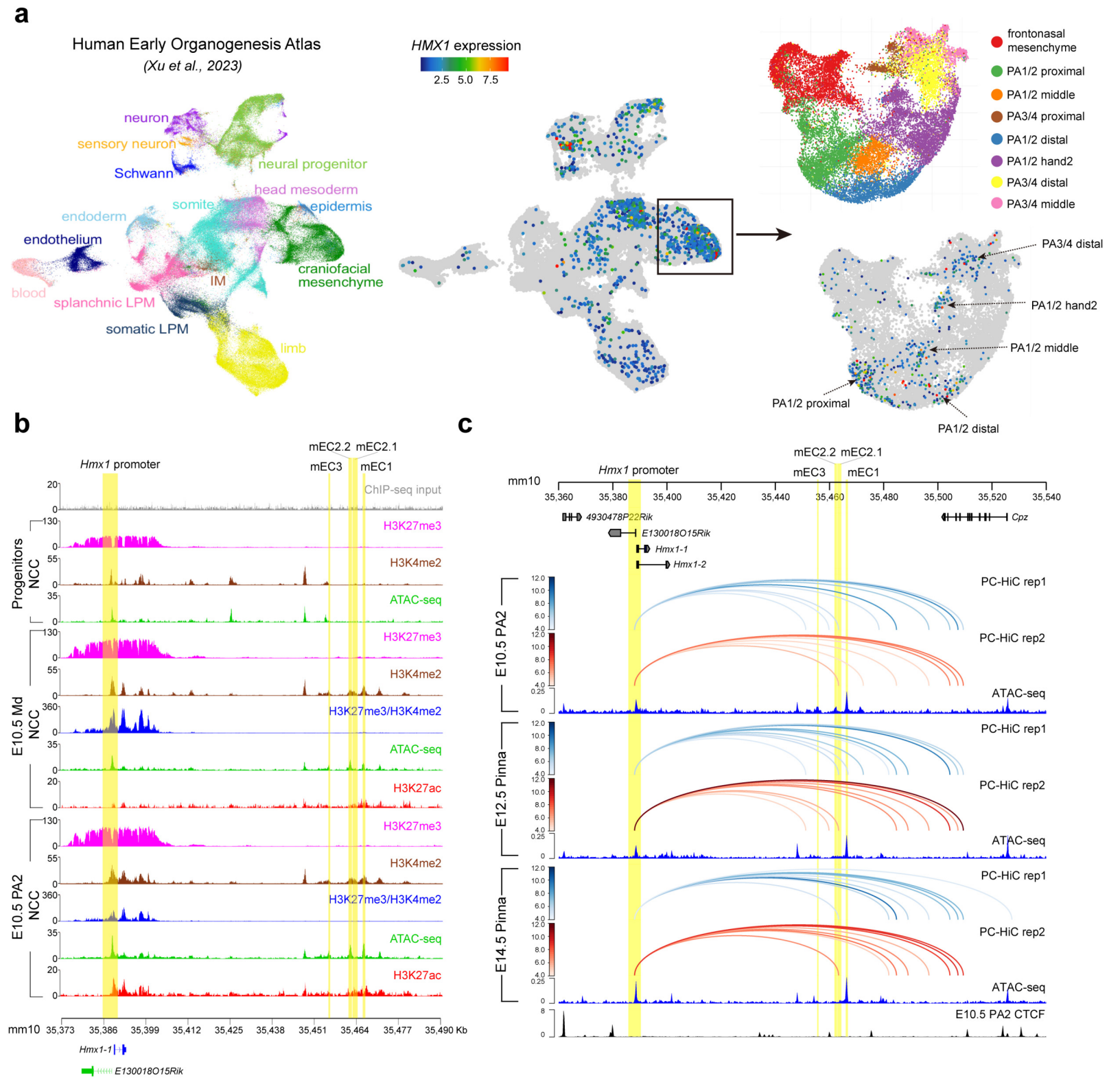

a
BCE loucs
CNCCs and pinna development
early hCNCCs
EC1 EC2 EC3 HMX1
PA-like hCNCCs
EC1 EC2 EC3 HMX1
b
PI-HEC patterning in PA2 and pinna
E9.5
PA2
OV
E11.5
Md
PA2
EC3-unstable (1/5)
E14.5
Pinna
EC1
EC1+EC2
EC1+EC2+EC3
EC2.1
EC2
EC3
Activity
weak
strong
c
HOXA2
PBX1
MEIS1
TCF7L1/2
TWIST1
TCF4
EC1 code
Specificity +
Activity
Md
PA2
d
Constricted outer ear
CNCCs (Hmx1+) mainly for Fibroblasts
Duplications
EC1 EC2 EC3 EC1 EC2 EC3 HMX1
Pinna other zones
Hmx1 positional expansion
Disrupted gene network (cartilage, muscle, epidermis)
Outer ear malformation

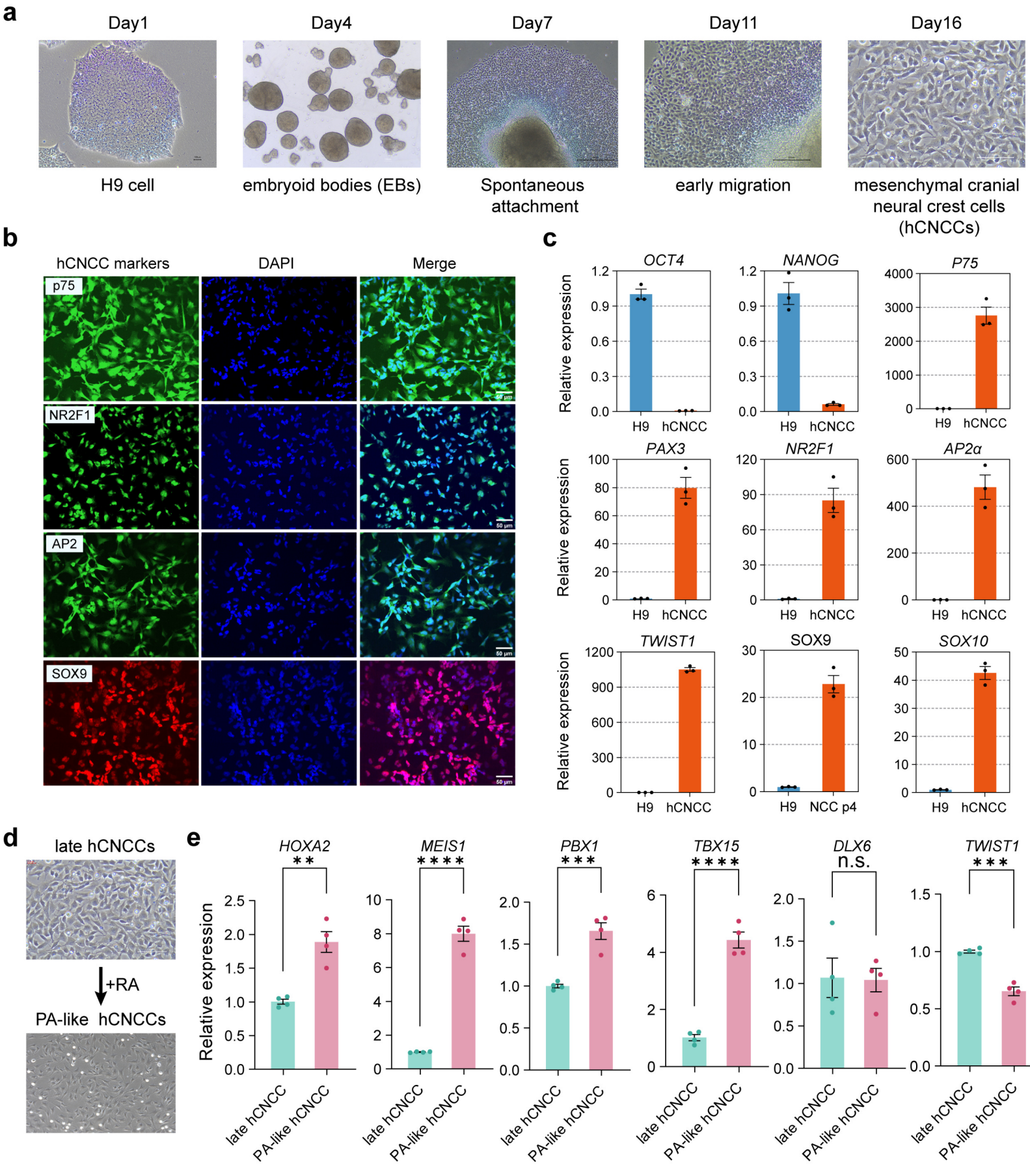
Figure 7



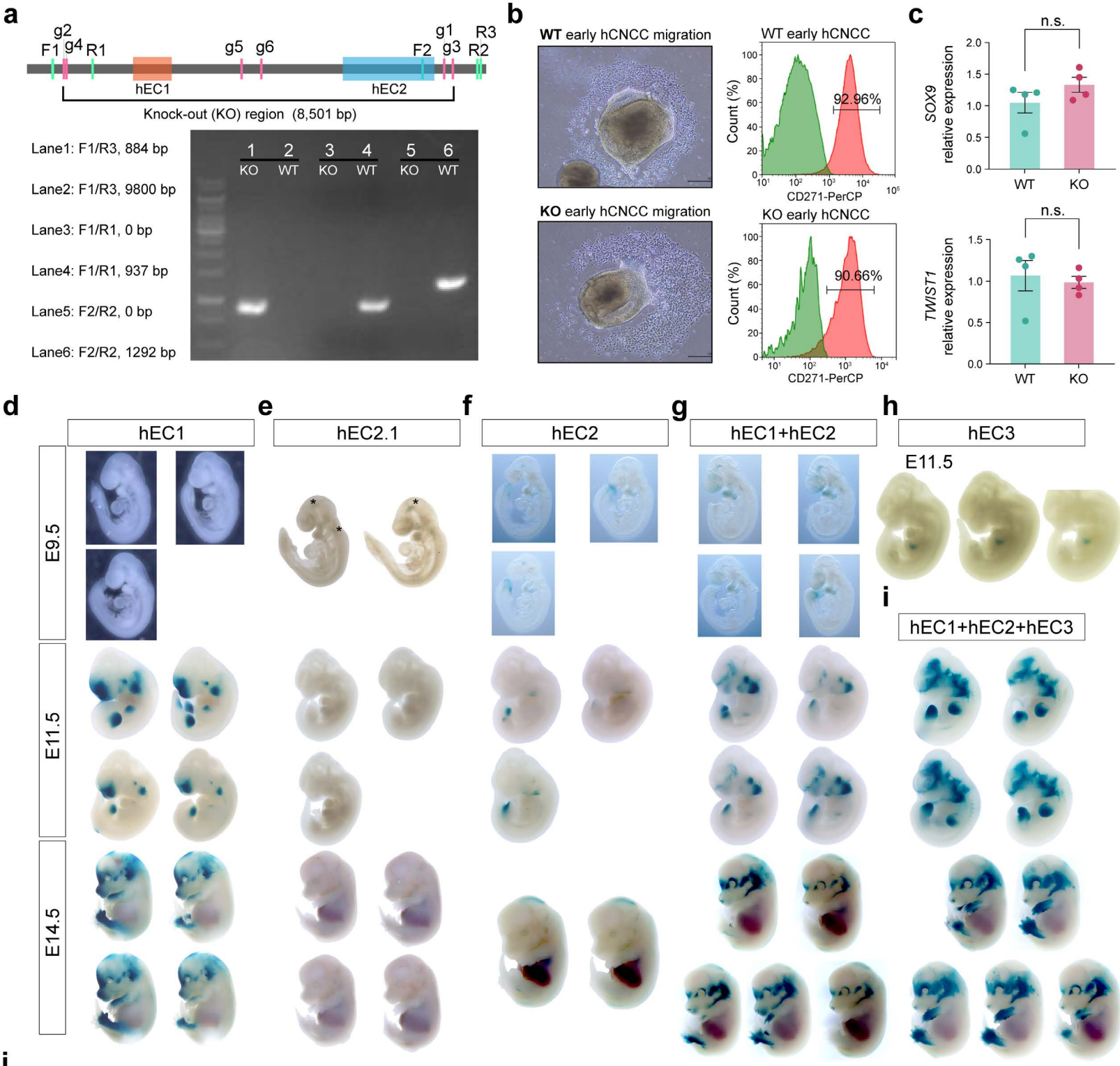
Extended Data Figure 1



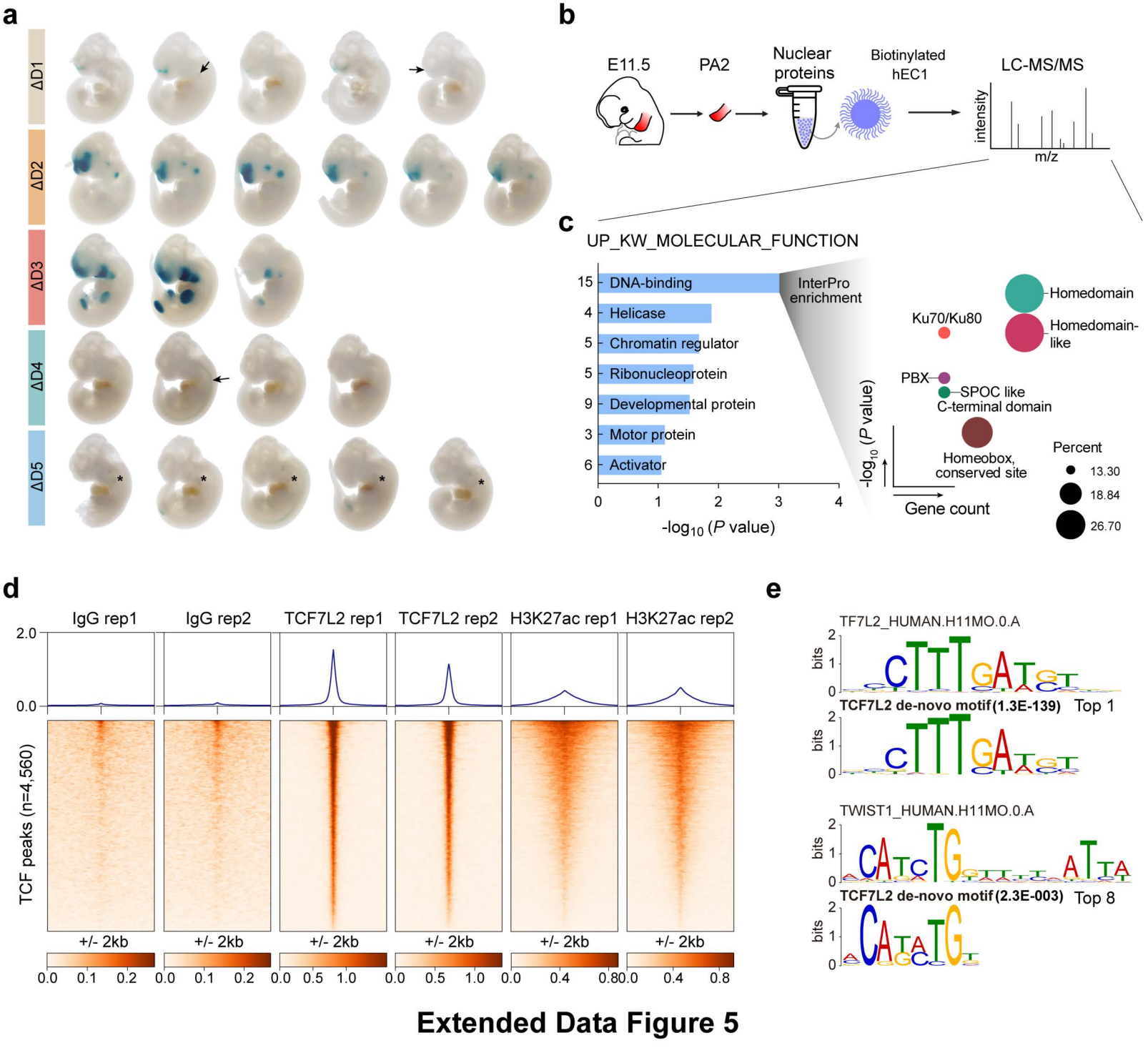
Extended Data Figure 2



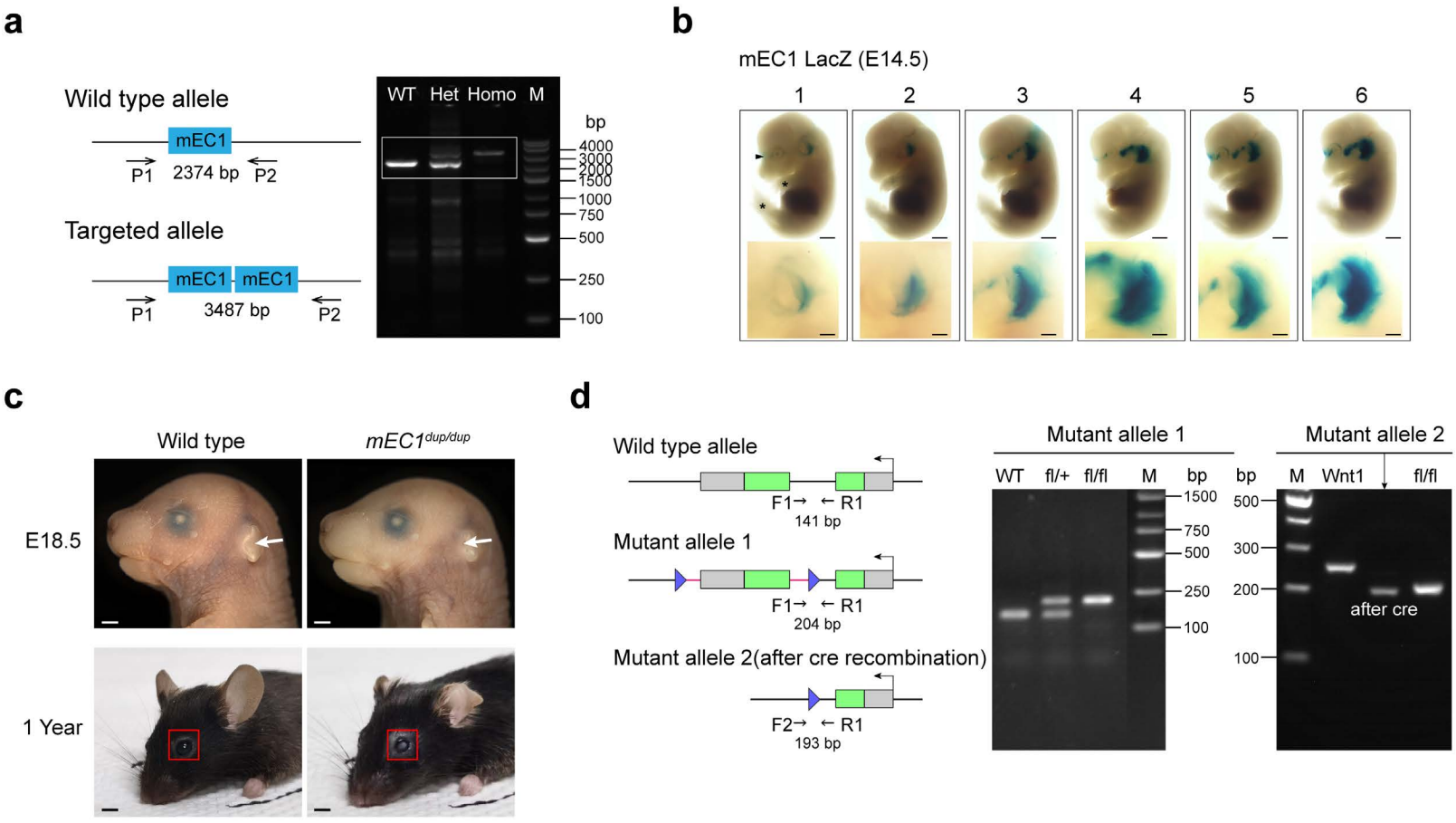
Extended Data Figure 3



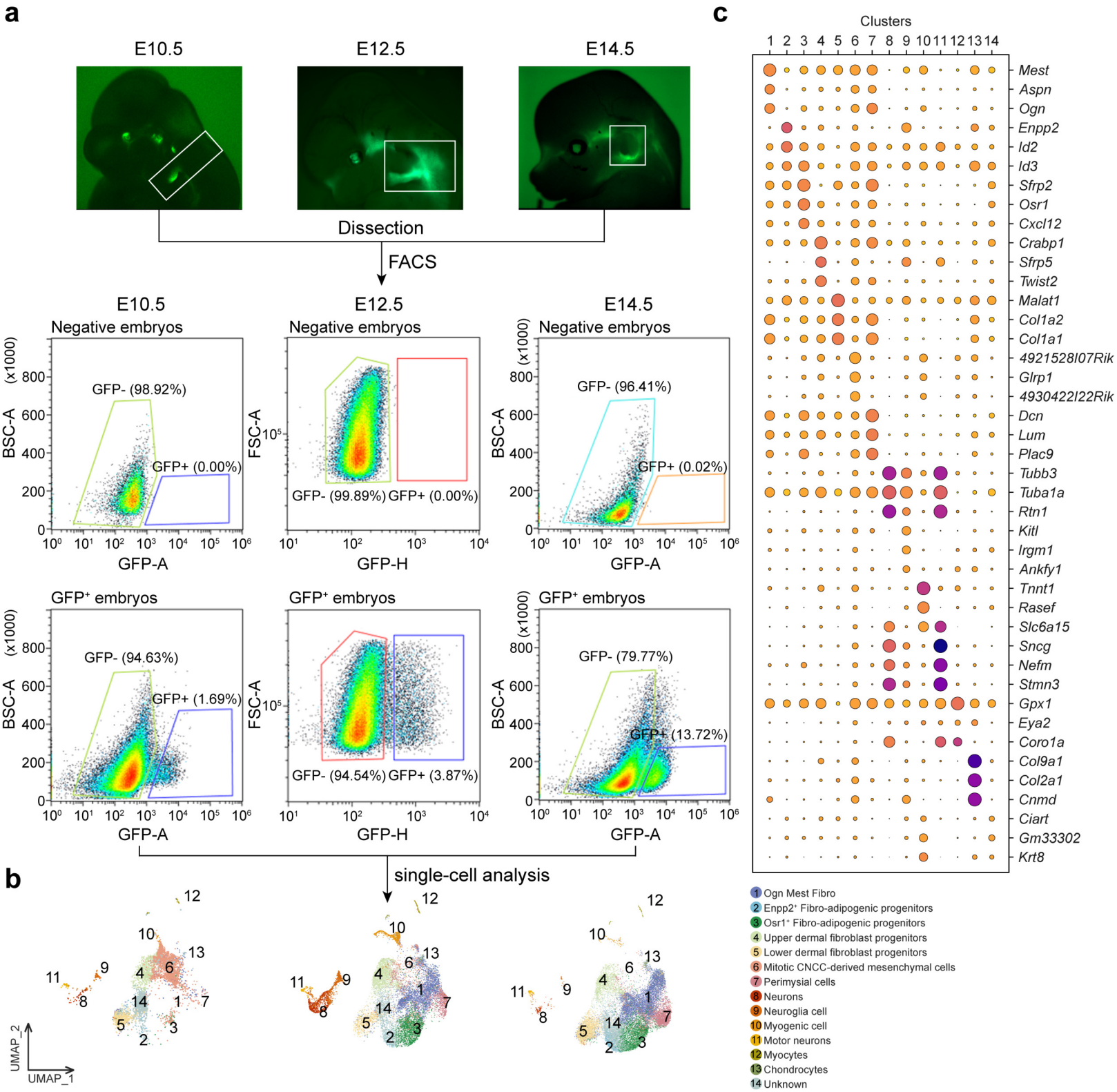
Extended Data Figure 4



Extended Data Figure 5



Extended Data Figure 6



Extended Data Figure 7