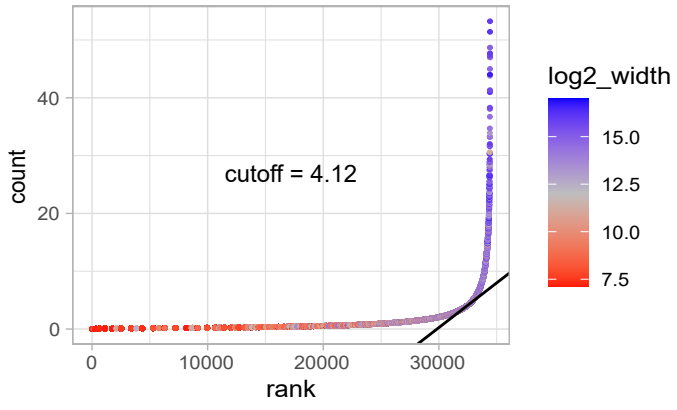


A multiple super-enhancer region establishes inter-TAD interactions and controls *Hoxa* function in cranial neural crest

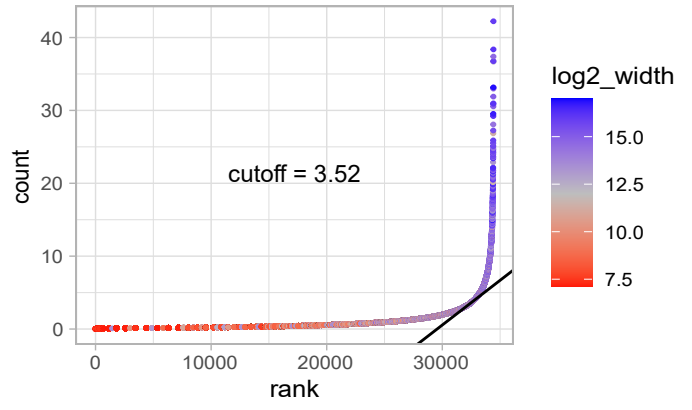
Kessler et al.

Supplementary Information

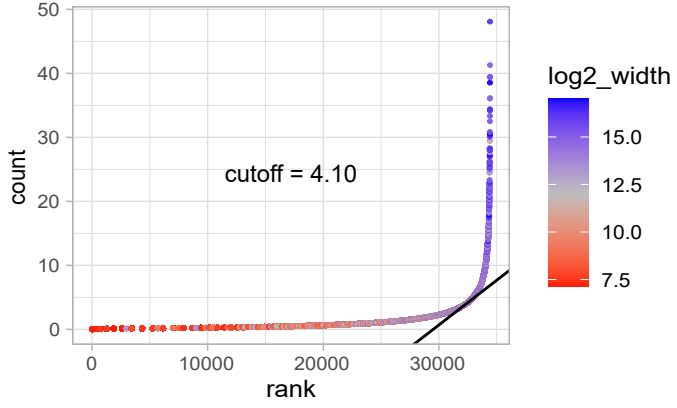
FNP



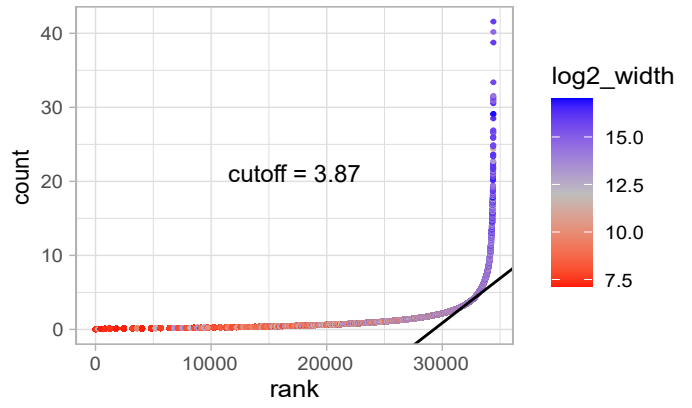
Mx



Md

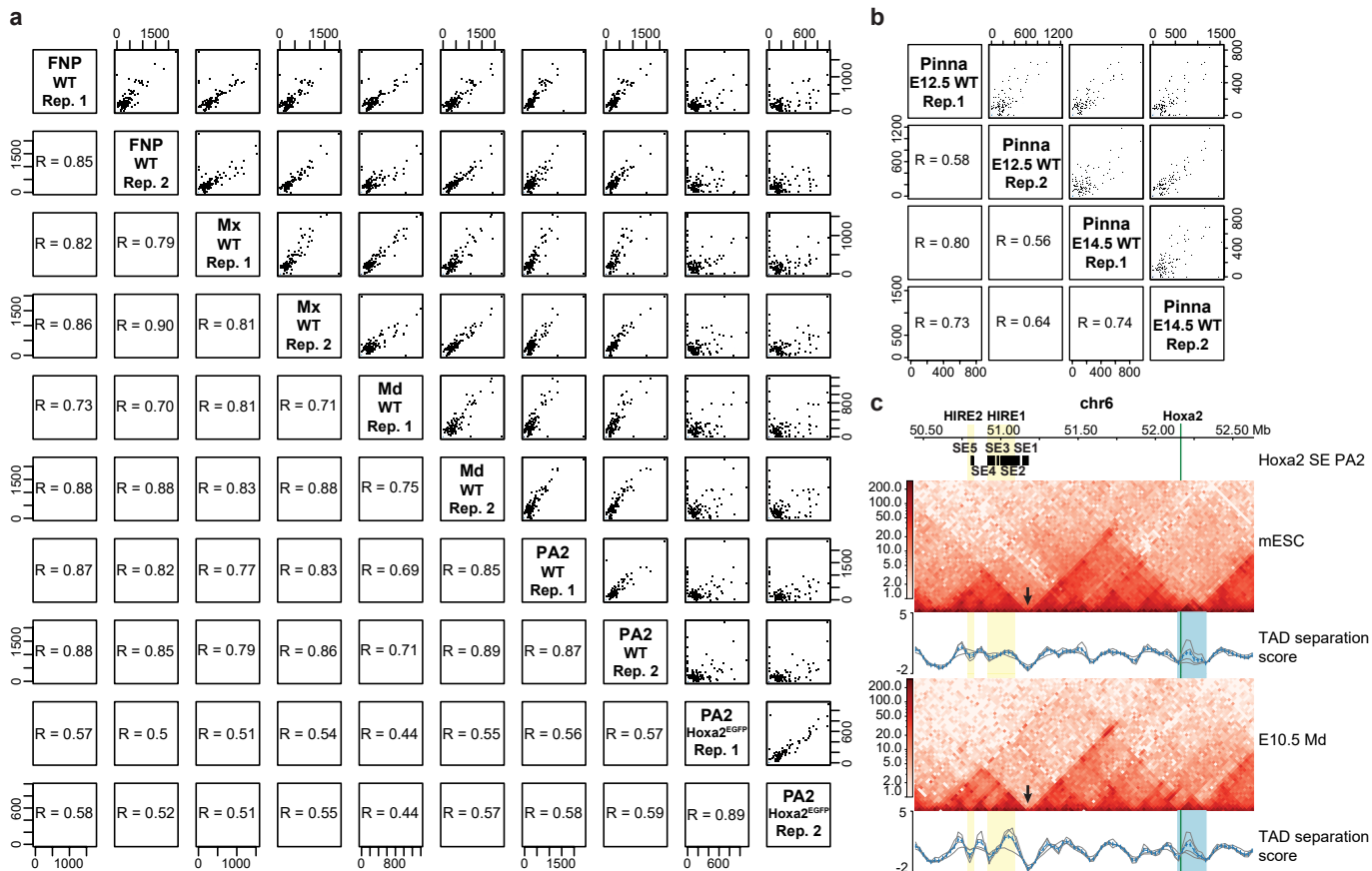


PA2



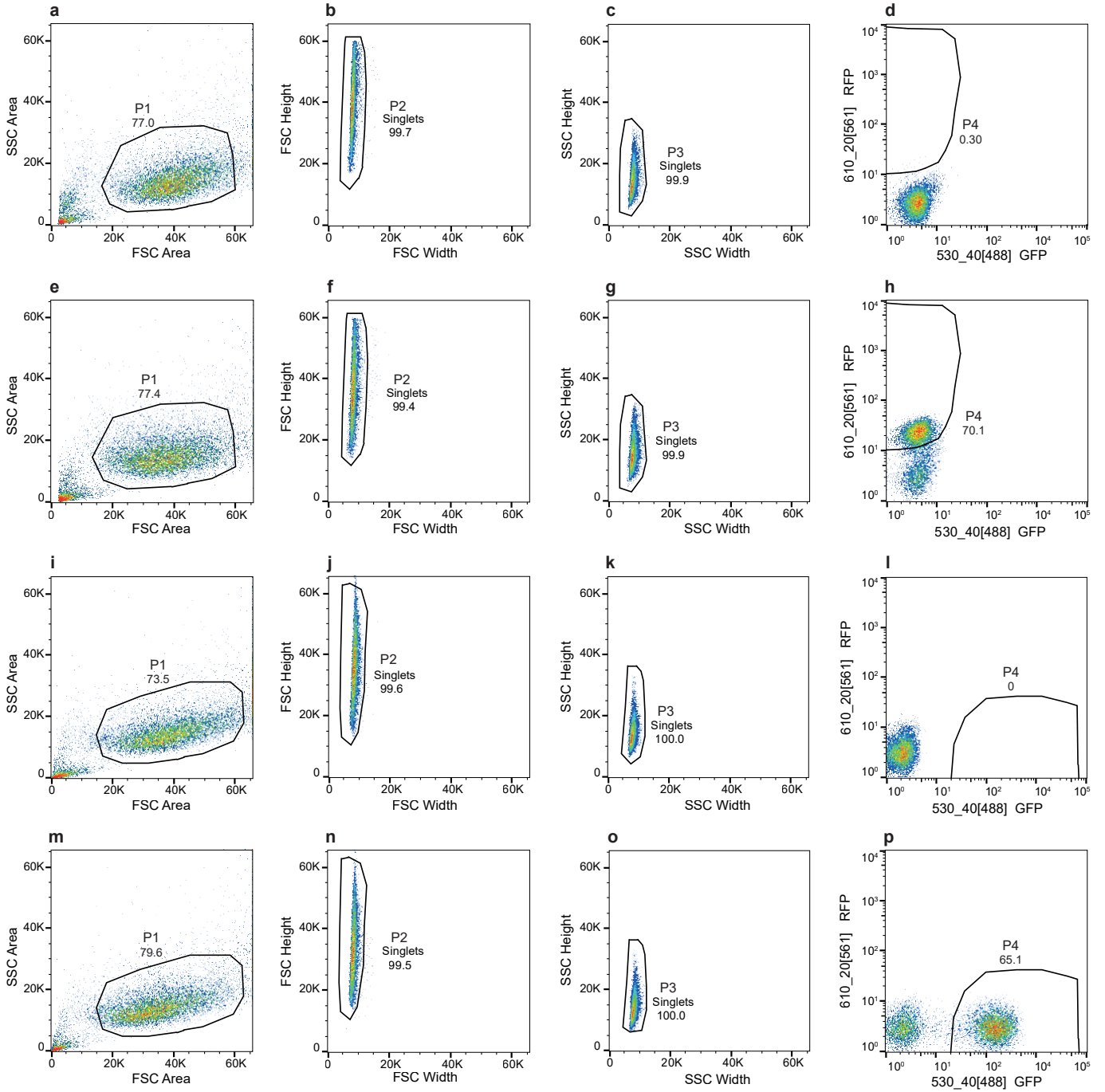
Supplementary Fig. 1 Identification of craniofacial super-enhancers.

Hockey stick plots displaying the H3K27 acetylation levels within merged peaks as counts per million on the y-axis and the according rank on the x-axis, separately for each cranial neural crest cell (CNCC) subpopulation. The cutoff value for the SE category was determined as described by the ROSE algorithm using the tangent (shown in black) and is indicated in the figure. The color reflects the log₂ transformed width of the merged regions.



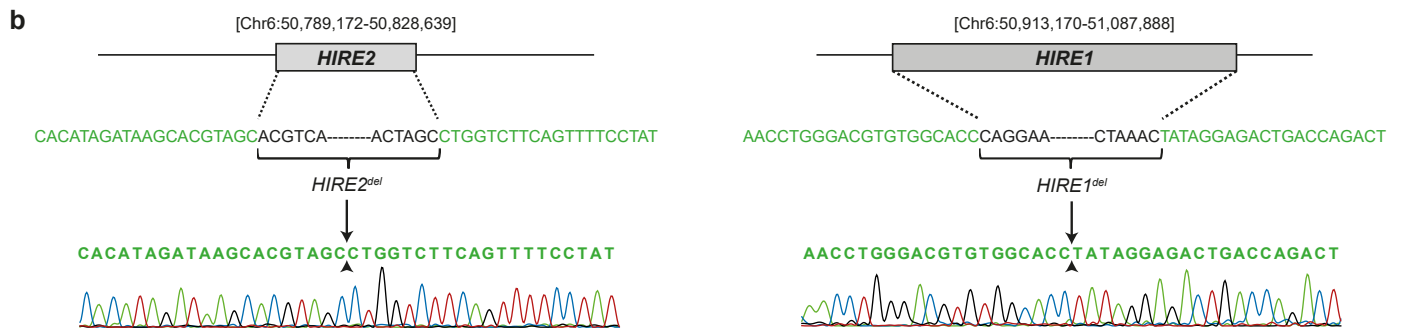
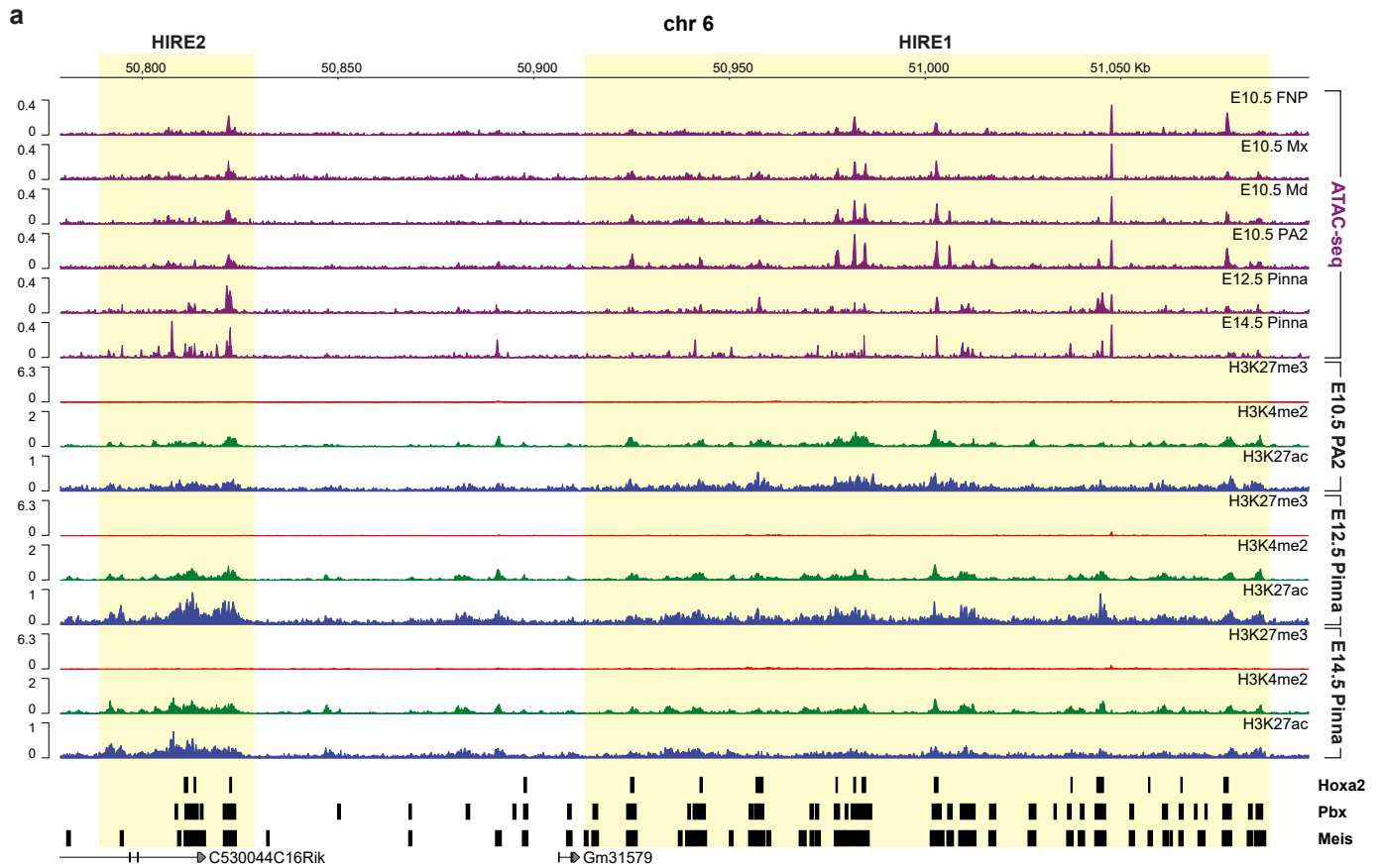
Supplementary Fig. 2 Promoter-capture Hi-C replicates, Hi-C in mESC and Md, and CTCF binding in PA2 and Md at E10.5.

a. Pairwise correlation scatter plots with Pearson's correlation coefficients between promoter capture Hi-C (PCHi-C) replicates from wild type (WT) frontonasal (FNP), maxillary (Mx), mandibular (Md), and second pharyngeal arch (PA2) derived cranial neural crest cells (CNCCs), as well as *Hoxa2*^{EGFP/EGFP} PA2 derived CNCCs at E10.5. **b.** Pairwise correlation scatter plots with Pearson's correlation coefficients between PCHi-C replicates of pinna derived CNCCs at E12.5 and E14.5. **c.** Hi-C interaction heatmaps at 25 kb resolution in a 2.2 Mb region of chromosome 6 including the *Hoxa* cluster (50,442,417-52,636,150 bp) in mouse embryonic stem cells (mESC) and in Md derived CNCCs at E10.5. TAD separation scores are called with HiCExplorer's hicFindTADs and are shown as blue lines. The additional gray lines show the TAD score for different window sizes. The blue highlight marks the polycomb domain encompassing the *Hoxa* cluster and *Evx1* (chr6:52,145,433-52,327,518). Arrows indicate the location of the TAD boundary between *Hoxa2* and HIRE1/HIRE2. **d.** CTCF ChIP-seq tracks in Md and PA2 derived CNCCs at E10.5. CTCF motif orientation at CTCF peaks is indicated with red (minus strand) and blue (plus strand) arrowheads. The pink box highlights the CTCF binding sites at the TAD boundary between *Hoxa2* and HIRE1 / HIRE2. A zoom in on the TAD boundary is shown at the bottom. In panels **c** and **d**, yellow boxes highlight HIRE1/HIRE2 and a green (**c**) or yellow (**d**) line highlights the *Hoxa2* locus.



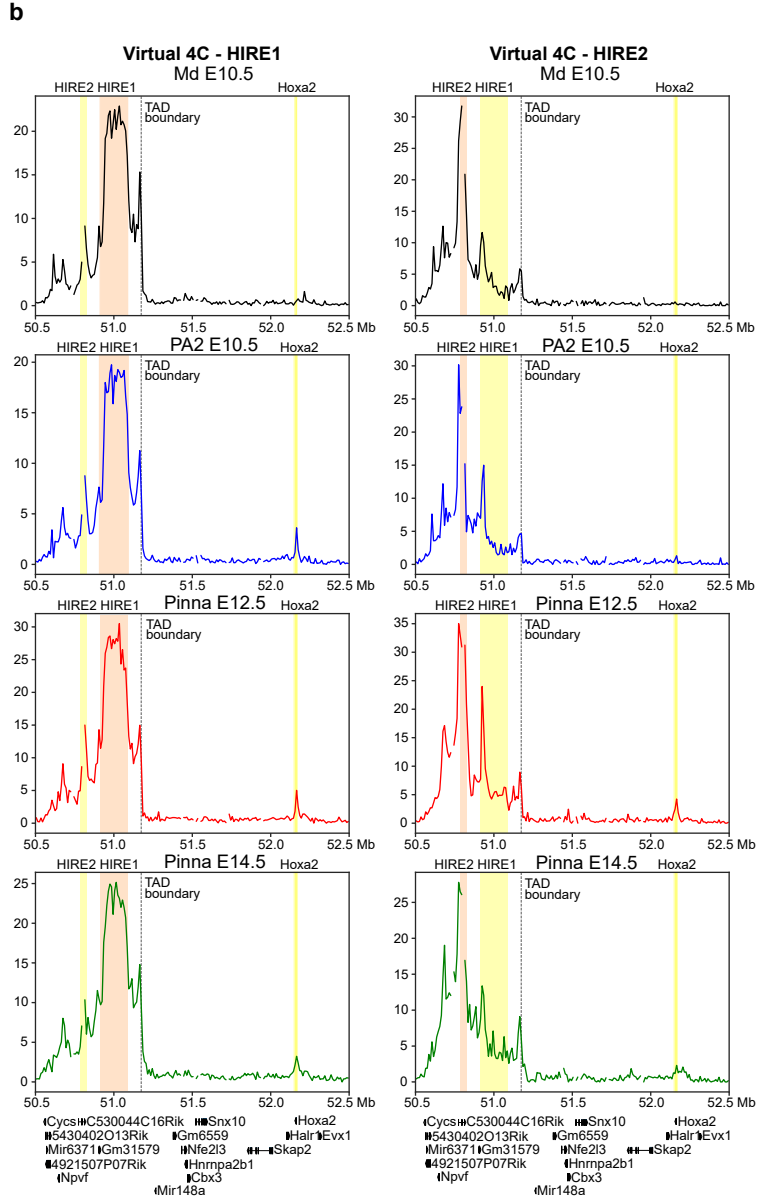
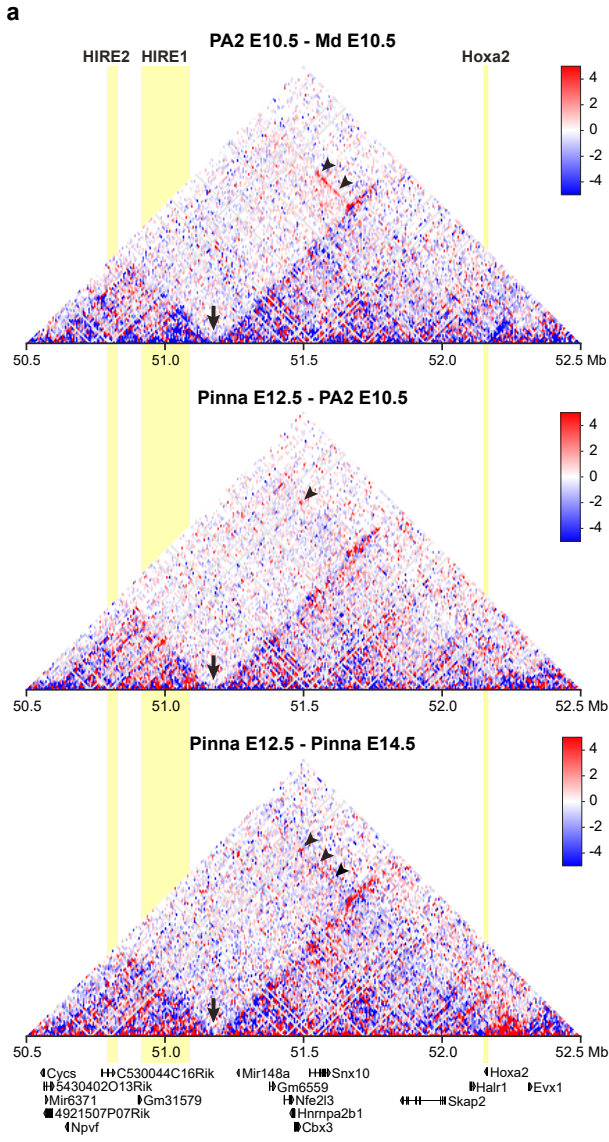
Supplementary Fig. 3 Fluorescence activated cell sorting (FACS) gating strategy.

a-p. FACS of Red Fluorescent Protein (RFP)-positive cranial neural crest cells (CNCCs) from *Wnt1::Cre;ROSA^{RFP}* embryos (**a-h**) and FACS of Green Fluorescent Protein (GFP)-positive cells from the second arch of E10.5 *Hoxa2^{EGFP/EGFP}* embryos (**i-p**). Living cells are first discriminated from debris by scatter areas P1 (FSC Area versus SSC Area) (**a, e, i, m**). Single cells are discriminated from doublets by firstly assessing forward scatter (FSC Width versus FSC Height; P2) (**b, f, j, n**) and next by assessing side scatter (SSC Width versus SSC Height; P3) (**c, g, k, o**). Subsequently, cells were sorted based on fluorescent markers of interest, GFP (**h**) or RFP (**p**). The baseline control fluorescent signal is from a wild-type (**d, l**) littermate (negative control).



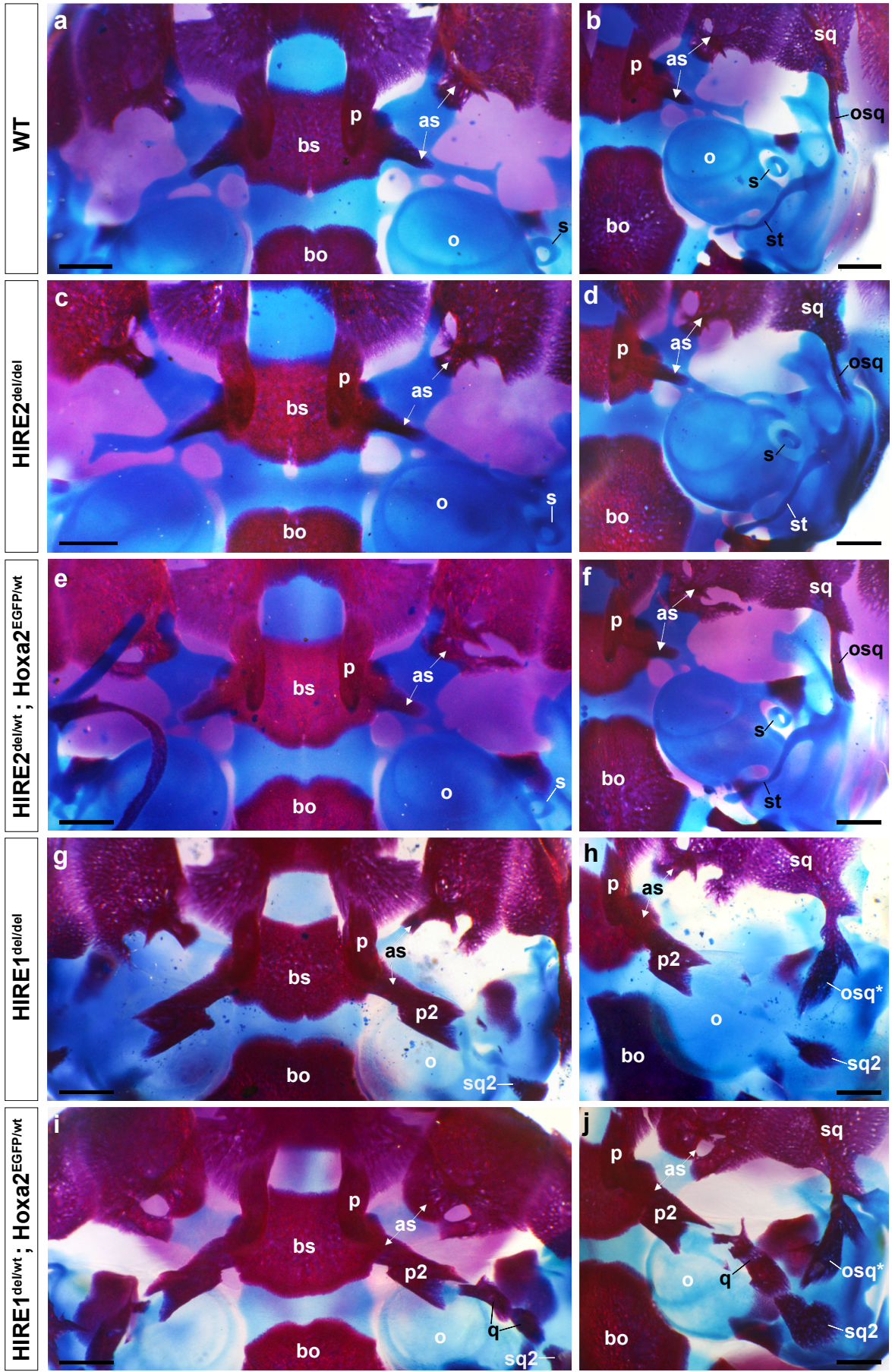
Supplementary Fig. 4 Chromatin landscape of HIRE1 and HIRE2.

a. Genome browser view on HIRE1 and HIRE2 (chr6:50,779,104 – 51,097,965) with chromatin accessibility (ATAC-seq, purple) from the frontonasal process (FNP), the maxillary (Mx), the mandibular (Md) and the second pharyngeal arch (PA2) cranial neural crest cells (CNCCs) at E10.5, and from the pinna CNCCs at E12.5 and E14.5. ChIP-seq profiles for H3K27me3 (red), H3K4me2 (green), and H3K27ac (blue) from PA2 CNCCs at E10.5, and from the pinna CNCCs at E12.5 and E14.5 are also shown. At the bottom are binding sites identified by ChIP-seq, for Hoxa2¹, Pbx, and Meis² in PA2 at E11.5. Highlighted in yellow are HIRE1 and HIRE2. **b.** Illustration representing HIRE1 [Chr6:50,913,170-51,087,888] and HIRE2 [Chr6:50,789,172-50,828,639] deletions. HIRE1 and HIRE2 regions were deleted in mutant mice using the CRISPR-Cas9 system, with regions preserved after deletion represented in green, and regions deleted represented in black. *HIRE1^{del}* and *HIRE2^{del}* mutations were validated using sanger sequencing.



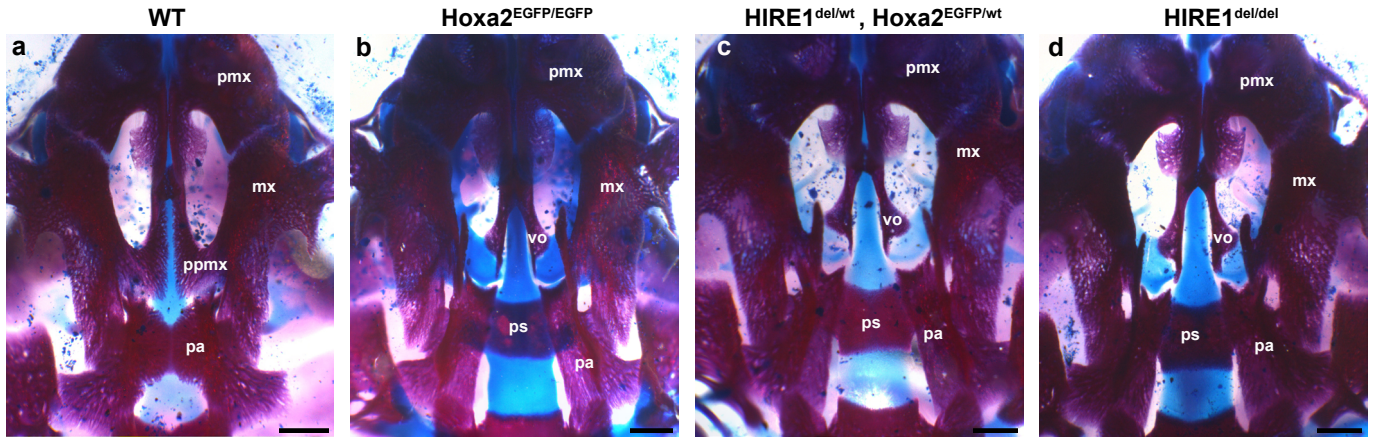
Supplementary Fig. 5 Interactions at HIRE1 and HIRE2

a. Differential HiC maps at 10 kb resolution are shown for a ~2Mb region of chromosome 6 including the *Hoxa* cluster, comparing interactions between E10.5 second pharyngeal arch (PA2) and E10.5 mandibular (Md) cranial neural crest cells (CNCCs) (top), between E12.5 pinna CNCCs and E10.5 PA2 CNCCs (middle), and between E12.5 pinna CNCCs and E14.5 pinna CNCCs (bottom). Shown below are the positions of genes. For better readability, only the genomic position of the *Hoxa2* gene is shown within the *Hoxa* cluster. The yellow rectangles show the HIRE1/HIRE2 regions and *Hoxa2* gene. The arrowheads on the differential map highlight the “stripe” or “spots” of increased interactions between *Hoxa2* and HIRE1/HIRE2 (red) in E10.5 PA2 as compared to Md (top) in E12.5 pinna as compared to E10.5 PA2 (middle) and in E12.5 pinna as compared to E14.5 pinna (bottom). A black arrow between two TADs indicates the presence of the TAD border. **b.** Virtual 4C plots from HIRE1 (left side) and HIRE2 (right side) for E10.5 Md, E10.5 PA2, E12.5 pinna and E14.5 pinna CNCCs, derived from Hi-C data. The orange rectangles highlight the viewpoints whereas the yellow rectangles highlight the *Hoxa2* gene, HIRE2 (left side) and HIRE1 (right side).



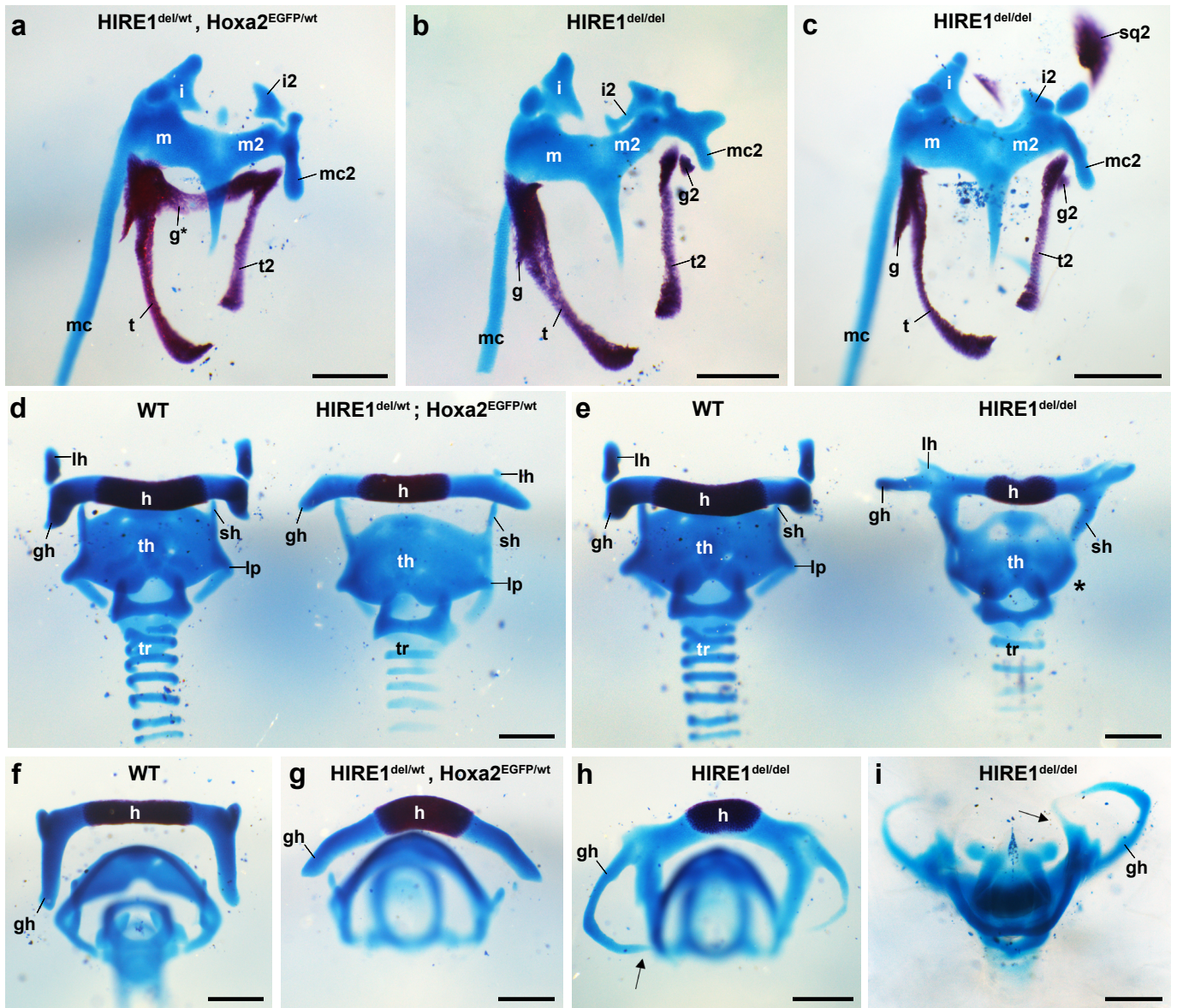
Supplementary Fig. 6 Cranial skeletal changes in HIRE1 and HIRE2 mutant fetuses.

Ventral (**a, c, e, g, i**) and ventrolateral (**b, d, f, h, j**) views of the cranial base of E18.5 wild-type (WT) (**a, b**), *HIRE2^{del/del}* homozygous mutant (**c, d**), *HIRE2^{del/wt};Hoxa2^{EGFP/wt}* trans-heterozygote mutant (**e, f**), *HIRE1^{del/del}* homozygous mutant (**g, h**) and *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* trans-heterozygous mutant (**i, j**) fetuses. **a** to **j** are representative images of n=4/4 WT (**a, b**), n=6/6 *HIRE2^{del/del}* (**c, d**), n=4/4 *HIRE2^{del/wt};Hoxa2^{EGFP/wt}* (**e, f**), n=4/4 *HIRE1^{del/del}* (**g, h**) and n=4/4 *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* (**i, j**) fetuses. **a** represents the alisphenoid bone; **b** represents the basioccipital bone; **bs** represents the basisphenoid bone; **o** represents the cartilaginous otic capsule; **p** represents the pterygoid bone; **p2** represents the ectopic pterygoid bone; **osq** and **osq*** represent respectively the normal and modified retrotympenic (otic) process of the squamosal bone; **s** represents the stapes; **sq** represents the squamosal bone; **sq2** represents the ectopic squamosal bone; **st** represents the styloid process; **q** represents the atavistic quadrate bone, ectopically appearing in *Hoxa2^{-/-}* mutants³ and in the *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* trans-heterozygous mutants. Scale bars represent 500 μ m.



Supplementary Fig. 7 Cleft secondary palate in HIRE1 mutant fetuses

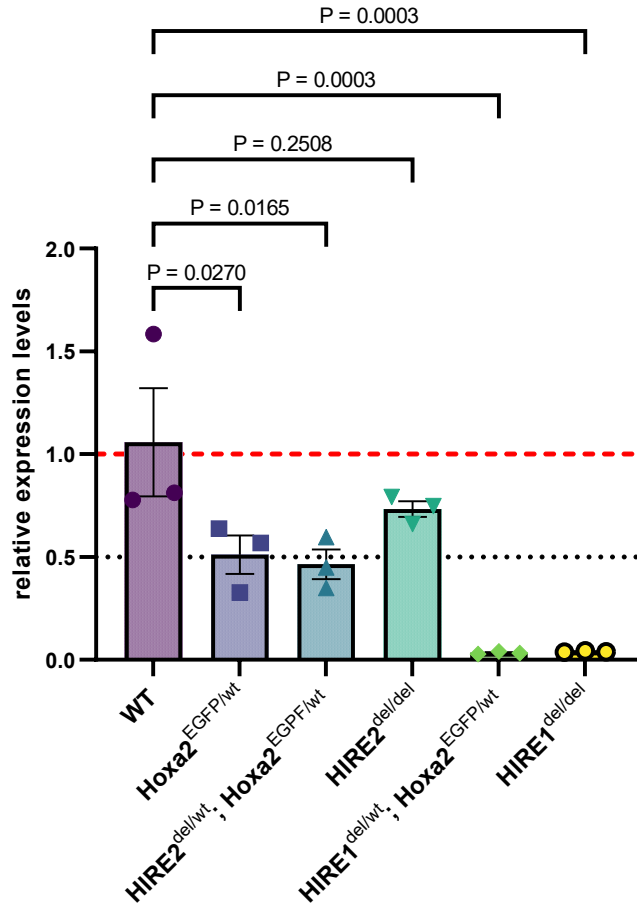
Skeletal preparations of the palate area after the removal of the lower jaw from E18.5 wild-type (WT) (a), *Hoxa2*^{EGFP/EGFP} (b), *HIRE1*^{del/wt};*Hoxa2*^{EGFP/wt} (c) and *HIRE1*^{del/del} (d) fetuses. a to d are representative images of WT (n=4/4) (a), *Hoxa2*^{EGFP/EGFP} (n=6/8) (b), *HIRE1*^{del/wt};*Hoxa2*^{EGFP/wt} (n=14/24) (c) and *HIRE1*^{del/del} (n=8/10) (d) fetuses. The nose is toward the top. mx represents the maxilla, pa the palatal bone, pmx the premaxilla, ppmx the palatal process of maxilla, ps the presphenoid bone, and vo the vomer. Note that vo and ps in (b, c, d) are not visible in (a) because they are underneath Mx and Pa. Scale bars represent 500 μ m.



Supplementary Fig. 8 Middle ear and hyoid bone skeletal phenotypes upon deletion of HIRE1

Middle ear (**a-c**) and hyoid (**d-i**) skeletal preparations from E18.5 wild-type (WT) (**d, f**), *HIRE1^{del/del}* homozygous mutant (**b, c, e, h, i**) and *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* trans-heterozygous mutant (**a, d, g**) fetuses. In **a** and **b**, the absence of fusion between the incus and its duplicated counterpart is representative of n=1/8 fetus sides for *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* (**a**) and n=1/8 fetus sides for *HIRE1^{del/del}* (**b**). g* represents the transformed gonial bone, as described in *Hoxa2*^{-/-} mutant fetuses³. The transformed gonial bone presented in **a** is representative of n=2/8 *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* fetus sides. **d, i** and **j** are representative of WT (n=4/4), *HIRE1^{del/wt};Hoxa2^{EGFP/wt}* (n=4/4) and *HIRE1^{del/del}* (n=4/4) fetuses. In **e**, the asterisk displays the absence of the lateral process (lp) of the laryngeal cartilage in *HIRE1^{del/del}* homozygous mutant fetuses. The black arrows in **h** and **i** display the fusion between the elongated greater horn (gh) of the hyoid bone and the posterior border of the thyroid cartilage (th) observed in *HIRE1^{del/del}* homozygous mutant fetuses. g and g2 represent the WT and duplicated gonial bones, respectively; h represents the body of the hyoid bone; i and i2 represent the WT and duplicated incus respectively; lh represents the lesser horn of the hyoid bone; m and m2 represent the WT and duplicated mallei, respectively; mc and mc2 represent the WT and partially duplicated Meckel's cartilages respectively; sh represents the superior horn of the thyroid cartilage; sq2 represents the ectopic squamosal bone; t and t2 represent the WT and duplicated tympanic bones respectively; tr represents the trachea. Scale bars represent 500 μm.

E10.5 PA2



Supplementary Fig. 9 *Hoxa2* down-regulation in E10.5 second pharyngeal arch upon HIRE1 or HIRE2 deletion.

Relative expression levels of *Hoxa2*, detected by qRT-PCR analysis, in E10.5 second pharyngeal arch (PA2) of WT and *Hoxa2*^{EGFP/wt}, *HIRE2*^{del/wt};*Hoxa2*^{EGFP/wt}, *HIRE2*^{del/del}, *HIRE1*^{del/wt};*Hoxa2*^{EGFP/wt}, and *HIRE1*^{del/del} mutant embryos (n = 3 per genotype). Data are represented as means ± SEM. Significance is based on a one-way ANOVA followed by Dunnett's multiple comparisons test with WT as the control sample.

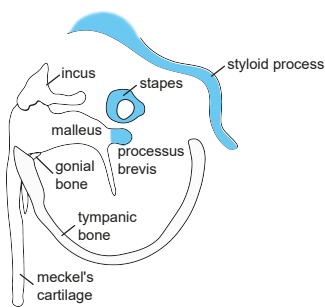
PA2 CNCC-derived skeletal and cartilaginous structures

Control



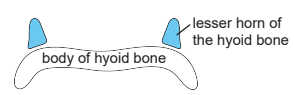
Hoxa2 transcript levels in PA2 at E10.5:
WT : 100 %
Hoxa2^{EGFP/wt} : about 50 %
(See supplementary figure 9)

Middle ear



(See figure 6a, 6b)

Hyoid



(See figure 6c)

Pinna

(See figure 5b, 5e, 5f)

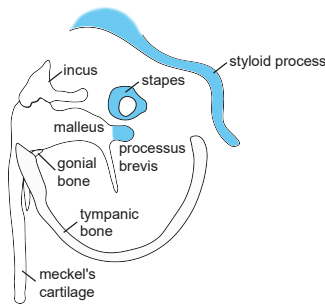
■ PA2 derived

HIRE2^{del/del}



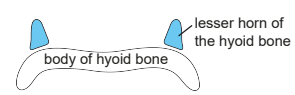
Hoxa2 transcript levels in PA2 at E10.5:
about 75% compared to WT
(See supplementary figure 9)

Middle ear



No visible malformations
(See figure 6d, 6e)

Hyoid



No visible malformation
(See figure 6f)

Pinna

No visible malformations at birth or in adult animals
(See figure 5c)

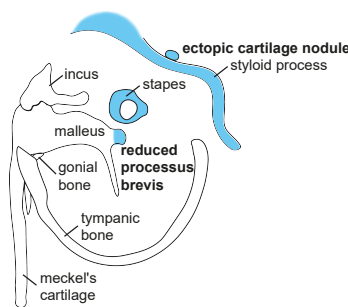
■ PA2 derived

HIRE2^{del/wt};Hoxa2^{EGFP/wt}



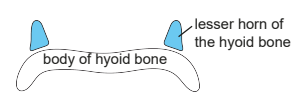
Hoxa2 transcript levels in PA2 at E10.5:
about 46% compared to WT
(See supplementary figure 9)

Middle ear



(See figure 6g, 6h)

Hyoid



No visible malformations
(See figure 6i)

Pinna

No visible malformations at birth
Microtia in adult animals
(See figure 5d, 5e, 5g)

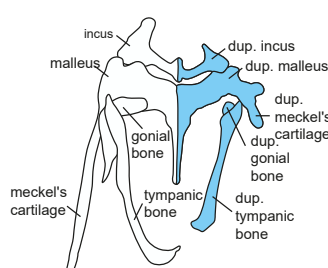
■ PA2 derived

HIRE1^{del/del}



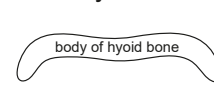
Hoxa2 transcript levels in PA2 at E10.5:
about 0-4% compared to WT
(See supplementary figure 9)

Middle ear



Homeotic transformation of PA2 skeletal elements to PA1-like morphology
(See figure 6j, 6k)

Hyoid



Absence of the lesser horns
(See figure 6l)

Pinna

Absent
(See figure 5h)

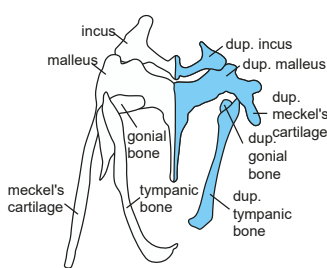
■ PA2 derived

HIRE1^{del/wt};Hoxa2^{EGFP/wt}



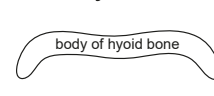
Hoxa2 transcript levels in PA2 at E10.5:
about 0-3% compared to WT
(See supplementary figure 9)

Middle ear



Homeotic transformation of PA2 skeletal elements to PA1-like morphology
(See figure 6m, 6n)

Hyoid



Absence of the lesser horns
(See figure 6o)

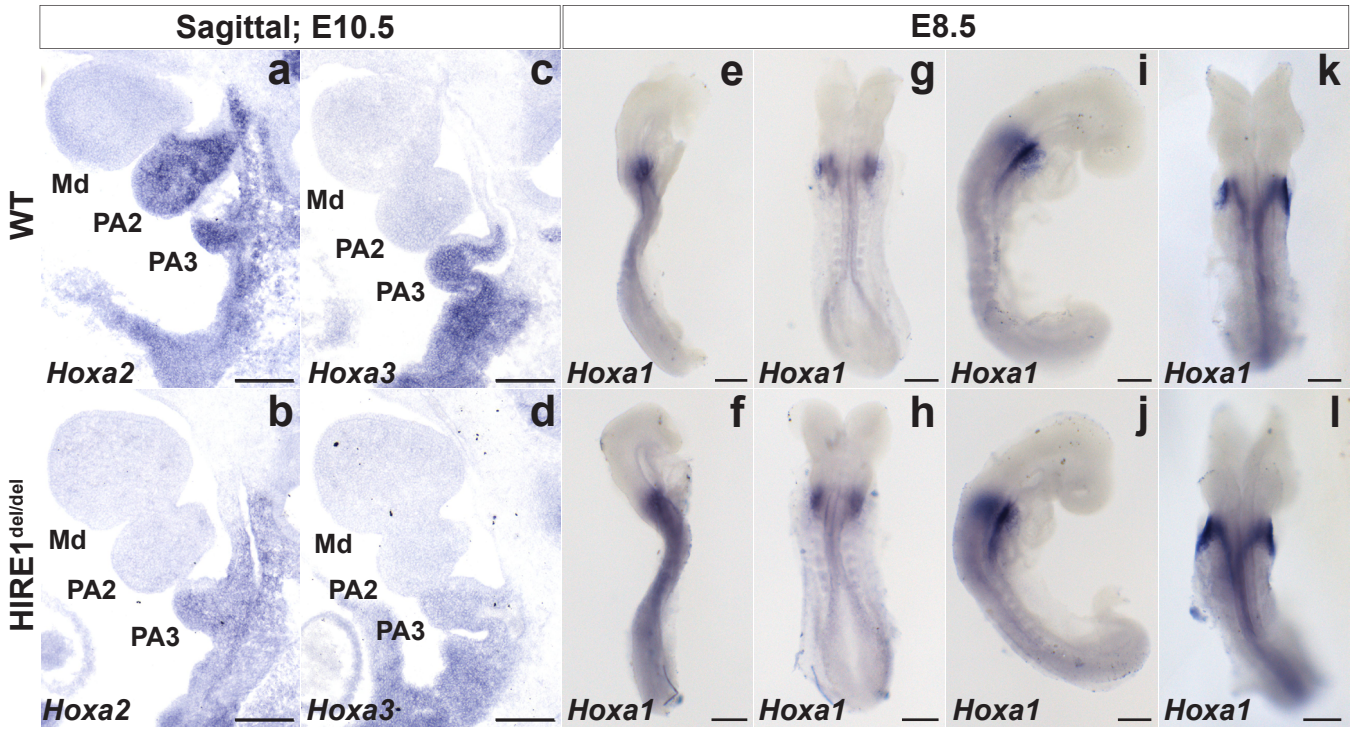
Pinna

Absent
(See figure 5i)

■ PA2 derived

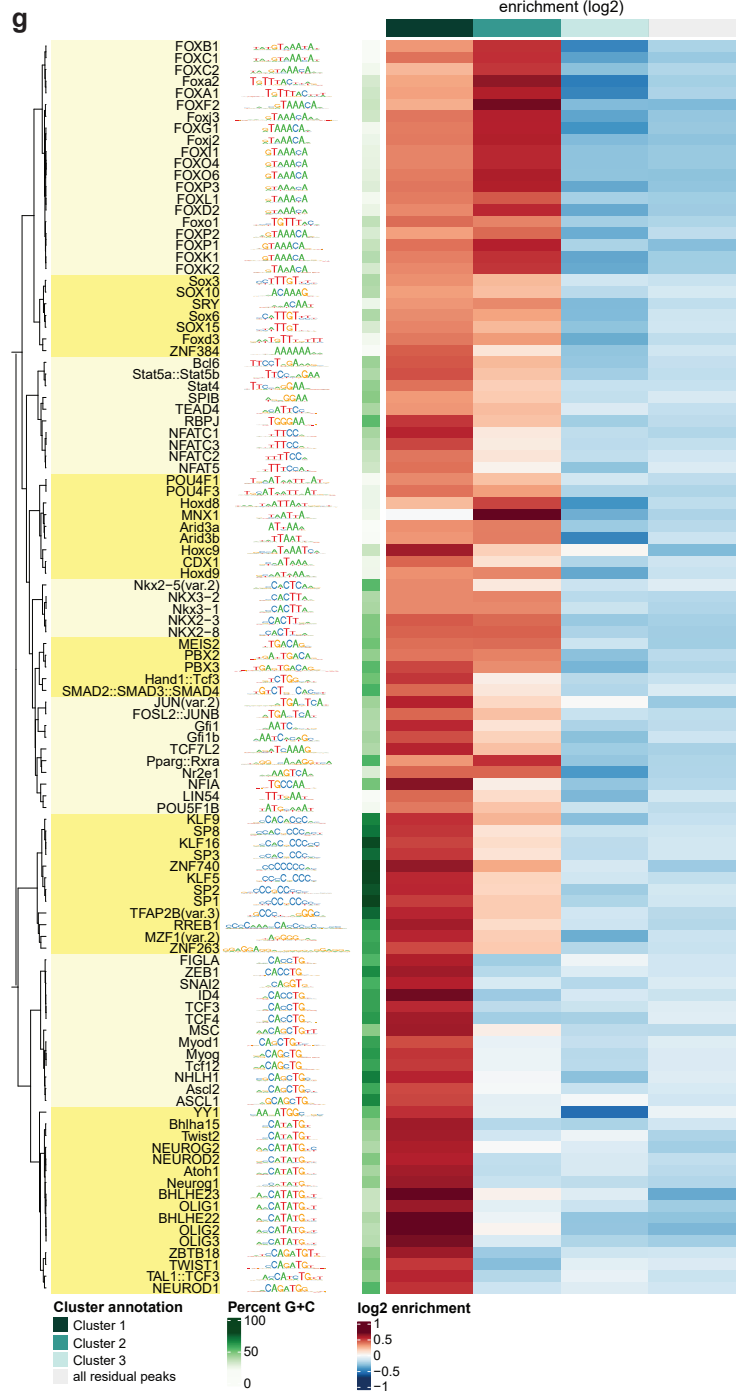
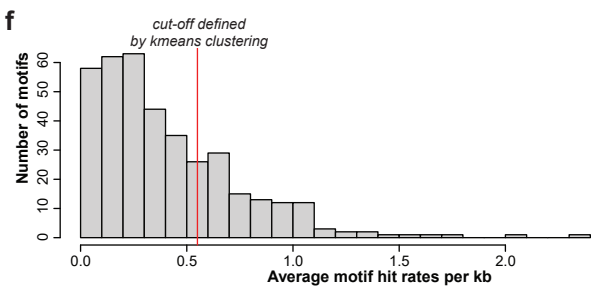
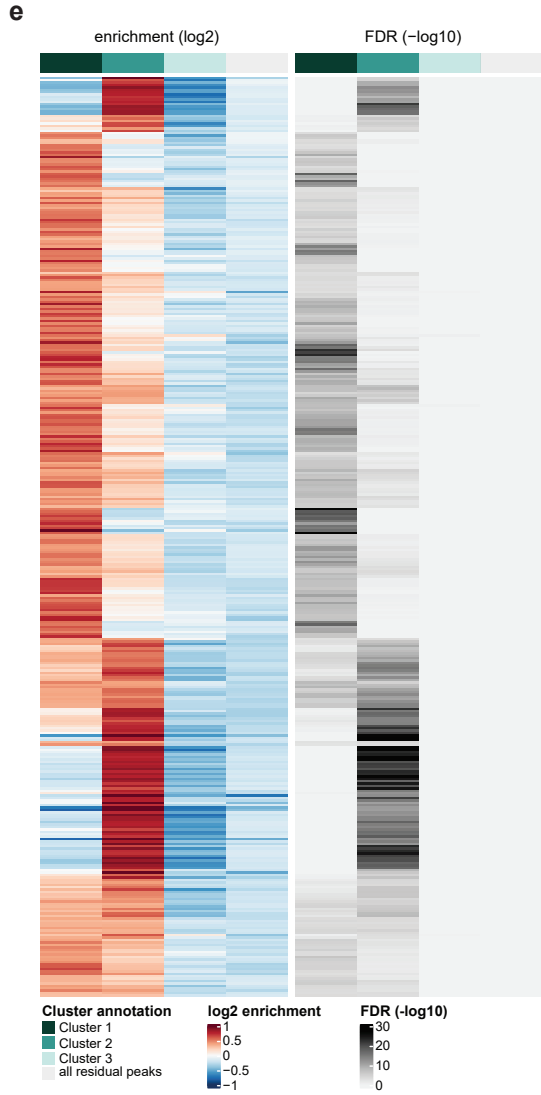
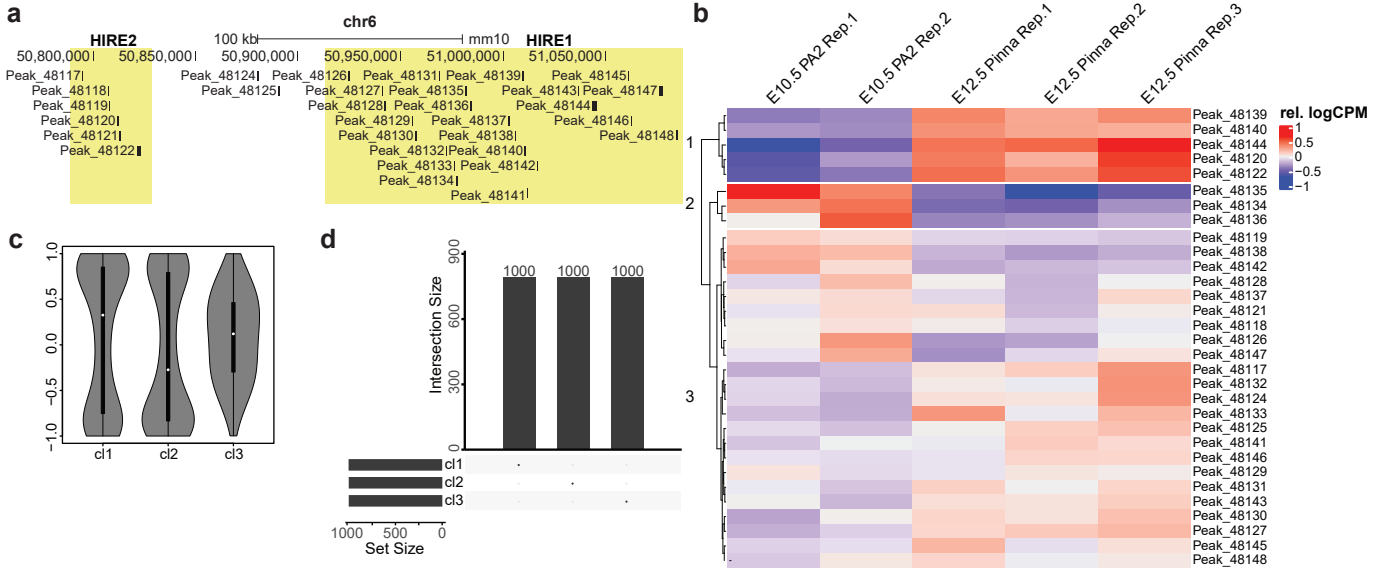
Supplementary Fig. 10 Morphological changes of second arch-derived skeletal and cartilaginous structures correlate to reduction of *Hoxa2* transcript levels.

(Left) *Hoxa2* transcript levels in E10.5 second pharyngeal arch (PA2) of *Hoxa2*^{EGFP/wt}, *HIRE2*^{del/del}, *HIRE2*^{del/wt};*Hoxa2*^{EGFP/wt}, *HIRE1*^{del/del} and *HIRE1*^{del/wt};*Hoxa2*^{EGFP/wt} mutant embryos, as compared to wild-type (WT), and obtained by qRT-PCR analysis (see suppl. fig. 9). (Right) For each genotype, the corresponding phenotype at PA2 cranial neural crest cells (CNCC)-derived skeletal and cartilaginous structures (blue) is represented. Note that no visible malformations are observed in *Hoxa2*^{EGFP/wt} fetuses.



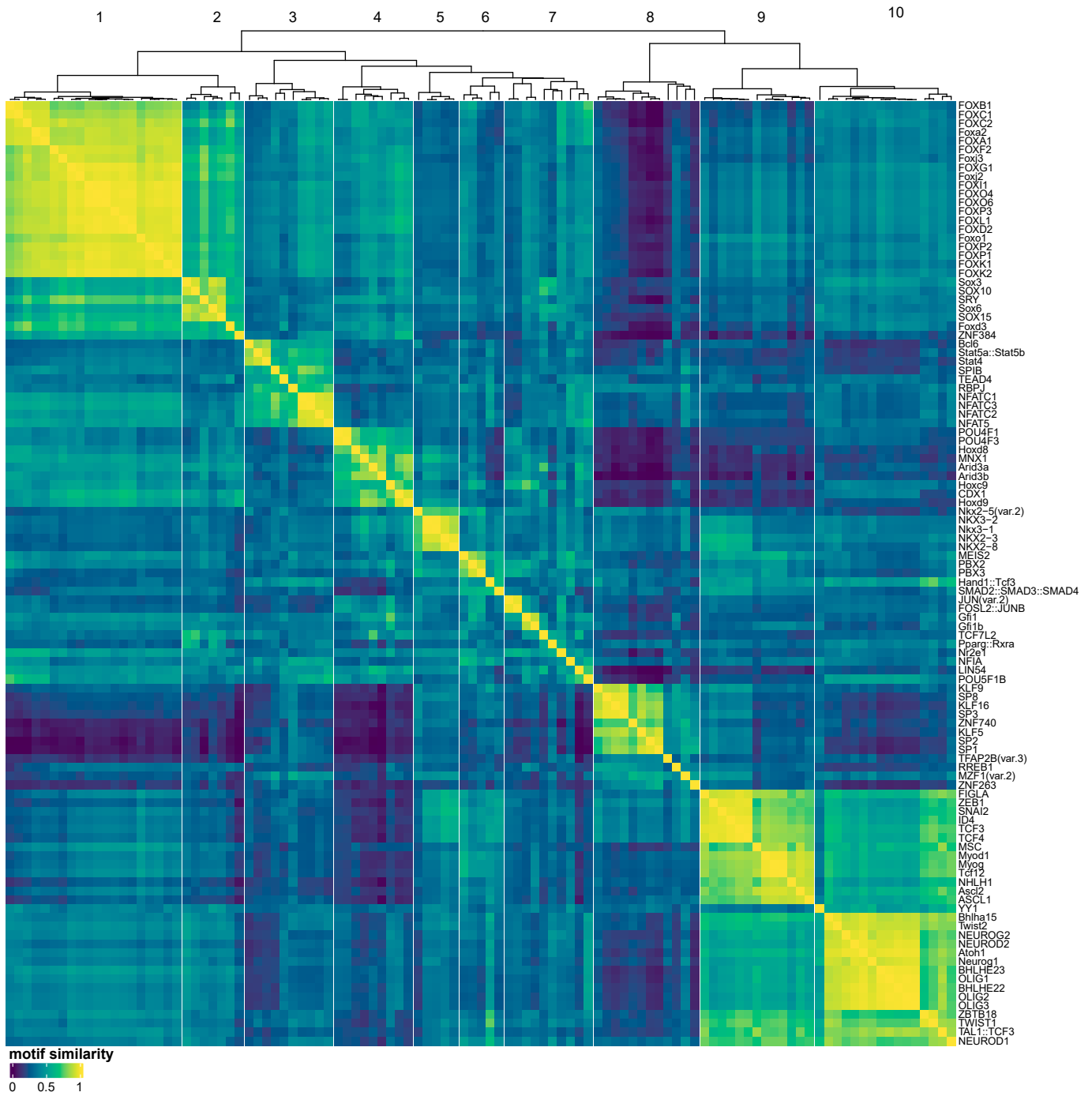
Supplementary Fig. 11 Effect of HIRE1 deletion on anterior *Hoxa* gene expression.

In-situ hybridization on sagittal sections of wild-type (WT) (**a, c**) and *HIRE1*^{del/del} (**b, d**) embryos at E10.5 using *Hoxa2* (**a, b**) and *Hoxa3* (**c, d**) antisense probes. **a-d** are representative images of serial sections from n=4 WT and n=4 *HIRE1*^{del/del} embryos. *Hoxa2* expression is undetectable in the second pharyngeal arch (PA2) and strongly reduced in the third pharyngeal arch (PA3) of *HIRE1*^{del/del} embryos (**b**) compared to WT (**a**). *Hoxa3* expression is strongly reduced in PA3 of *HIRE1*^{del/del} embryos (**d**) compared to WT (**c**). Whole mount *in-situ* hybridization on WT (**e, g, i, k**; n=5 embryos) and *HIRE1*^{del/del} (**f, h, j, l**; n=2 embryos) embryos at E8.5 using *Hoxa1* antisense probes. No differences in *Hoxa1* expression could be observed between WT and *HIRE1*^{del/del} embryos. Md represents the mandibular process. Scale bars represent 200 μ m.



Supplementary Fig.12 Transcription factor motif enrichment analysis.

a. Schematic representation of a genomic window, depicting the locations and names of 31 merged ATAC-seq peaks overlapping HIRE1 and HIRE2 from the second arch (PA2) derived CNCCs at E10.5 and from pinna derived CNCCs at E12.5. **b.** Clustering of ATAC-seq peaks overlapping HIRE1 and HIRE2 (chr6:50789172-51087888, mouse GRCm38/mm10) from PA2 derived CNCCs at E10.5 and from pinna derived CNCCs at E12.5 according to their relative accessibility levels at E10.5 and E12.5. Peaks were classified into three groups using hierarchical clustering. **c.** Violin plot depicting the distribution of Pearson's correlation coefficients of the ATAC-seq signal at all ATAC-seq peaks in E10.5 and E12.5 ($n = 133466$ ATAC-seq peaks) to the average accessibility profile of each of the three clusters. For clusters 1 and 2, the distribution of correlation coefficients was bimodal, corresponding to peaks that were highly correlated or anti-correlated with the profiles identified from HIRE2 to HIRE1 peaks. Plots extend from the data minima to the maxima; the white dot indicates the median, the box shows the interquartile range and whiskers extend to the most extreme data points. **d.** UpSet plot showing the top 1000 ATAC-seq peaks (ranked by their correlation coefficient) for each of the three clusters. There is no overlap between the top 1000 peaks for each cluster. **e.** Transcription factor motif enrichment heatmap. Shown are the \log_2 -fold enrichments and $-\log_{10}$ (FDR) values for all transcription factor motifs with an FDR less than 0.001 in any of the three clusters ($n = 382$). Each row depicts one transcription factor motif. While there are many significantly enriched motifs for clusters 1 and 2, no specific motifs are enriched in cluster 3 or all residual ATAC-seq peaks. **f.** Histogram of the average motif hit rates per kilobase (kb). The 31 ATAC-seq peaks overlapping HIRE1 and HIRE2 were scanned for motif hits using all transcription factor motifs with an FDR less than 0.001 in any of the three clusters as input (see Methods). To select motifs with a high frequency of predicted binding sites, the average motif hit rates per kb were calculated and clustered into two sets with k-means clustering ($k=2$). All motifs from the set with higher hit-rates ($n = 107$) were used in a motif similarity analysis. **g.** Transcription factor motif enrichment heatmap. Shown are the \log_2 -fold enrichments for all 107 transcription factor motifs with high hit-rates from **f** and an FDR less than 0.001 in any of the three clusters. The rows are ordered according to the motif similarity analysis shown in Supplementary Fig. 13. and yellow highlights indicate the different motif similarity clusters defined in Supplementary Fig. 13.



Supplementary Fig. 13 Motif similarity analysis. All 107 transcription factor motifs with high hit-rates defined in Supplementary Fig. 12f were used to perform a motif similarity analysis using *monaLisa*⁴. See Methods for details.

Supplementary Table 1 Genomic locations of the 2,232 identified CNCC super-enhancers

The table contains the genomic locations (mm10) of all super-enhancers identified in CNCCs at E10.5 based on their H3K27ac signal.

Supplementary Table 2 SEs interacting with genes encoding for transcription factor genes differentially expressed in CNCCs subpopulations at E10.5.

The table contains the genomic coordinates (mm10) of the 147 super-enhancers (SEs) with at least one significant interaction to an expressed positional transcription factor (TF) gene as depicted in figure 1, including the information to which TF each SE is linked. In addition, the mean strengths of all interactions from the positional TF promoter to restriction fragment(s) that overlap with a SE are given for each population. The last column indicates the number of the k-means cluster as shown on the left side of the heatmap. The rows follow the same order as the heatmap of figure 1. Note that the table contains 148 unique SE-promoter pairs (rows) because one SE is shared by 2 different promoters (*Twist2* and *Hes6*), and thus appears twice.

Supplementary Table 3 Differential gene expression analysis. The table provides on five separate sheets the differential gene expression analysis between second pharyngeal arch (PA2) derived CNCCs at E10.5 and pinna derived CNCCs at E12.5, between PA2 derived CNCCs at E10.5 and pinna derived CNCCs at E14.5, between pinna derived CNCCs at E12.5 and pinna derived CNCCs at E14.5, between PA2 cells from *HIRE1^{del/del}* and wild-type embryos at E10.5, and between PA3 cells from *HIRE1^{del/del}* and wild-type embryos at E10.5. Each sheet contains the following information in the columns: Entrez gene ID, gene width, ENSEMBL gene ID, gene symbol, log₂ fold change (column name logFC), mean expression level (log₂ counts per million, column name logCPM), quasi-likelihood F-test score (F), raw P value (PValue), and false discovery rate (FDR). P-values and FDR were calculated by *edgeR*⁵. See Methods for details on statistical analysis.

Supplementary Table 4 Transcription factor motif enrichment analysis. The table contains the results of the transcription factor motif analysis. Each row stands for one of the motifs of the JASPAR2018 database. The columns contain information about the motif ID, motif name, motif family, and motif class, as well as the log₂-fold enrichments and -log₁₀(FDR) values for each of the ATAC-seq peak clusters (as defined in Supplementary Fig. 12b-d) and all residual peaks. P-values and FDR were calculated by *monaLisa*⁴. See Methods for details.

Supplementary Table 5 Annotation of merged ATAC-seq peaks from PA2 derived CNCCs at E10.5 and pinna derived CNCCs at E12.5 overlapping the region spanning from HIRE2 to HIRE1 with predicted motifs. The number of occurrences of all transcription factor motifs with an FDR less than 0.001 in any of the three cluster profiles (Suppl. Fig. 8D) was predicted within each of the ATAC-seq peaks from second pharyngeal arch (PA2) derived cranial neural crest cells (CNCCs) at E10.5 and pinna derived CNCCs at E12.5, which overlap the region spanning from HIRE2 to HIRE1 (chr6:50789172-51087888, mouse GRCm38/mm10). See Methods for details.

Supplementary Table 6 List of selected representative motifs clustered by motif similarity analysis. Each row in the table stands for one of the motifs identified in the motif enrichment analysis (see Methods). The columns contain information about the motif similarity cluster that a motif was assigned to in Supplementary Fig. 13, the motif ID, motif name, motif family and motif class.

Supplementary Table 7 Sample information. The type of experiments performed on the different cell types, the protocols used for the libraries, the sample names, and for previously published datasets the GEO accession number are indicated.

Supplementary References

1. Donaldson IJ, *et al.* Genome-wide occupancy links *Hoxa2* to Wnt-beta-catenin signaling in mouse embryonic development. *Nucleic Acids Res* **40**, 3990-4001 (2012).
2. Amin S, *et al.* *Hoxa2* selectively enhances Meis binding to change a branchial arch ground state. *Dev Cell* **32**, 265-277 (2015).
3. Rijli FM, Mark M, Lakkaraju S, Dierich A, Dolle P, Chambon P. A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. *Cell* **75**, 1333-1349 (1993).
4. Machlab D, Burger L, Sonesson C, Rijli FM, Schubeler D, Stadler MB. monaLisa: an R/Bioconductor package for identifying regulatory motifs. *Bioinformatics*, (2022).
5. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010).